JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

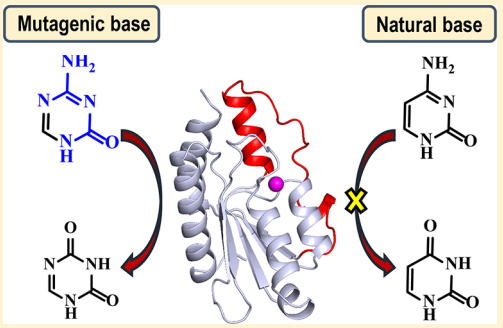
Selective Deamination of Mutagens by a Mycobacterial Enzyme

Vandana Gaded and Ruchi Anand*

Department of Chemistry, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

Supporting Information

ABSTRACT: Structure-based methods are powerful tools that are being exploited to unravel new functions with therapeutic advantage. Here, we report the discovery of a new class of deaminases, predominantly found in mycobacterial species that act on the commercially important s-triazine class of compounds. The enzyme Msd from *Mycobacterium smegmatis* was taken as a representative candidate from an evolutionarily conserved subgroup that possesses high density of *Mycobacterium* deaminases. Biochemical investigation reveals that Msd specifically acts on mutagenic nucleobases such as 5-azacytosine and isoguanine and does not accept natural bases as substrates. Determination of the X-ray structure of Msd to a resolution of 1.9 Å shows that Msd has fine-tuned its active site such that it is a hybrid of a cytosine as well as a guanine deaminase, thereby conferring Msd the ability to expand its repertoire to both purine and pyrimidine-like mutagens. Mapping of active site residues along with X-ray structures with a series of triazine analogues aids in deciphering the mechanism by which Msd proofreads the base milieu for mutagens. The genome location of the enzyme reveals that Msd is part of a conserved cluster that confers the organism with innate resistance toward select xenobiotics by triggering their efflux.



INTRODUCTION

Drug resistance is one of the most challenging problems of the 21st century.[1] It is projected that if no alternative therapies are undertaken by 2050, more than 300 million deaths will be attributed to drug resistance.[2] Particularly, *Mycobacterium tuberculosis*, the causative agent of tuberculosis, infects one-third of the human population and has turned out to be the world's most deadly pathogen.[3] Due to excessive drug abuse, most of the current drugs such as fluoroquinolones and rifampin, which target the replication and transcription machineries, and others, such as isoniazid, that cripple its cell wall have failed.[4] In addition, this pathogenic mycobacterial species has the exceptional ability to persist and adapt to the changing environment.[4,5] Therefore, in order to develop new therapies, it is necessary to focus on other pathways that are unique to the pathogenic species. The nucleobase catabolic pathway is one such pathway that varies among eukaryotes and prokaryotes. Especially nucleobase deaminases, an integral part of this pathway, are evolutionarily divergent between human and bacterial species.[6] Hence, these enzymes pose a great target for the design of future drug scaffolds. In fact because of this structural disparity, bacterial cytosine deaminases have been used as successful enzyme–prodrug systems for cancer treatment.[6,7]

Nucleobase deaminases are categorized into two major superfamilies, the amidohydrolase (AHS) and cytidine deaminase-like (CDA) superfamilies, based on the fold they possess.[8] While AHS harbors a TIM barrel fold, CDAs have an $\alpha/\beta/\alpha$ sandwich core fold. Both these deaminases employ a metal such as zinc, iron, or cobalt to hydrolyze the amino group of the base.[8,9] CDA superfamily proteins encompass a broad spectrum of substrates ranging from simple bases such as cytosine and guanine to whole tRNA molecules.[7,10,11] To avoid introduction of mutagenic bases into the regenerated DNA and RNA pool in the cell, nature has conferred these deaminases with specificity of function.[12−14]

In recent years, due to the advent of genome sequencing projects, there has been a tremendous increase in the number of curated genes.[15] This, along with a surge in development of bioinformatics tools, has initiated efficient curation of the available enzymes into common orthologous groups (COGs).[16,17] Generally, enzymes that fall in a particular COG have related reactions or functions, and therefore this curation strategy provides clues into the probable function of unassigned proteins.[17] Moreover, conjunction of structural, biochemical, and systems biology analysis can help unearth enzymes with new functions as well as novel pathways unique to a particular organism. Such a strategy has been successfully employed to discover the function of several AHS family deaminases.[12,18,19]

It is rather surprising that nucleobase deaminases have not been well characterized in the *Mycobacterium* genus.[20] Creation of a sequence similarity network[21] of cog0590 that contains a high concentration of CDA superfamily deaminases illustrated that Mycobacterial enzymes (inclusive of *M. tuberculosis*) separate out into a evolutionarily distinct, novel subgroup (Figure 1). Homologue organisms of *M. tuberculosis* have earlier helped as useful surrogates in drug discovery; hence Msmeg3575 (abbreviated as Msd) from *Mycobacterium smegmatis* was selected as a representative enzyme. Extensive biochemical and structural evaluations in conjunction with mapping of the active

Received: May 14, 2017
Published: July 14, 2017

DOI: 10.1021/jacs.7b04967
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

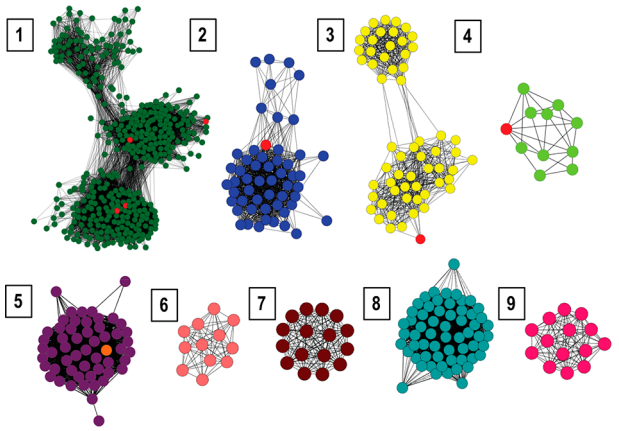


Figure 1. Cytoscape representation of the sequence similarity network of cog0590 at an *e* value cutoff of 10^{-40}. In each group, the nodes represent the proteins and the edges represent the BLASTP linkages. The representative crystal structures are colored in red in each group, and the groups are named based on the characterized protein present in each group (group 1, tRNA adenosine deaminases; groups 2 and 4, guanine deaminases; group 3, cytosine deaminase); other groups remain unannotated. The protein under study, Msmeg3575/Msd, falls in an unknown group 5, which is colored orange.

