Crystal Structure and Biochemical Characterization of a 
*Mycobacterium smegmatis* AAA-Type Nucleoside Triphosphatase Phosphohydrolase (Msm0858)

Mihaela-Carmen Unciuleac,Paul C. Smith,Stewart Shuman

Molecular Biology Program, Sloan-Kettering Institute, New York, New York, USA; Chemistry Department, Fordham University, Bronx, New York, USA

ABSTRACT

AAA proteins (ATPases associated with various cellular activities) use the energy of ATP hydrolysis to drive conformational changes in diverse macromolecular targets. Here, we report the biochemical characterization and 2.5-Å crystal structure of a *Mycobacterium smegmatis* AAA protein Msm0858, the ortholog of *Mycobacterium tuberculosis* Rv0435c. Msm0858 is a magnesium-dependent ATPase and is active with all nucleoside triphosphates (NTPs) and deoxynucleoside triphosphates (dNTPs) as substrates. The Msm0858 structure comprises (i) an N-terminal domain (amino acids [aa] 17 to 201) composed of two β-barrel modules and (ii) two AAA domains, D1 (aa 212 to 473) and D2 (aa 476 to 744), each of which has ADP in the active site. Msm0858-ADP is a monomer in solution and in crystallized form. Msm0858 domains are structurally homologous to the corresponding modules of mammalian p97. However, the position of the N-domain modules relative to the AAA domains in the Msm0858-ADP tertiary structure is different and would impede the formation of a p97-like hexameric quaternary structure. Mutational analysis of the A-box and B-box motifs indicated that the D1 and D2 AAA domains are both capable of ATP hydrolysis. Simultaneous mutations of the D1 and D2 active-site motifs were required to abolish ATPase activity. ATPase activity was effaced by mutation of the putative D2 arginine finger, suggesting that Msm0858 might oligomerize during the ATPase reaction cycle. A truncated variant Msm0858 (aa 212 to 745) that lacks the N domain was characterized as a catalytically active homodimer.

IMPORTANT

Recent studies have underscored the importance of AAA proteins (ATPases associated with various cellular activities) in the physiology of mycobacteria. This study reports the ATPase activity and crystal structure of a previously uncharacterized mycobacterial AAA protein, Msm0858. Msm0858 consists of an N-terminal β-barrel domain and two AAA domains, each with ADP bound in the active site. Msm0858 is a structural homolog of mammalian p97, with respect to the linear order and tertiary structures of their domains.

Members of the AAA protein family (ATPases associated with various cellular activities) are prevalent in all taxa. They comprise a subdivision of the P-loop nucleoside triphosphate (NTP) phosphohydrolase superfamily, named for the GXGK (T/S) motif (also known as an A-box) that engages the NTP β- and γ-phosphates. AAA proteins are composed of one or two AAA domains, a conserved fold that defines the family. The AAA domain consists of two modules: an N-terminal α/β-subdomain (wherein resides the P-loop), and an α-helical C-terminal subdomain (1, 2). As the acronym implies, AAA proteins have diverse biological functions embracing protein quality control, proteolysis, membrane fusion, DNA replication and recombination, transcription, etc. (1, 2). Many of the well-studied AAA proteins form homo-oligomeric or hetero-oligomeric assemblies. Additional protein-protein interactions may spatiotemporally regulate AAA protein activities (3).

Recent studies have highlighted the importance of AAA proteins in the physiology of *Mycobacterium*, the genus that includes the human pathogen *M. tuberculosis* and its avirulent cousin *M. smegmatis*. The mycobacterial AAA proteins that have been characterized biochemically and/or genetically include (i) the essential zinc metalloprotease FtsH (4, 5); (ii) the proteasome ATPase Rv2115c, which is required for virulence (6, 7); (iii) the essential ClpC1 and ClpX protein “unfoldases” that are promising therapeutic targets for tuberculosis (8–10); and (iv) the Rv3870, Rv3871, and Rv3868 components of the ESX-1 system, which secretes virulence factors (11, 12). Mycobacterial FtsH, Rv2115c, Rv3868, and ClpX are single-AAA-domain polypeptides; ClpC1 and Rv3871 have two tandem AAA domains.

In the present study, we focused on a previously uncharacterized *M. smegmatis* AAA family protein, Msm0858. The 745-­amino-acid Msm0858 polypeptide and its 728-amino-acid *M. tuberculosis* ortholog Rv0435c share 543 positions of amino acid side chain identity. Msm0858 and Rv0435c include two tandem AAA-like domains. Our interest in these proteins was sparked by two factors. First, Rv0435c was identified in a genome-wide yeast two-hybrid screen for potential interactors with *M. tuberculosis* DNA
ligase D, a central agent of the mycobacterial nonhomologous end-joining pathway of DNA double-strand break repair (13; N. Stephanou and M. Glickman, unpublished data). Second, Rv0435c and Msm0858 are candidate mycobacterial homologs of eukaryal p97 (also known as valosin-containing protein [VCP], or Cdc48), which is one of the most extensively studied AAA family ATPases, structurally and biochemically (14–20). p97 is of acute interest in light of its diverse cellular functions (especially in protein quality control and protein mobilization from large assemblies), its two-tiered homohexameric quaternary structure, its many macromolecular interactions, and the fact that human p97 mutations cause protein aggregation disease syndromes (3).

