Published online 27 May 2012 *Nucleic Acids Research, 2012, Vol. 40, No. 15* ***7465–7475***
doi:10.1093/nar/gks417

Mycobacterium smegmatis SftH exemplifies a distinctive clade of superfamily II DNA-dependent ATPases with 3′ to 5′ translocase and helicase activities

Lyudmila Yakovleva and Stewart Shuman*

Molecular Biology Program, Sloan-Kettering Institute, New York, NY 10065, USA

Received March 26, 2012; Revised April 20, 2012; Accepted April 23, 2012

ABSTRACT

Bacterial DNA helicases are nucleic acid-dependent NTPases that play important roles in DNA replication, recombination and repair. We are interested in the DNA helicases of Mycobacteria, a genus of the phylum *Actinobacteria*, which includes the human pathogen *Mycobacterium tuberculosis* and its avirulent relative *Mycobacterium smegmatis*. Here, we identify and characterize *M. smegmatis* SftH, a superfamily II helicase with a distinctive domain structure, comprising an N-terminal NTPase domain and a C-terminal DUF1998 domain (containing a putative tetracysteine metal-binding motif). We show that SftH is a monomeric DNA-dependent ATPase/dATPase that translocates 3′ to 5′ on single-stranded DNA and has 3′ to 5′ helicase activity. SftH homologs are found in bacteria representing 12 different phyla, being especially prevalent in *Actinobacteria* (including *M. tuberculosis*). SftH homologs are evident in more than 30 genera of *Archaea*. Among eukarya, SftH homologs are present in plants and fungi.

INTRODUCTION

Helicase motor proteins use the chemical energy of NTP hydrolysis to effect mechanical changes in the conformation of nucleic acids and protein–nucleic acid complexes. They are involved in a wide variety of nucleic acid transactions and functionally disabling mutations of helicase motors are associated with many human genetic diseases. Helicases have been classified into superfamilies and families according to their distinctive primary, tertiary and quaternary structures, and their particular biochemical specificities: i.e. their NTP preference, nucleic acid preference and directionality of translocation or unwinding (1). Within the helicase families, there are remarkable structural and functional diversities, via the acquisition of accessory domains (and interactions with collaborating proteins) that direct the helicase motor to a specific biological pathway. Also, even ostensibly homologous helicases may be devoted to distinct biological pathways in different taxa, as the presence of other helicases in the proteome frees them up to take on new or additional functions.

An example of species divergence of helicase function was encountered by Gupta *et al.* (2) in their studies on the recombination/repair function of DNA helicases in Mycobacteria, a genus that includes the slow-growing human pathogen *Mycobacterium tuberculosis* and its faster growing avirulent cousin *Mycobacterium smegmatis*. Mycobacteria have two pathway options for the homology-directed repair of DNA double-strand breaks (DSBs): (i) RecA-dependent homologous recombination (HR) and (ii) RecA-independent homology-directed single-strand annealing (SSA) (2). The HR and SSA mechanisms involve the resection of DSB ends by the coordinated actions of a helicase and a nuclease. Mycobacteria have a distinctive division of helicase labor between the two homology-dependent pathways, whereby the mycobacterial helicase–nuclease machine AdnAB is dedicated to the HR pathway, whereas the RecBCD machine is responsible for SSA (2). Deletion of mycobacterial RecBCD has no impact on the ability of the organism to withstand DNA damage induced by UV or ionizing radiation (2). In contrast, in the model bacterium *Escherichia coli* (that lacks a homolog of AdnAB), RecBCD is the principal helicase–nuclease, which drives HR and RecBCD is also critical for resistance of *E. coli* to a variety of DNA-damaging agents (3).

Such differences in the roster of DNA helicases between bacterial taxa provide useful clues to the evolution and diversification of DNA replication/repair strategy. Moreover, the distinctive helicase rosters of animals versus infectious pathogens (bacteria, viruses, protozoa

*To whom correspondence should be addressed. Tel: +1 212 639 7145; Fax: +1 212 717 3623; Email: shuman@ski.mskcc.org

and fungi) can suggest anti-infective drug targets. With these points in mind, we are focused on the ensemble of DNA helicases in mycobacteria. To date, five mycobacterial DNA helicases have been purified and characterized: AdnAB, UvrD1 and UvrD2 are superfamily I (SF1) helicases, whereas XBP and RqlH are superfamily II (SF2) helicases.

Mycobacterial AdnAB is a heterodimer of two subunits, both consisting of an N-terminal SF1 motor domain and a C-terminal nuclease domain (4). AdnAB hydrolyzes purine NTPs and dNTPs in the presence of single-stranded DNA or linear duplex DNA with a free DSB end. The 3′–5′ translocase/helicase activity of the AdnB subunit drives duplex unwinding from a DSB end, during which the AdnA and AdnB nucleases incise the displaced 5′- and 3′ strands, respectively (4–6). AdnAB-type helicases are present in the proteomes of many bacterial genera within the *Actinomycetales* order of the phylum *Actinobacteria*.

Mycobacterial UvrD1 is homologous to *E. coli* UvrD, a prototypal superfamily I DNA helicase. UvrD1 is a monomeric ssDNA-dependent ATPase/dATPase that unwinds 3′-tailed duplex DNAs by loading onto the 3′ single-strand and translocating in the 3′–5′ direction (7–10). A distinctive property of UvrD1 is that it interacts physically and functionally with Ku, a DNA end-binding protein that orchestrates the mycobacterial non-homologous end-joining pathway of DSB repair (7,11).

Mycobacterial UvrD2 is a monomeric ssDNA-dependent ATPase/dATPase with 3′ to 5′ duplex unwinding activity (7,12,13). UvrD2 has an unusual domain organization, insofar as its N-terminal ATPase domain resembles the SF1 helicase UvrD, yet it has a C-terminal HRDC domain, which is a feature of RecQ-type SF2 DNA helicases. The ATPase and HRDC domains are connected by a CxxC-(14)-CxxC tetracysteine module that defines a new clade of UvrD2-like bacterial helicases found only in *Actinomycetales* (12).

Mycobacterial XBP was identified as a homolog of the eukaryal XBP subunit of transcription factor IIH; it is composed of a C-terminal SF2 motor domain fused to an N-terminal XBP-specific domain module (14). Recombinant *M. tuberculosis* XBP is a monomeric ssDNA-dependent ATPase with 3′ to 5′ helicase activity on 3′-tailed duplex DNAs (15). XBP homologs are widely prevalent in the phylum *Actinobacteria* and are also found sporadically in other bacterial phyla.

Mycobacterium smegmatis RqlH is a 'RecQ-like' helicase (16). RqlH resembles *E. coli* RecQ in possessing DNA-dependent ATPase and ATP-dependent 3′–5′ helicase activities, but is distinguished from 'classic' RecQ proteins by its unique domain composition (16). RqlH consists of a RecQ-like ATPase domain and tetracysteine zinc-binding domain, separated by an RqlH-specific linker. RqlH lacks the C-terminal HRDC domain found in *E. coli* RecQ. Rather, the RqlH C-domain resembles bacterial ComF proteins and includes a phosphoribosyltransferase-like module. RqlH homologs are distributed widely among bacterial taxa. Within the genus *Mycobacterium*, RqlH is present in species *smegmatis, vanabaalenii*, *gilvum*, *KMS, JL* and *JDM601*, none of which has a homolog of classic *E. coli* RecQ. Yet, neither RqlH nor a classical RecQ is found in the proteome of the human pathogen *M. tuberculosis*. However, RqlH is present in the proteomes of more than 50 other species of the phylum *Actinobacteria*, many of which have both RqlH and RecQ.