site residues were performed to establish structure−activity relationships. Results revealed that enzymes of this subgroup not only are unique to *Mycobacterium* genus but also possess the exceptional ability to specifically catalyze the deamination of select mutagenic bases. This study opens the door toward exploration of the mechanism of inherent resistance to mutagens possessed by this subgroup of pathogens.

RESULTS AND DISCUSSION

Crystal Structure of Msd. cog0590 is one of the largest CDA superfamily groups and contains a diversity of enzymes that deaminate cytosine, guanine, and adenosine at the wobble position of tRNA. Due to low sequence identity of Msd with the known deaminases, its native structure (PDB ID: 5XKO) was determined by the zinc single-wavelength anomalous dispersion (Zn-SAD) method to a resolution of 1.89 Å (Figure 2A). The aim of the structure solution was to employ this information to predict its function (data and refinement statistics in Table 1). The structure revealed that Msd adopts the standard CDA $\alpha/\beta/\alpha$ sandwich fold with the central five-stranded mixed β sheet order of 2, 1, 3, 4, and 5, wherein the β1 strand runs in an antiparallel direction in comparison to the remaining four strands.[6] Helices αB and αC and their attendant loops coordinate the catalytic zinc ion, forming tetrahedral coordination with H56 (2.1 Å), C86 (2.3 Å), C89 (2.3 Å), and OH group (1.9 Å) of the catalytic water. A comparison of the structure of Msd with other members revealed that although all basic CDA features remain the same, there are marked variations (Figure 2B, C, and D). For instance, the extra C-terminal helix that renders yeast cytosine deaminase (yCD) active site compact and stabilizes the cytosine moiety is absent in Msd,[6] nor, like NE0047 guanine deaminase (GD), does it possess a C-terminal flap that serves as a gate, during the course of the reaction.[13] Instead, Msd harbors a

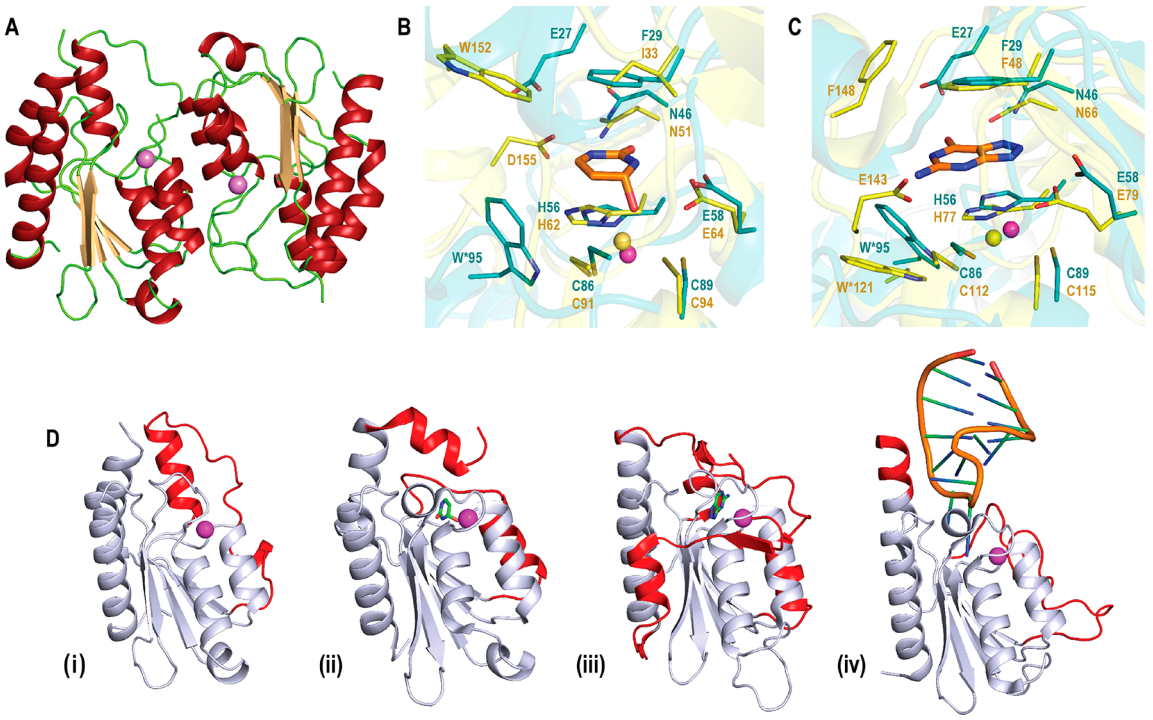


Figure 2. Structural analysis of Msd. (A) Cartoon representation of Msd dimer with α-helices in red, β-strands in light orange, and loop regions in green. (B) Active site superposition of Msd and dihydrouracil-bound yeast cytosine deaminase (PDB 1UAQ). (C) Active site superposition of Msd and NE0047−8-azaguanine complex (PDB 4HRQ). Carbon atoms of Msd are colored cyan, while those in 1UAQ and 4HRQ in yellow. Oxygen and nitrogen atoms are in red and blue, respectively. The zinc is colored magenta in Msd, whereas it is yellow in 1UAQ and 4HRQ. The carbon atoms of the bound ligand are colored orange in both the complexes, and * indicates a residue from the adjacent subunit. (D) Comparison of Msd with the representative members of cog0590 of the CDA superfamily. (i) Msd. (ii) Yeast cytosine deaminase (PDB- 1UAQ). (iii) *Nitrosomonas europaea* guanine deaminase (PDB-4HRQ). (iv) tRNA adenosine deaminase (PDB-2B3J). The CDA fold is highlighted in light blue, and the unique structural element in red. In all the structures, the carbon atoms of the bound ligand are colored green and zinc is depicted as a magenta sphere.

