CHARACTERIZATION OF MYCOBACTERIAL FLAP ENDONUCLEASE FenA AND RNA HELICASE HelY

A Dissertation

Presented to the Faculty of the Weill Cornell Graduate School

of Medical Sciences

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

By

Maria Loressa Lagula Uson

August 2018

CHARACTERIZATION OF MYCOBACTERIAL FLAP ENDONUCLEASE FenA AND RNA HELICASE HelY

Maria Loressa Lagula Uson, Ph.D.

Cornell University 2018

The genus *Mycobacterium* encompasses over 170 species, including the disease-causing *Mycobacterium tuberculosis* and its nonpathogenic relative *Mycobacterium smegmatis*. My thesis research focuses on mycobacterial proteins from two classes of enzymes critical for nucleic acid processing: helicases and nucleases. I purified and characterized FenA as a flap endonuclease and 5' exonuclease and HelY as an RNA helicase.

Structure-specific DNA nucleases are found in all domains of life and play important roles in genome replication and maintenance, yet little is known about such nucleases in mycobacteria. I characterized *M. smegmatis* FenA as a manganese-dependent nucleic acid phosphodiesterase with flap endonuclease and 5’ exonuclease activities. FenA efficiently removes the 5’ App(dN) terminus of an aborted nick ligation reaction intermediate, which is formed under a variety of circumstances by bacterial ligases and especially by mycobacterial ligases D and C.

We reported the 1.8 Å crystal structure of FenA, which assimilates three manganese ions in its active site. The three manganese ions are coordinated, directly and via waters, to a constellation of eight carboxylate side chains. I mutated each of the metal-coordinating residues individually to alanine and

tested the mutants for flap endonuclease and 5’ exonuclease activities. I found that the carboxylate contacts to all three manganese ions are essential for FenA’s activities. A fourth manganese ion is engaged by a structure unique to FenA that comprises a 3_{10} helix and a surface β-loop.

Mycobacteria possess a large ensemble of DNA helicases that function in DNA replication, repair and recombination. In comparison, little attention has been paid to the roster of RNA helicases in mycobacteria or their roles in RNA metabolism. We identified *M. smegmatis* HelY as a bacterial homolog of the eukaryotic Ski2 and Mtr4 helicases that regulate RNA 3' processing and turnover by the exosome. I showed that HelY is an RNA-stimulated ATPase/dATPase and an ATP/dATP-dependent 3'-to-5' helicase. HelY requires a 3' single-stranded RNA tail to displace the complementary strand of a RNA:DNA hybrid or RNA duplex. The biochemical properties of HelY are consistent with a role in mycobacterial RNA transactions, and we speculate that HelY functions in RNA catabolism.

BIOGRAPHICAL SKETCH

Maria Loressa Uson graduated from the University of California, Irvine in 2010 with a Bachelor of Science in Biological Sciences. From 2007 to 2010 she conducted undergraduate research in the laboratory of Dr. Lisa Flanagan, investigating the proliferation and differentiation of neural stem cells in three-dimensional matrices. From 2010 to 2012 Loressa worked in the lab of Dr. Peter Kuhn at The Scripps Research Institute. There, she focused on the detection of circulating tumor cells and structural characterization of EphA3, a tyrosine kinase receptor implicated in cancer. In 2012, after working briefly at the Genomics Institute for the Novartis Research Foundation, Loressa began her Ph.D. studies at the Weill Cornell Graduate School of Medical Sciences. She conducted her thesis research at Memorial Sloan Kettering in the laboratory of Dr. Stewart Shuman, which she joined in 2014.

For my mom

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my mentor, Dr. Stewart Shuman. I joined his laboratory at the end of my second year, and he has been supportive ever since. Stewart has a passion for science and teaching that is always evident in his mentoring, and he has been a great advisor throughout my time in his lab. I would also like to thank Dr. Michael Glickman and Dr. Christopher Lima for being on my thesis committee and providing valuable insight and support. I thank Dr. Dirk Remus for acting as my thesis examining committee chairperson.

I would like to thank the former and current members of the Shuman laboratory that I have had the privilege of working alongside, particularly Dr. Heather Ordonez for her guidance. I would also like to acknowledge my rotation student, Ayala Carl, for her hard work, spirit, and determination. I thank her and Yehuda Goldgur for their contributions to my thesis research.

I am forever grateful for my family and friends. They have voiced and demonstrated their support in numerous ways over the years, and I would not be where I am today without them. I am especially thankful for the friends and family members that brought home a little closer by visiting me in New York City or by traveling with me abroad on our crazy adventures.

Lastly, I would like to thank Mark Rauch, my high school A.P. Biology teacher. His enthusiasm for the subject piqued my interest in biology and his encouragement led me to my first experience working in a research laboratory.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

TB – tuberculosis
ROS – reactive oxygen species
NO – nitric oxide
RNI – reactive nitrogen intermediates
BER – base excision repair
AP – apurinic/apyrimidinic
8-oxoG – 8-oxo-7,8-dihydroguanine
NER – nucleotide excision repair
DSB – double strand break
HR – homologous recombination
NHEJ – non-homologous end joining
SSA – single-strand annealing
ssDNA – single-stranded DNA
dsDNA – double-stranded DNA
HJ – Holliday junction
LigD – Ligase D
LigD-POL – Ligase D polymerase domain
LigD-PE – Ligase D phosphoesterase domain
LigD-LIG – Ligase D ligase domain
AppDNA – adenylylated DNA
LigC – Ligase C
Pol1 – polymerase I
FEN – flap endonuclease
SF - superfamily
MMC – mitomycin C
PNPase – polynucleotide phosphorylase

CHAPTER ONE

INTRODUCTION

Mycobacteria

The genus *Mycobacterium* is composed of over 170 species, some of which cause disease in animals and humans. Mycobacteria are rod-shaped and distinguished by their elaborate cell walls, which consist of mycolic acids, peptidoglycan, and arabinogalactan (1). Most mycobacterial species thrive in water and soil, and several species, such as *Mycobacterium avium* and *Mycobacterium abscessus*, can cause opportunistic infections in humans. In contrast, the etiological agent of leprosy, *Mycobacterium leprae*, is an obligate intracellular pathogen.

DNA damage repair in mycobacteria

Organisms are exposed to exogenous and endogenous factors that can damage DNA, and repair of such damage to maintain genomic integrity is essential for survival. Mycobacteria have evolved multiple DNA damage repair pathways, and their utilization is dependent on the type of DNA damage. Modifications of DNA bases are predominantly repaired by base excision repair or nucleotide excision repair. Mycobacteria are not thought to employ mismatch repair, as they lack the canonical components, though a recent study of an archaeal-type endonuclease encoded by mycobacteria (2) is challenging this long-held view. Repair of double-strand breaks in mycobacteria is achieved through homologous recombination, non-

homologous end joining or single-strand annealing. The latter two pathways are not found in *Escherichia coli* and were previously not thought to occur in bacteria. In addition, studies of all the aforementioned pathways have uncovered differences in DNA damage repair between mycobacteria and *E. coli*. Thus, mycobacteria can serve as another bacterial model system for investigating the mechanisms of DNA damage repair.

Mycobacterium tuberculosis

DNA repair is also critical for bacterial pathogenesis, as infectious bacteria must evade the DNA-damaging attacks of the host immune system. Mechanisms of DNA damage repair have been studied extensively in *Mycobacterium tuberculosis*, the primary causative agent of TB, which remains one of the leading causes of death worldwide. Humans are the only known natural host of *M. tuberculosis*, which spreads when an infected individual coughs, becoming airborne in droplets that are then inhaled by the next host. The bacteria travel to the lungs, the most common site of infection, where they are engulfed by macrophages. Within the macrophage, *M. tuberculosis* is exposed to a variety of DNA-damaging assaults, including reactive oxygen species (ROS), acidification, nitric oxide (NO) and NO-derived reactive nitrogen intermediates (RNI), and nutrient starvation (3). To withstand this hostile landscape *M. tuberculosis* possesses strong DNA damage repair systems, and gene expression studies of *M. tuberculosis* in macrophages (4) or similarly antagonistic conditions *in vitro* (5,6) reveal up-regulation of the expression of genes involved in DNA replication, recombination and repair.

Critical to its pathogenicity, DNA repair in *M. tuberculosis* not only maintains genome integrity, it also mediates antibiotic resistance.

Base Excision Repair

Damage of DNA bases due to oxidation, deamination or alkylation is primarily repaired by the base excision repair (BER) pathway. Briefly, DNA glycosylases recognize and excise damaged bases, resulting in an abasic (apurinic/apyrimidinic) site. DNA glycosylases that possess apurinic/apyrimidinic (AP) lyase activity can further modify the abasic site by cleaving the sugar phosphate backbone. In the absence of AP lyase activity, an AP endonuclease generates the backbone nick. Following end processing by phosphonucleotide kinase or deoxyribose-phosphodiesterase, the site is then filled in by DNA polymerase and sealed by DNA ligase (7).

The high GC content of the mycobacterial genome (~65%) renders it vulnerable to the oxidation of guanine as well as the deamination of cytosine to uracil. *M. tuberculosis* has evolved a remarkable redundancy of BER that addresses these susceptibilities, particularly against the oxidized guanine product 8-oxo-7,8-dihydroguanine (8-oxoG) (8). 8-oxoG is one of the most stable and frequent DNA base legions, and adenine is frequently incorporated across an 8-oxoG, leading to G→T/C→A transversions. Mycobacteria encode for two DNA glycosylases: MutM, which excises 8-oxoG; and MutY, which excises A paired against 8-oxoG (9,10). Once in the nucleotide pool, MutT hydrolyzes 8-oxoG. Four homologs of MutT have been found in the mycobacterial genome, though only two have been characterized thus far as

hydrolyzing 8-oxoG (11). Additionally, the two AP endonucleases involved in mycobacterial BER, End and XthA, preferentially select for sites against C, emphasizing the bias for 8-oxoG removal in the BER pathway (12).

Cytosine can spontaneously deaminate, resulting in the formation of uracil. RNI may also promote the deamination of cytosine. *M. tuberculosis* possesses two DNA glycosylases, Ung and UdgB, for the excision of uracil from DNA (13,14), and two dUTPases, a canonical and an archaeal-type, for the hydrolysis of intracellular dUTP (15,16). In addition to cytosine deamination, RNI can form potent DNA-alkylating agents. Mycobacteria encode for multiple DNA glycosylases that excise alkylated DNA bases.

Nucleotide Excision Repair

Nucleotide excision repair (NER) excises a stretch of nucleotides flanking, and including, damaged DNA. In comparison to BER, NER is a lower specificity repair pathway that recognizes a more extensive range of substrates. A critical component of NER is the UvrABC exinuclease: a complex of UvrA and UvrB recognizes damaged DNA and recruits the endonuclease UvrC (17-19). If DNA damage stalls RNA polymerase during transcription, transcription-repair coupling factor, Mfd, displaces the transcription machinery and recruits UvrABC exinuclease (20). After UvrC excises 12 to 13 nucleotides around the DNA damage, UvrD, a helicase, disassembles the cleaved DNA. Mycobacteria encode for two homologs of UvrD, UvrD1 and UvrD2. Deletion of UvrD1 sensitizes the bacteria to DNA damaging agents, whereas UvrD2 is essential in *M. tuberculosis* (21). It is

suggested that the functions of UvrD1 and UvrD2 overlap, with UvrD2 distinctly capable of performing all essential activities. Following the action of UvrD, DNA polymerase fills the gap and ends are sealed by DNA ligase. Mutation of the NER pathway in mycobacteria resulted in a higher sensitivity to DNA damaging agents than a strain deficient for BER (22), highlighting its importance in DNA damage recognition and repair.

Double-strand break repair

Double-strand breaks (DSBs) are the most lethal form of DNA damage, and their repair is crucial for bacterial survival. Mycobacteria employ three distinct pathways to repair DSBs: (i) RecA-dependent homologous recombination (HR), which faithfully restores the chromosome; (ii) Ku/LigD-dependent non-homologous end-joining (NHEJ) and (iii) RecA-independent single-strand annealing (SSA), both of which are mutagenic (Figure 1.1).

Homologous Recombination

In HR, an unbroken homologous DNA strand serves as a template for DNA synthesis across the DSB. The process consists of three stages: (i) resection of DSB ends to create a 3' single-stranded DNA (ssDNA) overhang, (ii) Rec-A mediated strand invasion of the homologous strand followed by DNA synthesis and (iii) resolution of the Holliday junction, resulting in faithful repair of the DSB.

HR is the only DSB repair pathway utilized by *E. coli*. In *E. coli*, the helicase-nuclease RecBCD initiates recombination. The heterotrimer is comprised of one subunit each of RecB, RecC, and RecD. Helicase domains

are present in RecD and RecB, which also contains a C-terminal nuclease domain (23). RecD translocates along the 5' strand while RecB translocates along the 3' strand, and both activities require ATP hydrolysis (24,25). RecB and RecD do not contact each other directly, but each subunit makes protein-

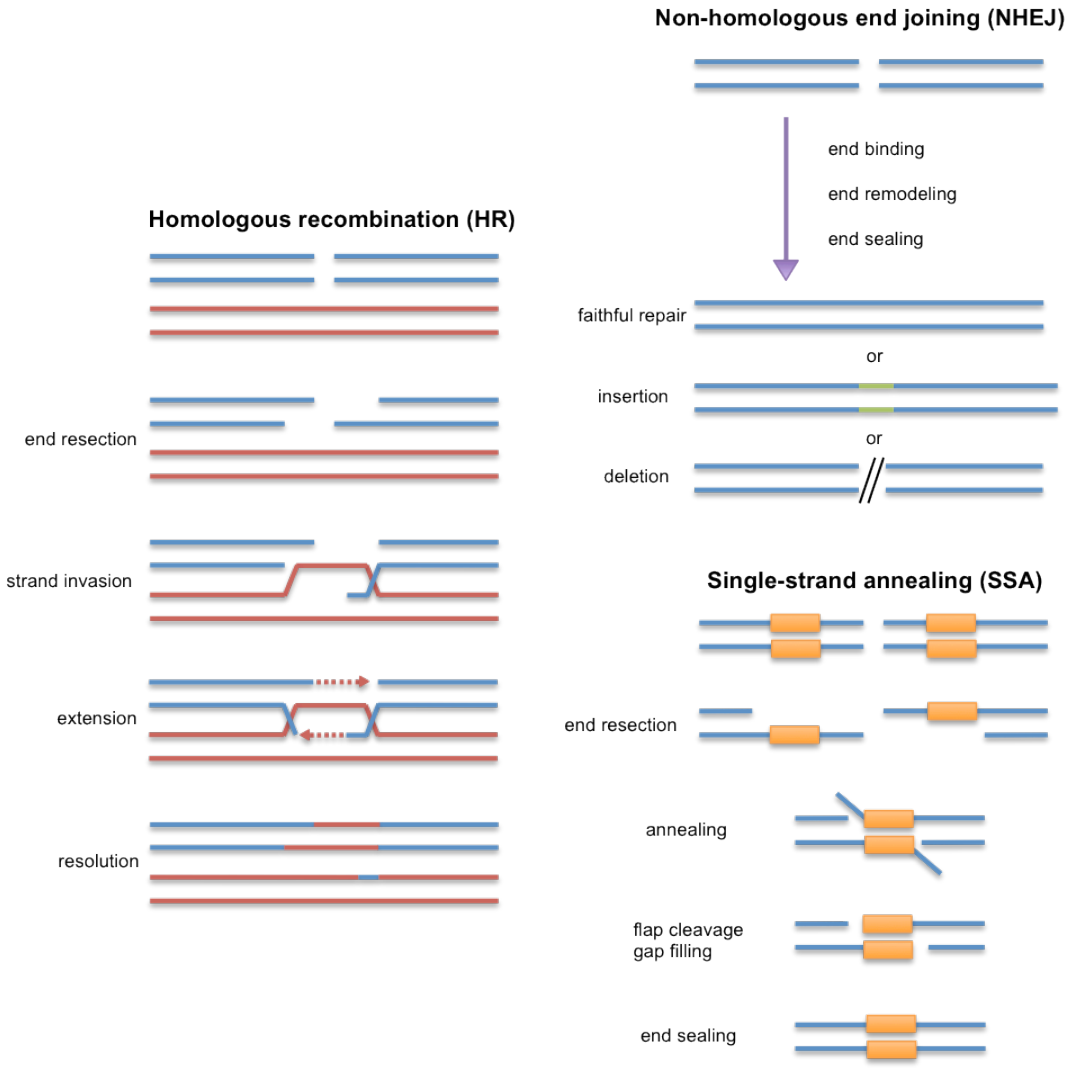


Figure 1.1 Pathways of double-strand break repair in mycobacteria. Mycobacteria rely on three distinct pathways for double-strand break (DSB) repair: homologous recombination (HR), non-homologous end joining (NHEJ) and single strand annealing (SSA). The major DNA processing events for each pathway are depicted. For SSA, the orange squares indicate two repeats flanking a DSB. Figure adapted from (26).

protein contacts with RecC. These interactions thread nascent ssDNA through RecC, which also recognizes the Chi site, an 8-bp stretch of DNA that activates the 5'-to-3' exonuclease activity of RecB (27). This activity generates a 3' overhang that becomes coated in single-stranded DNA-binding protein (SSB). Prior to strand invasion of homologous DNA, a second DNA binding protein, the ATPase RecA, must be exchanged for SSB. Following end resection RecBCD coordinates the loading of RecA onto ssDNA (28).

RecBCD is the predominant resection machinery in *E. coli* HR. However, in cells lacking RecBCD and the nuclease SbcBCD, RecF-mediated resection can compensate. This pathway includes the proteins RecQ (helicase), RecJ (5' exonuclease) and RecFOR, which exchanges SSB for RecA on ssDNA (29).

Mycobacteria encode two helicase-nucleases: RecBCD and AdnAB. Unlike its role in *E. coli*, mycobacterial RecBCD is part of the SSA repair pathway. AdnAB has been implicated in mycobacterial HR and cells lacking AdnAB are sensitive to a variety of clastogens, particularly ionizing radiation and hydrogen peroxide, which induce DSBs as well as other types of DNA damage (30). AdnAB is a heterodimer of AdnA and AdnB subunits, both of which contain a N-terminal helicase/motor domain and RecB-like C-terminal nuclease domain. This composition imparts a distinction between AdnAB and RecBCD (one nuclease, two helicase domains).

Biochemical studies have elucidated the helicase-nuclease mechanism of mycobacterial AdnAB. Binding of AdnAB to a DSB initiates translocation of

the motor domains in tandem along the same DNA strand in the 3'-to-5' direction (31,32). This translocation requires ATP and is dependent on the ATPase activity of the leading AdnB motor domain. As AdnAB moves along the DNA, the 3' ssDNA that is produced is threaded through the lagging AdnA motor domain and into the AdnB nuclease domain, while the 5' ssDNA is threaded through the AdnA nuclease domain.

Recent *in vivo* studies have revealed which functions of either AdnA or AdnB are critical for DNA repair. *Mycobacterium smegmatis*, a faster growing and non-pathogenic relative of *M. tuberculosis*, is commonly manipulated for genetic studies to elucidate the physiological functions of mycobacterial proteins. *M. smegmatis* strains with separate helicase- or nuclease-dead mutations in AdnA or AdnB were generated and assayed for HR activity using an inducible chromosomal DSB reporter system (33). In this system the endonuclease I-sceI induces a single chromosomal DSB, and repair by each DSB repair pathway (HR, NHEJ or SSA) results in a specific phenotype (of bacterial colonies on solid medium) and can thus be scored (30). Cells lacking AdnA or AdnB nuclease activity were capable of repairing I-sceI-induced DSBs by HR. Surprisingly, a double nuclease-dead AdnAB mutant also sustained HR *in vivo*. This study unveiled that the critical function of AdnAB during HR is its helicase motor activity, specifically of the AdnB subunit. This begs the question of whether 5' end resection and generation of a 3' ssDNA overhang is required for RecA-mediated strand invasion, though the possibility remains that mycobacteria encode a functionally redundant 5' exonuclease.

As in *E. coli*, RecFOR in mycobacteria has been implicated in an alternative HR pathway. As assayed by the chromosomal reporter system, deletion of AdnAB decreases but does not abrogate DSB repair by HR (30). However, loss of either RecO or RecF in combination with AdnAB abolishes HR events (34). In contrast, RecR is required for both AdnAB- and RecFO-mediated HR. As discussed earlier, *E. coli* RecBCD and RecFOR are capable of loading RecA onto ssDNA in preparation for strand invasion. The study of mycobacterial RecFOR suggests that AdnAB does not have inherent RecA loading activity, hence the requirement for RecR in HR.

In comparison with the compensatory RecFOR HR pathway in *E. coli*, neither a RecQ helicase nor a RecJ nuclease homolog is present in mycobacteria. Analysis of the *M. tuberculosis* proteome revealed Rv2090, a putative 5' exonuclease. I purified the *M. smegmatis* homolog of Rv2090 as a recombinant protein from *E. coli*. Biochemical characterization of the protein uncovered flap endonuclease and 5' exonuclease activities. Consequently, we named the protein Flap endonuclease A, or FenA. FenA does not cleave ssDNA, but rather, it utilizes its 5' exonuclease activity on the 5' end of a nick in duplex DNA. If the nicked DNA duplex contains a 5' flap, it endonucleolytically cleaves the flap. These activities rule out a role of FenA in HR. Instead, the functions of FenA point to possible roles in NHEJ, SSA, and lagging strand synthesis, as discussed below.

After DSB end resection, RecA mediates strand invasion of intact DNA and searches for homology within the undamaged DNA duplex. RecA

protomers bind ssDNA in an ATP-dependent manner, forming a nucleoprotein filament. Structural studies of RecA in complex with DNA and an ATP analogue show that ssDNA and ATP bind cooperatively to RecA-RecA interfaces (35). Subsequently, the RecA-ssDNA nucleoprotein filament binds double-stranded DNA (dsDNA) at a secondary DNA binding site. When the RecA-ssDNA filament encounters homology, strand exchange results in the formation of a heteroduplex in which the ssDNA is paired with the complementary strand of the double-stranded DNA (dsDNA). This displaces one strand of the donor duplex, which stays bound at the secondary DNA binding site (36). RecA makes few contacts to the complementary strand, and formation of the heteroduplex relies on correct base pairing with the original ssDNA (35). Additionally, RecA imposes a B-DNA-like conformation of the heteroduplex, excluding non-Watson-Crick hydrogen bonds and ensuring fidelity. Hydrolysis of RecA-bound ATP results in the loss of RecA DNA binding.

RecA-mediated strand invasion leads to the generation of a four-stranded DNA structure called a Holliday junction (HJ). As DNA is synthesized opposite the undamaged homologous DNA, the HJ migrates. In *E. coli* this migration is mediated by RecG or RuvAB. Biochemical characterization of RecG and RuvAB from *M. tuberculosis* revealed similar enzymatic functions. Mycobacterial RecG is a DNA-dependent ATPase and helicase that binds HJs, induces structural distortions, and catalyzes efficient branch migration in the presence of ATP (37,38). Similarly, mycobacterial RuvA binds and causes

structural distortions of HJs (39), and together with RuvB it promotes branch migration of HJs (40). After DNA synthesis, the endonuclease/resolvase RuvC catalyzes HJ resolution. Mycobacteria encode for RuvC and a second endonuclease/resolvase, RuvX, both of which catalyze HJ resolution *in vitro* (41).

Non-homologous end-joining

One of the most studied prokaryotic NHEJ systems is that of mycobacteria. Ku and Ligase D (LigD) are the key players of mycobacterial NHEJ, and deletion of either in *M. smegmatis* results in the loss of NHEJ in the chromosomal reporter assay (30). At a DSB Ku binds and bridges the broken ends and recruits LigD for end sealing. LigD is composed of three autonomous domains: polymerase (POL), phosphoesterase (PE) and an ATP-dependent ligase (LIG).

LigD can seal a DSB faithfully but often modifies the broken ends before end sealing. Mycobacterial LigD-POL is capable of fill-in synthesis across a 5' single-stranded overhang or the addition of single nucleotides to blunt-ended DSBs (42). Non-templated synthesis by LigD-POL has been demonstrated, as has synthesis opposite an abasic site, highlighting the mutagenic nature of NHEJ. In addition, LigD-POL prefers to incorporate ribonucleotides at DSBs (42). This may be a reflection of the non-replicating cellular environment where the dNTP pool might be limited. Inactivation of LigD-POL increased NHEJ fidelity in *M. smegmatis,* exposing LigD as a driver of mutagenesis (43).

LigD-PE of *Pseudomonas aeruginosa* has been characterized as a 3' ribonuclease and 3' phosphatase. On a primer-template substrate LigD-PE resects a stretch of 3' ribonucleotides until a single 3' ribonucleotide remains. The remaining ribonucleotide has a 3' PO_4 end that LigD-PE converts to a 3' OH per its phosphatase activity (44). Although these functions have yet to be demonstrated by mycobacterial LigD-PE, the amino acid residues critical for these activities are present in the mycobacterial homolog.

LigD-LIG demonstrates weak end-sealing activity *in vitro* of a nicked DNA duplex, though it is noted that this substrate differs from a DSB. Bacterial LigD preferentially ligates a nick with a 3' ribonucleotide (45). This activity coincides with the tendencies of LigD-POL to incorporate ribonucleotides and LigD-PE to reduce a ribonucleotide tract to a single 3' ribonucleotide.

LigD-LIG seals the 3' OH and 5' PO_4 ends of a nick via a three-step mechanism employed by all classic polynucleotide ligases. First, ligase reacts with ATP (or NAD^+) to form a ligase-AMP intermediate and release pyrophosphate (or nicotinamide mononucleotide). Next, AMP is transferred from the ligase-AMP intermediate to the 5' PO_4 terminus of the DNA nick, resulting in a DNA-adenylate (AppDNA). In the final step, ligase catalyzes the formation of a phosphodiester bond to seal the nick via attack of the 3' OH terminus on the AppDNA, releasing AMP in the process.

As discussed earlier, deleting Ku or LigD in *M. smegmatis* results in the loss of NHEJ repair of a chromosomal DSB, which is instead repaired by HR or SSA. However, a plasmid-based assay, in which a DSB (that cannot be

repaired using HR or SSA) is induced on plasmid DNA, revealed backup components of the NHEJ machinery: Ligase C (LigC) as well as multiple polymerases, including PolD1 and PolD2 (42,46). In contrast to LigD, LigC only contains an ATP-dependent ligase domain. Biochemical characterization of LigC and LigD showed that in the presence of ATP, both ligases have a tendency to dissociate from AppDNA prior to the final end-sealing step (47). Following dissociation from AppDNA, it is thought that the ligase immediately forms another ligase-AMP intermediate that cannot interact with AppDNA, thereby preventing ligation of the nick.

AppDNA itself is a form of DNA damage and requires repair. Eukarya possess a repair enzyme, aprataxin, to remove the AMP moiety from AppDNA (48). This DNA repair activity is crucial, and mutation of aprataxin in humans causes a neurological disorder called ataxia-ocular apraxia-1 (49). To our knowledge, no bacterial enzyme had yet been associated with a DNA-adenylate repair activity. In light of its 5' exonuclease activity at a DNA nick, it was hypothesized that FenA might be capable of removing the blocked 5' end of an AppDNA ligation intermediate. To investigate this possibility, I adenylylated a 5' ^{32}P-labeled 18-mer DNA strand. This strand and an unlabeled 18-mer DNA strand were annealed to a 36-mer DNA strand to form a singly nicked DNA-adenylate. Reaction of FenA with the DNA-adenylate resulted in the removal of the 5' App(dN) terminus, highlighting a possible role of FenA in NHEJ.

Single-strand annealing

A third DSB repair pathway, SSA, is utilized when a DSB is flanked by repeats. In SSA, bidirectional resection at the DSB exposes complementary repeats that anneal, forming flaps in the DNA that are then cleaved (Figure 1.1). Flap cleavage is followed by DNA ligation.

In contrast to its role in HR in *E. coli*, the helicase-nuclease RecBCD is essential for mycobacterial SSA, as assayed by the chromosomal reporter system in *M. smegmatis*. The precise function of RecBCD has yet to be demonstrated, but *M. tuberculosis* RecD has been characterized as a DNA-dependent ATPase and DNA helicase (50). It possesses robust 5'-to-3' and weak 3'-to-5' unwinding activities. Presumably, mycobacterial RecBCD performs the initial resection of the DSB (similar to its *E. coli* homolog during HR) and/or flap cleavage once the complementary repeats anneal. However, given its flap endonuclease activity, FenA is also a prime candidate for removal of the resulting flap DNA.

Interestingly, RecFOR has also been implicated in SSA. Deletion of RecF abolishes DSB repair by SSA in the chromosomal reporter assay, whereas deletion of RecO and RecR results in a 10-fold and 50-fold decrease in SSA frequency, respectively (34). It is postulated that bacterial RecF recognizes a ssDNA-dsDNA junction and RecR is a protein scaffold that serves to stabilize the RecFOR complex. Mycobacterial RecO binds ssDNA and promotes annealing of SSB-coated ssDNA (51). One can speculate that after RecBCD resection of the DSB, RecF recognizes a ssDNA intermediate that can be bound by RecO.

DNA polymerase I and 5' nucleases

Discovered by Arthur Kornberg in 1956, *E. coli* DNA polymerase I (Pol1) was the first enzyme demonstrated to catalyze template-directed nucleotide polymerization (52). But it soon became apparent that DNA Pol1 was not the primary replicative polymerase, nor was it essential for *E. coli* viability. However, *E. coli* mutants of DNA Pol1 were defective in DNA damage repair following UV radiation or exposure to alkylating agents (53). Additionally, the joining of Okazaki fragments during lagging strand synthesis was retarded in DNA Pol1 mutants (54). The critical function of DNA Pol1 is now considered to be its gap filling activity: (i) of gaps during DNA damage repair; and (ii) between Okazaki fragments during DNA replication.

DNA Pol1 has three distinct enzymatic functions: DNA polymerase activity, 3' exonuclease activity and 5' nuclease activity. Limited proteolysis experiments revealed the polymerase and 3' exonuclease activities reside on a large C-terminal fragment, whereas the 5' nuclease is contained in a small N-terminal fragment (55). This 5' nuclease possesses 5' exonuclease and flap endonuclease activities (56-58). Both functions are consistent with the role of DNA Pol1 in DNA repair and replication: the enzyme can degrade the 5' terminus of a nick or Okazaki fragment while the 3' end is extended by its polymerase activity. But, if strand displacement synthesis unanneals the 5' end of the nick or preceding Okazaki fragment, the enzyme can also cleave the 5' flap that is generated. Accordingly, these nuclease activities are structure specific, rather than based on DNA sequence. Some bacteria,

including *E. coli*, encode for a stand-alone homolog of the N-terminal 5' nuclease domain of DNA Pol1. Deletion of Pol1 is tolerated in bacteria that possess a separate 5' nuclease, but not in bacteria that have no such stand-alone homolog (59). Thus, bacteria require 5' exonuclease/flap endonuclease activity for viability.

FenA is homologous to the N-terminal 5' nuclease domain of DNA Pol1 in mycobacteria, which also demonstrates flap endonuclease and 5' exonuclease activities. Unlike its *E. coli* homolog, *M. tuberculosis* DNA Pol1 does not possess 3' exonuclease activity. Disruption of the DNA Pol1 C-terminal polymerase domain resulted in a viable *M. smegmatis* strain, but the essentiality of the 5' nuclease activity has not yet been established (60). Our laboratory is currently investigating the requirement of DNA Pol1 5' nuclease and FenA in mycobacteria.

5' nucleases are found in all domains of life and are critical to the maintenance of genome integrity. In contrast to bacterial DNA Pol1, no known eukaryotic polymerase possesses 5' nuclease activity. Instead, 5' nucleases, named as flap endonucleases (FENs) in eukarya, are encoded as separate proteins. Structural studies of FENs, from bacteriophage T5 to human, as well as other structure-specific nucleases, have provided insight on mechanisms for substrate recognition and cleavage (61-65). The nuclease core consists of a β-sheet, usually composed of seven strands, flanked on both sides by α-helical regions. Binding sites for dsDNA are located on either side of the β-sheet, including a helix-turn-helix motif that is characteristic of FENs. A helical

“wedge” enforces a 90-100° bend of the “template” DNA strand (the strand that is not nicked and is analogous to the DNA template strand during lagging strand synthesis) (Figure 1.2). This allows for easy identification of nicked duplex DNA, which contains a junction that allows for such flexibility around a phosphodiester bond. A single active site, which is highly conserved among FENs, is used for both exo- and endonucleolytic cleavages. The active site contains seven or eight acidic amino acids that coordinate the divalent metal

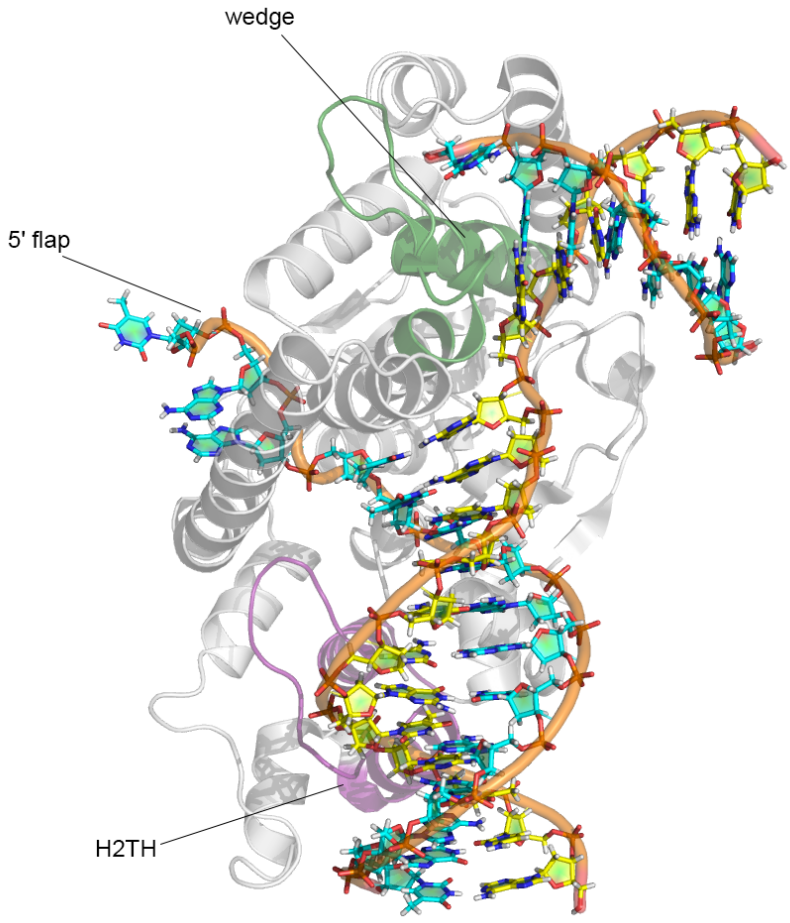


Figure 1.2 Overview of the structure of FEN in complex with substrate DNA. Top view of human Flap Endonuclease 1 structure bound to substrate DNA (PDB 5KSE). The substrate consists of three strands: the template strand (yellow), the 5' flap, and the 3' strand (both in cyan). The wedge motif is shaded dark green and the helix-2turn-helix (H2TH) is shaded in purple.

ions (magnesium or manganese) required for catalysis. A crystal structure of FenA in complex with manganese was solved, revealing at its active site eight acidic resides coordinating three manganese ions. Mutational analysis unveiled the essentiality of each metal ion for FenA nuclease activity.

Mycobacterial helicases

Helicases are ubiquitous enzymes that couple ATP hydrolysis to translocation along DNA or RNA, resulting in unwinding of nucleic acid duplexes or protein displacement. They are generally categorized into six superfamilies (SF) based on shared amino acid sequence motifs, tertiary and quaternary protein structures, and biochemical properties (66). Helicases in SF3 through SF6 form oligomeric, toroidal structures whereas those in SF1 and SF2 do not. SF1 and SF2 helicases are predominantly monomeric, with a core structure composed of two RecA-like domains. SF2 is the largest superfamily of helicases, and many SF1 and SF2 proteins are well characterized.

As described above, many mycobacterial helicases function in DNA damage repair. SF1 helicases include UvrD1, UvrD2, AdnAB, and RecD. RecG belongs to SF2 and RuvB is a SF6 helicase. Although the expression of these helicases (excepting UvrD1 and RecD) is up-regulated after DNA damage in *M. tuberculosis* (4,5), only UvrD2 is essential. Mycobacteria encode for several other helicases that are differentially expressed in *M. tuberculosis* under genotoxic stress (Table 1.1). Lhr is a SF2 helicase that translocates 3'-to-5' along DNA, unwinding both DNA:RNA and DNA:DNA

duplexes (67). It is proposed to be the founding member of a novel clade of SF2 helicases due to its unique biochemistry and domain organization: its ATPase domain is contained within the N-terminal third of the protein, but the C-terminal two thirds of Lhr has no motifs or similarities to proteins other than Lhr homologs. Although its physiological role has yet to be determined, Lhr is upregulated in *M. tuberculosis* exposed to UV irradiation or mitomycin C (MMC) (5,68).

Table 1.1 DNA helicases in mycobacteria

Helicase	Superfamily (SF)	Directionality	Function	Up-regulated in *M. tuberculosis* under stress? [1]	Essentiality in *M. tuberculosis*[3]
AdnAB	SF1	3'-to-5'	HR, end resection	Yes	Not essential
DnaB	SF4	5'-to-3'	DNA replication	Yes	Essential
DinG	SF2	5'-to-3'	?	No	Not essential
HelZ	SF2	?	?	Yes	Not essential
Lhr	SF2	3'-to-5'	?	Yes	Not essential
RecD	SF1	5'-to-3' 3'-to-5'	SSA	No	Not essential
RecG	SF2	5'-to-3' 3'-to-5'	HR, branch migration	Yes[2]	Not essential
RqlH	SF2	3'-to-5'	?	No	Not essential
RuvAB	SF6	5'-to-3'?	HR, branch migration	Yes	Not essential
SftH	SF2	3'-to-5'	?	No	Not essential
UvrD1	SF1	3'-to-5'	NER	No	Not essential
UvrD2	SF1	3'-to-5'	NER	Yes	Essential
XPB	SF2	3'-to-5'	?	No	Not essential

[1] Namouchi, A. *et al.* (6)
[2] Schnappinger, D. *et al.* (5)
[3] As tested by transposon mutagenesis (71)

Two other helicases up-regulated following MMC exposure are DnaB and HelZ. DnaB belongs to SF4, forming a hexameric ring that tracks 5'-to-3' along DNA (69,70). Bacterial DnaB is the replicative helicase and the same function is presumed of the *M. tuberculosis* homolog, as it is essential for the organism's viability (71). Mycobacterial HelZ has yet to be characterized and its role is unknown.

