Enzymatic Deamination of the Epigenetic Base *N*-6-Methyladenine

Siddhesh S. Kamat[Ψ], **Hao Fan**[φ], **J. Michael Sauder**[‡], **Stephen K. Burley**[‡], **Brian K. Shoichet**[§], **Andrej Sali**[φ], and **Frank M. Raushel**[*,Ψ]

[Ψ]Department of Chemistry, Texas A&M University, P.O. 30012, College Station, Texas 77842-3012

[‡]Eli Lilly & Co., Lilly Biotechnology Center, 10300 Campus Point Dr., Suite 200, San Diego, California 92121

[φ]Department of Bioengineering and Therapeutic Sciences, California Institute for Quantitative Biosciences, University of California, San Francisco, 1700 4th St., San Francisco, CA 94158

[§]Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, University of California, San Francisco, 1700 4th St., San Francisco, CA 94158

Abstract

Two enzymes of unknown function from the amidohydrolase superfamily were discovered to catalyze the deamination of *N*-6-methyladenine to hypoxanthine and methyl amine. The methylation of adenine in bacterial DNA is a common modification for the protection of host DNA against restriction endonucleases. The enzyme from *Bacillus halodurans*, Bh0637, catalyzes the deamination of *N*-6-methyladenine with a k_{cat} of 185 s^{-1} and a k_{cat}/K_m of 2.5×10^6 M^{-1} s^{-1}. Bh0637 catalyzes the deamination of *N*-6-methyladenine two orders of magnitude faster than adenine. A comparative model of Bh0637 was computed using the three-dimensional structure of Atu4426 (PDB code: 3NQB) as a structural template and computational docking was used to rationalize the preferential utilization of *N*-6-methyladenine over adenine. This is the first identification of an *N*-6-methyladenine deaminase (6-MAD).

NCBI has classified proteins from completely sequenced bacterial genomes that perform the same or similar functions into Clusters of Orthologous Groups (COG).[1] One of these clusters, cog1001, consists of ~250 proteins, all of which are members of the amidohydrolase superfamily (AHS). The AHS is an ensemble of evolutionarily related enzymes capable of hydrolyzing amide, amine, or ester functional groups at carbon and phosphorus centers.[2] Enzymes within this superfamily possess a mononuclear or binuclear metal center embedded within a $(\beta/\alpha)_8$-barrel structural fold.[2]

Cog1001

A sequence similarity network for cog1001 is presented in Figure 1 at an E-value cutoff of 10^{-80}.[3,4] Adenine deaminase (ADE) is the predominant enzyme in cog1001 and is represented in Figure 1 by subgroups **1** and **4**. Subgroups **3** and **5** consist of enzymes that are

currently annotated as isoaspartyl dipeptidases and enamidases, respectively.[5,6] The functional annotations for most members of subgroup **2** in Figure 1 are currently listed as unknown in the NCBI database but some of them have been labeled as adenine deaminases without apparent experimental validation. Proteins in subgroup **2** have less than 35% sequence identity to the prototypical adenine deaminases found in subgroups **1** and **4**.

There are sequences available for ~32 proteins from subgroup **2**. These enzymes are found in various species of *Bacillus*. At least a dozen of these microorganisms also have an authentic adenine deaminase from subgroup **1** in their genome. Most of the genes in subgroup **2** from cog1001 lie adjacent to the *pur* operon for the biosynthesis of purine nucleotides. Moreover, in *Bacillus halodurans*, the genes for ADE from subgroup **1** (Bh0640) and Bh0637 from subgroup **2** are nearly adjacent to one another. The co-localization of these two genes suggests that these enzymes may use the same or similar substrates and that the products of their reactions are similar. Based on these observations, we have thus predicted that Bh0637 is likely to have an adenine-like substrate and that the reaction product resembles hypoxanthine.