DOI: 10.1021/jacs.7b04967
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Table 1. Data and Refinement Statistics

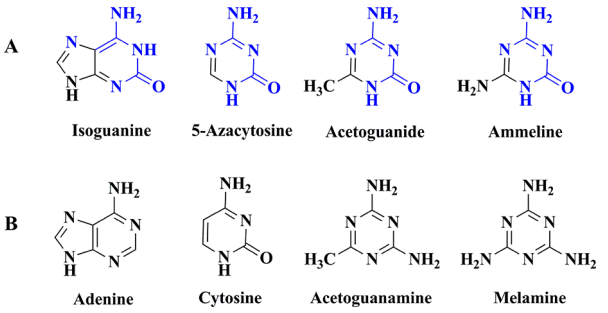
protein	Msd native (PDB ID: 5XKO)	Msd−ammeline complex (PDB ID: 5XKQ)	Msd−5-azacytosine complex (PDB ID: 5XKP)	Msd− benzoguanamine complex (PDB ID: 5XKR)
space group	$P2_12_12$	$P2_1$	$P2_1$	$P2_1$
resolution (Å)	1.89	2.68	3.0	1.38
multiplicity	39.8 (38.6)	4.1 (4.0)	5.0 (4.0)	3.9 (3.7)
completeness (%)[a]	99.8 (100)	99.1 (100)	99.1 (90.5)	98.3 (95.5)
R_{sym} (%)	7.3 (38.1)	9.4 (47.4)	10.6 (41.6)	4.5 (23.1)
I/σ	73 (16.1)	12.9 (2.1)	10.6 (2.5)	25.7 (4.1)
total no. of reflns	1 193 684	78 758	69 504	563 664
no. of unique reflns	29 961	19 362	13 940	145 157
Refinement				
resolution range (Å)	47.8−1.89	109.4−2.70	62−3.0	108−1.38
no. of reflns total	27 732	17 737	12 845	136 748
no. of reflns test set	1988	832	653	7144
R_{work}/R_{free} (%)	13.8/17.8	19.3/25.3	17.9/24.2	12.7/15.2
no. of atoms				
total	2721	4695	4760	6382
protein	2379	4605	4664	4993
ligands	44	52	60	120
ion	2	4	4	4
water	296	34	32	1265
rmsd				
bond lengths (Å)	0.021	0.012	0.012	0.017
bond angles (deg)	1.89	1.45	1.56	1.82
Ramachandran plot				
most favored region (%)	97.73	98.18	96.76	97.6
additionally allowed region (%)	2.27	1.82	3.07	2.31
outliers (%)	0.0	0.0	0.16	0.0

[a]Values in parentheses represent the data in the highest-resolution shell.

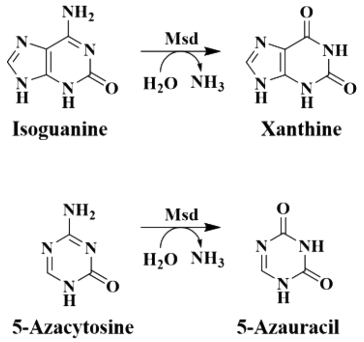
topologically as well as structurally unique helix loop insertion (αD) that forms a flap above the catalytic site (Figure 2D). Attempts to superimpose its active site residues with other known deaminases revealed that the active site architecture is not in congruence with any of the other known deaminases in this COG (Figure 2B and C). For example, the positioning of the two catalytically essential glutamic acid residues in GD as well as the aspartic acid in the case of yCD is not in correspondence with similarly charged residues in Msd (Figure 2B and C).[6,13] Therefore, it became difficult to ascertain its function via structural analysis. Interestingly, the catalytic site of Msd consists of residues from the adjacent monomer that interweaves and forms part of the active site cavity. Another noticeable feature of Msd is that it possesses a relatively larger active site pocket (area 460.9 Å^2 and volume 561.5 Å^3 calculated using the CASTp server),[22] which is 3 times that of yCD and double that of GD.

Substrate Selectivity of Msd. Since structure solution was not helpful in immediately ascertaining its function, a combination of docking (Supporting Information Tables S3 and S4) and screening strategies were simultaneously undertaken. After rigorous screening, it was concluded that Msd does not respond to any of the natural bases, nucleosides, or nucleotides. The only purine, which underwent deamination, was the mutagenic base isoguanine. Msd also showed deaminase activity toward select triazine compounds such as acetoguanide, 5-azacytosine, and ammeline (Schemes 1 and 2). The kinetic constants for the deamination of isoguanine and triazine compounds are listed in Table 2. This was a surprising discovery, as Msd categorically does not deaminate any natural base and is designed to accept only mutagenic bases. Earlier studies by Raushel and co-workers have shown that certain AHS super-

Scheme 1. (A) Substrates Catalyzed by Msd with the Basic Prerequisite Chemical Scaffold Highlighted in Blue; (B) Compounds That Do Not Exhibit Deamination by Msd



Scheme 2. Representative Reactions Catalyzed by Msd



DOI: 10.1021/jacs.7b04967
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Table 2. Kinetic Constants of Msd for Different Substrates

substrate	K_m (μM)	k_{cat} (s^{-1})	$[k_{cat}/K_m$ (M^{-1} s^{-1})]/10^3
acetoguanide	755 ± 13	11.14 ± 0.07	14.70 ± 0.03
5-azacytosine	838 ± 26	2.61 ± 0.03	3.07 ± 0.06
ammeline	473 ± 12	0.86 ± 0.01	1.81 ± 0.02
isoguanine	518 ± 6	0.30 ± 0.02	0.60 ± 0.01

family deaminases can not only catalyze deamination of natural bases but also accept mutagenic and epigenetic modified bases as substrates. For instance, *E. coli* cytosine deaminase was found to deaminate both isoguanine and cytosine, whereas *P. aeruginosa* Pa0142 exhibited a broad substrate profile and acts on a spectrum of nucleobases inclusive of 8-oxoguanine and isocytosine.[12,23,24] Unlike the trend displayed by the AHS superfamily members, in the CDA superfamily, Msd appears to be more stringent in proofreading for mutagens.

