BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

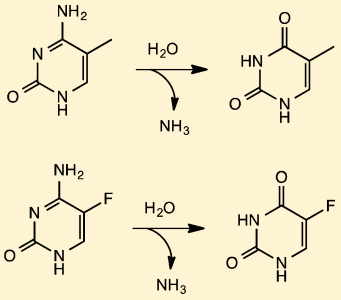
Discovery of a Bacterial 5-Methylcytosine Deaminase

Daniel S. Hitchcock,[†] Alexander A. Fedorov,[§] Elena V. Fedorov,[§] Steven C. Almo,*[,§] and Frank M. Raushel*[,†,‡]

[†]Department of Biochemistry & Biophysics and [‡]Department of Chemistry, Texas A&M University, College Station, Texas 77843, United States

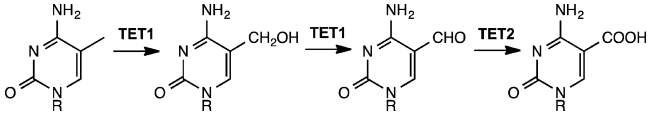
[§]Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, United States

ABSTRACT: 5-Methylcytosine is found in all domains of life, but the bacterial cytosine deaminase from *Escherichia coli* (CodA) will not accept 5-methylcytosine as a substrate. Since significant amounts of 5-methylcytosine are produced in both prokaryotes and eukaryotes, this compound must eventually be catabolized and the fragments recycled by enzymes that have yet to be identified. We therefore initiated a comprehensive phylogenetic screen for enzymes that may be capable of deaminating 5-methylcytosine to thymine. From a systematic analysis of sequence homologues of CodA from thousands of bacterial species, we identified putative cytosine deaminases where a "discriminating" residue in the active site, corresponding to Asp-314 in CodA from *E. coli*, was no longer conserved. Representative examples from *Klebsiella pneumoniae* (locus tag: Kpn00632), *Rhodobacter sphaeroides* (locus tag: Rsp0341), and *Corynebacterium glutamicum* (locus tag: NCgl0075) were demonstrated to efficiently deaminate 5-methylcytosine to thymine with values of k_{cat}/K_m of 1.4×10^5, 2.9×10^4, and 1.1×10^3 M^{-1} s^{-1}, respectively. These three enzymes also catalyze the deamination of 5-fluorocytosine to 5-fluorouracil with values of k_{cat}/K_m of 1.2×10^5, 6.8×10^4, and 2.0×10^2 M^{-1} s^{-1}, respectively. The three-dimensional structure of Kpn00632 was determined by X-ray diffraction methods with 5-methylcytosine (PDB id: 4R85), 5-fluorocytosine (PDB id: 4R88), and phosphonocytosine (PDB id: 4R7W) bound in the active site. When thymine auxotrophs of *E. coli* express these enzymes, they are capable of growth in media lacking thymine when supplemented with 5-methylcytosine. Expression of these enzymes in *E. coli* is toxic in the presence of 5-fluorocytosine, due to the efficient transformation to 5-fluorouracil.



5-Methylcytosine is a modified nucleobase formed by the methylation of cytosine in DNA. The synthesis of 5-methylcytosine is catalyzed by DNA methyltransferases, and in animals, plants, and fungi this modification functions as an epigenetic marker.[1−3] In mammals, methylation occurs predominantly at CpG sites in ∼1% of the human genome.[4] In *Escherichia coli* and related bacteria, methylation occurs at CC(A/T)GG sites by the dcm methylase.[5] Methylation of cytosine in the DNA of bacteria is part of the restriction/modification system and has also been implicated in controlling gene expression during stationary phase.[6] There is no known direct demethylation reaction to form cytosine from 5-methylcytosine in DNA. Instead, the methyl group is first hydroxylated and then oxidized to form 5-carboxycytosine, which is excised from DNA by base excision repair (Scheme 1).[7,8] This process is initiated by

Scheme 1



methylcytosine dioxygenase 1 (TET1) to produce 5-hydroxymethylcytosine.[9] Further oxidation of hydroxymethyl cytosine by TET1 and methylcytosine dioxygenase 2 (TET2) yields 5-formylcytosine and 5-carboxycytosine, respectively.[10]

The deamination of the cytosine moiety in nucleotides and nucleic acids is a conserved metabolic step for the recycling of pyrimidines across all domains of life. This reaction may occur through the deamination of cytosine,[11,12] cytidine,[13] cytidine monophosphate,[14,15] or cytidine triphosphate.[16,17] Cytidine deaminases from cog0295 are found in both prokaryotes and eukaryotes.[18,19] The cytidine deaminases from *E. coli* and yeast have been studied in some detail, and the *E. coli* enzyme has been shown to be catalytically active with both cytidine and 5-methylcytidine. At least two variants of cytosine deaminase exist. The yeast cytosine deaminase can deaminate 5-methylcytosine in addition to cytosine and the active site of this enzyme is similar to that of cytidine deaminase.[12,20] These enzymes are members of the cytidine deaminase-like superfamily and cog0590. In contrast, the unrelated bacterial cytosine deaminase (CodA) from *E. coli* (locus tag: b0337) will not deaminate 5-methylcytosine at appreciable rates.[21]

CodA from *E. coli* is a member of cog0402 and the amidohydrolase superfamily (AHS).[22,23] Other deaminases from this Cluster of Orthologous Groups (COG) include guanine deaminase,[24] *S*-adenosylhomocysteine deaminase,[25] and 8-oxoguanine deaminase[26] among others. The structure of CodA from *E. coli* has been determined in the absence of

Received: October 11, 2014
Revised: November 7, 2014
Published: November 10, 2014

   dx.doi.org/10.1021/bi5012767 | *Biochemistry* 2014, 53, 7426−7435

bound ligands (PDB id: 1K6W), and also in the presence of isoguanine (PDB id: 3RN6) and a phosphonate mimic of the transition-state (PDB id: 3O7U; Figure 1). Substrate binding