Here, we identify, purify and biochemically characterize a new mycobacterial superfamily II helicase (hereafter named SftH): the 776-amino acid protein encoded by *M. smegmatis* gene MSMEG_6160. SftH is composed of an N-terminal SF2 motor domain and a signature C-terminal 'DUF1998' module, which includes a putative metal-binding cysteine cluster (Figure 1). We report that SftH is a monomeric enzyme possessing DNA-dependent ATPase, ATP-dependent 3′–5′ translocase and helicase activities. SftH homologs with DUF1998 domains comprise a distinct SF2 helicase subfamily distributed widely among bacteria, archaea, fungi and plants.

MATERIALS AND METHODS

SftH proteins

The open reading frame encoding *M. smegmatis* SftH (MSMEG_6160) was PCR amplified from *M. smegmatis* genomic DNA with primers that introduced an NdeI site at the start codon and a BamHI site downstream of the stop codon. The PCR product was digested using NdeI and BamHI and inserted between NdeI and BamHI sites in pET16b to generate an expression plasmid encoding the SftH polypeptide fused to an N-terminal His_{10} tag. Alanine or serine substitution mutations were introduced into the *sftH* ORF by two-stage overlap extension PCR with mutagenic primers. The plasmid inserts were sequenced to exclude the acquisition of unwanted coding changes during amplification or cloning.

The pET-His_{10}SftH plasmids were transformed into *E. coli* BL21-CodonPlus(DE3)-RIL. Cultures (1-l) derived from single transformants were grown at 37°C in Luria-Bertani medium containing 100 μg/ml ampicillin and 50 μg/ml chloramphenicol until the A_{600} reached 0.6. The cultures were chilled on ice for 1 h, adjusted to 2% (v/v) ethanol and 0.2 mM isopropyl-β-D-thiogalactopyranoside, and then incubated at 17°C for 16 h with constant shaking. The cells were harvested by centrifugation, and the pellets were stored at −80°C. All subsequent procedures were performed at 4°C. Thawed bacteria were resuspended in 50 ml of buffer A (50 mM Tris–HCl, pH 7.5, 250 mM NaCl, 10% sucrose) containing one protease inhibitor cocktail tablet (Roche). Lysozyme and Triton X-100 were added to final concentrations of 0.5 mg/ml and 0.1%, respectively. After incubation for 30 min, the lysate was sonicated to reduce viscosity and insoluble material was removed by centrifugation at 38 000*g* for 1 h. The soluble extracts were mixed for 1 h with 6 ml of Ni-NTA agarose resin (Qiagen) that had been equilibrated with buffer A. The resins were recovered by centrifugation, resuspended in buffer A, poured into columns and then washed with buffer B (50 mM Tris–HCl, pH 7.5, 250 mM NaCl, 10% glycerol) containing 25 mM imidazole. Bound material was eluted stepwise with 4 ml

Figure 1. Domain organization of *M. smegmatis* SftH. The 776-amino acid *M. smegmatis* SftH polypeptide is depicted in the top panel as a linear array (with the N-terminus at left and the C-terminus at right) and the known or imputed domains drawn as cylinders spanning their segments of the primary structure. RqlH is composed of an N-terminal ATPase domain (beige) and a C-terminal DUF1998 domain (green) that includes a potential metal-binding cysteine cluster. The amino acid sequence of the SftH ATPase domain is aligned to that of the homologous portion of *E. coli* RecQ (a prototypal superfamily II DNA helicase) in the middle panel. Positions of side chain identity/similarity are indicated by dot above the sequences. Gaps in the alignments are denoted by dashes. The superfamily II ATPase motifs I, II, III and VI are highlighted in beige. The bottom panel shows an alignment of the amino acid sequence of the *M. smegmatis* (Msm) SftH DUF1998 domain to the homologous segments of putative helicases from the archaea *M. labreanum* (Mla), *M. thermautotrophicus* (Mth), *T. pendens* (Tpe) and *D. kamchatkensis* (Dka), the bacterium *S. thermophilus* (Sth) and the plant *O. sativa* (Osa). The four conserved cysteines are highlighted in green.

aliquots of buffer B containing 100, 200, 300 and 500 mM imidazole. The polypeptide compositions of the eluate fractions were monitored by SDS–PAGE. The recombinant SftH proteins were eluted in the 100, 200 and 300 mM imidazole fractions, which were pooled and then dialyzed against buffer C (50 mM Tris–HCl, pH 7.5, 1 mM dithiothreitol (DTT), 1 mM EDTA, 10% glycerol) containing 100 mM NaCl. The dialysates were applied to 3 ml DEAE–Sephacel columns that had been equilibrated with the same buffer. The columns were washed with 100 mM NaCl in buffer C and then eluted stepwise with 0.25, 0.5 and 1.0 M NaCl in buffer C. SftH was recovered predominantly in 0.25 and 0.5 M NaCl fractions, which were pooled. Protein concentrations were determined by using the BioRad dye reagent with BSA as the standard.

Velocity sedimentation

Aliquots (200 μg in 0.2 ml) of SftH, either alone or mixed with catalase (50 μg), BSA (50 μg) and cytochrome *c* (100 μg), were applied to 4.8-ml 15–30% glycerol gradients containing 50 mM Tris–HCl, pH 7.5, 250 mM NaCl, 1 mM DTT, 1 mM EDTA, 0.05% Triton X-100. The gradients were centrifuged at 50 000 rpm for 18 h at 4°C in

a Beckman SW55Ti rotor. Fractions (0.2 ml) were collected from the bottoms of the tubes.

Nucleoside triphosphatase assay

Reaction mixtures containing (per 10 μl) 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 1 mM [α^{32}P]ATP (Perkin-Elmer Life Sciences) and DNA and SftH as specified were incubated at 37°C. The reactions were quenched by adding 2 μl of 5 M formic acid. An aliquot (2 μl) of the mixture was applied to a polyethyleneimine–cellulose thin layer chromatography (TLC) plate, which was developed with 0.45 M ammonium sulfate. The radiolabeled material was visualized by autoradiography and ^{32}P-ADP formation was quantified by scanning the TLC plate with a Fujix BAS2500 imager.

Helicase assay

The 5′ ^{32}P-labeled strand was prepared by reaction of a synthetic oligodeoxynucleotide with T4 polynucleotide kinase and [γ^{32}P]ATP. The labeled DNA was purified by electrophoresis through a native 17% polyacrylamide gel and then annealed to a 4-fold excess of a complementary DNA strand to form the various helicase substrates shown in the figures. Helicase reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 0.5 pmol (50 nM) radiolabeled DNA and SftH as specified were preincubated for 5 min at room temperature. The reactions were initiated by adding 1 mM ATP and 5 pmol of an unlabeled oligonucleotide identical to the labeled strand of the helicase substrate. Addition of excess unlabeled strand prevents the spontaneous reannealing of the unwound ^{32}P-labeled DNA strand. The reaction mixtures were incubated for 30 min at 37°C and then quenched by adding 2 μl of a solution containing 2% SDS, 200 mM EDTA, 40% glycerol, 0.3% bromophenol blue. A control reaction mixture containing no protein was heated for 5 min at 95°C. The reaction products were analyzed by electrophoresis through a 15-cm 17% polyacrylamide gel in 89 mM Tris–borate, 2.5 mM EDTA. The products were visualized by autoradiography.