On close examination of the chemical structure of the substrates, it became apparent that all catalyzed compounds by Msd had a common scaffold and s-triazines exhibited conjugation chemistry similar to isoguanine (Scheme 1). The substrates possess strategically placed keto and amino functional groups, which are situated between two electron-withdrawing nitrogen atoms. This scaffold is also present in other nucleobases that were tested such as cytosine, isocytosine, 5-methylcytosine, guanine, 8-oxoguanine, and pterin. However, they were inert to Msd, as they lack the additional nitrogen atom present exclusively in the *s*-triazine (azapyrimidine) scaffold that further partakes in facilitating efficient reaction.

The substrates of Msd, *s*-triazine (1,3,5-triazine) compounds, are an interesting and the oldest class of compounds with multiple implications due to their functionalization and structural imitation of natural nucleobases.[25] Among the triazines, 5-azacytosine was one of the best compounds for Msd, and the other substrates were acetoguanide and ammeline. These compounds are commercially important. For example, 5-azacytosine and its derivatives are reported to have several applications in medicine and are used as chemotherapeutic agents,[25,26] whereas acetoguanide is a major degradation product of several *s*-triazines, which are an extensively used class of herbicides.[27] Moreover, the precursor of ammeline, melamine, is a prominent food adulterant that has been associated with several pet and infant deaths and is considered an environmental hazard.[28] It appears that the *Mycobacterium* genus and other soil bacteria in group 5 have developed specific machineries to combat against these mutagens, thereby becoming resistant to these causative agents.

To better understand the organism's need for this activity, the region of the genome where this enzyme is situated was explored. Investigation reveals that in several organisms inclusive of other soil bacteria, which are also part of this group, a similar genome organization is present with analogous genes flanking Msd homologues. Analyses of the surrounding gene cluster shows that the Msd gene is adjacent to membrane-associated proteins and putative tetracycline regulators, a combination of genes that are generally involved in formation of efflux pumps (Figure 3). For instance, the Msd gene cluster includes genes such as Msmeg3570 (fusaric acid resistance) and Msmeg3584 (drug exporter of the resistance−nodulation−division family), which are a part of the efflux pump family and are known to be involved in the expulsion of toxic compounds.[29] This cluster also consists of genes such as the tetratricopeptide domain and trehalose *o*-mycolyltransferase, both of which play a central role in outer

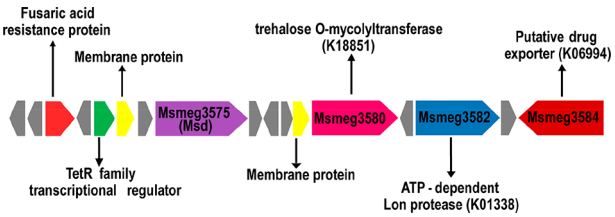


Figure 3. Schematic diagram of the Msd gene cluster. For the Msmeg3580, 3582, and 3584 genes, functions are annotated as per Kegg Orthology.

membrane assembly and pathogenesis in mycobacterium.[30,31] Overall, it appears that the Msd gene is clustered such that it could be part of a proofreading mechanism, which scans the nucleobase pool for xenobiotics and common mutagenic base analogues, thereby preventing their incorporation into the genetic code.

Structural Basis of Substrate Specificity. To unravel the structural basis of the mechanism of action, crystal structures were solved with a subset of substrates. As deamination results in replacement of an amino group by a keto group, using crystallography it is challenging to determine whether a bound ligand species is the reactant or the product. Therefore, to predominantly trap the ligand in their reactant form, all soaks were performed for short duration (~1−2 min) using excess ligand concentration. Crystal structures with ammeline (PDB ID: 5XKQ) and 5-azacytosine (PDB ID: 5XKP) bound to Msd reveal that both ligands are stabilized in the active site cleft via a π-stacking interaction and are sandwiched between F29 and H56. The amine group to be deaminated is within hydrogen-bonding distance of the catalytic water molecule that is also hydrogen bonded to the highly conserved glutamic acid residue. It is this glutamic acid (E58) that, along with the tetrahedrally coordinated zinc atom, activates the water molecule for attack on the carbon atom adjacent to the amino group to be deaminated (Figure 4A and B). In addition, residues N46 and

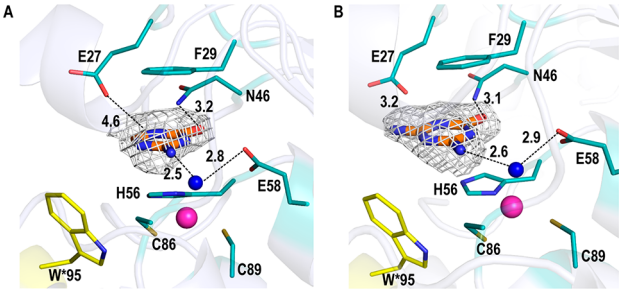


Figure 4. Active site representation of Msd in complex with 5-azacytosine (A) and ammeline (B). Carbon atoms of the ligand binding residues are shown in cyan except W*95, where * indicates a residue from the adjacent subunit. The zinc atom is shown in magenta, and the zinc-coordinated water molecule is shown as a blue sphere. Carbon atoms of both the ligands are shown in orange. Electron density (F_o − F_c) maps are contoured at 3σ for both ligands. Oxygen and nitrogen atoms are in red and blue, respectively.