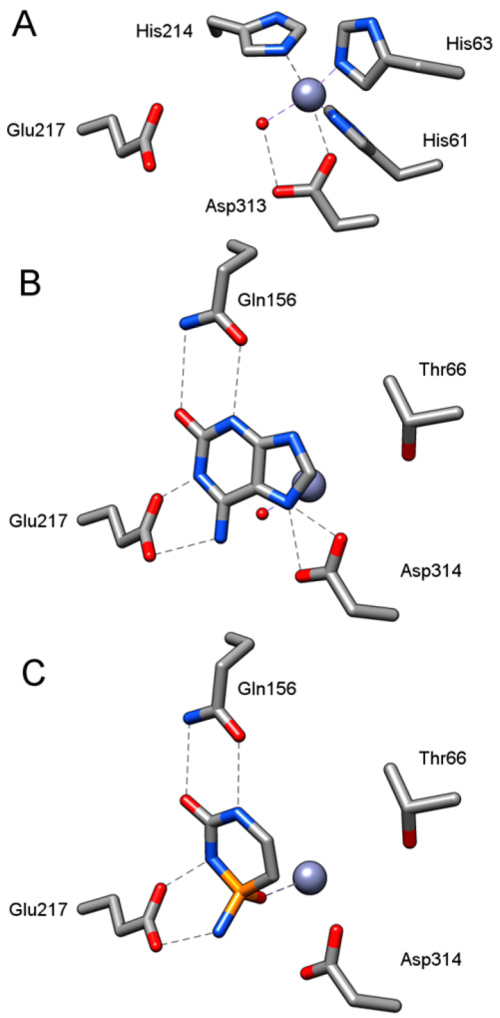


Figure 1. Active site structure of CodA from *E. coli*. (A) Residues involved in the binding of the divalent cation in the active site are conserved in all enzymes from cog0402 of the amidohydrolase superfamily (PDB id: 1K6W). (B) Mode of binding of isoguanine in the active site of CodA (PDB id: 3RN6). (C) Mode of binding of phosphonocytosine in the active site of CodA (PDB id: 3O7U).

relies on Gln-156, which forms a pair of hydrogen bonds with the carboxamide moiety of the pyrimidine or purine base. Glu-217 participates in substrate recognition and catalysis by a direct interaction with the amidine moiety of the substrate (Figure 1B and C). In this active site, Asp-314 provides an apparent steric boundary for the binding of cytosine as a substrate, and participates in a hydrogen bond to N7 of the purine ring for recognition of isoguanine (Figure 1B). CodA can accept pyrimidine (cytosine) and purine (isoguanine) substrates but the active site is apparently not configured to deaminate structurally related compounds such as 5-methylcytosine and 5-fluorocytosine.[21,27]

Since significant amounts of 5-methylcytosine are produced in both prokaryotes and eukaryotes, this compound must eventually be catabolized and the fragments recycled. We therefore initiated a search for enzymes of unknown function related to the bacterial cytosine deaminase with the catalytic ability to deaminate 5-methylcytosine. By analyzing sequence homologues of CodA from thousands of bacterial species, we identified groups of putative cytosine deaminases where the "discriminating" residue corresponding to Asp-314 in CodA from *E. coli* was no longer conserved. Representative examples of these enzymes were purified and found to efficiently deaminate cytosine, 5-methylcytosine, and 5-fluorocytosine. Expression of this enzyme in thymine auxotrophs of *E. coli* rescued growth in the presence of 5-methylcytosine. Expression of this enzyme was toxic in the presence of 5-fluorocytosine in strains of *E. coli* that also expressed uracil phosphoribosyltransferase.

MATERIALS AND METHODS

***E. coli* Cell Lines.** Two gene knockout strains of *E. coli* were obtained from the Coli Genetic Stock Center (CGSC) at Yale University. Both cell lines lack the genes for the metabolism of arabinose, allowing the use of arabinose-inducible plasmids. The pyrimidine auxotroph (CGSC-9145) lacks the gene for orotidine-5-phosphate decarboxylase (F^-, Δ*(araD-araB)567*, Δ*lacZ4787*(::rrnB-3), λ^-, Δ*pyrF789::kan*, *rph-1*, Δ*(rhaD-rhaB)-568*, *hsdR514*).[28] This cell line expresses pyrimidine phosphoribosyltransferase and will convert added pyrimidine nucleotides to their monophosphorylated counterparts, including 5-fluorocytosine and 5-fluorouracil. The thymine auxotroph (CGSC-4091) lacks the gene for thymidylate synthetase (F-, *araBAD-1*?, *tsx-77*, *thyA40*, *deoB15*).

Cloning and Purification of Kpn00632, Rsp0371, NCgl0075, and CodA. The genes for Kpn00632 from *Klebsiella pneumoniae* subsp. pneumoniae (gil152969203), Rsp0371 from *Rhodobacter sphaeroides* 2.4.1 (gil77463913), and NCgl0075 from *Corynebacterium glutamicum* ATCC 13032 (gil19551325) were cloned into pET28 with a C-terminal His_6-tag using standard cloning practices. The plasmids were transformed into *E. coli* BL-21 DE3 cells and plated onto LB-agarose containing 50 μg/mL kanamycin. Cultures of 1 L (LB broth, kanamycin 50 μg/mL) were inoculated with the resulting colonies and grown at 37 °C until the optical density at 600 nm reached 0.6. Protein expression was induced with 1.0 mM isopropyl β-D-1-thiogalactopyranoside, and the cultures were shaken for 20 h at 25 °C. The cultures were centrifuged at 7000 rpm for 15 min, and the isolated cell pellets disrupted by sonication in 35 mL of buffer A (20 mM HEPES, 250 mM NaCl, 250 mM NH_4SO_4, 20 mM imidazole, pH 7.5) containing 2.5 mg of phenylmethylsulfonyl fluoride (PMSF). DNA was removed by dropwise addition of 100 mg of protamine sulfate dissolved in 15 mL of buffer A. Proteins with a His_6-tag were loaded onto a 5 mL HisTrap column (GE Heathcare) with running buffer A and eluted with a linear gradient of elution buffer B (20 mM HEPES, 250 mM NaCl, 250 mM NH_4SO_4, 500 mM imidazole, pH 7.5).