A number of other mycobacterial DNA helicases have been characterized, though their functions *in vivo* remain to be interrogated (Table 1.1). RqlH, SftH and XPB are DNA-dependent ATPases and ATP-dependent 3'-to-5' DNA helicases (72-74). They all belong to SF2. Mycobacterial RqlH is a RecQ-like helicase whose enzymatic activities resemble those of *E. coli* RecQ, though RqlH possesses a distinct protein domain composition that differentiates it from other RecQ helicases. RqlH contains a RqlH-specific linker region and a C-terminal domain composed of a phosphoribosyltransferase-like module (73). Similarly, mycobacterial SftH exemplifies a novel clade of SF2 helicases that contain a signature C-terminal domain of unknown function, which includes a putative metal binding site (74). *M. tuberculosis* XPB is a homolog of human XBP, which is a critical component of the eukaryotic transcription factor IIH (75). Eukaryotic XBP functions in transcription as well as NER, and mutation of XBP leads to a number of genetic disorders in humans, including xeroderma pigmentosum type B.

Another SF2 helicase from *M. tuberculosis* that has been characterized is DinG, a 5'-to-3' helicase. It unwinds an assortment of branched DNA substrates, including those that mimic replication forks and HJs (76). This unwinding requires ATPase activity and points toward roles of DinG in DNA replication and recombination.

In contrast, little is known about mycobacterial RNA helicases. RNA helicases participate in RNA degradation, maturation and quality control (77). To our knowledge the only mycobacterial RNA translocase/helicase that had been characterized was the transcription termination factor Rho, which belongs to SF5. Bacterial Rho binds the nascent transcript, translocates along the RNA and enforces dissociation of the transcription machinery. Rho from *M. tuberculosis* is an RNA-binding NTPase that efficiently terminates transcription *in vitro* (78,79), and is essential for bacterial viability (71).

I purified and characterized a putative mycobacterial RNA helicase that had been annotated in *M. tuberculosis* as HelY. *M. smegmatis* HelY hydrolyzes ATP or dATP to unwind RNA:DNA or RNA:RNA duplexes. The enzyme tracks 3'-to-5' on RNA, and duplex unwinding requires a 3' tail. The ATPase/dATPase activity of HelY is stimulated by RNA but not DNA. Accordingly, HelY is unable to translocate on DNA.

Ski2 and Mtr4

HelY was chosen due to its homology to two eukaryotic RNA helicases, Ski2 and Mtr4. Both are SF2 helicases, and Ski2 is the founding member of the Ski2-like family, which includes Mtr4. Ski2 and Mtr4 are RNA-dependent

ATPases that unwind 3'-tailed RNA duplexes in the 3'-to-5' direction (80,81). The structures of *Saccharomyces cerevisiae* Ski2 and Mtr4 have been solved, and they share a similar overall protein structure. Figure 1.3 depicts the crystal structure of Ski2. Forming the core of Ski2 and Mtr4 are two RecA-like domains, a winged-helix domain and a C-terminal ratchet domain. An insertion domain splits the winged-helix domain and protrudes from the protein core. The amino acid sequences of the insertion domains are poorly conserved between Ski2 and Mtr4, but they share secondary structure elements. The

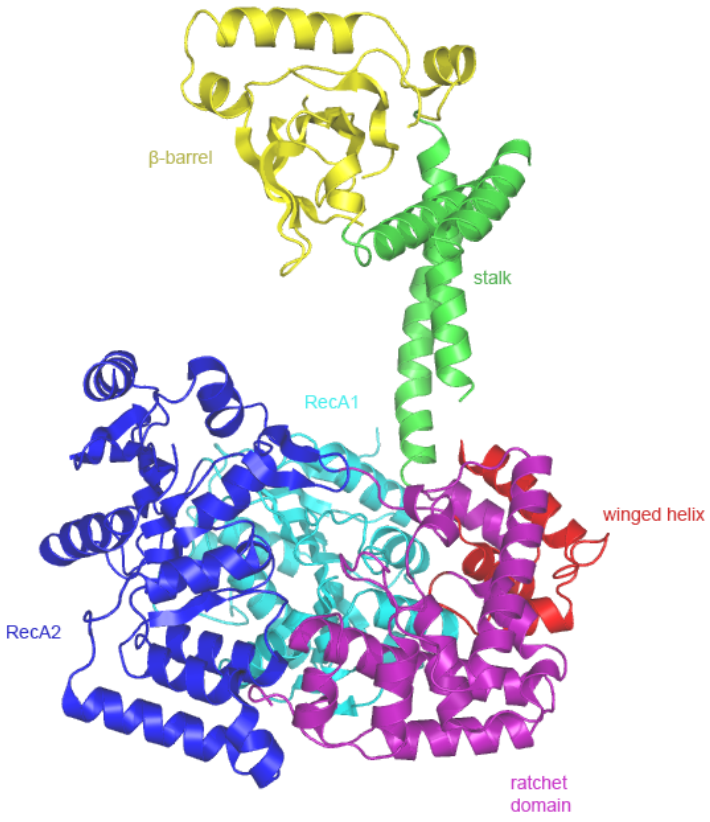


Figure 1.3 Overview of the Ski2 crystal structure. The structure of *S. cerevisiae* Ski2-ΔN, lacking the N-terminal 295 residues (PDB 4A4Z), shown from a side view. Domain names and structural features are colored as indicated.

insertion forms a flexible arch composed of an elongated helical stalk and β-barrel. The domain contributes to RNA binding (82,83), but it is not required for helicase activity in either Ski2 or Mtr4 (81,84).

Ski2 and Mtr4 function in the context of multi-subunit complexes (Ski2 in the Ski complex and Mtr4 in the TRAMP complex) to modulate the activities of the eukaryotic RNA exosome in the cytoplasm and nuclease, respectively. The exosome is a multi-protein complex that catalyzes degradation of 3' RNA ends (85,86). In the nucleus, the exosome functions in maturation of rRNA, snRNA and snoRNA as well as overall RNA quality control. The predominant roles of the cytoplasmic exosome are in mRNA surveillance and general mRNA turnover. The exosome core forms a doughnut-shaped structure with a central channel through which only single-stranded RNA can thread (87,88). Ski2 and Mtr4 presumably unwind RNA secondary structures or displace RNA-bound proteins prior to RNA entry into the exosome.

The Ski complex is a heterotetramer of Ski2, Ski3 and two copies of Ski8 (89), and all three proteins are required for 3'-to-5' RNA decay in *S. cerevisiae* (90). Ski3 and Ski8 are RNA binding proteins, with Ski3 serving as the protein scaffold for the complex. Another protein, Ski7, bridges the Ski complex and RNA exosome (91).

The TRAMP complex includes Mtr4, a noncanonical poly(A) polymerase (Trf4 or Trf5) and a zinc knuckle protein (Air1 or Air2). Mtr4 preferentially binds substrates with short poly(A) 3' tails (80), which are added to RNA by Trf4/5. Trf4/5 and Air1/2 bind Mtr4 as a heterodimer (92), and Mtr4

physically interacts with components of the exosome (93). Mtr4 may also act in isolation outside of the TRAMP complex to unwind substrates for the exosome, and the helicase is essential for *S. cerevisiae* viability (94).

The bacterial degradosome

The biochemical properties of HelY resemble those of Ski2 and Mtr4 and suggest a role for HelY in mycobacterial RNA catabolism. Bacteria possess a structural homolog of the eukaryotic exosome, the 3' exoribonuclease polynucleotide phosphorylase (PNPase). *E. coli* PNPase is a critical component of the RNA degradosome, a multi-protein complex that functions in 3' RNA end processing and degradation in many bacterial species. The enzyme organization of the degradosome differs between evolutionary divergent bacteria, but the most common composition includes an endoribonuclease, an exoribonuclease, a metabolic enzyme and a SF2 helicase, specifically a DEAD-box helicase (77). DEAD-box helicases contain several characteristic motifs, including an aspartate-glutamate-alanine-aspartate (D-E-A-D) motif involved in ATP hydrolysis (95). The Ski2-like family (including Ski2, Mtr4 and HelY) is closely related and shares most of the characteristic DEAD-box helicase family motifs. However in place of the DEAD sequence, Ski2-like helicases contain a DExH motif. In HelY, Ski2, and Mtr4, this sequence is DEVH.

The most extensively studied degradosome is that of *E. coli*. The proteins of the *E. coli* degradosome assemble around the endoribonuclease, RNase E (Figure 1.4). The catalytic domain of RNase E is located at its N-

terminus, while the largely unstructured C-terminus binds RNA and the other major components of the degradosome: PNPase, the DEAD-box helicase RhlB and the glycolytic enzyme enolase (96,97). Although the N-terminal domain of RNase E is required for *E. coli* viability, its enzymatic activity is not required for *in vitro* reconstitution of the degradosome. In fact, a mutant of RNase E truncated for most of its N-terminal domain was adequate for

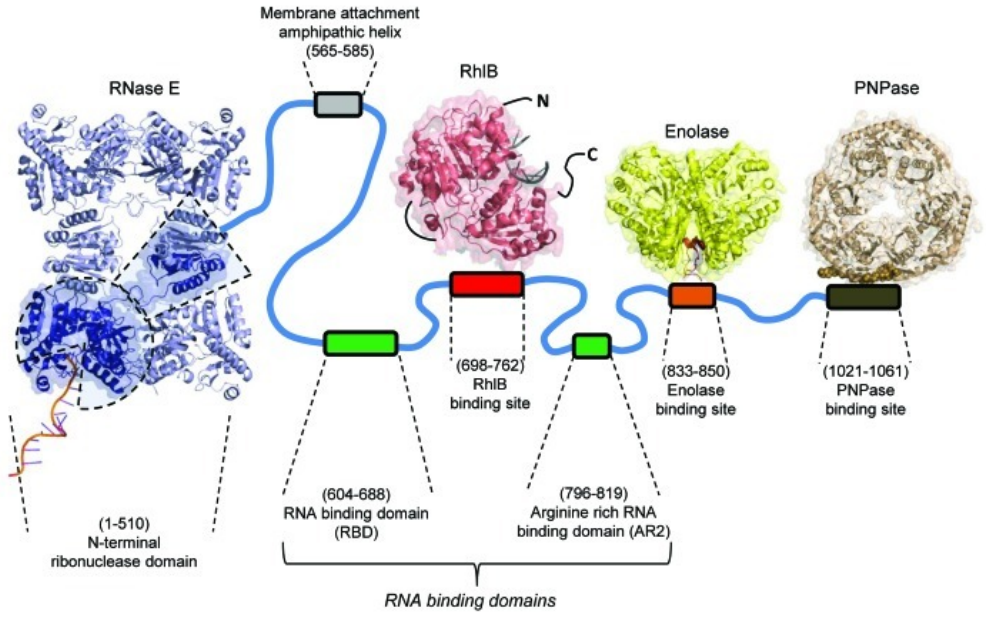


Figure 1.4 Domain structure of RNase E with the binding sites for RNA, helicase, enolase and PNPase. One putative protomer of the degradosome assembly is shown. (Four such protomers might form the core of the canonical degradosome.) The disordered C-terminal domain of RNase E is depicted as the thick, blue curvy line. The recognition sites for monomeric RhlB, dimeric enolase and trimeric PNPase are small segments of 20 to 40 amino acids represented as solid blocks on the C-terminus or RNase E. The catalytic N-terminal domain of RNase E is a tetramer (blue), and one protomer is highlighted with a dotted outline. The large domain harbors the catalytic site and is within the dashed circle, with the missing wedge engaging the RNA substrate; the small domain that organizes the dimer-of-dimer interfaces is outlined with the trapezoid. PDB codes for the structrures presented: RNase E, 2BX2; Vasa, the Drosophila homolog of *E. coli* RhlB, 2DB3; enolase, 3H8A; PNPase, 3GME. Figure is from (77).

degradosome activity *in vitro* (98), suggesting that its primary function in the protein complex is to serve as a scaffold for the other degradosome constituents. The nucleolytic activity of the *E. coli* degradosome is provided by PNPase, which degrades RNA from the 3' end. Like the eukaryotic exosome, PNPase has a toroidal structure that can encircle single-stranded but not double-stranded RNA (99,100). Thus, the role of RhlB in the degradosome is analogous to that of Ski2 and Mtr4 in the exosome. RNase E stimulates the ATPase activity of RhlB *in vitro* (101), and RhlB is required for PNPase degradation of transcripts with stable stem-loop structures (98). A minimal *E. coli* RNA degradosome of RNase E, PNPase and RhlB can be reconstituted *in vitro* without enolase, whose function in the degradosome is unknown. It is hypothesized that enolase may serve as a bridge between the metabolic status of the cell and post-transcriptional gene regulation.

Mycobacteria do not encode for a homolog of RhlB, but they possess a closely related DEAD-box helicase, RhlE. *E. coli* RhlE binds RNase E and can substitute for RhlB in the minimal degradosome *in vitro* (102). The physiological role of RhlE is unknown, though it has been implicated in the regulation of two other DEAD-box helicases in *E. coli* (103). Mycobacterial PNPase and RNase E have been purified and biochemically characterized (104-106), but the biochemical properties of mycobacterial RhlE were unknown. I purified recombinant *M. tuberculosis* RhlE and found it to have little ATPase or helicase activity. Regardless, I attempted to reconstitute a

mycobacterial degradosome *in vitro* with endoribonuclease-dead RNase E, PNPase and either HelY or RhlE. My attempts were unsuccessful.

The DEAD-box helicase CsdA has also been found to associate with the *E. coli* degradosome. CsdA is involved in ribosome biogenesis and its expression is up-regulated in *E. coli* following a cold shock (107,108). The helicase co-purifies with the RNA degradosome (including RhlB) from *E. coli* cultures grown at 15°C and can replace RhlB *in vitro* in the minimal degradosome (109). Mycobacteria encode a homolog of CsdA, which I have purified and characterized as an ATPase. Further investigation of its helicase activity and possible role in the degradosome is warranted.

REFERENCES

1. Hett, E.C. and Rubin, E.J. (2008) Bacterial growth and cell division: a mycobacterial perspective. *Microbiology and molecular biology reviews : MMBR*, **72**, 126-156.

2. Castaneda-Garcia, A., Prieto, A.I., Rodriguez-Beltran, J., Alonso, N., Cantillon, D., Costas, C., Perez-Lago, L., Zegeye, E.D., Herranz, M., Plocinski, P. *et al.* (2017) A non-canonical mismatch repair pathway in prokaryotes. *Nature communications*, **8**, 14246.

3. Stallings, C.L. and Glickman, M.S. (2010) Is Mycobacterium tuberculosis stressed out? A critical assessment of the genetic evidence. *Microbes and infection*, **12**, 1091-1101.

4. Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C. *et al.* (2003) Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment. *The Journal of experimental medicine*, **198**, 693-704.

5. Namouchi, A., Gomez-Munoz, M., Frye, S.A., Moen, L.V., Rognes, T., Tonjum, T. and Balasingham, S.V. (2016) The Mycobacterium tuberculosis transcriptional landscape under genotoxic stress. *BMC genomics*, **17**, 791.

6. Rand, L., Hinds, J., Springer, B., Sander, P., Buxton, R.S. and Davis, E.O. (2003) The majority of inducible DNA repair genes in Mycobacterium tuberculosis are induced independently of RecA. *Molecular microbiology*, **50**, 1031-1042.

7. Dos Vultos, T., Mestre, O., Tonjum, T. and Gicquel, B. (2009) DNA repair in Mycobacterium tuberculosis revisited. *FEMS microbiology reviews*, **33**, 471-487.

8. Singh, A. (2017) Guardians of the mycobacterial genome: A review on DNA repair systems in Mycobacterium tuberculosis. *Microbiology*, **163**, 1740-1758.

9. Jain, R., Kumar, P. and Varshney, U. (2007) A distinct role of formamidopyrimidine DNA glycosylase (MutM) in down-regulation of accumulation of G, C mutations and protection against oxidative stress in mycobacteria. *DNA repair*, **6**, 1774-1785.

10. Kurthkoti, K., Srinath, T., Kumar, P., Malshetty, V.S., Sang, P.B., Jain, R., Manjunath, R. and Varshney, U. (2010) A distinct physiological role of MutY in mutation prevention in mycobacteria. *Microbiology*, **156**, 88-93.

11. Dos Vultos, T., Blazquez, J., Rauzier, J., Matic, I. and Gicquel, B. (2006) Identification of Nudix hydrolase family members with an antimutator role in Mycobacterium tuberculosis and Mycobacterium smegmatis. *Journal of bacteriology*, **188**, 3159-3161.

12. Puri, R.V., Singh, N., Gupta, R.K. and Tyagi, A.K. (2013) Endonuclease IV Is the major apurinic/apyrimidinic endonuclease in Mycobacterium tuberculosis and is important for protection against oxidative damage. *PloS one*, **8**, e71535.

13. Purnapatre, K. and Varshney, U. (1998) Uracil DNA glycosylase from Mycobacterium smegmatis and its distinct biochemical properties. *European journal of biochemistry*, **256**, 580-588.

14. Srinath, T., Bharti, S.K. and Varshney, U. (2007) Substrate specificities and functional characterization of a thermo-tolerant uracil DNA glycosylase (UdgB) from Mycobacterium tuberculosis. *DNA repair*, **6**, 1517-1528.

15. Chan, S., Segelke, B., Lekin, T., Krupka, H., Cho, U.S., Kim, M.Y., So, M., Kim, C.Y., Naranjo, C.M., Rogers, Y.C. *et al.* (2004) Crystal structure of the Mycobacterium tuberculosis dUTPase: insights into the catalytic mechanism. *Journal of molecular biology*, **341**, 503-517.

16. Helt, S.S., Thymark, M., Harris, P., Aagaard, C., Dietrich, J., Larsen, S. and Willemoes, M. (2008) Mechanism of dTTP inhibition of the bifunctional dCTP deaminase:dUTPase encoded by Mycobacterium tuberculosis. *Journal of molecular biology*, **376**, 554-569.

17. Verhoeven, E.E., Wyman, C., Moolenaar, G.F. and Goosen, N. (2002) The presence of two UvrB subunits in the UvrAB complex ensures

damage detection in both DNA strands. *The EMBO journal*, **21**, 4196-4205.

18. Webster, M.P., Jukes, R., Zamfir, V.S., Kay, C.W., Bagneris, C. and Barrett, T. (2012) Crystal structure of the UvrB dimer: insights into the nature and functioning of the UvrAB damage engagement and UvrB-DNA complexes. *Nucleic acids research*, **40**, 8743-8758.

19. Moolenaar, G.F., Franken, K.L., Dijkstra, D.M., Thomas-Oates, J.E., Visse, R., van de Putte, P. and Goosen, N. (1995) The C-terminal region of the UvrB protein of Escherichia coli contains an important determinant for UvrC binding to the preincision complex but not the catalytic site for 3'-incision. *The Journal of biological chemistry*, **270**, 30508-30515.

20. Roberts, J. and Park, J.S. (2004) Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Current opinion in microbiology*, **7**, 120-125.

21. Williams, A., Guthlein, C., Beresford, N., Bottger, E.C., Springer, B. and Davis, E.O. (2011) UvrD2 is essential in Mycobacterium tuberculosis, but its helicase activity is not required. *Journal of bacteriology*, **193**, 4487-4494.

22. Kurthkoti, K., Kumar, P., Jain, R. and Varshney, U. (2008) Important role of the nucleotide excision repair pathway in Mycobacterium smegmatis in conferring protection against commonly encountered DNA-damaging agents. *Microbiology*, **154**, 2776-2785.

23. Dillingham, M.S. and Kowalczykowski, S.C. (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiology and molecular biology reviews : MMBR*, **72**, 642-671, Table of Contents.

24. Dillingham, M.S., Spies, M. and Kowalczykowski, S.C. (2003) RecBCD enzyme is a bipolar DNA helicase. *Nature*, **423**, 893-897.

25. Taylor, A.F. and Smith, G.R. (2003) RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. *Nature*, **423**, 889-893.

26. Glickman, M.S. (2014) Double-Strand DNA Break Repair in Mycobacteria. *Microbiology spectrum*, **2**.

27. Singleton, M.R., Dillingham, M.S., Gaudier, M., Kowalczykowski, S.C. and Wigley, D.B. (2004) Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature*, **432**, 187-193.

28. Anderson, D.G. and Kowalczykowski, S.C. (1997) The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell*, **90**, 77-86.

29. Kowalczykowski, S.C. (2000) Initiation of genetic recombination and recombination-dependent replication. *Trends in biochemical sciences*, **25**, 156-165.

30. 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. *Molecular microbiology*, **79**, 316-330.

31. Sinha, K.M., Unciuleac, M.C., Glickman, M.S. and Shuman, S. (2009) AdnAB: a new DSB-resecting motor-nuclease from mycobacteria. *Genes & development*, **23**, 1423-1437.

32. Unciuleac, M.C. and Shuman, S. (2010) Characterization of the mycobacterial AdnAB DNA motor provides insights into the evolution of bacterial motor-nuclease machines. *The Journal of biological chemistry*, **285**, 2632-2641.

33. Gupta, R., Unciuleac, M.C., Shuman, S. and Glickman, M.S. (2017) Homologous recombination mediated by the mycobacterial AdnAB helicase without end resection by the AdnAB nucleases. *Nucleic acids research*, **45**, 762-774.

34. Gupta, R., Shuman, S. and Glickman, M.S. (2015) RecF and RecR Play Critical Roles in the Homologous Recombination and Single-Strand Annealing Pathways of Mycobacteria. *Journal of bacteriology*, **197**, 3121-3132.

35. Chen, Z., Yang, H. and Pavletich, N.P. (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature*, **453**, 489-484.

36. Mazin, A.V. and Kowalczykowski, S.C. (1998) The function of the secondary DNA-binding site of RecA protein during DNA strand exchange. *The EMBO journal*, **17**, 1161-1168.

37. Zegeye, E.D., Balasingham, S.V., Laerdahl, J.K., Homberset, H. and Tonjum, T. (2012) Mycobacterium tuberculosis RecG binds and unwinds model DNA substrates with a preference for Holliday junctions. *Microbiology*, **158**, 1982-1993.

38. Thakur, R.S., Basavaraju, S., Somyajit, K., Jain, A., Subramanya, S., Muniyappa, K. and Nagaraju, G. (2013) Evidence for the role of Mycobacterium tuberculosis RecG helicase in DNA repair and recombination. *The FEBS journal*, **280**, 1841-1860.

39. Khanduja, J.S., Tripathi, P. and Muniyappa, K. (2009) Mycobacterium tuberculosis RuvA induces two distinct types of structural distortions between the homologous and heterologous Holliday junctions. *Biochemistry*, **48**, 27-40.

40. Khanduja, J.S. and Muniyappa, K. (2012) Functional analysis of DNA replication fork reversal catalyzed by Mycobacterium tuberculosis RuvAB proteins. *The Journal of biological chemistry*, **287**, 1345-1360.

41. Nautiyal, A., Rani, P.S., Sharples, G.J. and Muniyappa, K. (2016) Mycobacterium tuberculosis RuvX is a Holliday junction resolvase formed by dimerisation of the monomeric YqgF nuclease domain. *Molecular microbiology*, **100**, 656-674.

42. Gong, C., Bongiorno, P., Martins, A., Stephanou, N.C., Zhu, H., Shuman, S. and Glickman, M.S. (2005) Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nature structural & molecular biology*, **12**, 304-312.

43. Shuman, S. and Glickman, M.S. (2007) Bacterial DNA repair by non-homologous end joining. *Nature reviews. Microbiology*, **5**, 852-861.

44. Zhu, H. and Shuman, S. (2005) Novel 3'-ribonuclease and 3'-phosphatase activities of the bacterial non-homologous end-joining protein, DNA ligase D. *The Journal of biological chemistry*, **280**, 25973-25981.

45. Zhu, H. and Shuman, S. (2008) Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-OH monoribonucleotide. *The Journal of biological chemistry*, **283**, 8331-8339.

46. Zhu, H., Bhattarai, H., Yan, H.G., Shuman, S. and Glickman, M.S. (2012) Characterization of Mycobacterium smegmatis PolD2 and PolD1 as RNA/DNA polymerases homologous to the POL domain of bacterial DNA ligase D. *Biochemistry*, **51**, 10147-10158.

47. Gong, C., Martins, A., Bongiorno, P., Glickman, M. and Shuman, S. (2004) Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *The Journal of biological chemistry*, **279**, 20594-20606.

48. Ahel, I., Rass, U., El-Khamisy, S.F., Katyal, S., Clements, P.M., McKinnon, P.J., Caldecott, K.W. and West, S.C. (2006) The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature*, **443**, 713-716.

49. Moreira, M.C., Barbot, C., Tachi, N., Kozuka, N., Uchida, E., Gibson, T., Mendonca, P., Costa, M., Barros, J., Yanagisawa, T. *et al.* (2001) The gene mutated in ataxia-ocular apraxia 1 encodes the new HIT/Zn-finger protein aprataxin. *Nature genetics*, **29**, 189-193.

50. Dewhare, S.S., Umesh, T.G. and Muniyappa, K. (2015) Molecular and Functional Characterization of RecD, a Novel Member of the SF1 Family of Helicases, from Mycobacterium tuberculosis. *The Journal of biological chemistry*, **290**, 11948-11968.

51. Gupta, R., Ryzhikov, M., Koroleva, O., Unciuleac, M., Shuman, S., Korolev, S. and Glickman, M.S. (2013) A dual role for mycobacterial RecO in RecA-dependent homologous recombination and RecA-independent single-strand annealing. *Nucleic acids research*, **41**, 2284-2295.

52. Lehman, I.R., Bessman, M.J., Simms, E.S. and Kornberg, A. (1958) Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. *The Journal of biological chemistry*, **233**, 163-170.

53. De Lucia, P. and Cairns, J. (1969) Isolation of an E. coli strain with a mutation affecting DNA polymerase. *Nature*, **224**, 1164-1166.

54. Okazaki, R., Arisawa, M. and Sugino, A. (1971) Slow joining of newly replicated DNA chains in DNA polymerase I-deficient Escherichia coli mutants. *Proceedings of the National Academy of Sciences of the United States of America*, **68**, 2954-2957.

55. Brutlag, D., Atkinson, M.R., Setlow, P. and Kornberg, A. (1969) An active fragment of DNA polymerase produced by proteolytic cleavage. *Biochemical and biophysical research communications*, **37**, 982-989.

56. Klett, R.P., Cerami, A. and Reich, E. (1968) Exonuclease VI, a new nuclease activity associated with E. coli DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, **60**, 943-950.

57. Lyamichev, V., Brow, M.A. and Dahlberg, J.E. (1993) Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science*, **260**, 778-783.

58. Xu, Y., Derbyshire, V., Ng, K., Sun, X.C., Grindley, N.D. and Joyce, C.M. (1997) Biochemical and mutational studies of the 5'-3' exonuclease of DNA polymerase I of Escherichia coli. *Journal of molecular biology*, **268**, 284-302.

59. Fukushima, S., Itaya, M., Kato, H., Ogasawara, N. and Yoshikawa, H. (2007) Reassessment of the in vivo functions of DNA polymerase I and RNase H in bacterial cell growth. *Journal of bacteriology*, **189**, 8575-8583.

60. Gordhan, B.G., Andersen, S.J., De Meyer, A.R. and Mizrahi, V. (1996) Construction by homologous recombination and phenotypic characterization of a DNA polymerase domain polA mutant of Mycobacterium smegmatis. *Gene*, **178**, 125-130.

61. Hwang, K.Y., Baek, K., Kim, H.Y. and Cho, Y. (1998) The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. *Nature structural biology*, **5**, 707-713.

62. Tsutakawa, S.E., Classen, S., Chapados, B.R., Arvai, A.S., Finger, L.D., Guenther, G., Tomlinson, C.G., Thompson, P., Sarker, A.H., Shen, B. *et al.* (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell*, **145**, 198-211.

63. Orans, J., McSweeney, E.A., Iyer, R.R., Hast, M.A., Hellinga, H.W., Modrich, P. and Beese, L.S. (2011) Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell*, **145**, 212-223.

64. Anstey-Gilbert, C.S., Hemsworth, G.R., Flemming, C.S., Hodskinson, M.R., Zhang, J., Sedelnikova, S.E., Stillman, T.J., Sayers, J.R. and Artymiuk, P.J. (2013) The structure of Escherichia coli ExoIX--implications for DNA binding and catalysis in flap endonucleases. *Nucleic acids research*, **41**, 8357-8367.

65. AlMalki, F.A., Flemming, C.S., Zhang, J., Feng, M., Sedelnikova, S.E., Ceska, T., Rafferty, J.B., Sayers, J.R. and Artymiuk, P.J. (2016) Direct observation of DNA threading in flap endonuclease complexes. *Nature structural & molecular biology*, **23**, 640-646.

66. Singleton, M.R., Dillingham, M.S. and Wigley, D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annual review of biochemistry*, **76**, 23-50.

67. Ordonez, H. and Shuman, S. (2013) Mycobacterium smegmatis Lhr Is a DNA-dependent ATPase and a 3'-to-5' DNA translocase and helicase that prefers to unwind 3'-tailed RNA:DNA hybrids. *The Journal of biological chemistry*, **288**, 14125-14134.

68. Boshoff, H.I., Reed, M.B., Barry, C.E., 3rd and Mizrahi, V. (2003) DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis. *Cell*, **113**, 183-193.

69. Biswas, T. and Tsodikov, O.V. (2008) Hexameric ring structure of the N-terminal domain of Mycobacterium tuberculosis DnaB helicase. *The FEBS journal*, **275**, 3064-3071.

70. Zhang, H., Zhang, Z., Yang, J. and He, Z.G. (2014) Functional characterization of DnaB helicase and its modulation by single-stranded DNA binding protein in Mycobacterium tuberculosis. *The FEBS journal*, **281**, 1256-1266.

71. DeJesus, M.A., Gerrick, E.R., Xu, W., Park, S.W., Long, J.E., Boutte, C.C., Rubin, E.J., Schnappinger, D., Ehrt, S., Fortune, S.M. *et al.* (2017) Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. *mBio*, **8**.

72. Balasingham, S.V., Zegeye, E.D., Homberset, H., Rossi, M.L., Laerdahl, J.K., Bohr, V.A. and Tonjum, T. (2012) Enzymatic activities and DNA substrate specificity of Mycobacterium tuberculosis DNA helicase XPB. *PloS one*, **7**, e36960.

73. 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 research*, **40**, 4604-4614.

74. Yakovleva, L. and Shuman, S. (2012) Mycobacterium smegmatis SftH exemplifies a distinctive clade of superfamily II DNA-dependent ATPases with 3' to 5' translocase and helicase activities. *Nucleic acids research*, **40**, 7465-7475.

75. Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J.H., Chambon, P. and Egly, J.M. (1993) DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science*, **260**, 58-63.

76. Thakur, R.S., Desingu, A., Basavaraju, S., Subramanya, S., Rao, D.N. and Nagaraju, G. (2014) Mycobacterium tuberculosis DinG is a structure-specific helicase that unwinds G4 DNA: implications for targeting G4 DNA as a novel therapeutic approach. *The Journal of biological chemistry*, **289**, 25112-25136.

77. Hardwick, S.W. and Luisi, B.F. (2013) Rarely at rest: RNA helicases and their busy contributions to RNA degradation, regulation and quality control. *RNA biology*, **10**, 56-70.

78. Kalarickal, N.C., Ranjan, A., Kalyani, B.S., Wal, M. and Sen, R. (2010) A bacterial transcription terminator with inefficient molecular motor action but with a robust transcription termination function. *Journal of molecular biology*, **395**, 966-982.

79. Mitra, A., Misquitta, R. and Nagaraja, V. (2014) Mycobacterium tuberculosis Rho is an NTPase with distinct kinetic properties and a novel RNA-binding subdomain. *PloS one*, **9**, e107474.

80. Bernstein, J., Patterson, D.N., Wilson, G.M. and Toth, E.A. (2008) Characterization of the essential activities of Saccharomyces cerevisiae Mtr4p, a 3'->5' helicase partner of the nuclear exosome. *The Journal of biological chemistry*, **283**, 4930-4942.

81. Halbach, F., Reichelt, P., Rode, M. and Conti, E. (2013) The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell*, **154**, 814-826.

82. Weir, J.R., Bonneau, F., Hentschel, J. and Conti, E. (2010) Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 12139-12144.

83. Halbach, F., Rode, M. and Conti, E. (2012) The crystal structure of S. cerevisiae Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome. *Rna*, **18**, 124-134.

84. Jackson, R.N., Klauer, A.A., Hintze, B.J., Robinson, H., van Hoof, A. and Johnson, S.J. (2010) The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *The EMBO journal*, **29**, 2205-2216.

85. Chlebowski, A., Lubas, M., Jensen, T.H. and Dziembowski, A. (2013) RNA decay machines: the exosome. *Biochimica et biophysica acta*, **1829**, 552-560.

86. Zinder, J.C. and Lima, C.D. (2017) Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes & development*, **31**, 88-100.

87. Liu, Q., Greimann, J.C. and Lima, C.D. (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell*, **127**, 1223-1237.

88. Lorentzen, E., Dziembowski, A., Lindner, D., Seraphin, B. and Conti, E. (2007) RNA channelling by the archaeal exosome. *EMBO reports*, **8**, 470-476.

89. Synowsky, S.A. and Heck, A.J. (2008) The yeast Ski complex is a hetero-tetramer. *Protein science : a publication of the Protein Society*, **17**, 119-125.

90. Anderson, J.S. and Parker, R.P. (1998) The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *The EMBO journal*, **17**, 1497-1506.

91. Araki, Y., Takahashi, S., Kobayashi, T., Kajiho, H., Hoshino, S. and Katada, T. (2001) Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast. *The EMBO journal*, **20**, 4684-4693.

92. Falk, S., Weir, J.R., Hentschel, J., Reichelt, P., Bonneau, F. and Conti, E. (2014) The molecular architecture of the TRAMP complex reveals the organization and interplay of its two catalytic activities. *Molecular cell*, **55**, 856-867.

93. Schuch, B., Feigenbutz, M., Makino, D.L., Falk, S., Basquin, C., Mitchell, P. and Conti, E. (2014) The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. *The EMBO journal*, **33**, 2829-2846.

94. Liang, S., Hitomi, M., Hu, Y.H., Liu, Y. and Tartakoff, A.M. (1996) A DEAD-box-family protein is required for nucleocytoplasmic transport of yeast mRNA. *Molecular and cellular biology*, **16**, 5139-5146.

95. Fairman-Williams, M.E., Guenther, U.P. and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Current opinion in structural biology*, **20**, 313-324.

96. Carpousis, A.J., Van Houwe, G., Ehretsmann, C. and Krisch, H.M. (1994) Copurification of E. coli RNAase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell*, **76**, 889-900.

97. Py, B., Higgins, C.F., Krisch, H.M. and Carpousis, A.J. (1996) A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. *Nature*, **381**, 169-172.

98. Coburn, G.A., Miao, X., Briant, D.J. and Mackie, G.A. (1999) Reconstitution of a minimal RNA degradosome demonstrates functional coordination between a 3' exonuclease and a DEAD-box RNA helicase. *Genes & development*, **13**, 2594-2603.

99. Symmons, M.F., Jones, G.H. and Luisi, B.F. (2000) A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure*, **8**, 1215-1226.

100. Shi, Z., Yang, W.Z., Lin-Chao, S., Chak, K.F. and Yuan, H.S. (2008) Crystal structure of Escherichia coli PNPase: central channel residues are involved in processive RNA degradation. *Rna*, **14**, 2361-2371.

101. Vanzo, N.F., Li, Y.S., Py, B., Blum, E., Higgins, C.F., Raynal, L.C., Krisch, H.M. and Carpousis, A.J. (1998) Ribonuclease E organizes the protein interactions in the Escherichia coli RNA degradosome. *Genes & development*, **12**, 2770-2781.

102. Khemici, V., Toesca, I., Poljak, L., Vanzo, N.F. and Carpousis, A.J. (2004) The RNase E of Escherichia coli has at least two binding sites for DEAD-box RNA helicases: functional replacement of RhlB by RhlE. *Molecular microbiology*, **54**, 1422-1430.

103. Jain, C. (2008) The E. coli RhlE RNA helicase regulates the function of related RNA helicases during ribosome assembly. *Rna*, **14**, 381-389.

104. Unciuleac, M.C. and Shuman, S. (2013) Distinctive effects of domain deletions on the manganese-dependent DNA polymerase and DNA phosphorylase activities of Mycobacterium smegmatis polynucleotide phosphorylase. *Biochemistry*, **52**, 2967-2981.

105. Unciuleac, M.C. and Shuman, S. (2013) Discrimination of RNA from DNA by polynucleotide phosphorylase. *Biochemistry*, **52**, 6702-6711.

106. Zeller, M.E., Csanadi, A., Miczak, A., Rose, T., Bizebard, T. and Kaberdin, V.R. (2007) Quaternary structure and biochemical properties of mycobacterial RNase E/G. *The Biochemical journal*, **403**, 207-215.

107. Jones, P.G., Mitta, M., Kim, Y., Jiang, W. and Inouye, M. (1996) Cold shock induces a major ribosomal-associated protein that unwinds double-stranded RNA in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 76-80.

108. Charollais, J., Dreyfus, M. and Iost, I. (2004) CsdA, a cold-shock RNA helicase from Escherichia coli, is involved in the biogenesis of 50S ribosomal subunit. *Nucleic acids research*, **32**, 2751-2759.

109. Prud'homme-Genereux, A., Beran, R.K., Iost, I., Ramey, C.S., Mackie, G.A. and Simons, R.W. (2004) Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: evidence for a 'cold shock degradosome'. *Molecular microbiology*, **54**, 1409-1421.

CHAPTER TWO

Mycobacterium smegmatis HelY is an RNA-activated ATPase/dATPase that unwinds 3'-tailed RNA duplexes and RNA:DNA hybrids[*]

INTRODUCTION

Helicases are nucleic acid-dependent nucleoside triphosphatases (NTPases) that play important roles in diverse nucleic acid transactions. DNA helicases orchestrate replication, recombination, and repair. RNA helicases choreograph transcription, RNA processing, ribosome biogenesis, translation, and RNA turnover. Helicases use the chemical energy of nucleoside triphosphate (NTP) hydrolysis either to effect mechanical changes in the secondary structure of nucleic acids or to remodel (or disrupt) the structures of protein-nucleic acid complexes. Helicases are classified into superfamilies, families, and subfamilies according to their distinctive primary, tertiary, and quaternary structures and their biochemical specificities, i.e., NTP preference, nucleic acid preference (duplex DNA, duplex RNA, or RNA:DNA hybrids), and directionality of translocation or unwinding (5′ to 3′ or 3′ to 5′) (1).

Differences in the roster of helicases between taxa offer useful clues to the evolution and diversification of DNA replication/repair and RNA synthesis/processing strategies. Where the helicase rosters diverge in animals versus pathogens, they can suggest anti-infective drug targets. With this in

[*] Uson, M.L., Ordonez, H. and Shuman, S. (2015) *Mycobacterium smegmatis* HelY is an RNA-activated ATPase/dATPase and 3′-to-5′ helicase that unwinds 3′-tailed RNA duplexes and RNA:DNA hybrids. *J. Bacteriol.*, **197**, 3057-3065.

mind, we are focused on the helicases of members of the genus *Mycobacterium*, a genus of the phylum *Actinobacteria* that includes the human pathogen *Mycobacterium tuberculosis* and its avirulent relative, *M. smegmatis*. To date, we have purified and characterized six mycobacterial DNA helicases: AdnAB (2-4), UvrD1 (5, 6), UvrD2 (7), SftH (8), RqlH (9), and Lhr (10). Mycobacterial DNA helicases XBP (11, 12), RecG (13-15), RecD (16), RuvB (17), and DnaB (18) have been purified and characterized by other investigators.