There are a limited number of purine moieties in Nature that closely resemble adenine. A prime substrate candidate for enzymes in subgroup **2** is *N*-6-methyladenine because this compound is found in the genomes of all bacteria.[7] This compound is often considered as the 6th DNA base with 5-methylcytosine being the 5th base. The occurrence of *N*-6-methyladenine in most bacterial genomes varies from 1.4 – 2.0% of the total adenine content.[8] This nucleotide modification functions to protect host DNA against restriction endonucleases and is critical for cell viability.[8] More recently, *N*-6-methyladenine has been implicated in DNA repair and replication, regulation of gene expression, control of transposition, and host-pathogen interactions.[7,8] In bacteria, DNA methylases and cell-cycle regulated methyltransferases are responsible for the methylation of adenine in DNA.[7,8] Previous studies with cell free extracts from various sources have implicated a cryptic enzyme activity for the conversion of *N*-6-methyladenine to hypoxanthine.[9,10] However, no enzyme is currently known to metabolize *N*-6-methyladenine. Other modified adenine moieties that could serve as potential substrates include 7-methyladenine and 2-hydroxyadenine (isoguanine).[11,12]

Enzyme Purification

Bh0637 from *B. halodurans* C-125 was selected from enzymes clustered in subgroup **2** for substrate interrogation. The gene for Bh0637 was cloned into a pET30a(+) plasmid (Novagen) with *Nde*I and *Hin*dIII restriction sites, expressed in *Escherichia coli* and purified to homogeneity. Bsu06560 from *Bacillus subtilis* (gi|16077724) was also purified to homogeneity. These two proteins share 71% sequence identity. The gene for Bsu06560 was codon-optimized for expression in *E. coli*, synthesized, and then cloned into an expression vector that yields a protein with a non-cleavable C-terminal histidine tag. The sequence and expression/purification details for Bsu06560 are available from PepcDB (http://pepcdb.pdb.org/) under target ID "NYSGXRC-9208a", clone 9208a4BCt13p1. The proteins were expressed with an iron chelator, 2,2′-dipyridyl, to prevent oxidation of active site sidechains. ICP-MS of the purified enzymes established the presence of ~2 equivalents of manganese per protein monomer. Apo-Bh0637 was prepared by dialysis against 10 mM *o*-phenanthroline at pH 7.5. Zn^{2+} and Fe^{2+} were added to reconstitute the enzymatic activity of Bh0637. Iron was added anaerobically to prevent air-oxidation prior to reconstitution.

Enzymatic Assays

Bh0637 and Bsu06560 were screened against a small highly focused library of approximately 20 adenine-like compounds (Supporting Information). All of the compounds

in the compound library were monitored for changes in the UV spectrum. The compounds (100-150 µM) were dissolved in 20 mM HEPES, pH 7.5, and incubated with 1.0 µM Bh0637 or Bsu06560 for 2 hours. The only compounds showing any significant absorbance changes were adenine, *N*-6-methyladenine, 6-methoxypurine and 6-methylmercaptopurine.

The kinetic constants for the deamination of adenine were measured using a coupled assay for the formation of ammonia.[13] The deamination of *N*-6-methyladenine was followed directly at 270 nm ($\Delta\epsilon = 15{,}000$ M^{-1} cm^{-1}).[8] The cleavage of methanol from 6-methoxypurine was monitored at 270 nm ($\Delta\epsilon = 3760$ M^{-1} cm^{-1}) and the loss of methylthiol from 6-methylmercaptopurine was determined at 290 nm ($\Delta\epsilon = 20{,}200$ M^{-1} cm^{-1}).[14,15] The formation of hypoxanthine was confirmed using UV absorbance spectroscopy. The production of methylamine from *N*-6-methyladenine was demonstrated by mass spectrometry and the release of methanol from 6-methoxypurine was verified with alcohol dehydrogenase. DTNB (5,5′-dithiobis(2-nitro-benzoic acid)) was utilized to monitor the formation of methylthiol from 6-methylmercaptopurine. The kinetic constants for the various substrates with the two enzymes are presented in Table 1. Since similar kinetic constants are observed for the Zn-, Mn-, and Fe-reconstituted forms of Bh0637 and Bsu06560, it is difficult to predict the metal used by either enzyme *in vivo*. The overall reaction for the deamination of *N*-6-methyladenine is represented in Scheme 1.