The gene for cytosine deaminase (CodA, b0337) from *E. coli* K12 (gil16128322) was cloned into pBAD322C without purification tags.[29] *E. coli* cells lacking arabinose metabolizing genes (CGSC 9145) were transformed with the plasmid and plated onto LB-agarose containing 25 μg/mL chloramphenicol. A 1 L culture (LB broth, chloramphenicol 25 μg/mL) was inoculated from the resulting colonies and grown at 37 °C until the optical density at 600 nm reached 0.6. Protein expression was induced with 1.0 mM L-arabinose, and the culture was shaken for 20 h at 25 °C. The sample was centrifuged at 7000 rpm for 15 min and the cell pellet was disrupted by sonication in 35 mL of running buffer (50 mM HEPES, pH 7.5) and 2.5 mg of PMSF. DNA was removed by the dropwise

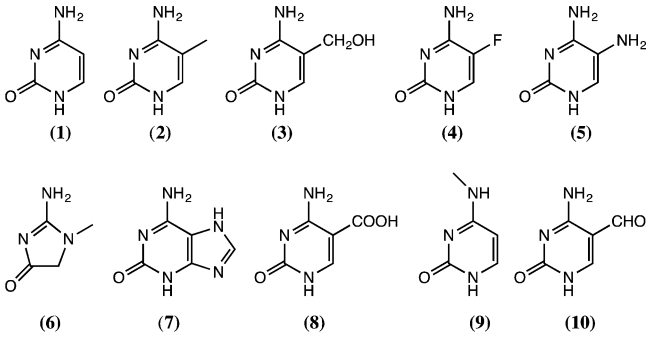
dx.doi.org/10.1021/bi5012767 | *Biochemistry* 2014, 53, 7426−7435

addition of 100 mg protamine sulfate dissolved in 15 mL of running buffer. The protein was precipitated with 70% saturated NH_4SO_4, and the pellet resuspended in 4.0 mL of running buffer. The protein solution was loaded onto a HiLoad 26/600 Superdex 200 gel filtration column. Fractions containing active enzyme were pooled and concentrated.

Cloning of Enzymes for in Vivo Assays. The genes for Kpn00632 and Rsp0341 were cloned into pBAD322C using standard cloning practices without purification tags. The D314A mutant of cytosine deaminase from *E. coli* was constructed by standard site directed mutagenesis methods using primer overlap extension from the pBAD322c-b0337 plasmid with the primers 5′-CGTCTGCTTTGGTCACGATGCTGT-CTTCGATCCGTGGTATCC-3′ and 5′- GGATACCACGGA-TCGAAGACAGCATCGTGACCAAAGCAGACG-3′.[30]

Activity Screening and Determination of Kinetic Constants. Purified enzyme was incubated in 50 mM HEPES, pH 7.5, at 25 °C for 1 h in a 96-well UV−vis quartz plate with a small library of potential substrates (Scheme 2). The spectra

Scheme 2



were monitored as a function of time from 240 to 350 nm. The substrate library included cytosine (**1**), 5-methylcytosine (**2**), 5-hydroxymethylcytosine (**3**), 5-fluorocytosine (**4**), 5-aminocytosine (**5**), creatinine (**6**), isoguanine (**7**), cytosine-5-carboxylate (**8**), *N*-methylcytosine (**9**), and 5-formylcytosine (**10**). Initial reaction velocities were measured by a direct UV−vis assay in a 96-well quartz plate using various enzyme/substrate combinations in 20 mM HEPES, pH 7.5 at 30 °C. Product formation was monitored at the following wavelengths using experimentally derived differential molar extinction coefficients for the following substrates: cytosine (255 nm, $\Delta\varepsilon$ = 2600 M^{-1} cm^{-1}), 5-methylcytosine (262 nm, $\Delta\varepsilon$ = 3700 M^{-1} cm^{-1}), creatinine (240 nm, $\Delta\varepsilon$ = 6100 M^{-1} cm^{-1}), isoguanine (294 nm, $\Delta\varepsilon$ = −6600 M^{-1} cm^{-1}), 5-fluorocytosine (262 nm, $\Delta\varepsilon$ = 3200 M^{-1} cm^{-1}), 5-aminocytosine (235 nm, $\Delta\varepsilon$ = −2000 M^{-1} cm^{-1}), and 5-hydroxymethylcytosine (258 nm, $\Delta\varepsilon$ = 3700 M^{-1} cm^{-1}). Values of k_{cat} and k_{cat}/K_m were determined by fitting the data to eq 1 using SigmaPlot 11, where v is the initial velocity, E_t is enzyme concentration, and A is the substrate concentration.

$$v/E_t = k_{cat}A/(A + K_a) \quad (1)$$

Selection Medium for in Vivo Experiments. The selection medium contained M9 minimal salts, 1.96 g/L yeast synthetic dropout medium without uracil, 0.1% glycerol, 100 μM $CaCl_2$, 1.0 mM $MgSO_4$, and trace elements. The 5000x mixture of trace elements was prepared with 95 mL of H_2O, 5.0 g of citric acid, 5.0 g of $ZnSO_4·(7\ H_2O)$, 4.75 g of $FeSO_4·(7\ H_2O)$, 1.0 g of $Fe(NH_4)_2(SO_4)_2·(6\ H_2O)$, 250 mg of $CuSO_4·(H_2O)$, and 50 mg each of $MnSO_4·(H_2O)$ and $Na_2MoO_4·(2H_2O)$. Yeast synthetic dropout medium with glycerol (and cytosine when supplemented) was autoclaved separately from the M9 medium. $CaCl_2$ and $MgSO_4$ were sterile filtered and the mixture of trace elements was autoclaved before all components were combined.

Determination of in Vivo 5-Methylcytosine Deaminase Activity. The thymine auxotroph of *E. coli* was transformed with pBAD322C and the plasmids containing cytosine deaminase from *E. coli*, Kpn00326, and Rsp0341. These cells were plated onto LB-agar with chloramphenicol (25 μg/mL) supplemented with 200 μM thymine. Single colonies were picked, and 5 mL cultures of LB containing 100 μM thymine and 25 μg/mL chloramphenicol were grown for 14 h overnight. Cultures of selection medium (25 mL) containing 100 μM $MnSO_4$, 100 μM zinc acetate, and 25 μg/mL chloramphenicol were prepared in 100 mL flasks with the following experimental conditions: (a) 400 μM 5-fluorocystosine (b) 100 μM arabinose, (c) 100 μM thymine and 100 μM arabinose, (d) 100 μM 5-methylcytosine and 100 μM arabinose, and (e) 500 μM 5-methylcytosine and 100 μM arabinose. The cultures were inoculated with 25 μL from the overnight cultures and shaken at 200 rpm for 12 h at 37 °C. The absorbance at 600 nm was measured every 2 h.