Streptavidin displacement assay of SftH translocation on DNA

Synthetic 34-mer oligodeoxynucleotides of otherwise identical nucleobase sequence containing a Biotin–ON internucleotide spacer either at the fourth position from the 5′-terminus or the second position from the 3′-terminus were 5′-end labeled with [γ^{32}P]ATP by using T4 polynucleotide kinase and then purified by electrophoresis through a native 17% polyacrylamide gel. Steptavidin–DNA (SA–DNA) complexes were formed by pre-incubating 50 nM biotinylated ^{32}P-DNA with 2 μM streptavidin (Sigma) in 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$ and 1 mM ATP for 10 min at room temperature. The mixtures were supplemented with 20 μM free biotin (Fisher) and the displacement reactions (10 μl, containing 0.5 pmol biotinylated ^{32}P-DNA) were initiated by adding 3.2 pmol SftH. After incubation for 1 min at 37°C, the reactions were quenched by adding 3 μl of a solution containing 200 mM EDTA, 0.6% SDS, 25% glycerol and 15 μM of a unlabeled single-stranded DNA (the 24-mer oligodeoxynucleotide shown in Figure 6) to mask any binding of SftH to ^{32}P-DNA released from the SA–DNA complex. The reaction products were analyzed by electrophoresis through a 15-cm native 17% polyacrylamide gel containing 89 mM Tris–borate, 2.5 mM EDTA. The free ^{32}P-labeled 34-mer DNA and the SA–DNA complexes were visualized by autoradiography.

RESULTS AND DISCUSSION

Distinctive domain organization of *M. smegmatis* SftH, a putative SF2 helicase

The genomic *M. smegmatis* open reading frame MSMEG_6160 encodes a 776-amino acid polypeptide with an N-terminal domain that resembles the NTPase domains of SF2 helicases, as defined by the following characteristic motifs (highlighted in Figure 1): (i) motif I GKS (the P-loop or Walker A-box), the lysine side chain of which contacts the β and γ phosphates of the NTP substrate and the serine of which coordinates the metal cofactor; (ii) motif II DExH (the Walker B-box), which coordinates the metal cofactor; (iii) motif VI QxxGRxGR, which coordinates the NTP γ phosphate and the water nucleophile for NTP hydrolysis; and (iv) motif III SAT, which makes bridging contacts to motifs II and VI and couples NTP hydrolysis to motor activity (17,18). A pairwise alignment of SftH to *E. coli* RecQ, an exemplary SF2 DNA helicase with an established role in DNA repair and recombination, reveals that their primary structure similarity is limited to the NTPase domain, spanning amino acids 77–424 in SftH and amino acids 42–346 in RecQ, to the extent of 128 positions of side chain identity/similarity (Figure 1). Note that SftH lacks counterparts of the two downstream domains that are the signatures of *E. coli* RecQ and most members of the extended RecQ family, these being: a tetracysteine zinc-binding domain with a distinctive fold (17) and a C-terminal HRDC domain (19). SftH also lacks the C-terminal ComF domain found in the *M. smegmatis* RqlH helicase (16).

Rather, the C-terminal 116 amino acids segment of SftH corresponds to DUF1998, a conserved domain of unknown function found in a large number of putative SF2 helicases distributed widely among bacterial and archaeal taxa. Figure 1 shows an alignment of the C-terminal domain of *M. smegmatis* SftH to the homologous domains of putative helicases from the proteomes of a bacterium from a different phylum—*Sphaerobacter thermophilus* (phylum *Chloroflexi*)—and several archaea (*Methanocorpusculum labreanum*, *Methanothermobacter thermautotrophicus*, *Thermofilum pendens*, *Desulfurococcus kamchatkensis*). A eukaryal SftH homolog from the rice plant *Oryza sativa* is also included in the DUF1998 alignment (Figure 1). The signature feature of the DUF1998 domain is a cluster of four cysteines, suggestive of a metal-binding motif (Figure 1, highlighted in green). Mycobacterial SftH proteins have four other potential metal-binding residues in the DUF1998 domains (two histidines and two cysteines,