Structural Model for Bh0637

A comparative model of the three-dimensional structure for Bh0637 was constructed. The amino acid sequence of Bh0637 was first submitted to the PSI-BLAST server at NCBI to identify suitable template structures.[16] The crystal structure of adenine deaminase (Atu4426, PDB code: 3NQB) was the only template that produced a significant alignment with an E-value (6×10^{-20}) better than the default threshold of 6×10^{-3} (Supporting Information). The sequence alignment between Bh0637 and Atu4426 was computed by profile-profile alignment as implemented in the "profile.scan" routine of MODELLER-9v8.[17] A total of 500 comparative models were generated with the standard "automodel" class in MODELLER, and the model with the best DOPE score[18] was selected for refinement of backbone and side-chain conformations of residues near the binuclear metal center (Ser-464 to Gly-469, Ser-552 to Pro-557) using the "loopmodel" class in MODELLER. Sidechains of residues in these two loops and Phe-92, Arg-253, Thr-467, Phe-550 and His-555 were subsequently refined using the "sidechain prediction" protocol in PLOP.[19] This refinement protocol resulted in eight representative models of Bh0637.

Computational Docking to Bh0637

The high-energy intermediate (HEI) library of KEGG molecules was docked to all eight comparative models of Bh0637 by DOCK 3.5.54.[20-26] The model yielding the highest enrichment for *N*-6-methyladenine was selected to represent the most likely binding mode of this substrate in Bh0637. A comparison of the binding site in the ligand-free template structure of Atu4426 (PDB: 3NQB) and the binding site in the comparative model of Bh0637 is presented in Figure 2. In the Bh0637 model, Glu-233 and the six residues that coordinate the binuclear metal center (His-88, His-90, Glu-180, His-211, His-232 and Asp-284) overlap with the aligned residues from the structural template. Several residues that probably interact with *N*-6-methyladenine are not conserved between the target and the modeling template. Ser-97, His-122, Glu-123, Asp-290 and Gly-563 in Atu4426 are replaced by Phe-92, Leu-117, Met-118, Gly-285 and Phe-550 in Bh0637, respectively. The substrate binding site of Bh0637 is therefore more hydrophobic than that of Atu4426.

A model for the binding of the intermediate that would be formed by the attack of hydroxide on *N*-6-methyladenine in the active site of Bh0637 is presented in Figure 2. The tetrahedral

intermediate was formed by attack of the bridging hydroxide on the C6 carbon from the *re*-face of *N*-6-methyladenine.[26] In this model, the N6 amino group hydrogen bonds with the side chain of Asp-284. The protonated N1 nitrogen hydrogen bonds with the sidechain of Glu-233. In addition, the N7 nitrogen coordinates the α-metal ion and the N9 nitrogen hydrogen bonds with the sidechain of Thr-467. The N3 nitrogen is adjacent to the Nε nitrogen from His-555 (3.9 Å), which was treated as protonated during the virtual screening calculations. The methyl group of the docked substrate is not predicted to be in close contact with any of the neighboring residues. However, if Gly-285 is replaced by a larger residue, such as the aspartate found at the equivalent position in Atu4426, the methyl group of the substrate would fall within 4 Å of the sidechain of the larger residue.

In addition to *N*-6-methyladenine, the top ranked compounds that could undergo nucleophilic attack by the hydroxide on the *re*-face were inspected visually. The docked pose of adenine is nearly identical to that of *N*-6-methyladenine but the relative rank of adenine (ranked 282) is worse than that of *N*-6-methyladenine (ranked 176). The prioritization of *N*-6-methyladenine over adenine appears to reflect the more favorable van der Waals interactions between the *N*-6-methyl group and hydrophobic residues in the substrate binding site, such as Phe-92 and Phe-550. Five other compounds that scored well in the computational docking were subsequently tested as substrates but none were found to be deaminated at measureable rates. These compounds included 5-amino-4-imidazolecarboxyamide, toxopyrimidine, 5-acetylamino-6-amino-3-methyluracil, 2-hydroxy-4-amino-6-(cyclopropylamino)-1,3,5-triazine, and deisopropylhydroxyatrazine (Supporting Information).

Physiological Relevance

For Bh0637 and Bsu06560 *N*-6-methyladenine is a 100-fold better substrate than the next best candidate compound based on the relative values of k_{cat}/K_m. *N*-6-methyladenine is a known metabolite and is widely prevalent in all bacteria. Bh0637 and Bsu06560 therefore can support the recycling of *N*-6-methyladenine to form hypoxanthine, which is presumably taken up by the purine salvage pathway to form guanine nucleotides from inosine.[27] These two enzymes can also use adenine to form hypoxanthine but not nearly at the same rate as *N*-6-methyladenine. We conclude that *N*-6-methyadenine is the likely physiological substrate for members of subgroup **2** in cog1001 and thus these enzymes should now be annotated as 6-methyladenine deaminases (6-MAD). Consistent with this view, the prototypical adenine deaminases in *B. halodurans* (Bh0640) and *B. subtilus* (Bsu14520) cannot use *N*-6-methyladenine as substrates.[27] The kinetic constants for the deamination of adenine by Bh0640 are 120 ± 4 s^{-1}, 0.30 ± 0.04 mM, and $(4 \pm 0.3) \times 10^5$ M^{-1} s^{-1} for k_{cat}, K_m and k_{cat}/K_m respectively. The kinetic constants for the deamination of adenine by Bsu14520 are 140 ± 6 s^{-1}, 0.38 ± 0.04 mM, and $(3.8 \pm 0.3) \times 10^5$ M^{-1} s^{-1} for k_{cat}, K_m and k_{cat}/K_m respectively.

Strategy for Functional Annotation

It has been estimated that nearly one third of the currently sequenced genes have an unknown, uncertain, or incorrect functional annotation.[28] We have demonstrated that Bh0437 and Bsu06560 catalyze the deamination of *N*-6-methyladenine. Conservation of residues suggest that the rest of the enzymes depicted in subgroup **2** of Figure 1 will catalyze the same enzymatic reaction (see Supporting Information for a list of annotated N-6-methyladenine deaminases in cog1001). The key observations that led to the successful experimental annotation of these proteins included: recognition of adenine deaminase as the closest functional homologue, localization of the gene for this enzyme near the operon for purine synthesis, and identification of residues within the substrate binding site that could

accommodate larger substituents attached to the N6 nitrogen. These insights enabled experimental identification of the enzyme substrate using a relatively modest compound library. The construction of a comparative model and the docking of *N*-6-methyl adenine in the active site of Bh0637 elucidated the structural rationale for the preferential deamination of this compound relative to the poorer substrate, adenine.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

(1). Tatusov RL, Galperin MY, Natale DA, Koonin EV. Nucleic Acids Research 2000;28:33. [PubMed: 10592175]

(2). Seibert CM, Raushel FM. Biochemistry 2005;44:6383. [PubMed: 15850372]

(3). Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Genome Res 2003;13:2498. [PubMed: 14597658]

(4). Atkinson HJ, Morris JH, Ferrin TE, Babbit PC. PLoS ONE 2009;4:4345.

(5). Marti-Arbona R, Fresquet V, Thoden JB, Davis ML, Holden HM, Raushel FM. Biochemistry 2005;44:7115. [PubMed: 15882050]

(6). Kreb D, Alhapel A, Pierek A, Essem LO. J. Mol. Biol 2008;384:837. [PubMed: 18805424]

(7). Wion D, Casadesus J. Nature Reviews/Microbiology 2006;4:183.

(8). Ratel D, Ravanat JL, Berger F, Wion D. Bioessays 28 (Wiley Periodicals) 2006:309.

(9). Remy CN. J. Biol. Chem 1961;236:2999. [PubMed: 14491430]

(10). Duggan DE, Weigert MG, Grieb WE, Titus EO. Biochim. Biophys. Acta 1963;68:519.

(11). Trewick SC, Henshaw TF, Hausinger RP, Lindahl T, Sedgwick B. Nature 2002;419:174. [PubMed: 12226667]

(12). Hashiguchi K, Zhang QM, Sugiyama H, Ikeda S, Yonei S. Int. J. Radiation Biol 2002;78:585.

(13). Marti-Arbona R, Xu C, Steele S, Weeks A, Kuty GF, Seibert CM, Raushel FM. Biochemistry 2006;45:1997. [PubMed: 16475788]

(14). Cercignani G. Anal. Biochem 1987;166:418. [PubMed: 3124669]

(15). Quiggle K, Wejrowski ML, Chladek S. Biochemistry 1978;17:94. [PubMed: 618551]

(16). Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. Nucleic Acids Res 1997;25:3389. [PubMed: 9254694]

(17). Sali A, Blundell TL. Journal of Molecular Biology 1993;234:779. [PubMed: 8254673]

(18). Shen MY, Sali A. Protein Sci 2006;15:2507. [PubMed: 17075131]

(19). Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. J Med Chem 2006;49:534. [PubMed: 16420040]

(20). Hermann JC, Ghanem E, Li YC, Raushel FM, Irwin JJ, Shoichet BK. J Am Chem Soc 2006;128:15882. [PubMed: 17147401]

(21). Hermann JC, Marti-Arbona R, Fedorov AA, Fedorov E, Almo SC, Shoichet BK, Raushel FM. Nature 2007;448:775. [PubMed: 17603473]

(22). Kanehisa M, Goto S. Nucleic Acids Res 2000;28:27. [PubMed: 10592173]

(23). Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. Nucleic Acids Res 2006;34:354.

(24). Lorber DM, Shoichet BK. Current Topics in Medicinal Chemistry 2005;5:739. [PubMed: 16101414]

(25). Meng EC, Shoichet BK, Kuntz ID. Journal of Computational Chemistry 1992;13:505.

(26). Xiang DF, Kolb P, Fedorov AA, Meier MM, Fedorov LV, Nguyen TT, Sterner R, Almo SC, Shoichet BK, Raushel FM. Biochemistry 2009;48:2237. [PubMed: 19159332]

(27). Nygaard P, Duckert P, Saxild HH. J. Bact 1996;178:846. [PubMed: 8550522]

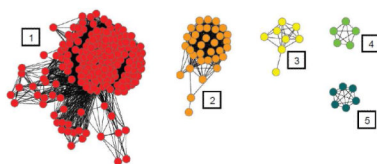(28). Schones AM, Brown SD, Dodevski I, Babbitt PC. PLoS Comp. Bio 2009;12:e1000605.