E27 also appear to be key players important for positioning of the substrate and in stabilizing the buildup of negative charge during the course of the reaction (Figure 4A and B). Overall, E58 initiates the reaction by abstracting a proton from the zinc-bound water. This leads to the formation of a hydroxide ion that undergoes a nucleophilic attack on the C4 carbon atom, resulting

DOI: 10.1021/jacs.7b04967
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

in the formation of a tetrahedral intermediate that collapses to subsequently release ammonia.[6,32]

A pertinent question to address was the ability of Msd to screen out natural bases. For example, it exhibits deaminase activity toward 5-azacytosine; however, Msd is completely inert to cytosine. Similarly, while Msd hydrolyzes isoguanine, it exhibits no activity toward adenine or guanine bases. In the azapyrimidine scaffold, the presence of the nitrogen atom at the 5-position renders the C4 carbon atom more electrophilic. This facilitates the nucleophilic addition at this center, which is the rate-limiting step in nucleophilic aromatic substitution reactions. Moreover, the electrophilic character of the tetrahedral intermediate is enhanced by the presence of a keto group at position 2. Due to the absence of the N5 nitrogen atom in the natural base cytosine, the C4 atom is relatively less electrophilic, and as a consequence, Msd is unable to activate cytosine for effective deamination. For isoguanine, the presence of the N7 atom in addition to the keto group at position 2 of the purine scaffold together facilitates the first step of deamination. However, since N7 of purine is farther as compared to N5 in the pyrimidine scaffold, the rate of the reaction is slowed down, as evident from the 5-fold reduction of catalytic efficiency for isoguanine.

Furthermore, to understand the structural basis of such specificity, the structure of the Msd−5-azacytosine complex was compared with the structure of yCD (group 3 of cog0590; PDB ID: 1UAQ) in complex with a cytosine analogue 3, 4-dihydrouracil (DHU). As mentioned earlier, unlike yCD, the Msd active site is more open and also lacks the C-terminal tail that harbors a conserved aspartic acid.[6] This negatively charged residue in yCD has been implicated as being important toward stabilizing the cytosine moiety and likely determines substrate specificity. Comparison of the active site shows that DHU is unable to make sufficient interactions in Msd, and due to lack of a proper anchor, it is unable to undergo deamination (Figure 2B). However, Msd accepts a similarly sized 5-azacytosine as a substrate, as the presence of the additional nitrogen atom in the azapyrimidine scaffold alters the energetics of the reaction, thereby making deamination more facile. Curiously enough apart from 5-azacytosine, Msd also deaminates ammeline. Guanine deaminases have been earlier shown to possess moonlighting activity toward ammeline.[13,28] A comparison reveals that the active site of Msd has partial resemblance to GD. For instance, the residues involved in hydrophobic packing such as F29 and H56 are conserved in both. Msd also contains an acidic residue in the active site E27, although it is differentially located (Figure 2C). This rearrangement may be the prime reason that Msd does not deaminate guanine but acts on isoguanine. Overall, the active site of Msd appears to be a hybrid between the cytosine and guanine deaminases, resulting in Msd deaminating neither of the pure bases, but recognizes alternate mutagenic scaffolds. Msd functions by allowing compounds with a specific conjugation pattern such that deamination in a less protected environment is viable. GD and yCD, which require complete exclusion of solvent and harbor compact active sites, have been modified by this intelligent organism to proofread for mutagens and export them out of the cell, so they do not linger in the nucleobase pool.

Mutational Analysis of Key Active Site Residues. To determine the relative importance and specific roles of the active site residues, they were systematically mutated. Results clearly show that the zinc-bound proton shuttling E58 is essential for activity. For all four substrates, E58A mutation resulted in complete loss of activity (Figure 5). Interestingly, the other

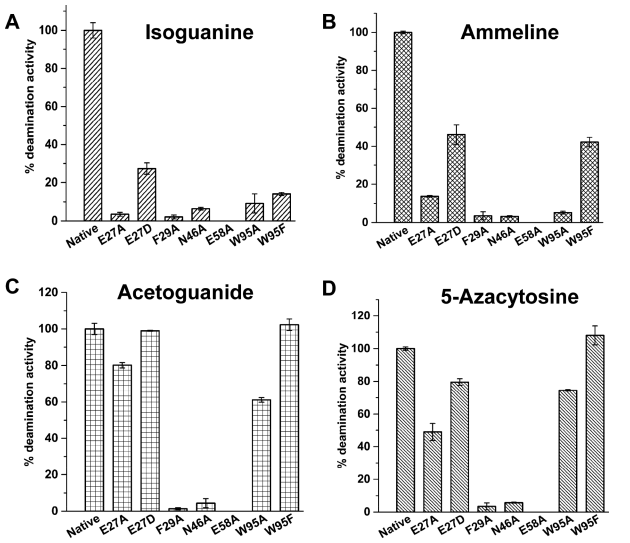


Figure 5. Activity assay of various mutants of Msd toward isoguanine (A), ammeline (B), acetoguanide (C), and 5-azacytosine (D).