Determination of in Vivo 5-Fluorocytosine Deaminase Activity. The *E. coli* pyrimidine auxotroph was transformed with pBAD322C and the plasmids containing the genes for cytosine deaminase (CodA, b0337), Kpn00326, and the D314A mutant of cytosine deaminase from *E. coli*. Cells were grown on LB-agar plates in the presence of 25 μg/mL chloramphenicol. Single colonies were picked and overnight cultures (5 mL) were grown in selection medium containing 150 μM cytosine, 25 μg/mL chloramphenicol, 100 μM $MnSO_4$, and 100 μM zinc acetate in 100 mL flasks prepared with the following experimental conditions: (a) 100 μM arabinose, (b) 50 μM 5-fluorocytosine and 100 μM arabinose (c) 500 μM 5-fluorocytosine and 100 μM arabinose, (d) 5.0 μM 5-fluorouracil and 100 μM arabinose. The cultures were inoculated with 25 μL of the overnight cultures and shaken at 200 rpm for 12 h at 37 °C. The absorbance at 600 nm was measured every 2 h.

Crystallization and Data Collection. Crystals of Kpn00632 from *Klebsiella pneumoniae* liganded with Fe^{2+} and 5-methylcytosine were grown by the sitting drop method at room temperature. The protein solution contained Kpn00632 (18.6 mg/mL) in 20 mM HEPES (pH 7.5), 200 mM imidazole, 250 mM NaCl, 250 mM ammonium sulfate, 1.0 mM $FeCl_2$, and 40 mM 5-methylcytosine. The precipitant contained 20% PEG 3350, 0.15 M malic acid (pH 7.0) and 1.0 mM $FeCl_2$. Crystals appeared in 2 weeks and exhibited diffraction consistent with the space group $P2_1$, with six polypeptides per asymmetric unit. Crystals of Kpn00632 liganded with Fe^{2+} and 5-fluorocytosine were also grown by the sitting drop method at room temperature. The protein solution contained enzyme (18.6 mg/mL) in 20 mM HEPES (pH 7.5), 200 mM imidazole, 250 mM NaCl, 250 mM ammonium sulfate, 1.0 mM $FeCl_2$, and 60 mM 5-fluorocytosine; the precipitant contained 30% PEG 3350, 0.1 M sodium citrate (pH 5.6), 0.2 M ammonium acetate, and 1.0 mM $FeCl_2$. Crystals appeared in 5 days and exhibited diffraction consistent with the space group $P2_12_12_1$, with six polypeptides per asymmetric unit. Crystals of Kpn00632 liganded with Fe^{2+} and phosphonocytosine[21] were grown by

dx.doi.org/10.1021/bi5012767 | *Biochemistry* 2014, 53, 7426−7435

Table 1. Data Collection and Refinement Statistics for Kpn00632

	5-methylcytosine	5-fluorocytosine	phosphonocytosine
data collection			
space group	$P2_1$	$P2_12_12_1$	$P2_12_12_1$
molecules in asym. unit	6	6	6
cell dimensions			
a (Å)	117.477	101.957	102.169
b (Å)	137.665	162.430	147.807
c (Å)	112.096	184.051	185.327
β°	118.03		
resolution (Å)	1.80	2.0	1.90
unique reflections	265794	198447	214497
R_{merge}	0.089	0.091	0.114
completeness (%)	96.8	96.07	97.60
refinement			
resolution (Å)	25.0−1.80	25.0−2.0	25.0−1.90
R_{cryst}	0.159	0.163	0.198
R_{free}	0.188	0.202	0.243
no. atoms			
protein	19377	19416	19334
waters	1743	1160	1523
ligand RMS deviations	234	254	195
bond lengths (Å)	0.007	0.007	0.007
bond angles (deg)	1.06	1.05	1.05
PDB entry	4R85	4R88	4R7W

the sitting drop method at room temperature. The protein solution contained enzyme (15 mg/mL) in 20 mM HEPES (pH 7.5), 1.0 mM $FeCl_2$, and 10 mM phosphonocytosine. The precipitant contained 20% PEG 3350, 0.15 M malic acid (pH 7.0), and 1.0 mM $FeCl_2$. Crystals appeared in 2 days and exhibited diffraction consistent with the space group $P2_12_12_1$, with six polypeptides per asymmetric unit.

Prior to data collection, crystals of the three forms of Kpn00632 were transferred to cryoprotectant solutions composed of their mother liquids supplemented with 20% glycerol, and flash-cooled in a nitrogen stream. Three X-ray diffraction data sets were collected at the NSLS X29A beamline (Brookhaven National Laboratory) on the 315Q CCD detector. Diffraction intensities were integrated and scaled with programs DENZO and SCALEPACK.[31] The data collection statistics are given in Table 1.

Structure Determination and Model Refinement. The three Kpn00632 structures were determined by molecular replacement with BALBES,[32] using only input diffraction and sequence data. BALBES used the structure of cytosine deaminase from *E. coli* complexed with phosphonocytosine (PDB id: 3O7U) as the search model. Partially refined structures of the three Kpn00632 crystal forms were generated by BALBES. Subsequent iterative cycles of refinement were performed for each crystal form including model rebuilding with COOT,[33] refinement with PHENIX,[34] and automatic model rebuilding with ARP.[35] The quality of the final structures was verified with omit maps. The stereochemistry was checked with WHATCHECK[36] and MOLPOBITY.[37] Program LSQKAB[38] was used for structural superposition. Figures with electron density maps were prepared using PYMOL.[39] The final model of the monoclinic crystal form of Kpn00632 with 5-methylcytosine contains residues 1−411 of the enzyme in all six polypeptides located in the asymmetric unit. The C-terminal His-tag residues are not included in the final model. The ferrous ion and ligand are well-defined in the active site of every monomer.

The final models of the two orthorhombic crystal forms of Kpn00632 with 5-fluorocytosine and phosphonocytosine in the active site contain residues 1−412 of the enzyme in all six polypeptides of these crystal forms. The C-terminal His-tag residues have weak electron density and are not included in the final models. The ferrous ions and the corresponding ligand molecules are well-defined and bound in the active sites of the corresponding structures. The final crystallographic refinement statistics for all three structures of Kpn00632 are provided in Table 1.