We report that purified recombinant Msm0858 is a monomeric protein in solution. It has intrinsic magnesium-dependent ATPase activity, and it is broadly active with NTPs and deoxyribo nucleoside triphosphates (dNTPs) as phosphohydrolase substrate. The 2.5-Å crystal structure of Msm0858 highlights (i) tandem D1 and D2 AAA domains, each of which has ADP in the phosphohydrolase site, and (ii) an N-terminal domain comprising two discrete β-barrel modules. Whereas the folds of the individual Msm0858 N-terminal and AAA domains resemble the corresponding modules of mammalian p97, the relative positions of the domains in the tertiary structures differ substantially. Simultaneous mutations of the active-site motifs of the tandem D1 and D2 domains are required to efface ATPase activity. These studies establish a foundation for interrogating the role of Msm0858/Rv0435c in mycobacterial physiology.

MATERIALS AND METHODS
Recombinant Msm0858. The open reading frame encoding full-length Msm0858 was PCR amplified from M. smegmatis genomic DNA with primers that introduced a BglII site at the start codon and immediately downstream of the stop codon. The PCR product was digested with BglII and ligated into pET28b-His10Smt3 that had been digested with BamHI. The resulting expression plasmid encodes the Msm0858 polypeptide. The concentration of each Msm0858 protein was determined by SDS-PAGE analysis of protein aliquots (5.8 mg/ml kanamycin and 0.05 mg/ml DL-SeMet (Acros Organics) until the A600 reached 0.5, adjusted to 0.5 mM IPTG and 2% ethanol, and incubated for 20 h at 17°C. The purification was performed as described above for the native protein.

ATPase assay. Reaction mixtures (10 µl) containing 40 mM Tris-HCl (pH 8.0), 20 mM MgCl2, 2 mM [α-32P]ATP, and Msm0858 were incubated at 37°C. The reactions were quenched by adding 2 µl of 5 N formic acid. An aliquot of the mixture was applied to a polyethyleneimine-cellulose thin-layer chromatography (TLC) plate, which was developed with 0.45 M ammonium sulfate. [32P]ATP and [32P]ADP were quantified by scanning the TLC plate with a Fujix BAS2500 imager. ATPase activity (expressed as nanomoles of ADP formed) was calculated from the percent conversion of [32P]ATP to [32P]ADP (i.e., ADP/[ADP + ATP]) in a 10-µl reaction mixture containing 20 nmol input ATP.

NTPase assay. Reaction mixtures (10 µl) containing 40 mM Tris-HCl (pH 8.0), 20 mM MgCl2, 38 pmol Msm0858, and 2 mM NTP (ATP, GTP, or CTP) or dNTP (dATP, dGTP, dCTP, or dTTP) substrate were incubated at 4 °C prior to crystallization. Crystals of native Msm0858 were grown by the hanging drop vapor diffusion method at room temperature by mixing the protein (2 µl) with an equal volume of 0.3 M ammonium tartrate dibasic and 25% polyethylene glycol 3350 (PEG 3350). Crystals grew to their full size within a week. Crystals of SeMet-Msm0858 were recovered in the flowthrough fraction. The total protein concentration was measured with the Bio-Rad dye reagent using bovine serum albumin (BSA) as the standard. The yield of wild-type Msm0858 was ~14 mg from a 300-ml bacterial culture. The concentrations of the wild-type and mutant Msm0858 polypeptides were measured by parallel SDS-PAGE analysis of protein aliquots (5.8 µg) and scanning of the Coomassie blue-stained gel with a Kodak ImagePro 4000R instrument. The Image Gauge software was then used to quantify the purity with respect to the Msm0858 polypeptide. The concentration of each Msm0858 protein (versus total protein, as determined by dye binding) was adjusted accordingly.

Selenomethionine (SeMet)-substituted Msm0858 was purified from a 2-liter culture of E. coli B834 that had been transformed with the expression plasmid, grown at 37°C in modified M9 medium containing 0.05 mg/ml kanamycin and 0.05 mg/ml DL-SeMet (Acros Organics) until the A600 reached 0.5, adjusted to 0.5 mM IPTG and 2% ethanol, and incubated for 20 h at 17°C. The purification was performed as described above for the native protein.

Crystallization. Native and SeMet-substituted Msm0858 were concentrated by centrifugal ultrafiltration to 6.5 mg/ml and dialyzed against 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 10 mM dithiothreitol (DTT) for 3 h at 4°C prior to crystallization. Crystals of native Msm0858 were grown by the hanging drop vapor diffusion method at room temperature by mixing the protein (2 µl) with an equal volume of 0.3 M ammonium tartrate dibasic and 25% polyethylene glycol 3350 (PEG 3350). Crystals grew to their full size within a week. Crystals of SeMet-Msm0858 were grown under identical conditions after streak-seeding of the drops using a thin fiber that had been dipped in a suspension of native Msm0858 microcrystals that were pulverized by vortexing with glass beads. Single crystals were harvested and cryoprotected by suspending in 0.1 M ammonium tartrate dibasic, 25% PEG 3350, and 15% ethylene glycol and then flash-frozen in liquid nitrogen.

Diffraction data collection. Data for a single native Msm0858 crystal were collected at the Advanced Photon Source (APS) using beamline 24-ID-C equipped with an ADSC-Q315 charged-coupled-device (CCD) detector using 200 consecutive 1° oscillations. The crystal contained an orthorhombic lattice, in space group P212121, and diffracted X rays to a maximum resolution of 2.5 Å. Data for a single SeMet-Msm0858 crystal were obtained at the National Synchrotron Light Source (NSLS) using beamline X12-C equipped with an ADSC-Q210 CCD detector. The SeMet-Msm0858 data consisted of two 370° sweeps collected using two SeMet-Msm0858 crystals that were pulverized by vortexing with glass beads. Single crystals were harvested and cryoprotected by suspending in 0.1 M ammonium tartrate dibasic, 25% PEG 3350, and 15% ethylene glycol and then flash-frozen in liquid nitrogen.

Structure determination. Phases for the SeMet-Msm0858 crystal structure were obtained using an iterative process of single-wavelength
TABLE 1 Crystallographic data and refinement statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value(s) or data for:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection statistics</td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Unit cell dimensions at 130 K (a, b, c) (Å)</td>
<td>88.36, 109.26, 154.95</td>
</tr>
<tr>
<td>Radiation source</td>
<td>NSLS-X12C λ = 0.979 Å</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>40.0–3.2 (3.26–3.2)</td>
</tr>
<tr>
<td>Rsym (%)</td>
<td>14.7 (53.7)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>46,746 (2,226)</td>
</tr>
<tr>
<td>Mean redundancy</td>
<td>9.1 (4.5)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.1 (93.3)</td>
</tr>
<tr>
<td>Mean I/σI</td>
<td>11.1 (2.4)</td>
</tr>
<tr>
<td>Experimental phasing data</td>
<td></td>
</tr>
<tr>
<td>Phasing method</td>
<td>SeMet SAD/partial structure</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>40.0–3.2</td>
</tr>
<tr>
<td>Anomalous signal</td>
<td>0.04 (all data)/0.23 (to 5.5 Å)</td>
</tr>
<tr>
<td>Partial structure</td>
<td>460 alanine residues via BRUTEPTF</td>
</tr>
<tr>
<td>FOM (SOLVE/RESOLVE/PHASER)</td>
<td>0.16/0.83/0.79</td>
</tr>
<tr>
<td>Δφ (SOLVE/RESOLVE/PHASER) (°)</td>
<td>82.5/75.0/55.0</td>
</tr>
<tr>
<td>Refinement statistics (I &gt; σI)</td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>29.5–2.50 (2.55–2.50)</td>
</tr>
<tr>
<td>Rw/Refree</td>
<td>0.206/0.251 (0.290/0.338)</td>
</tr>
<tr>
<td>Estimated coordinate error</td>
<td>0.71 Å</td>
</tr>
<tr>
<td>Model statistics</td>
<td></td>
</tr>
<tr>
<td>Model contents</td>
<td>1,423 residues in two protomers, 3 glycerol, 2 triethylene glycol, 1 tetraethylene glycol, 4 ADP, and 332 water molecules</td>
</tr>
<tr>
<td>RMSD for bond length (Å)/bond angle (°)</td>
<td>0.003/0.77</td>
</tr>
<tr>
<td>NCS RMSD (Å)</td>
<td>0.48</td>
</tr>
<tr>
<td>Ramachandran favored regions (%)/outliers</td>
<td>97.9/none</td>
</tr>
<tr>
<td>B-factors (Å²): Wilson/overall</td>
<td>37.8/49.7</td>
</tr>
<tr>
<td>Heteroatoms</td>
<td>47.2</td>
</tr>
<tr>
<td>TLS anisotropy</td>
<td>0.28 (10 groups)</td>
</tr>
</tbody>
</table>

* Standard definitions are used for all of the parameters. Values in parentheses refer to data for the highest-resolution bin. The data collection statistics come from SCALEPACK, and the refinement and geometric statistics come from PHENIX.

 anomalous diffraction (SAD) maximum-likelihood phasing, in conjunction with phases derived from partial-structure information. Recombinant Msm0858 contained 10 methionine residues per protomer. Based on a solvent content of 50%, the asymmetric unit of the orthorhombic crystals was expected to contain two protomers and a total of 20 selenium sites. Consistent with this prediction, analysis of the native Patterson map revealed a pseudotranslation vector of (0, 0.5, 0.26). Coordinates for selenium positions were obtained using anomalous difference Patterson methods implemented in SOLVE (21), with a resolution cutoff of 5.5 Å, which found 18 candidate selenium sites, 12 of which obeyed the expected translational noncrystallographic symmetry. Despite the promising nature of this heavy-atom solution, the resulting electron density maps were not interpretable, even after inverting the selenium site pattern and recalculating phases. However, density modification in RESOLVE (22) of the electron density calculated using the inverted heavy-atom coordinates, aided by the use of noncrystallographic symmetry, revealed a clear solvent envelope and regions of density with the rough appearance of protein secondary structure.

Because these initial maps were not of sufficient quality to trace a polypeptide, an initial protein model was assembled with the aid of real-space molecular replacement, as follows. Four copies of the closest sequence homolog for which a structure is known, residues 240 to 355 of p97 from Mus musculus (Protein Data Bank [PDB] ID 1E32), were first truncated to polyalanine and then placed into the electron density map using BRUTEPTF (23). Once placed into the electron density map, the fold of these minimal ATPase domains matched up well with the crude electron density in the experimentally phased maps, particularly in the helices bracketing the core β-sheet of each AAA domain. This model, consisting of 460 alanine residues, was edited manually in COOT (24) so as to better position strands and helices into the electron density map via rigid body rotation and translation. Loops not matching the electron density map were removed. Additional polyalanine strands and helices in regions outside the AAA core domain were added, as justified by the density. Once assembled, this model then served as a source of partial structure phase information, in conjunction with SAD phasing, using the previously determined site pattern, as implemented in the CCP4 version of PHASER (25). The resulting electron density maps were much improved with respect to the original maps derived exclusively from experi-
mental phases; they revealed interpretable density for the N-terminal β-domains and helical extensions of the AAA domains. Sequence assignment was aided by sharp peaks in an anomalous difference Fourier map calculated using only experimental (not model) phases corresponding to the selenium atoms of SteMet residues. The improved PHASER maps were used to build an initial model in COOT, which was refined in PHENIX (26). Each refined model was then reused for partial structure information in PHASER, and the resulting improved electron density maps were used to further improve the protein model. After refining the SteMet-Msm0858 model to an R-factor of <0.40, it was used as a starting model for molecular replacement for the nonisomorphic native Msm0858 data set. (The protein components within the two lattices are related by a rigid body translation of 5.8 Å and a rotation of 2.6°.) This molecular replacement solution was then rebuilt using PHENIX.autobuild, further edited by hand in COOT, and finally refined using PHENIX to R/Rfree values of 0.206/0.251, with excellent geometry.

The mean figures of merit (FOM) and phase differences (Δφ) between the phases derived from the final refined native Msm0858 model (with respect to the SteMet-Msm0858 data set) and the phases as determined by SOLVE (inverted site pattern), RESOLVE (with noncrystallographic symmetry averaging), and PHASER (with partial structure information derived from real-space molecular replacement, followed by a second round of density modification) are given in Table 1. Δφ is defined as the absolute difference between calculated and observed phases calculated on a common origin (|φcalc − φobs|) via CCP4-PHASE-MATCH, where φobs refers to phases output from the program indicated, and φcalc values refer to model phases calculated from the final refined model placed into the SteMet-Msm0858 unit cell.

RESULTS
Recombinant Msm0858 is a monomer in solution. To evaluate the biochemical properties of Msm0858, we produced the protein in E. coli as a His10-Smt3 fusion and isolated it from a soluble extract by nickel-agarose chromatography. The His10-Smt3 tag was removed with the Smt3-specific protease Ulp1, and the native Msm0858 protein was separated from the tag by a second round of Ni-agarose chromatography. The quaternary structure of Msm0858 was examined by zonal velocity sedimentation in a 15 to 30% glycerol gradient (Fig. 1A). The marker proteins catalase (native size, 248 kDa; a homotetramer of a 62-kDa polypeptide), ovalbumin (45-kDa monomer), and cytochrome c (12-kDa monomer) were included as internal standards in the gradient. The gradient was centrifuged for a relatively short interval (7 h) to permit the detection of potential oligomeric forms of Msm0858. The polypeptide compositions of the odd-numbered gradient fractions were analyzed by SDS-PAGE. The Coomassie blue-stained gel is shown. The identities of the polypeptides are shown on the right. Msm0858 alone (700 μg in 0.2 ml) was applied to a 4.8-ml 15 to 30% glycerol gradient containing 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 1 mM DTT, 1 mM EDTA, and 0.05% Triton X-100. The gradient was centrifuged at 50,000 rpm for 7 h at 4°C in a Beckman SW55Ti rotor. Fractions (~0.19 ml) were collected from the bottom of the tube. Aliquots (20 μl) of odd-numbered fractions were analyzed by SDS-PAGE. The Coomassie blue-stained gel is shown. The identities of the polypeptides are shown on the right. The Msm0858 alone (700 μg in 0.2 ml) was applied to a 4.8-ml 15 to 30% glycerol gradient and centrifuged at 50,000 rpm for 17 h at 4°C in a Beckman SW55Ti rotor. Aliquots (20 μl) of odd-numbered gradient fractions were analyzed by SDS-PAGE. The Coomassie blue-stained gel is shown in the top panel. The bottom panel shows the ATPase activity profile of the same gradient. Reaction mixtures (10 μl) containing 40 mM Tris-HCl (pH 8.0), 20 mM MgCl2, 2 mM [α-32P]ATP, and 4 μl of the indicated gradient fractions were incubated for 30 min at 37°C.

mM Mg2+, we estimated a turnover number of 30 ATPs min−1 enzyme−1. When various metals were tested at a concentration of 20 mM, the cofactor requirement for ATP hydrolysis was satisfied best by magnesium. Calcium and manganese were 79% and 32% as effective, respectively, as magnesium (Fig. 2B). Copper, cadmium, cobalt, nickel, and zinc were ineffective (Fig. 2B). ATP hydrolysis displayed a nonlinear (sigmoidal) dependence on input Msm0858 enzyme (Fig. 2C). (It is conceivable that initially monomeric Msm0858 undergoes an oligomerization step under the
From the extent of ATP hydrolysis by 19 pmol Msm0858, we estimated a turnover number of 28 ATPs min⁻¹ enzyme⁻¹. The time course of ATP hydrolysis by 3.8 μM Msm0858 (Fig. 2D) revealed a linear accumulation of ADP product from 2 to 15 min, at a rate of 33 ATPs min⁻¹ enzyme⁻¹. After a 20-min reaction, 93% of the input ATP was converted to ADP (Fig. 2D). NTP substrate specificity was examined by colorimetric assay.
of the release of Pi from unlabeled ribonucleotide ATP, GTP, CTP, or UTP and deoxynucleotide dATP, dGTP, dCTP, or dTTP, each at a 2 mM concentration (Fig. 2E). Msm0858 was capable of hydrolyzing each of the NTP and dNTP substrates.

**Crystallization of Msm0858 and structure determination.** A native Msm0858 crystal (grown as described in Materials and Methods) was in space group P2₁2₁2₁ and diffracted X rays to 2.5 Å resolution. A SeMet-Msm0858 crystal in a nonisomorphic orthorhombic lattice diffracted to 3.2 Å. Phases for the SeMet-Msm0858 crystal structure were obtained via SAD methods in conjunction with phases derived from partial structure information for the archetypal AAA protein murine p97. A SeMet-Msm0858 model (refined to an $R$ factor of 0.40) was used to solve the native Msm0858 structure by molecular replacement. The native Msm0858 model, comprising two protomers in the asymmetric unit, was refined to $R$/$R_{free}$ of 0.206/0.251, with excellent geometry (Table 1). Protomer A consisted of three polypeptide segments from amino acids Thr17 to Asp151, Ser159 to Ala473, and Glu476 to Lys744. Protomer B comprised four polypeptide segments, from Pro18 to Asp151, Ser159 to Gly200, Pro210 to Arg593, and Gly601 to Lys744. The A and B protomers superimposed with a root mean square deviation (RMSD) of 0.5 Å at 702 Ca positions. Electron density maps revealed two molecules of ADP bound to each protomer. Because the protein was not exposed to exogenous nucleotides during purification or crystallization, we surmise that ADP had bound stably to Msm0858 during its production in *E. coli*. Analysis using PISA (27) calculated that the protomer interface in the crystal comprised 1,136 Å² of buried surface area of protomer B (out of 29,292 Å² total surface area) and 1,141 Å² of buried surface area of protomer A (out of 30,364 Å² total surface area), with a complex formation significance (CFS) score of 0.026 (scores can range from 0 to 1 as interface relevance to complex formation increases). The CFS score for Msm0858 implies that the interface between the A and B protomers is not significant for complex formation and may be solely a result of crystal packing, i.e., Msm0858 crystallized as a monomer. The descriptions and depictions of the Msm0058 structure below will refer to the A protomer.

**Overview of the Msm0858 structure.** A stereo view of the Msm0858 tertiary structure is shown in Fig. 3A. The protein comprises an N-terminal domain (N) with two $\beta$-barrel lobes (green and yellow), followed by two AAA domains, D1 and D2, each of which is composed of a proximal $\alpha/\beta$-fold module (magenta for D1 and blue for D2) and a distal $\alpha$-helical module (pink for D1 and cyan for D2). ADP nucleotides in the D1 and D2 active sites are depicted as stick models. (B) Stereo view of the p97 fold (from PDB ID 3CF3), color coded by domain as described for panel A.

![Figure 3](https://example.com/figure3.png)
The D1 and D2 AAA domains are each composed of a proximal \(\alpha/\beta\)-fold module (aa 212 to 391, colored magenta for D1, and aa 476 to 653, colored blue for D2, in Fig. 3A and 5) and a distal \(\alpha\)-helical module (aa 392 to 473, colored pink for D1, and aa 654 to 744, colored cyan for D2, in Fig. 3A and 5). The proximal AAA modules are built around a central five-stranded parallel \(\beta\)-sheet (supplemented in D1 by an additional antiparallel \(\beta\)-strand at the edge of the sheet). The distal AAA modules consist of four \(\alpha\)-helices common to D1 and D2; D2 has an additional \(\alpha\)-helix at its C terminus (Fig. 5). The secondary-structure elements of the tandem AAA domains are indicated above the Msm0858 amino acid sequence in Fig. 6. The A-box and B-box catalytic motifs are highlighted in yellow and green shading, respectively. The Msm0858 D1 and D2 domains superimposed in PyMOL with an RMSD of 2.5 Å at 220 C\(_\text{α}\) positions (Fig. 5).

The D1 and D2 domains of Msm0858 are structurally quite similar to p97 D1 and D2 (shown in gray in Fig. 5, superimposed in PyMOL on their Msm0858 counterparts, and then offset horizontally). The D1 domains of Msm0858 and p97 superimposed with an RMSD of 1.5 Å at 163 C\(_\text{α}\) positions. The D2 domains superimposed with an RMSD of 2.4 Å at 232 C\(_\text{α}\) positions. Alignment of the primary structures of the tandem D1 and D2 domains

FIG 4 N domain of Msm0858. (A) Aligned tertiary structures of the Msm0858 N domain (with modules colored as described in the legend to Fig. 3) and the p97 N domain (gray). The N- and C-terminal amino acids are indicated, as are the amino acids bordering a gap in the Msm0858 polypeptide chain. An \(\alpha\)-helix unique to Msm0858 is denoted by an asterisk. (B) Aligned primary and secondary structures of the Msm0858 and p97 N domains. Positions of amino acid side chain identity/similarity are denoted by dots above the alignment. Gaps in the alignment are denoted by dashes. Secondary structure elements (arrows for \(\beta\)-strands and cylinders for \(\alpha\)-helices) are displayed above and below the amino acid sequences. The secondary structure elements are colored as described for panel A.
highlights 181 positions of amino acid identity and 82 positions of side chain similarity between Msm0858 and p97 (Fig. 6).

**Distinct domain arrangements in Msm0858 versus p97.** Figure 3 provides a comparison of the tertiary structures of the ADP-bound forms of Msm0858 and p97 (PDB ID 3CF3) (16), superimposed with respect to their D2 AAA domains. In both structures, the D1 AAA domain is roughly parallel and to the right of D2 (i.e., the cyan and pink α-helical modules are situated above the blue and magenta α/β-modules in the views shown in Fig. 3). However, the D1 domains differ in their angular positions, such that the α-helical module of Msm0858 D1 is rotated forward (out of the plane of the image in Fig. 3A) relative to the recessed rotation of the p97 α-helical module (into the plane in Fig. 3B). This angular displacement is attributable to a drastic difference in the position of the N domains in the two protein structures. In p97, the N modules decorate the outer circumference of the D1 tier of
Msm0858 N module (asterisk in Fig. 4) lines up antiparallel to the (Fig. 3A). The proximal N module (green) is wedged between the D1, where they are packed against the D1 and D2 domains of p97. The p97 N domain makes no contact with the D2 AAA domain abuts D2, while the N-domain module (Fig. 3B). The proximal p97 N-domain /H9251 closes barrel module (yellow) touches the D1 /H9252 A-box module. (A) Stereo image of the full p97 hexamer (in gray) viewed down the central axis of the ring, with the bottom D2 tier facing forward. The p97 N-domain modules decorate the outer circumference of the D1-D2 hexamer. As aligned, the Msm0858 N-domain modules (in green and yellow) sterically clash with the p97 interprotomer interfaces. the stacked D1-D2 hexamer ring (Fig. 7B), such that the N domain is located on the forward face of the D1 AAA domain in the view shown in Fig. 3B. The proximal p97 N-domain β-barrel module (green) touches the D1 α/β-module, and the distal N-domain β-barrel module (yellow) touches the D1 α-helical module (Fig. 3B). The p97 N domain makes no contact with the D2 AAA domain (Fig. 3B).

In contrast, the Msm0858 N-domain modules are located behind D1, where they are packed against the D1 and D2 domains (Fig. 3A). The proximal N module (green) is wedged between the D1 and D2 α-helical modules, such that first α-helix of the N module abuts D2, while the β-barrel abuts D1 (Fig. 3A). The distal N-terminal module (yellow) is packed against the D1 α/β-module and the D1 α-helical module. The distinctive α-helix of the Msm0858 N module (asterisk in Fig. 4) lines up antiparallel to the third α-helix of the D1 α-helical module (Fig. 3A). The positions of the N-terminal modules with respect to the AAA domains that we see in the Msm0858 crystal structure would preclude the formation of a p97-like stacked hexameric quaternary structure, plausibly accounting for the monomeric state of Msm0858 in solution, as gauged by velocity sedimentation. This is illustrated in Fig. 7A, in which the Msm0858 protomer (colored by domain as in Fig. 3A) is superimposed with respect to the D2 AAA domain of a gray p97 protomer of the p97 hexameric ring (20). The adjacent p97 protomer of the ring is colored beige. The top tier of the p97 ring is stabilized by D1-D1 interactions and the bottom tier by D2-D2 interactions. The alignment places the Msm0858 N-domain modules (in green and yellow) so that they overlap and sterically block both of the p97 cross-protomer interfaces on the outer circumference of the ring. Figure 7B shows the superposition of the Msm0858 protomer in the context of the full p97 hexamer, viewed down the central axis of the ring, with the bottom tier facing forward.

**Phosphohydrolase sites in the Msm0858 AAA domains.** The AAA family active site typically comprises (i) a P-loop (A-box) motif, GXGXXGKTS, which engages the NTP phosphates via the lysine Nɛ and main-chain amides and the metal cofactor via threonine; (ii) a B-box motif, DEXD, which engages the metal cofactor and the water nucleophile; (iii) a sensor asparagine that positions the water nucleophile relative to the NTP γ-phosphate; and (iv) an arginine finger that coordinates the γ-phosphate (1, 2, 19). In oligomeric AAA proteins, the arginine finger in the active site is provided by the adjacent AAA protomer. The D1 and D2 domains of p97 both have consensus active-site motifs: the A and B boxes are shaded yellow and green in Fig. 6. The p97 sensor asparagines (Asn348 in D1 and Asn624 in D2) and the arginine fingers (Arg359 in D1 and Arg635 in D2) are indicated by arrowheads below the p97 sequence. The sequence alignment highlights that the catalytic motifs are present and conform to consensus in the Msm0858 D2 domain (Fig. 6). However, the Msm0858 D1 has deviant or missing motifs. Msm0858 D1 is missing the A-box threonine and the first B-box aspartate, and it lacks the sensor asparagine and the arginine finger; in each case, the position is occupied by alanine (Fig. 6).

The ADP binding site of the Msm0858 D2 domain is shown in Fig. 8A; the ADP site in the D1 domain is shown in Fig. 8B in a similar orientation. The adenosine nucleosides are in an anti conformation. There are no side chain contacts to the adenine nucleobase, nor are there any contacts to the ribose hydroxyls, which likely accounts for the lack of strict nucleobase or pentose sugar specificity of the Msm0858 NTP phosphohydrolase activity. In both nucleotide sites, the A-box lysine (Lys532 in D2 and Lys276 in D1) contacts the ADP β-phosphate. As is typical of Β-loop proteins, the main-chain amides of the loop and the ensuing α-helix (Gly529, Cys530, Gly531, Thr533, and Phe534 in D2; Gly273, Gly275, Lys276, Ala277, and Thr278 in D1) donate hydrogen bonds to the ADP α- and β-phosphate oxygens. The Thr278 hydroxyl donates an additional hydrogen bond to the α-phosphate in the D1 domain (Fig. 8B). The corresponding residue in the D2 domain is Phe534, which is packed against the ribose sugar (Fig. 8B).

The B-box acidic residues (Asp585, Glu586, and Asp588 in D2; Asp327 and Asp329 in D1) are near the bound α-phosphate oxygens. The corresponding residue in the D2 domain is Asp327, which is packed against the ribose sugar (Fig. 8B). The putative D2 Arg641 arginine finger is located on the protein surface 26 Å away from the ADP.

**Effects of A-box and B-box mutations on ATP hydrolysis.** The A-box lysines Lys276 in D1 and Lys532 in D2 were mutated individually to alanine, and the recombinant mutant proteins (Fig. 9A) were assayed for ATPase activity (Fig. 9B). The K276A and K332A mutants were each 50% as active as wild-type Msm0858. In light of these results, we simultaneously changed both A-box lysines to alanine. The recombinant K276A-K332A double mutant (Fig. 9A) was inert in ATP hydrolysis (Fig. 9B). In mutating the B-box (326ADVD329 in domain D1, 586DEID588 in
domain D2), we introduced alanine in lieu of Glu585 in domain D2 and its counterpart Asp327 in domain D1, singly and pairwise (Fig. 9A). Whereas the D327A and E585A single mutants retained ATPase activity, at 60% and 80% of the wild-type level, respectively, the D327A-E586A double mutant was inert in ATP hydrolysis (Fig. 9B). These results suggest that (i) the D1 and D2 AAA domains of Msm0858 are both catalytically active, and (ii) ATP hydrolysis by D2 is not contingent on an intact active site in D1, and vice versa.

Effect of deleting the N-domain modules. If Msm0858 does form a homo-oligomeric quaternary structure during the ATPase reaction cycle, and if multimerization is necessary for ATP hydrolysis, then activation of the enzyme would require movement of the N-domain modules away from their position in the Msm0858 crystal structure in order to expose a multimerization surface. To address this issue, we deleted the N-domain modules (aa 1 to 211) and produced the truncated Msm0858 (aa 212 to 745) protein (herein referred to as ΔN) in E. coli in parallel with full-length wild-type Msm0858. The recombinant tag-free proteins were purified by serial Ni affinity chromatography steps, as described above, followed by an additional Superdex 200 gel filtration step. SDS-PAGE affirmed the purity and expected sizes of the wild-type Msm0858 crystal structure in order to expose a multimerization surface. To address this issue, we deleted the N-domain modules (aa 1 to 211) and produced the truncated Msm0858 (aa 212 to 745) protein (herein referred to as ΔN) in E. coli in parallel with full-length wild-type Msm0858. The recombinant tag-free proteins were purified by serial Ni affinity chromatography steps, as described above, followed by an additional Superdex 200 gel filtration step. SDS-PAGE affirmed the purity and expected sizes of the wild-type (WT) (78-kDa) and ΔN (57-kDa) polypeptides (Fig. 10A). Enzyme titration experiments showed that the ΔN mutant had a more active ATPase than the WT (Fig. 10B). ATP hydrolysis by both ΔN and WT Msm0858 proteins displayed a sigmoidal dependence on the input enzyme. From the extents of ATP hydrolysis by 3 pmol ΔN mutant and 5 pmol WT Msm0858, we estimated turnover numbers of 300 and 145 ATPs min⁻¹ enzyme⁻¹, respectively. (The addition of the gel filtration purification step resulted in a more active WT enzyme preparation vis-à-vis the experiments shown in Fig. 2C. It is conceivable that gel filtration removed prebound ADP product from one or both of the Msm0858 active sites.) The key conclusions from the deletion analysis are that (i) the N-domain modules are not necessary for ATPase activity and (ii) their presence exerts a modest negative influence on catalysis.