glutamic acid residue, E27, seems to show differential behavior. For example, in the case of the bulkier purine isoguanine, E27A mutation led to 90% loss of deamination activity, whereas the E27D mutant retains 30% of its activity. Similarly, the E27A mutation exhibited only 50% loss of activity with 5-azacytosine and had little effect on acetoguanide deamination. Thus, from the activity assay and crystal structure results, it appears that E27 plays an important role in stabilization of the larger compounds such as isoguanine (purine) in contrast to *s*-triazine substrates. Analogously, it was observed that purines required more stabilization, and the mutation of the tryptophan residue from the neighboring subunit to an alanine residue results in complete loss of activity. Even replacement of the tryptophan residue with another hydrophobic residue such as phenylalanine did not help restore activity. However, such drastic effects were not observed for 5-azacytosine derivatives, which could be efficiently deaminated in a less protected environment, with the bulkiest ammeline being the most affected. N46A and F29A mutations also result in almost complete loss of activity for all compounds. While F29 seems to be the major stacking anchor for all compounds, N46 plays an integral role in hydrogen bonding the carbonyl group attached to the C2 atom of the azapyrimidine scaffold, and thereby stabilizes the transition state by pulling away the electrons from the ring. This asparagine residue is generally conserved in all CDA superfamily deaminases, and it has been reported that in enzymes that deaminate DNA bases this mutation leads to only partial loss of activity. For instance, in the case of GD, replacement of this asparagine only results in (60%) loss of activity. However, in Msd it appears to be absolutely necessary for activation, as the active site is more solvent accessible and deamination occurs in a less protected environment.

To further ascertain the role of the interaction of the carbonyl group and the asparagine residue in assisting catalysis, the structure of Msd was solved with the 5-azacytosine analogue, benzoguanamine (2,4-diamino-6-phenyl-1,3,5-triazine) (PDB ID: 5XKR). Benzoguanamine lacks the C2 carbonyl group and instead possesses an amino group in this position. This change of functional group results in benzoguanamine acting as an inhibitor of Msd. Corroborating ITC results shows that it binds Msd with a relatively high binding affinity of 7 μM (Figure 6D). Further, the

DOI: 10.1021/jacs.7b04967
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

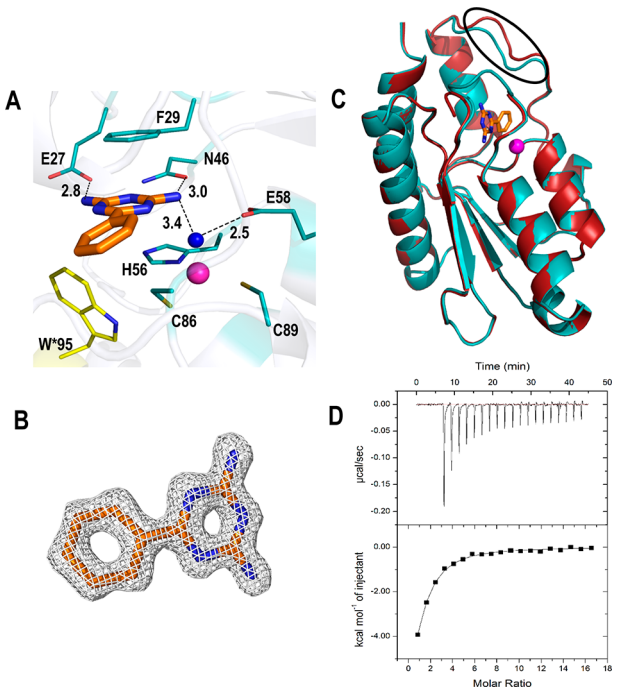


Figure 6. Analysis of Msd−benzoguanamine complex. (A) Interaction network of benzoguanamine in the active site. Carbon atoms of ligand binding residues are colored cyan except for W*95, where * indicates the residue from the adjacent subunit. Benzoguanamine is shown in orange. Zinc atom is shown in magenta, and the zinc-coordinated water molecule is shown as a blue sphere. Oxygen and nitrogen atoms are in red and blue, respectively. (B) Electron density ($F_o - F_c$) map contoured at 3σ for benzoguanamine. (C) Superposition of native Msd (red) and Msd−benzoguanamine complex (cyan), where changes in the structure are encircled. (D) ITC data of Msd with benzoguanamine. Data were fit using one set of sites model.

crystal structure shows that it snugly fits in the binding pocket and the appended phenyl ring in the triazine scaffold stacks with W*95. Both E27 and N46 also anchor the compound into the active site. The distance of the amine group to be deaminated is now 3.4 Å from the catalytic water, which is too far for efficient catalysis to occur (Figure 6A). Hence, either the energetics of the reaction alter due to the absence of the keto group in the core scaffold or perhaps the orientation of the molecule occludes close proximity of the reaction center from both the catalytic water and E58, the key players of deamination. The complete loss of activity as observed for the N46A mutant of Msd clearly indicates that interaction of the keto group and N46 is of paramount importance, and disturbing the scaffold in benzoguanamine is likely the dominant reason.

In addition to the positioning of benzoguanamine in the active site, structural differences (rmsd 2 Å) in the extended loop region could be seen prominently when compared with the native structure (Figure 6C). This flexibility of the loop indicates that this region may play an important role in catalysis and could serve as a cap above the active site. Many members of the CDA superfamily are shown to inherit either loops or a flexible structural element that assists in catalysis. These structural elements undergo conformational changes during the course of the reaction and shield the active site from the bulk solvent. For example, yCD harbors the C-terminal helix and has been shown to form a flexible lid on the active site.[6] Even the C-terminal loop of NE0047 GD has been reported to order upon ligand binding, thereby occluding the solvent. This loop in GD was shown to be very important for deamination, and its deletion led to complete loss of activity.[13,14]

In summary, here a new class of enzymes that specifically act on mutagenic nucleobases such as 5-azacytosine and isoguanine was discovered. These enzymes are unique to certain *Mycobacterium* and soil organisms and likely provide them with innate resistance toward several drugs of the triazine class. The genome location of the enzyme corroborates this conjecture, as the deaminase is situated in a conserved cluster that moderates efflux of mycotoxins such as fusaric acid. The structurally characterized enzyme does not act on any of the natural bases and has fine-tuned its active site such that it is a hybrid of cytosine as well as guanine deaminase and hence is able to accept both purine and pyrimidine-like mutagens. A series of mutagenesis and structures in complex with select substrates help in understanding how Msd moderates its activity and why an azapyrimidine scaffold is the most preferred substrate. Overall, the reported results suggest that mycobacterial deaminases have evolved to combat xenobiotic compounds by employing a series of defense mechanisms where deamination is one of the key reactions. These findings on the intragenus model organism of *M. tuberculosis* help in unraveling the mysteries of mycobacterial biology and provide a framework for rational drug design.