RESULTS

Phylogeny of the Cytosine Deaminase Group from cog0402. A BLAST search using the sequence of cytosine deaminase from *E. coli* (b0337, CodA) was submitted, and 1377 sequence homologues were identified. An all-by-all BLAST was subsequently performed with these protein sequences to create a sequence similarity network (SSN) at an *E*-value stringency of 10^{-140}.[40,41] At this level of protein sequence identity, three major subgroups and a number of minor subgroups could be identified as illustrated in Figure 2 and four representative proteins, one from each major cluster and one from a single minor cluster, were selected for functional characterization. These proteins included cystosine deaminase (CodA, b0337) from *E. coli* (subgroup-*a*); Kpn00632 from *Klebsiella pneumonia* (subgroup-*b*); Rsp0341 from *Rhodobacter sphaeroides* 2.4.1 (subgroup-*c*); and NCgl0075 from *Corynebacterium glutamicum* ATCC 13032 (subgroup-*d*). The sequence identity between any two proteins ranged from 35% (Rsp0341 and NCgl0075) to 58% (CodA and Kpn00632).

An amino acid sequence alignment of the proteins contained within each of the four groups of proteins selected for this investigation identified a striking difference in the amino acid residue that immediately follows the critical aspartate residue at the C-terminal end of β-strand 8. In the amidohydrolase superfamily (AHS) of enzymes, this invariant aspartate residue

dx.doi.org/10.1021/bi5012767 | *Biochemistry* 2014, 53, 7426−7435

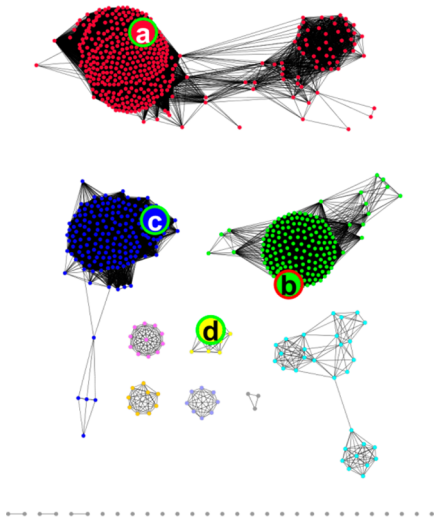


Figure 2. Sequence similarity network (SSN) diagram of CodA homologues at a BLAST *E*-value cutoff of 10^{-140}. Groups are labeled by their representative purified protein: (*a*) CodA (b0337); (*b*) Kpn00632; (*c*) Rsp0341; and (*d*) NCgl0057.

(Asp-313 in *E. coli* CodA) coordinates one of the divalent cations in the active site and functions in proton transfer reactions.[21] In subgroup-*a*, the residue that follows the invariant aspartate is also an aspartate, whereas in subgroups-*b*, -*c*, and -*d* this residue is a serine, cysteine, and serine, respectively (Figure 3).

Enzymatic Characterization. The genes for the four enzyme targets were cloned and expressed in *E. coli*, and the proteins purified to homogeneity. The kinetic constants for the purified proteins with a series of 10 potential substrates were determined at pH 7.5 and the results are presented in Tables 2 and 3. The prototypical cytosine deaminase from *E. coli* (CodA, b0337) was able to deaminate cytosine (**1**) and isoguanine (**7**) at appreciable rates (k_{cat}/K_m = 8.4 × 10^4 and 1.1 × 10^5 M^{-1} s^{-1}, respectively), whereas the ability to deaminate 5-methylcytosine (**2**) was barely detectable (k_{cat}/K_m = 2.2 × 10^1 M^{-1} s^{-1}). Kpn00632 deaminated 5-methylcytosine (**2**) approximately 4 orders-of-magnitude more efficiently than CodA from *E. coli* (k_{cat}/K_m = 3.3 × 10^5 M^{-1} s^{-1}). This enzyme also utilized 5-fluorocytosine (**4**) and 5-aminocystosine (**5**) as substrates with k_{cat}/K_m values greater than 10^5 M^{-1} s^{-1}. Rsp0341 was determined to have a similar catalytic profile as Kpn00632, except that this enzyme preferentially deaminates cytosine (**1**) relative to 5-methylcytosine (**2**). The best substrate for NCgl0075 was creatinine (k_{cat}/K_m = 6.3 × 10^4 M^{-1} s^{-1}). The D314A mutant of CodA of *E. coli* shared a similar substrate profile with Kpn00632, including the dramatic increase in the catalytic activity using 5-methylcytosine (k_{cat}/K_m = 9.7 × 10^4 10 M^{-1} s^{-1}). As previously reported, this enzyme also

```
CodA          1 SNNALQTIINARLPGEEGLWQIHLQDGKISAIDAQSGVMPIT--ENSLDAEQGLVIPPFVEPHIHLDTTQTAGQPNWNQSGTLFEGIE 86
Kpn00632      1 -----MKIINARLRRQEALFTLDLQDGIIHRITAQAAMQTAD--AGAIDAQGRLAIPPFVEPHIHLDATLTAGEPEWNRSGTLFEGIT 81
Rsp0371       1 ---MFDLIVKGGTLPDGRVADVGIRGDRIAAIGALGTDA-----ARLIEATGDLVSPAFVDPHFHMDATLSYGLPRVNASGTLLEGIR 80
NCgl0075      1 -----MRITNAQVKNYAELVDITIEGEKISSITPSSIRSEEDHRADDYDAAGRLVAPQFAEAHIHLDYANTAGIPRENSSGTLFEAIE 83

CodA         87 RWAERKA--LLTHDDVKQRAWQTLKWQIANGIQHVRTHVDVSDATLTALKAMLEVKQEVAPWIDLQIVAFPQEGILSYPNGEALLEEA 172
Kpn00632     82 RWSQRKA--SITPEDTRQRALKTIGMLRDFGVQHVRTHVDVTDPSLAALQALLAVKQEAADLIDLQIVAFPQEGIESYPNGRELMTRA 167
Rsp0371      81 LWGELKE--IVTVEAMVERALAYCDWAASMGLLAVRTHVDVCDDRLLGVEAMLAVREKVKDWMDLQLVAFPQDGLYRDPTARANLLRA 166
NCgl0075     84 IWADRKTQGFHIKEDIKAKALQAARRAAEHGVGFIRTHVDVTDPTFAGFEAIAELRDEVREWCDIQIVAFPQNGIYAYEGGQKLISDA 171

CodA        173 LRLGADVVGAIPHFEFTREYGVESLHKTFALAQKYDRLIDVHCDEIDDEQSRFVETVAALAHHEGMGARVTASHTTAMHSYNGAYTSR 260
Kpn00632    168 IEMGADVVGGIPHYENTRDKGVSSVMFLMDLAQRYGRLVDVHCDEIDDPQSRFLEVLAEEARVRGMGAQVTASHTCAMGSYDNAYCSK 255
Rsp0371     167 LDMGVDVVGGIPHFERTMADGAASVRDLCEIAADRGLPIDFHCDETDDPLSRHIETYAAEVLRTGLQGRAAAGHLTSMHSMDNYYVSK 254
NCgl0075    172 MSAGADVVGGIPHLEPTRDDGVESVKWLFDLAEKHSAPIDIHTDEIDDPHSRFVEVLAAEAAKRDMGAQTVVSHSVAMAYYSPGYMAR 259

CodA        261 LFRLLLKMSGINFVANPLVNIHLQG-RFDTYPKRRGITRVKEMLESGINVCFGHDDVFDPWYPLGTANMLQVLHMGLHVCQLMGYGQIN 347
Kpn00632    256 LFRLLLKASGINFISCPTESIHLQG-RFDSWPKRRGVTRVAELDRAGINVCFAQDSIQDPWYPLGNGNILRILDAGLHICHMLGYDDLQ 342
Rsp0371     255 LLPLIAEAGIAAIPNPLINIVLQG-RHDSFPKRRGLTRVKEMQAMGITVGWGQDCVLDPWYSLGTADMLDVAFMGLHVAQMTHPDEMR 341
NCgl0075    260 LLPKLAASKVRFAVCPNENLHLQGLGFQG-PVPRGVAPVKQLTEWGIPVSFCQDSLNDPFYPMGDGDLLRILDSGLHVSHMLTASHLK 346

CodA        348 DGLNLITHHSARTLNLQD-YGIAAGNSANLIILPAENGFDALRRQVPVRYSVRGGKVIASTQPAQTTVYLE-QPEAIDYKR-----  426
Kpn00632    343 RCLDFVTDNSARALCLGDNYGLAEGRPANLLILDAENDYEAVRRQARVLTSIRHGKVILQ----------R-EVEHIRYPA-----  412
Rsp0371     342 RCFDMVTVENARIMGLD--YGLREGAVASLVVLDAGNPVEALRLRPDRLCVIAKGQVVSEKARNDARLSLPGRPETVRRRHLLPQR  425
NCgl0075    347 NALSFITTNPAGNLGLDN-YDIAENSPANLLVLDASSEKEAVQRKASVLLSIHRGKKVLS----------R-EPEQVDWNI-----  415
```