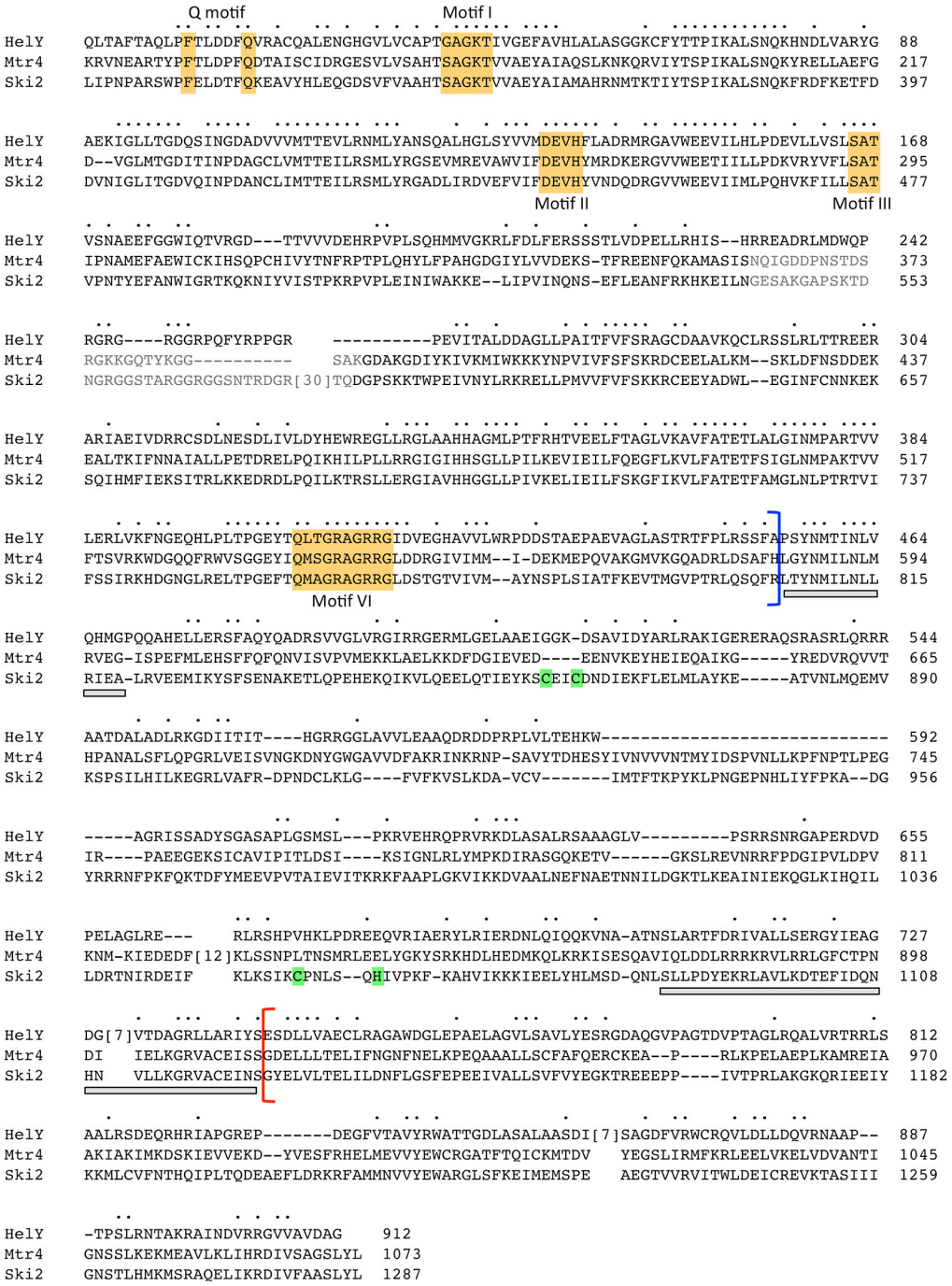
Comparatively little attention has been paid to the mycobacterial RNA helicase repertoire. The transcription termination factor Rho is, to our knowledge, the lone example of a biochemically characterized NTP-driven RNA translocase/helicase from a *Mycobacterium* species (19-21). In the present study, we aimed to biochemically characterize *M. smegmatis* MSMEG_3885, a 912-amino-acid (aa) protein and a plausible RNA helicase candidate. Its counterpart in *M. tuberculosis* is Rv2092c, a 906-aa protein that has been annotated as HelY (albeit in the absence of any functional information). The *M. smegmatis* and *M. tuberculosis* HelY polypeptides are homologous across their entire lengths, with 713 positions of amino acid side chain identity, and we presume that they are functionally orthologous.

Our interest in HelY was sparked by its similarity to two structurally characterized DEXH-family helicases implicated in eukaryal RNA processing: Mtr4 and Ski2 (22-25). Figure 2.1 shows a COBALT alignment of the primary structure of *M. smegmatis* HelY to Saccharomyces cerevisiae Mtr4 (1,073 aa)

and *S. cerevisiae* Ski2 (1,287 aa). HelY lacks a counterpart of the N-terminal extensions of the longer Mtr4 and Ski2 polypeptides that precede their respective NTPase domains. The N-terminal 454-aa segment of HelY strongly resembles the NTPase catalytic domains of Mtr4, Ski2, and other superfamily II helicases, which consist of two tandem RecA-like modules that contain a set of signature peptide motifs highlighted in Figure 2.1: (i) a Q motif (FXXXXQ) that engages the nucleobase of the NTP substrate; (ii) motif I (GXGKT; the P loop or Walker A box), the lysine side chain of which contacts the β and γ phosphates of the NTP substrate; (ii) motif II (DEXH; the Walker B-box), which coordinates the metal cofactor; (iii) motif VI (QXXGRXGRR), 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 (1, 22-25). The aligned NTPase domains of HelY, Mtr4, and Ski2 (demarcated by the blue bracket in Figure 2.1) include 193 positions of side chain identity/similarity in all three polypeptides.

The segment of Mtr4 and Ski2 distal to the NTPase domain is composed of several discrete tertiary structure modules, including (i) a split winged helix domain into which is inserted a large arch domain consisting of a bent helical stalk and a β barrel and (ii) a large C-terminal α-helical bundle domain, the so-called ratchet module. The putative equivalent of the ratchet domain in HelY (aa 750 to 912; demarcated by the red bracket in Figure 2.1) aligns to Mtr4 and Ski2 with 36 positions of side chain identity/similarity in all

Figure 2.1 *M. smegmatis* HelY is homologous to eukaryal Mtr4 and Ski2. The amino acid sequence of the 912-aa *M. smegmatis* HelY polypeptide is aligned to the sequences of the *S. cerevisiae* Mtr4 (1,073-aa) and Ski2 (1,287-aa) proteins. Positions of amino acid side chain identity/similarity in all three polypeptides are denoted by dots. Gaps in the alignments are denoted by dashes; numbers in brackets indicate the number of amino acids inserted. The distal margin of the NTPase domain is demarcated by the blue bracket. The superfamily II ATPase motifs Q, I, II, III, and VI of the NTPase domain are labeled and highlighted in gold shading. The two segments comprising the split winged helix domain are underscored by gray bars. The constituents of the CCCH zinc-binding motif in Ski2 are highlighted in green. The start of the C-terminal ratchet domain is indicated by the red bracket.


Q motif
Motif I
Motif II
Motif III
Motif VI
HelY
Mtr4
Ski2

three polypeptides. The putative segments that form the split winged helix domain in HelY (aa 455 to 468 and 706 to 749; underscored in Figure 2.1) align to Mtr4 and Ski2 with 17 positions of side chain identity/similarity in all three proteins. The putative arch domain in HelY is less well conserved among the three proteins and is punctuated by multiple gaps and insertions in the sequence alignment. Neither HelY nor Mtr4 has the zinc-binding CCCH motif located at the bend in the helical stalk of the Ski2 protein (highlighted in green in Figure 2.1) (24).

To our knowledge, there has been no prior characterization of a bacterial homolog of Mtr4/Ski2. To rectify this knowledge gap, we produced, purified, and characterized recombinant M. smegmatis HelY. The following were key questions of interest to us (i) Is HelY a nucleic acid-dependent phosphohydrolase, and, if so, what are its substrate and cofactor specificities? (ii) Can HelY couple NTP hydrolysis to duplex unwinding? (iii) Do the biochemical activities of HelY point to a role in DNA or RNA transactions?

RESULTS

Recombinant HelY is an ATPase/dATPase

To evaluate the enzymatic properties of HelY, we produced the protein in *E. coli* as a His_{10}Smt3 fusion and isolated it from a soluble extract by nickel-agarose chromatography. The His_{10}Smt3 tag was removed with the Smt3-specific protease Ulp1, and the native HelY protein was separated from the tag by a second round of Ni-agarose chromatography. The tag-free HelY was further purified by gel filtration, during which HelY eluted as a single peak that,

by reference to size standards, was monomeric (not shown). SDS-PAGE of the purified protein revealed a predominant ~100-kDa polypeptide corresponding to HelY (Figure 2.2A). Reaction of wild-type HelY with 1 mM [α-^{32}P]ATP in the presence of magnesium and poly(U) resulted in the hydrolysis of [α-^{32}P]ATP to [α-^{32}P]ADP (Figure 2.2B). Mutated versions of the HelY protein in which Lys48 in motif I or Asp135 in motif II was changed to alanine were prepared (Figure 2.2A). The K48A and D135A mutations effaced ATPase activity (Figure 2.2B), signifying that the observed phosphohydrolase activity inheres to the recombinant HelY protein.

NTP substrate specificity was examined by colorimetric assay of the release of P_i from unlabeled ribonucleotide ATP, CTP, GTP, or UTP and deoxyribonucleotide dATP, dCTP, dGTP, or dTTP, each at a 1 mM concentration. HelY hydrolyzed ATP and dATP but was unreactive with other nucleotides (Figure 2.2C). The adenine NTP specificity of HelY is consistent with the presence of a Q motif (^{19}FXXXXXQ^{25} in HelY) that, in the Mtr4/Ski2 structures (23, 24), engages the adenine nucleobase via π stacking of the phenylalanine on the purine ring and via hydrogen bonds from the eponymous glutamine to adenine N-7 and N-6.

Omission of magnesium from the phosphohydrolase reaction mixture reduced ATP hydrolysis by 5-fold (Figure 2.3A). ATPase activity was abolished when 5 mM EDTA was included in reaction mixtures lacking exogenous divalent cation, suggesting that residual metal might have been carried along with HelY during the enzyme preparation. When various metals

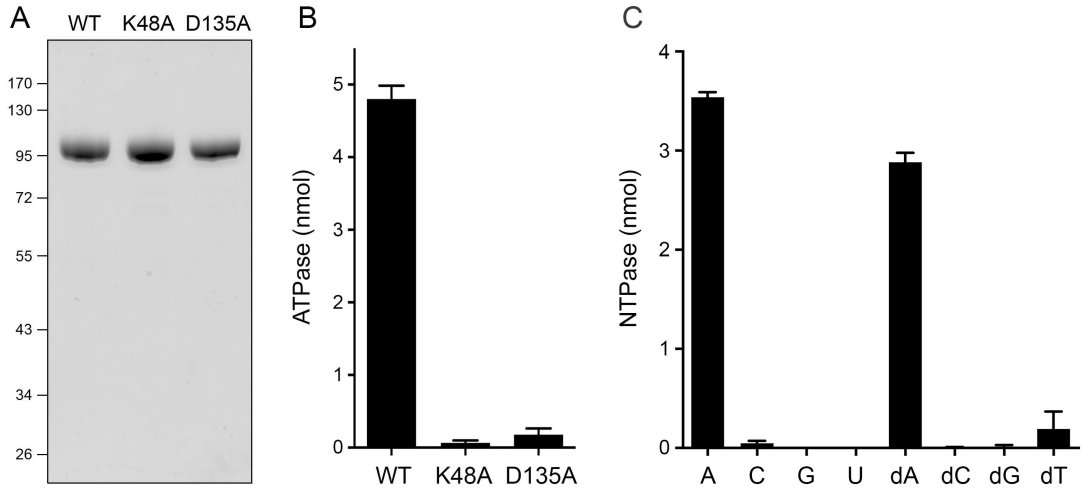


Figure 2.2 HelY is an ATPase/dATPase. (A) Aliquots (5 µg) of purified wild-type (WT) HelY and K48A and D135A mutant HelY were analyzed by SDS-PAGE. The Coomassie blue-stained gel is shown. The positions and sizes (in kilodaltons) of marker polypeptides are indicated on the left. (B) ATP hydrolysis. Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 1 mM [α-^{32}P]ATP (10 nmol ATP), 200 nM poly(U) (as UMP nucleotide), and 10 pmol (1 µM) wild-type or mutant HelY, as specified, were incubated at 37°C for 20 min. The extents of ATP hydrolysis are plotted. Each datum is the average ± SEM from three separate experiments. (C) NTP specificity. Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 200 nM poly(U), 10 pmol HelY, and 1 mM the indicated NTP or dNTP were incubated at 37°C for 20 min. The reactions were quenched with 990 µl of malachite green reagent (Biomol Research Laboratories). Phosphate release was quantified by measuring the A_{620} and interpolating the value into a phosphate standard curve. The values were corrected for the low levels of phosphate measured in control reaction mixtures containing 1 mM NTP/dNTP but no added enzyme. Data are the averages ± SEMs from three separate experiments.

were provided at a 5 mM concentration, magnesium proved to be the most effective cofactor (Figure 2.3A). Manganese was 68% as effective as magnesium. The ATPase activity in the presence of calcium, cobalt, copper, nickel, or zinc did not rise above the background level. In the presence of magnesium, HelY hydrolyzed ATP over a broad pH range (from pH 5.5 to 9.5 in Tris buffers). Activity declined sharply at pH 5.0 and was virtually nil at pH

4.5 (Figure 2.3B).

The extent of ATP hydrolysis during a 20-min reaction increased with an increase in the level of input HelY protein in a biphasic manner (Figure 2.4A). From the slope of the titration curve at low HelY levels, we derived a turnover number of 140 min^{-1}. The kinetic profile of the hydrolysis of 1 mM ATP by 1 μM HelY was also biphasic (Figure 2.4B). From the initial rate, we calculated a turnover number of 136 min^{-1}. The biphasic patterns, whereby ATP consumption diminished as the reaction proceeded, suggested that product accumulation (either ADP or P_i) might inhibit the ATPase reaction of HelY. Indeed, we found that inclusion of 0.25 to 4 mM ADP in the reaction mixture (containing 1 mM ATP and 10 mM magnesium) elicited a progressive

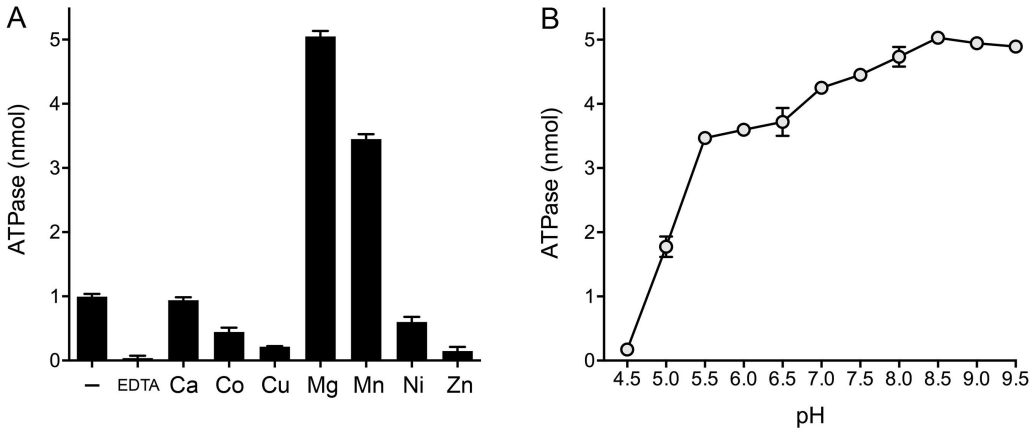
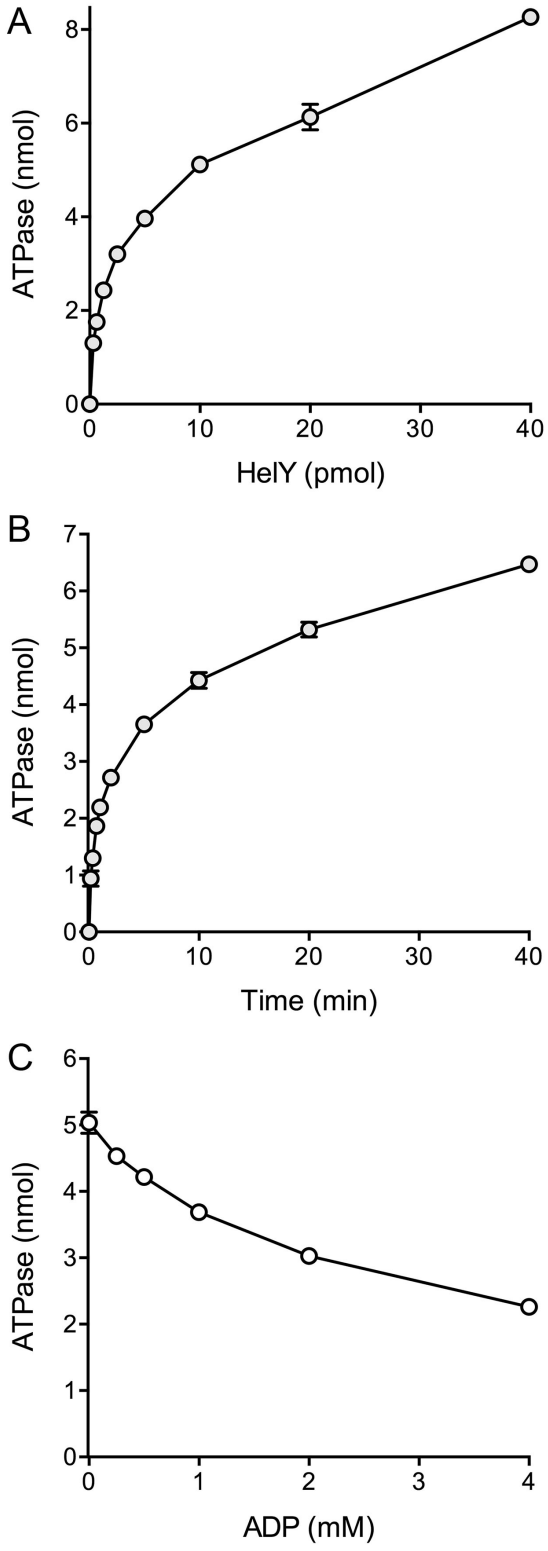


Figure 2.3 Divalent cation specificity and pH profile. (A) Reaction mixtures (10 μl) containing 20 mM Tris-HCl, pH 8.0, 1 mM [α-^{32}P]ATP, 200 nM poly(U), 10 pmol HelY, and either no added divalent cation (−), no added divalent cation plus 5 mM EDTA (EDTA), or 5 mM the indicated divalent cation (chloride salt) were incubated at 37°C for 20 min. (B) Reaction mixtures (10 μl) containing 20 mM Tris buffer (Tris acetate at pH 4.5, 5.0, 5.5, 6.0 or 6.5; Tris-HCl at pH 7.0, 7.5, 8.0, 8.5, 9.0, or 9.5), 5 mM $MgCl_2$, 1 mM [α-^{32}P]ATP, 200 nM poly(U), and 10 pmol HelY were incubated for 20 min. The extents of ATP hydrolysis are plotted. Data are the averages ± SEMs from three separate experiments.

Figure 2.4 Characterization of HelY ATPase activity. (A) Enzyme titration. Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 1 mM [α-^{32}P]ATP, 200 nM poly(U), and HelY as indicated were incubated at 37°C for 20 min. (B) Time course. A reaction mixture containing 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 1 mM [α-^{32}P]ATP, 200 nM poly(U), and 1 µM HelY was incubated at 37°C. Aliquots (10 µl containing 10 nmol ATP) were withdrawn at the times specified and quenched with formic acid. (C) Product inhibition by ADP. Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 1 mM [α-^{32}P]ATP, 200 nM poly(U), 10 pmol HelY, and ADP at the indicated concentration were incubated at 37°C for 20 min. The extents of ATP hydrolysis are plotted. Each datum is the average ± SEM from three separate experiments.


A
ATPase (nmol)
0
2
4
6
8
0
10
20
30
40
HelY (pmol)
B
ATPase (nmol)
0
1
2
3
4
5
6
7
0
10
20
30
40
Time (min)
C
ATPase (nmol)
0
1
2
3
4
5
6
0
1
2
3
4
ADP (mM)

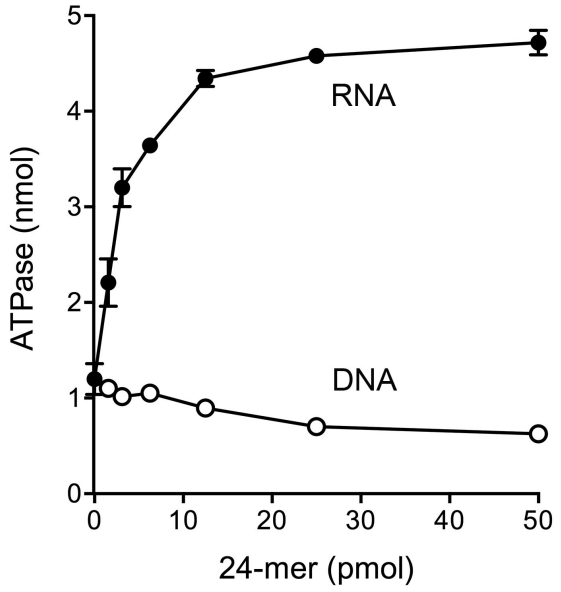
decrease in the extent of ATP hydrolysis (Figure 2.4C). In contrast, inclusion of inorganic phosphate up to 20 mM had no effect on the extent of ATP hydrolysis (data not shown).

Nucleic acid cofactor dependence of ATP hydrolysis

We tested 24-mer RNA and DNA oligonucleotides of identical nucleobase sequence (except T for U in DNA) as cofactors for the HelY ATPase. A basal level of ATP hydrolysis was observed in the absence of exogenous nucleic acid. Activity was enhanced by the 24-mer RNA in a concentration-dependent manner, eliciting a 4-fold increase in ATP hydrolysis at a saturating RNA concentration (Figure 2.5). In contrast, the 24-mer DNA had no salutary effect (Figure 2.5).

HelY is a helicase that requires a 3'-tailed RNA loading strand

The role of RNA as a cofactor for the ATPase activity of HelY raises the question of whether HelY can unwind RNA-containing duplexes. To address this issue, we prepared a series of tailed duplex helicase substrates, by annealing a 5′ ^{32}P-labeled 24-mer (either RNA or DNA) to a complementary 39-mer (either RNA or DNA) to form a 24-bp duplex with a 15-nucleotide 3′ single-strand tail, which was either oligo$(U)_{15}$ or oligo$(dT)_{15}$ (Figure 2.6). The helicase assay format that we used entailed preincubation of HelY with the tailed duplex substrate, after which the helicase reaction was initiated by simultaneous addition of ATP and a trap of excess unlabeled 24-mer strand. The trap strand minimizes reannealing of any ^{32}P-labeled 24-mer that was unwound by HelY and competes for binding to any free HelY or HelY that



RNA 5′ GGGUCGCAAUUGUAUUCCGAUAGUG

DNA 5′ GGGTCGCAATTGTATTCCGATAGTG

Figure 2.5 Nucleic acid cofactor dependence of ATP hydrolysis. Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 1 mM [α-^{32}P]ATP, 10 pmol HelY, and 24-mer single-strand DNA or RNA oligonucleotide, as specified, were incubated at 37°C for 20 min.

dissociated from the labeled substrate without unwinding it. Consequently, the assay predominantly gauges a single round of strand displacement by HelY bound to the labeled 3′-tailed duplex prior to the onset of ATP hydrolysis.

The instructive initial findings were that HelY unwound the 3′-tailed RNA:RNA and RNA (39-mer):DNA (24-mer) duplexes to yield a radiolabeled free RNA or DNA single strand that comigrated during native PAGE with the free 24-mers generated by thermal denaturation of the substrates (Figure 2.6). HelY failed to unwind a 3′-tailed DNA:DNA duplex or a DNA (39-mer):RNA (24-mer) hybrid in which the tailed strand was DNA and the labeled strand was RNA (Figure 2.6).

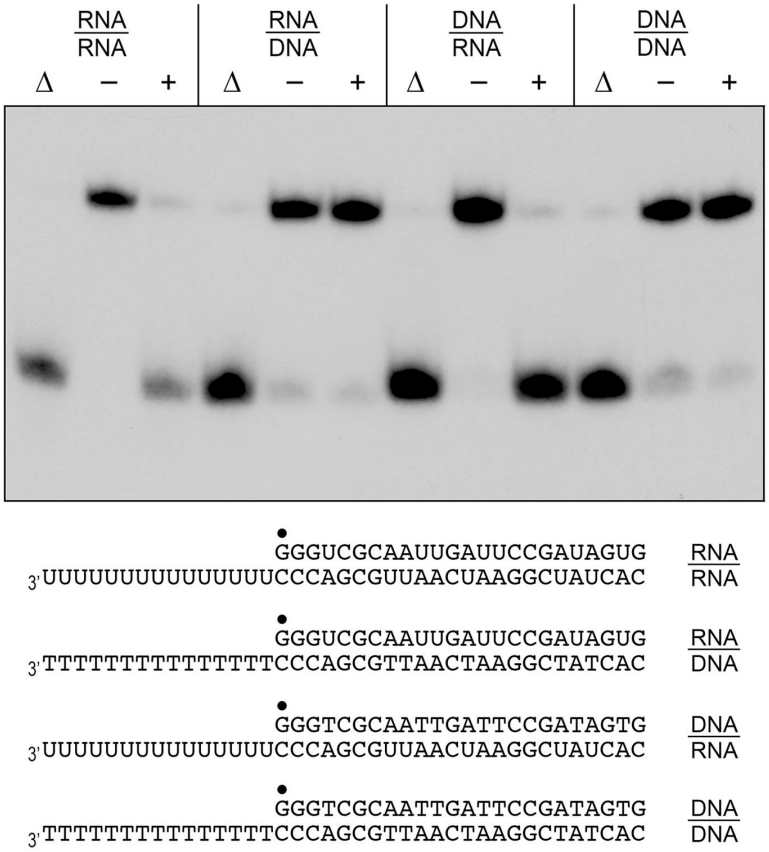


Figure 2.6 HelY helicase activity. Helicase reaction mixtures (10 µl) contained 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 1 pmol (100 nM) of the indicated 3′-tailed RNA:RNA, RNA:DNA, DNA:RNA, or DNA:DNA duplex substrates (depicted at the bottom, with the 5′ ^{32}P label denoted by •), 1 mM ATP, and either no enzyme (lanes −) or 10 pmol (1 µM) HelY (lanes +). The mixtures were analyzed by native PAGE, and radiolabeled nucleic acids were visualized by autoradiography. Reaction mixtures lacking enzyme that were heat denatured prior to PAGE were analyzed in lanes Δ. (24-mer) hybrid in which the tailed strand was DNA and the labeled strand was RNA.

To investigate the role of the single-strand tail in duplex unwinding, we prepared a set of three substrates, each with a 24-bp ^{32}P-labeled DNA:RNA duplex segment, that had either (i) a 3′ oligo$(U)_{15}$ single-strand tail on the RNA strand; (ii) a 5′ oligo$(U)_{15}$ single-strand tail on the RNA strand; or (iii) no single-strand tail, i.e., a blunt duplex (Figure 2.7). Whereas HelY efficiently unwound

the 3′-tailed duplex, it failed to unwind either the blunt duplex or the 5′-tailed duplex substrates (Figure 2.7).

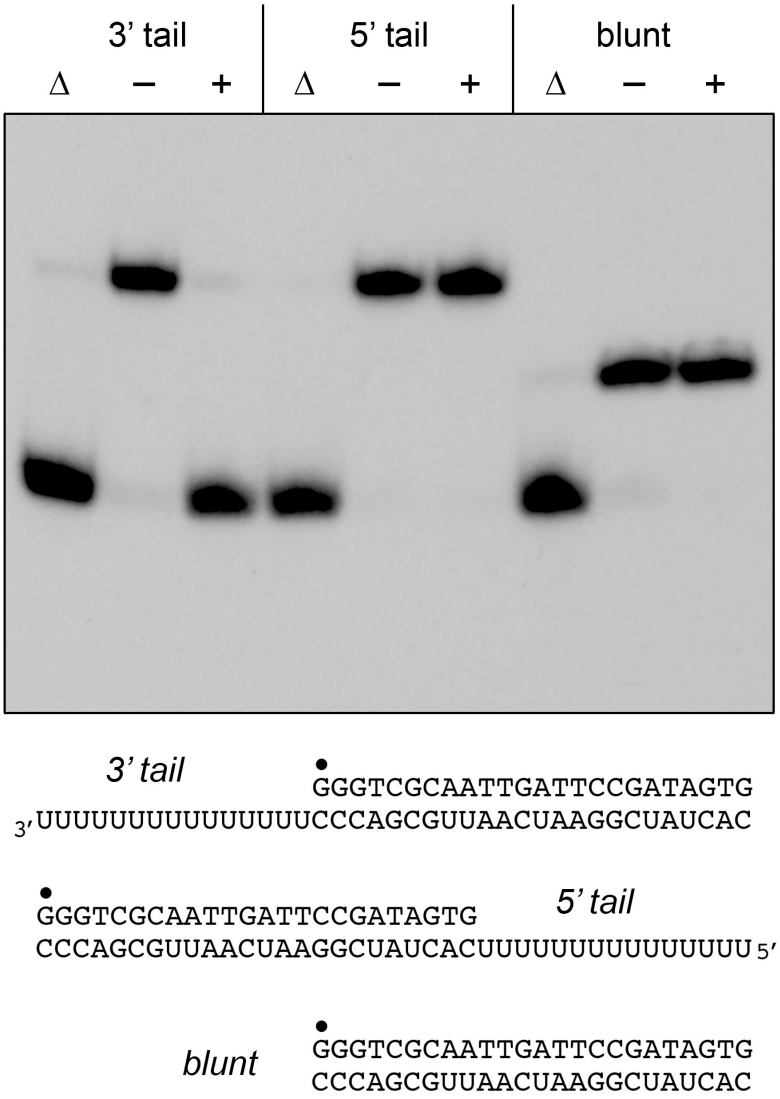


Figure 2.7 Requirement for a 3′ single-strand tail on the RNA strand. Helicase reaction mixtures (10 µl) contained 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 1 mM ATP, 1 pmol (100 nM) of the indicated 3′-tailed, 5′-tailed, or blunt ^{32}P-labeled DNA:RNA hybrid substrates, and either no enzyme (lanes −) or 10 pmol (1 µM) HelY (lanes +). The mixtures were analyzed by native PAGE, and radiolabeled nucleic acids were visualized by autoradiography. Reaction mixtures lacking enzyme that were heat denatured prior to PAGE were analyzed in lanes Δ.

The extent of unwinding of the 3′-tailed ^{32}P-labeled DNA:RNA duplex in the single-turnover assay format was proportional to the level of input HelY (Figure 2.8A). Helicase activity was abolished by the K48A and D135A mutations (Figure 2.8B). No duplex unwinding was observed when ATP was omitted from the reaction mixture (Figure 2.9). The NTP requirement for the HelY helicase was satisfied by ATP or dATP but not GTP, CTP, UTP, dGTP, dCTP, or dTTP (Figure 2.9).

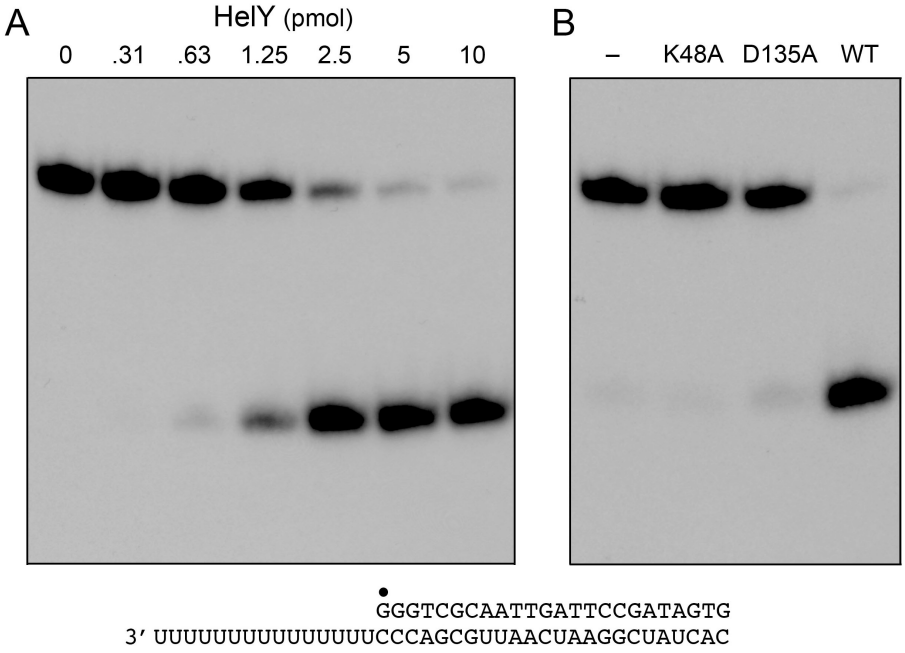


Figure 2.8 HelY titration and mutational inactivation of HelY helicase. (A) Titration. Reaction mixtures (10 µl) contained 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 1 mM ATP, 1 pmol (100 nM) 3′-tailed ^{32}P-labeled DNA (24-mer):RNA (39-mer) substrate, and HelY as specified. (B) Mutational inactivation. Reaction mixtures (10 µl) contained 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 1 mM ATP, 1 pmol (100 nM) 3′-tailed ^{32}P-labeled DNA:RNA substrate, and 10 pmol (1 µM) wild-type or mutant HelY, as specified. HelY was omitted from the control reaction mixture in lane −. The mixtures were analyzed by native PAGE, and radiolabeled nucleic acids were visualized by autoradiography.

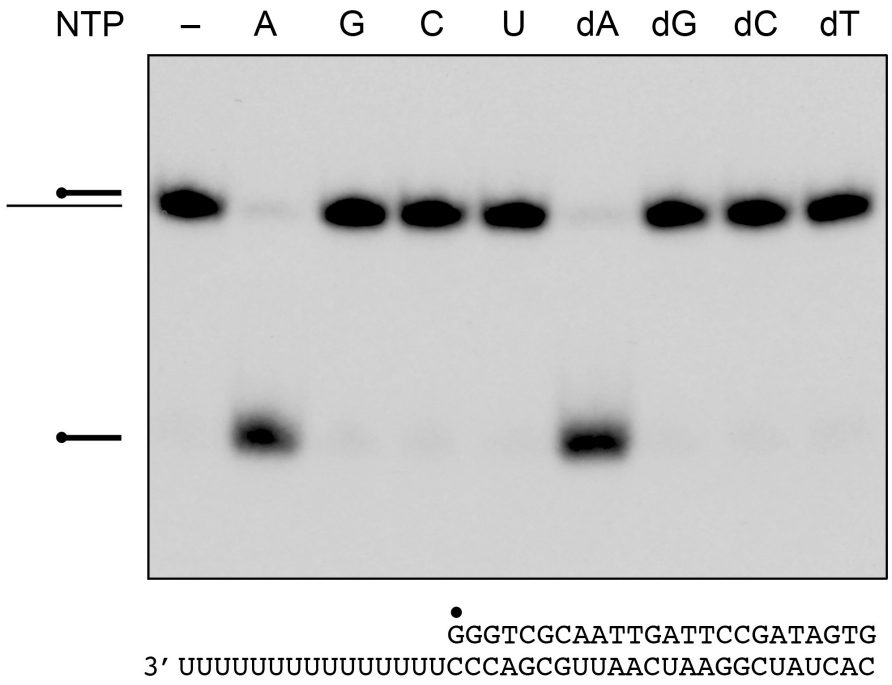


Figure 2.9 NTP requirement of the HelY helicase. Reaction mixtures (10 µl) contained 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 1 pmol (100 nM) 3′-tailed ^{32}P-labeled DNA (24-mer):RNA (39-mer) substrate, 10 pmol (1 µM) HelY, and either no added NTP (lane −) or 1 mM the indicated NTP or dNTP. The reaction products were analyzed by native PAGE and visualized by autoradiography.

DISCUSSION

The results presented here indicate that *M. smegmatis* HelY is a unidirectional motor, powered by RNA-dependent ATP/dATP hydrolysis, that tracks 3′ to 5′ along a loading RNA strand to displace the complementary strand of a tailed RNA:RNA or RNA:DNA duplex. The findings that HelY ATPase is unresponsive to a DNA polynucleotide cofactor and that HelY is unable to unwind a 3′-tailed duplex in which the loading strand is DNA distinguish HelY from all other mycobacterial NTPases/helicases characterized previously. Whereas many of the other mycobacterial NTPases/helicases are implicated biochemically and genetically in DNA

replication, recombination, or repair (2, 5, 7, 14, 26-29), the biochemical activities of HelY are consistent with a role in mycobacterial RNA metabolism.

The structural homology of HelY to eukaryal Mtr4 and Ski2, inferred from their aligned amino acid sequences (Figure 2.1), accords with shared biochemical parameters. Recombinant yeast Mtr4, like HelY, is an RNA-dependent ATPase/dATPase that unwinds a tailed RNA duplex with 3′-to-5′ polarity (30). The observed turnover numbers for ATP hydrolysis by Mtr4 (120 to 163 min^{-1}) (30, 31) are similar to what we observe for HelY (140 min^{-1}). Similarly, recombinant yeast Ski2 is an RNA-dependent ATPase (turnover number, 180 min^{-1}) that can unwind a 3′-tailed RNA:RNA duplex (32).

The eukaryal Mtr4 and Ski2 helicases, respectively, modulate the activity of the nuclear and cytoplasmic forms of the eukaryal exosome, a macromolecular machine that resects RNA 3′ ends in the service of either RNA maturation or RNA decay (25). A parsimonious model posits that Mtr4/Ski2 catalyzes unwinding of RNA secondary structures or displacement of RNA-bound proteins that might otherwise impede the exosome. Both Mtr4 and Ski2 function *in vivo* in the context of multisubunit complexes (TRAMP and the Ski complex, respectively) (25).

The bacterial exosome analog is polynucleotide phosphorylase (PNPase) (33). *E. coli* PNPase comprises part of a larger macromolecular RNA processing machine, the RNA degradosome, that includes an RNA helicase, RhlB (34, 35). Recombinant mycobacterial PNPase has been purified and characterized biochemically (36, 37), but its potential interaction

partners and genetic dissection of its physiology are uncharted territory. Indeed, little is known about RNA turnover in mycobacteria. Transposon mutagenesis showed that M. tuberculosis HelY (Rv2092c) is inessential for bacterial growth in culture (38, 39) and suggested that PNPase is essential. It is conceivable that HelY participates in aspects of RNA metabolism that are optional under laboratory growth conditions; alternatively, mycobacteria might encode other RNA helicases that are functionally redundant with HelY.

Homologs of HelY are evident in the proteomes of most species of *Mycobacterium*, including *M. leprae*. HelY is widely prevalent among the *Actinobacteria* (*Nocardia*, *Streptomyces*, *Rhodococcus*, *Gordonia*, etc.) and *Cyanobacteria* (*Nostoc*, *Anabaena*, *Synechococcus*, *Prochlorococcus*, etc.). HelY homologs are found sporadically elsewhere in the bacterial domain: in the *Chloroflexi* (*Sphaerobacter*, *Thermomicrobium*, *Nitrolancea*, and *Thermorudis*) *Proteobacteria* (*Burkholderia* and *Mesorhizobium*), *Firmicutes* (*Dehalobacter*), and *Chlamydiae* (*Chlamydia trachomatis*).

EXPERIMENTAL PROCEDURES

Recombinant *M. smegmatis* HelY

The open reading frame encoding full-length *M. smegmatis* HelY (MSMEG_3885) was PCR amplified from *M. smegmatis* genomic DNA with primers that introduced a BglII site immediately flanking the start codon and a HindIII site downstream of the stop codon. The PCR product was digested with BglII and HindIII and ligated into pET28b-His_{10}Smt3 that had been digested with BamHI and HindIII. The resulting pET28b-His_{10}Smt3-HelY

expression plasmid encodes the HelY polypeptide fused to an N-terminal His_{10}Smt3 tag under the transcriptional control of a T7 RNA polymerase promoter. Plasmids encoding mutants HelY-(K48A) and HelY-(D135A) were obtained by site-directed mutagenesis using the QuikChange method with NEB Phusion polymerase. The plasmid inserts were sequenced to verify that no unintended coding changes were acquired during amplification and cloning.

The pET28b-His_{10}Smt3-HelY plasmids were transformed into *Escherichia coli* BL21(DE3) cells. Cultures (1 liter) amplified from single kanamycin-resistant transformants were grown at 37°C in Terrific broth (TB) containing 0.4% glycerol and 60 μg/ml kanamycin until the A_{600} reached 0.8. The cultures were chilled on ice for 1 h and then adjusted to 2% (vol/vol) ethanol and 0.3 mM isopropyl-β-D-thiogalactopyranoside and incubated for 16 to 18 h at 17°C with constant shaking. All subsequent steps were performed at 4°C. Cells were harvested by centrifugation and resuspended in 25 ml of buffer A (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% sucrose, 20 mM imidazole) containing 1 protease inhibitor cocktail tablet (Roche) and 1 mM dithiothreitol (DTT). Lysozyme was added to a concentration of 1 mg/ml. After incubation for 1 h, the lysate was sonicated to reduce viscosity and the insoluble material was pelleted by centrifugation at 38,000 × g for 35 min. The supernatant was mixed for 1 h with 3 ml of Ni-nitrilotriacetic acid (NTA)-agarose resin (Qiagen) that had been equilibrated with buffer A. The resin was recovered by centrifugation and resuspended in 30 ml of buffer A. The washed resin was then recovered by centrifugation, resuspended in 20 ml of buffer B

(50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol) containing 20 mM imidazole, and poured into a column. The column was washed serially with 4 column volumes of 3 M KCl, 5 column volumes of buffer B containing 20 mM imidazole, and 4 column volumes of 50 mM imidazole in buffer B. The bound protein was eluted with 500 mM imidazole in buffer B. The polypeptide compositions of the fractions were monitored by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE). The 500 mM imidazole eluates containing His_{10}Smt3-HelY were supplemented with the Smt3-specific protease Ulp1 (60 μg) and then dialyzed against 2 liters of buffer C (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM DTT, 20 mM imidazole, 10% glycerol) for 16 h, during which time the His_{10}Smt3 tag was cleaved. The dialysates were mixed with 3 ml of Ni-NTA-agarose equilibrated with buffer C. The mixtures were nutated for 30 min and then poured into a column; the tag-free HelY proteins were recovered in the flowthrough fractions. EDTA was added to a final concentration of 10 mM, and the HelY protein preparations were purified further by gel filtration through a Superdex-200 column equilibrated in buffer D (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM DTT, 1 mM EDTA, 10% glycerol). Peak HelY fractions were pooled and stored at −80°C. Protein concentrations were determined by measuring the A_{280} with a NanoDrop 2000 spectrophotometer and applying an extinction coefficient of 73,840 M^{-1} cm^{-1}, as calculated by use of the ProtParam tool. The yields of HelY and the corresponding mutants were ~5 mg per liter of culture.

Nucleoside triphosphatase assay

Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 1 mM [α-^{32}P]ATP, HelY, and RNA or DNA, as specified below, were incubated for 20 min at 37°C. The reactions were quenched by adding 2 µl of 5 N 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. [^{32}P]ATP and [^{32}P]ADP were quantified by scanning the TLC plate with a Typhoon FLA7000 imager. ATPase activity (expressed as nanomoles of ATP hydrolysis) was calculated from the percent conversion of [^{32}P]ATP to [^{32}P]ADP [i.e., amount of ADP/(amount of ADP + amount of ATP)] in a 10-µl reaction mixture containing 10 nmol input ATP.

Helicase assay

The 5′ ^{32}P-labeled DNA or RNA strand was prepared by reaction of a synthetic oligonucleotide with T4 polynucleotide kinase and [γ-^{32}P]ATP. The labeled DNA or RNA was heated to 95°C to inactivate the kinase and then annealed to a 3-fold excess of a cRNA or DNA strand to form the various substrates indicated in the relevant figures. The annealed substrates were purified by electrophoresis through a native 12% polyacrylamide gel and then eluted from an excised gel slice by incubation for 16 h at 4°C in 400 µl of 10 mM Tris-HCl, pH 6.8, 1 mM EDTA, 50 mM NaCl. Helicase reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, ^{32}P-labeled nucleic acid (as specified below), and HelY (as specified below) were preincubated for 10 min at room temperature. The unwinding reactions were initiated by adding

1 mM ATP and a 10-fold excess of an unlabeled oligonucleotide identical to the labeled strand of the helicase substrate. (The unlabeled trap strand prevents the spontaneous reannealing of the unwound ^{32}P-labeled strand.) The reaction mixtures were incubated for 30 min at 37°C and then quenched by adding 1 µl of a solution containing 5% SDS, 100 mM EDTA. The mixtures were supplemented with 4 µl of 50% glycerol, 0.3% bromphenol blue. The reaction products were analyzed by electrophoresis through a 15-cm 10% polyacrylamide gel in 89 mM Tris-borate, 2.5 mM EDTA. The products were visualized by autoradiography.

ACKNOWLEDGEMENTS

Heather Ordonez performed the initial cloning of wild-type HelY.

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. 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.

3. 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.

4. 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.

5. 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.

6. 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.

7. 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.

8. Yakovleva, L. and Shuman, S. (2012) *Mycobacterium smegmatis* SftH exemplifies a distinctive clade of superfamily II DNA-dependent ATPases with 3' to 5' translocase and helicase activities. *Nucleic Acids Res*., **40,** 7465-7475.

9. Ordonez, H., Unciuleac, M.C. 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.

10. Ordonez, H. and Shuman, S. (2013) *Mycobacterium smegmatis* Lhr Is a DNA-dependent ATPase and a 3'-to-5' DNA translocase and helicase that prefers to unwind 3'-tailed RNA:DNA hybrids. *J. Biol. Chem*., **288**, 14125–-14134.

11. 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.

12. Balasingham, S.V., Zegeye, E.D., Homberset, H., Rossi, M.L., Laerdahl, J.K., Bohr, V.A. and Tønjum, T. (2012) Enzymatic activities and DNA substrate specificity of *Mycobacterium tuberculosis* DNA helicase XPB. *PLOS ONE*, **7,** 36860.

13. Zegeye, E.D., Balasaingham, S.V., Laerdahl, J.K., Homberset, H. and Tønjum, T. (2012) *Mycobacterium tuberculosis* RecG binds and unwinds model DNA substrates with a preference for Holliday junctions. *Microbiology*, **158**, 1982-1993.

14. Thakur, R.S., Basavaraju, S., Somyajit, K., Jain, A., Subramanya, S., Muniyappa, K. and Nagaraju, G. (2013) Evidence for the role of *Mycobacterium tuberculosis* RecG helicase in DNA repair and recombination. *FEBS J*., **280**, 1841-1860.

15. Zegeye, E.D., Balasingham, S.V., Laerdahl, J.K., Homberset, H., Kristiansen, P.E. and Tønjum, T. (2014) Effects of conserved residues and naturally occurring mutations on *Mycobacterium tuberculosis* RecG helicase activity. *Microbiology*, 160, 217-227.

16. Dewhare, S.S., Umesh, T.G. and Muniyappa, K. (2015) Molecular and functional characterization of RecD, a novel member of the SF1 family of helicases, from *Mycobacterium tuberculosis*. *J Biol Chem*., **290**, 11948-11968.

17. Khanjuja, J.S. and Muniyappa, K. (2012) Functional analysis of DNA replication fork reversal catalyzed by *Mycobacterium tuberculosis* RuvAB proteins. *J Biol Chem*., **287**, 1345-1360.

18. Zhang, H., Zhang, Z., Yang, J. and He, Z.G. (2014) Functional characterization of DnaB helicase and its modulation by single-stranded DNA binding protein in *Mycobacterium tuberculosis*. *FEBS J*., **281**, 1256-1266.

19. Kalarickal, N.C., Ranjan, A., Kalyani, S., Wal, M. and Sen, R. (2010) A bacteria transcription terminator with inefficient molecular motor action but with a robut transcription termination function. *J. Mol. Biol*., **395**, 966-982.

20. Mitra, A., Misquitta, R. and Nagaraja, V. (2014) *Mycobacterium tuberculosis* Rho is an NTPase with distinct kinetic properties and a novel RNA-binding subdomain. *PLOS ONE*, **9**, e107474.

21. D'Heygere, F., Schwartz, A., Coste, F., Castaing, B. and Boudvillain, M. (2015) Monitoring RNA unwinding by the transcription termination factor Rho from *Mycobacterium tuberculosis*. *Meth. Mol. Biol*., **1259**, 293-311.

22. Jackson, R.N., Klauer, A.A., Hintze, B.J., Rosinson, H., van Hoof, A. and Johnson, S.J. (2010) The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *EMBO J*., **29**, 2205-2216.

23. Weir, J.R., Bonneau, F., Hentschel, J. and Conti, E. (2010) Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proc. Natl. Acad. Sci. USA*., **107**, 12139-12144.

24. Halbach, F., Rode, M. and Conti, E. (2012) The crystal structure of *S. cerevisiae* Ski2, a DExH helicase associated with the cytoplasmic function of the exosome. *RNA*, **18**, 124-134.

25. Johnson, S.J. and Jackson, R.N. (2013) Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol*., **10**, 33-43.

26. 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.

27. 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.

28. Houghton, J., Townsend, C., Williams, A.R., Rodgers, A., Rand, L., Walker, K.B., Böttger, E.C., Springer, B. and Davis, E.O. (2012) Important role for *Mycobacterium tuberculosis* UvrD1 in pathogenesis and persistence apart from its function in nucleotide excision repair. *J. Bacteriol*., **194**, 2916-2923.

29. 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.

30. Bernstein, J., Patterson, D.N., Wilson, G.M. and Toth, E.A. (2008) Characterization of the essential activities of *Saccharomyces cerevisiae* Mtr4p, a 3'-to-5' helicase partner of the nuclear exosome. *J. Biol. Chem*., **283**, 4930-4942.

31. Bernstein, J., Ballin, J.D., Patterson, D.N., Wilson, G.M. and Toth, E.A. (2010) Unique properties of the Mtr4p–poly(A) complex suggest a role in substrate targeting. *Biochemistry*, **49**, 10357-10370.

32. Halback, F., Reichelt, P., Rode, M. and Conti, E. (2013) The yeast Ski complex: crystal structure and RNA channeling to the exosome complex. *Cell*, **154**, 814-825.

33. Hardwick, S.W., Gubbey, T., Hug, I., Jenal, U. and Luisi, B.F. (2012) Crystal structure of *Caulobacter crescentus* polynucleotide phosphorylase reveals a mechanism of RNA substrate channeling and RNA degradosome assembly. *Open Biol*., **2**, 120028.

34. Py, B., Higgins, C.F., Krisch, H.M. and Carpousis, A.J. (1996) A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature*, **381**, 169-172.

35. Liou, G.G., Chang, H.Y., Lin, C.S. and Lin-Chao, S. (2002) DEAD box RnlB RNA helicase physically associates with exoribonuclease PNPase to degrade double-stranded RNA independent of the degradosome-assembling region of RNase E. *J. Biol. Chem*., **277**, 41157-41162.

36. Unciuleac, M. and Shuman, S. (2013) Distinctive effects of domain deletions on the manganese-dependent DNA polymerase and DNA phosphorylase activities of *Mycobacterium smegmatis* polynucleotide phosphorylase. *Biochemistry*. **52**, 2967-2981.

37. Unciuleac, M. and Shuman, S. (2013) Discrimination of RNA versus DNA by polynucleotide phosphorylase. *Biochemistry*, **52**, 6702-6711.

38. Laminchhane, G., Zignol, M., Blades, N.J., Geiman, D.E., Dougherty, A., Grosset, J., Broman, K.W. and Bishai, W.R. (2003) A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA*, **100**, 7213-7218.

39. Griffin, J.E., Gawronski, J.D., DeJesus, M.A., Ioerger, T.R., Akerley, B.J. and Sassetti, C.M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathogens*, **7**, e1002251.

CHAPTER THREE.

The DNA repair repertoire of *Mycobacterium smegmatis* FenA includes the incision of DNA 5' flaps and the removal of 5' adenylylated products of aborted nick ligation*

INTRODUCTION

Homologous recombination (HR) in mycobacteria initiates via resection of a DNA double-strand break (DSB) by the AdnAB helicase-nuclease to generate a 3′ single-strand DNA substrate for RecA-mediated strand invasion. AdnAB consists of two subunits, AdnA and AdnB, each composed of an N-terminal ATPase domain and a C-terminal nuclease domain (1–3). DSB unwinding by AdnAB *in vitro* is stringently dependent on the ATPase activity of the AdnB motor translocating on the 3′ DNA strand. Unwinding drives the 3′ DNA strand into the AdnB nuclease domain and threads the 5′ DNA strand into the AdnA nuclease domain. Recent genetic analyses in *Mycobacterium smegmatis* revealed that a nuclease-dead AdnAB enzyme can sustain mycobacterial HR *in vivo*, as long as its AdnB motor is intact (4). Thus, AdnAB's processive DSB unwinding activity suffices for AdnAB function in HR. These findings raised two possibilities: (i) mycobacteria have a backup nuclease that resects the 5′ DNA strand unwound by AdnAB, or (ii) HR can

* Uson,M.L., Ghosh,S. and Shuman S. (2017) The DNA repair reptoire of *Mycobacterium smegmatis* FenA includes the incision of DNA 5' flaps and the removal of 5' adenylylated products of aborted nick ligation. *J. Bacteriol.*, **199**, e00304-17.

proceed via DSB unwinding and capture of the displaced strands by single-strand binding protein without the need for end resection.

In *Escherichia coli*, the 5′-to-3′ DNA exonuclease RecJ can substitute for the RecBCD nuclease in certain mutant backgrounds (5, 6). RecJ, which functions in a RecFOR pathway of DNA break repair, belongs to the DHH phosphodiesterase enzyme superfamily (7). *M. smegmatis* HR in the absence of the AdnAB nucleases requires RecO and RecR (4). Whereas the *M. smegmatis* proteome includes a DHH protein (MSMEG_2630), it has been characterized biochemically as an oligoribonuclease with 3′-to-5′ exonuclease activity (8), i.e., it is not RecJ-like. To our knowledge, there has been no report to date of a mycobacterial DNA single-strand 5′-to-3′ exonuclease enzyme.

Besides AdnAB, the only mycobacterial DNA 5′ exonuclease that has been characterized is the 5′ exonuclease of *M. tuberculosis* DNA polymerase I (MtuPol1) (9, 10). Purified recombinant MtuPol1 lacks a proofreading 3′ exonuclease activity (9) but possesses a 5′ exonuclease activity that resects the 5′-PO_4 strand of a nicked DNA substrate. Mutational analysis assigned the 5′ exonuclease activity to the N-terminal segment of the 904-amino-acid (aa) MtuPol1 protein and identified nine individual amino acids as essential for 5′ exonuclease function but not for DNA polymerase activity (9).

In pursuit of other candidate 5′ exonucleases in mycobacteria, our inspection of the *M. tuberculosis* proteome revealed the 318-amino-acid Rv2090 polypeptide (GenBank accession number WP_009935270.1) as a stand-alone homolog of the N-terminal 5′ exonuclease domain of

mycobacterial Pol1. The *M. smegmatis* proteome includes a 319-amino-acid homolog (GenBank accession number AFP40251.1) of the Pol1 5′ exonuclease domain. An alignment of the amino acid sequences of the *M. tuberculosis* and *M. smegmatis* putative 5′ exonucleases to the N-terminal domain of *M. smegmatis* Pol1 highlights 130 positions of side chain identity/similarity (Figure 3.1A). The nine amino acids identified by the Mizrahi laboratory (9) as essential for mycobacterial Pol1 5′ exonuclease activity (indicated by arrowheads in Figure 3.1A) are strictly conserved in the *M. tuberculosis* and *M. smegmatis* stand-alone homologs.

In order to evaluate this novel candidate mycobacterial nuclease, we have (i) produced and purified the recombinant *M. smegmatis* protein, (ii) determined that it is a DNA 5′-flap endonuclease, and (iii) educed its reaction requirements, substrate preferences, and parameters affecting cleavage site choice. Accordingly, we have named this mycobacterial enzyme FenA (flap endonuclease A). The homologous N-terminal 303-amino-acid segment of *M. smegmatis* Pol1 is also shown here to be a DNA 5′-flap endonuclease.

We report that the FenA catalytic repertoire includes the efficient excision of DNA-adenylate intermediate (AppDNA) products of abortive nick ligation, a function not previously associated with any specific bacterial DNA repair enzyme. We discuss the properties of mycobacterial FenA in light of insightful structural studies of eukaryal flap endonucleases (11, 12) and its potential biological roles in mycobacterial DNA repair.

RESULTS

Flap endonuclease activity of *M. smegmatis* FenA

We produced the 319-aa FenA protein in *E. coli* as a His_{10}Smt3 fusion and isolated it from a soluble extract by nickel-agarose chromatography. The His_{10}Smt3 tag was removed with the Smt3-specific protease Ulp1, and native FenA was separated from the tag by a second round of Ni-agarose chromatography. During subsequent gel filtration, FenA eluted as a single peak that, by reference to size standards, was monomeric (data not shown). SDS-PAGE of the purified protein revealed a predominant ~34-kDa polypeptide corresponding to FenA (Figure 3.1B). In parallel, we produced and purified a mutated version in which Asp125 (essential in Pol1 [9] and a predicted metal-binding residue) was changed to alanine (Figure 3.1B).

The isolated N-terminal 5′ exonuclease domains of *E. coli* DNA polymerase I and *Thermus aquaticus* DNA polymerase, to which FenA is homologous, have been shown to be structure-specific endonucleases that incise 5′ flap junctions (13–15). The analogous eukaryal structure-specific endonuclease, FEN-1, is a free-standing flap endonuclease devoid of DNA polymerase activity (16). To assay for flap endonuclease activity of FenA, a substrate was prepared by annealing a 5′ ^{32}P-labeled 24-mer DNA strand and an unlabeled 18-mer DNA_{OH} strand to a 36-mer template DNA strand to form a singly nicked duplex with a 6-nucleotide (nt) 5′ homooligodeoxyadenylate flap (Figure 3.1C). Reaction of 100 nM FenA with 100 nM nick flap substrate in the presence of 1 mM manganese resulted in quantitative conversion of the input ^{32}P-labeled 24-mer DNA to a shorter oligonucleotide product that was resolved

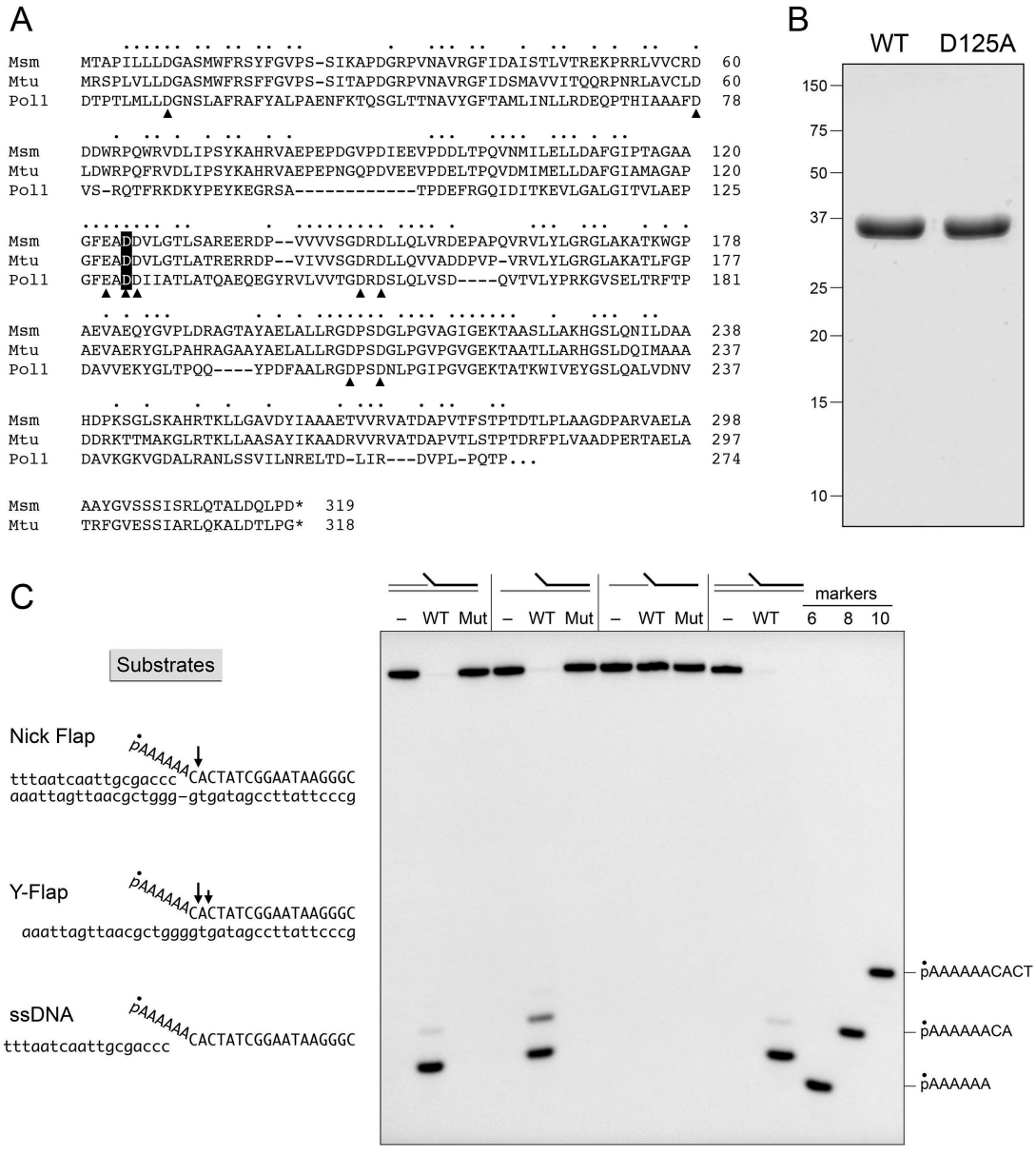
from the substrate by urea-PAGE (Figure 3.1C). In order to accurately assign the site of DNA cleavage, we analyzed the FenA reaction product alongside 5′ ^{32}P-labeled 6-, 8-, and 10-nucleotide markers of identical sequence to the 5′ end of the radiolabeled substrate strand. We thereby determined that the predominant ^{32}P-labeled FenA product derived from an incision of the flap strand at the seventh internucleotide phosphodiester (the site indicated by the arrowhead in Figure 3.1C), i.e., between the first and second nucleotides of the duplex portion of the flap strand (Figure 3.1C)

FenA was also adept at cleaving a Y-flap substrate (Figure 3.1C) composed of the same 5′ ^{32}P-labeled 24-mer DNA strand and 36-mer template DNA strand, which yielded the same predominant ^{32}P-labeled 7-mer product as well as a minor 8-mer product derived by incision between the second and third nucleotides of the duplex segment (the site indicated by the short arrow in Figure 3.1C). No incision was detected when the 36-mer template DNA strand was omitted, indicating that FenA does not act on single-stranded DNA. The D125A mutation abolished nuclease activity with the nick flap and Y-flap substrates (Figure 3.1C), signifying that the observed flap endonuclease activity inheres to the recombinant FenA protein.

Effect of varying 5′ flap length on FenA nuclease activity

To interrogate substrate specificity, we generated a series of nick flap substrates with 5′ ^{32}P-labeled 24-mer, 28-mer, or 34-mer flap strands that extruded nonhomooligomeric 5′ flaps of 6, 10, or 16 nucleotides in length (Figure 3.2A). FenA quantitatively cleaved the flap strands to generate a

Figure 3.1 Recombinant FenA is a flap endonuclease. (A) The amino acid sequence of the 319-aa *M. smegmatis* (Msm) FenA polypeptide is aligned to the sequences of the homologous M. tuberculosis (Mtu) protein Rv2090 and the N-terminal 5′ exonuclease domain of *M. smegmatis* DNA polymerase I (Pol1). Gaps in the alignment are denoted by dashes. Positions of side chain identity/similarity in all three polypeptides are denoted by dots. Amino acids essential for Pol1 5′ exonuclease activity (9) are indicated by arrowheads. The mutated aspartate residue (Asp125 in MsmFenA) is highlighted in white type on black background. (B) Purification. Aliquots (5 μg) of the Superdex-200 preparations of wild-type (WT) FenA and D125A mutant FenA were analyzed by SDS-PAGE. The Coomassie blue-stained gel is shown. The positions and sizes (in kilodaltons) of marker polypeptides are indicated on the left. (C) Nuclease activity. Reaction mixtures (10 μl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, and 1 pmol (100 nM) of ^{32}P-labeled nick flap, Y-flap duplex DNA, or single-stranded DNA (depicted to the left, with the 5′ ^{32}P label denoted by a filled circle) and either no enzyme (lanes −) or 1 pmol (100 nM) WT or mutant (Mut) FenA were incubated for 30 min at 37°C. The products were analyzed by urea-PAGE and visualized by autoradiography. The predominant sites of cleavage are indicated on the substrate by arrows, with a minor site indicated by the shorter arrow (left). The positions and sequences of 5′ ^{32}P-labeled marker oligonucleotides are indicated on the right.


A
Msm MTAPILLLDGASMWFRSYFGVPS-SIKAPDGRPVNAVRGFIDAISTLVTREKPRRLVVCRD 60
Mtu MRSPLVLLDGASMWFRSFFGVPS-SITAPDGRPVNAVRGFIDSMAVVITQQRPNRLAVCLD 60
Pol1 DTPTLMLLDGNSLAFRAFYALPAENFKTQSGLTTNAVYGFTAMLINLLRDEQPTHIAAAFD 78
Msm DDWRPQWRVDLIPSYKAHRVAEPEPDGVPDIEEVPDDLTPQVNMILELLDAFGIPTAGAA 120
Mtu LDWRPQFRVDLIPSYKAHRVAEPEPNGQPDVEEVPDELTPQVDMIMELLDAFGIAMAGAP 120
Pol1 VS-RQTFRKDKYPEYKEGRSA----------TPDEFRGQIDITKEVLGALGITVLAEP 125
Msm GFEADDVLGTLSAREERDP--VVVVSGDRDLLQLVRDEPAPQVRVLYLGRGLAKATKWGP 178
Mtu GFEADDVLGTLATRERRDP--VIVVSGDRDLLQVVADDPVP-VRVLYLGRGLAKATLFGP 177
Pol1 GFEADDIIATLATQAEQEGYRVLVVTGDRDSLQLVSD----QVTVLYPRKGVSELTRFTP 181
Msm AEVAEQYGVPLDRAGTAYAELALLRGDPSDGLPGVAGIGEKTAASLLAKHGSLQNILDAA 238
Mtu AEVAERYGLPAHRAGAAYAELALLRGDPSDGLPGVPGVGEKTAATLLARHGSLDQIMAAA 237
Pol1 DAVVEKYGLTPQQ----YPDFAALRGDPSDNLPGIPGVGEKTATKWIVEYGSLQALVDNV 237
Msm HDPKSGLSKAHRTKLLGAVDYIAAAETVVRVATDAPVTFSTPTDTLPLAAGDPARVAELA 298
Mtu DDRKTTMAKGLRTKLLAASAYIKAADRVVRVATDAPVTLSTPTDRFPLVAADPERTAELA 297
Pol1 DAVKGKVGDALRANLSSVILNRELTD-LIR---DVPL-PQTP... 274
Msm AAYGVSSSISRLQTALDQLPD* 319
Mtu TRFGVESSIARLQKALDTLPG* 318
B
WT
D125A
150
75
50
37
25
20
15
10
C
Substrates
Nick Flap
pAAAAAA
CACTATCGGAATAAGGGC
tttaatcaattgcgaccc
aaattagttaacgctggg-gtgatagccttattcccg
Y-Flap
pAAAAAA
CACTATCGGAATAAGGGC
aaattagttaacgctggggtgatagccttattcccg
ssDNA
pAAAAAA
CACTATCGGAATAAGGGC
tttaatcaattgcgaccc
– WT Mut – WT Mut – WT Mut – WT
markers
6 8 10
pAAAAAACACT
pAAAAAACA
pAAAAAA

discrete ^{32}P-labeled product of a length that increased in lockstep with the length of the extruded flap segment (Figure 3.2A). The ^{32}P-labeled product of FenA cleavage of the 10-mer flap comigrated with a 5′ ^{32}P-labeled 11-mer oligonucleotide, pTAACGCTGGGC, corresponding to the 5′-terminal 11-mer of the input ^{32}P-labeled 28-mer strand of the substrate (not shown), signifying that FenA incised the flap strand between the first and second nucleotides of the duplex segment, independent of the length of the 5′ flap. The 10-mer flap substrate was used routinely in the ensuing characterization of the FenA nuclease.

Divalent cation specificity

FenA-mediated cleavage of the 10-mer flap nick required an exogenous divalent cation cofactor (Figure 3.2B). When various metals were tested at 1 mM concentration in the absence of dithiothreitol (DTT) (Figure 3.2B, black bars), the divalent cation requirement was best satisfied by manganese and cobalt, which supported 92% and 80% incision of the flap strand, respectively. Magnesium (20% cleavage), zinc (25% cleavage), and nickel (21% cleavage) were less effective. Calcium and copper were ineffective (≤1% cleavage). However, when these metals were tested at 5 mM concentration (Figure 3.2B, gray bars), the preferred cofactor was manganese, which supported near-quantitative incision of the flap strand (97% cleavage). Magnesium (57% cleavage), cobalt (41% cleavage), and nickel (15% cleavage) were progressively less effective, whereas 5 mM calcium, copper, and zinc were inactive (<1% cleavage) (Figure 3.2B). Changing the metal cofactor or the

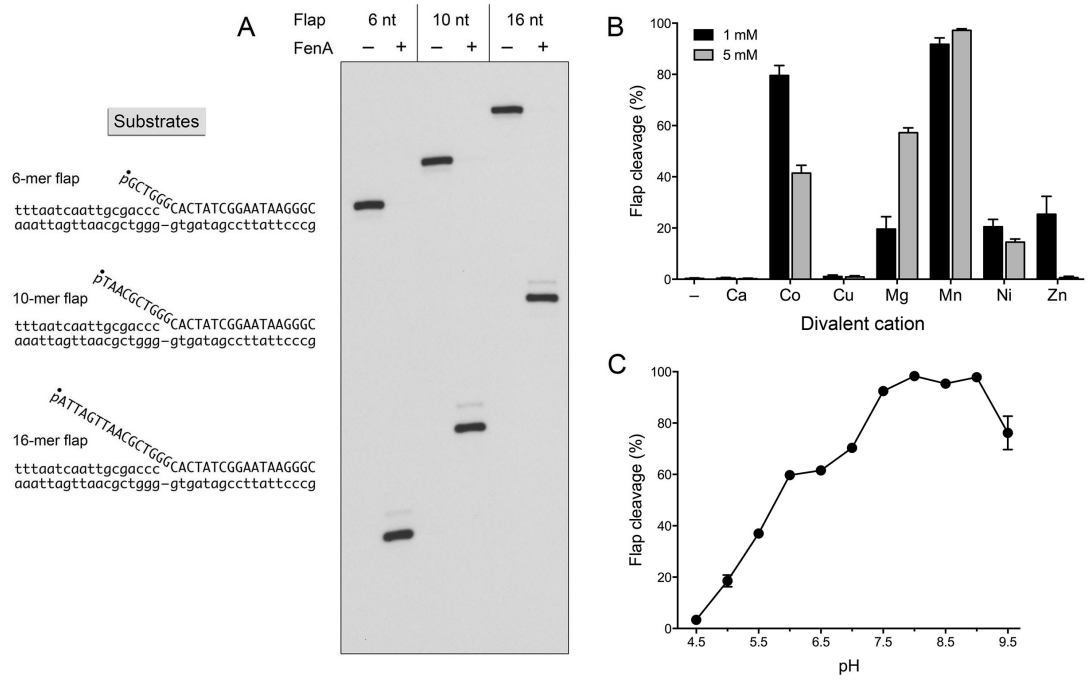


Figure 3.2 Effect of varying 5′-flap length, metal cofactor, and pH on FenA nuclease activity. (A) 5′-flap length. Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, 1 pmol (100 nM) of the indicated ^{32}P-labeled nick flap duplex substrate (depicted to the left, with the 5′ ^{32}P label denoted by a filled circle) and either no enzyme (lanes −) or 1 pmol (100 nM) FenA (lanes +) were incubated for 30 min at 37°C. The products were analyzed by urea-PAGE and visualized by autoradiography. (B) Divalent cation specificity. Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 pmol (100 nM) ^{32}P-labeled 10-mer flap nick duplex substrate, 1 pmol (100 nM) FenA, and either no added divalent cation (−) or 1 mM or 5 mM the indicated divalent cation (as the chloride salt) were incubated for 30 min at 37°C. The extents of substrate cleavage are plotted in bar graph format. Each datum is the average from three independent experiments ± standard error of the mean (SEM). (C) Reaction mixtures (10 µl) containing 50 mM NaCl, 1 mM DTT, 1 mM $MnCl_2$, 1 pmol (100 nM) of ^{32}P-labeled 10-mer flap nick duplex substrate, 1 pmol (100 nM) FenA, and 20 mM Tris buffer (Tris acetate at pH 4.5, 5.0, 5.5, 6.0, or 6.5; Tris-HCl at pH 7.0, 7.5, 8.0, 8.5, 9.0, or 9.5) were incubated for 30 min at 37°C. The extents of substrate cleavage are plotted as a function of pH. Each datum is the average from three independent experiments ± SEM.

concentration did not alter the site of cleavage of the flap strand between the first and second nucleotides of the duplex segment (not shown). Additional manganese and magnesium titration experiments in the presence of 1 mM DTT showed that FenA cleaved 91% of the input 10-mer flap nick duplex substrate during a 30-min reaction in the presence of 0.5 mM manganese and 96 to 98% in 1, 2, 5 and 10 mM manganese. In contrast, 0.5, 1, 2, 5, and 10 mM magnesium elicited only 2%, 13%, 33%, 45%, and 49% cleavage, respectively. All subsequent experiments were performed with 1 mM manganese as the metal cofactor.

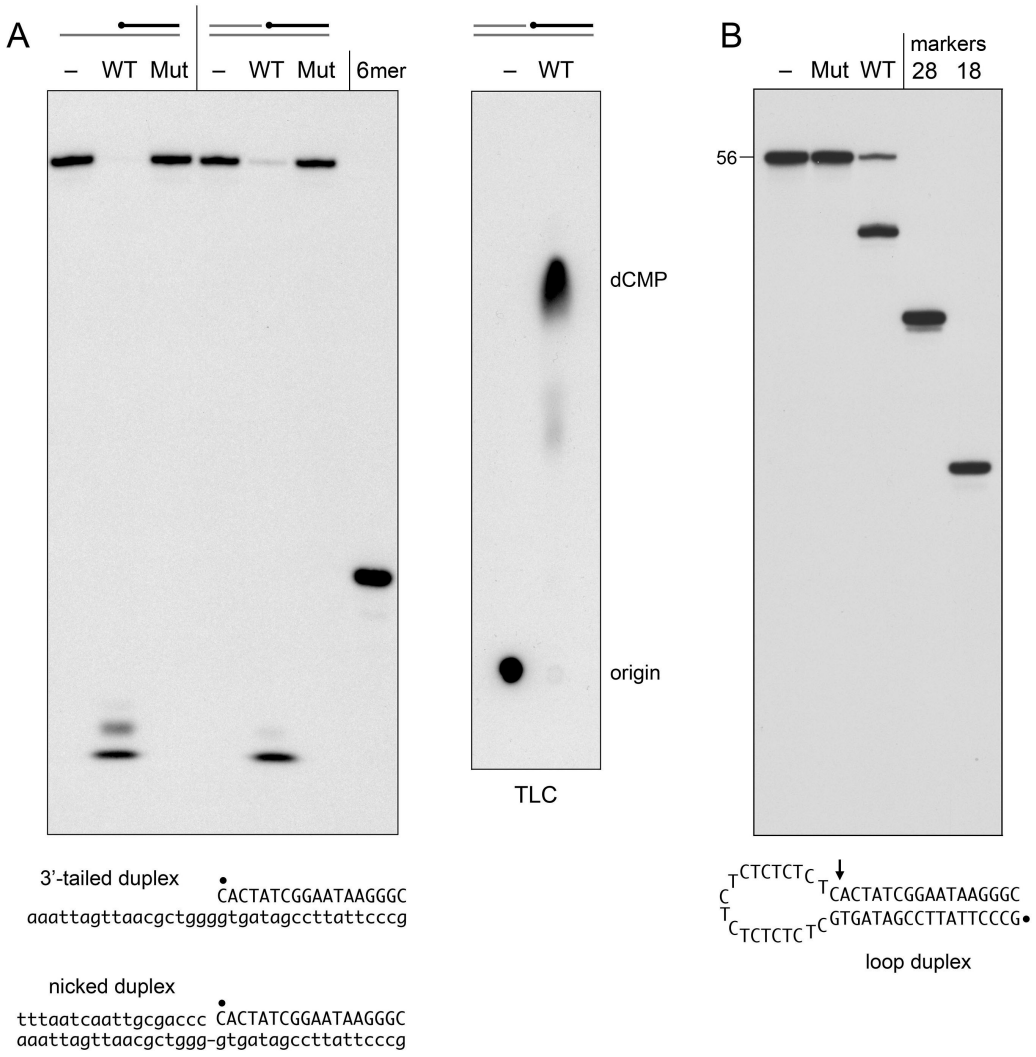
Effect of pH and potassium

Incision of the 10-mer flap nick by FenA was optimal from pH 7.5 to 9.0 in Tris-HCl buffer (93 to 98% cleavage) (Figure 3.2C). Activity was diminished at pH 9.5 (76% cleavage; higher pH values were not tested). FenA incision declined progressively as the pH was decreased below neutrality to 61% at pH 6.5, 37% at pH 5.5, 19% at pH 5.0, and 3% at pH 4.5 (Figure 3.2C). All ensuing assays were performed at pH 8.0. Because several previously characterized flap endonucleases were found to bind a potassium ion at the enzyme-DNA interface that promotes DNA binding (11, 12, 17, 18), we evaluated the impact of exogenous potassium on the activity of FenA. We found that 100 mM KCl had no effect on the rate of cleavage of the 10-mer flap nick when reactions were performed in the presence of 1 mM manganese or on the extent of cleavage when 5 mM magnesium was the metal cofactor (not shown).