MATERIALS AND METHODS

Sequence Similarity Network. To generate a sequence similarity network, protein sequences belonging to cog0590 were retrieved in FASTA format from NCBI database. An all-by-all BLAST with an *e*-value cutoff of 10^{-40} was performed using BLAST+ and BLAST2 sim plugin software from the NCBI site.[21] This generates a BLAST on BLAST file for each pair, and the output can be loaded into Cytoscape to visualize the network as evolutionarily distinct groups.[21,33]

Purification, Crystallization, and Structure Determination of Msd. The Msd gene was cloned from *Mycobacterium smegmatis* MC2 155 genomic DNA (courtesy of Prof. Sarika Mehra, IIT Bombay) using standard PCR techniques and cloned into the pET28a vector. All the site-directed mutants (E27A, E27D, F29A, N46A, E58A, W95A, and W95F) were made using Phusion DNA polymerase from New England Biolabs. The Msd clone was transformed in *E. coli* Rosetta (DE3) cells, overexpressed with 1 mM IPTG at 25 °C as His-tagged protein, and purified using the standard Ni-NTA affinity chromatography method (Supporting Information). The protein was concentrated to ∼9 mg/mL in the sample buffer (25 mM HEPES, pH 7.5; 200 mM NaCl). Initial crystallization conditions (0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodylate trihydrate, and 15% PEG 8000) were optimized, and crystals were cryo-protected with a solution composed of mother liquor and 20% glycerol. The diffraction data for native Msd were collected at the BM14 beamline at the European Synchrotron Radiation Facility (ESRF), which has a MAR CCD detector. SAD data were collected at the zinc absorption edge of 1.28 Å and energy level of 9.67 keV, and the crystals diffracted to 1.89 Å. The intensities were indexed, integrated, and scaled using the HKL2000[34] program (space group $P2_12_12$; unit-cell parameters, a = 63.26 Å, b = 108.33 Å, c = 53.26 Å, α, β, γ = 90°). The crystal contained two molecules per asymmetric unit with a solvent content of 53.5%. Phases were obtained from the experimental SAD data using the SHELX[35] program, and the model was built using the Autosol program from PHENIX.[36] The resulting native Msd structure (PDB ID: 5XKO) was further built in COOT[37] and refined using REFMAC.[38] Data collection and refinement statistics are summarized in Table 1. Msd complexes with 5-azacytosine (PDB ID: 5XKP), ammeline (PDB ID: 5XKQ), and benzoguanamine (PDB ID: 5XKR) were obtained by rapid soaking native Msd crystal in 10 mM ligand concentration for 1−2 min and were immediately flash frozen for diffraction. The complexed structures were determined by the molecular replacement method (CCP4i[35] and Autorickshaw[39]) using the phases from the native Msd structure. All the figures were prepared using PyMOL.

DOI: 10.1021/jacs.7b04967
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Deamination Assay and Determination of Kinetic Constants. The activity screening of a variety of substrates for this uncharacterized deaminase (Msd) was performed using the Berthelot reaction as described earlier.[13,40] The list of compounds tested for activity are presented in the Supporting Information (Tables S1 and S2). Further, kinetics was performed on a subset of compounds that exhibited appreciable activity (details in the Supporting Information). The experiments were performed in triplicates, and the presented data show the mean of three independent experiments (Table 2).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b04967.

Supporting materials and methods; Tables S1−S4 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel: 91-25767165. Fax: 91-25767152. E-mail: ruchi@chem.iitb.ac.in.

ORCID

Ruchi Anand: 0000-0002-2045-3758

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Professors Pradeep Kumar P.I., R. B. Sunoj, D. Maiti, and A. Chowdhury for helpful discussions. This work was funded by DST, Government of India (Grant Numbers EMR/2015/002121 and DST/TM/WTI/2K16/252). We also thank Protein crystallography facility at IIT Bombay for helping in initial crystallization trials and screening.