The quaternary structure of the ΔN mutant was examined by sedimentation for 12 h in a 15 to 30% glycerol gradient, along with internal standards aldolase (native size, 160 kDa; a tetramer of a
40-kDa polypeptide), BSA (a 66-kDa monomer), ovalbumin (a 45-kDa monomer), and cytochrome c (a 12-kDa monomer). Figure 10C shows SDS-PAGE analysis of the polypeptide composition of the protein mixture loaded onto the gradient (lane L) and of gradient fractions 12 to 25, which contained all the size-resolved polypeptides. No polypeptides were detected in fractions 1 to 11, which sedimented heavier than the 160-kDa aldolase standard (not shown); thus, deleting the N-domain modules of Msm0858 did not trigger p97-like hexamerization. The instructive finding from the sedimentation experiment was that the peak fraction of the 57-kDa N/H9004 polypeptide (in gradient fraction 16) was heavier than the 66-kDa BSA peak (in fractions 17 and 18) but lighter than the 160-kDa aldolase peak (in fraction 13), consistent with the majority of the N/H9004 preparation being a homodimer. However, the distribution of the N/H9004 polypeptide revealed an asymmetric shoulder on the lighter side of the peak, between the 66-kDa BSA and 45-kDa ovalbumin markers, consistent with a minority fraction of the N/H9004 preparation being a monomer. This asymmetry was also evident in the ATPase activity profile of the N/H9004 glycerol gradient (Fig. 10C). We conclude that subtraction of the Msm0858 N-domain modules facilitates dimerization of the C-terminal tandem AAA domains. That this was a specific attribute of the N/H9004 protein was verified by an additional control experiment in which full-length WT Msm0858 was analyzed by glycerol gradient centrifugation under the same conditions with the same set of internal standards. To wit, the 78-kDa Msm0858 polypeptide sedimented with the leading edge of the 66-kDa BSA peak (i.e., as an Msm0858 monomer), and no Msm0858 polypeptide was detected cosedimenting with the aldolase peak (the position expected for an Msm0858 dimer) or in fractions heavier than aldolase.

Effect of D2 arginine finger mutation on ATP hydrolysis. Arg641 is the predicted arginine finger of the D2 AAA domain based on the structural alignment of Msm0858 to p97 (Fig. 6). We mutated Arg641 to alanine, purified the recombinant R641A protein in parallel with wild-type Msm0858 (Fig. 11A), and assayed them for ATPase activity. R641A was virtually inert in ATP hydrolysis at a level of input protein that sufficed for hydrolysis of 95% of the ATP substrate by wild-type Msm0858 (Fig. 11B). Because Arg641 is situated far away from the ADP nucleotides in the monomeric Msm0858-ADP structure, we infer that Msm0858 is likely to undergo a structural change in the ATP-bound state (presumably oligomerization) that enables the D2 arginine finger to coordinate an ATP γ-phosphate.

DISCUSSION

Here, we report the NTP phosphohydrolase activity of a mycobacterial AAA protein, Msm0858, and a crystal structure of Msm0858 bound to ADP. Our findings highlight that Msm0858 is a struc-
Incubating purified Msm0858 with 20 mM Mg2+ and ATP-PNP. In the same vein, preincubation of the ΔN protein with 20 mM Mg2+ and 2 mM AMP-PNP did not change its sedimentation behavior versus what we report in Fig. 10C. The key point is that we saw no p97-like hexamerization of either Msm0858 or the ΔN protein in the presence of Mg2+/AMP-PNP. Thus, the hexamerization question, and the issue of homodimerization of full-length Msm0858, remains open until conditions can be found that promote stabilization and capture of higher-order Msm0858 quaternary structures.

An initial mutational analysis of A-box and B-box amino acids in the D1 and D2 domains of Msm0858 suggests that both AAA domains are active for ATP hydrolysis, apparently independently, insofar as alanine substitutions for the P-loop lysine in D1 or D2 reduce activity by half, but subtraction of both lysines simultaneously abolishes ATPase activity. The alanine mutations in the Msm0858 D1 and D2 B-boxes behave similarly.

Our studies here expand the roster of bona fide mycobacterial AAA ATPases and provide a foundation to help address the role of Msm0858 in mycobacterial physiology. Genome-wide transposon mutagenesis revealed that the homologous Rv0435c AAA protein is not essential for growth of M. tuberculosis under laboratory conditions (28, 29). To gain clues to cellular processes in which these AAA proteins participate, it would be fruitful to identify interacting proteins, either by purifying Msm0858 from the native mycobacterial source or performing genome-wide interaction screens using Msm0858 as bait.

Finally, a BLAST search of bacterial proteomes with Msm0858 revealed full-length homologous proteins in multiple species of the genus Mycobacterium (which belongs to the phylum Actinobacteria) and in multiple other genera of Actinobacteria (e.g., Rhodococcus, Nocardia, Gordonia, Tsukamurella, Dietzia, Glycomyces, Streptomyces, Micromonospora, Actinosynnema, etc.). Full-length Msm0858 homologs were also present sporadically in genera of Actinobacteria (e.g., Microvirga, Sphingomonas, Citromicrobium, Methylobacterium, Roseomonas, Erythrobacter, Roseivivax, Agrobacterium, Rhizobium, Burkholderia, Psychromonas, Myxococcus, and Geobacter), phylum Chloroflexi (e.g., Oscillochloris, Kedonobacter, and Dehalococcoidia), phylum Bacteroidetes (Polarbacter), phylum Firmicutes (Thermoanaerobacter and Moorella), phylum Planctomycetes (Isophaera), and phylum Synergistetes (Dethiosulfovibrio, Aminobacterium, and Pyrrolobacter). Thus, bacterial p97-like AAA proteins comprise a large clade with a broad phylogenetic distribution.

ACKNOWLEDGMENTS

This research was supported by NIH grant AI64693 and by MSKCC core grant P30-CA008748.

FUNDING INFORMATION

This work, including the efforts of Mihaela-Carmen Uciumeac, Paul Smith, and Stewart Shuman, was funded by HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID) (AI64693).

REFERENCES

2. Erzberger JP, Berger JM. 2006. Evolutionary relationships and structural