FenA exonuclease activity at a DNA nick and a recessed 5′ end

To query whether FenA is able to act as a nuclease in the absence of a 5′ flap, we reacted the enzyme with a nicked 36-bp duplex ^{32}P labeled on the 5′-PO_4 side of the nick. (Note that the nicked DNA substrate was composed of strands identical in nucleotide sequence to the nicked flap substrates used in the preceding experiments.) Wild-type FenA cleaved the 5′-phosphate DNA (pDNA) strand of the nick to liberate a discrete product that migrated during urea-PAGE well ahead of a labeled 6-mer marker strand (Figure 3.3A, left). Analysis of the reaction mixtures by polyethyleneimine (PEI)-cellulose thin-layer chromatography (TLC) showed that FenA converted the ^{32}P-labeled 18-mer pDNA strand, which remained at the origin, to a predominant mononucleotide product that comigrated with dCMP (Figure 3.3A, right). Thus, the principal site of FenA incision of the nick was between the first two nucleotides of the nick 5′ terminus (just as was observed for the nick flap substrates). A minor product migrating more slowly than [^{32}P]dCMP was also detected by TLC, which we presume arises via cleavage between the second and third nucleotides of the pDNA strand. FenA was also capable of cleaving the recessed 5′ end of a 3′-tailed duplex composed of the same ^{32}P-labeled 18-mer pDNA and unlabeled 36-mer template DNA strands to yield a mixture of two products that we termed pC and pCpA (mimicking the cleavage sites seen with the Y-flap substrate). The FenA-D125A mutant was inert with the nicked duplex and 3′-tailed duplex substrates (Figure 3.3A), signifying that the flap-independent 5′ exonuclease activity was inherent to FenA.

Figure 3.3 FenA activity at a DNA nick, a recessed 5′ end, and a loop duplex. (A) Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, 1 pmol (100 nM) of ^{32}P-labeled 3′ tailed or nicked duplex (depicted below, with the 5′ ^{32}P label denoted by a filled circle) and either no enzyme (lanes −) or 1 pmol (100 nM) wild-type (WT) or mutant (Mut) FenA were incubated for 30 min at 37°C. The products were analyzed by urea-PAGE and visualized by autoradiography, along with a 5′ ^{32}P-labeled marker DNA oligonucleotide, pAAAAAA (lane 6mer). For TLC, the same reaction mixtures (10 µl) with either no enzyme (lanes −) or 1 pmol (100 nM) WT FenA were incubated for 30 min at 37°C and quenched with 2 µl of 10 mM EDTA. The products were analyzed by PEI-cellulose TLC. The mobile phase was 0.75 M LiCl. An autoradiograph of the chromatogram is shown. The origin and position of an unlabeled dCMP marker are indicated on the right. (B) Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, 1 pmol (100 nM) of ^{32}P-labeled loop duplex (depicted below, with the 5′ ^{32}P label denoted by a filled circle) and either no enzyme (lanes −) or 1 pmol (100 nM) WT or Mut FenA were incubated for 30 min at 37°C. The products were analyzed by urea-PAGE and visualized by autoradiography, along with 5′ ^{32}P-labeled 28-nt (pTAACGCTGGGCACTATCGGAATAAGGGC) and 18-nt (pCACTATCGGAATAAGGGC) marker DNA oligonucleotides (lanes at the right).


A
– WT Mut – WT Mut 6mer
– WT
dCMP
origin
TLC
B
markers
– Mut WT 28 18
56
3'-tailed duplex
CACTATCGGAATAAGGGC
aaattagttaacgctggggtgatagccttattcccg
nicked duplex
tttaatcaattgcgaccc CACTATCGGAATAAGGGC
aaattagttaacgctggg-gtgatagccttattcccg
CTCTCTC
TCACTATCGGAATAAGGGC
TCTCTC
GTGATAGCCTTATTCCCG
loop duplex

FenA cleaves a loop duplex DNA without a free flap 5′ end

Studies of other flap endonucleases have suggested a model for flap substrate engagement whereby the enzyme binds the double-stranded DNA (dsDNA) of a nick flap duplex and then threads the 5′ flap through a hole in the protein formed by a helical arch. Indeed, the structure of a complex of bacteriophage T5 FEN with a 5′ flap duplex visualizes this threading event, although the hole that encircles the flap in this structure can accommodate single-strand DNA but not duplex DNA (17). Threading through such a hole would require a free 5′ end on the single-stranded flap. To test if such a requirement applies to FenA, we reacted it with a self-complementary 5′ ^{32}P-labeled 56-mer DNA oligonucleotide that (after heating and quick cooling) forms a loop duplex in which the ends of an 18-bp stem are connected by a 20-nucleotide single-strand loop (Figure 3.3B). Wild-type FenA, but not the D125A mutant, incised the loop duplex to form a discrete ^{32}P-labeled product of a size consistent with incision at the 5′ single-strand/double-strand junction (Figure 3.3B). We conclude that single-strand threading of a free 5′ end through a narrow aperture is not an obligate feature of flap recognition by FenA. Recent models posit that the diameter of the aperture is flexible via disorder and can expand to allow threading of short 5′ duplexes or loops before adopting an ordered arch conformation that promotes cleavage (19).

Effect of varying the length of the 3′ tail of the template strand at a Y-flap

We gauged the effect of the length of the 3′ single-stranded tail on the efficiency and site of FenA incision. Two series of Y-flap substrates were

prepared, containing identical 18-bp duplex segments but differing with respect to the lengths of the 3′ tails of the template strand (18, 12, 6, or 0 nucleotides) and the nature of the 5′ flap, which was a 10-mer mixed sequence in one series (Figure 3.4, left) or a 6-mer homo-(dA) sequence in the other series (Figure 3.4, right). FenA efficiently incised both series of Y-flaps with 18-nt, 12-nt, or 6-nt 3′ tails at the same two sites within the duplex segment: a predominant cleavage between the first and second nucleotides and a minor cleavage between the second and third nucleotides of the duplex (indicated by long and short arrows in Figure 3.4). The salient finding was that elimination of the 3′ tail (0 nucleotides), which converts a Y-flap into a 5′-tailed duplex, lowered the extent of incision compared to the 6-nt tailed DNA while also shifting the sites of incision by one nucleotide to the right. This was evinced by (i) suppression of cleavage at the major CpA step, characteristic of the Y-flaps, (ii) enhanced incision at the now-predominant downstream ApC step, and (iii) appearance, in the 10-mer 5′ flap series, of a new minor cleavage site between the third and fourth nucleotides of the duplex segment (Figure 3.4). We surmise that the template 3′ single strand is an important determinant of cleavage site preference of FenA.

Multistep action of FenA at a 5′ flap nick

The extent of incision of the 5′ ^{32}P-labeled 28-mer strand of 100 nM 10-mer flap nick substrate to yield a 5′ ^{32}P-labeled 11-mer product was 62% at 12.5 nM FenA (average from three experiments ±4%, i.e., five DNA substrates cleaved per enzyme) and was virtually quantitative at ≥50 nM FenA (Figure

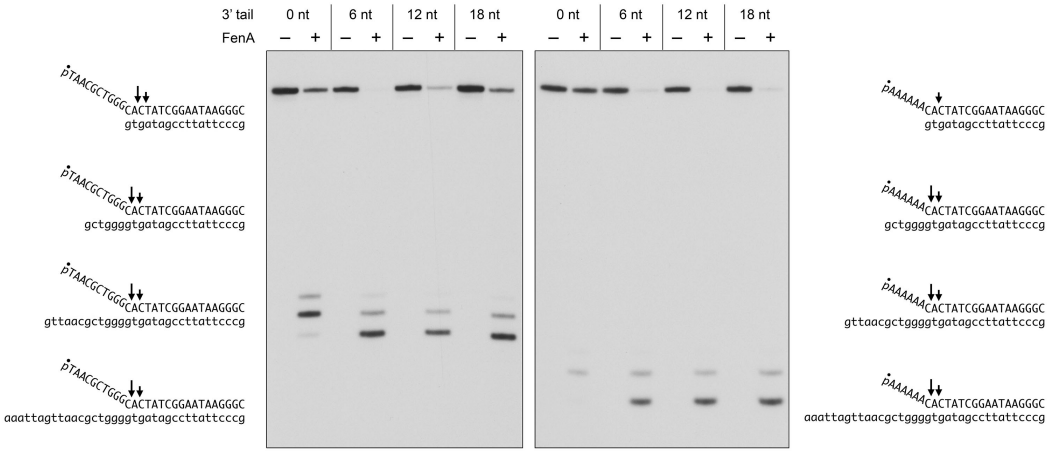


Figure 3.4 Effect of varying the length of the 3′ tail of the template strand at a Y-flap. Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, 1 pmol (100 nM) ^{32}P-labeled substrate (depicted at left and right, with the 5′ ^{32}P label denoted by a filled circle) and either no enzyme (lanes −) or 1 pmol (100 nM) FenA (lanes +) were incubated for 30 min at 37°C. The products were analyzed by urea-PAGE and visualized by autoradiography. The initial sites of cleavage are indicated on the substrates by long arrows; secondary sites are denoted by short arrows.

3.5A). The use of a 5′-labeled substrate restricts the analysis of the FenA reaction to the first cleavage event. To investigate whether the initial incision is followed by subsequent cleavages, we employed an otherwise identical 10-mer flap nick substrate composed of a 3′ [^{32}P]dAMP-labeled 29-mer flap strand annealed to an unlabeled 18-mer DNA_{OH} strand and a 36-mer template DNA strand (Figure 3.5). At limiting FenA concentrations, when not all of the input 3′ [^{32}P]dAMP labeled 29-mer was consumed, the predominant product was a 3′ ^{32}P-labeled 18-mer, corresponding to cleavage of the flap plus one nucleotide into the DNA duplex (the site indicated by the long arrow in Figure 3.5B, confirmed by comigration with a pACTATCGGAATAAGGCA marker oligonucleotide [not shown]). At 50 nM FenA, when all of the input 3′

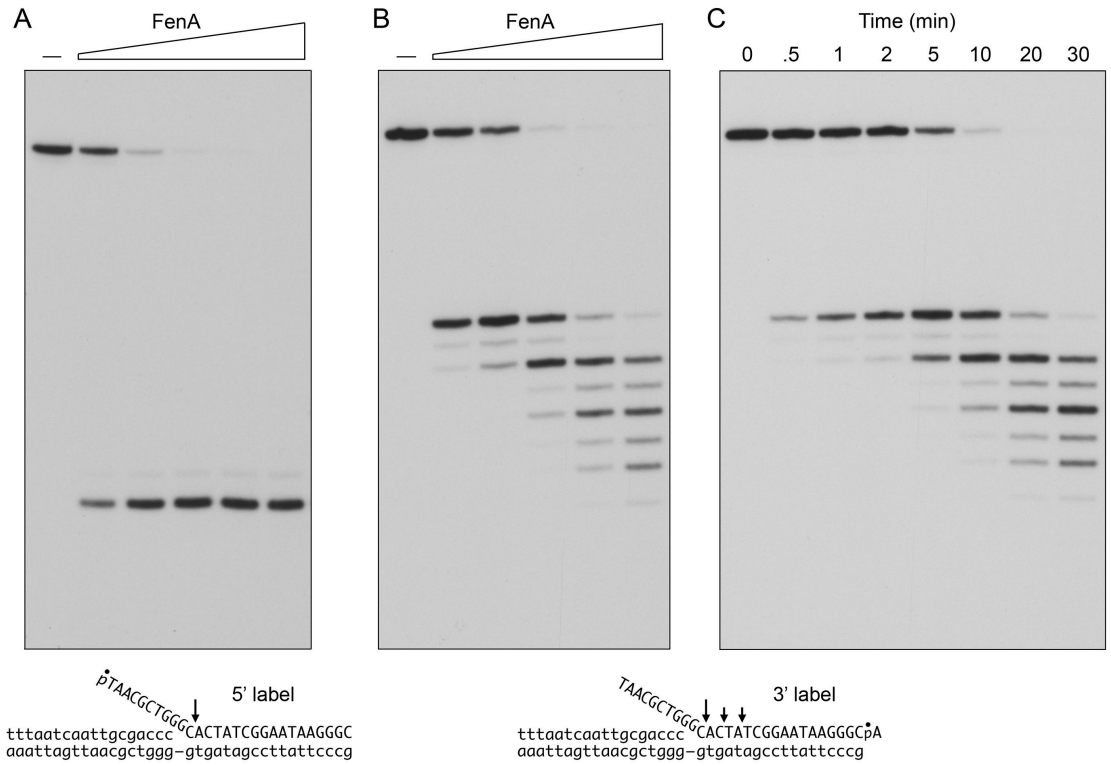


Figure 3.5 Multistep action of FenA at a 5′-flap nick. (A and B) Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, either 1 pmol (100 nM) 5′ ^{32}P-labeled 10-mer flap nick duplex substrate (A) or 1 pmol (100 nM) 3′ ^{32}P-labeled 10-mer flap nick duplex substrate (B), and 0, 0.125, 0.25, 0.5, 1, or 2.5 pmol (0. 12.5, 25, 50, 100, or 250 nM) FenA, from left to right in each titration series, were incubated for 30 min at 37°C. The products were analyzed by urea-PAGE and visualized by autoradiography. The sizes of cleavage products are indicated on the right. The predominant cleavage site is indicated on the substrate by an arrow, with minor sites denoted by shorter arrows. (C) A reaction mixture containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, 1 pmol (100 nM) 3′ ^{32}P-labeled 10-mer flap duplex substrate, and 2.5 pmol (250 nM) FenA was incubated at 37°C. Aliquots (10 µl) were withdrawn at the times specified and quenched with 90% formamide, 50 mM EDTA.

[^{32}P]dAMP-labeled 29-mer was consumed, we detected the formation of a novel 3′ ^{32}P-labeled secondary 16-mer cleavage product that was two nucleotides shorter than the primary cleavage fragment. Increasing FenA to 100 nM eliminated the primary cleavage fragment and elicited the appearance

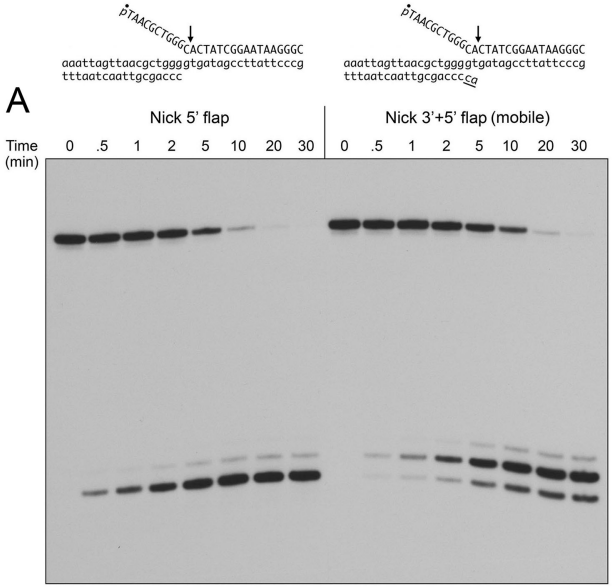
of a tertiary cleavage product, a 14-mer that was two nucleotides shorter than the secondary product (Figure 3.5B). The concentration dependence of the product distribution suggested a precursor-product relationship, whereby FenA further processes the cleaved flap (equivalent to a 1-nucleotide gapped duplex) by iterative cleavages in predominantly dinucleotide increments.

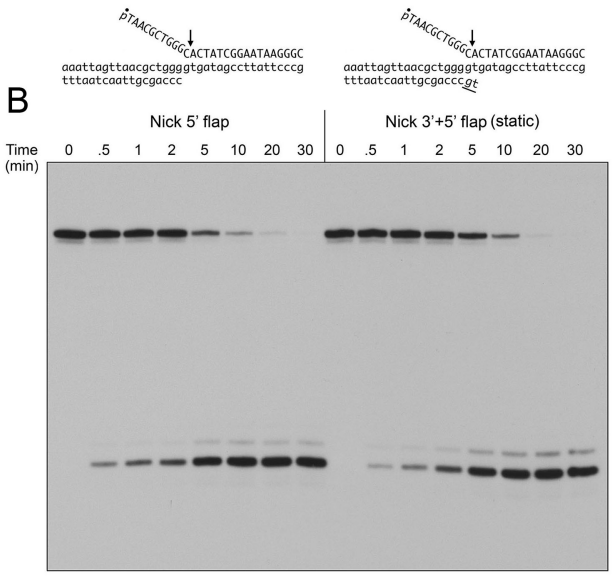
The kinetic profile of the reaction of 250 nM FenA with 100 nM 3′ ^{32}P-labeled nick flap substrate affirmed the precursor-product relationships and the dinucleotide steps of secondary and tertiary cleavage (Figure 3.5C). A model to explain this periodicity of FenA cleavage is considered in Discussion.

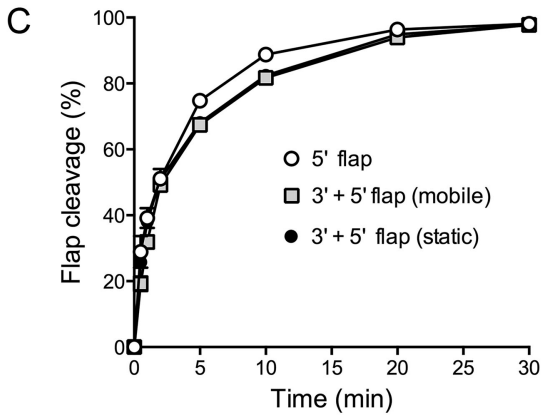
FenA cleavage of mobile and static double-flap substrates

Mycobacterial DNA repair polymerases such as PolD1 and PolD2 are capable of limited strand displacement synthesis during gap filling (20); this reaction can allow formation of a double-flap structure with a mobile branch point. An example of a mobile double flap is shown in Figure 3.6A, in which the 3′-terminal CA dinucleotide of the unlabeled DNA_{OH} strand is identical to the CA dinucleotide of the labeled flap strand that comprises the 5′ margin of the duplex segment of the flap strand-template strand hybrid. The base-pairing state shown in Figure 3.6A is one of three potential conformations of the flap branch point in the mobile double-flap substrate. A comparison of the time course of reaction of FenA with 5′ ^{32}P-labeled nick 5′ flap and mobile double-flap substrates revealed that whereas the rates of substrate incision were similar (Figure 3.6A and C), the presence of the extra 3′-terminal CA dinucleotide of the DNA_{OH} strand elicited a shift in the site of incision of the 5′

Figure 3.6 FenA cleavage of mobile and static double-flap substrates. Reaction mixtures containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, 1 pmol (100 nM) ^{32}P-labeled 5′ flap nick (A and B), mobile double flap nick (A) or static flap nick (B) substrates (depicted above, with the 5′ ^{32}P label denoted by a filled circle), and 2.5 pmol (250 nM) FenA were incubated at 37°C. Aliquots (10 µl) were withdrawn at the times specified and quenched with formamide and EDTA. The products were analyzed by urea-PAGE and visualized by autoradiography. The predominant cleavage site is indicated on each substrate with an arrow. (C) The extents of cleavage of the substrates are plotted as a function of reaction time. Each datum is the average from three independent time course experiments ± SEM.


A
Nick 5' flap
Nick 3'+5' flap (mobile)
Time (min)
0 .5 1 2 5 10 20 30
0 .5 1 2 5 10 20 30
pTAACGCTGGGCACTATCGGAATAAGGGC
aaattagttaacgctggggtgatagccttattcccg
tttaatcaattgcgaccc
pTAACGCTGGGCACTATCGGAATAAGGGC
aaattagttaacgctggggtgatagccttattcccg
tttaatcaattgcgacccca


B
Nick 5' flap
Nick 3'+5' flap (static)
Time (min)
0 .5 1 2 5 10 20 30
0 .5 1 2 5 10 20 30
pTAACGCTGGGCACTATCGGAATAAGGGC
aaattagttaacgctggggtgatagccttattcccg
tttaatcaattgcgaccc
pTAACGCTGGGCACTATCGGAATAAGGGC
aaattagttaacgctggggtgatagccttattcccg
tttaatcaattgcgacccgt


C
Flap cleavage (%)
100
80
60
40
20
0
0 5 10 15 20 25 30
Time (min)
5' flap
3' + 5' flap (mobile)
3' + 5' flap (static)

flap strand by one nucleotide in the 3′ direction (Figure 3.6A). Given that FenA strongly prefers to incise a nick 5′ flap strand between the first and second nucleotides of the duplex segment, we conclude that FenA preferentially engages the mobile double-flap substrate in a 3′-mononucleotide flap conformation in which the 3′-dA of the DNA_{OH} strand is unpaired and the penultimate dC nucleotide is paired to the template. The product of this incision is a nicked DNA with correctly base-paired 3′-OH and 5′-PO_4 terminal nucleotides that is set up for sealing by a DNA ligase.

To gauge the role of double-flap mobility in cleavage site choice, we tested a static double-flap substrate in which the 3′-terminal GT dinucleotide of the DNA_{OH} strand is unable to pair with the template strand, ensuring that the DNA will adopt the conformation depicted in Figure 3.6B, in which the substrate has a 10-nucleotide 5′ flap and a 2-nucleotide 3′ flap. The rate of static double-flap substrate incision was similar to that of the nick 5′ flap DNA (Figure 3.6B and C) and the sites of incision of the 5′ flap strand were identical, i.e., between the first and second nucleotides of the duplex segment (Figure 3.6B).

FenA can remove nicked AppDNA products of aborted ligation

Classic polynucleotide ligases join 3′-OH and 5′-PO_4 termini at nicks via a series of three nucleotidyl transfer steps. In step 1, ligase reacts with either ATP or NAD^+ to form a covalent ligase-(lysyl-Nζ)-AMP intermediate and release pyrophosphate (PP_i) or nicotinamide mononucleotide (NMN). In step 2, AMP is transferred from ligase-adenylate to the 5′-PO_4 DNA end to form a

DNA-adenylate intermediate (AppDNA). In step 3, ligase catalyzes attack by the nick 3′-OH on the DNA-adenylate to seal the two ends via a phosphodiester bond and release AMP. Various nucleic acid lesions (e.g., 1-nucleotide gaps, 3′-OH base mispairs, abasic sites, 5′-PO_4 ribonucleotides) can interfere with step 3 of the ligase pathway and lead to the accumulation of AppDNA as products of aborted ligation (21–24). AppDNA is itself a deleterious lesion that necessitates repair. In eukarya, AppDNA ends are deadenylylated by the repair enzyme aprataxin, a member of the histidine triad (HIT) superfamily of nucleotidyltransferases (24–26). Whereas HIT proteins are found in bacterial proteomes, they are not specifically homologous to eukaryal aprataxin and have not been demonstrated to resolve abortive ligation products or function in nucleic acid repair.

Bacterial DNA ligases that participate in nonhomologous end joining (NHEJ) share two distinctive properties: (i) they preferentially seal DNA nicks with a 3′-OH monoribonucleotide terminus (27) and (ii) they generate high levels of abortive AppDNA products during attempted sealing at a correctly base-paired all-DNA nick (27–29). Because mycobacteria have two NHEJ ligases of this type (LigD and LigC), it is of interest to understand how mycobacteria might contend with AppDNA nicks. We considered FenA to be a plausible candidate to remove the blocked 5′ end of an AppDNA nick.

To prepare nicked AppDNA, we adenylylated the 5′ ^{32}P-labeled 18-mer pDNA strand by reaction with *E. coli* RtcA and cold ATP (30). The purified radiolabeled AppDNA strand migrated slower during electrophoresis than the

18-mer pDNA (Figure 3.7A). Whereas the labeled 5′ phosphate of the 18-mer pDNA was hydrolyzed by alkaline phosphatase (calf intestinal phosphatase [CIP]), the label in the AppDNA strand was refractory to CIP, affirming that the 5′ end was blocked (Figure 3.7A). The AppDNA strand was annealed with an unlabeled 18-mer DNA_{OH} strand to a 36-mer DNA template strand to form a nicked DNA adenylate. FenA incised the nicked AppDNA substrate to yield a product that migrated more slowly during urea-PAGE than the mononucleotide product of FenA digestion of the pDNA strand of nicked duplex DNA (Figure 3.7A).

The products of FenA digestion of the nicked DNA and nicked AppDNA substrates were analyzed by PEI-cellulose TLC in parallel with nuclease P1-digested AppDNA (Figure 3.7B). Digestion of the AppDNA strand with nuclease P1 (which cleaves the 3′-O-P bond) liberated a single ^{32}P-labeled species from AppDNA, predicted to be a dinucleotide App(dC), which migrated faster than cold dCMP marker during TLC (Figure 3.7B). Whereas FenA action on the nicked DNA yielded primarily [^{32}P]dCMP, the predominant product of FenA incision of nicked AppDNA comigrated with the nuclease P1 product (Figure 3.7B). A minor FenA reaction product was detected that migrated more slowly than dCMP, and we presume it arose from incision at the next downstream phosphodiester to release App(dC)p(dA) (Figure 3.7B). Thus, FenA displays the same preference for AppDNA cleavage between the first and second nucleotides of the nick duplex segment as it does for a 5′ nick and a 5′-flap nick.

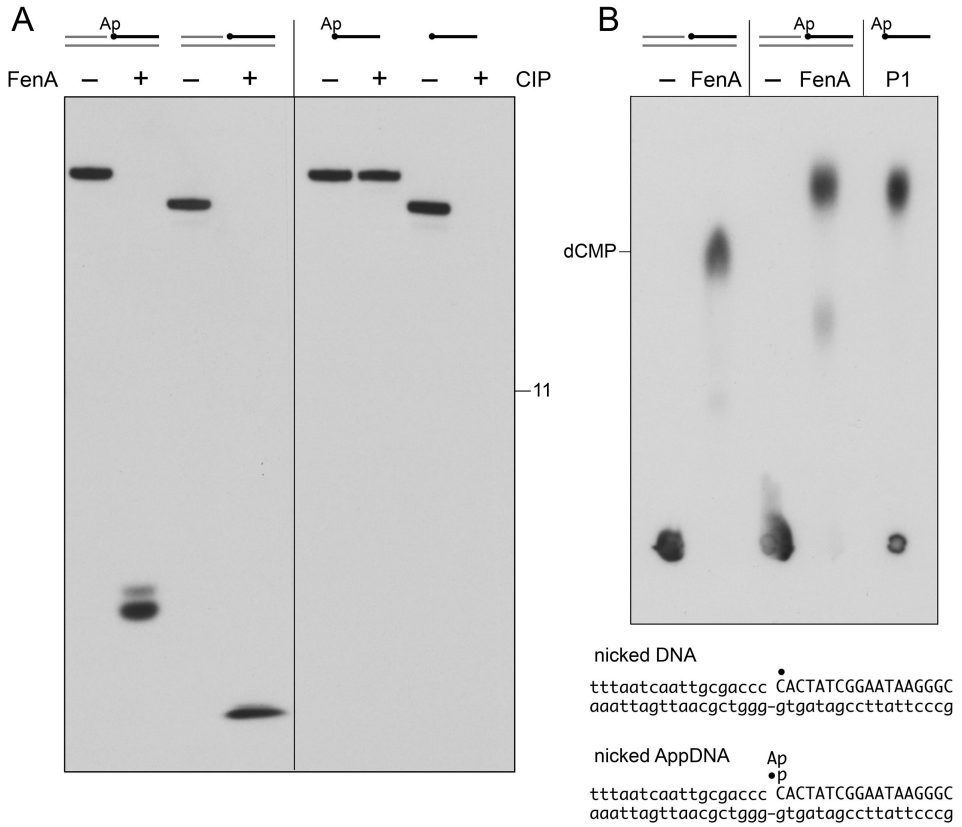


Figure 3.7 FenA can incise nicked AppDNA products of aborted ligation. The ^{32}P-labeled nicked DNA and nicked AppDNA substrates are depicted at bottom right with the 5′ ^{32}P label denoted by a filled circle. (A) Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, 100 nM ^{32}P-labeled nicked DNA or nicked AppDNA, and either no enzyme (lanes −) or 1 pmol (100 nM) FenA (lanes +) were incubated for 30 min at 37°C. For CIP treatment, reaction mixtures (10 µl) containing 20 mM Tris-acetate (Tris-OAc), pH 7.9, 50 mM KOAc, 10 mM MgOAc, 100 µg/ml bovine serum albumin (BSA), 100 nM ^{32}P-labeled pDNA or AppDNA single strands, and either no enzyme (lanes −) or 10 U CIP (lanes +) (New England BioLabs) were incubated for 30 min at 37°C. The products were analyzed by urea-PAGE and visualized by autoradiography. (B) Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, 100 nM ^{32}P-labeled nicked DNA or nicked AppDNA, and either no enzyme (lanes −) or 1 pmol (100 nM) FenA were incubated for 30 min at 37°C. For nuclease P1 digestion, a mixture (10 µl) containing 50 mM NaOAc, pH 5.2, 1 mM $ZnCl_2$, 100 nM ^{32}P-labeled AppDNA single strand, and 1 U nuclease P1 (U.S. Biological) was incubated for 20 min at 50°C. Reactions were quenched with 2 µl of 10 mM EDTA. The products were analyzed by PEI-cellulose TLC. The mobile phase was 0.5 M LiCl. An autoradiograph of the chromatogram is shown. The position of an unlabeled dCMP marker is indicated on the left.

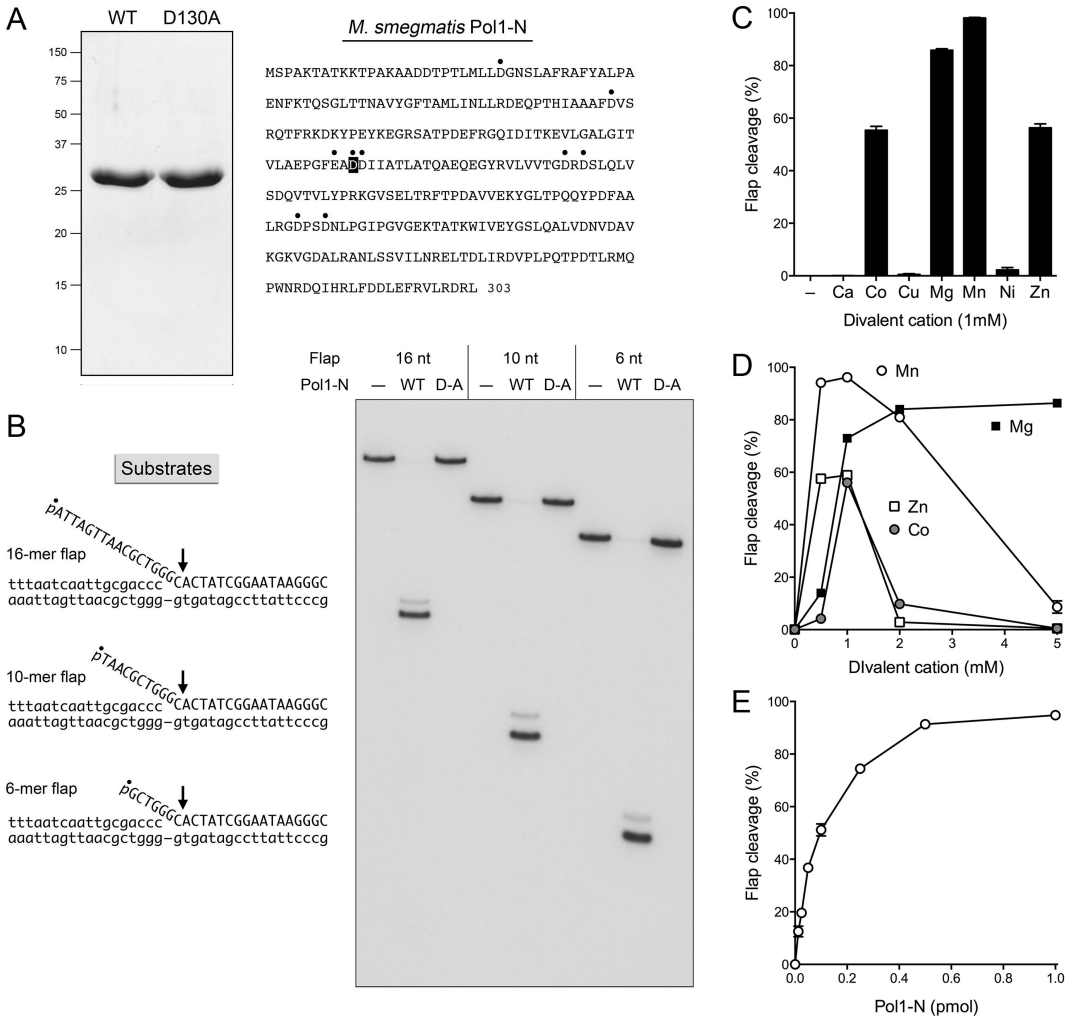
The N-terminal domain of *M. smegmatis* Pol1 is a flap endonuclease

It is well established that mycobacterial Pol1 has 5′ exonuclease activity that can resect the 5′ ends of gapped and nicked duplex DNAs and that the 5′ exonuclease of Pol1 relies on catalytic amino acids within the N-terminal portion of the polypeptide (9). In light of the above characterization of mycobacterial FenA, we asked whether the homologous N-terminal segment of *M. smegmatis* Pol1 is a flap endonuclease and whether the N-terminal Pol1 is autonomous with respect to nuclease activity. We produced an N-terminal 303-aa fragment of Pol1 (referred to here as Pol1-N; Figure 3.8A depicts its amino acid sequence) in *E. coli* as a His_{10}Smt3 fusion, isolated it from a soluble extract by nickel-agarose chromatography, cleaved the tag with Ulp1, and recovered tag-free Pol1-N in the flowthrough after a second nickel affinity step to obtain pure Pol1-N (Figure 3.8A). In parallel, we produced and purified a mutated version of Pol1-N, predicted to be catalytically dead based on studies of *M. tuberculosis* Pol1 (9), in which the Asp130 was changed to alanine (Figure 3.8A). Wild-type Pol1-N quantitatively cleaved the 5′ ^{32}P-labeled 34-mer, 28-mer, or 24-mer flap strands of the nick flap substrates with 5′ flaps of 16, 10, or 6 nucleotides, respectively (Figure 3.8B). The Pol1-N reaction products comigrated exactly with the products of incision of the same DNA flap substrates by FenA (not shown), signifying that Pol1-N also incises a 5′ flap nick between the first and second nucleotides of the duplex segment. The D130A mutation abolished Pol1-N flap endonuclease activity (Figure 3.8B). Subsequent assays of Pol1-N were performed using the 10-mer flap

nick substrate.

Pol1-N flap cleavage required an exogenous divalent cation cofactor (Figure 3.8C). Tests of various metals at 1 mM concentration indicated that the requirement was best satisfied by manganese (98% incision) and magnesium (86%), followed by zinc (56%) and cobalt (5%) (Figure 3.8C). Calcium and copper were ineffective (≤1% cleavage), as was nickel (2%). The four metals that supported Pol1-N activity were then titrated across a concentration range from 0.5 to 5 mM (Figure 3.8D). Manganese-dependent activity was best at 0.5 and 1 mM (94 to 96% cleavage), declined modestly at 2 mM (81%), and declined precipitously at 5 mM (8%). Conversely, magnesium was weakly active at 0.5 mM (14% cleavage) but improved at 1 mM (73%) and plateaued at 2 to 5 mM (84 to 86%). Zinc supported activity at 0.5 and 1 mM (58% cleavage) but was inactive at 2 and 5 mM. Cobalt was narrowly active at 1 mM (56% cleavage) but ineffective at 0.5, 2, and 5 mM concentrations (Figure 3.8D). Pol1-N was active in flap incision from pH 5.5 to pH 8.5 in Tris-acetate or Tris-HCl buffers; activity declined at higher and lower pH values (not shown). Pol1-N activity flap endonuclease activity increased with input protein (Figure 3.8E); at the lowest enzyme concentration tested (1.25 nM), Pol1-N cleaved an ~10-fold molar excess of the 10-mer flap substrate. In agreement with prior studies of full-length Pol1 (9), we found that Pol1-N resected the 5′ termini of nicked duplex and 3′-tailed duplex DNAs (the substrates shown in Figure 3.3) and that this exonuclease activity was effaced by the D130A mutation (not shown). Finally, we found that Pol1-N could resect the App(dN)

Figure 3.8 N-terminal domain of mycobacterial Pol1 is a flap endonuclease. (A) The amino acid sequence of the 303-aa M. smegmatis Pol1-N polypeptide is shown at the right, with amino acids essential for M. tuberculosis Pol1 5′ exonuclease activity indicated by dots. The mutated aspartate (Asp130) is highlighted in white type on a black background. Aliquots (10 μg) of the recombinant wild-type (WT) Pol1-N and D130A mutant were analyzed by SDS-PAGE. The Coomassie blue-stained gel is shown at the left. The positions and sizes (in kilodaltons) of marker polypeptides are indicated. (B) Flap endonuclease activity. Reaction mixtures (10 μl) containing 20 mM Tris-HCl, pH 8.0, 1 mM $MnCl_2$, 1 mM DTT, 1 pmol (100 nM) of the indicated ^{32}P-labeled nick flap duplex substrate (depicted to the left, with the 5′ ^{32}P label denoted by a filled circle), and either no enzyme (lanes −), 1 pmol (100 nM) Pol1-N (lanes WT), or 1 pmol (100 nM) D130A (lanes D-A) were incubated for 30 min at 37°C. The products were analyzed by urea-PAGE and visualized by autoradiography. (C) Divalent cation requirement. Reaction mixtures (10 μl) containing 20 mM Tris-HCl, pH 8.0, 1 pmol (100 nM) ^{32}P-labeled 10-mer flap nick duplex substrate, 0.5 pmol (50 nM) Pol1-N, and either no added divalent cation (−) or 1 mM the indicated divalent cation (as the chloride salt) were incubated for 30 min at 37°C. The extents of substrate cleavage are plotted in bar graph format. Each datum is the average from three independent experiments ± SEM. (D) Divalent cation concentration dependence. Reaction mixtures (10 μl) containing 20 mM Tris-HCl, pH 8.0, 1 pmol (100 nM) ^{32}P-labeled 10-mer flap nick duplex substrate, 0.5 pmol (50 nM) Pol1-N, and 0, 0.5, 1, 2, or 5 mM $MnCl_2$, $MgCl_2$, $CoCl_2$, or $ZnCl_2$ were incubated for 30 min at 37°C. Flap cleavage is plotted as a function of divalent ion concentration; each datum is the average from three independent experiments ± SEM. (E) Enzyme titration. Reaction mixtures (10 μl) containing 20 mM Tris-HCl, pH 8.0, 1 mM $MnCl_2$, 1 mM DTT, 1 pmol (100 nM) 5′ ^{32}P-labeled 10-mer flap nick duplex substrate, and 0, 0.0125, 0.025, 0.05, 0.1, 0.25, 0.5, or 1 pmol (0, 1.25, 2.5, 5,10, 12.5, 25, 50, or 100 nM) Pol1-N were incubated for 30 min at 37°C. Flap cleavage is plotted as a function of input enzyme. Each datum is the average from three independent experiments ± SEM.


A
WT
D130A
150
75
50
37
25
20
15
10
M. smegmatis Pol1-N
MSPAKTATKKTPAKAADDTPTLMLLDGNSLAFRAFYALPA
ENFKTQSGLTTNAVYGFTAMLINLLRDEQPTHIAAAFDVS
RQTFRKDKYPEYKEGRSATPDEFRGQIDITKEVLGALGIT
VLAEPGFEADDIIATLATQAEQEGYRVLVVTGDRDSLQLV
SDQVTVLYPRKGVSELTRFTPDAVVEKYGLTPQQYPDFAA
LRGDPSDNLPGIPGVGEKTATKWIVEYGSLQALVDNVDAV
KGKVGDALRANLSSVILNRELTDLIRDVPLPQTPDTLRMQ
PWNRDQIHRLFDDLEFRVLRDRL 303
Flap
16 nt
10 nt
6 nt
Pol1-N
—
WT
D-A
B
Substrates
pATTAGTTAACGCTGGGC
16-mer flap
tttaatcaattgcgaccc CACTATCGGAATAAGGGC
aaattagttaacgctggg-gtgatagccttattcccg
pTAACGCTGGGC
10-mer flap
tttaatcaattgcgaccc CACTATCGGAATAAGGGC
aaattagttaacgctggg-gtgatagccttattcccg
pGCTGGGC
6-mer flap
tttaatcaattgcgaccc CACTATCGGAATAAGGGC
aaattagttaacgctggg-gtgatagccttattcccg
C
Flap cleavage (%)
Ca
Co
Cu
Mg
Mn
Ni
Zn
Divalent cation (1mM)
D
Mn
Mg
Zn
Co
Flap cleavage (%)
DIvalent cation (mM)
E
Flap cleavage (%)
Pol1-N (pmol)

5′ end of nicked AppDNA (not shown).

DISCUSSION

Although our original motivation was to study FenA as a candidate backup 5′ exonuclease to AdnAB, the biochemical specificity of FenA illuminated here, particularly its specificity for flap junctions and its inability to digest single-strand DNA, militates against a RecJ-like end-resection function in mycobacterial HR. Rather, the flap endonuclease and AppDNA removal activities of FenA point toward a role in gap repair during lagging-strand DNA replication or NHEJ. The flap removal activity of FenA might also come into play in the single-strand annealing (SSA) pathway of DNA double-strand break (DSB) repair that entails large DNA deletions between homologous regions flanking the DSB (31).

Transposon mutagenesis showed that *M. tuberculosis* FenA (Rv2090) is inessential for bacterial growth in culture (32). It is conceivable that FenA overlaps functionally with the 5′ exonuclease of mycobacterial Pol1. Genetic analysis of *M. smegmatis* Pol1 by the Mizrahi laboratory has shown that a *polA* strain with targeted gene disruption within the polymerase domain is viable, albeit sensitized to UV irradiation (10). The *M. smegmatis* polymerase insertion strain would, in principle, still produce the N-terminal exonuclease domain of Pol1, which we show here is an autonomous catalytically active flap endonuclease. To our knowledge, it has not been reported whether the 5′ exonuclease activity of Pol1 is essential or dispensable for mycobacterial viability. It will be of interest in future studies to evaluate the impact of

replacing the wild-type *M. smegmatis polA* gene with an allele (e.g., *D130A*) encoding a full-length Pol1 protein that is specifically defective for 5′ exonuclease and flap endonuclease functions and to do so in strain backgrounds in which FenA is present or deleted. Genetic studies in *Bacillus subtilis* and *E. coli* have shown that the 5′ exonuclease activity of either Pol1 or the stand-alone exonuclease YpcP/Xni is essential in those organisms (33).

Mycobacterial FenA shares in several respects the biochemical properties of other flap endonucleases from diverse sources (reviewed in reference 34), particularly its preference for incision 1 nucleotide into the double-stranded segment of a nicked 5′ flap or nicked duplex substrate. Recent highly informative crystal structures of DNA-bound human FEN1 and human Exo1 suggest a unified mechanism for substrate recognition and catalysis (11, 12, 35). In brief, the enzyme engages a 12-bp duplex DNA segment flanking the flap and including the scissile phosphodiester, enforces a sharp bend in the template DNA strand at the flap/nick junction (by insertion of a protein α-helix between the splayed-apart template strand nucleobases), and positions the scissile phosphodiester for metal-catalyzed hydrolysis. Thus, in accord with this mechanism, the “first cut” made by FenA on a 5′ flap nick substrate generates a 1-nucleotide gapped duplex product (Figure 3.9A).

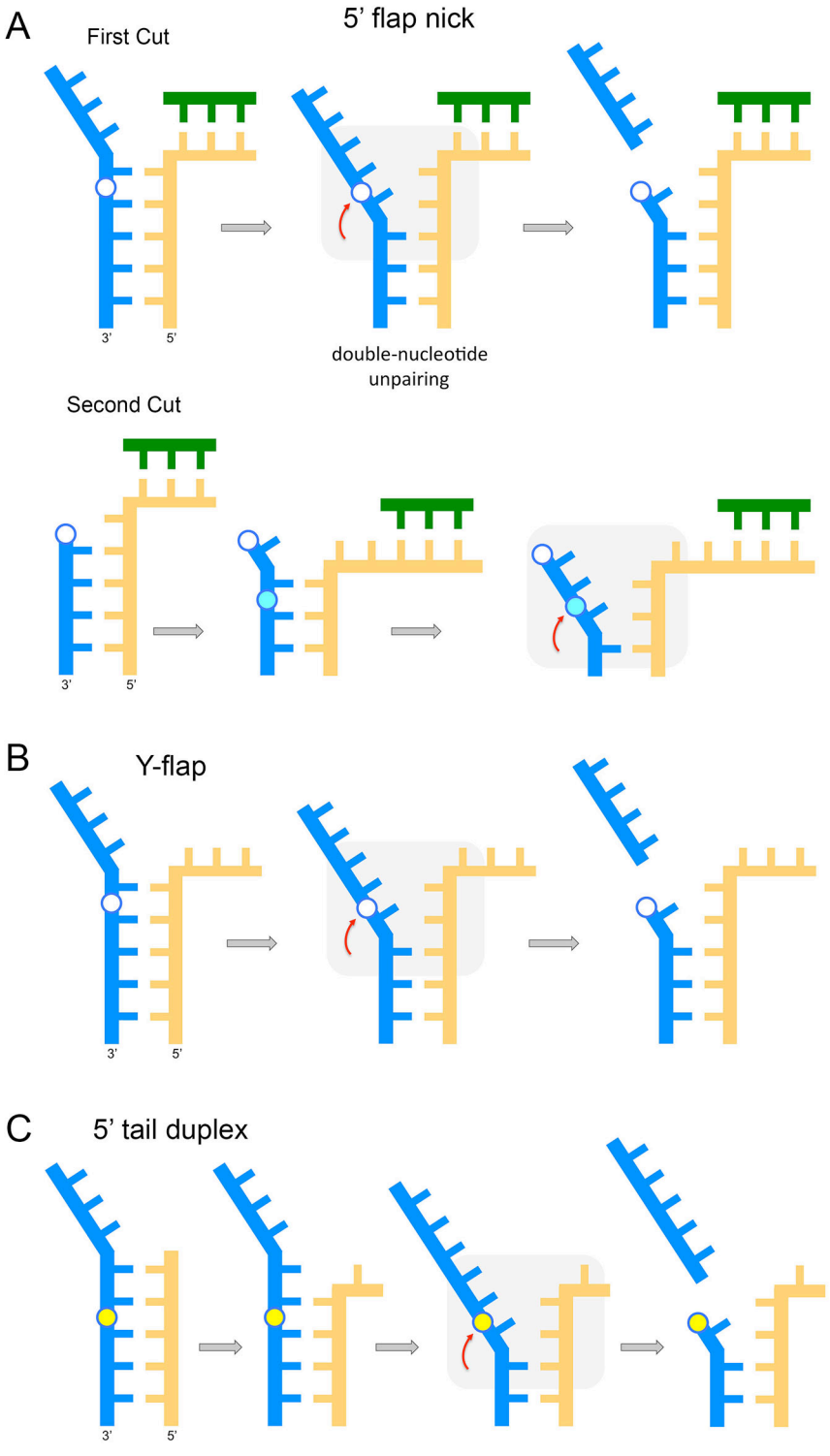
Our experiments here revealed that FenA takes serial second and third “bites” at the 1-nucleotide gap DNA and does so in two-nucleotide cleavage increments. (In contrast, the first cut by FenA at a perfect DNA nick generates a predominant 5′ mononucleotide product.) This dinucleotide stepping is a

novel feature, although it is one that is rationalized by the DNA 3′ template bending model, as illustrated in Figure 3.9A. We envision that the second cut is preceded by a conformational rearrangement of the 1-nucleotide gap product of the first cut, entailing a ratcheting forward of the template strand bend point to a position 1 nucleotide into the duplex. Starting from this new ground state, FenA hydrolyzes the (blue) phosphodiester two steps downstream from the (white) 5′-phosphate generated by the first cut. The observation that iterative cleavages of the 5′ flap nick substrate by FenA cease after two or three dinucleotide bites is likely due to shortening of the residual duplex segment to the point that it cannot satisfy key DNA contacts between the conserved helix-hairpin-helix motif of the enzyme and upstream phosphates of the template strand (11, 12, 18, 35).

The behavior of FenA in the cleavage of Y-flap substrates is also in keeping with the model, as depicted in Figure 3.9B, where the preferred site of incision is between the first and second nucleotides of the duplex segment. However, this preference was less stringent for Y-flaps than for 5′ flap nicks, insofar as the Y-flaps were also incised at a low but appreciable level at an alternative site between the second and third nucleotides of the duplex.

An informative shift in FenA site specificity occurred when the 3′ single strand was eliminated so that a Y-flap was transformed into a 5′-tailed duplex, which was now incised primarily between the second and third nucleotides of the duplex segment. (In contrast, it was reported that *E. coli* Pol1 and *Taq* DNA polymerase incise a 5′-tailed duplex between the first and second

Figure 3.9 DNA template strand bending model as applied to FenA. (A) Sequential cleavages at a 5′ flap nick. (First cut) The nick flap DNA duplex is depicted with a yellow DNA template strand, blue 5′ flap DNA strand, and green DNA_{OH} strand. The scissile phosphodiester of the 5′ flap strand is indicated by the white circle. FenA elicits a sharp bend in the template strand and positions the scissile phosphodiester for metal-catalyzed hydrolysis, indicated by the red arrow. The product of the first cut is a 1-nucleotide gap duplex. (Second cut) The phosphodiester (blue circle) two steps downstream from the 5′-phosphate (white circle) of the 1-nucleotide gap duplex is targeted for hydrolysis after a proposed conformational rearrangement of the bend point of the template DNA strand. (B) Incision of a Y-flap. FenA cleaves the Y-flap via the DNA template strand-bending mechanism, whereby the preferred scissile phosphodiester (white circle) is between the first and second nucleotides of the duplex segment. (C) Incision of a 5′-tailed duplex. FenA engages the 5′-tailed duplex in a conformation in which the terminal base pair is melted and the enzyme can trap the terminal 3′ nucleotide of the template strand in the bent state. FenA then induces hydrolysis of the scissile phosphodiester (yellow circle) between the second and third nucleotides of the original duplex end.


A
First Cut
5’ flap nick
3’
5’
double-nucleotide
unpairing
Second Cut
3’
5’
B
Y-flap
3’
5’
C
5’ tail duplex
3’
5’

nucleotides of the duplex [14], i.e., at the same position at which they cleave a 5′-flap nick.) To account for this behavior, we posit that FenA selects for engagement with the 5′-tailed duplex in a conformation in which the terminal base pair is melted and the enzyme can trap the terminal 3′ nucleotide of the template strand in the bent state, as illustrated in Figure 3.9C. From this ground state, FenA induces hydrolysis of the scissile phosphodiester between the second and third nucleotides of the original duplex end. Our results underscore that the imputed 3′ bend in the template strand is critical for FenA incision of the scissile 5′ strand.

Finally, our characterization of FenA activity in removing the 5′ App(dN) terminus of an aborted nick ligation reaction intermediate highlights FenA as an agent of repair of such lesions in mycobacteria. This is a topic that has received little attention in bacterial DNA repair, although it has been investigated intensively in eukarya, where aprataxin is a key catalyst of deadenylylation of nicked AppDNAs (24–26). Our findings regarding mycobacterial FenA resection of App(dN) resonate with those of the Wilson laboratory, who showed that human FEN1 can remove an App(deoxyribose) adduct resulting from aborted ligation at a 5′-phosphate-terminated abasic lesion in a nicked or 1-nucleotide gapped duplex and also an App(dN) adduct at a conventional nick (36). Human Fen1, like mycobacterial FenA, incised the App(dN) terminus of a conventional nick between the first and second nucleotides of the duplex segment.

Whereas FenA is the first mycobacterial enzyme shown to repair the

adenylylated end of an aborted nick ligation reaction, it is unknown if mycobacteria have other enzymes that can perform this function, similar to aprataxin. In that vein, it is noteworthy that the *M. smegmatis* (and *M. tuberculosis*) proteome includes three members of the HIT enzyme superfamily. MSMEG_2932 (Rv2613c) is a homotetrameric ApppA phosphorylase (Protein Data Bank [PDB] code 3ANO) (37, 38). MSMEG_5871 (Rv0759c) has not been characterized. MSMEG_5028 (Rv1262c) has been characterized structurally (PDB entry 3O0M), but its biochemical activity is uncharted.

EXPERIMENTAL PROCEDURES

Recombinant *M. smegmatis* FenA

The open reading frame encoding *M. smegmatis* FenA was PCR amplified from *M. smegmatis* genomic DNA with primers that introduced a BamHI site immediately flanking the start codon and a HindIII site downstream of the stop codon. The PCR product was digested with BamHI and HindIII and ligated into pET28b-His_{10}Smt3 that had been digested with BamHI and HindIII. The resulting pET28b-His_{10}Smt3-FenA expression plasmid encodes the FenA polypeptide fused to an N-terminal His_{10}Smt3 tag under the transcriptional control of a T7 RNA polymerase promoter. A plasmid encoding mutant FenA (D125A) was obtained by site-directed mutagenesis using the QuikChange method with NED Phusion polymerase. The plasmid inserts were sequenced to verify that no unintended coding changes were acquired during amplification and cloning.

The pET28b-His_{10}Smt3-FenA plasmids were introduced into *Escherichia coli* BL21(DE3) cells. Cultures (1 liter) amplified from single kanamycin-resistant transformants were grown at 37°C in Terrific broth containing 0.4% (vol/vol) glycerol and 60 μg/ml kanamycin until the A_{600} reached 0.8. The cultures were chilled on ice for 1 h and then adjusted to 2% (vol/vol) ethanol and 0.4 mM isopropyl-β-d-thiogalactopyranoside and incubated for 16 to 18 h at 17°C with constant shaking. All subsequent steps were performed at 4°C. Cells were harvested by centrifugation and resuspended in 25 ml of buffer A (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% sucrose, 20 mM imidazole) containing 1 protease inhibitor cocktail tablet (Roche) and 1 mM dithiothreitol (DTT). Lysozyme was added to a concentration of 1 mg/ml. After incubation for 1 h, the lysate was sonicated to reduce viscosity and the insoluble material was pelleted by centrifugation at 38,000 × *g* for 35 min. The supernatant was mixed for 1 h with 3.5 ml of Ni-nitrilotriacetic acid (Ni-NTA)-agarose resin (Qiagen) that had been equilibrated with buffer A. The resin was recovered by centrifugation and resuspended in 30 ml of buffer A. The washed resin was then recovered by centrifugation, resuspended in 20 ml of buffer B (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol) containing 20 mM imidazole, and poured into a column. The column was washed serially with 4 column volumes of 3 M KCl and 5 column volumes of buffer B containing 20 mM imidazole. The bound protein was eluted serially with 100 mM and 200 mM imidazole in buffer B. The polypeptide compositions of the fractions were monitored by SDS-PAGE. The 100 mM and 200 mM

imidazole eluates containing His_{10}Smt3-FenA were pooled and supplemented with the Smt3-specific protease Ulp1 (60 μg) and then dialyzed against 2 liters of buffer C (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM DTT, 20 mM imidazole, 10% glycerol) for 16 h, during which time the His_{10}Smt3 was cleaved. The dialysates were mixed with 3.5 ml of Ni-NTA-agarose equilibrated with buffer C. The mixtures were nutated for 30 min and then poured into a column; the tag-free FenA proteins were recovered in the flowthrough fractions. The FenA protein preparations were purified further by gel filtration through a Superdex-200 column equilibrated in buffer D (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM DTT, 1 mM EDTA, 10% glycerol). FenA eluted as a single discrete peak at an elution volume consistent with it being a monomer in solution. The ratio of A_{260} to A_{280} of the FenA peak was 0.56, indicative of pure protein without nucleic acid contamination. Indeed, a very small A_{260} peak was seen in the void volume, signifying that any residual nucleic acid was eliminated during gel filtration. Peak FenA fractions were pooled and stored at −80°C. Protein concentrations were determined by measuring the A_{280} with a NanoDrop 2000 spectrophotometer and applying an extinction coefficient of 32,430 M^{-1} cm^{-1}, as calculated by ProtParam. The yields of wild-type FenA and D125A mutant were ~5 mg per liter of bacterial culture.

Nuclease substrates

The nicked and flap-nicked substrates were composed of an 18- or 20-mer DNA_{OH} strand and a 5′ ^{32}P-labeled 18- to 34-mer pDNA strand annealed

to a 36-mer DNA template strand. The 5′ ^{32}P-labeled DNA was prepared by reaction of a synthetic oligonucleotide with T4 polynucleotide kinase and [γ-^{32}P]ATP. The DNA was separated from free ATP by electrophoresis through a nondenaturing 18% polyacrylamide gel and then eluted from an excised gel slice. To form the nicked and flap-nicked duplexes, the radiolabeled pDNA strand, DNA_{OH} strand, and DNA template strand were annealed at a 1:5:2 molar ratio in 200 mM NaCl, 10 mM Tris-HCl, pH 6.8, 1 mM EDTA by incubating for 10 min at 65°C, 15 min at 37°C, and then 30 min at 22°C. Y-flap substrates lacked the DNA_{OH} strand.

The 3′ ^{32}P-labeled nuclease substrate was prepared as follows. A 28-mer oligonucleotide 5′-TAACGCTGGGCACTATCGGAATAAGGGC was mixed with a 2-fold molar excess of a complementary 16-mer DNA strand 3′-ATAGCCTTATTCCCGT to form a primer-template with a 15-nucleotide duplex, a 13-nucleotide 5′ tail on the 28-mer strand, and a 1-nucleotide 5′-T overhang on the 16-mer strand that served to template [α-^{32}P]dATP incorporation by Klenow DNA polymerase. The 3′ [^{32}P]dAMP-labeled 29-mer DNA strand was gel purified and then mixed with an 18-mer DNA_{OH} strand and 36-mer DNA template strand at a 1:5:2 molar ratio in 200 mM NaCl, 10 mM Tris-HCl, pH 6.8, 1 mM EDTA. The mixture was incubated for 10 min at 65°C, 15 min at 37°C, and then 30 min at 22°C.

The 5′ ^{32}P-labeled AppDNA strand was prepared by incubating 5′ ^{32}P-labeled pDNA with recombinant E. coli RtcA (30) and 1 mM ATP for 30 min at 37°C. The gel-purified AppDNA strand was mixed with an 18-mer DNA_{OH}

strand and 36-mer DNA template strand at a 1:5:2 molar ratio in 200 mM NaCl, 10 mM Tris-HCl, pH 6.8, 1 mM EDTA and then serially incubated for 10 min at 65°C, 15 min at 37°C, and 30 min at 22°C.

Nuclease assay

Reaction mixtures (10 μl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, ^{32}P-labeled DNA substrate, and FenA (as specified in the figure legends) were incubated at 37°C for 30 min. The reactions were quenched by adding 10 μl of 90% formamide, 50 mM EDTA. The samples were heated at 95°C for 5 min and then analyzed by electrophoresis through a 40-cm 18% polyacrylamide gel containing 7.5 M urea in 44.5 mM Tris-borate, pH 8.3, 1 mM EDTA. The products were visualized by autoradiography and, where specified, quantified by using a Typhoon FLA7000 imager. For analysis by thin-layer chromatography (TLC), the reactions were quenched with 2 μl of 10 mM EDTA. Aliquots (5 μl) of the mixtures were applied to a polyethyleneimine-cellulose TLC plate; each sample was overlaid with 1 μl of a 10 mM solution of cold dCMP. The mobile phase for ascending TLC was 0.5 M or 0.75 M LiCl. The radiolabeled species were visualized by autoradiography, and the cold dCMP markers were localized by UV illumination.

Recombinant *M. smegmatis* Pol1-N

The open reading frame of amino acids 1 to 303 of *M. smegmatis* DNA polymerase I (MSMEG_3839) was PCR amplified with primers that introduced a BglII site immediately flanking the start codon and an XhoI site downstream

of the stop codon. The PCR product was digested with BglII and XhoI and ligated into pET28B-His_{10}Smt3 vector that had been digested with BamHI and XhoI. A plasmid encoding mutant Pol1-N-(D130A) was obtained by site-directed mutagenesis using the QuikChange method with Agilent *Pfu* Turbo DNA polymerase. The plasmid inserts were sequenced to verify that no unintended coding changes were acquired during amplification and cloning. The pET28B-His_{10}Smt3-Pol1-N plasmids were introduced into *E. coli* BL21(DE3) cells. Cultures (1 liter) amplified from single kanamycin-resistant transformants were grown at 37°C in Luria-Bertani medium containing 60 μg/ml kanamycin until the A_{600} reached 0.6. The cultures were brought to room temperature and then adjusted with 2% (vol/vol) ethanol and 0.4 mM isopropyl-β-d-thiogalactopyranoside and incubated for 3 h at 30°C with constant shaking at 225 rpm. Cells were harvested by centrifugation and the pellets were stored at −80°C. All subsequent steps were performed at 4°C. Lysate preparation, nickel affinity chromatography, cleavage of the His_{10}Smt3 tag, and recovery of tag-free Pol1-N proteins in the flowthrough fraction during a second nickel affinity step were performed as described above for FenA purification, with minor modifications. The Pol1-N preparations were frozen and stored at −80°C. Protein concentrations were determined by using the Bio-Rad dye reagent, with bovine serum albumin as the standard. The yields of wild-type Pol1-N and D130A mutant were ∼1.6 mg and ∼6.4 mg per liter of bacterial culture, respectively.

ACKNOWLEDGEMENTS

Shreya Gosh conducted the research on recombinant *M. smegmatis* Pol1-N.

REFERENCES

1. 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**, 1423–37.

2. Unciuleac, M.C. and Shuman, S. (2010) Double strand break unwinding and resection by the mycobacterial helicase-nuclease AdnAB in the presence of single strand DNA-binding protein (SSB). *J. Biol. Chem*., **285**, 34319–34329.

3. Unciuleac, M.C. and Shuman, S. (2010) Characterization of the mycobacterial AdnAB DNA motor provides insights into the evolution of bacterial motor-nuclease machines. *J. Biol. Chem*., *285*, 2632–2641.

4. Gupta, R., Unciuleac, M.C., Shuman, S. and Glickman, M.S. (2017) Homologous recombination mediated by the mycobacterial AdnAB helicase without end resection by the AdnAB nucleases. *Nucleic Acids Res*., **45**, 762-774.

5. Vlasić, I., Ivancić-Baće, I., Imesek, M., Mihaljević, B. and Brcić-Kostić, K. (2008) RecJ nuclease is required for SOS induction after introduction of a double-strand break in a RecA loading deficient *recB* mutant of *Escherichia coli*. *Biochimie*., **90**, 1347-1355.

6. Ivancic-Bace, I., Salaj-Smic, E. and Brcic-Kostic, K. (2005) Effects of *recJ*, *recQ*, and *recFOR* mutations on recombination in nuclease-deficient *recB recD* double mutants of *Escherichia coli*. *J. Bacteriol*., **187**, 1350-1356.

7. Cheng, K., Xu, H., Chen, X., Wang, L., Tian, B., Zhao, Y. and Hua, Y. (2016) Structural basis for DNA 5´-end resection by RecJ. *eLife*, **5**, e14294.

8. Srivastav, R., Kumar, D., Grover, A., Singh, A., Manjasetty, B.A., Sharma, R. and Taneja, B. (2014) Unique subunit packing in mycobacterial nanoRNase leads to alternate substrate recognitions in DHH phosphodiesterases. *Nucleic Acids Res*., **42**, 7894-7910.

9. Mizrahi, V. and Huberts, P. (1996) Deoxy- and dideoxynucleotide discrimination and identification of critical 5' nuclease domain residues of

the DNA polymerase I from *Mycobacterium tuberculosis*. *Nucleic Acids Res.*, **24**, 4845-4852.

10. Tsutakawa, S.E., Classen, S., Chapados, B.R., Arvai, A.S., Finger, L.D., Guenther, G., Tomlinson, C.G., Thompson, P., Sarker, A.H., Shen, B. *et al.* (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell*, **145**, 198-211.

11. Orans, J., McSweeney, E.A., Iyer, R.R., Hast, M.A., Hellinga, H.W., Modrich, P. and Beese, L. (2011) Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell*, **145**, 212-223.

12. Lyamichev, V., Brow, M.A. and Dahlberg, J.E. (1993) Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science*, **260**, 778-783.

13. Xu, Y., Derbyshire, V., Ng, K., Sun, X.C., Grindley, N.D. and Joyce, C.M. (1997) Biochemical and mutational studies of the 5'-3' exonuclease of DNA polymerase I of *Escherichia coli*. *J. Mol. Biol.*, **268**, 284-302.

14. Lyamichev, V., Brow, M.A., Varvel, V.E. and Dahlberg, J.E. (1999) Comparison of the 5' nuclease activities of *Taq* DNA polymerase and its isolated nuclease domain. *Proc. Natl. Acad. Sci. USA*, **96**, 6143-6148.

15. Harrington, J.J. and Lieber, M.R. (1994) The characterization of a mammalian DNA structure-specific endonuclease. *EMBO J.*, **13**, 1235-1246.

16. AlMalki, F.A., Flemming, C.S., Zhang, J., Feng, M., Sedelnikova, S.E., Ceska, T., Rafferty, J.B., Sayers, J.R. and Artymiuk, P.J. (2016) Direct observation of DNA threading in flap endonuclease complexes. *Nat. Struct. Mol. Biol.*, **23**, 640-646.

17. Anstey-Gilbert, C.S., Hemsworth, G.R., Flemming, C.S., Hodskinson, M.R., Zhang, J., Sedelnikova, S.E., Stillman, T.J., Sayers, J.R. and Artymiuk, P.J. (2013) The structure of *Escherichia coli* ExoIX – implications for DNA binding and catalysis in flap endonucleases. *Nucleic Acids Res.*, **41**, 8357-8367.

18. Patel, N., Atack, J.M., Finger, L.D., Exell, J.C., Thompson, P., Tsutakawa, S., Tainer, J.A., Williams, D.M. and Grasby, J.A. (2012) Flap endonucleases pass 5'-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5'-ends. *Nucleic Acids Res*., **40**, 4507-4519.

19. Zhu, H., Bhattarai, H., Yan, H., Shuman, S. and Glickman, M.S. (2012) Characterization of *Mycobacterium smegmatis* PolD2 and PolD1 as RNA/DNA polymerases homologous to the POL domain of bacterial DNA ligase D. *Biochemistry*, **51**, 10147–10158.

20. Shuman, S. (1995) Vaccinia DNA ligase: specificity, fidelity, and inhibition. *Biochemistry*, **34**, 16138-16147.

21. Sriskanda, V. and Shuman, S. (1998) Mutational analysis of *Chlorella* virus DNA ligase: catalytic roles of domain I and motif VI. *Nucleic Acids Res*., **26**, 4618-4625.

22. Chauleau, M, and Shuman, S. (2016) Kinetic mechanism and fidelity of *Escherichia coli* NAD^+-dependent DNA ligase (LigA). *Nucleic Acids Res*., **44**, 2298-2309.

23. Tumbale, P., Williams, J.S., Schellenberg, M.J., Kunkel, T.A. and Williams, R.S. (2014) Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity. *Nature*, **506**,111-115.

24. Ahel, I., Rass, U., El-Khamisy, S.F., Katyal, S., Clements, P.M., McKinnon, P.J., Caldecott, K.W. and West, S.C. (2006) The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature*, **443**, 713-716.

25. Tumbale, P., Appel, C.D., Kraehenbuehl, R., Robertson, P.D., Williams, J.S., Krahn, J., Ahel, I. and Williams, R.S. (2011) Structure of an aprataxin-DNA complex with insights into AOA1 neurodegenerative disease. *Nat. Struct. Mol. Biol*., **18**, 1189-1195.

26. Zhu, H. and Shuman, S. (2008) Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-OH monoribonucleotide. *J. Biol. Chem*., **283**, 8331-8339.

27. Gong, C., Martins, A., Bongiorno, P., Glickman, M. and Shuman, S. (2004) Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *J. Biol. Chem*., **279**, 20594-20606.

28. Zhu, H. and Shuman, S. (2007) Characterization of *Agrobacterium tumefaciens* DNA ligases C and D. *Nucleic Acids Res*., **35**, 3631-3645.

29. Chakravarty, A.K. and Shuman, S. (2011) RNA 3'-phosphate cyclase (RtcA) catalyzes ligase-like adenylylation of DNA and RNA 5'-monophosphate ends. *J. Biol. Chem*., **286**, 4117-4122.

30. 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.

31. Griffin, J.E., Gawronski, J.D., DeJesus, M.A., Ioerger, T.R., Akerley, B.J., Sassetti, C.M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathogens*., **7**, e1002251.

32. Gordhan, B.G., Andersen, S.J., De Meyer, A.R. and Mizrahi, V. (1996) Construction by homologous recombination and phenotypic characterization of a DNA polymerase domain polA mutant of *Mycobacterium smegmatis*. *Gene*, **178**, 125-130.

33. Fukushima, S., Itaya, M., Kato, H., Ogasawara, N. and Yoshikawa, H. (2007) Reassessment of the in vivo function of DNA polymerase I and RNase H in bacterial cell growth. *J. Bacteriol*., **189**, 8575-8583.

34. Finger, L.D., Atack, J.M., Tsutakawa, S., Classen, S., Tainer, J., Grasby, J. and Shen, B. (2012) The wonders of flap endonucleases: structure, function, mechanism and regulation. *Subcell. Biochem*., **62**, 301-326.

35. Shi, Y., Hellinga, H.W. and Beese, L.S. (2017) Interplay of catalysis, fidelity, threading, and processivity in the exo- and endonucleolytic reactions of human exonuclease I. *Proc. Natl. Acad. Sci. USA*, **114**, 6010-6015.

36. Cağlayan, M., Batra, V.K., Sassa, A., Prasad, R. and Wilson, S.H. (2014) Role of polymerase β in complementing aprataxin deficiency during abasic-site base excision repair. *Nat. Struct. Mol. Biol*., **21**, 497-499.

37. Mori, S., Shibayama, K., Wachino, J. and Arakawa, Y. (2011) Structural insights into the novel diadenosine 5',5-P1,P4-tetraphosphate phosphorylase from *Mycobacterium tuberculosis* H37Rv. *J. Mol. Biol*., **410**, 93-104.

38. Honda, N., Kim, H., Rimbara, E., Kato, A., Shibayama, K. and Mori, S. (2015) Purification and functional characterization of diadenosine 5',5'''-P(1),P(4)-tetraphosphate phosphorylases from *Mycobacterium smegmatis* and *Mycobacterium avium*. *Protein Expr. Purif*., **112**, 37-42.

CHAPTER FOUR

Crystal structure and mutational analysis of *Mycobacterium smegmatis* FenA highlight active site amino acids and three metal ions essential for flap endonuclease and 5' exonuclease activities*

INTRODUCTION

Structure-specific DNA endonucleases of the FEN (flap endonuclease) and FEN-like family are widely distributed in Nature and play important roles in DNA replication, repair and recombination (1, 2). FEN enzymes that incise 5′ flap junctions are exemplified by: (i) the N-terminal 5′ exonuclease domains of bacterial DNA polymerase I (3–7); (ii) free-standing bacterial FEN enzymes such as *Escherichia coli* ExoIX, *Bacillus subtilis* YpcP and *Mycobacterium smegmatis* FenA (7–9); (iii) bacteriophage enzymes T5 Fen (10, 11) and T4 Fen/RNaseH (12, 13); (iv) eukaryal FEN1 (14–16); and (v) archaeal FENs (17–19). Human ExoI is an exemplar of the FEN-like clade (20). Other members of the FEN/FEN-like family include the nucleotide excision repair nuclease Rad2 (21) and the Holliday junction resolvase GEN1 (22, 23). The FEN/FEN-like family are metal-dependent phosphodiesterases. Crystal structures of family members highlight a conserved protein fold and an active site that includes either two or three divalent cations (8, 10–13, 16–24). Structures of T5 Fen, T4 RNaseH, human FEN1 and human ExoI in

* Uson,M.L., Carl,A., Goldgur,Y. and Shuman,S. (2018) Crystal structure and mutational analysis of *Mycobacterium smegmatis* FenA highlight active site amino acids and three metal ions essential for flap endonuclease and 5′ exonuclease activities. *Nucleic Acids Res.*, **46**, 4164-4175.

complexes with DNA (11, 13, 16, 20, 25, 26) collectively indicate that the enzyme engages a 12-bp duplex DNA segment flanking the flap or nick via contacts to a helix-turn-helix motif, enforces a sharp bend in the template DNA strand at the flap/nick junction by insertion of a protein α-helix between the splayed-apart template strand nucleobases, and positions the scissile phosphodiester for metal-catalyzed hydrolysis. The 5′ flap of the scissile strand is threaded through a hole in the protein formed by a helical arch.

We recently identified *M. smegmatis* FenA (Figure 4.1B) as a stand-alone 5′ flap endonuclease on the basis of its homology to the N-terminal 5′ exonuclease domain of mycobacterial DNA polymerase I (Pol1) (7) (Figure 4.2B). A stand-alone FenA is also present in the human pathogen *Mycobacterium tuberculosis* (Figure 4.2B). The purified recombinant 319-aa FenA protein catalyzes manganese-dependent endonucleolytic incision of a nicked DNA with a 5′ flap, predominantly between the first and second nucleotides of the duplex segment of the flap strand. Initial FenA cleavage of the 5′-PO_4 strand at a DNA nick (i.e. its exonuclease activity) also occurs between the first and second nucleotides of the duplex. In addition, FenA removes the 5′ App(dN) terminus of an aborted nick ligation reaction intermediate (7). Such lesions are formed under a variety of circumstances by bacterial NAD^+-dependent DNA ligases (27) and especially by the bacterial non-homologous end-joining (NHEJ) ligases, DNA ligases D and C (28–30). The flap endonuclease and AppDNA removal activities of FenA point toward a role in gap repair during lagging strand DNA replication or NHEJ. The flap

removal activity of FenA might also come into play in the mycobacterial single-strand annealing pathway of DNA double-strand break (DSB) repair that entails large DNA deletions between homologous regions flanking the DSB (31).

Our goal in the present study was to determine the atomic structure of mycobacterial FenA in complex with its manganese cofactor, in order to gain insights into its mechanism and its relatedness to other FEN/FEN-like enzymes. We report that the FenA active site contains three manganese ions (M1, M2 and M3), similar to the three-magnesium structure determined for T5 Fen (11). A distinctive feature of FenA is that it does not have the helical arch module found in many other FEN/FEN-like enzymes. Rather, the analogous segment of FenA comprises a short 3_{10} helix and surface β-loop that engages a fourth manganese ion (M4). We find that the direct and water-bridged side chain carboxylate contacts to all three active site manganese ions are essential for FenA's flap endonuclease and 5′ exonuclease activities. Determining atomic structures of three nuclease-dead FenA Asp-to-Asn mutants revealed how these changes elicit a failure to bind one of the three active-site Mn^{2+} ions, in a distinctive fashion for each Asn change, without affecting overall protein structure. The structure of FenA with a phosphate anion coordinated by M1 and M2 in a state mimetic of a product complex suggests a mechanism for metal-catalyzed phosphodiester hydrolysis.

RESULTS

Structure of mycobacterial FenA in complex with manganese

The structure of FenA was determined using SAD phases obtained for a crystal of SeMet-substituted FenA in space group $P2_1$ grown from a solution of 0.1 mM SeMet-FenA and 10 mM $MnCl_2$. The refined model at 1.8 Å resolution (R_{work}/R_{free} 0.174/0.205; Table 4.1) comprised a continuous FenA polypeptide from amino acids 2–318 (Figure 1A). The secondary structure elements (12 β strands, 13 α helices and 3 3_{10} helices) are displayed over the primary structure in Figure 1B. The tertiary structure includes a 7-strand β-sheet with topology β8↓•β7↑•β6↑•β1↑•β2↑•β5↑•β9↓, proceeding from front to back in the view in Figure 1A. The sheet is flanked on the upper right side by a five-helix bundle (α1, α2, α4, α12, α13) and on the lower left side by the other eight α helices, the 3_{10} helices, and a surface β3–β4 loop (Figure 4.1A). The FenA model includes three Mn^{2+} ions in the active site (M1, M2 and M3) and a fourth Mn^{2+} (M4) engaged by the β3–β4 loop (Figure 4.1A). Eight acidic amino acids that coordinate the three manganese ions in the active site are denoted by red dots below the primary structure in Figure 4.1B.