■ REFERENCES

(1) Davies, J.; Davies, D. *Microbiol Mol. Biol. Rev.* **2010**, *74*, 417.
(2) Bragginton, E. C.; Piddock, L. J. V. *Lancet Infect. Dis.* **2014**, *14*, 857.
(3) Honer zu Bentrup, K.; Russell, D. G. *Trends Microbiol.* **2001**, *9*, 597.
(4) Boehme, C. C.; Nabeta, P.; Hillemann, D.; Nicol, M. P.; Shenai, S.; Krapp, F.; Allen, J.; Tahirli, R.; Blakemore, R.; Rustomjee, R.; Milovic, A.; Jones, M.; O'Brien, S. M.; Persing, D. H.; Ruesch-Gerdes, S.; Gotuzzo, E.; Rodrigues, C.; Alland, D.; Perkins, M. D. *N. Engl. J. Med.* **2010**, *363*, 1005.
(5) Somoskovi, A.; Parsons, L. M.; Salfinger, M. *Respir Res.* **2001**, *2*, 164.
(6) Ko, T. P.; Lin, J. J.; Hu, C. Y.; Hsu, Y. H.; Wang, A. H.; Liaw, S. H. *J. Biol. Chem.* **2003**, *278*, 19111.
(7) Ireton, G. C.; Black, M. E.; Stoddard, B. L. *Structure* **2003**, *11*, 961.
(8) Iyer, L. M.; Zhang, D.; Rogozin, I. B.; Aravind, L. *Nucleic Acids Res.* **2011**, *39*, 9473.
(9) Kamat, S. S.; Bagaria, A.; Kumaran, D.; Holmes-Hampton, G. P.; Fan, H.; Sali, A.; Sauder, J. M.; Burley, S. K.; Lindahl, P. A.; Swaminathan, S.; Raushel, F. M. *Biochemistry* **2011**, *50*, 1917.
(10) Liaw, S.-H.; Chang, Y.-J.; Lai, C.-T.; Chang, H.-C.; Chang, G.-G. *J. Biol. Chem.* **2004**, *279*, 35479.
(11) Torres, A. G.; Piñeyro, D.; Filonava, L.; Stracker, T. H.; Batlle, E.; Ribas de Pouplana, L. *FEBS Lett.* **2014**, *588*, 4279.
(12) Hall, R. S.; Fedorov, A. A.; Marti-Arbona, R.; Fedorov, E. V.; Kolb, P.; Sauder, J. M.; Burley, S. K.; Shoichet, B. K.; Almo, S. C.; Raushel, F. M. *J. Am. Chem. Soc.* **2010**, *132*, 1762.
(13) Bitra, A.; Hussain, B.; Tanwar, A. S.; Anand, R. *Biochemistry* **2013**, *52*, 3512.
(14) Bitra, A.; Biswas, A.; Anand, R. *Biochemistry* **2013**, *52*, 8106.
(15) Baker, D.; Sali, A. *Science* **2001**, *294*, 93.
(16) Dunbrack, R. L., Jr *Curr. Opin. Struct. Biol.* **2006**, *16*, 374.
(17) Tatusov, R. L.; Galperin, M. Y.; Natale, D. A.; Koonin, E. V. *Nucleic Acids Res.* **2000**, *28*, 33.
(18) Hermann, J. C.; Marti-Arbona, R.; Fedorov, A. A.; Fedorov, E.; Almo, S. C.; Shoichet, B. K.; Raushel, F. M. *Nature* **2007**, *448*, 775.
(19) Hitchcock, D. S.; Fan, H.; Kim, J.; Vetting, M.; Hillerich, B.; Seidel, R. D.; Almo, S. C.; Shoichet, B. K.; Sali, A.; Raushel, F. M. *J. Am. Chem. Soc.* **2013**, *135*, 13927.
(20) Baugh, L.; Phan, I.; Begley, D. W.; Clifton, M. C.; Armour, B.; Dranow, D. M.; Taylor, B. M.; Muruthi, M. M.; Abendroth, J.; Fairman, J. W.; Fox, D., 3rd; Dieterich, S. H.; Staker, B. L.; Gardberg, A. S.; Choi, R.; Hewitt, S. N.; Napuli, A. J.; Myers, J.; Barrett, L. K.; Zhang, Y.; Ferrell, M.; Mundt, E.; Thompkins, K.; Tran, N.; Lyons-Abbott, S.; Abramov, A.; Sekar, A.; Serbzhinskiy, D.; Lorimer, D.; Buchko, G. W.; Stacy, R.; Stewart, L. J.; Edwards, T. E.; Van Voorhis, W. C.; Myler, P. J. *Tuberculosis (Oxford, U. K.)* **2015**, *95*, 142.
(21) Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C. *PLoS One* **2009**, *4*, No. e4345.
(22) Binkowski, T. A.; Naghibzadeh, S.; Liang, J. *Nucleic Acids Res.* **2003**, *31*, 3352.
(23) Hitchcock, D. S.; Fedorov, A. A.; Fedorov, E. V.; Dangott, L. J.; Almo, S. C.; Raushel, F. M. *Biochemistry* **2011**, *50*, 5555.
(24) Hitchcock, D. S.; Fedorov, A. A.; Fedorov, E. V.; Almo, S. C.; Raushel, F. M. *Biochemistry* **2014**, *53*, 7426.
(25) Banerjee, R.; Pace, N. J.; Brown, D. R.; Weerapana, E. *J. Am. Chem. Soc.* **2013**, *135*, 2497.
(26) Christman, J. K.; Schneiderman, N.; Acs, G. *J. Biol. Chem.* **1985**, *260*, 4059.
(27) Rouchaud, J.; Neus, O.; Moulard, C. *Pest Manage. Sci.* **2003**, *59*, 940.
(28) Seffernick, J. L.; Dodge, A. G.; Sadowsky, M. J.; Bumpus, J. A.; Wackett, L. P. *J. Bacteriol.* **2010**, *192*, 1106.
(29) Youenou, B.; Favre-Bonte, S.; Bodilis, J.; Brothier, E.; Dubost, A.; Muller, D.; Nazaret, S. *Genome Biol. Evol.* **2015**, *7*, 2484.
(30) Zeytuni, N.; Zarivach, R. *Structure* **2012**, *20*, 397.
(31) Yamaryo-Botte, Y.; Rainczuk, A. K.; Lea-Smith, D. J.; Brammananth, R.; van der Peet, P. L.; Meikle, P.; Ralton, J. E.; Rupasinghe, T. W. T.; Williams, S. J.; Coppel, R. L.; Crellin, P. K.; McConville, M. J. *ACS Chem. Biol.* **2015**, *10*, 734.
(32) Sklenak, S.; Yao, L.; Cukier, R. I.; Yan, H. *J. Am. Chem. Soc.* **2004**, *126*, 14879.
(33) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res.* **2003**, *13*, 2498.
(34) Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307.
(35) *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 76010.1107/S0907444994003112.
(36) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213.
(37) Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126.
(38) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 240.
(39) Panjikar, S.; Parthasarathy, V.; Lamzin, V. S.; Weiss, M. S.; Tucker, P. A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 449.
(40) Caraway, W. T. *Clin. Chem.* **1966**, *12*, 187.

DOI: 10.1021/jacs.7b04967
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX