

# Unique subunit packing in mycobacterial nanoRNase leads to alternate substrate recognitions in DHH phosphodiesterases

Rajpal Srivastav<sup>1,†</sup>, Dilip Kumar<sup>1,†</sup>, Amit Grover<sup>1</sup>, Ajit Singh<sup>1</sup>, Babu A. Manjasetty<sup>2,3</sup>, Rakesh Sharma<sup>1,\*</sup> and Bhupesh Taneja<sup>1,\*</sup>

<sup>1</sup>CSIR-IGIB, Institute of Genomics and Integrative Biology, South Campus Mathura Road, New Delhi 110020, India,

<sup>2</sup>European Molecular Biology Laboratory, Grenoble Outstation, 6 rue Jules Horowitz, Grenoble 38042, France and

<sup>3</sup>Unit for Virus Host-Cell Interactions, University Grenoble Alpes-EMBL-CNRS, 6 rue Jules Horowitz, Grenoble 38042, France

Received October 17, 2013; Revised May 02, 2014; Accepted May 2, 2014

## ABSTRACT

DHH superfamily includes RecJ, nanoRNases (NrnA), cyclic nucleotide phosphodiesterases and pyrophosphatases. In this study, we have carried out *in vitro* and *in vivo* investigations on the bifunctional NrnA-homolog from *Mycobacterium smegmatis*, MSMEG\_2630. The crystal structure of MSMEG\_2630 was determined to 2.2-Å resolution and reveals a dimer consisting of two identical subunits with each subunit folding into an N-terminal DHH domain and a C-terminal DHHA1 domain. The overall structure and fold of the individual domains is similar to other members of DHH superfamily. However, MSMEG\_2630 exhibits a distinct quaternary structure in contrast to other DHH phosphodiesterases. This novel mode of subunit packing and variations in the linker region that enlarge the domain interface are responsible for alternate recognitions of substrates in the bifunctional nanoRNases. MSMEG\_2630 exhibits bifunctional 3'-5' exonuclease [on both deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) substrates] as well as CysQ-like phosphatase activity (on pAp) *in vitro* with a preference for nanoRNA substrates over single-stranded DNA of equivalent lengths. A transposon disruption of MSMEG\_2630 in *M. smegmatis* causes growth impairment in the presence of various DNA-damaging agents. Further phylogenetic analysis and genome organization reveals clustering of bacterial nanoRNases into two distinct

subfamilies with possible role in transcriptional and translational events during stress.

## INTRODUCTION

NanoRNA (2–5 nucleotides) is generated in several cellular processes such as messenger ribonucleic acid (mRNA) degradation, abortive transcription initiation, RNA cleavage during transcription elongation and cyclic nucleotide degradation (1,2). NanoRNA has the ability to affect the physiological state of the cell and must be degraded completely. Accumulation of nanoRNA leads to altered priming of transcription initiation and hence changes in expression levels of various genes that could be lethal for the normal functioning of cell (3). These small-sized nanoRNAs are processed by special enzymes in bacterial cells. In *Escherichia coli*, the degradation of nanoRNA is carried out by a highly conserved enzyme oligoribonuclease, ORN, deletion of which is detrimental toward growth (4). Despite its important physiological role, ORN is not present in all bacteria. ORN is replaced by a redundant nanoRNase family, first identified in *Bacillus subtilis*, homologs of which are also found in several other bacterial groups (5).

Several classes of nanoRNases, NrnA and NrnB in *B. subtilis* and NrnC in *Bartonella birtlesii*, have been identified (5–7) with distinct sequence features. Apart from conserved sequence motifs necessary for activity, the sequences of nanoRNase families vary significantly. Consequently, sequence similarity of NrnC with NrnB and NrnA is only 16% and 11%, respectively (7). While NrnAs are bifunctional enzymes with nanoRNase and CysQ activity (313 residues in *B. subtilis* NrnA), only a nanoRNase activity has been identified in NrnB (399 residues in *B. subtilis*)

\*To whom correspondence should be addressed: Tel: +91 11 29879324; Fax: +91 11 27667471; Email: btaneja@igib.res.in

Correspondence may also be addressed to Rakesh Sharma. Tel: +91 11 29879222; Fax: +91 11 27667471; Email: rsharma@igib.res.in

<sup>†</sup>The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Present address:

Amit Grover, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK.

and NrnC (181 residues in *B. birtlesii*). NrnA and NrnB belong to DHH superfamily of phosphodiesterases, which also include RecJ-like exonucleases, inorganic exopolyphosphatases, family II pyrophosphatases, while NrnC belongs to DEDD family of exonucleases (7). The phylogenetic origin of the multiple classes of nanoRNases and their physiological roles remain unclear. The recent crystal structure of NrnA-homolog from *Bacteroides fragilis* (8) at 2.95-Å resolution in ligand-bound form (with GMP) shows interactions with several residues in the C-terminal domain (CTD). However, the molecular basis of selective recognition of nanoRNA by nanoRNases over pyrophosphates is still not clear. The crystal structure of ORN from *Xanthomonas campestris* (9) is also available but exhibits no structural similarity with *B. fragilis* NrnA and the two proteins contain completely different structural folds. This large amount of variation in sequence as well as in structure among various ORN and nanoRNase families suggests independent evolutionary origin of these enzymes despite related functions.

NrnA is a member of DHH phosphodiesterase superfamily widely distributed across several bacterial groups that do not harbor *orn* homologs (5,10). Actinobacteria is the only bacterial group identified so far, where both ORN and NrnA are present (5). The presence of both ORN and nanoRNase having similar kind of functions suggests that RNA processing pathway in Actinobacteria may differ substantially as compared to other bacteria and needs to be studied in detail. Recently, Rv2837c of *Mycobacterium tuberculosis*, a 336 residue protein, has been reported to be the bifunctional NrnA counterpart that can complement *E. coli orn*<sup>-</sup> or *cysQ*<sup>-</sup> strains (10).

Here, we identify MSMEG\_2630 of *Mycobacterium smegmatis* as an ortholog of Rv2837c and investigate the role of this nanoRNase through a combination of structural, *in vivo* and *in vitro* biochemical studies. Like previously characterized NrnA members, MSMEG\_2630 exhibits bifunctional 3'-5' exoribonuclease on RNA substrates as well as CysQ-like pAp phosphatase activity *in vitro*. In addition, we find MSMEG\_2630 also degrades single-stranded deoxyribonucleic acid (ssDNA) in 3'-5' direction *in vitro*, although nanoRNA substrates appear to be the preferred substrate. A transposon mutant of MSMEG\_2630 in *M. smegmatis* is not lethal but shows growth impairment in the presence of various DNA-damaging agents. We have also determined the crystal structure of this mycobacterial nanoRNase to 2.2-Å resolution and investigate its structural features to atomic details. The structure of *M. smegmatis* nanoRNase reveals a dimer consisting of two identical subunits. However, the subunit packing is completely different from other DHH phosphodiesterase structures, suggesting that mycobacteria and other microbes harboring nanoRNases utilize alternate packaging of the same well-characterized domains to obtain specific functional requirements in RNA processing (ORN or nanoribonuclease), DNA repair (RecJ exonucleases) or phosphatase (pyrophosphatases) activity. Further sequence analysis and difference in genomic location of MSMEG\_2630 reveal that nanoRNases of different bacteria are present in two functionally distinct subfamilies. Our findings indicate that *M.*

*smegmatis* MSMEG\_2630 belongs to a unique group of nanoRNases with possible role in transcriptional and translational events during stress.

## MATERIALS AND METHODS

### Cloning, expression and purification of MSMEG\_2630

The coding region of MSMEG\_2630 was polymerase chain reaction (PCR) amplified from genomic DNA of *M. smegmatis* mc<sup>2</sup>155 using forward and reverse primers (Supplementary Table S1). The forward primer introduced a NcoI site at the start codon at the 5' end, whereas the reverse primer introduced a HindIII site at the 3' end. The PCR products were digested with NcoI and HindIII restriction enzymes and cloned in a similarly digested pET-28a (Novagen) vector to give the expression plasmid pRPS25. pRPS25 encodes full-length MSMEG\_2630 fused to a C-terminal 6 × His-tag separated by a linker peptide. pRPS25 was transformed into *E. coli* BL21(DE3) for expression. The transformed cells were grown at 37°C to A<sub>600</sub> ~ 0.5 in Luria Bertani (LB) medium and the culture was induced for protein expression with 0.1-mM IPTG (isopropyl β-D-1-thiogalactopyranoside) for 12 h with constant shaking at 18°C. Subsequently, the cells were harvested at 8000 g for 10 min at 4°C. The pellet was resuspended in sonication buffer (50-mM Tris-HCl, pH 8.0) and lysed by sonication. The cell debris was removed by centrifugation and the His-tagged protein was purified over nickel-nitrilotriacetate (Ni-NTA) agarose beads (Invitrogen) as per manufacturer's protocol. The protein was eluted with 250-mM imidazole (in sodium phosphate buffer, pH 6.0, and 500-mM NaCl). Fractions containing significant amounts of purified protein, as examined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (10%), were pooled. Recombinant His-tagged MSMEG\_2630 thus purified contains full-length MSMEG\_2630 polypeptide with 13 additional residues at the C-terminus comprising the 6 × His-tag and the linker peptide and was designated 'rMSMEG\_2630-A'. Buffer exchange was carried out by gel filtration and the protein was finally suspended into 20-mM Tris-HCl, pH 7.5. Protein purity was assessed on 10% SDS-PAGE and concentration was estimated using Bicinchoninic acid (Sigma). All activity assays described below were carried out using rMSMEG\_2630-A, unless indicated otherwise.

### Construct for crystallization experiments

The coding region of MSMEG\_2630 was PCR amplified from pRPS25 using forward (Supplementary Table S1) and reverse primers and cloned in pET-28-His10-Smt3 vector at BamHI and XhoI sites to give expression plasmid pRPS26. pRPS26 encodes full-length MSMEG\_2630 fused to an N-terminal His10-Smt3 tag. The recombinant construct was transformed into *E. coli* BL21(DE3) and the tagged protein was purified as described previously (11). The eluted protein was incubated overnight with Smt3-specific protease Ulp1 (Ulp1: protein ratio of 1:500) at 4°C to cleave the His10-Smt3 tag. Tag-free protein was recovered in flow through by passage over Ni-NTA agarose column (Qiagen). Removal of the Smt3 tag left one serine residue at the N-terminus

of the purified protein. The recombinant protein thus purified was dialyzed against low salt buffer (20-mM Tris-HCl, pH 8.5, 25-mM NaCl, 5% glycerol, 2-mM MgCl<sub>2</sub>). Further downstream purification over an anion exchange column (MonoQ, GE Healthcare) followed by a size-exclusion step over Superdex 75 column (GE Healthcare) was necessary to remove other contaminants and obtain purified protein, designated 'rMSMEG\_2630-B' for reproducible crystallizations. The purity of rMSMEG\_2630-B was checked on a 12% SDS-PAGE and the protein was concentrated to 30 mg/ml using Amicon ultra 3-kDa molecular weight cutoff filter units and stored at 4°C until further use.

Constructs expressing proteins corresponding to residues 1–188, 1–211 and 212–340 of MSMEG\_2630 fused to an N-terminal His10-Smt3 tag were similarly prepared and designated MS-NTD-A, MS-NTD-B and MS-CTD, respectively. The proteins were purified as for full-length native protein described above. Activity assays with these constructs were carried out for the respective proteins in the presence of the in-frame fused tag as indicated. Full-length MSMEG\_2630 purified along with the in-frame fused tag (and designated Ntag\_MSMEG\_2630) was used as controls in the activity assays to confirm that the tag did not affect activity.

#### Expression and purification of selenomethionine derivatized MSMEG\_2630

pRPS26 was transformed into *E. coli* BL21(DE3) and grown on selenomethionine (Se-Met) minimal media (Molecular Dimensions) as described by the manufacturer's protocol. Se-Met derivatized rMSMEG\_2630-B was purified as native protein, except 2-mM TCEP was added to all buffers and the protein concentrated to 25 mg/ml before crystallizations.

#### Protein crystallization, data collection and structure refinement

Native rMSMEG\_2630-B (30 mg/ml) (in 20-mM Tris-HCl, pH 8.5, 25-mM NaCl, 5% glycerol, 2-mM MgCl<sub>2</sub>) was crystallized by sitting drop vapor diffusion against crystallization buffer (100-mM HEPES, pH 7.0, 200-mM LiCl, and 26% PEG 6000) at 24°C as well as 10°C. Needle-shaped crystals were obtained in 4–5 days at both temperatures. The crystals were further refined to larger diffraction quality crystals by microseeding. For microseeding, crystals were pooled from eight to ten crystallization drops and washed multiple times with the crystallization buffer. Protein crystals were then crushed with the help of crystallization microtools (Hampton Research) and resuspended in 50 µl of crystallization buffer and used as seed stock. Serial dilutions of the seed stock (1:100, 1:200, 1:400 and 1:1000) were prepared and used for crystallization. Large rod-shaped crystals were obtained after incubation for another 5–6 days with the seed dilutions of 1:400 and 1:1000. Se-Met derivatized rMSMEG\_2630-B was also crystallized by microseeding using the same native seeds in crystallizations. The crystals were cryoprotected in a solution containing 15% glycerol in addition to the crystallization buffer and flash frozen into liquid nitrogen until data collection.