Figure 3. Sequence alignment of CodA, Kpn00632, Rsp0371, and NCgl0075. Residues presented in Figures 1, 4, and 5 are highlighted.

Table 2. Values of k_{cat}/K_m for Enzymes Purified for This Investigation (M^{-1} s^{-1})[a]

substrate	CodA	CodA-D314A	Kpn00632	Rsp0341	NCgl0075
cytosine	8.4 (0.9) × 10^4	2.2 (0.2) × 10^4	2.9 (0.2) × 10^4	1.4 (0.2) × 10^5	1.1 (0.2) × 10^3
5-methylcytosine	2.2 (0.1) × 10^1	9.7 (0.7) × 10^4	3.3 (0.6) × 10^5	2.0 (0.3) × 10^4	1.5 (0.1) × 10^3
creatinine	1.3 (0.1) × 10^2	2.8 (0.6) × 10^1	6.8 (1.3) × 10^3	1.3 (0.1) × 10^2	6.3 (0.3) × 10^4
isoguanine	1.1 (0.1) × 10^5	1.9 (0.2) × 10^4	3.8 (0.5) × 10^4	1.4 (0.1) × 10^2	5.9 (0.3) × 10^2
5-fluorocytosine	2.5 (0.5) × 10^2	9.9 (0.7) × 10^3	1.2 (0.1) × 10^5	6.8 (0.5) × 10^4	2.0 (0.2) × 10^2
5-aminocytosine	8.0 (0.5) × 10^3	9 (1) × 10^4	5.8 (0.3) × 10^5	<1 × 10^1	1.0 (0.1) × 10^3
N6-methylcytosine	4.0 (0.5) × 10^1	<1 × 10^1	<1 × 10^1	<1 × 10^1	<1 × 10^1
5-hydroxymethylcytosine	<1 × 10^1	2.3 (0.1) × 10^3	2.2 (0.2) × 10^3	<1 × 10^1	<1 × 10^1

[a]30 °C, pH 7.5.

dx.doi.org/10.1021/bi5012767 | *Biochemistry* 2014, 53, 7426−7435

Table 3. Values of k_{cat} for Enzymes Purified for this Investigation (s^{-1})[a]

substrate	CodA	CodA-D314A	Kpn00632	Rsp0341	NCgl0075
cytosine	33 ± 7	15 ± 2	13 ± 2	30 ± 5	0.35 ± 0.05
5-methylcytosine	ND[b]	14 ± 1	8 ± 1	4.8 ± 0.8	1.3 ± 0.2
creatinine	0.11 ± 0.03	0.012 ± 0.004	ND	ND	32 ± 4
isoguanine	3.9 ± 0.1	0.68 ± 0.02	4.5 ± 0.5	0.035 ± 0.003	0.23 ± 0.04
5-fluorocytosine	ND	>17	6.5 ± 0.2	25 ± 3	0.055 ± 0.005
5-aminocytosine	ND	24 ± 3	160 ± 10	ND	ND
*N*6-methylcytosine	0.0017 ± 0.0001	ND	ND	ND	ND
5-Hydroxymethylcytosine	ND	>3	ND	ND	ND

[a]30 °C, pH 7.5. [b]ND: not determined.

exhibited a significant increase in the rate of deamination of 5-fluorocytosine, relative to the wild-type enzyme (9.9×10^3 M^{-1} s^{-1}).[27]

Three-Dimensional Structure of Kpn00632. The three-dimensional structure of Kpn00632 was determined in the presence of 5-methylcytosine (**2**) (PDB id: 4R85), 5-fluorocytosine (**4**) (PDB id: 4R88), and phosphonocytosine (PDB id: 4R7W). In each case, the enzyme adopts a distorted $(\beta/\alpha)_8$-barrel with a mononuclear metal center at the C-terminal end of the β-barrel. In the complexes formed with 5-methylcytosine (Figure 4A), 5-fluorocytosine (Figure 4B), and phosphonocytosine[21] (Figure 4C), the metal ion is coordinated to His-58 and His-60 from the C-terminal end of β-strand 1, His-209 from the C-terminal end of β-strand 5, and Asp-308 from the C-terminal end of β-strand 8. In all three structures, the substrates and inhibitor form a pair of hydrogen bonds between the carboxamide functional group and the side chain of Gln-151 (Figures 5 A-C). Additionally, Glu-212 interacts with the —NH═C—NH_2 moiety contained within each of these three ligands.

Rescue of Thymine Auxotrophs with 5-Methylcytosine. Strains of *E. coli* that lack thymidylate synthase cannot grow without the addition of exogenous thymine. These cells cannot grow in the presence of added 5-methylcytosine, since the wild-type cytosine deaminase of *E. coli* lacks the ability to deaminate 5-methylcytosine to thymine.[21] The *E. coli* thymine auxotroph was transformed with pBAD322c vectors containing the genes for b0337, Rsp0341, and Kpn00632 to determine whether the expression of enzymes capable of deaminating 5-methylcytosine to thymine was sufficient to enable *E. coli* to grow in the absence of added thymine. All four examples were capable of growth when supplemented with 100 μM thymine (Figure 6). However, neither the empty vector nor additional CodA from *E. coli* was capable of rescuing growth when supplemented with 400 μM 5-methylcytosine (Figure 6A and B). Expression of either Kpn00632 or Rsp0341 allowed growth in the presence of 400 μM 5-methylcytosine, and to a lesser extent in the presence of 100 μM added 5-methylcytosine for Kpn00632 (Figure 6C and D).

Enhanced Toxicity of 5-Fluorocytosine. The ability of Kpn00632 and the D314A mutant of CodA of *E. coli* to deaminate 5-fluorocytosine to 5-fluorouracil in vivo was tested in pyrimidine auxotrophs, lacking pyrF. This cell line will phosphoribosylate exogenous pyrimidines and in the presence of 5-fluorouracil will form 5-fluorouridine monophosphate, a potent inhibitor of thymidylate synthase. This strain of *E. coli* cannot grow in the presence of 5 μM 5-fluorouracil (Figure 7). The growth rate is not affected by the addition of 50 μM 5-fluorocytosine, but is measurably reduced in the presence of 500 μM 5-fluorocytosine (Figure 7A). Overexpression of the wild-type cytosine deaminase from *E. coli* in these cells reduces the growth rate in the presence of 50 and 500 μM 5-fluorocytosine (Figure 7B) and is reduced even further when the D314A mutant of CodA is expressed (Figure 7C). Expression of Kpn00632 in these cells completely abolished growth in the presence of 500 μM added 5-fluorocytosine and

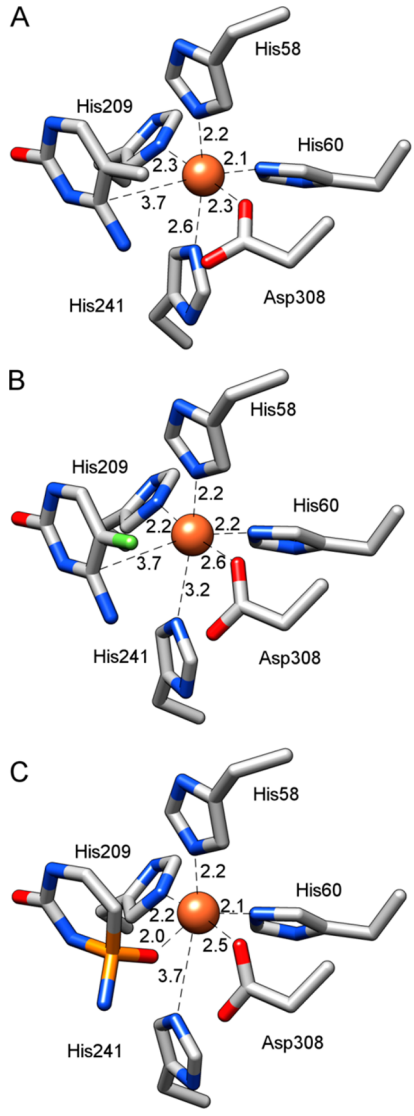


Figure 4. Metal center of Kpn00632 with various ligands bound in the active site. (A) 5-methylcytosine (PDB id: 4R85); (B) 5-fluorocytosine (PDB id: 4R88; and (C) phosphonocytosine (PDB id: 4R7W).