Figure 1.
Sequence similarity network representation of cog1001 obtained at an E-value cutoff 10^{-80}. Subgroups **1** and **4** are annotated as adenine deaminases. Subgroups **3** and **5** are annotated as isoaspartyl didpeptidases and enamidases, respectively. The reactions catalyzed by enzymes in subgroup **2** are unknown.
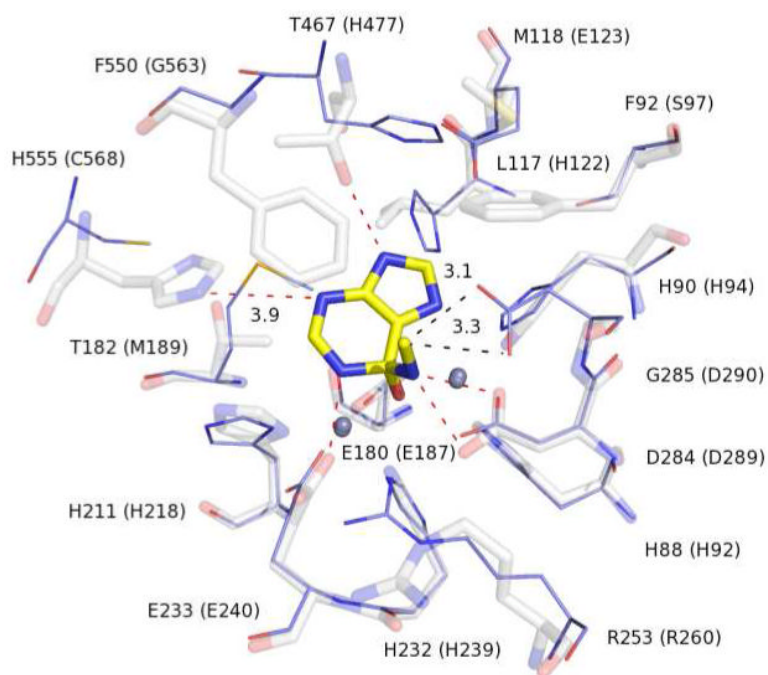
Figure 2.
The predicted binding pose of the high energy intermediate for the deamination of *N*-6-methyladenine in the modeled binding site of Bh0637 (transparent stick representation) superposed on the apo-crystal structure of Atu4426 (solid stick representation). Polar interactions between *N*-6-methyladenine and the Bh0637 model are shown with red dashed lines. The residue numbering for Bh0637 is indicated with the numbers for the corresponding residues in Atu4426 given in parentheses.
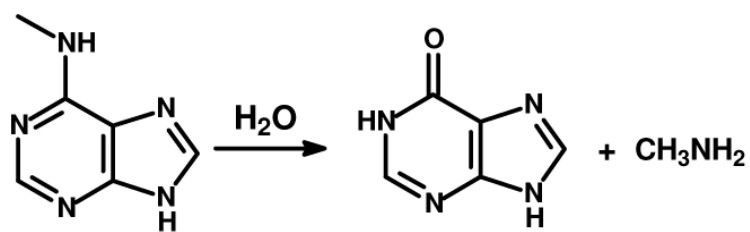
Scheme 1.
Reaction of *N*-6-methyladenine deaminase (6-MAD).

Table 1

Kinetic constants for Bh0637 and Bsu06560

Enzyme	Metal	Substrate	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
Bh0637	Mn	*N*-6-MeAd	135 ± 5	(2.9 ± 0.2) × 10^6
Bh0637	Fe	*N*-6-MeAd	151 ± 3	(3.8 ± 0.3) × 10^6
Bh0637	Zn	*N*-6-MeAd	185 ± 3	(2.5 ± 0.2) × 10^6
Bh0637	Mn	adenine	4.7 ± 0.2	(1.1 ± 0.2) × 10^4
Bh0637	Mn	6-MOAd	1.7 ± 0.1	(1.5 ± 0.1) × 10^4
Bh0637	Mn	6-MMCP	5.8 ± 0.3	(6.1 ± 0.6) × 10^4
Bsu06560	Mn	*N*-6-MeAd	154 ± 3	(2.5 ± 0.2) × 10^6
Bsu06560	Mn	adenine	3.3 ± 0.3	(8.0 ± 1.0) × 10^3

N-6-MeAd = *N*-6-methyladenine, 6-MOAd = 6-methoxyadenine, 6-MMCP = 6-methylmercaptopurine.