Se-Met-incorporated crystal was subjected to a fluorescence scan at the Se *K* edge on beamline BM14 under attenuation at the European Synchrotron Radiation Facility, Grenoble, and the data were analyzed with CHOOCH (12). Diffraction data were collected on the same crystal at Se-peak wavelength of 0.9795 Å. The data were processed using MOSFLM, data quality assessed (14) and scaled using SCALA within the CCP4 package (15,16). The crystals belong to the space group P1 2<sub>1</sub> 1 and contain two molecules in the crystallographic asymmetric unit. Details of data collection and data processing are summarized in Table 1.

#### X-ray structure determination

The structure of rMSMEG\_2630-B was solved by a single-wavelength anomalous dispersion (SAD) phasing method with data collected at the peak energy of X-ray absorption spectrum of selenium, using the automated structure determination platform of AutoRickshaw (17). Further model building on the solution obtained was carried out in Coot (18) and refined in Refmac5 (19) and Buster 2.10.0 (20). Non-crystallographic symmetry restraints were released in order to capture domain movements in the different subunits and translation/libration/screw (TLS) refinement (21) was performed during final stages with three TLS groups (residues 22–189, 190–240 and 241–340) for each subunit. Details of model building and refinement statistics are given in Table 1. The final coordinates have been deposited with PDB with the accession code 4LS9.

#### N-terminal sequencing

N-terminal sequencing of rMSMEG\_2630-B was carried out on a Procise 491A protein sequencer (Applied Biosystems) by the Edman degradation method. The crystals were pooled from three to four crystallization drops and washed thoroughly with the crystallization buffer to remove any residual uncrystallized protein in the drop. The crystals were then crushed with the help of crystallization microtools and the resultant protein resuspended in 30 µl of solubilization buffer (20-mM Tris-HCl, pH 8.5, 25-mM NaCl, 2-mM MgCl<sub>2</sub> and 5% glycerol). The protein was then applied on a 15% SDS-PAGE and transferred to polyvinylidene difluoride (PVDF) membrane for Edman sequencing.

#### Transposon mutagenesis and complementation

A transposon mutant, L154, having disruption in *MSMEG\_2630* in *M. smegmatis* mc<sup>2</sup>155 was isolated for genetic studies from the existing mini-mariner-disrupted transposon mutant library (22). The transposon mutant was confirmed by Southern blotting and the location of transposon insertion was identified by sequencing the transposon flanking region.

*MSMEG\_2630* or *rv2837c* was PCR amplified from genomic DNA of *M. smegmatis* mc<sup>2</sup>155 or *M. tuberculosis* H37Rv, respectively, using respective forward and reverse primers (Supplementary Table S1). The PCR products were digested with PstI and HindIII restriction enzymes and cloned in *E. coli*-*Mycobacterium* shuttle vector, pSMT3 (23), to create pRPS16 and pRPS31, respectively. L154 cells

**Table 1.** Data collection and model refinement statistics

<b>Crystal and data collection</b>	
Space group	$P1\ 2_1\ 1$
Cell dimensions	$a = 65.87\ \text{\AA}, b = 86.10\ \text{\AA}, c = 66.03\ \text{\AA}, a = 90^\circ, b = 118.37, \gamma = 90^\circ$
Resolution ( $\text{\AA}$ ) <sup>a</sup>	34.59–2.20
Number of unique reflections	32833 (4744)
Completeness (%)	99.5 (99.2)
Multiplicity	7.8 (7.6)
$R_{\text{meas}}$ (%) <sup>b</sup>	11.1 (40.8)
$\langle\langle I \rangle\rangle / \sigma(\langle I \rangle)$ <sup>c</sup>	12.9 (5.3)
<b>Refinement</b>	
Resolution ( $\text{\AA}$ )	19.38–2.20
Number of reflections working/test	32831/1666
$R_{\text{work}}$ (%) <sup>d</sup>	0.176 (0.183)
$R_{\text{free}}$ (%) <sup>e</sup>	0.225 (0.243)
Total protein residues/atoms <sup>f</sup>	638/4779
Total water molecules	268
Wilson B factor ( $\text{\AA}^2$ )	31.7
<b>Average B factor (<math>\text{\AA}^2</math>)</b>	
Protein atoms (chain)	39.0
Water molecules	39.3
<b>rms deviations from ideal</b>	
Bond lengths ( $\text{\AA}$ )	0.01
Bond angles ( $^\circ$ )	1.1
<b>Ramachandran plot</b>	
Most favored regions (%)	97.9
Disallowed regions (%)	0.0

<sup>a</sup>Numbers in parentheses correspond to the highest resolution shell.

<sup>b</sup> $R_{\text{meas}}$  as described in (14).

<sup>c</sup> $\langle\langle I \rangle\rangle / \sigma(\langle I \rangle) = \text{mean } I_h \text{ over the standard deviation of the mean } I_h \text{ averaged over all reflections in a resolution shell.}$

<sup>d</sup> $R_{\text{work}} = \sum \|F_o\| - |F_c| / \sum |F_o|$ , where  $|F_o|$  is the observed structure factor amplitude and  $|F_c|$  is the calculated structure factor amplitude.

<sup>e</sup> $R_{\text{free}}$ :  $R_{\text{factor}}$  based on 5% of the data excluded from refinement.

<sup>f</sup>Total number of protein atoms, including those in alternate conformations.

were transformed with pRPS16 or pRPS31 by electroporation. The resultant transformants were used for complementation and growth experiments described below.

### Complementation and growth of *M. smegmatis* mutant L154

Effect of deletion of *MSMEG\_2630* on growth in *M. smegmatis* mutant L154 was monitored by OD<sub>600</sub>. Briefly, 1:500 dilutions of freshly growing log-phase *M. smegmatis* mc<sup>2</sup> 155 (WT)/pSMT3, L154/pSMT3 and L154/pRPS16 cells were inoculated separately into 50-ml LB medium at 37°C. Aliquots of cultures were withdrawn after intervals of 12 h and growth monitored by OD<sub>600</sub>. To study the effect of oxidative stress and DNA-damaging agents, minimum inhibitory concentration (MIC) was estimated for various extrinsic agents. For MIC estimation, *M. smegmatis* mc<sup>2</sup> 155 (WT)/pSMT3, L154/pSMT3 and L154/pRPS16 cells were grown in Middlebrook 7H9 liquid medium (supplemented with 0.05% Tween-80 and 0.2% glycerol) in the presence of increasing concentrations of H<sub>2</sub>O<sub>2</sub>, menadione, mitomycin C, ciprofloxacin, nalidixic acid, zeocin, hydroxyurea, ethyl methanesulfonate, methyl methanesulfonate and ethidium bromide (EtBr). The cell growth was monitored after 48 h by OD<sub>600</sub>. MIC for DNA intercalating agent, EtBr, was also estimated as above on L154/pRPS31.

### Reverse transcriptase PCR reaction

Total RNA from wild-type (WT) *M. smegmatis* was isolated using RNeasy kit (Qiagen). First strand complementary DNA (cDNA) synthesis was carried out using 1–2- $\mu\text{g}$  total RNA and random hexamers as primers (Omni-script kit, Qiagen). A 15- $\mu\text{l}$  aliquot of the cDNA synthesis reaction was amplified with primers For-RT1 and Rev-RT1 for *MSMEG\_2628* and *MSMEG\_2629* intergenic region, For-RT2 and Rev-RT2 for *MSMEG\_2629* and *MSMEG\_2630* intergenic region and For-RT3 and Rev-RT3 for *MSMEG\_2630* and *MSMEG\_2631* intergenic region. The oligonucleotide sequences are shown in Supplementary Table S1. For each sample, a reverse transcriptase (RT) negative control was also performed to rule out DNA contamination. RT-PCR positive control was carried out for sigma factor A, *sigA*.

### Enzyme activity assays

Activity assays on nano-oligonucleotides were performed using custom-made RNA and DNA 5-mers labeled at their 5'-end with sulfoindocyanine succinimidyl ester cyanine 5 (Cy5) as described previously (5). A 10- $\mu\text{l}$  reaction mix containing 12.5- $\mu\text{M}$  Cy5-labeled oligos as substrate and 4- $\mu\text{M}$  rMSMEG\_2630-A in 50-mM HEPES, pH 7.5, 5-mM MnCl<sub>2</sub>, was incubated at 37°C for indicated time intervals in

different vials. The reaction was stopped by heating at 95°C for 5 min and the reaction mixture was stored at -20°C. For analysis of the reaction products, the samples were applied on a 22% polyacrylamide gel (containing 8-M urea) and run in 1× Tris/Borate/ethylenediaminetetraacetic acid (EDTA) electrophoresis buffer. Fluorescent oligos were visualized using a Typhoon FLA 7000 (GE Healthcare) with excitation and emission at 635 nm and 670 nm, respectively. Quantification of the data was done by defining the total amount of fluorescence measured in the substrate (5-mer) and the reaction products (4-mer, 3-mer, 2-mer and 1-mer for ribonucleotides or 4-mer, 3-mer and 2-mer for DNA oligos) for each time point as 100% and expressing the amount of each reaction product as fraction of the total.

Exonuclease activity of rMSMEG\_2630-A was also analyzed with long synthetic radiolabeled oligonucleotides as substrates. 49-mer DNA oligos (Supplementary Table S1) 5'-end-labeled with [ $\gamma$ -<sup>32</sup>P]ATP by T4 polynucleotide kinase or 3'-end-labeled with [ $\alpha$ -<sup>32</sup>P]dATP by terminal deoxynucleotide transferase and a similarly 5'-end-labeled 35-mer ribonucleotide (Supplementary Table S1) were used as substrates. Synthetic high performance liquid chromatography (HPLC)-purified 5'-end-labeled oligonucleotides of lengths 5-mer, 10-mer, 20-mer, 30-mer, 40-mer and 50-mer were used as markers to analyze the length of degradation products. The exonuclease activity assay was carried out by incubating (280 nM) of 5'-end-labeled ssDNA or RNA with 2.8  $\mu$ M of purified rMSMEG\_2630-A in 20-mM Tris-HCl, pH 7.5, 100-mM NaCl, 10-mM MgCl<sub>2</sub> and 1-mM dithiothreitol (DTT) at 37°C for different time periods. The exonuclease activity was terminated by addition of 50-mM EDTA and heating the reaction mix at 95°C for 15 min. The products were analyzed on 15% polyacrylamide gel (containing 8-M urea) and visualized on a Phosphor-Imager (Bio-Rad).

The phosphatase activity of rMSMEG\_2630-A was estimated with 3'-phosphoadenosine phosphate (pAp), sodium pyrophosphate and sodium triphosphate (Sigma Aldrich) as substrates. Briefly, 1-mM substrate (in 100-mM Tris-HCl, pH 8.3, 100-mM KCl and 5-mM MgCl<sub>2</sub>) was incubated with 250 nM of rMSMEG\_2630-A at 37°C for 2 h. The release of inorganic phosphate (Pi), due to phosphatase activity of the protein, was estimated by malachite green inorganic phosphate assay and monitored colorimetrically at  $\lambda_{610}$ .

Phosphodiesterase activity of rMSMEG\_2630-A was tested using bis-p-nitrophenol phosphate (bis-pNpp) as a substrate. The standard reaction included 1-mM substrate, 100-mM Tris-HCl, pH 8.3, 100-mM KCl, 5-mM MgCl<sub>2</sub> and 250 nM of purified protein. The reaction was incubated at 37°C for 30 min and the release of p-nitrophenol as product was monitored colorimetrically at  $\lambda_{405}$ . Activity assays of mutants (described below) and N-terminal domain (NTD) constructs were similarly estimated.

### Mutagenesis of conserved residues

A conserved aspartate residue (D134) and a conserved histidine residue (H316) in pRPS25 were substituted with alanine by site-directed mutagenesis (Quik-change XL II Stratagene) as per manufacturer's instructions. The presence of the desired mutation was confirmed by DNA se-

quencing. Mutant plasmids thus obtained, pD134A and pH316A, were transformed into *E. coli* BL21(DE3). The mutant proteins MS-D134A and MS-H316A were expressed and purified as the WT rMSMEG\_2630-A.

### Circular dichroism spectroscopy measurements

Far-ultraviolet circular dichroism (CD) spectra were collected on a Jasco J815 CD spectrometer in a quartz cuvette with a path length of 0.1 cm at room temperature in the range of 190–250 nm. CD measurements were carried out in 5-mM Tris-HCl, pH 8.5, 5-mM NaCl, 0.25-mM MgCl<sub>2</sub>, 0.125-mM DTT and 1.25% glycerol at protein concentrations of 0.6 mg/ml for MS-D134A and MS-H316A, 0.8 mg/ml for WT rMSMEG\_2630-A, 0.25 mg/ml for MS-NTD-A, MS-NTD-B and 0.3 mg/ml for Ntag\_MSMEG\_2630. Each spectrum was recorded as an average of four scans. In all experiments, contributions of the buffer to the spectra were subtracted, and mean residue ellipticities were determined before plotting the spectra.

### Sequence analysis

Sequence homologs, as bidirectional best hits from Basic Local Alignment Search Tool (BLAST) for the previously characterized *B. subtilis* NrnA, from different bacterial groups were obtained from NCBI (<http://ncbi.nlm.nih.gov/protein>) (10). A phylogenetic tree was constructed using MEGA5.1 (24) with 1000 cycles of bootstrap for selected sequences. Genetic location and neighboring genes were identified from GenBank.

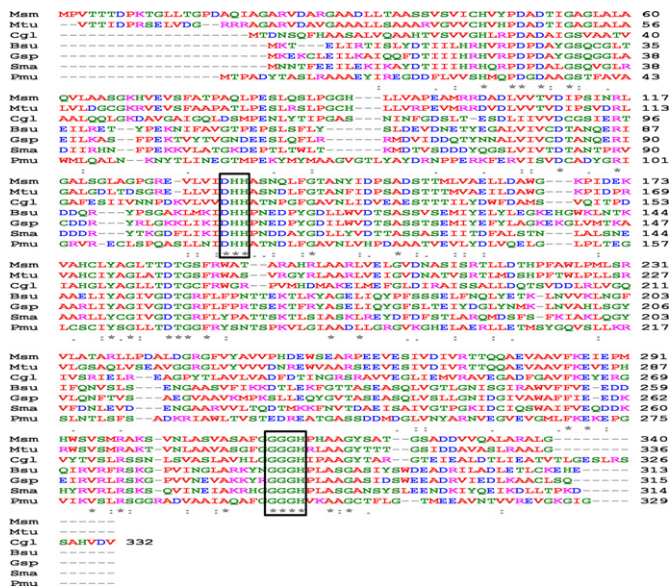
## RESULTS

### Sequence/phylogenetic analysis of NrnA

A BLAST search with Rv2837c, described earlier as a nanoRNase in *M. tuberculosis* (10), identified MSMEG\_2630 as the ortholog in *M. smegmatis* mc<sup>2</sup>155 genome with an *e*-value of 10<sup>-168</sup>. MSMEG\_2630 consists of 340 amino acids and when compared with Rv2837c has sequence identity and similarity of 73% and 83%, respectively. MSMEG\_2630 is the best bidirectional BLAST hit of NrnA in *M. smegmatis* genome, confirming it as the NrnA-ortholog in *M. smegmatis*. The identity and similarity of MSMEG\_2630 with NrnA of *B. subtilis* are 25% and 44%, respectively.

A sequence alignment of nanoRNases from various organisms revealed the presence of several conserved sequence features among all nanoRNase sequences examined, suggesting that these residues play important roles in its function (Figure 1). A conserved <sup>134</sup>DHH<sup>136</sup> motif in the NTD and <sup>313</sup>GGGH<sup>316</sup> motif in the CTD, which are characteristic features of DHH/DHHA1 subfamily, are also conserved in MSMEG\_2630 (Figure 1). In addition, actinobacterial NrnA sequences possess a short stretch of highly variable residues at their N-terminus not observed in any other sequences (Figure 1).

An examination of genomic loci of *nrnA* and *MSMEG\_2630*, however, revealed striking differences in patterns of neighboring genes of nanoRNase homologs. Translation initiation factor 2 (*infB*) (*MSMEG\_2628*) and



**Figure 1.** Multiple sequence alignment of NanoRNases. Multiple sequence alignment of selected nanoRNase sequences (Msm: *M. smegmatis* MSMEG\_2630; Mtu: *M. tuberculosis* Rv2837c; Cgl: *Corynebacterium glutamicum* WP\_003861692; Bsu: *B. subtilis* NrnA (Ytqj); Gsp: *Geobacillus* sp. YP\_001126761; Sma: *Streptococcus macedonicus* YP\_005094369 and Pmu: *Paenibacillus mucilaginosus* YP\_004644163). The conserved DHH motif in the NTD and GGGH motif in the C-terminal DHHA1 domain are shown in rectangular boxes. The mycobacterial sequences contain a unique stretch of residues at the N-terminus (see the text for details).

ribosome-binding factor A (*rbfA*) (*MSMEG\_2629*) are upstream neighboring genes of *MSMEG\_2630* and appear to share a transcription start site with *MSMEG\_2630* (Figure 2a). A gene encoding a multidrug and toxic compound extrusion protein (*MATE*)/DNA-damage-inducible protein F (*dinF*) (*MSMEG\_2631*) present downstream of *MSMEG\_2630* also appears to be transcribed from the same transcription start site. While *MATE* is a neighboring gene of mycobacterial sequences only, *nusA*, a transcriptional regulator present further downstream of *MSMEG\_2630*, is also commonly found in this locus in all the examined sequences. On the other hand, *dnaE*, acetyl CoA carboxylase (*accA*), 6-phosphofructokinase (*pfkA*) and pyruvate kinase (*pyk*) are the frequently observed neighbors of *nrrnA* of *B. subtilis* (Figure 2b). A phylogenetic tree constructed for nanoRNase sequences showed clear clustering into two major clades, with *infB*, *rbfA* and *nusA* as neighboring genes in one clade and *dnaE*, *accA*, *pfkA* and *pyk* as neighboring genes in the other (Figure 2c). *MSMEG\_2630*, like Rv2837c, is distinct from *B. subtilis* NrnA and is part of a separate clade (Figure 2c). We propose that members of this separate clade be designated NrnA'. The nanoRNase activity of *MSMEG\_2630*, a member of NrnA', may have specific roles in regulation of RNA-mediated transcriptional and/or translational-related activities and was investigated further during the course of this work.

**MSMEG\_2630 is transcribed with *infB* and *rbfA* as a single operon**

The genetic organization of *MSMEG\_2630* in *M. smegmatis* mc<sup>2</sup>155 predicts that *MSMEG\_2630* locus would be transcribed as an operon with additional members. To confirm the transcription of *MSMEG\_2630* as a single operon with these members and explore the possible physiological role of *MSMEG\_2630*, the genic boundaries were identified by RT-PCR. The amplification with intergenic primer pairs for each junction between two genes demonstrated that expression of *MSMEG\_2630* is transcriptionally linked to two upstream (*MSMEG\_2628*, *infB* and *MSMEG\_2629*, *rbfA*) and one downstream genes (*MSMEG\_2631*, *dinF*) (Figure 2d) confirming that these genes are transcribed as a polycistronic mRNA. This polycistronic message encoding *MSMEG\_2630* (NrnA') along with regulators of translation is suggestive of related function. *MATE* efflux family protein (*MSMEG\_2631*) has been shown to be involved in stress tolerance in mycobacteria (25). NanoRNase function of *MSMEG\_2630* may hence be linked to stress tolerance in mycobacteria.

**MSMEG\_2630 mutant is growth defective**

The transposon mutant, L154, with a single mini-mariner transposon insertion at 441 bp of *MSMEG\_2630* (this study) was used to probe the role of *MSMEG\_2630* in growth of cells. Disruption of *MSMEG\_2630* was not lethal but had a slight effect on growth when the mutant was grown in LB medium (Figure 3a). At a 48-h interval, there was a growth difference of ~1 OD<sub>600</sub> between WT and L154. Constitutive expression of *MSMEG\_2630* from a multicopy plasmid (pRPS16) under *hsp60* promoter restored growth of the mutant. This is somewhat different from the ortholog (Rv2837c), which has been previously reported to be essential for growth of *M. tuberculosis* H37Rv (26,27).

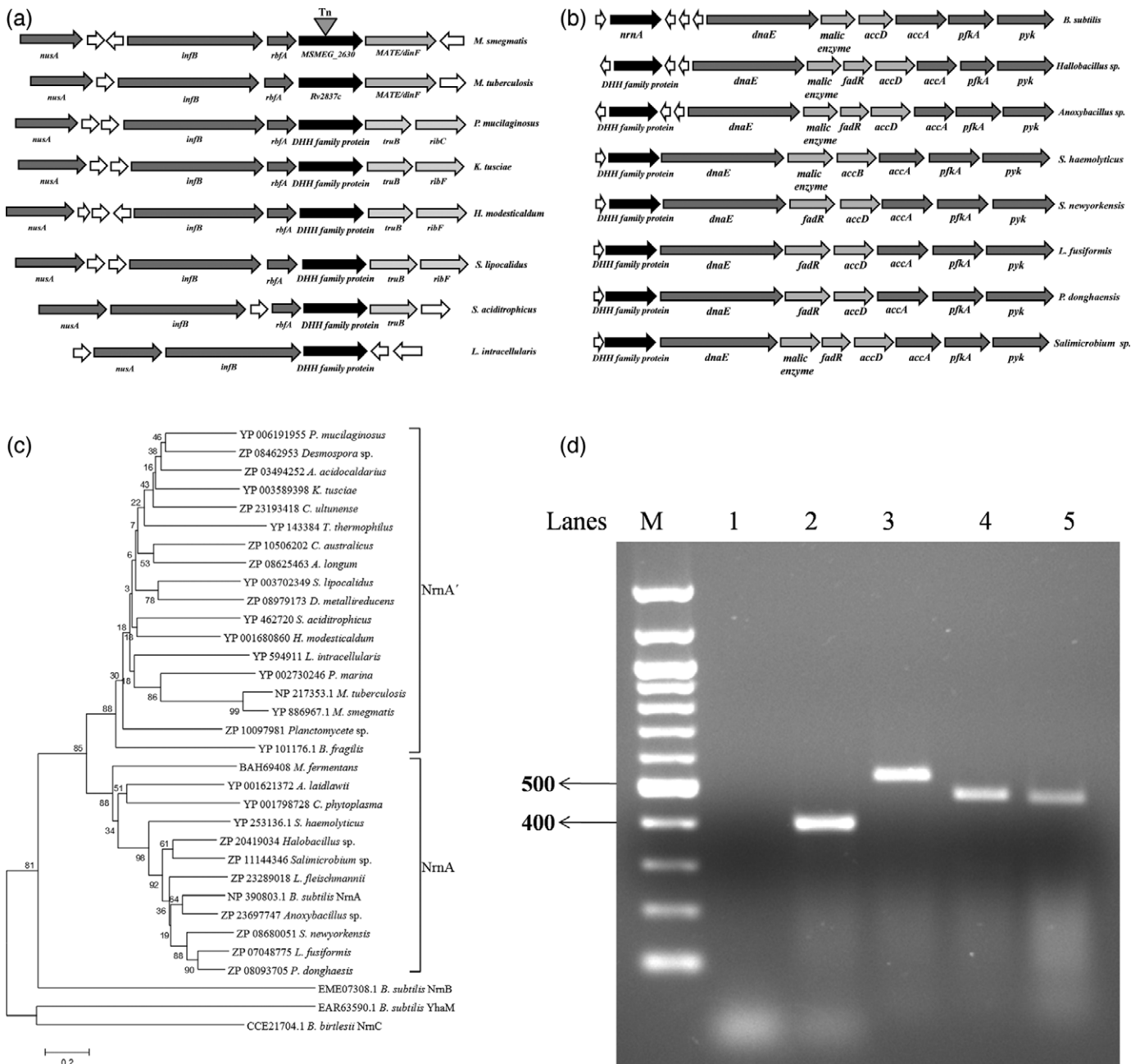
**Role of MSMEG\_2630 in survival under stress conditions**

In order to investigate the survival of *MSMEG\_2630* disruptant under DNA-damaging conditions, L154 cells were grown with different concentrations of oxidizing and DNA-damaging agents such as hydrogen peroxide, EtBr, menadione and mitomycin C (Figure 3b–e). Expression of *MSMEG\_2630* from pRPS16 restored growth of the mutant in all cases, confirming that the phenotype was due to the disruption of *MSMEG\_2630* and not due to polar effect. Other agents, namely, ciprofloxacin, nalidixic acid, zeocin, hydroxyurea, ethyl methanesulfonate and methyl methanesulfonate had no effect on growth of the mutant (data not shown).

L154 cells transformed with pRPS31, carrying the *M. tuberculosis* ortholog *rv2837c* could also complement the deletion of *MSMEG\_2630* when tested for potential DNA damage with EtBr (Figure 3c).

**MSMEG\_2630 is a 3'-5' exonuclease**

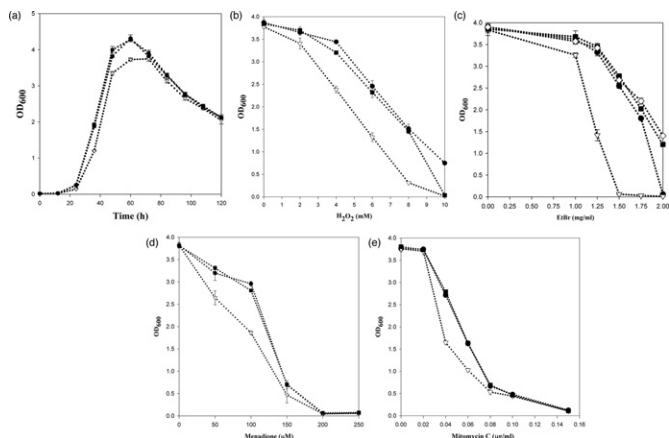
Purified recombinant *MSMEG\_2630* with an N-terminal His-tag (rMSMEG\_2630-A) was tested for nanoRNase ac-



**Figure 2.** Synteny conservation and phylogeny of nanoRNases are indicative of diverse families. Gene neighbors of (a) *MSMEG\_2630* and (b) *B. subtilis* NrnA show two distinct groups based on synteny conservation. The location of insertion of mini-mariner transposon (Tn) at 441 bp of *MSMEG\_2630* is indicated with a triangle. The indicated gene sizes are suggestive and not drawn to scale. (c) Phylogenetic tree for selected bacterial nanoRNases (NrnA, NrnB and NrnC). (d) RT-PCR confirms the presence of the indicated neighbors of *MSMEG\_2630* in a single operon. A 15- $\mu$ l aliquot of the cDNA synthesis reaction was amplified with primers for *MSMEG\_2628* and *MSMEG\_2629* intergenic region (lane 3), *MSMEG\_2629* and *MSMEG\_2630* intergenic region (lane 4), and *MSMEG\_2630* and *MSMEG\_2631* intergenic region (lane 5). An RT negative control (lane 1) was also performed to rule out DNA contamination and RT-PCR positive control was carried out for sigma factor A, *sigA* (lane 2). Lane M contains a standard DNA ladder (100 bp).

tivity. The nanoRNase activity was tested on different 5'-Cy5-labeled 5-mer RNA oligos, varying in their sequences, namely, 5'-Cy5-rCrArCrCrA, 5'-Cy-rCrCrCrCrC and 5'-Cy-rArArArArA. As shown in Figure 4a, rMSMEG\_2630-A could degrade all three 5-mer nanoRNAs tested. Turnover rates were measured as disappearance of the initial 5-mer substrate. As the initial rates of conversion of 5-mer were linear for a short period (less than 1 min), the turnover

rates provide only an approximate rate of conversion. The turnover rates, however, did not show any significant difference in sequence preference and were calculated to be in the range of 32–44 pmol/ $\mu$ g/min for the three different nanoRNAs (Figure 4b). The turnover rates for disappearance of the 5-mer RNA are similar to those reported for *M. tuberculosis* homolog Rv2837c (9–28 pmol/ $\mu$ g/min) (10). Like Rv2837c, 2-mer was found to be a preferred substrate, as



**Figure 3.** Role of MSMEG\_2630 in bacterial growth and stress tolerance. (a) Growth of mutant L154/pSMT3 (open triangle) is slow as compared to wild-type *M. smegmatis*/pSMT3 (WT) (filled circles). L154/pRPS16 (filled squares) has growth similar to WT. Mutant L154/pSMT3 is sensitive to stress from external agents, namely, (b) H<sub>2</sub>O<sub>2</sub>, (c) EtBr, (d) menadione and (e) mitomycin C. The stress is relieved by complementation and growth is restored to that of WT cells in all cases. L154/pSMT3 (open triangles), *M. smegmatis*/pSMT3 (WT) (filled circles) and L154/pRPS16 (filled squares). The growth defect in mutant L154 by EtBr (C) could also be complemented by *rv2837c* of *M. tuberculosis*, L154/pRPS31 (open diamonds).

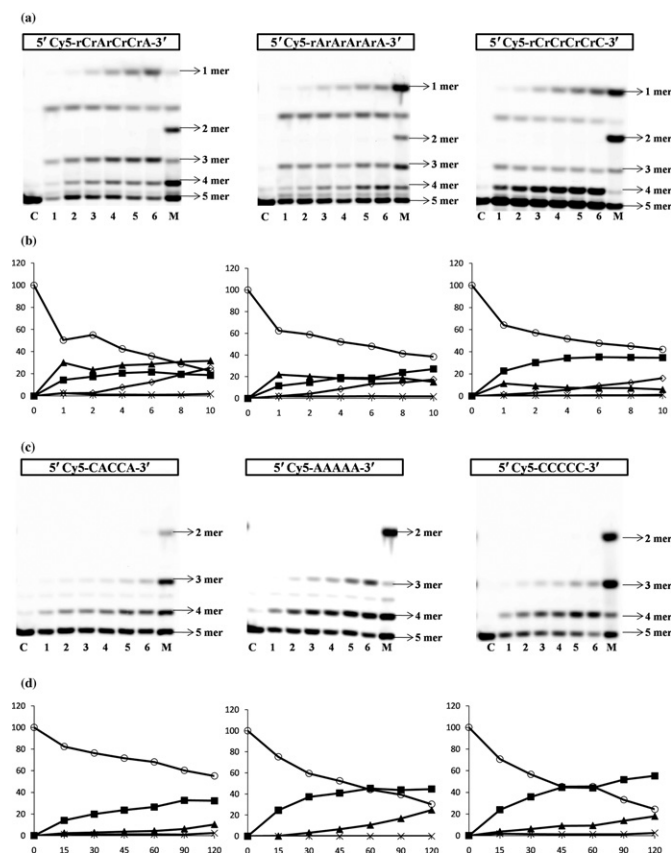
this intermediate was missing and could not be trapped in any of the reactions (Figure 4a).

The activity of rMSMEG\_2630-A was also tested on DNA oligos (Figure 4c). rMSMEG\_2630-A was able to degrade ssDNA oligos, namely, 5'Cy5-CACCA, 5'Cy5-CCCCC and 5'Cy5-AAAAA without any sequence preference with a 20- to 40-fold slower rate of turnover (calculated between 1.1 and 1.7 pmol/μg/min) in comparison to nanoRNA, when measured as disappearance of the initial 5-mer substrate (Figure 4d). In addition, no preference was observed for shorter ssDNA oligos.

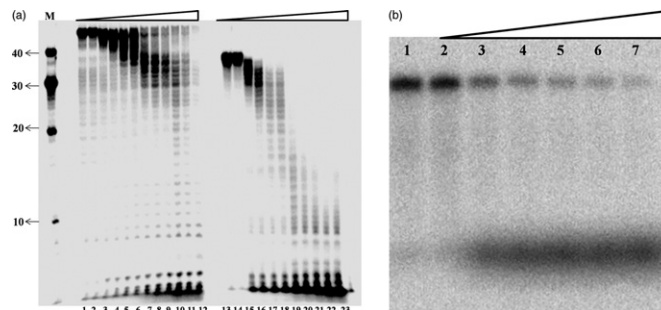
We then probed whether rMSMEG\_2630-A could degrade longer substrates as well. rMSMEG\_2630-A exhibited *in vitro* exonuclease activity on both a 49-mer ssDNA (Figure 5a and b) and a 35-mer RNA (Figure 5a) in the 3'-5' direction. No detectable activity could be identified on a 49-mer double-stranded DNA molecule (data not shown), indicating rMSMEG\_2630-A acts only on single-stranded oligonucleotides as substrates. Next, we probed the *in vitro* phosphodiesterase activity of rMSMEG\_2630-A. Phosphodiesterase activity of rMSMEG\_2630-A was tested with bis-pNpp as a substrate was found to be 16.1 pmol/μg/min (Figure 6a).

#### rMSMEG\_2630-A shows *in vitro* pAp phosphatase activity

Purified rMSMEG\_2630-A was tested for its CysQ-like phosphatase activity on pAp. rMSMEG\_2630-A could hydrolyze pAp *in vitro*, resulting in release of inorganic phosphate and the specific activity calculated to be 20.9 pmol/μg/min (Figure 6b). rMSMEG\_2630-A appears to be highly selective in recognition of this small substrate as no *in vitro* phosphatase activity could be detected with sodium pyrophosphate or sodium triphosphate under our experimental conditions.

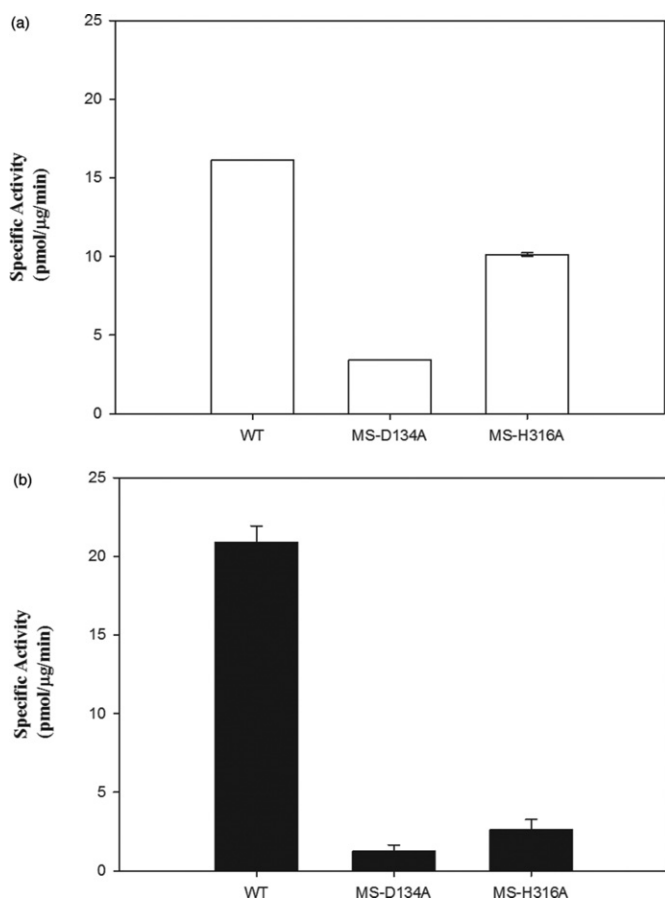


**Figure 4.** 3'-5' exonuclease activity of rMSMEG\_2630-A on Cy5-labeled short oligo substrates. rMSMEG\_2630-A catalyzed degradation of (a) Cy5-labeled 5-mer nanoRNA or (c) Cy5-labeled 5-mer ssDNA (as indicated). Reaction products were separated on a 22% polyacrylamide gel (containing 8-M urea). Cy5 oligos exhibit a reverse migration phenomenon (5). This effect is due to cyanine dyes having a lower net negative charge than nucleic acids; thus, removing nucleotides reduces the charge relative to the mass of the oligonucleotide and causes it to shift up instead of down. The different gel mobility as compared with RNA is caused by the one missing charge of DNA as DNA monomer cannot enter the gel. M indicates a size marker obtained by an *E. coli* Orn catalyzed reaction in all panels. Background bands observed in some panels were ignored for calculation of turnover rates. Quantification of (b) nanoRNA or (d) ssDNA reaction products and degradation intermediates. Open circles, closed squares, closed triangles, crosses and open diamonds indicate 5-mer, 4-mer, 3-mer, 2-mer and 1-mer, respectively.



**Figure 5.** Exonuclease activity of rMSMEG\_2630-A on longer substrates. (a) *In vitro* exonuclease activity of rMSMEG\_2630-A on 49-mer ssDNA 5'-end-labeled with [γ-<sup>32</sup>P] ATP was monitored at 0, 5, 10, 15, 30, 45, 60, 90, 120, 240 or 360 min (lanes 2–12) and 35-mer RNA, 5'-end-labeled with [γ-<sup>32</sup>P] ATP (lanes 13–23). Lane 1 is a control reaction (without enzyme). (b) The *in vitro* exonuclease activity of rMSMEG\_2630-A on 49-mer ssDNA 3'-end-labeled with [α-<sup>32</sup>P] dATP (lanes 2–7) monitored at 0, 15, 30, 45, 60, 120 confirms the 3'-5' direction of degradation by the enzyme. Lane 1 is control reaction without rMSMEG\_2630-A.

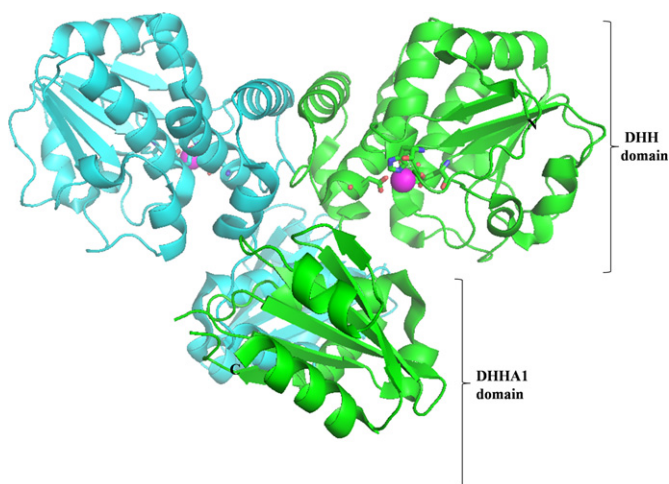




**Figure 6.** Phosphodiesterase and CysQ-like phosphatase activity of rMSMEG\_2630-A and mutant proteins. (a) Phosphodiesterase activity (unfilled bars) or (b) CysQ-like pAp phosphatase (filled bars) of rMSMEG\_2630-A, MS-D134A and MS-H316A was monitored with 1-mM bis-pNpp or 1-mM pAp as substrates, respectively, and specific activity was plotted (see the text for more details).

### Overall structure

In order to understand the molecular basis of selective recognition of nanoRNA by MSMEG\_2630 and the architecture of its constituent domains, the structure of recombinant tag-free MSMEG-2630 (rMSMEG\_2630-B) was determined to 2.2-Å resolution by Se-Met SAD method and refined to final  $R_{\text{work}}$  of 0.176 and  $R_{\text{free}}$  of 0.225 (Table 1). There are two molecules in the asymmetric unit forming a homodimer of identical subunits (Figure 7). The quaternary structure of rMSMEG\_2630-B was also confirmed by size-exclusion chromatography (Supplementary Figure S1). The overall structure of each subunit can be divided into two distinct domains, an NTD comprising residues 22–206 and a shorter CTD comprising residues 224–340. The two domains are joined by a short linker region consisting of residues 207–223. The unique N-terminal stretch in mycobacterial sequences (Figure 1) comprising residues 1–21 of rMSMEG\_2630-B was found to be disordered in both subunits; their presence in the protein crystal confirmed by N-terminal sequencing. The final model hence consists of 319 residues in each subunit.



**Figure 7.** Overall structure. The overall structure of MSMEG\_2630 indicates a homodimer with each subunit consisting of two distinct domains: the N-terminal DHH domain and the C-terminal DHHA1 domain. The two subunits are shown in different colors. A Mn<sup>2+</sup> bound at catalytic center is shown as a sphere (magenta) while the conserved residues involved in metal coordination are shown in stick mode. The N- and C- termini are labeled.

The NTD of rMSMEG\_2630-B consists of a twisted five-stranded parallel  $\beta$ -sheet with  $\alpha$ -helices on both sides (Figure 7 and Supplementary Figure S2A). The topology resembles the six-stranded parallel  $\beta$ -sheet of nucleotide binding  $\alpha/\beta$ -fold or Rossmann-fold with differences in the crossover helices as also confirmed by Dali (for instance, a Z-score of 10.6 and rmsd of 2.8 Å for 133 C $\alpha$  atoms were obtained for a proton-translocating transhydrogenase (PDB ID: 1DJL)). The NTD harbors the conserved <sup>134</sup>DHH<sup>136</sup> motif and hence designated the DHH domain. One Mn<sup>2+</sup> was located in the electron density map, coordinated to conserved residues in the DHH domain. The CTD consists of a central mixed  $\beta$ -sheet surrounded by  $\alpha$ -helices on both sides. The CTD harbors conserved residues of the DHHA1 motif and designated the DHHA1 domain of rMSMEG\_2630-B. In the following discussion in the rest of this manuscript, the NTD and CTD are hence referred to as DHH and DHHA1 domains, respectively. The average B-factor of the DHHA1 domain (49.8 Å<sup>2</sup> or 41.6 Å<sup>2</sup> in the two subunits) is higher than that of the DHH domain (27.9 Å<sup>2</sup> or 34.6 Å<sup>2</sup>) in both subunits, suggestive of larger flexibility in the DHHA1 domain. The DHH or DHHA1 domain in one subunit has similar conformation to the corresponding domain in the other subunit with low rmsd values of 0.29 Å (185 C $\alpha$  atoms) and 0.30 Å (114 C $\alpha$  atoms), respectively (Supplementary Figure S2B and C). However, superposition of complete subunits (rmsd: 2.59 Å for 295 C $\alpha$  atoms) reveals different orientations, indicating both DHH and DHHA1 domains can move relative to each other due to flexibility around the linker region, and provides multiple snapshots of the enzyme (Supplementary Figure S2D).

### Active site

The active site of MSMEG\_2630 is located at the domain interface and defined by an Mn<sup>2+</sup> coordinated to conserved

residues in the DHH domain.  $Mn^{2+}$  is present in an octahedral geometry coordinated to Asp-51, His-135 and Asp-185 of the protein. In addition, Asp-110 is coordinated to  $Mn^{2+}$  in a bidentate fashion along with two water molecules at distances of 2.33 Å and 2.75 Å to complete the coordination state (Figure 8a). Although  $Mn^{2+}$  is generally present in an octahedral geometry with six-coordination state, the seven-coordination state (with a bidentate coordination with at least one residue) has been previously seen for  $Mn^{2+}$  in several hydrolases (PDB ID: 1D3V) (28) and phosphatases (PDB ID: 1G5B) (29) in PDB.

The catalysis by DHH phosphoesterases including NrnA (8), RecJ (30) or pyrophosphatases (31–33) follows a two-metal ion mechanism. However, we could identify only one  $Mn^{2+}$  in each subunit in our apo structure. In the absence of a bound substrate, the position for the second metal ion is not occupied, suggesting that the second metal ion is not structurally required and may bind during catalysis.

The presence of  $Mn^{2+}$  in the structure was somewhat surprising as  $Mg^{2+}$  was used in protein purification and crystallization. An intrinsic bound  $Mn^{2+}$  in the protein active site is not exchanged despite excess of  $Mg^{2+}$  in the buffers in pyrophosphatases as well (32). The coordination geometry of both  $Mg^{2+}$  and  $Mn^{2+}$  is very similar and their coordination distances differ only by 0.1 Å and measured accurately in our high-resolution structure in agreement with  $Mn^{2+}$  rather than  $Mg^{2+}$  coordination distances (34). The metal ion in the crystal structure was further confirmed as  $Mn^{2+}$  through several other lines of evidence. Firstly, the local B-factors during refinement were found to be closer to the neighboring residues for a modeled  $Mn^{2+}$  (B-factor: 30.1 or 36.1 Å<sup>2</sup> in the two subunits) than a modeled  $Mg^{2+}$  (B-factor: 10.1 or 12.1 Å<sup>2</sup>; Supplementary Figure S3). Moreover, the coordination state of  $Mn^{2+}$  in the active site of Family II pyrophosphatases has been suggested to change from five to six upon substrate binding. Varied coordination states are possible for transition metals viz,  $Mn^{2+}$  but not  $Mg^{2+}$  (33,35).  $Mn^{2+}$  is hence conducive for catalysis due to its alterations between different coordination states, necessary for catalysis (33). Thirdly,  $Mn^{2+}$  has been suggested to accommodate bidentate coordination by carboxylate groups (Asp-110 of MSMEG\_2630) better than  $Mg^{2+}$  due to its larger ionic radius (32). Lastly,  $Mn^{2+}$  is a stronger Lewis acid than  $Mg^{2+}$  and is likely to activate a water molecule more effectively for a nucleophilic attack for catalysis (32). Activity assays indicate that although  $Mn^{2+}$  was the preferred ion for catalysis at low concentrations (Figure 8b), at saturating concentrations, both metal ions can substitute each other for exonuclease activity *in vitro*. However,  $Ca^{2+}$ ,  $Ni^{2+}$ ,  $Co^{2+}$ ,  $Zn^{2+}$  and  $Cu^{2+}$  could not serve as cofactors when used at saturating concentrations and exhibit no activity (Figure 8c).

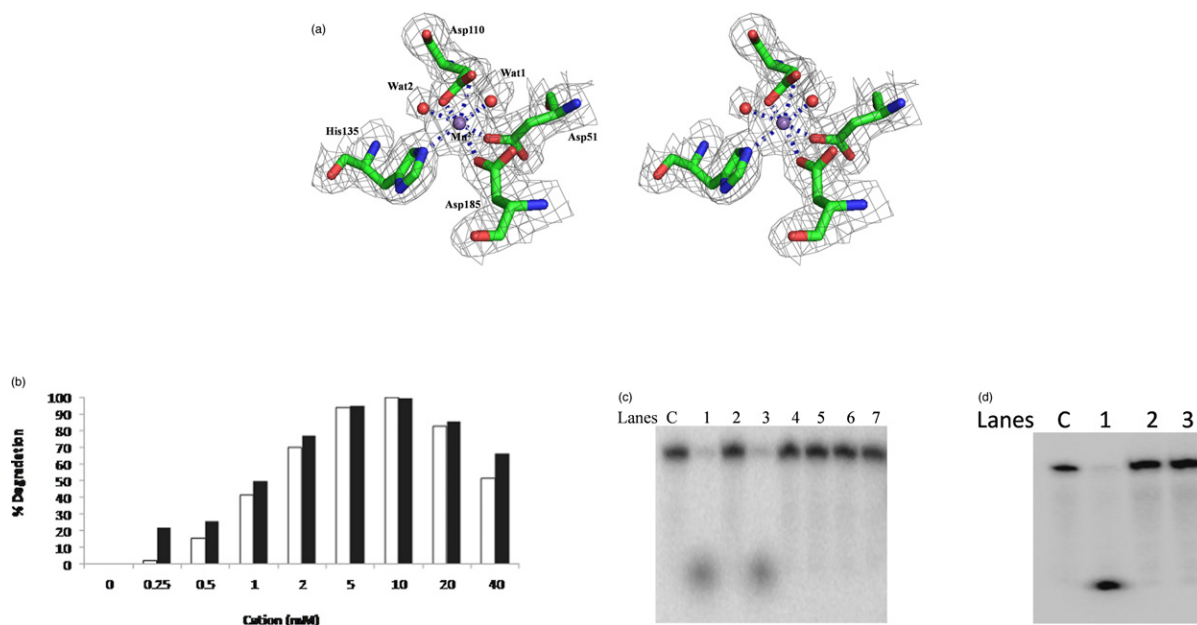
**Role of DHH motif.** DHH motif is conserved in all members of DHH superfamily. However, not all of the conserved residues of the motif are involved in metal coordination. While His-135 from the conserved <sup>134</sup>DHH<sup>136</sup> motif (or the equivalent histidine of the motif in other structures) is directly involved in metal coordination, the role of the other two residues is unclear. Examination of the structure sug-

gested Asp-134 and His-136 to be structurally important in arranging the active site for catalysis. Oδ1 atom of Asp-134 is hydrogen bonded to Nδ1 of His-136 and positions His-135 for coordination to  $Mn^{2+}$ . Any change in this region is likely to perturb the coordination state of  $Mn^{2+}$  and affect activity. A similar arrangement of the conserved aspartate with respect to the other conserved histidine of DHH motif to stabilize the bimetal catalytic center has been seen earlier in pyrophosphatases (31,33) but not other members of the DHH superfamily. The role of conserved aspartate was probed with a D134A mutant of MSMEG\_2630. Mutation of D134A in MS-D134A led to loss of both exonuclease (Figure 8d) and pAp phosphatase activity (Figure 6b) *in vitro*. A CD spectrum of MS-D134A (Supplementary Figure S4A) is similar to that of WT rMSEG\_2630-A and confirmed the overall folded state of the protein. Any potential local perturbations in MS-D134A structure appear to be minor and possibly limited to the active site and not directly discernible in the CD spectra, suggesting no other major structural changes took place in MS-D134A.

### Structural comparisons

A search for structural homologs was carried out with the final coordinates of rMSMEG\_2630-B dimer with Dali (36) and reveals hits from all the major members of DHH superfamily, namely, (i) NrnA (DHH/DHHA1 subfamily), (ii) RecJ exonucleases (also belonging to DHH/DHHA1 subfamily) and (iii) metal-dependent inorganic pyrophosphatases (belonging to DHH/DHHA2 subfamily). The closest structural homologs of MSMEG\_2630 are *B. fragilis* nanoRNase structures (PDB ID: 3W5W, Z-score 29.6, rmsd: 4.5 Å for 340 C<sup>α</sup> atoms and PDB ID: 3DMA, Z-score 30.1, rmsd: 4.2 Å for 340 C<sup>α</sup> atoms). *Staphylococcus haemolyticus* exopolyphosphatase (PDB ID: 3DEV, Z-score 28.0, rmsd: 3.2 Å for 316 C<sup>α</sup> atoms), an uncharacterized protein from Northeast Structural Genomics Consortium, also shows high Z-score and appears to belong to NrnA family.

The overall structure and arrangement of subunits in rMSMEG\_2630-B are quite similar to the GMP-bound nanoRNase structure of *B. fragilis* (PDB ID: 3W5W) (Figure 9). The superposition of individual subunits as given by Dali, however, is poor. In addition, a short stretch of residues at the N-terminus (residues 22–27 in MSMEG\_2630 and 2–11 in *B. fragilis* nanoRNase) does not align in the structures, indicating variability in the N-terminal region of nanoRNases. A closer examination of individual subunit structures indicates that the CTD in GMP-bound 3W5W moves upward toward NTD shortening the cavity at the domain interface (Figure 9a). In MSMEG\_2630, the cavity at the domain interface remains in a relatively open conformation resulting in large deviations when aligning the structures of MSMEG\_2630 with *B. fragilis* nanoRNase. In the MSMEG\_2630 apo structure, the distance between H316 of the DHHA1 motif (<sup>313</sup>GGGH<sup>316</sup>) and  $Mn^{2+}$  is 14.6 Å while in the ligand-bound ‘closed conformation’ in 3W5W, this distance is 6.6 Å. The individual domains, however, are superposed much better (Figure 9b and c) with rmsd of 1.61 Å (173 C<sup>α</sup>



**Figure 8.** Role of metal ion. (a) Stereo view of Mn<sup>2+</sup> coordinated at the active site. The residues involved in coordination, namely, Asp51, Asp110, His135 and Asp185 are conserved across all nanoRNase sequences (see Figure 1 and text for details). In addition, two water molecules (Wat1 and Wat2) are present in the active site to complete the coordination state of Mn<sup>2+</sup>. The manganese ion is shown as a purple sphere while water molecules are shown as red spheres. A 2Fo-Fc map generated at 1.0 sigma level indicates a good fit. (b) Exonuclease activity of rMSMEG\_2630-A as a function of increasing concentrations of Mn<sup>2+</sup> (filled bars) or Mg<sup>2+</sup> (unfilled bars) as indicated. (c) Exonuclease activity of rMSMEG\_2630-A on a 3'-P<sup>32</sup>-end-labeled ssDNA (49-mer) in the presence of 10 mM of indicated metal ion (lane 1: Mg<sup>2+</sup>; lane 2: Ca<sup>2+</sup>; lane 3: Mn<sup>2+</sup>; lane 4: Ni<sup>2+</sup>; lane 5: Co<sup>2+</sup>; lane 6: Zn<sup>2+</sup>; lane 7: Cu<sup>2+</sup>). No extraneous metal ion was added in the control reaction (lane C). (d) An activity assay with MS-D134A (lane 2) or MS-H316A (lane 3) on 3'-P<sup>32</sup>-end-labeled ssDNA (49-mer) showed no detectable exonuclease activity. Lane 1 is an assay with rMSMEG\_2630-A as a positive control while lane C is a 'no-protein' negative control.

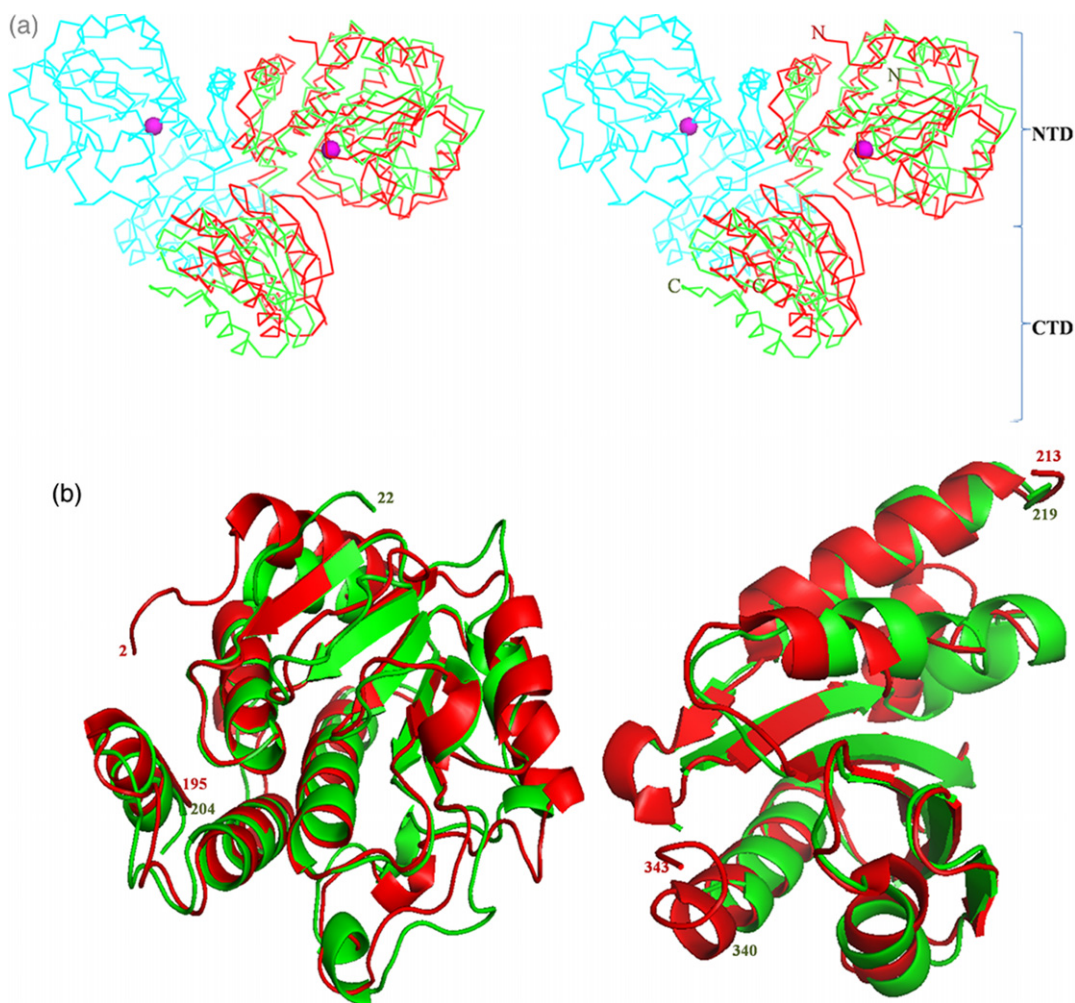
atoms) and 1.62 Å (110 C<sup>α</sup> atoms) for N-terminal DHH and C-terminal DHHA1 domains, respectively, confirming thereby that larger deviations in the overall structures of rMSMEG\_2630-B and *B. fragilis* nanoRNase are due to domain closure in the GMP-bound state. Overall similarities in the individual domains and conservation of conserved residues in equivalent positions, however, suggest that NrnA and NrnA' members bind to and degrade nucleotides through similar mechanisms.

A Dali search also revealed structural similarities between MSMEG\_2630 and RecJ exonucleases (PDB ID: 1IR6 and 2ZXP) (30,37), both members of DHH/DHHA1 subfamily. The structures of the N-terminal DHH domain and the C-terminal DHHA1 domain of both proteins are very similar to each other and superpose well with rmsd of 1.82 Å (for 185 C<sup>α</sup> atoms) and 2.83 Å (for 104 C<sup>α</sup> atoms), respectively (Supplementary Figure S5). However, while all other top structural homologs consist of biological dimers, RecJ family exists as a monomer. An additional OB-fold domain present in RecJ (30) but absent in nanoRNases confers binding ability to longer oligonucleotide substrates.

#### Differential subunit packing in DHH/DHHA1 and DHH/DHHA2 subfamilies

The third major family showing structural similarities with MSMEG\_2630 (belonging to DHH/DHHA1 subfamily) is family II pyrophosphatases and polyphosphatases (belonging to DHH/DHHA2 subfamily) (Supplementary Table S2). Structural comparison of MSMEG\_2630 with *Strep-*

*tococcus mutans* family II pyrophosphatase (32) provides interesting insights. Both MSMEG\_2630 NrnA' and the *S. mutans* pyrophosphatase form homodimers with each monomer folding into two distinct domains. Despite variable sequences in the N-terminal DHH and C-terminal DHHA1 or DHHA2 domains of MSMEG\_2630 and *S. mutans* pyrophosphatase (PDB ID: 1I74), respectively, individual subunits (rmsd of 2.92 Å for 309 C<sup>α</sup> atoms) and domains (rmsd of 2.19 Å for 146 C<sup>α</sup> atoms of N-terminal DHH domain and 2.80 Å for 93 C<sup>α</sup> atoms of C-terminal DHHA1 versus DHHA2 domain) of these two subfamilies have similar folds and show overall similar structures (Figure 10a and Supplementary Figure S6). However, there is a striking difference in the subunit arrangement in dimers in these two classes of proteins (Figure 10b and c). The family II pyrophosphatase homodimer is formed by the two N-terminal (DHH) domains packing together making a hydrophobic dimer interface of two antiparallel β-strands, one contributed by each monomer. Together, the six β-strands from each N-terminal DHH domain, hence, form a single continuous 12-stranded, twisted β-sheet running through the protein (31). In MSMEG\_2630, the terminal α-helix of the DHH domain (residues 194–205) and the first α-helix of the DHHA1 domain (residues 227–234) apart from the linker region form the dimer interface through interactions with corresponding regions of the other subunit and bury 1985 Å<sup>2</sup> area at the dimer interface. In contrast to *S. mutans* pyrophosphatase, MSMEG\_2630 subunits are oriented in a perpendicular direction to each other (Figure



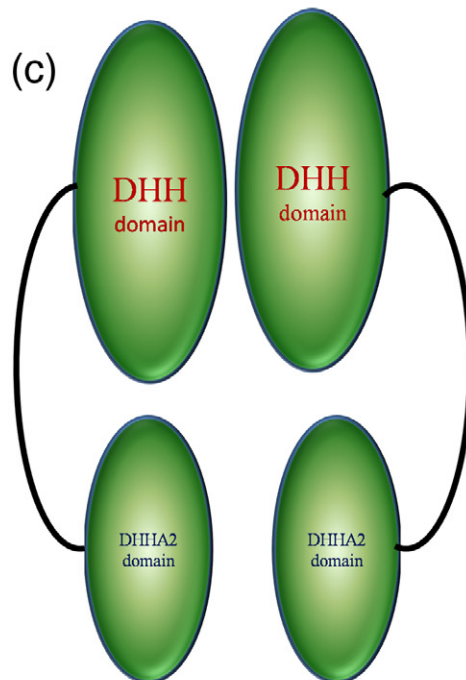
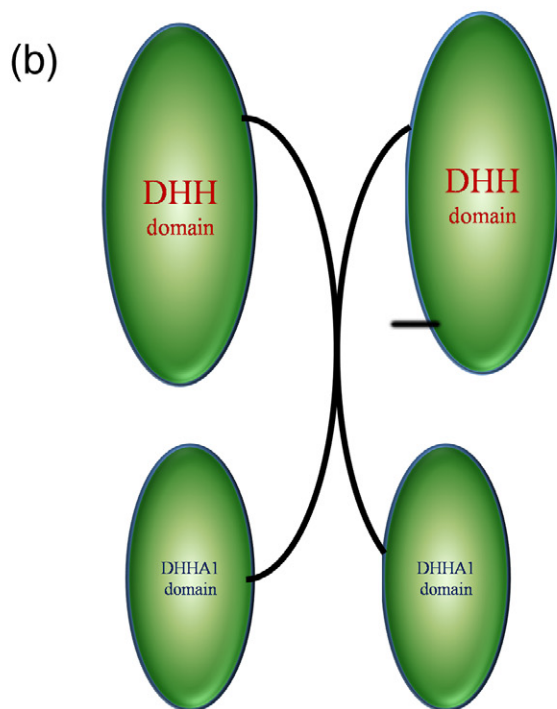
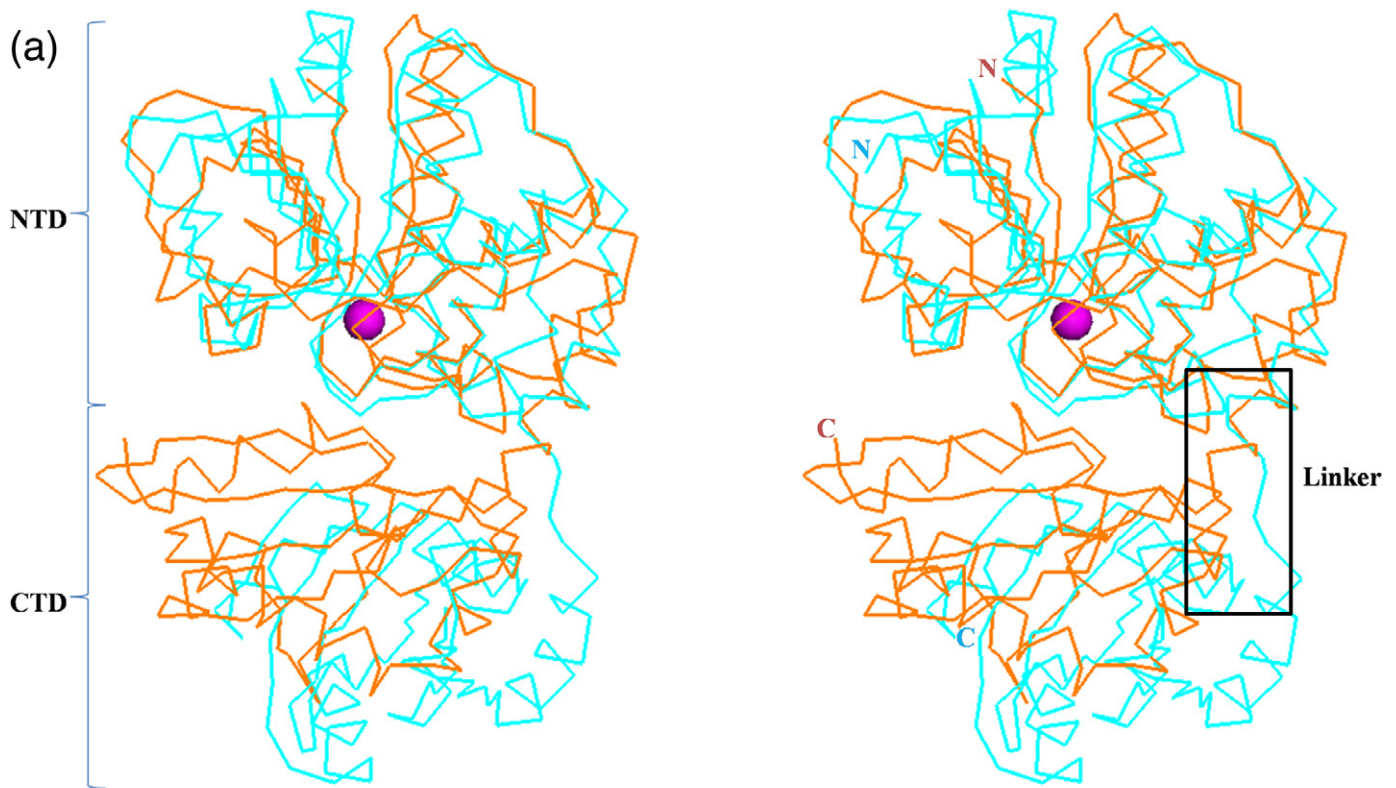
**Figure 9.** Structural comparison of MSMEG\_2630 with *B. fragilis* nanoRNase. (a) Stereoview of superposition of MSMEG\_2630 (subunits in green and cyan) with nanoRNase of *B. fragilis* (PDB ID: 3W5W) (red) indicates CTD of *B. fragilis* nanoRNase is slightly shifted toward catalytic center. A bound  $Mn^{2+}$  (magenta sphere) marks the active site of MSMEG\_2630. Superposition of (b) the N-terminal DHH domain and (c) the C-terminal DHHA1 domain alone of the two proteins shows better alignment indicating overall similar structures of the individual domains.

10b and c). This orientation presents two independent active sites on opposite faces.

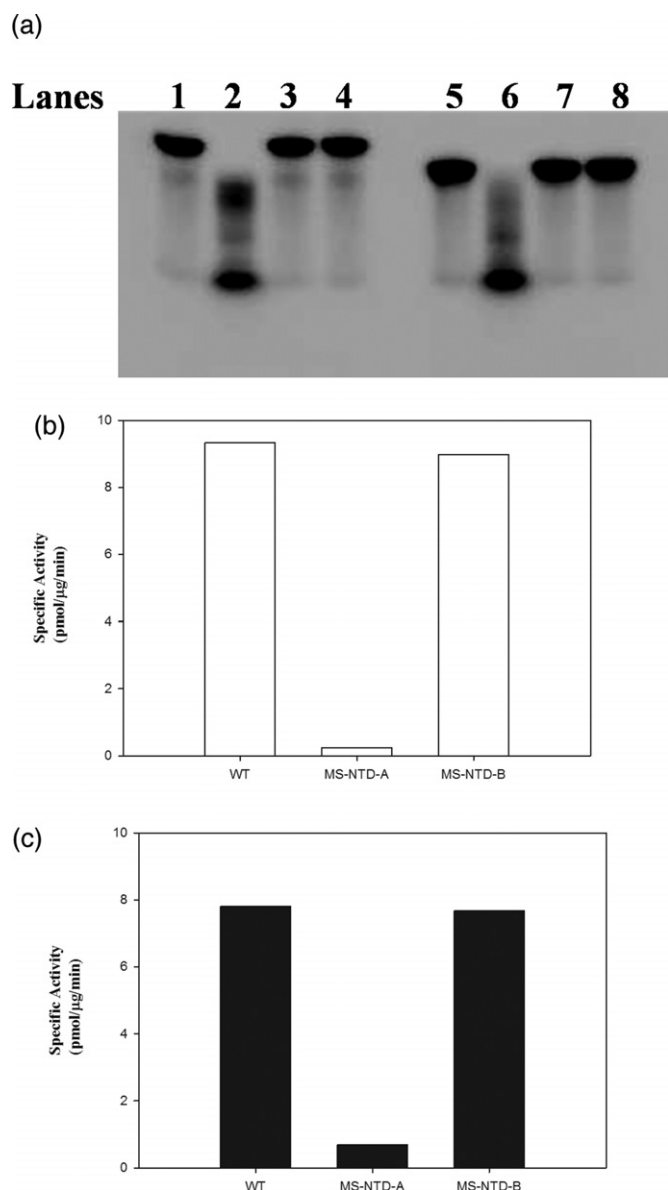
The linker region in pyrophosphatases is relatively small when compared to nanoRNases (Figure 10a). Consequently, the cavity in rMSMEG\_2630-B appears to be enlarged when compared to pyrophosphatases. Although rigid body domain movements reveal an ‘open’ conformation in the active site of *B. subtilis* pyrophosphatase (31), the enlarged cavity between the DHH and DHHA1 domains of rMSMEG\_2630-B may be able to accommodate comparatively larger substrate (oligonucleotides) as compared to pyrophosphatases even in the ‘closed’ state. Hence, despite overall structural similarities in the two domains of rMSMEG\_2630-B and pyrophosphatases, rMSMEG\_2630-B may be able to utilize DHH domains in an alternate subunit arrangement in contrast to pyrophosphatases to reveal a completely different quaternary structure of the two proteins (Figure 10b and c).

### Role of DHH and DHHA1 domains

In order to investigate the role of DHH and DHHA1 domains, we made constructs expressing individual domains. Deletion of the DHHA1 domain in MS-NTD-A (1–188) and MS-NTD-B (1–211) resulted in loss of *in vitro* exonuclease activity on synthetic 49-mer ssDNA substrates (Figure 11a). A CD spectrum of MS-NTD-A and MS-NTD-B (Supplementary Figure S4B) confirmed the overall folded state of the protein. However, MS-NTD-B exhibited both pAp phosphatase and phosphodiesterase activity *in vitro* (Figure 11b and c) suggesting that the DHH domain is capable of catalysis on its own but requires the DHHA1 domain (212–340) for binding to longer oligonucleotide substrates. Both the phosphodiesterase activity of MS-NTD-B (8.9 pmol/ $\mu$ g/min) and pAp phosphatase activity (7.7 pmol/ $\mu$ g/min) were similar to that of full-length Ntag\_MSMEG\_2630 (9.3 nmol/ $\mu$ g/min and 7.8 pmol/ $\mu$ g/min, respectively). The absence of pAp phosphatase and phosphodiesterase activity in MS-NTD-A was somewhat peculiar. Asp-185 of MSMEG\_2630 is required



**Figure 10.** Structural comparison of MSMEG\_2630 with Class II pyrophosphatases. (a) Stereoview of superposition of MSMEG\_2630 (cyan) with *S. mutans* pyrophosphatase (PDB ID: 1I74) (orange). The longer linker region in MSMEG\_2630 enlarging the cavity at the domain interface to enable binding of larger oligonucleotide substrates is indicated. Cartoon representation of subunit arrangement in dimers of (b) MSMEG\_2630 and (c) family II pyrophosphatases.



**Figure 11.** Role of the N-terminal DHH domain of MSMEG\_2630. (a) Exonuclease activity was monitored on ssDNA (lanes 1–4) or ssRNA (lanes 5–8). Deletion of CTD leads to loss of exonuclease activity on 3'-P<sup>32</sup> ssDNA or RNA in MS-NTD-A (lane 3 and lane 7) as well as MS-NTD-B (lane 4 and lane 8). Lanes 1 and 5: control reactions without MSMEG\_2630; lanes 2 and 6: Ntag\_MSMEG\_2630. (b) Phosphodiesterase activity (unfilled bars) or (c) CysQ-like pAp phosphatase (filled bars) of MS-NTD-A and MS-NTD-B was monitored with 1-mM bis-pNpp or 1-mM pAp as substrates, respectively, and specific activity was plotted (see the text for more details). WT is reaction with Ntag\_MSMEG\_2630.

for coordination of Mn<sup>2+</sup> and maintenance of the active site (Figure 8a). Asp-185 is present close to the C-terminus of MS-NTD-A construct, which consists of 188 residues. We presume rearrangements in the C-terminus of MS-NTD-A affected the coordination state of the bound metal ion, thereby leading to loss of activity even on small substrates, unlike MS-NTD-B.

C-terminal DHHA1 domain of MSMEG\_2630 is devoid of known active site residues. In order to explore the role of DHHA1 domain, a Dali search with the C-terminal DHHA1 domain alone was carried out. Dali search shows top hits from CTDs of DHH subfamilies (NrnA, pyrophosphatases and RecJ). Structural similarity was also observed with the C-ala domain of alanyl t-RNA synthetase involved in single-stranded nucleotide binding (PDB ID: 3G98, Z-score: 8.8, rmsd: 2.8 Å for 111 C<sup>α</sup> atoms) (38). The overall structural similarity of DHHA1 domain with the C-ala domain suggests its role in binding to longer single-stranded oligonucleotide substrates as suggested earlier for both RecJ (38) and NrnA exonuclease (8). However, no detectable binding to ssDNA or RNA could be identified with the DHHA1 domain (MS-CTD) under our experimental conditions (data not shown).

The role of DHHA1 domain and the residues therein was hence investigated by mutagenesis of H316 in the conserved <sup>313</sup>GGGH<sup>316</sup> motif of full-length protein. An activity assay with MS-H316A showed no phosphatase activity on pAp (Figure 6b) or exonuclease activity (Figure 8d). In addition, the specific activity with bis-pNpp as a substrate was found to be 10.1 pmol/μg/min, an ~40% loss in the *in vitro* phosphodiesterase activity (Figure 6a). A structurally equivalent His311 of *B. fragilis* nanoRNase has been proposed earlier to be a likely residue in substrate recognition (8). A GMP ligand binds close to the conserved His311 of *B. fragilis* nanoRNase structure, although no direct interactions were observed between GMP and residues in the motif (8). Mutation of equivalent conserved histidine resulted in decreased exonuclease activity in *E. coli* RecJ as well (39). A complete loss of exonuclease activity in the MS-H316A in contrast to *E. coli* RecJ could be due to an additional OB-fold domain available in RecJ for oligonucleotide substrate binding. Like MS-NTD-B, MS-H316A appears to retain phosphodiesterase activity on bis-pNpp as it contains all the catalytic residues. We could, however, not explain the loss of pAp phosphatase activity of MS-H316A.

## DISCUSSION

Members of DHH superfamily include 5'-3' exonucleases [RecJ (30,37)], nanoRNases [NrnA, (5,6,8,10)], cyclic nucleotide phosphodiesterases [YybT in *B. subtilis*, (40)] and eukaryotic prune proteins (41), or phosphatases [exopolyphosphatase, family II pyrophosphatase (31–32,35,42)]. Despite the versatile nature of their specific substrates, all DHH superfamily members play significant physiological roles and are important in maintenance of homeostasis. DHH superfamily in bacteria can be divided into three distinct subfamilies on the basis of specific substrate recognition: (i) RecJ or RecJ-like 5'-3' exonucleases (DHH/DHHA1 subfamily), (ii) manganese-dependent family II pyrophosphatases (DHH/DHHA2 subfamily) and (iii) the recently discovered nanoRNases (DHH/DHHA1 subfamily).

Multiple nanoRNases of DHH/DHHA1 subfamily [NrnA and NrnB in *B. subtilis* (5,6), Mpn140 in *M. pneumoniae*, Rv2837c in *M. tuberculosis* (10), nanoRNase of *B. fragilis* (8)] and DEDD subfamily (NrnC in *Bartonella henselae*) (7) have been identified. However, the phylo-

genetic origin of these multiple nanoRNase families in prokaryotes is unclear. On the basis of its gene location and synteny conservation, we show for the first time that DHH/DHHA1 subfamily of NrnA may have another level of functional divergence and it may have evolved to carry out distinct *in vivo* functions.

### NanoRNases of different bacteria are present in two distinct families

Rv2837c is the only bifunctional enzyme with both nanoRNase and pAp phosphatase activity identified in *M. tuberculosis* so far (10). We have characterized MSMEG\_2630 of *M. smegmatis* and attributed this enzyme with similar biochemical properties. However, mycobacteria and a few other actinobacteria are unique in encoding a nanoRNase, an ORN and CysQ in their genomes (5,43-44). Differing substrate specificities in multiple nanoRNA degrading enzymes in bacteria have been suggested earlier (1). In order to seek greater insights into the physiological role of different nanoRNA-processing enzymes in mycobacteria (or Actinomycetes), we revisited the locus of NrnA in different organisms for clues to its possible physiological function. Phylogenetic analysis and examination of neighboring genes revealed clustering of sequences into two distinct groups: operons containing *infB* and *rbfA* genes and those lacking these. In addition to *infB* and *rbfA*, *MATE/dinF* encoding for MATE efflux pump is present on the same operon in both *M. tuberculosis* and *M. smegmatis* genomes. MATE efflux pump is involved in the efflux of toxic compounds (25). The rescue of sensitive phenotype in *M. smegmatis*  $\Delta$ MSMEG\_2630 to toxic compounds by complementation with MSMEG\_2630 of *M. smegmatis* or *rv2837c* of *M. tuberculosis* (in case of EtBr) suggests all these genes may be involved in related functions of DNA damage tolerance and repair. The redundant presence of ORN and nanoRNase only among Actinobacteria leads to a somewhat provocative assumption that while the role of ORN in mycobacteria may be in generalized nanoRNA degradation, NrnA' may be specialized to be involved in DNA repair-related activities. In a recent study, the possible coevolution of nanoRNases and their association with the RNA degradation machinery of the bacterial cell has been suggested (45). The coevolution with RNA degradasome and synteny conservation of MSMEG\_2630 and Rv2837c with *infB* and *rbfA* hence suggest that mycobacterial nanoRNases may have a dual role in DNA-repair related as well as in the transcription/translation fidelity in the cell.

### New packaging of an old fold for mechanistic conservation

Despite a large number of structures in PDB, the total number of folds is limited (46). Proteins use these limited folds in different permutations to achieve a functional and mechanistic convergence for related activities. Functional convergence in several enzymes through conservation of active site residues despite completely different structures is a common feature in bacterial enzymes. For instance, despite completely different sequence and structure, family I and family II pyrophosphatases remain mechanistically analogous to each other through conservation of their active sites

(32). Similar examples of convergent evolution to conserve their mechanisms mean that Trypsin- or subtilisin-families use the same catalytic triad for protein hydrolysis (47). Similarly, despite completely different three-dimensional structures, the active site and catalytic mechanism of topoisomerase and site-specific recombinases are similar (48).

NanoRNases achieve a functional conservation and multiple activities in a unique way. The three-dimensional structures of either the NTD or CTD of NrnA, RecJ and family II pyrophosphatases are very similar. The loss of exonuclease activity on DNA while retaining phosphodiesterase as well as phosphatase activity by the DHH domain (residues 1–211) of MSMEG\_2630 suggests that the DHH domain confers activity while the DHHA1 domain confers substrate recognition in these proteins. However, completely different packing of the subunits in the three proteins results in recognition of different substrates in a highly specific manner. This is achieved through several minor alterations: (i) alternate packing of subunits via additional interactions in the N-terminal DHH domain, C-terminal DHHA1 domain and linker region of MSMEG\_2630 against  $\beta$ -sheet extension in the N-terminal DHH domain of pyrophosphatases, giving rise to completely different quaternary structures. (ii) Extension of the linker between DHH and DHHA1 domains of MSMEG\_2630, to enlarge the substrate binding cavity for binding of oligonucleotides. (iii) Insertion of an additional OB-fold between the DHH and DHHA1 domains of RecJ to enlarge domain interface forcing the protein to remain a monomer. Intriguingly, apart from the above modifications, all these proteins harbor structurally similar N-terminal DHH or C-terminal DHHA1/DHHA2 domains. Hence, through combinations of previously existing DHH and DHHA1 domains, nanoRNases achieve a new quaternary arrangement enabling it to achieve multiple functions in a mechanistically similar manner. This mode of conservation of mechanism is also evolutionarily advantageous to the cell by avoiding encoding of multiple folds in its genome.

### CONCLUSIONS

Permutations of few available structural and sequence motifs are utilized by bacteria to achieve multiple, functionally distinct activities. The structure of MSMEG\_2630 nanoRNase reveals a new way of oligomerization in exopolyphosphatase-like subunits to achieve a similar metal-dependent activity on longer oligonucleotide substrates.

### ACCESSION NUMBERS

The coordinates of MSMEG\_2630 have been deposited with PDB with the accession code 4LS9. Accession codes of other PDB IDs analyzed: 1DJL, 1D3V, 1G5B, 3W5W, 3DMA, 3DEV, 11R6, 2ZXP, 1174 and 3G98.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGMENTS

We thank Dr Beena Pillai and Dr Chetana Sachidanandan for critically reading the manuscript; Dr Rajesh Gokhale for his support for the project; Prof. Stewart Shuman and Dr Krishna Murari Sinha for pET28-His10-Smt3 plasmid; Santosh Kumar and Prachi Agarwala for help with CD experiments; Sonam Nain for cloning *E. coli* oligoribonuclease; Dr Atul Kumar, Prof. Antoine Danchin and Prof. Undine Mechold for discussions.

## FUNDING

Council of Scientific and Industrial Research (CSIR) Project Grant [FAC-02 to B.T.]; Senior Research Fellowships from CSIR [to R.S., D.K., A.G.]; Department of Biotechnology (DBT), India (BM14 project) - a collaboration between DBT (India), EMBL and ESRF for data collection. Funding for open access charge: CSIR funds.  
*Conflict of interest statement.* None declared.

## REFERENCES

- Nickels, B.E. and Dove, S.L. (2011) NanoRNAs: a class of small RNAs that can prime transcription initiation in bacteria. *J. Mol. Biol.*, **412**, 772–781.
- Vvedenskaya, I.O., Sharp, J.S., Goldman, S.R., Kanabar, P.N., Livny, J., Dove, S.L. and Nickels, B.E. (2012) Growth phase-dependent control of transcription start site selection and gene expression by nanoRNAs. *Genes Dev.*, **26**, 1498–1507.
- Goldman, S.R., Sharp, J.S., Vvedenskaya, I.O., Livny, J., Dove, S.L. and Nickels, B.E. (2011) NanoRNAs prime transcription initiation in vivo. *Mol. Cell*, **42**, 817–825.
- Zhang, X., Zhu, L. and Deutscher, M.P. (1998) Oligoribonuclease is encoded by a highly conserved gene in the 3'-5' exonuclease superfamily. *J. Bacteriol.*, **180**, 2779–2781.
- Mechold, U., Fang, G., Ngo, S., Ogryzko, V. and Danchin, A. (2007) YtqI from *Bacillus subtilis* has both oligoribonuclease and pAp-phosphatase activity. *Nucleic Acids Res.*, **35**, 4552–4561.
- Fang, M., Zeisberg, W.M., Condon, C., Ogryzko, V., Danchin, A. and Mechold, U. (2009) Degradation of nanoRNA is performed by multiple redundant RNases in *Bacillus subtilis*. *Nucleic Acids Res.*, **37**, 5114–5125.
- Liu, M.F., Cescau, S., Mechold, U., Wang, J., Cohen, D., Danchin, A., Boulouis, H.J. and Biville, F. (2012) Identification of a novel nanoRNase in *Bartonella*. *Microbiology*, **158**, 886–895.
- Uemura, Y., Nakagawa, N., Wakamatsu, T., Kim, K., Montelione, G.T., Hunt, J.F., Kuramitsu, S. and Masui, R. (2013) Crystal structure of the ligand-binding form of nanoRNase from *Bacteroides fragilis*, a member of the DHH/DHHA1 phosphoesterase family of proteins. *FEBS Lett.*, **587**, 2669–2674.
- Chin, K.H., Yang, C.Y., Chou, C.C., Wang, A.H. and Chou, S.H. (2006) The crystal structure of XC847 from *Xanthomonas campestris*: a 3'-5' oligoribonuclease of DnaQ fold family with a novel opposingly shifted helix. *Proteins*, **65**, 1036–1040.
- Postic, G., Danchin, A. and Mechold, U. (2012) Characterization of RrnA homologs from *Mycobacterium tuberculosis* and *Mycoplasma pneumoniae*. *RNA*, **18**, 155–165.
- Kumar, A., Saigal, K., Malhotra, K., Sinha, K.M. and Taneja, B. (2011) Structural and functional characterization of Rv2966c protein reveals an RsmD-like methyltransferase from *Mycobacterium tuberculosis* and the role of its N-terminal domain in target recognition. *J. Biol. Chem.*, **286**, 19652–19661.
- Evans, G. and Pettifer, R.F. (2001) CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J. Appl. Cryst.*, **34**, 82–86.
- Leslie, A.G.W. and Powell, H.R. (2007) Processing diffraction data with mosflm. In: Read, R.J. and Sussman, J.L. (eds). *Evolving Methods for Macromolecular Crystallography*. Springer, Netherlands, Vol. **245**, pp. 41–51.
- Diederichs, K. and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct. Biol.*, **4**, 269–275.
- Potterton, E., Briggs, P., Turkenburg, M. and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Cryst.*, **D67**, 1131–1137.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A. et al. (2011) Overview of the CCP4 suite and current developments. *Acta Cryst.*, **D67**, 235–242.
- Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S. and Tucker, P.A. (2009) On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Cryst.*, **D65**, 1089–1097.
- Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Cryst.*, **D66**, 486–501.
- Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R. A., Winn, M. D., Long, F. and Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst.*, **D67**, 355–367.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O.S., Vornrhein, C. et al. (2011) *BUSTER 2.10.0*. Global Phasing Ltd, Cambridge.
- Painter, J. and Merritt, E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Cryst.*, **D62**, 439–450.
- Grover, A. and Sharma, R. (2006) Identification and characterization of a major Zn(II) resistance determinant of *Mycobacterium smegmatis*. *J. Bacteriol.*, **188**, 7026–7032.
- Gaora, P.O., Barnini, S., Hayward, C., Filley, E., Rook, G., Young, D. and Thole, J. (1997) Mycobacteria as immunogens: development of expression vectors for use in multiple mycobacterial species. *Med. Princ. Pract.*, **6**, 91–96.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, **28**, 2731–2739.
- Mishra, M.N. and Daniels, L. (2013) Characterization of the MSMEG\_2631 gene (mmp) encoding a multidrug and toxic compound extrusion (MATE) family protein in *Mycobacterium smegmatis* and exploration of its polyspecific nature using biologic phenotype microarray. *J. Bacteriol.*, **195**, 1610–1621.
- Sasseti, C.M., Boyd, D.H. and Rubin, E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.*, **48**, 77–84.
- Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J. and Sasseti, C.M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.*, **7**, e1002251.
- Cox, J.D., Kim, N.N., Traish, A.M. and Christianson, D.W. (1999) Arginase-boronic acid complex highlights a physiological role in erectile function. *Nat. Struct. Biol.*, **6**, 1043–1047.
- Voegtli, W.C., White, D.J., Reiter, N.J., Rusnak, F. and Rosenzweig, A.C. (2000) Structure of the bacteriophage lambda Ser/Thr protein phosphatase with sulfate ion bound in two coordination modes. *Biochemistry*, **39**, 15365–15374.
- Wakamatsu, T., Kitamura, Y., Kotera, Y., Nakagawa, N., Kuramitsu, S. and Masui, R. (2010) Structure of RecJ exonuclease defines its specificity for single-stranded DNA. *J. Biol. Chem.*, **285**, 9762–9769.
- Ahn, S., Milner, A.J., Futterer, K., Konopka, M., Ilias, M., Young, T.W. and White, S.A. (2001) The “open” and “closed” structures of the type-C inorganic pyrophosphatases from *Bacillus subtilis* and *Streptococcus gordonii*. *J. Mol. Biol.*, **313**, 797–811.
- Merckel, M.C., Fabrichny, I.P., Salminen, A., Kalkkinen, N., Baykov, A.A., Lahti, R. and Goldman, A. (2001) Crystal structure of *Streptococcus mutans* pyrophosphatase: a new fold for an old mechanism. *Structure*, **9**, 289–297.
- Fabrichny, I.P., Lehtio, L., Tammenkoski, M., Zyryanov, A.B., Oksanen, E., Baykov, A.A., Lahti, R. and Goldman, A. (2007) A trimetal site and substrate distortion in a family II inorganic pyrophosphatase. *J. Biol. Chem.*, **282**, 1422–1431.
- Harding, M.M. (2001) Geometry of metal-ligand interactions in proteins. *Acta Cryst.*, **D57**, 401–411.



35. Fabrichny, I.P., Lehtio, L., Salminen, A., Zyryanov, A.B., Baykov, A.A., Lahti, R. and Goldman, A. (2004) Structural studies of metal ions in family II pyrophosphatases: the requirement for a Janus ion. *Biochemistry*, **43**, 14403–14411.
36. Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W549.
37. Yamagata, A., Kakuta, Y., Masui, R. and Fukuyama, K. (2002) The crystal structure of exonuclease RecJ bound to  $Mn^{2+}$  ion suggests how its characteristic motifs are involved in exonuclease activity. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 5908–5912.
38. Guo, M., Chong, Y.E., Beebe, K., Shapiro, R., Yang, X.L. and Schimmel, P. (2009) The C-Ala domain brings together editing and aminoacylation functions on one tRNA. *Science*, **325**, 744–747.
39. Sutera, V.A. Jr, Han, E.S., Rajman, L.A. and Lovett, S.T. (1999) Mutational analysis of the RecJ exonuclease of *Escherichia coli*: identification of phosphoesterase motifs. *J. Bacteriol.*, **181**, 6098–6102.
40. Rao, F., See, R.Y., Zhang, D., Toh, D.C., Ji, Q. and Liang, Z.X. (2010) YybT is a signaling protein that contains a cyclic dinucleotide phosphodiesterase domain and a GGDEF domain with ATPase activity. *J. Biol. Chem.*, **285**, 473–482.
41. Aravind, L. and Koonin, E.V. (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem. Sci.*, **23**, 469–472.
42. Ugochukwu, E., Lovering, A.L., Mather, O.C., Young, T.W. and White, S.A. (2007) The crystal structure of the cytosolic exopolyphosphatase from *Saccharomyces cerevisiae* reveals the basis for substrate specificity. *J. Mol. Biol.*, **371**, 1007–1021.
43. Hatzios, S.K., Iavarone, A.T. and Bertozzi, C.R. (2008) Rv2131c from *Mycobacterium tuberculosis* is a CysQ 3'-phosphoadenosine-5'-phosphatase. *Biochemistry*, **47**, 5823–5831.
44. Young, P.A., Elvin, C.M., Hamdan, S.M., Wood, R.J., Liyou, N.E., Hamwood, T.E., Jennings, P.A. and Dixon, N.E. (2008) Hydrolysis of the 5'-p-nitrophenyl ester of TMP by oligoribonucleases (ORN) from *Escherichia coli*, *Mycobacterium smegmatis*, and human. *Protein Expr. Purif.*, **57**, 180–187.
45. Engelen, S., Vallenet, D., Médigue, C. and Danchin, A. (2012) Distinct co-evolution patterns of genes associated to DNA polymerase III DnaE and PolC. *BMC Genomics*, **13**, 69–83.
46. Oberai, A., Ihm, Y., Kim, S. and Bowie, J.U. (2006) A limited universe of membrane protein families and folds. *Protein Sci.*, **15**, 1723–1734.
47. Wallace, A.C., Laskowski, R.A. and Thornton, J.M. (1996) Derivation of 3D coordinate templates for searching structural databases: application to Ser-His-Asp catalytic triads in the serine proteinases and lipases. *Protein Sci.*, **5**, 1001–1013.
48. Cheng, C., Kussie, P., Pavletich, N. and Shuman, S. (1998) Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell*, **92**, 841–850.