Relatedness of FenA to other FEN family members

A DALI search (37) with the FenA structure identified other FEN/FEN-like proteins as homologs with *Z*-scores > 10 (Table 4.2). The top DALI hits, with Z scores greater than 20, were: *E. coli* ExoIX (Z score 25.1; 2.0 Å rmsd at 243 Cα positions with 25% amino acid identity); the N-terminal 5′ exonuclease domain of *Taq* DNA polymerase (Z score 23.1; 3.1 Å rmsd at 254 Cα positions with 26% amino acid identity); and bacteriophage T5 Fen (*Z*-score 22.5; 2.7 Å rmsd at 242 Cα positions with 21% amino acid identity). Figure 4.2A shows a

Table 4.1 Crystallographic data and refinement statistics

	SeMet-WT	D125N	D148N	D208N
Data collection				
Beamline	APS 24-ID-C	APS 24-ID-E	APS 24-ID-E	APS 24-ID-C
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions a, b, c (Å) α, β, γ (°)	 63.1, 39.8, 68.2 90, 108.9, 90	 63.2, 39.9, 68.4, 90, 108.8, 90	 63.4, 40.1, 68.4, 90, 109.1, 90	 63.2, 39.9, 68.4, 90, 109.1, 90
Resolution (Å)	50–1.8(1.85-1.8)	50–2.2 (2.25–2.2)	50–1.8 (1.85-1.8)	50–1.9 (1.95-1.9)
Wavelength (Å)	0.9791	0.9792	0.9792	0.9791
R_{pim}	0.060 (0.447)	0.075 (0.420)	0.031 (0.133)	0.068 (0.378)
CC(1/2)	0.997 (0.600)	0.995 (0.810)	0.995 (0.915)	0.996 (0.620)
$\langle I \rangle / \langle \sigma I \rangle$	20.9 (2.0)	5.9 (1.4)	22.2 (3.0)	10.0 (1.5)
Completeness (%)	98.8 (97.4)	99.0 (99.9)	98.0 (87.2)	98.9 (97.1)
Redundancy	5.4 (3.1)	4.1 (4.1)	3.9 (2.8)	4.2 (3.3)
Unique reflections	29804	16681	29647	25196
Phasing				
HA sites	2 Se, 1 Mn			
Figure of merit	0.338			
Refinement				
R_{work} / R_{free}	0.174 / 0.205	0.177 / 0.226	0.157 / 0.193	0.185 / 0.220
B-factors (Å^2) Average/Wilson	28.5 / 23.8	37.7 / 31.3	23.9 / 19.1	24.1 / 20.3
RMS deviations bond lengths (Å) bond angles (°)	 0.006 0.887	 0.007 0.938	 0.006 0.903	 0.007 0.977
Ramachandran plot % favored % allowed % outliers	 99.4 0.6 0	 98.1 1.6 0.3	 97.8 2.2 0	 98.7 1.3 0
Model contents				
Protomers / ASU	1	1	1	1
Protein residues	316	316	316	316
Ions	4 Mn	3 Mn	3 Mn	3 Mn, 1 phosphate
Water	271	251	431	292
PDB ID	6C33	6C34	6C35	6C36

Values in parentheses refer to the highest resolution shell.

R_{free} set consists of 10% of data chosen randomly against which structures was not refined.

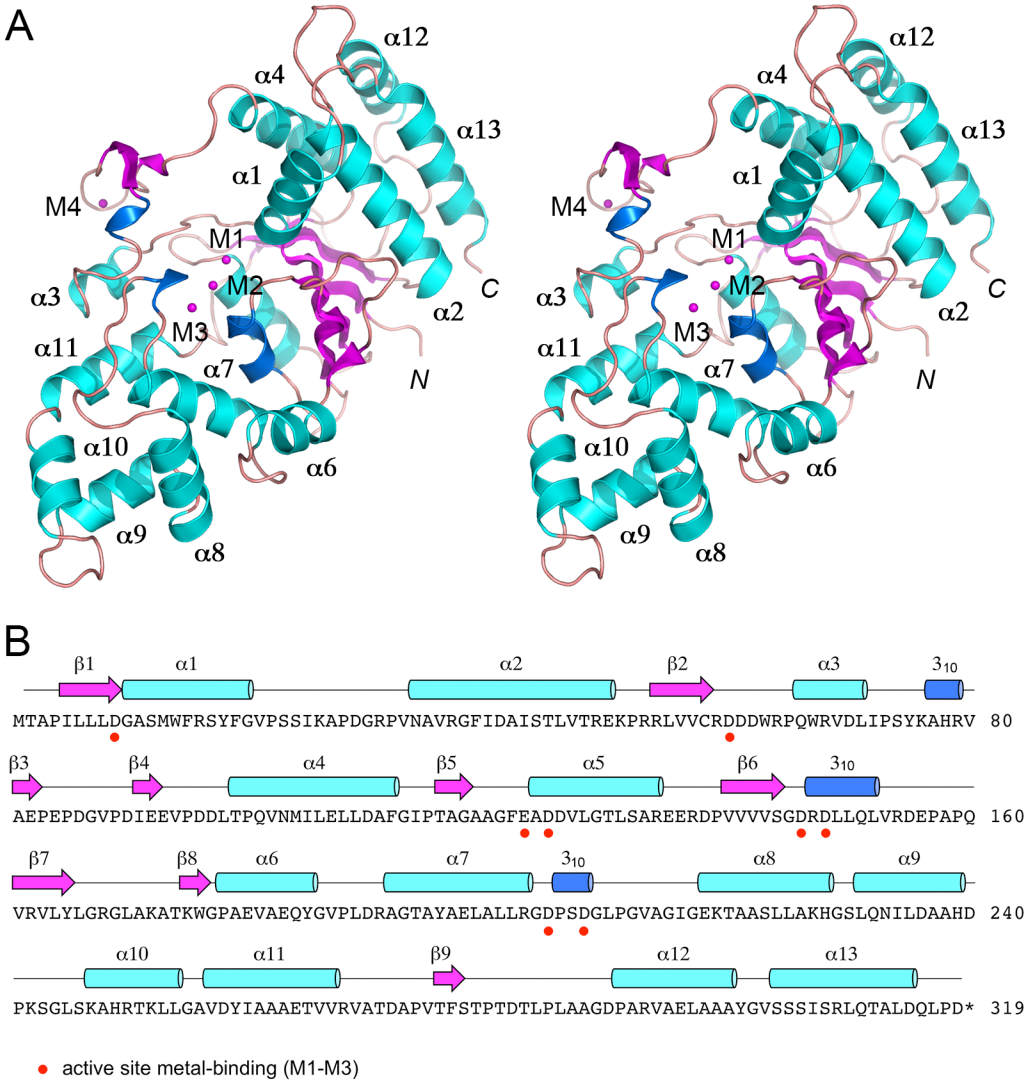


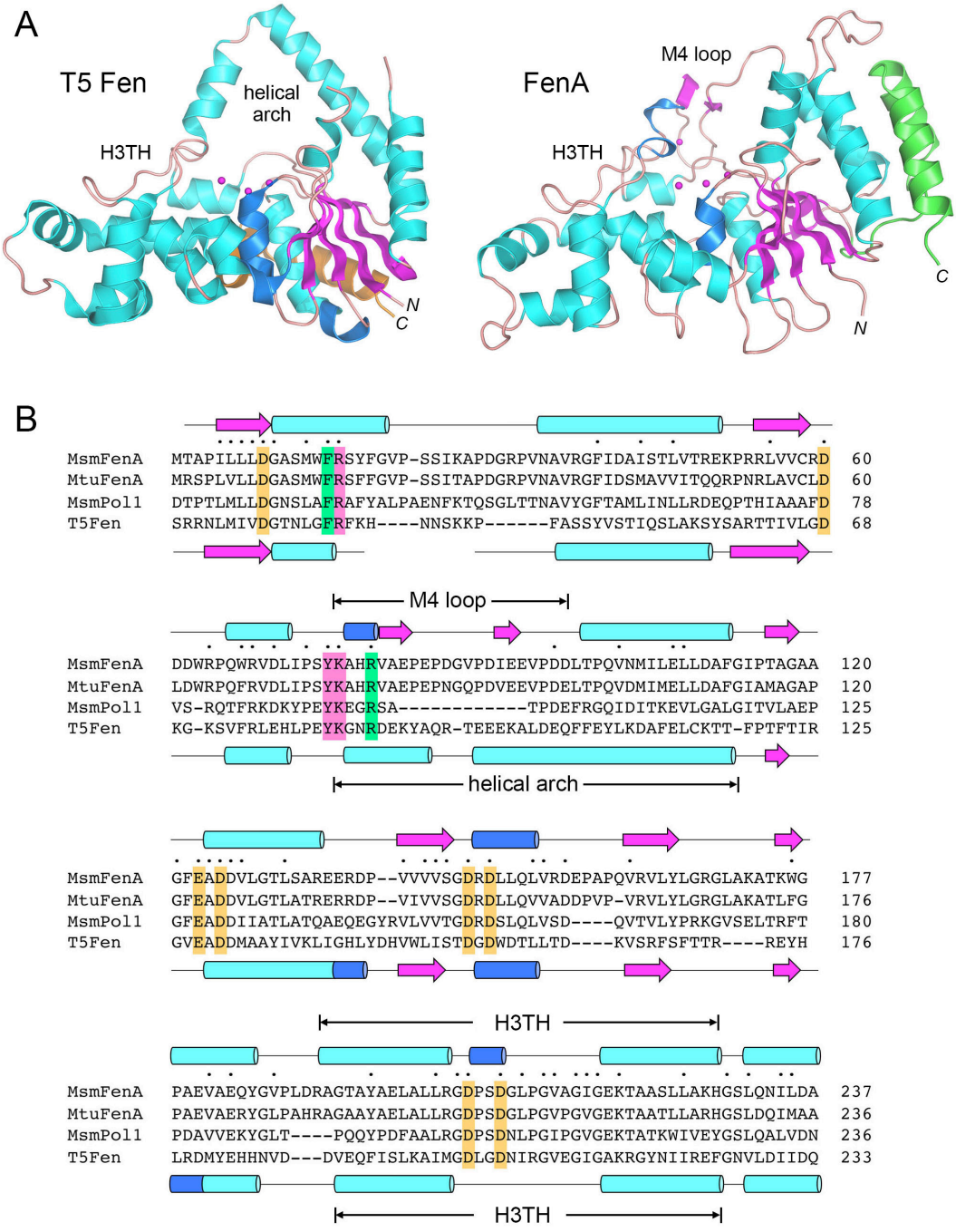
Figure 4.1 Overview of the FenA structure. (A) Stereo view of the tertiary structure of FenA, depicted as a cartoon model with magenta β strands, cyan α helices (numbered sequentially), and blue 3_{10} helices. The *N* and *C* termini are indicated in italics. Three manganese ions in the active site (M1, M2, M3) and a fourth manganese (M4) engaged by the β3–β4 loop are shown as magenta spheres. (B) The secondary structure elements of FenA are shown above the FenA amino acid sequence, labeled sequentially and colored as in panel A. The eight acidic amino acid side chains that bind the active site metals are denoted by red dots below FenA the primary structure.

side-by-side superposition of the structure of the 276-aa T5 Fen protein with three Mg^{2+} ions in the active site (11) on the present structure of FenA with three Mn^{2+} ions in the active site. The homology spans the N-terminal 240-aa segment of *M. smegmatis* FenA, the primary structure of which is aligned in Figure 4.2B to the corresponding segments of *M. tuberculosis* FenA, the 5′ exonuclease/Fen domain of mycobacterial Pol1 (6, 7), and T5 Fen, revealing 64 positions of amino acid identity/similarity in all four proteins. The eight acidic amino acids that bind the three active site metal ions are strictly conserved and are highlighted in gold shading in Figure 4.2B. The secondary structure elements of FenA and T5 Fen are arrayed above and below their primary structures in Figure 4.2B and are mostly conserved, with exceptions as noted below.

Table 4.2 Closest structural homologs of FenA

Protein	pdb ID	Z score	root mean square deviation	% identity
Escherichia coli ExoIX	3ZDD	25.1	2.0 Å at 243 Cα positions	25
Taq Polymerase	1TAU	23.1	3.1 Å at 254 Cα positions	26
T5 Fen	1EXN	22.5	2.7 Å at 242 Cα positions	21
Methanopyrus kandleri FEN1	4WA8	18.7	3.3 Å at 251 Cα positions	20
Archaeoglobus fulgidus FEN1	1RXV	18.3	3.9 Å at 254 Cα positions	17
T4 Fen (RNaseH)	2IHN	18.2	3.1 Å at 247 Cα positions	16
Methanococcus jannaschii FEN1	1A77	17.6	3.9 Å at 253 Cα positions	18
Human FEN1	3Q8M	17.5	4.3 Å at 264 Cα positions	16
Pyrococcus furiosus FEN1	1B43	17.2	3.5 Å at 249 Cα positions	16
Sulfolobus solfataricus FEN1	2IZO	15.9	3.2 Å at 217 Cα positions	18
Saccharomyces cerevisiae Rad2	4Q0Z	15.9	4.2 Å at 250 Cα positions	18
Human GEN1	5T9J	15.5	3.4 Å at 253 Cα positions	15
Human ExoI	5V0D	13.6	3.7 Å at 227 Cα positions	13
Caenorhabditis elegans XRN2	5FIR	10.7	4.3 Å at 247 Cα positions	11

Figure 4.2 Comparison of mycobacterial FenA to T5 Fen. (A) Side-by-side superposition of the structures of bacteriophage T5 Fen with three Mg^{2+} ions in the active site (at left, from accession code: 5HMM) and FenA with three Mn^{2+} ions in the active site (at right). Cartoon models of the tertiary structures are shown colored according to secondary structure, except for the two C-terminal α helices that are colored gold in T5 Fen and green in FenA to highlight their different spatial locations. The N and C termini are labeled in italics. The conserved DNA-binding H3TH motifs are labeled. The helical arch of T5 Fen is not conserved in FenA, which has a distinctive M4 loop as shown. (B)The amino acid sequence of *M. smegmatis* (Msm) FenA from aa 1 to 237 is aligned to the corresponding sequences of the homologous *M. tuberculosis* FenA (aa 1 to 236), the N-terminal 5′ exonuclease domain of *M. smegmatis* Pol1 (aa 18–233) and bacteriophage T5 Fen (aa 3–233). Gaps in the alignment are denoted by dashes. Positions of side chain identity/similarity in all four polypeptides are denoted by dots above the MsmFenA sequence. Secondary structure elements of FenA and T5 Fen (colored as in Figure 4.1) are shown above and below their amino acid sequences. Amino acids subjected to mutational analysis in the present study are highlighted in shaded boxes: gold for amino acids coordinating metals M1, M2 and M3; magenta for nearby amino acids in the active site; green for amino acids that, in T5 Fen, make contacts to DNA.


A
T5 Fen
helical
arch
H3TH
N
C
FenA
M4 loop
H3TH
C
N
B
MsmFenA MTAPILLLDGASMWFRSYFGVP-SSIKAPDGRPVNAVRGFIDAISTLVTREKPRRLVVCRD 60
MtuFenA MRSPLVLLDGASMWFRSFFGVP-SSITAPDGRPVNAVRGFIDSMAVVITQQRPNRLAVCLD 60
MsmPol1 DTPTLMLLDGNSLAFRAFYALPAENFKTQSGLTTNAVYGFTAMLINLLRDEQPTHIAAAFD 78
T5Fen SRRNLMIVDGTNLGFRFKH----NNSKKP------FASSYVSTIQSLAKSYSARTTIVLGD 68
M4 loop
MsmFenA DDWRPQWRVDLIPSYKAHRVAEPEPDGVPDIEEVPDDLTPQVNMILELLDAFGIPTAGAA 120
MtuFenA LDWRPQFRVDLIPSYKAHRVAEPEPNGQPDVEEVPDELTPQVDMIMELLDAFGIAMAGAP 120
MsmPol1 VS-RQTFRKDKYPEYKEGRSA-----------TPDEFRGQIDITKEVLGALGITVLAEP 125
T5Fen KG-KSVFRLEHLPEYKGNRDEKYAQR-TEEEKALDEQFFEYLKDAFELCKTT-FPTFTIR 125
helical arch
MsmFenA GFEADDVLGTLSAREERDP--VVVVSGDRDLLQLVRDEPAPQVRVLYLGRGLAKATKWG 177
MtuFenA GFEADDVLGTLATRERRDP--VIVVSGDRDLLQVVADDPVP-VRVLYLGRGLAKATLFG 176
MsmPol1 GFEADDIIATLATQAEQEGYRVLVVTGDRDSLQLVSD----QVTVLYPRKGVSELTRFT 180
T5Fen GVEADDMAAYIVKLIGHLYDHVWLISTDGDWDTLLTD----KVSRFSFTTR----REYH 176
H3TH
MsmFenA PAEVAEQYGVPLDRAGTAYAELALLRGDPSDGLPGVAGIGEKTAASLLAKHGSLQNILDA 237
MtuFenA PAEVAERYGLPAHRAGAAYAELALLRGDPSDGLPGVPGVGEKTAATLLARHGSLDQIMAA 236
MsmPol1 PDAVVEKYGLT----PQQYPDFAALRGDPSDNLPGIPGVGEKTATKWIVEYGSLQALVDN 236
T5Fen LRDMYEHHNVD---DVEQFISLKAIMGDLGDNIRGVEGIGAKRGYNIIREFGNVLDIIDQ 233
H3TH

FenA shares with T5 Fen and other FEN/FEN-like enzymes the helix–turn–helix motif (indicated by H3TH in Figure 4.2A and B) that engages the duplex DNA segment preceding the duplex-flap junction (11, 13, 16, 20) as well as the 'wedge' helix (α1 in FenA) that enforces a sharp bend in the template DNA strand at the flap/nick junction (13, 16, 20). FenA deviates from T5 Fen in that it lacks the helical arch (Figure 4.2A) that creates an aperture through which the 5′ flap DNA is shown to thread, e.g. in T5 Fen, human FEN1 and human ExoI (11, 16, 20). Whereas the structures of some FEN family members that were crystallized without DNA reveal that segments corresponding to the helical arch are disordered (12, 38), this is not the case in FenA. Rather, the FenA segment of primary structure between α3 and α4 adopts a distinctive secondary structure and fold (i.e. *vis à vis* T5 Fen), comprising a short 3_{10} helix and a surface β-loop that coordinates the M4 manganese ion. The M4 β-loop is conserved in length and amino acid sequence in *M. tuberculosis* FenA, but is missing from the 5′ exonuclease domain of mycobacterial Pol1 (Figure 4.2B).

Another distinction between FenA and T5 Fen is that the two-helix bundles at their respective C-termini reside on different surfaces of the enzymes. These C-terminal α helices in T5 Fen (aa 262–290, colored gold in Figure 2A) pack against the α5 helix (aa 128–140) on the back surface of the protein as viewed in Figure 4.2A. By contrast, the C-terminal α12 and α13 helices of FenA (aa 291–318, colored green in Figure 4.2A) lie directly over the α2 and α4 helices on the rightward face of the enzyme as viewed in

Figures 4.1A and 4.2A; the C-terminal helices are connected to the rest of the FenA polypeptide by a long loop (aa 268–290) that drapes across the back surface of the protein (Figure 4.1A and B). Equivalents of the FenA α12 and α13 helices are present in the same spatial position in human FEN1 and yeast Rad2 (16, 21). However, the 251-aa *E. coli* ExoIX enzyme (the best homology hit to FenA per DALI) has no such C-terminal helices (8), nor does the 352-aa catalytic domain of human Exo1 (20).

Metal binding in the FenA active site

We observed three Mn^{2+} ions in the active site, affirmed by three discrete peaks of anomalous scattering in the SeMet-FenA dataset (Figure 4.3) and in a native FenA dataset collected at the wavelength corresponding to Mn absorption peak (not shown). Each manganese ion is coordinated by a rich network of direct and water-mediated contacts to the carboxylate oxygens of conserved acidic side chains (shaded in gold in Figure 4.2B). The eight metal-binding residues in FenA are: Asp9, Asp60, Glu123 and Asp125 (site M1); Asp125, Asp146, Asp148 and Asp208 (site M2); Asp148, Asp205 and Asp208 (site M3) (Figure 4.4A). Sites M1 and M2 comprise what had been referred to previously as the ‘Cat1’ site. The M1 and M2 metal ions, spaced 3.6 Å apart, coordinate a bridging water. Asp125 makes direct bidentate contact to the M1 and M2 metals. The four remaining sites in the octahedral M1 complex are filled by waters bridged to side chains Asp9, Asp60 and Glu123. The four remaining positions in the octahedral M2 complex are occupied by the Asp146 and Asp148 Oδ2 atoms and by two waters—engaged

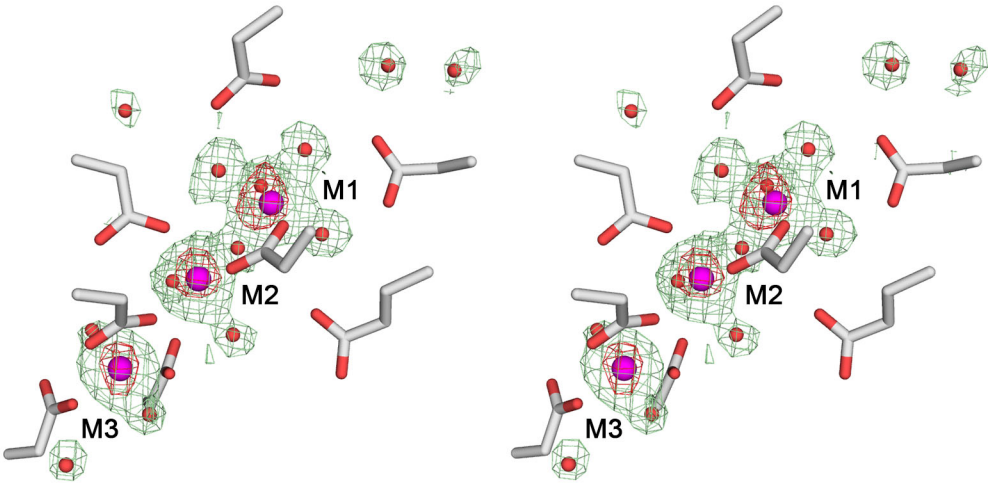


Figure 4.3 Manganese ions in the FenA active site. Stereo view of a simulated annealing Fo-Fc omit map (green mesh contoured at 3σ) and an anomalous difference map (red mesh contoured at 4σ) affirm the presence of three manganese ions (M1, M2, and M3, magenta spheres) in the wild-type FenA active site. Waters are depicted as red spheres. The eight conserved acidic side chains that engage the metal ions are shown as stick models.

to Asp208 and Glu123 Oϵ1, respectively (Figure 4.4A). The M3 site (previously named 'Cat2') is conserved among a subset of FEN/FEN-like family members, e.g. it is not present in *E. coli* ExoIX (8), notwithstanding the close overall homolog of ExoIX to FenA. The M3 manganese in FenA is located 4.1 Å away from M2. The M3 coordination complex is occupied by Asp148 Oδ1 and Oδ2, Asp205 Oδ2, Asp208 Oδ2 and two waters. Other conserved FenA side chains make noteworthy atomic contacts to the outer shell of the metal complexes. The Tyr75 OH makes bifurcated hydrogen bonds to an M2 water and Glu123 Oϵ2 and the Lys76 Nζ makes an ion pair with Glu123 Oϵ1 (Figure 4.4A). [The equivalent Tyr82 and Lys83 side chains in T5 Fen (Figure 4.2B; shaded in magenta) make the same contacts to the conserved metal-binding Glu128.] FenA Arg79 makes a salt bridge to Asp208

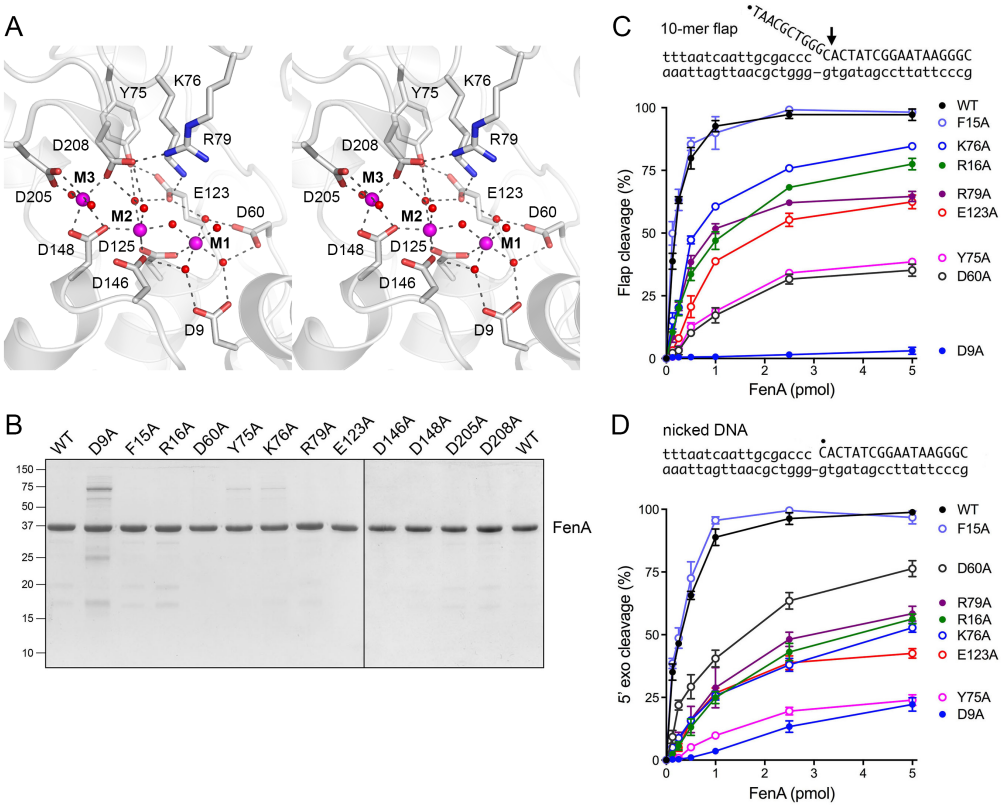


Figure 4.4 Active site architecture and structure-guided mutagenesis. (A) Stereo view of the FenA active site highlighting the octahedral coordination complexes of the manganese ions M1, M2 and M3 (depicted as magenta spheres). Selected amino acid side chains that coordinate the metals, the metal-bound waters (red spheres), and the metal-binding carboxylates are rendered as stick models against a cartoon trace of the FenA tertiary structure. Atomic contacts are denoted by dashed lines. (B) FenA mutants. Aliquots (5 μg) of the preparations of WT FenA and the indicated alanine mutants were analyzed by SDS-PAGE. The Coomassie blue-stained gel is shown. The positions and sizes (in kilodaltons) of marker polypeptides are indicated on the left. (**C** and **D**) Mutational effects on flap endonuclease (C) and 5′ exonuclease activity (D). Reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT and 1 pmol (100 nM) of either ^{32}P-labeled 10-mer flap-nicked duplex DNA (panel C) or ^{32}P-labeled nicked duplex DNA (panel D), and increasing amounts of WT or mutant FenA were incubated for 30 min at 37°C. The DNA substrates are depicted above the graphs, with the 5′ ^{32}P label denoted by •. The reaction products were analyzed by urea-PAGE and visualized and quantified by scanning the gel with a phosphorimager. (The predominant site of cleavage is indicated on the flap substrate by an arrow in C. The same phosphodiester is cleaved in the nicked DNA substrate in D.). The extents of cleavage of the substrates are plotted as a function of input FenA. Each datum is the average of at least three independent titration experiments ± SEM.

(Figure 4.4A). In the structure of T5 Fen bound to a flap duplex DNA, the equivalent arginine (Arg76; shaded in green in Figure 4.2B) makes a network of hydrogen bonds from its guanidinium nitrogens to the non-bridging phosphate oxygens of the first three nucleotides of the flap strand adjacent to the scissile phosphodiester (11).

Effect of alanine mutations on FenA activity

To gauge the contribution of the metal-coordinating carboxylate residues to FenA activity, the Asp9, Asp60, Glu123, Asp146, Asp148, Asp205 and Asp208 side chains were mutated individually to alanine. (We showed previously that changing Asp125 to alanine abolishes FenA flap endonuclease and 5′ exonuclease activities (7)). We also introduced alanine in lieu of Tyr75, Lys76 and Arg79 in the active site (Figure 4.4A, discussed above), as well as two other conserved residues: Phe15 and Arg16 in the α1 helix (Figure 4.2B). The equivalent of Phe15 in T5 Fen (Phe32) demarcates the junction between the duplex and flap segments of the scissile strand, via a π-stack of the phenylalanine on the 5′-most paired nucleobase of the duplex (11). The counterpart of Arg16 in T5 Fen (Arg33) is located next to the 3′ end of the template strand of the flap duplex (11) and is a plausible candidate to interact with the 3′ template or duplex DNA segment of a flap-nick or Y-flap substrate.

Twelve FenA-Ala mutants were produced in *E. coli* in parallel with WT FenA. SDS-PAGE analysis of the recombinant protein preparations is shown in Figure 4.4B. To assay for flap endonuclease activity of FenA, a substrate was prepared by annealing a 5′ ^{32}P-labeled 28-mer DNA strand and an

unlabeled 18-mer DNA_{OH} strand to a 36-mer template DNA strand, to form a singly nicked duplex with a 10-nt 5′ flap (Figure 4.4C). Reaction of WT FenA with 1 pmol (100 nM) nick flap substrate in the presence of 1 mM manganese resulted in FenA concentration-dependent conversion of the input ^{32}P-labeled 28-mer DNA to a predominant ^{32}P-labeled 11-nt product (via incision between the first and second nucleotides of the duplex segment) and a minor 12-nt species (cleaved at the vicinal phosphodiester) that were resolved from the substrate by urea-PAGE (7); the extent of flap cleavage as a function of FenA concentration is plotted in Figure 4.4C for WT FenA and 8 of the FenA-Ala mutants for which product formation was detectable at up to 5 pmol of input FenA. The D146A, D148A, D205A and D208A mutants, which formed no flap cleavage product at 5 pmol of input enzyme, were titrated up to 50 pmol of input FenA. The flap endonuclease-specific activities of the mutants, determined from the slope of the titration profile in the linear range as calculated by linear regression in Prism and then normalized to that of WT FenA, are shown in Table 4.3.

In similar fashion, we assayed 5′ exonuclease activity of WT FenA and FenA-Ala mutants on a nicked DNA substrate, 5′ ^{32}P-labeled at the nick (Figure 4.4D). FenA exonuclease action liberates ^{32}P-dCMP as the predominant product, along with a minor dinucleotide species pCpA (7). The extents of 5′ exonuclease cleavage as a function of input FenA are plotted in Figure 4.4D. Again, the D146A, D148A, D205A and D208A mutants, which formed little or no exonuclease cleavage products at 5 pmol of input enzyme,

were titrated up to 50 pmol of input FenA. The normalized 5′ exonuclease specific activities of the mutants are compiled in Table 4.3.

The salient findings were that loss of the carboxylate functional groups of Asp146 (M2 ligand), Asp148 (M2 and M3 ligand), Asp205 (M3 ligand) and Asp208 (M3 ligand and water-bridged M2 ligand) abolished flap endonuclease and 5′ exonuclease activities (i.e. the alanine mutants had ≤0.1% of the WT-specific activity), signifying that the M2 and M3 coordination complexes are essential for FenA function. Alanine mutation of Asp9, which coordinates two of the M1 waters, virtually eliminated the flap endonuclease (0.2% specific

Table 4.3 Effects of FenA alanine mutations on flap endonuclease and 5′ exonuclease activities

FenA mutant	Specific Activity (% of wild-type FenA)	
	Flap endonuclease	5’ exonuclease
D9A	0.2	3
F15A	75	109
R16A	27	17
D60A	5	44
Y75A	6	5
K76A	35	18
R79A	29	20
E123A	15	18
D146A	≤0.1	≤0.1
D148A	≤0.1	≤0.1
D205A	≤0.1	≤0.1
D208A	≤0.1	≤0.1
D60A-E123A	≤0.1	0.1

activity) and reduced 5′ exonuclease to 3% of WT activity. Yet, alanine substitutions of the other M1 water ligands—Asp60 and Glu123—elicited less severe reductions in flap endonuclease specific activity (to 5 and 15% of WT for D60A and E123A, respectively) and 5′ exonuclease-specific activity (to 44 and 18% of WT for D60A and E123A, respectively). Because Asp60 and Glu123 make atomic contacts to the same water molecule in the M1 coordination complex (Figure 4.4A), we considered the possibility that these two amino acids might be functionally redundant (at least in part). To address this point, we produced and purified a D60A-E123A double-mutant (Figure 4.5A) and found that the simultaneous loss of both carboxylates abolished flap endonuclease and 5′ exonuclease activities (≤0.1% of WT) (Table 4.3). We conclude that all three active site manganese complexes are essential for FenA nuclease activity.

Alanine substitutions for conserved residues Tyr75, Lys76, and Arg79 that interact with the outer spheres of the metal complexes also resulted in lower flap endonuclease and 5′ exonuclease specific activity: 6 and 5% for Y75A; 35 and 18% for K76A and 29 and 20% for R79A (Table 4.3). The R16A mutation in α1 reduced endonuclease and exonuclease activity to 27 and 17% of WT, respectively. However, the F15A mutation was benign.

Effect of Asp to Asn substitutions

The six aspartates deemed essential for FenA activity based on alanine scanning (Asp9, Asp125, Asp146, Asp148, Asp205 and Asp208) were mutated conservatively to asparagine. The purified recombinant Asn mutants

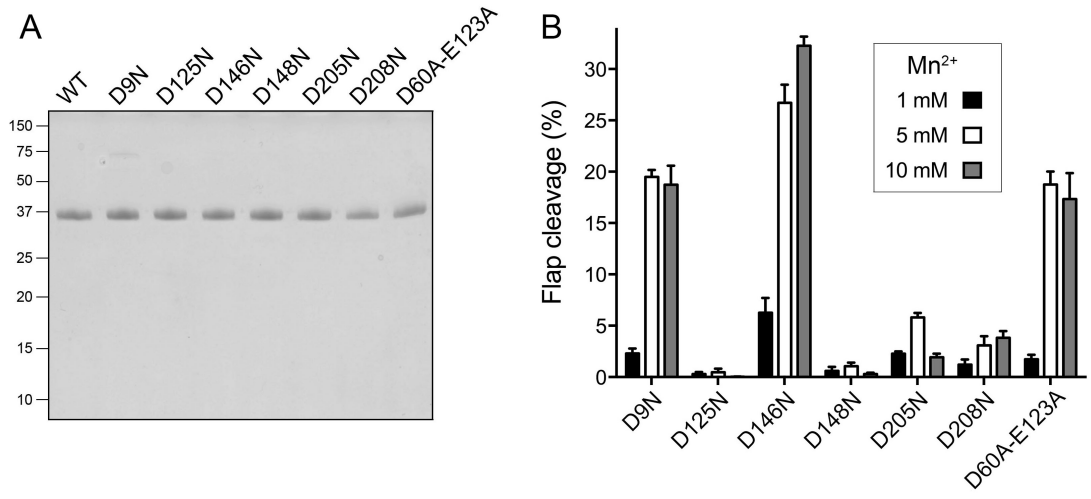


Figure 4.5 FenA Asp-to-Asn mutants and tests of rescue by increasing manganese concentration. (A) Aliquots (5 μg) of the preparations of wild-type (WT) FenA and the indicated Asp-to-Asn mutants as well as the D60A-E123A double-alanine mutant were analyzed by SDSPAGE. The Coomassie blue-stained gel is shown. The positions and sizes (in kilodaltons) of marker polypeptides are indicated on the left. (B) Reaction mixtures (10 μl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1, 5 or 10 mM MnCl2 as specified, 1 mM DTT, 1 pmol (100 nM) 32P-labeled flap-nick substrate, and 10 pmol (1 μM) FenA mutant were incubated for 30 min at 37°C. The extents of flap cleavage by each mutant are plotted in bar graph format as a function of increasing manganese concentration. Each datum is the average of three separate experiments ±SEM.

(Figure 4.5A) were either inert or virtually inert when assayed for flap endonuclease and 5′ exonuclease activity (Table 4.4), i.e. the isosteric asparagine substitutions mimicked the effects of alanine substitution. The simplest explanation for the loss of function is that Asp-to-Asn mutation precludes binding of one or more of the active site manganese ions. Alternatively, Asn mutation might reduce affinity of manganese binding at sites M1, M2 or M3, such that the sites are not filled at the concentration of 1 mM Mn^{2+} established previously as optimal for WT FenA activity (7). To address the latter prospect, we tested whether increasing the manganese

concentration in the flap endonuclease reaction mixtures to 5 and 10 mM might elicit a gain of activity for any of the 'dead' Asn mutants or for the D60A-E123A double-mutant. To achieve sensitivity in this experiment, the level of input mutant enzyme (10 pmol) was ~10-fold greater than that sufficient for complete cleavage of the flap substrate by WT FenA. As shown in Figure 4.5B, increasing manganese concentration to 5 and 10 mM had no significant rescuing effect on the D125N, D148N, D205N or D208N mutants. (Note that the four Mn^{2+} concentration-unresponsive mutations were at FenA aspartates that make direct atomic contact with M1, M2 or M3.) By contrast, the D9N and D60A-E123A proteins (mutated at acidic residues that make only water-mediated contacts to M1 or M2) were 8-fold and 11-fold more active in the presence of 5 and 10 mM Mn^{2+} than at 1 mM Mn^{2+}, resulting in cleavage of 19% of the flap substrate (Figure 4.5B). The activity of the D146N mutant (at an aspartate that contacts M2 directly) was increased 5-fold (to 32% flap cleavage) at 10 mM Mn^{2+} (Figure 4.5B). That said, the apparent specific

Table 4.4 Effects of Asp-to-Asn mutations on flap endonuclease and 5′ exonuclease activities

FenA mutant	Specific Activity (% of wild-type FenA)	
	Flap endonuclease	5'exonuclease
D9N	≤0.1	0.2
D125N	≤0.1	≤0.1
D146N	0.4	0.3
D148N	≤0.1	≤0.1
D205N	≤0.1	≤0.1
D208N	≤0.1	≤0.1

activities of the D9N, D60A-E123A and D146N proteins at higher manganese concentrations were still ≤1.2% of the flap endonuclease activity of WT FenA.

Crystal structures of FenA Asp-to-Asn mutants

Structures of nuclease-dead FenA mutants D125N (at 2.2 Å resolution), D148N (1.8 Å resolution) and D208N (1.9 Å resolution) were determined by diffraction of crystals in space group $P2_1$ grown from FenA mutant protein solutions containing 10 mM $MnCl_2$. Pairwise DALI comparison of the tertiary structure of WT FenA to D125N (rmsd 0.3 Å), D148N (rmsd 0.3 Å), and D208N (rmsd 0.1 Å) indicated they were virtually identical. Superposition of the active sites revealed that each of the Asn mutants lacked one of the three active site manganese ions. To wit, D125N was missing M2 whereas D148N and D208N were missing M3 (Figure 4.6). The presence of only two active site Mn^{2+} in the Asn mutants was affirmed by anomalous peaks coincident with the modeled metal ions (not shown). The effects of the Asn mutations on active site architecture are discussed in detail below.