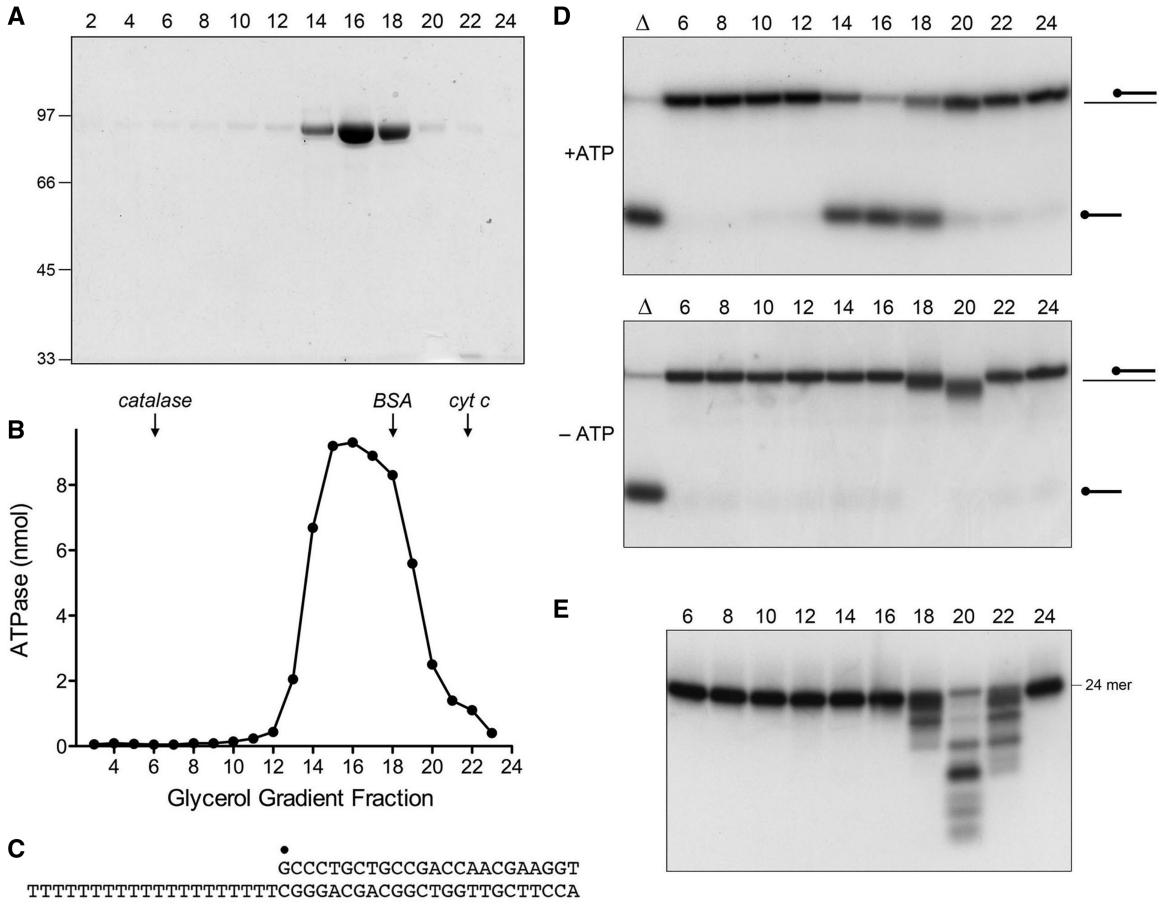

Figure 2. SftH is a monomeric ATPase and helicase. (**A**) SftH was sedimented through a 15–30% glycerol gradient. Aliquots (15 µl) of the even-numbered fractions (fraction 2 being at the bottom of the gradient) were analyzed by SDS–PAGE. The Coomassie blue-stained gel is shown; the positions and sizes (kDa) of marker polypeptides are indicated on the left. (**B**) ATPase reaction mixtures (10 µl) containing 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 1 mM [α^{32}P]ATP, 1 µg salmon sperm DNA and 0.2 µl of the indicated glycerol gradient fractions (i.e. 1 µl of a 5-fold dilution) were incubated at 37°C for 30 min. The activity profile is plotted. The peak positions of the internal standards catalase, BSA and cytochrome *c* analyzed in a parallel glycerol gradient are indicated by vertical arrows. (**C**) The 3′-tailed duplex helicase substrate is shown with the 5′ ^{32}P-label denoted by filled circle. (**D**) Helicase reaction mixtures (10 µl) contained 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 0.5 pmol of 3′-tailed duplex DNA substrate, either 1 mM ATP (top panel) or no ATP (bottom panel) and 0.2 µl of the indicated glycerol gradient fractions. The products were analyzed by native PAGE and visualized by autoradiography. Control reactions lacking enzyme that were heat denatured prior to PAGE are shown in lanes Δ. The positions of the tailed duplex and unwound single strand are indicated at right. (**E**) Reaction mixtures (10 µl) containing 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 0.5 pmol 5′ ^{32}P-labeled 24-mer single-stranded DNA oligonucleotide (the top strand in panel C) and 1 µl of the indicated glycerol gradient fractions were incubated for 15 min at 37°C. The reaction products were analyzed by electrophoresis through a denaturing 17% polyacrylamide gel containing 7 M urea in 45 mM Tris–borate, 1.2 mM EDTA. An autoradiogram of the gel is shown.

highlighted in yellow in Figure 1), but these are not conserved *en bloc* among DUF1998-containing helicase-like proteins.

To our knowledge, there has been no prior characterization of a bacterial enzyme with the distinctive domain composition seen in *M. smegmatis* SftH. Key questions of interest to us were: (i) is SftH a nucleic acid-dependent phosphohydrolase and, if so, what is its substrate and cofactor specificity? (ii) can SftH couple NTP hydrolysis to mechanical work, especially duplex unwinding? and (iii) is the C-terminal domain and its tetracysteine cluster required for such activities?

Recombinant SftH is a DNA-dependent ATPase and helicase

To evaluate the enzymatic and physical properties of SftH, we produced the protein in *E. coli* as a His_{10} fusion and purified it from a soluble extract by nickel–agarose chromatography, DEAE–Sephacel chromatography and zonal velocity sedimentation through a 15–30% glycerol gradient. The SftH preparation was sedimented alone and, in a parallel gradient, as a mixture with marker proteins catalase (248 kDa), BSA (66 kDa) and cytochrome *c* (13 kDa). SDS–PAGE analysis of the even-numbered fractions of the 'SftH-alone' glycerol gradient revealed that the predominant 85 kDa SftH polypeptide sedimented as a single component peaking in fraction 16 (Figure 2A). Aliquots of each gradient fraction were incubated with 1 mM [α^{32}P]ATP in the presence of magnesium and salmon sperm DNA and the hydrolysis of [α^{32}P]ATP to [α^{32}P]ADP was monitored by polyethyleneimine-cellulose TLC. The ATPase activity profile paralleled the abundance of the StfH protein, peaking in fractions 15–17, with nearly complete

conversion of input ATP to ADP (Figure 2B). The sedimentation peaks of the three marker proteins are denoted by arrows on the ATPase activity profile, placing the SftH peak on the 'heavy' side of BSA (Figure 2B). We surmise that SftH is a monomer in solution.

The gradient fractions were also tested for DNA helicase activity with a substrate consisting of a 24-bp duplex with a 20-nt 3′-$(dT)_{20}$ tail to serve as a potential 'loading strand' (Figure 2C). The tailed duplex was formed by annealing a 5′ ^{32}P-labeled 24-mer DNA oligonucleotide to a complementary unlabeled 44-mer DNA. The helicase assay format we used entailed preincubation of the glycerol gradient fractions with labeled DNA, followed by initiation of unwinding by addition of ATP, with simultaneous addition of a 'trap' of excess unlabeled 24-mer displaced strand that: (i) minimizes reannealing of any ^{32}P-labeled 24-mer that was unwound and (ii) competes with the loading strand for binding to any free enzyme that dissociated from the labeled DNA without unwinding it. Consequently, the assay predominantly gauges a single round of strand displacement by enzyme bound to the labeled 3′-tailed duplex prior to the onset of ATP hydrolysis. We observed a single peak of DNA unwinding activity in the SftH-containing fractions 14, 16 and 18 that yielded a radiolabeled free single strand that migrated faster than the input tailed duplex during native PAGE (Figure 2D, top panel); the helicase reaction product comigrated with free 24-mer generated by thermal denaturation of the substrate (lane Δ). No duplex unwinding was observed when the same glycerol gradient fractions were reacted with the helicase substrate in the absence of added ATP (Figure 2D, bottom panel). However, we did note that the incubation with fraction 20 resulted in a slight increase in the electrophoretic mobility of the tailed duplex substrate during native PAGE (Figure 2D, bottom panel), suggesting that the gradient purification might have separated the SftH ATPase/helicase from a contaminating nuclease. This was verified by reacting the glycerol gradient fractions with the ^{32}P-labeled 24-mer strand in the presence of magnesium and analyzing the products by denaturing PAGE, whereby a single component of nuclease activity was detected in fractions 18–22, peaking in fraction 20 (Figure 2E). No nuclease activity was evident in the SftH peak fraction 16 (Figure 2E); thus fraction 16 (or an equivalent fraction from other glycerol gradients) was used for all further biochemical characterizations of StfH.

SftH is a 3′ to 5′ helicase

Whereas SftH readily unwound the 3′-tailed duplex DNA substrate, it failed to unwind a 24-bp blunt duplex DNA substrate (Figure 3), thereby attesting to a requirement for a single-strand tail to serve as a loading strand for the helicase. SftH also failed to unwind a 5′-tailed duplex substrate, consisting of a 24-bp duplex with a 20-nt 5′-$(dT)_{20}$ tail (Figure 3). These results indicate that SftH is a unidirectional 3′–5′ helicase that binds to and tracks along the loading strand and unwinds duplex DNA as it advances.

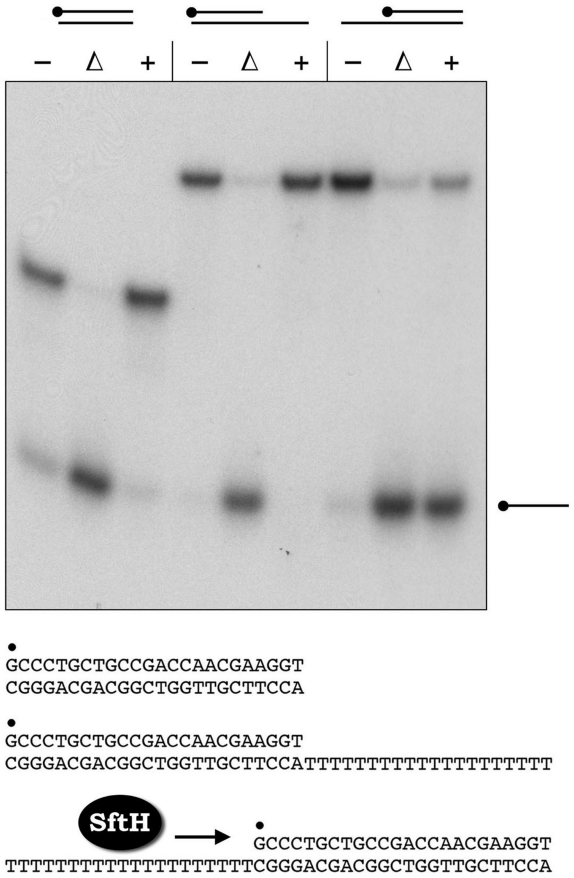

Figure 3. Directionality of duplex unwinding and requirement for a loading strand. Helicase reaction mixtures (10 μl) contained 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 1 pmol of the indicated ^{32}P-labeled DNA substrate, 1 mM ATP and (where indicated by +) 4.5 pmol SftH. Control reactions from which SftH was omitted are shown in lanes −. The reaction products were analyzed by native PAGE and visualized by autoradiography (top panel). Reaction mixtures lacking SftH that were heat denatured prior to PAGE are included in lanes Δ. The 5′ tailed, 3′ tailed, and blunt duplex substrates are shown in the bottom panel, with the 5′ ^{32}P label denoted by filled circle. The results indicate that SftH is a 3′ to 5′ helicase that requires a single-stranded tail to serve as a loading strand.

Effects of motif I and motif II mutations

Mutated versions of the SftH protein were prepared in which Lys90 in motif I or Asp195 and Glu196 in motif II were changed individually to alanine. The three mutants were purified in parallel with wild-type SftH and recovered as single monomeric peaks after glycerol gradient sedimentation. SDS–PAGE showed comparable purity of the wild-type, K90A, D195A and E196A glycerol gradient fractions (Figure 4A). The motifs I and II mutations did away with ATP hydrolysis (Figure 4B) and DNA unwinding (Figure 4C). These results verify that the phosphohydrolase and helicase activities are intrinsic to the recombinant SftH protein.

NTP substrate specificity

NTP substrate specificity was examined by colorimetric assay of the release of P_i from unlabeled ribonucleotides ATP, GTP, CTP or UTP and deoxynucleotides dATP,

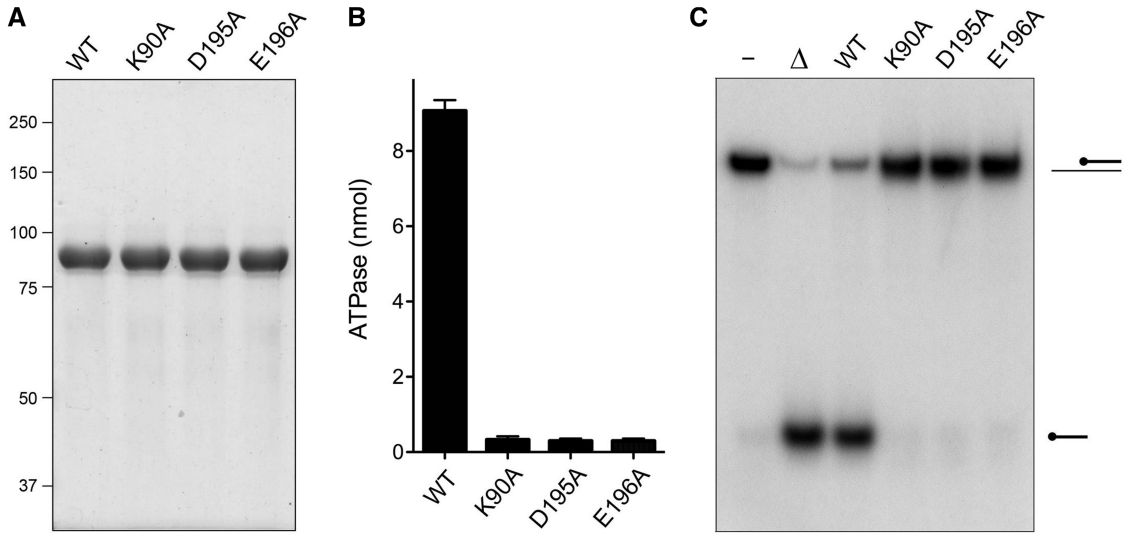

Figure 4. Effects of motifs I and II mutations. (**A**) Aliquots (3 μg) of glycerol gradient-purified wild-type SftH and mutants K90A, D195A and E196A were analyzed by SDS–PAGE. The Coomassie blue-stained gel is shown. The positions and sizes (kDa) of marker polypeptides are indicated on the left. (**B**) ATPase reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 1 mM [α^{32}P]ATP, 1 μg salmon sperm DNA and 4.5 pmol of wild-type or mutant SftH as specified were incubated at 37°C for 30 min. The extent of ATP hydrolysis is plotted. Each datum is the average of three separate experiments ± SEM. (**C**) Helicase reaction mixtures (10 μl) contained 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 0.5 pmol of 3′-tailed duplex substrate (Fig. 2C), 1 mM ATP and 4.5 pmol of wild-type or mutant SftH as specified. SftH was omitted from the control reaction in lane −. The reaction products were analyzed by native PAGE and visualized by autoradiography. A reaction mixture lacking SftH that was heat denatured prior to PAGE is included in lane Δ.

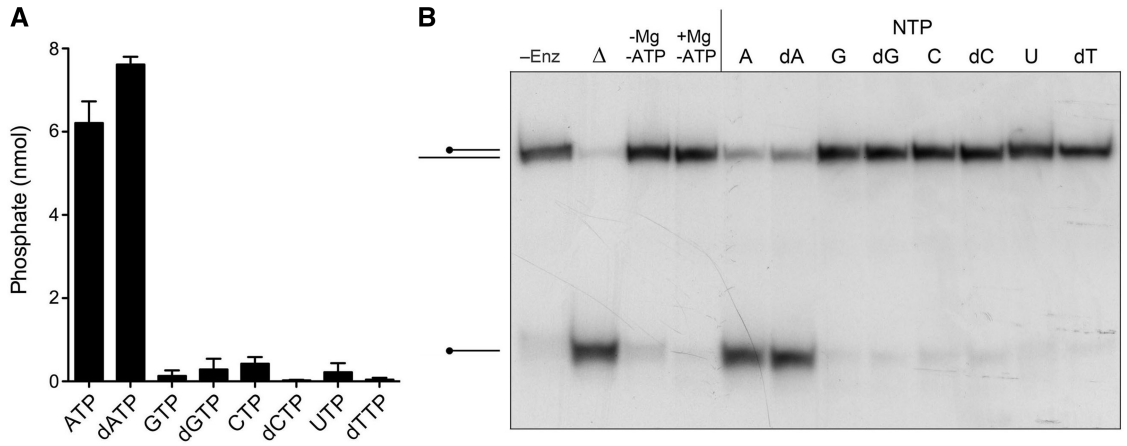

Figure 5. NTP specificity. (**A**) Phosphohydrolase reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 1 μg salmon sperm DNA, 4.5 pmol SftH and 1 mM of the indicated nucleoside triphosphate were incubated for 30 min at 37°C. The reactions were quenched with 1 ml of malachite green reagent (Biomol Research Laboratories). Phosphate release was quantified by measuring A_{620} and interpolating the value to a phosphate standard curve. The values were corrected for the low levels of phosphate measured in control reaction mixtures containing 1 mM of the indicated NTP/dNTP but no added enzyme. Each datum is the average of three separate experiments ± SEM. (**B**) Helicase reaction mixtures (10 μl) contained 0.5 pmol of 3′-tailed duplex substrate, 2.2 pmol SftH and 5 mM $MgCl_2$ plus 1 mM of the indicated NTP. The reaction products were analyzed by native PAGE and visualized by autoradiography. Control reactions lacking SftH (−Enz), or NTP (−ATP), or divalent cation (−$MgCl_2$) or that were heat denatured prior to PAGE (Δ) are included in the lanes at left.

dGTP, dCTP and dTTP, each at 1 mM concentration. SftH specifically hydrolyzed ATP and dATP (Figure 5A). Similarly, the NTP requirement for the SftH helicase reaction was satisfied by ATP and dATP; the other rNTPs and dNTPs were ineffective (Figure 5B).

DNA cofactor requirement and metal cofactor specificity

We tested the ability of single-stranded DNAs of varying length to activate ATP hydrolysis by SftH. Titration of the oligonucleotides revealed a hyperbolic dependence of ATP hydrolysis on the amount of 44-, 36-, 30-, 24- or 12-mer strands (Figure 6). Nonlinear regression fitting of the data to a one-site binding model in Prism yielded apparent K_d values as follows: 0.05 μM 44-mer; 0.11 μM 36-mer; 0.14 μM 24-mer and 0.67 μM 12-mer. A 6-mer oligonucleotide was ineffective at up to 2.5 μM concentration (Figure 6).

The ATPase activity of SftH in the presence of the 44-mer DNA was optimal from pH 5.0 to 7.0 in Tris buffer; activity declined sharply at pH 4.5 and was nil at

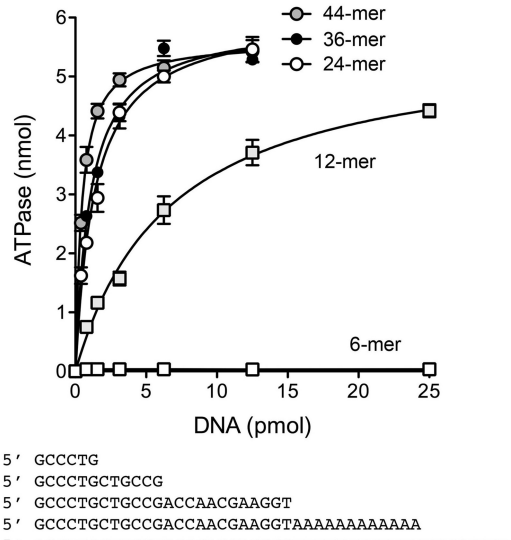

Figure 6. DNA dependence of ATP hydrolysis. Reaction mixtures (10 µl) containing 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 1 mM (10 nmol) [α^{32}P]ATP, 3.2 pmol SftH and increasing amounts of 44-, 36-, 24-, 12- or 6-mer DNA oligonucleotides as specified were incubated at 37°C for 15 min. The nucleobase sequences of the oligonucleotides are shown at bottom. The extent of ATP hydrolysis is plotted as a function of added DNA (pmol). Each datum is the average of three separate DNA titration experiments ± SEM.

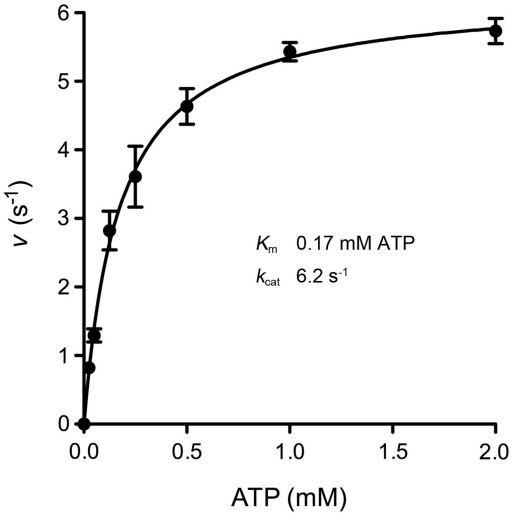

Figure 7. Steady-state kinetics. Reaction mixtures (30 µl) containing 20 mM Tris–HCl, pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 1.25 µM 44-mer oligonucleotide (shown in Fig. 6), 21.6 pmol SftH and either 0.025, 0.05, 0.125, 0.25, 0.5, 1.0 or 2.0 mM [α^{32}P]ATP were incubated at 37°C. Aliquots (5 µl) were withdrawn at either 10, 20, 30 and 60 s (0.025 and 0.05 mM ATP reactions), 15, 30, 60 and 120 s (0.125, 0.25 and 0.5 mM ATP reactions), or 30, 60, 120 and 300 s (1 and 2 mM ATP reactions) and quenched immediately with formic acid. The extent of ATP hydrolysis was plotted as a function of time for each ATP concentration and the initial rates were derived by linear regression analysis in Prism. The initial rates (pmol•s^{-1}) were divided by the molar amount of input enzyme to obtain a turnover number v (s^{-1}), which is plotted in the figure as a function of ATP concentration. Each datum is the average of three separate time-course experiments ± SEM. A nonlinear regression curve fit of the data to the Michaelis–Menten equation (in Prism) is shown. The K_m and k_{cat} values are indicated.

pH 4.0 (Supplementary Figure S1A). No ATP hydrolysis was detected when magnesium was omitted. The divalent cation requirement for ATP hydrolysis by SftH was satisfied by 1 mM magnesium, manganese or calcium, whereas cobalt was half as active (Supplementary Figure S1B). In contrast, 1 mM cadmium, copper, nickel or zinc was ineffective (Supplementary Figure S1B). Activity was broadly optimal between 0.5 and 5 mM magnesium, manganese or calcium (data not shown).

Steady-state kinetics of ATP hydrolysis

We determined steady-state kinetic parameters by measuring the velocity of ATP hydrolysis as a function of ATP concentration in the presence of 1.25 µM 44-mer single-stranded DNA cofactor (Figure 7). From a nonlinear regression curve fit of the data to the Michaelis–Menten equation, we calculated that SftH has a K_m of 0.17 mM ATP and a k_{cat} of 6.2 s^{-1}.

SftH translocates unidirectionally on single-stranded DNA

NTP hydrolysis by nucleic acid-dependent phosphohydrolases is often coupled to mechanical work—either duplex unwinding or displacement of protein–nucleic acid complexes—as a consequence of translocation of the phosphohydrolase enzyme along the nucleic acid. To address whether SftH has translocase activity, we employed a streptavidin (SA) displacement assay (5,16,20–22), as follows. ^{32}P-labeled 34-mer DNA oligonucleotides containing a single biotin moiety at the fourth internucleotide from the 5′-end or the second internucleotide from the 3′-end were preincubated with excess SA to form a stable SA–DNA complex that was easily resolved from free biotinylated 34-mer DNA during native PAGE (Figure 8). The translocation assay scores the motor-dependent displacement of SA from the DNA in the presence of ATP and excess free biotin, which instantly binds to free SA and precludes SA rebinding to the labeled DNA. The rationale of the assay is that directional tracking of the motor along the DNA single strand will displace SA from one DNA end, but not the other. When moving 3′ to 5′, it can displace SA as it collides with the 5′-biotin–SA. In contrast, a 3′-biotin–SA is not expected to be displaced by a 3′–5′ translocase, because the motor moves away from the SA and simply falls off the free 5′-end. The converse outcomes apply to a 5′–3′ translocase; it displaces a 3′ SA, but not a 5′ SA. The instructive finding was that SftH displaced SA from a 5′ biotin–SA complex on the 34-mer single-stranded DNA to yield the free ^{32}P-labeled 34-mer strand, but was unable to displace SA from a 3′ biotin–SA complex tested in parallel (Figure 8A). Stripping of the 5′ biotin–SA complex by SftH to liberate free DNA depended on ATP and magnesium (Figure 8B). The ATPase-defective SftH mutants K90A and D195A displayed no detectable 3′–5′ translocase activity (Figure 8B).

Deletion and mutation of the DUF1998 domain

To address the role of the DUF1998 domain, if any, and the possible sufficiency of the ATPase domain, we

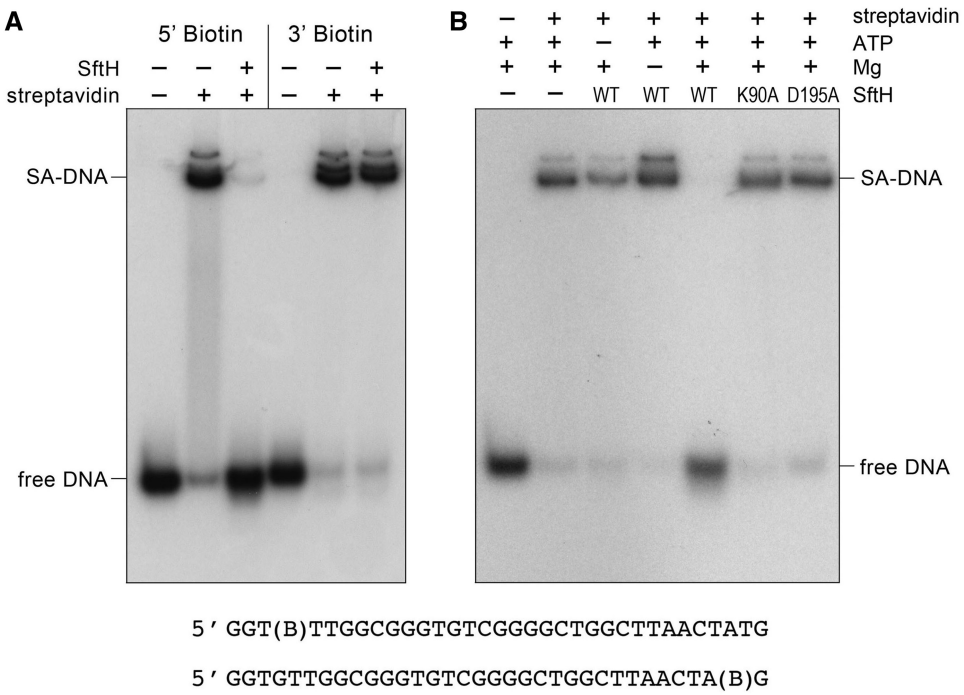

Figure 8. Directionality of SftH translocation on ssDNA. Translocase assays were performed as described under Methods. Native PAGE analysis of the reaction products are shown. The species corresponding to SA–DNA complex and free DNA are indicated. The nucleobase sequences of the 5′ or 3′ biotinylated 34-mer single-stranded DNAs are shown at bottom with (**B**) signifying the position of the biotin spacer. The complete translocase reaction mixtures contained 1 mM ATP (+), 5 mM $MgCl_2$ (+), 0.5 pmol ^{32}P-labeled 3′ or 5′ biotinylated DNA attached to streptavidin (+), and 1 pmol of either wild-type SftH (+ or WT) or mutants K190A or D195A as specified. Streptavidin, ATP, magnesium and/or SftH were omitted (−) from control reactions as indicated above the lanes. The reactions in panel A were performed with the 5′ biotinylated or 3' biotinylated DNAs as indicated. All of the reactions in panel B were performed with the 5′ biotinylated DNA.

attempted to produce a series of C-terminal truncation mutants: SftH-(1-701), SftH-(1-660), SftH-(1-600) and SftH-(1-500). Each was intractably insoluble when expressed in *E. coli* and therefore not amenable to purification and characterization. As an alternative, we sought to evaluate the contributions of the putative metal-binding motif of the DUF1998 domain, by introducing double alanine substitutions C736A–C738A and C742A–C745A in the context of the full-length SftH polypeptide, but the C736A–C738A and C742A–C745A proteins were again insoluble when produced in *E. coli*. Then, we replaced individual cysteines with serine, but once again the C736S, C738S, C742S and C745S proteins were insoluble. In contrast, alanine mutations at two other cysteines in the DUF1998 domain (Cys681 and Cys751; highlighted in yellow in Figure 1) did not adversely affect solubility and the isolated SftH C681A and C751A proteins retained DNA-dependent ATPase activity (Supplementary Figure S2). We surmise that the C736–C738–C742–C745 tetracysteine motif of the DUF1998 domain is needed for proper folding of the SftH protein, possibly via metal coordination.

Phylogenetic distribution of SftH helicases

The biochemical properties of *M. smegmatis* SftH elucidated above are consonant with those of the mycobacterial helicases UvrD1, UvrD2 and RqlH, insofar as these enzymes are monomeric single-strand DNA-dependent ATPase/dATPase motors with 3′ to 5′ helicase activity on tailed duplex DNA substrates (28–31). Proper folding and function of SftH apparently requires the C-terminal DUF1998 domain, and the putative metal-binding tetracysteine cluster composed of C736, C738, C742 and C745. The mycobacterial RqlH enzyme has a structurally distinct RecQ-type tetracysteine zinc-binding domain that is also essential for its proper folding (16). In contrast, the signature tetracysteine domain of UvrD2 is not required for its DNA-dependent ATPase activity, but is critical for helicase activity (12).

In our view, the domain organization of mycobacterial SftH is sufficiently diverged from other *bona fide* helicases to warrant the designation of SftH as the exemplar of a new SF2 clade. A search of the NCBI database with *M. smegmatis* SftH recovered scores of SftH-like proteins, composed of the same SF2 motor and DUF1998 domains, arrayed in the same order, and with similar inter-domain spacing. SftH is present in the available mycobacterial proteomes, with the exception of *M. leprae* (Supplementary Table S1). The *M. tuberculosis* SftH homolog is a 771-amino acid polypeptide encoded by the Rv3649 gene (Supplementary Figure S3). SftH is apparently nonessential for mycobacterial growth, insofar as Sassetti *et al.* (23) have identified two viable mutant strains of *M. tuberculosis* that have transposons interrupting the 5′ half of the Rv3649 ORF; the most proximal transposon insert is located at nucleotide 322 and should ablate SftH enzymatic activity.

SftH is present in the proteomes of 49 other genera of the phylum *Actinobacteria* (Supplementary Table S1). Indeed, SftH homologs are distributed widely in the bacterial domain of life, being present in species belonging to the phyla *Proteobacteria*. *Firmicutes*, *Cyanobacteria*, *Chloroflexi*, *Deinococcus-Thermus*, *Deferribacteres*, *Acidobacteria*, *Synergistetes*, *Aquificae*, *Planctomycetes* and *Spirochaetales* (Supplementary Table S2). Among the *Proteobacteria*, SftH was prevalent in species of the delta-subdivision, but was lacking in taxa of the *Enterobacteriales* order of the gamma-subdivision (which includes *E. coli*). SftH homologs were not retrieved from the phyla *Bacterioidetes*, *Chlamydia* or *Mollicutes*.

SftH homologs were found in 35 different genera of *Archaea* (Supplementary Table S3). In eukarya, SftH homologs were seemingly limited to fungi and plants (Supplementary Table S3). An alignment of amino acid sequence of *M. smegmatis* SftH to its homolog from the rice plant *O. sativa* is shown in Supplementary Figure S4 and highlights conservation of the motor domain, the DUF1998 domain (including the putative metal-binding cysteines) and the interdomain linker. The SftH homolog of the budding yeast *S. cerevisiae* is named Hrq1 and is inessential for yeast vegetative growth (24). The Hrq1 homolog of the fission yeast *S. pombe* was recently reported to have 3′ to 5′ helicase activity and to play a role in resistance of fission yeast to killing by mitomycin C and cisplatin (25).

In conclusion, we unveil SftH as a member of a novel clade of SF2 helicases found in diverse bacterial, archaeal and eukaryal taxa. Our initial biochemical analysis and phylogenetic analyses provide impetus for future studies, including: (i) characterization of the *in vivo* effects of ablating *sftH* in *M. smegmatis*, singly and in combination with other potentially redundant inessential *M. smegmatis* helicase genes (*adnAB*, *uvrD1*, *rqlH*, *recBCD*) and (ii) structural analyses, especially of the distinctive C-terminal DUF1998 domain.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–4 and Supplementary Figures 1–4.

FUNDING

US National Institutes of Health (NIH) [AI64693]; American Cancer Society Research Professorship (to S.S.). Funding for open access charge: NIH [AI64693].

Conflict of interest statement. None declared.

REFERENCES

1. Singleton,M.R., Dillingham,M.S. and Wigley,D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, **76**, 23–50.
2. Gupta,R., Barkan,D., Redelman-Sidi,G., Shuman,S. and Glickman,M.S. (2011) Mycobacteria exploit three genetically distinct DNA double-strand break repair pathways. *Mol. Microbiol.*, **79**, 316–330.
3. Dillingham,M.S. and Kowalczykowski,S.C. (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.*, **72**, 642–671.
4. Sinha,K.M., Unciuleac,M.C., Glickman,M.S. and Shuman,S. (2009) AdnAB: a new DSB-resecting motor-nuclease from Mycobacteria. *Genes Dev.*, **23**, 1423–1437.
5. Unciuleac,M.C. and Shuman,S. (2010) Characterization of the mycobacterial AdnAB DNA motor provides insights to the evolution of bacterial motor-nuclease machines. *J. Biol. Chem.*, **285**, 2632–2641.
6. Unciuleac,M.C. and Shuman,S. (2010) Double-strand break unwinding and resection by the mycobacterial helicase-nuclease AdnAB in the presence of mycobacterial SSB. *J. Biol. Chem.*, **285**, 34319–34329.
7. Sinha,K.M., Stephanou,N.C., Gao,F., Glickman,M.S. and Shuman,S. (2007) Mycobacterial UvrD1 is a Ku-dependent DNA helicase that plays a role in multiple DNA repair events, including double-strand break repair. *J. Biol. Chem.*, **282**, 15114–15125.
8. Sinha,K.M., Glickman,M.S. and Shuman,S. (2009) Mutational analysis of Mycobacterium UvrD1 identifies functional groups required for ATP hydrolysis, DNA unwinding, and chemomechanical coupling. *Biochemistry*, **48**, 4019–4030.
9. Curti,E., Smerdon,S.J. and Davis,E.O. (2007) Characterization of the helicase activity and substrate specificity of *Mycobacterium tuberculosis* UvrD. *J. Bacteriol.*, **189**, 1542–1555.
10. Güthlein,C., Wanner,R.M., Sander,P., Davis,E.O., Bosshard,M., Jiricny,J., Böttger,E.C. and Springer,B. (2009) Characterization of the mycobacterial NER system reveals novel functions of the *uvrD1* helicase. *J. Bacteriol.*, **191**, 555–562.
11. Aniukwu,J., Glickman,M.S. and Shuman,S. (2008) The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev.*, **22**, 512–527.
12. Sinha,K.M., Stephanou,N.C., Unciuleac,M.C., Glickman,M.S. and Shuman,S. (2008) Domain requirements for DNA unwinding by mycobacterial UvrD2, an essential DNA helicase. *Biochemistry*, **47**, 9355–9364.
13. Williams,A., Güthlein,C., Beresford,N., Böttger,E.C., Springer,B. and Davis,E.O. (2011) UvrD2 is essential in Mycobacterium tuberculosis, but its helicase activity is not required. *J. Bacteriol.*, **193**, 4487–4494.
14. Poterszman,A., Lamour,V., Egly,J.M., Moras,D., Thierry,J.C. and Poch,O. (1997) A eukaryotic XBP/ERCC3-like helicase in Mycobacterium leprae? *TIBS*, **22**, 418–419.
15. Biswas,T., Pero,J.M., Joseph,C.G. and Tsodikov,O.V. (2009) DNA-dependent ATPase activity of bacterial XBP helicases. *Biochemistry*, **48**, 2839–2848.
16. Ordonez,H., Unciuleac,M. and Shuman,S. (2012) *Mycobacterium smegmatis* RqlH defines a novel clade of bacterial RecQ-like DNA helicases with ATP-dependent 3'-5' translocase and duplex unwinding activities. *Nucleic Acids Res.*, **40**, 4604–4614.
17. Bernstein,D.A., Zittel,M.C. and Keck,J.L. (2003) High-resolution structure of the *E. coli* RecQ helicase core. *EMBO J.*, **22**, 4910–4921.
18. Gu,M. and Rice,C.M. (2010) Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc. Natl Acad. Sci. USA*, **107**, 521–528.
19. Bernstein,D.A. and Keck,J.L. (2005) Conferring substrate specificity to DNA helicases: role of the RecQ HRDC domain. *Structure*, **13**, 1173–1182.
20. Morris,P.D. and Raney,K.D. (1999) DNA helicases displace streptavidin from biotin-labeled oligonucleotides. *Biochemistry*, **38**, 5164–5171.
21. Morris,P.D., Byrd,A.K., Tackett,A.J., Cameron,C.E., Tanega,P., Ott,R., Fanning,E. and Raney,K.D. (2002) Hepatitis C virus NS3 and simian virus 40 T antigen helicases displace streptavidin from 5'-biotinylated oligonucleotides but not from 3'-biotinylated oligonucleotides: evidence for directional bias in translocation on single-stranded DNA. *Biochemistry*, **41**, 2372–2378.
22. Byrd,A.K. and Raney,K.D. (2004) Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA. *Nat. Struct. Mol. Biol.*, **11**, 531–538.

23. Sassetti,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.
24. Shiratori,A., Shibata,T., Arisawa,M., Hanaoka,F., Murakami,Y. and Eki,T. (1999) Systematic identification, classification, and characterization of the open reading frames which encode novel helicase-related protein in *Saccharomyces cerevisiae* by gene disruption and Northern analysis. *Yeast*, **15**, 219–253.
25. Groocock,L.M., Prudden,J., Perry,J.P. and Boddy,M.N. (2012) The RecQ4 orthologue Hrq1 is critical for DNA interstrand cross-link repair and genome stability in fission yeast. *Mol. Cell. Biol.*, **32**, 276–287.