The D125N change preserves the M1 site in which the Asn125 Oδ makes the same contact to M1 as the Asp125 Oδ2, but fails to fill the M2 site because the Asn Nδ does not accommodate direct M2 coordination. Rather, the Asn125 amide group rotates so that Nδ makes a hydrogen bond to Asp146 (Figure 4.6). The M3 site is also affected in D125N, whereby Asp148 rotates so that it makes only one direct contact to M3 (instead of two as in WT FenA) and the vacated position in the M3 coordination complex is taken by water bridged to Asp146 (in lieu of the Asp146 contact to the missing M2

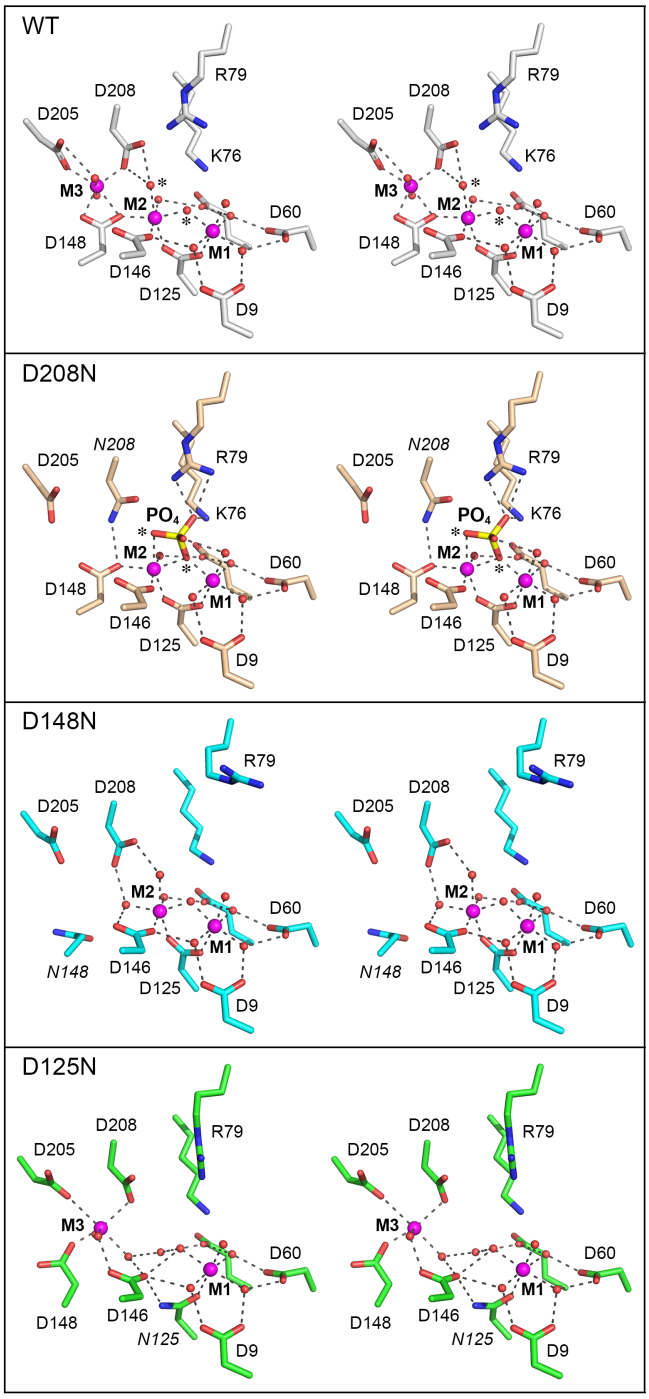


Figure 4.6 Active sites of FenA Asp-to-Asn mutants. The active sites of FenA WT (colored with gray carbons) and mutants D208N (beige carbons), D148N (cyan carbons) and D125N (green carbons) were superimposed and offset vertically. Stereo views are shown. The mutated Asn chains are labeled in italics. Manganese ions are magenta spheres and waters are red spheres. Atomic contacts are denoted by dashed lines.

manganese) (Figure 4.6). Asp208 maintains its M3 contact, but rotates absent the water-bridged interaction with M2, a consequence of which is that the Asp208-Arg79 salt bridge is lost and the Arg79 side chain rotates compared to the WT active site (Figure 4.6).

The D148N mutant fails to fill the M3 site and the Asn148 side chain rotates out of position so that it no longer makes the direct contact to the M2 manganese seen in WT FenA, in lieu of which the M2 coordination complex acquires a water bridged between Asp146 and Asp208 (Figure 4.6). The Arg79 side chain adopts a different rotamer in D148N whereby it points away from Asp208. The M1 site in D148N is unperturbed compared to WT FenA.

The D208N structure lacks M3 and instead the Asn208 Nδ makes a hydrogen bond to Asp148 (Figure 4.6). The salient feature of the D208N enzyme is the presence of a tetrahedron-shaped density in the active site that we modeled as a phosphate anion. [Note that phosphate was not included in the Fen-D208N plus manganese enzyme solution, nor was it supposed to be present in the commercial 0.2 M ammonium acetate, 20% PEG3350 precipitant solution used to grow the D208N crystals. However, by assaying all solutions for inorganic phosphate by the malachite green method, we determined that the commercial precipitant solution contained 0.4 mM inorganic phosphate.] The phosphate O1 oxygen is coordinated by the M2 manganese (at site in the M2 coordination complex occupied by a water in WT FenA) and its O2 oxygen is bridged between M2 and M1, in lieu of the bridging water seen in WT FenA. (The phosphate O1 and O2 atoms and the

waters they replace are indicated by asterisks in the top two panels in Figure 4.6.) The phosphate O3 atom is engaged by Lys76 Nζ and the terminal guanidinium nitrogens of Arg79. The phosphate O4 atom is bridged by a water to the M1 manganese (Figure 4.6). The D208N structure with phosphate is instructive with respect to the FenA mechanism, insofar as we construe it to be a plausible mimetic of a product complex of the enzyme with the 5′-PO_4 group of the incised flap strand, as discussed further below.

The M4 loop of FenA engages a fourth metal ion outside the active site

WT FenA and each of the Asn mutants that were crystallized contained a fourth Mn^{2+} engaged in the β3-β4 loop that packs against the FenA surface by virtue of metal-contacts and an Arg64-Asp90 salt bridge (Figure 4.7). The octahedral M4 complex is filled by Asp90 Oδ1, the Glu84 main-chain carbonyl oxygen, and four waters, one of which is bridged between Gln66 Oϵ and

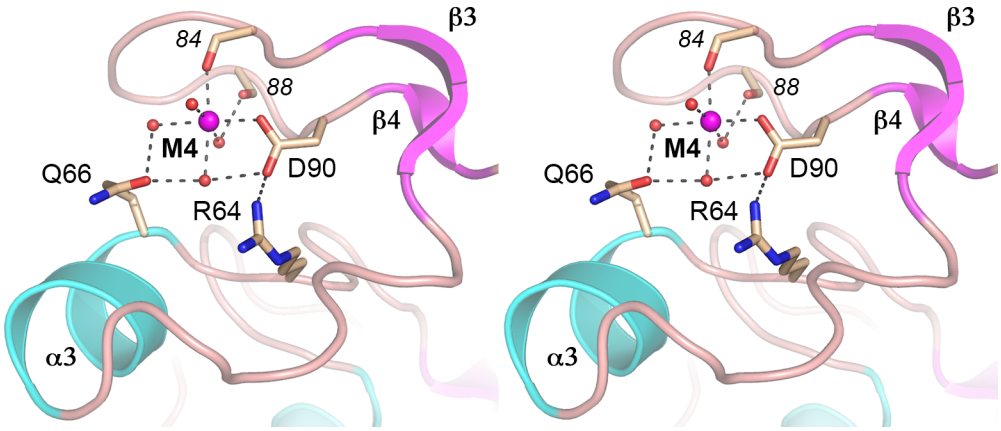


Figure 4.7 Manganese engagement at the M4 loop. Stereo view of the M4 loop, depicted as a cartoon trace colored and labeled according to secondary structure. The M4 manganese is depicted as a magenta sphere and waters as red spheres. Selected amino acid side chains (labeled in plain font) and main-chain carbonyls (in italics) are shown as stick models with beige carbons. Atomic contacts are indicated by dashed lines.

Asp90 Oδ2, while another is engaged to Gln66 Oϵ, and a third is coordinated to the Val88 main-chain carbonyl (Figure 4.7). The M4 side chain ligands are not conserved in ExoIX, Pol1, T5 Fen, T4 Fen/RNaseH or ExoI, consistent with this structural motif being a distinctive feature of FenA.

In order to gauge whether the M4-coordinating amino acids are important for FenA activity, we simultaneously replaced Gln66 and Asp90 with alanine and purified the recombinant Q66A-D90A protein in parallel with WT FenA (Figure 4.8A). Assays of the flap endonuclease activity (Figure 4.8B) and 5′ exonuclease activity (Figure 4.8C) as a function of input FenA showed that the titration profiles of the WT and Q66A-D90A enzymes were superimposable. The flap endonuclease specific activity of the Q66A-D90A mutant was 92% of WT; the 5′ exonuclease-specific activity of the Q66A-D90A mutant was 95% of WT. The benign effect of mutating the M4-coordinating amino acids stands in stark contrast to the severe defects accompanying mutations of the M1, M2 and M3 ligands. We infer that the distinctive M4 metal is not critical for FenA nuclease activity *in vitro*.

DISCUSSION

Mycobacterial FenA exemplifies a bacterial clade of stand-alone nucleic acid phosphodiesterases with flap endonuclease and 5′ exonuclease activities similar to those of the 5′ exonuclease domain of bacterial DNA Pol1, to which FenA is structurally homologous. The crystal structures of FenA reported here, together with the results of a structure-guided mutagenesis, firmly establish that FenA assimilates three manganese ions (its preferred metal cofactor) in

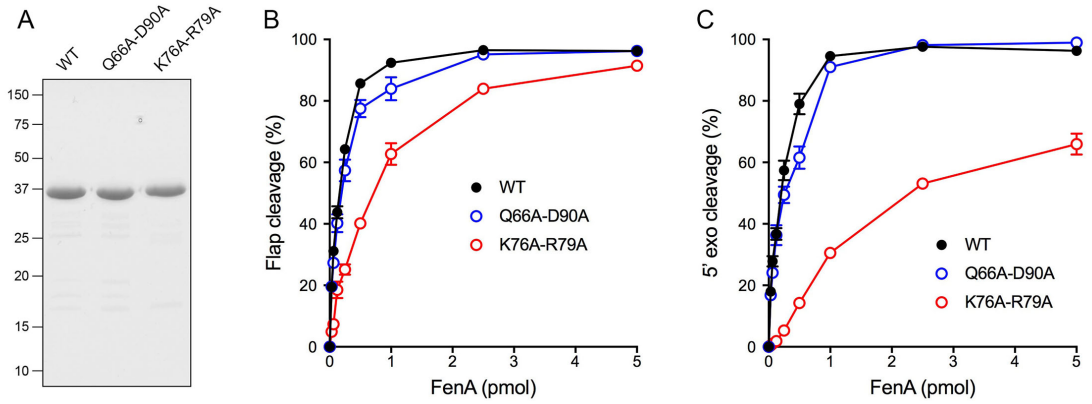


Figure 4.8 Effect of mutations of phosphate-binding and M4-binding amino acids on FenA activity. (A) Aliquots (5 µg) of the preparations of wild-type (WT) FenA and the indicated double alanine mutants were analyzed by SDS-PAGE. The Coomassie blue-stained gel is shown. The positions and sizes (in kilodaltons) of marker polypeptides are indicated on the left. (B and C) Mutational effects on flap endonuclease (B) and 5' exonuclease activity (C). Reaction mixtures (10 µl) containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM MnCl2, 1 mM DTT, and 1 pmol (100 nM) of either 32P-labeled 10-mer flap-nicked duplex DNA (panel B) or 32P-labeled nicked duplex DNA (panel C), and increasing amounts of wild-type (WT) or mutant FenA were incubated for 30 min at 37°C. The reaction products were analyzed by urea-PAGE and visualized and quantified by scanning the gel with a phosphorimager. The extents of cleavage of the substrates are plotted as a function of input FenA. Each datum is the average of three independent titration experiments ± SEM.

the active site and that all three metals are essential for FenA nuclease activity. As a three-metal enzyme, FenA resembles bacteriophage T5 Fen, the structure of which has been solved with three magnesium ions in its active site (11), and which makes many of the same atomic contacts to the M1, M2 and M3 metals observed here for FenA. FenA differs from *E. coli* ExoIX, its nearest tertiary structural homolog among stand-alone FEN proteins, insofar as ExoIX lacks an M3 site and relies on two active site metals, at the positions analogous to M1 and M2 in FenA (8). Human FEN1 and ExoI also adhere to a two-metal mechanism with corresponding M1 and M2 sites (16, 20).

The conservation of M1 and M2 in all FEN/FEN-like nucleases implies

that they are direct catalysts of phosphodiester hydrolysis. In that vein, our structure of FenA mutant D208N with its active site populated by the M1 and M2 manganese ions and a phosphate bound between them provides mechanistic insights to FenA catalysis, as follows. Engagement of the phosphate displaces both the M1–M2 bridging water and an M2 water (denoted by asterisks in Figure 4.6, top panel) and replaces them with phosphate oxygens. We propose that the state of the FenA active site in the M1-M2•PO_4 structure is an analog of the product state of the phosphodiesterase reaction, whereby the phosphate mimics the terminal 5′-PO_4 of the cleaved DNA strand in the duplex segment of the flap-nick substrate. Because the water that functions as the nucleophile in the cleavage reaction becomes one of the non-bridging oxygens in the 5′-PO_4 product, this means that either the M1-M2 bridging water or the M2 water is the attacking nucleophile.

A comparison of the FenA M1-M2•PO_4 structure to the structure of the product complex of the flap cleavage reaction of human ExoI, recently solved by the Beese lab (20), makes a strong case for the M2 water as the nucleophile for FenA. Figure 4.9 shows the active sites of the FenA product analog and the *bona fide* ExoI product complex superimposed and then offset vertically. The respective M1 and M2 metal coordination complexes (to Mn^{2+} in FenA and Mg^{2+} in Exo1) are virtually identical and the position and metal contacts of the FenA phosphate are the same as those of the DNA 5′-PO_4 in the ExoI product complex. The 3′-OH of the cleaved flap strand in the Exo1

product complex (i.e. the leaving group in the phosphodiesterase reaction) is engaged directly by the M1 metal, and the O3′ atom occupies the same position as the M1 water in the FenA structure. The key point is that the O3′ leaving group is nearly apical to the 5′ phosphate oxygen atom that is liganded to M2 (O–P–O3′ angle of 165°), signifying that this apical 5′ phosphate oxygen

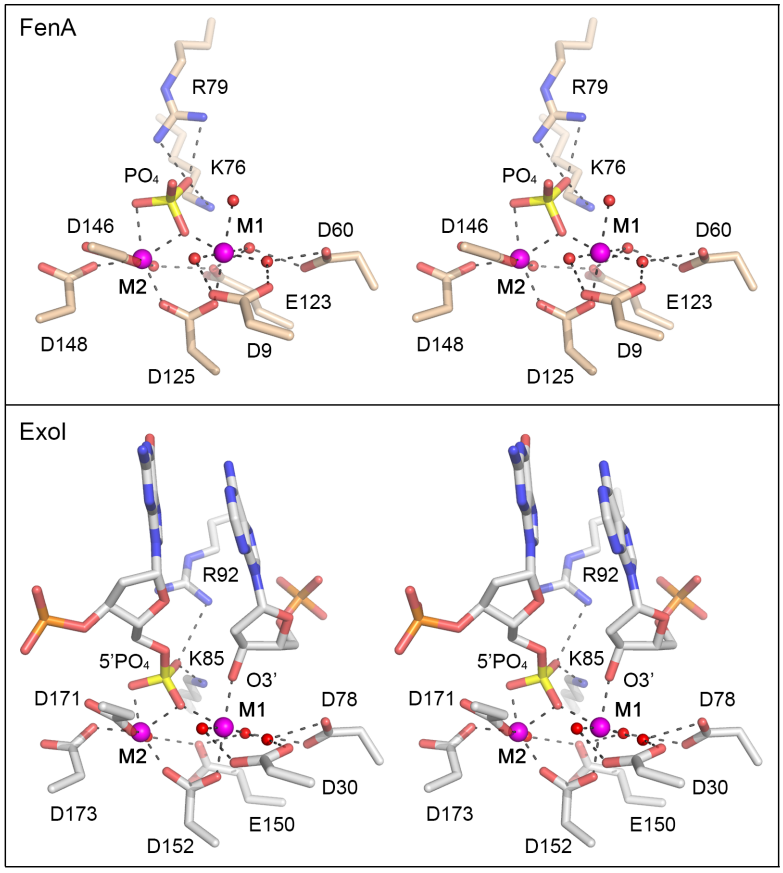


Figure 4.9 Comparison of FenA product analog and ExoI product complex structures. Shown is a stereo view of the active sites of FenA D208N with Mn^{2+} at sites M1 and M2 and a phosphate anion mimetic of the product 5′ phosphate (top panel) superimposed on the active site of human ExoI in a DNA cleavage product complex with Mg^{2+} at sites M1 and M2 (bottom panel; from accession code: 5V0D). The deoxynucleotides at the 5′-PO_4 and 3′-OH ends resulting from DNA cleavage are shown as stick models in the ExoI active site. M1 and M2 are depicted as magenta spheres and waters as red spheres. Atomic contacts are denoted by dashed lines.

derives from the water nucleophile. [By contrast, the 5′-phosphate oxygen atom that bridges M1 and M2 makes an O–P–O3′ angle of 57° in the ExoI product complex.] We surmise that the equivalent M2-bound phosphate oxygen atom in the FenA structure, which makes an O–P-wat angle of 167° to the M1 water (the putative product O3′), derives from the M2-bound water that is the likely nucleophile in the FenA reaction. This mechanism proposed for ExoI and invoked here for FenA differs from a model elaborated for T5 Fen, in which it is hypothesized that the water bridged between M1 and M2 in the T5 Fen active site is the nucleophile that attacks the scissile phosphodiester (11).

The FenA M1 and M2 manganese complexes *per se* could potentially catalyze the chemical steps of phosphodiester hydrolysis: orientation and activation of the nucleophilic water (by M2 and Asp208), stabilization and charge neutralization of the phosphorane transition state (by joint M1–M2 coordination of a trigonal planar phosphate oxygen) and stabilization/expulsion of the O3′ leaving group (by contact to M1). An open question is the degree to which non-metal constituents of the FenA active site drive catalysis, e.g. by interactions of basic side chains with the scissile phosphodiester that might also stabilize the transition state. FenA Lys76 and Arg79 make electrostatic contacts to the same phosphate oxygen in the FenA product-like complex (Figure 4.9) and this phosphate oxygen corresponds to one of the trigonal planar oxygens in the proposed transition state of the ExoI reaction (20). ExoI has an arginine side chain, Arg92, at a position equivalent to Arg79 in FenA and, indeed, ExoI Arg92 contacts the 5′-PO_4 oxygen atom that is not

associated with one of the metals in the product complex (Figure 4.9). ExoI Lys85, though not equivalent to FenA Lys76 with respect to its position in the tertiary structure, also contacts the same 5′-PO_4 oxygen as Arg92 (Figure 4.9). Whereas the ExoI Lys85 and FenA Lys76 Cα atoms to do not overlap, their Nζ atoms that contact the phosphates do coincide when the structures are aligned (Figure 4.9). Single-alanine mutations of Lys85 and Arg92 in human ExoI virtually abolished 5′ exonuclease activity (25). [The lysine and arginine equivalents in human FEN1, Lys93 and Arg100, also contact the scissile phosphate and mutating them singly to alanine reduced flap endonuclease activity by >400-fold (16, 39). Alanine mutation of the lysine equivalent in T5 Fen (Lys83) reduced flap endonuclease activity by 1000-fold (40).] In the case of FenA, our mutational analysis excludes an essential role for Arg79 or Lys76 *per se* in transition state stabilization, insofar as changing Arg79 or Lys76 singly to alanine elicited only an ∼3-fold decrement in flap endonuclease activity and an ∼5-fold reduction in 5′ exonuclease activity (Table 1). To test whether Arg79 and Lys76 might be functionally redundant with respect to FenA activity, we purified a K76A-R79A double mutant of FenA (Figure 4.8A) and assayed its flap endonuclease activity (Figure 4.8B) and 5′ exonuclease activity (Figure 4.8C) as a function of input enzyme. The flap endonuclease specific activity of K76A-R79A was 23% of WT, a value similar to the 29% activity seen for the R79A single mutant. The K76A-R79A 5′ exonuclease-specific activity was 9% of WT, which was half the specific activities of 18–20% of WT that we observed for the K76A and R79A single mutants. These

results indicate only a modest additive effect on 5′ exonuclease function upon loss of both phosphate contacts and show that FenA retains significant flap endonuclease activity in the absence of both contacts. Thus, Arg79 and Lys76 are not redundantly essential *à la* the M1-coordinating residues Asp60 and Glu123.

Although we do not have a structure of FenA bound to a flap DNA substrate, we can surmise from a comparison to the structure of a mutant T5 Fen bound to a flap DNA substrate, where T5 Fen contains M1 and M3 in the active site but lacks M2 (11), that the essential M3 manganese in FenA will contact the DNA phosphate group immediately 3′ of the scissile phosphodiester. An outstanding question anent FenA is how the enzyme engages a 5′ flap. The FenA protein structure does not have the helical arch module found in many other FEN/FEN-like enzymes. Instead it has a distinctive M4 loop in lieu of the first α helix of the helical arch. We envision two possible scenarios: (i) FenA accommodates the DNA 5′ flap in a unique manner involving the M4 loop; or (ii) the structure of the M4 loop is remodeled upon binding to DNA into a helical arch akin to that seen in T5 Fen, FEN1 and Exo1. The plausibility of the former model is suggested by a surface view of FenA, oriented in Figure 4.10A to highlight a through and through hole in the enzyme, the ‘roof’ of which is formed by the M4 loop as it reaches in an arch-like fashion to join the α4 helix (Figure 4.10B). This aperture is along the path predicted to be taken by the 5′ flap strand and could potentially flex to allow threading of the 5′ flap though the hole.

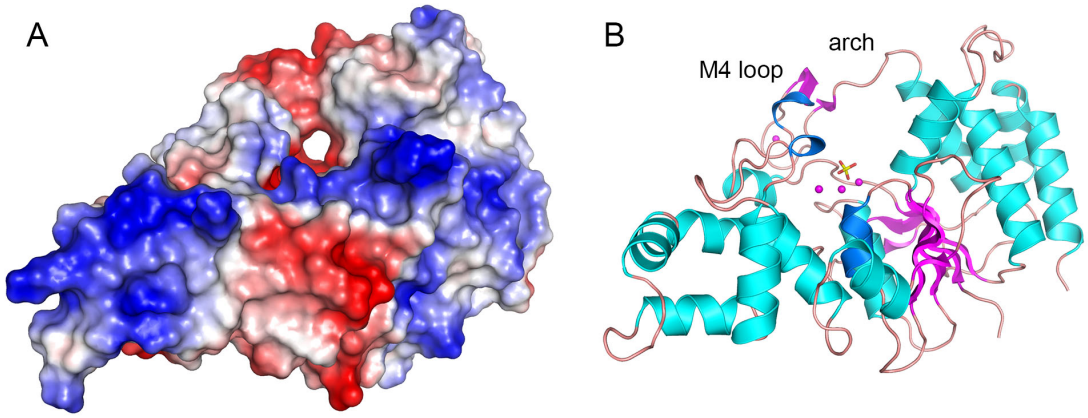


Figure 4.10 Surface model of FenA highlights a hole that might accommodate a 5' flap strand. Surface model of FenA highlights a hole that might accomodate a 5′ flap strand. Panel **A** shows a surface electrostatic model of WT FenA oriented so as to feature a through-and-through hole near the active site along the path predicted for the 5′ flap strand. Panel **B** shows a cartoon model of the tertiary structure in the same position used to generate the surface view. Manganese ions (magenta spheres) from the WT FenA structure, and a phosphate (stick model) imported from the D208N structure, are shown as active site landmarks in panel B.

EXPERIMENTAL PROCEDURES

Recombinant *M. smegmatis* FenA

We produced wild-type (WT) and mutant FenA proteins in *E. coli* as His_{10}Smt3 fusions and isolated them from soluble extracts of 3-liter bacterial cultures by nickel-agarose chromatography as described previously (7). The His_{10}Smt3 tags were removed with the Smt3-specific protease Ulp1, and native FenA proteins were separated from the tag by a second round of Ni-agarose chromatography. Proteins for crystallography were further purified by Superdex-200 gel filtration as described (7). Selenomethionine-(SeMet)-substituted WT FenA was produced as follows. Plasmid pET28-His_{10}Smt3-

FenA was transformed into *E. coli* B834(DE3), a methionine auxotroph. A single transformant was inoculated into 10 ml of LB medium containing 60 μg/ml kanamycin and incubated for 7 h at 37°C. The bacteria were harvested by centrifugation and then resuspended in 10 ml of complete LeMaster medium containing 60 μg/ml kanamycin. This step was repeated twice and the cells were finally resuspended in 400 ml of complete LeMaster medium (with kanamycin) containing 25 μg/ml SeMet (L-enantiomer). After overnight incubation, the culture volume was increased to 4 liters with fresh LeMaster medium containing kanamycin and SeMet, and growth was continued at 37°C with constant shaking until the A_{600} reached 0.4. The culture was adjusted to 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 2% ethanol and incubated for 16 h at 17°C with continuous shaking. SeMet-FenA was then purified following the procedures described above for native FenA. The protein concentrations of the Superdex FenA preparations were determined by using the BioRad dye reagent with bovine serum albumin as the standard.

FenA-Ala and FenA-Asn missense mutations were introduced into the pET28b-His_{10}Smt3-FenA expression plasmid by two-step overlap extension polymerase chain reaction (PCR) or by using the QuikChange method with NEB Phusion polymerase. The plasmid inserts were sequenced to verify that no unwanted coding changes were introduced during PCR amplification and cloning. WT and mutant FenA proteins used for nuclease assays were purified through the second Ni-agarose step described above, then concentrated by centrifugal ultrafiltration and stored at −80°C. Total protein concentrations of

the concentrated FenA preparations were measured by using the Bio-Rad dye reagent. The concentrations of the FenA polypeptides were determined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of 5 µg aliquots of each protein preparation in parallel with 1.25, 2.5, 5, 10 µg of bovine serum albumin (BSA) standards. The gels were stained with Coomassie blue, and the extents of dye binding to FenA and BSA polypeptides were quantified using Chemidoc XRS imager and ImageJ. The concentrations of the FenA proteins were calculated by interpolation to the BSA standard curve.

Crystallization, diffraction data collection and structure determination

FenA crystals were grown by sitting drop vapor diffusion at room temperature. A 1 µl aliquot of a solution of 4 mg/ml FenA and 10 mM $MnCl_2$ in buffer containing 20 mM Tris–HCl, pH 8.0, 150 mM NaCl, 2 mM DTT and 10% glycerol was mixed with 1 µl of precipitant solution and equilibrated in a 96-well plate against 150 µl of reservoir solution. The SeMet-WT FenA precipitant and reservoir solution solutions were 0.4 M NH_4Cl, 28% PEG3350. The D125N precipitant solution was 0.2 M NaCl, 20% PEG3350 and the reservoir solution was 0.5 M LiCl. The D148N precipitant solution was 0.2 M LiCl, 20% PEG3350 and the reservoir solution was 0.5 M LiCl. The D208N precipitant solution was 0.2 M ammonium acetate, 20% PEG3350 and the reservoir solution was 0.5 M LiCl. The crystals were cryoprotected in a solution of 70% sucrose, 50 mM $MnCl_2$ (for SeMet-WT, D125N and D148N) or 70% sucrose (D208N) before being flash-frozen in liquid nitrogen. X-ray diffraction data

were collected from single crystals at APS beamline 24ID-C for SeMet-WT and D208N and at beamline 24ID-E for D148N and D125N. The SeMet-WT, D208N and D148N data were indexed with HKL2000 (32); D125N data were indexed with XDS (33). The cutoff criteria for X-ray diffraction data were CC(1/2) ≥0.6 and R_{pim} < 0.5 in the highest resolution shell. All FenA crystals were in space group $P2_1$. Phases were obtained using single-wavelength anomalous dispersion (SAD) data from a single crystal of SeMet-WT FenA, processed with SHELX (34) to locate two Se sites and one Mn site. After solvent flattening in Phenix (35), interactive model building was done using O (36). Refinement was accomplished with Phenix. Data collection and refinement statistics are compiled in Table 4.1.

Nuclease assays

The nick and 5′ flap-nick substrates, prepared as described (7), were composed of an 18-mer DNA_{OH} strand and a 5′ ^{32}P-labeled 18-mer or 28-mer pDNA strand annealed to a 36-mer DNA template strand. Nuclease reaction mixtures (10 µl) containing 20 mM Tris–HCl, pH 8.0, 50 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, 1 pmol (0.1 µM) ^{32}P-labeled DNA substrate and WT or mutant FenA (as specified in the figure legends) were incubated at 37°C for 30 min. The reactions were quenched by adding 10 µl of 90% formamide, 50 mM ethylenediaminetetraacetic acid (EDTA). The samples were heated at 95°C for 5 min and then analyzed by electrophoresis through a 40-cm 18% polyacrylamide gel containing 7.5 M urea in 44.5 mM Tris-borate, pH 8.3, 1 mM EDTA. The products were visualized and quantified by scanning the gel

using a Typhoon FLA7000 imager. The extents of DNA cleavage were plotted as a function of input enzyme, with each datum being the average of at least three separate enzyme titration experiments (±SEM). The specific activities of the FenA preparations were calculated from the slopes of the titration curves in the linear range of FenA dependence, as implemented by linear regression in Prism. The standard errors of the linear regression fits for FenA proteins that had detectable activity varied from 2 to 16% of the calculated specific activity values. The specific activities of the mutant FenA enzymes were normalized to that of WT FenA (defined as 100%).

ACKNOWLEDGEMENTS

Ayala Carl set up wild-type and mutant FenA proteins for crystallization. Yehuda Goldgur collected diffraction data and solved the FenA crystal structures.

REFERENCES

1. Finger, L.D., Atack, J.M., Tsutakawa, S., Classen, S., Tainer, J., Grasby, J. and Shen, B. (2012) The wonders of flap endonucleases: structure, function, mechanism and regulation. *Subcell. Biochem*., **62**, 301-326.

2. Tomlinson, C.G., Atack, J.M., Chapados, B., Tainer, J.A. and Grasby, J.A. (2010) Substrate recognition and catalysis by flap endonucleases and related enzymes. *Biochem. Soc. Trans*., **38**, 433-437.

3. Lyamichev, V., Brow, M.A. and Dahlberg, J.E. (1993) Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science*, **260**, 778-783.

4. Xu, Y., Derbyshire, V., Ng, K., Sun, X.C., Grindley, N.D. and Joyce, C.M. (1997) Biochemical and mutational studies of the 5'-3' exonuclease of DNA polymerase I of *Escherichia coli*. J*. Mol. Biol*., **268**, 284-302.

5. Lyamichev, V., Brow, M.A., Varvel, V.E. and Dahlberg, J.E. (1999) Comparison of the 5' nuclease activities of *Taq* DNA polymerase and its isolated nuclease domain. *Proc. Natl. Acad. Sci. USA*, **96**, 6143-6148.

6. Mizrahi, V. and Huberts, P. (1996) Deoxy- and dideoxynucleotide discrimination and identification of critical 5' nuclease domain residues of the DNA polymerase I from *Mycobacterium tuberculosis*. *Nucleic Acids Res*., **24**, 4845-4852.

7. Uson, M.L., Ghosh, S. and Shuman, S. (2017) The DNA repair repertoire of *Mycobacterium smegmatis* FenA includes the incision of DNA 5' flaps and the removal of 5' adenylylated products of aborted nick ligation. *J. Bacteriol*., **199**, e00304-17.

8. Anstey-Gilbert, C.S., Hemsworth, G.R., Flemming, C.S., Hodskinson, M.R., Zhang, J., Sedelnikova, S.E., Stillman, T.J., Sayers, J.R.. and Artymiuk, P.J. (2013) The structure of *Escherichia coli* ExoIX – implications for DNA binding and catalysis in flap endonucleases. *Nucleic Acids Res*., **41**, 8357-8367.

9. Fukushima, S., Itaya, M., Kato, H., Ogasawara, N. and Yoshikawa, H. (2007) Reassessment of the in vivo function of DNA polymerase I and RNase H in bacterial cell growth. *J. Bacteriol*., **189**, 8575-8583.

10. Ceska, T.A., Sayers, J.R., Stier, G. and Suck, D. (1996) A helical arch allowing single-stranded DNA to thread through T5 5'-exonuclease. *Nature*, **382**, 90-93.

11. AlMalki, F.A., Flemming, C.S., Zhang, J., Feng, M., Sedelnikova, S.E., Ceska, T., Rafferty, J.B., Sayers, J.R. and Artymiuk, P.J. (2016) Direct observation of DNA threading in flap endonuclease complexes. *Nat. Struct. Mol. Biol*., **23**, 640-646.

12. Mueser, T.C., Nossal, N.G. and Hyde, C.C. (1996) Structure of bacteriophage T4 RNase H, a 5' to 3' RNA-DNA and DNA-DNA exonuclease with sequence similarity to the RAD2 family of eukaryotic proteins. *Cell*, **85**, 1101-1112.

13. Devos, J.M., Tomanicek, S.J., Jones, C.E., Nossal, N.G. and Mueser, T.C. (2007) Crystal structure of bacteriophage T4 5' nuclease in complex with a branched DNA reveals how flap endonuclease-1 family nucleases bind their substrates. *J. Biol. Chem*., **282**, 31713-31724.

14. Harrington, J.J. and Lieber, M.R. (1994) The characterization of a mammalian DNA structure-specific endonuclease. *EMBO J*., **13**, 1235-1246.

15. Balakrishnan, L. and Barbara, R.A. (2013) Flap endonuclease 1. *Annu. Rev. Biochem*., **82**, 119-138.

16. Tsutakawa, S.E., Thompson, M.J., Arvai, A.S., Neil, A.J., Shaw, S.J., Algasaier, S.I., Kim, J.C., Finger, L.D., Jardine, E., Gotham, V.J.B. *et al.* (2017) Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. *Nat. Commun*., **8**, 15855.

17. Hwang, K.Y., Baek, K., Kim, H.Y. and Cho, Y. (1998) The crystal structure of flap endonuclease-1 from *Methanococcus jannaschii*. *Nat. Struct. Biol*., **5**, 707-713.

18. Hosfield, D.J., Mol, C.D., Shen, B. and Tainer, J.A. (1998) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell*, **95**, 135-146.

19. Chapados, B.R., Hosfield, D.J., Han, S., Qiu, J., Yelent, B., Shen, B. and Tainer, J.A. (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell*, **116**, 39-50.

20. Shi, Y., Hellinga, H.W. and Beese, L.S. (2017) Interplay of catalysis, fidelity, threading, and processivity in the exo- and endonucleolytic reactions of human exonuclease I. *Proc. Natl. Acad. Sci. USA*, **114**, 6010-6015.

21. Mietus, M., Nowak, E., Jaciuk, M., Kustosz, P., Studnicka, J. and Nowotny, M. (2014) Crystal structure of the catalytic core of Rad2: insights into the mechanism of substrate binding. *Nucleic Acids Res*., **32**, 10762-10775.

22. Liu, Y., Freeman, A.D.J., Declais, A.C., Wilson, T.J., Gartner, A. and Lilley, D.M.J. (2015) Crystal structue of a eukaryotic GEN1 resolving enzyme bound to DNA. *Cell Reports*, **13**, 2565-2575.

23. Lee, S.H., Princz, L.N., Klügel, M.G., Habermann, B., Pfander, B. and Biertümpfel, C. (2015) Human Holliday junction resolvase GEN1 uses as chromodomain for efficient DNA recognition and cleavage. *eLife*, **4**, e12256.

24. Kim, Y., Eom, S.H., Wang, J., Lee, D.S., Suh, S.W. and Steitz, T.A. (1995) Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature*, **376**, 612-616.

25. Orans, J., McSweeney, E.A., Iyer, R.R., Hast, M.A., Hellinga, H.W., Modrich, P. and Beese, L. (2011) Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell*, **145**, 212-223.

26. Patel, N., Atack, J.M., Finger, L.D., Exell, J.C., Thompson, P., Tsutakawa, S., Tainer, J.A., Williams, D.M. and Grasby, J.A. (2012) Flap endonucleases pass 5'-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5'-ends. *Nucleic Acids Res*., **40**, 4507-4519.

27. Chauleau, M, and Shuman, S. (2016) Kinetic mechanism and fidelity of *Escherichia coli* NAD^+-dependent DNA ligase (LigA). *Nucleic Acids Res*., **44**, 2298-2309.

28. Zhu, H. and Shuman, S. (2008) Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-OH monoribonucleotide. *J. Biol. Chem*., **283**, 8331-8339.

29. Gong, C., Martins, A., Bongiorno, P., Glickman, M. and Shuman, S. (2004) Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *J. Biol. Chem*. **279**, 20594-20606.

30. Zhu, H. and Shuman, S. (2007) Characterization of *Agrobacterium tumefaciens* DNA ligases C and D. *Nucleic Acids Res*., **35**, 3631-3645.

31. 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.

32. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.*, **276**, 307-326.

33. Kabsch, W. (2010) XDS. *Acta Crystallograph*., **D66**, 125-132.

34. Sheldrick, G.M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Cryst.*, **D66**, 479-485.

35. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.*, **D66**, 213-221.

36. Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.*, **A47**, 110-119.

37. Holm, L., Kaariainen, S., Rosenstrom, P. and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**, 1780-2781.

38. Sakurai, S., Kitano, K., Yamguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H. and Hakoshima, T. (2005) Structural basis for recruitment of human flap endonuclease I to PCNA. *EMBO J.*, **24**, 683-693.

39. Tsutakawa, S.E., Classen, S., Chapados, B.R., Arvai, A.S., Finger, L.D., Guenther, G., Tomlinson, C.G., Thompson, P., Sarker, A.H., Shen, B. *et al.* (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell*, **145**, 198-211.

40. Pickering, T.J., Garforth, S.J., Thorpe, S.J., Sayers, J.R. and Grasby, J.A. (1999) A single cleavage assay for T5 5'→3' exonuclease: determination of the catalytic parameters for wild-type and mutant proteins. *Nucleic Acids Res.*, **27**, 730-735.