dx.doi.org/10.1021/bi5012767 | *Biochemistry* 2014, 53, 7426−7435

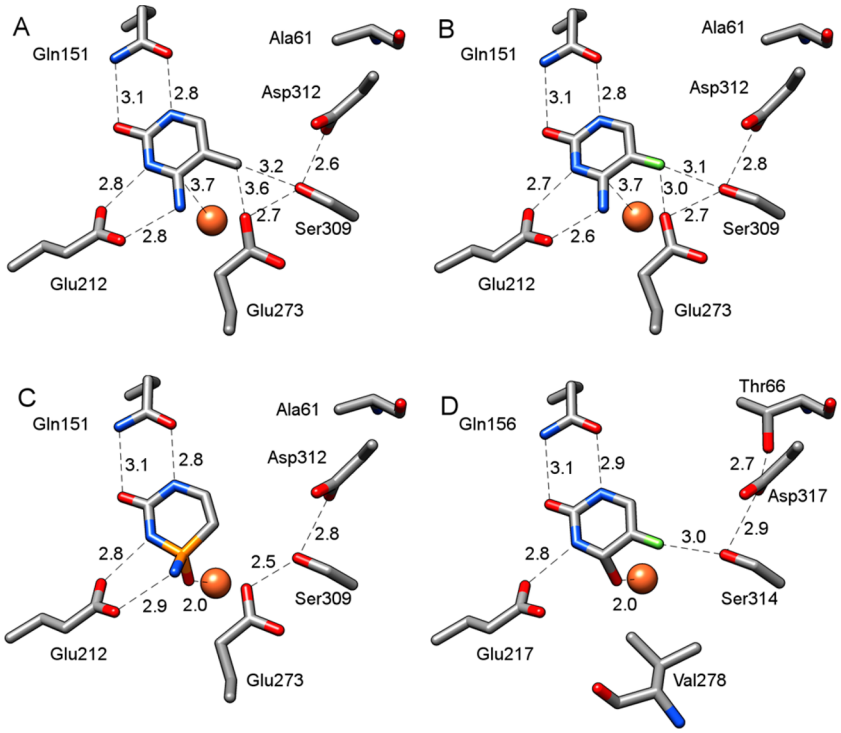


Figure 5. Active site and ligand binding residues of Kpn00632 and the D314S mutant of CodA from *E. coli*. (A) Kpn00632 bound with 5-methylcytosine; (B) Kpn00632 bound with 5-fluorocytosine; (C) Kpn00632 bound with phosphonocytosine; and (D) CodA-D314S (PDB: 1RAK) bound with 5-fluoro-4-*S*-hydroxy-3,4-dihydropyrimidine.

severely inhibited growth in the presence of 50 μM added 5-fluorocytosine (Figure 7D).

DISCUSSION

The methylation of the nucleobase cytosine is a common epigenetic modification of DNA in both eukaryotes and prokaryotes.[1,6] However, certain bacterial enzymes apparently actively exclude this metabolite from the active site. For example, wild-type cytosine deaminase from *E. coli* (CodA) catalyzes the deamination of 5-methylcytosine with a value of k_{cat}/K_m of ~22 M^{-1} s^{-1}. This rate constant is more than 3 orders of magnitude smaller than the value of k_{cat}/K_m for the deamination of cytosine (~10^5 M^{-1} s^{-1}) by the same enzyme. Since significant quantities of 5-methylcytosine are produced in bacterial cells during the modification of DNA, a metabolic pathway must exist for the catabolism of this compound. In our search to identify candidate enzymes that would be capable of deaminating 5-methylcytosine to thymine, we assumed that these enzymes would be quite similar in structure and sequence to the cytosine deaminase from *E. coli*. A similar approach has previously led to the successful identification of the first enzyme capable of deaminating 8-oxoguanine to uric acid.[26]

A search for sequence homologues to CodA from *E. coli* identified approximately 1400 candidate sequences. These sequences cluster into three major groups and numerous minor groups at an E-value threshold of 10^{-140} (Figure 2). The genes for three proteins, in addition to CodA from *E. coli*, were cloned and expressed, and the proteins purified to homogeneity. Kpn00632 from *Klebsiella pneumoniae* efficiently catalyzes the deamination of 5-methylcystosine to thymine with a value of k_{cat}/K_m that exceeds 10^5 M^{-1} s^{-1}. This rate constant is 4 orders of magnitude greater than the value of k_{cat}/K_m for the deamination of 5-methylcytosine by CodA from *E. coli*. Similar results were obtained with Rsp0341 from *Rhodobacter sphaeroides*.

The three-dimensional structure of Kpn00632 was determined in the presence of 5-methylcytosine, 5-fluorocytosine, and the phosphonate mimic of the putative tetrahedral intermediate bound in the active site of the enzyme. In the vicinity of the methyl- and fluoro-substituents at C5 of the bound ligands, the closest two residues are Glu-273 and Ser-309. In CodA from *E. coli* these residues correspond to Val-278 and Asp-314, respectively. These two residue positions are highly conserved within each of the three major subgroups identified in Figure 2. Sequence alignments indicate that these two residue positions correspond to isoleucine and cysteine in Rsp0371, and with glutamate and serine in NCgl0075.

Wild-type CodA from *E. coli* is essentially unable to catalyze the deamination of 5-methylcytosine. In Kpn00632, the residue that follows Asp-308 is a serine. When Asp-314 in *E. coli* is mutated to alanine, this enzyme is now capable of deaminating 5-methylcytosine, thus demonstrating that this residue is largely responsible for the discrimination between 5-methylcytosine and cytosine in the active site. These changes in the active site also enable these enzymes to deaminate 5-fluorocytosine to 5-fluorouracil, a highly toxic metabolite. Similar results have previously been observed by Mahan et al. when they demonstrated that the D314S mutation in CodA of *E. coli* enhanced the turnover ratio of 5-fluorocytosine/cytosine by 4-fold.[27]

The addition of Kpn00632 enables the *E. coli* thymine auxotroph to grow in the presence of 5-methylcytosine. The thymine auxotroph cannot produce thymidine, since it lacks thymidylate synthase and thus has no way to make thymine since wild-type CodA cannot effectively catalyze the deamination of 5-methylcytosine to thymine. In the presence of Kpn00632, these cells can catalyze the formation of thymine

dx.doi.org/10.1021/bi5012767 | *Biochemistry* 2014, 53, 7426−7435

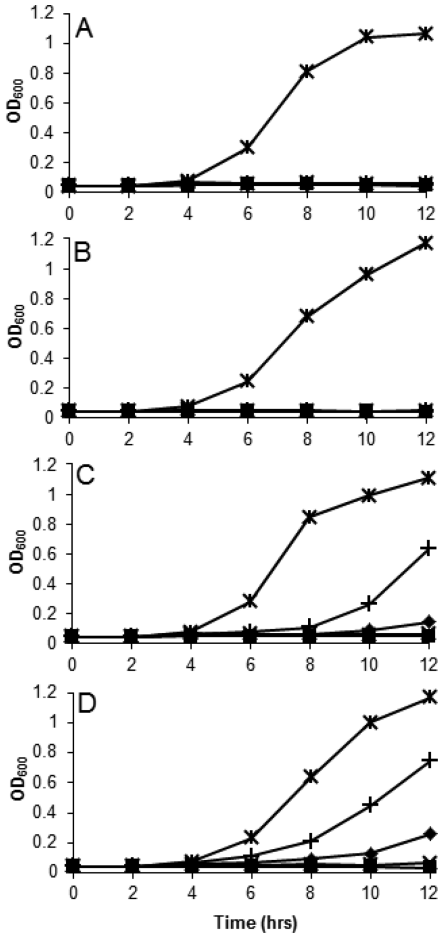


Figure 6. Growth of *E. coli* thymine auxotrophs supplemented with 5-methylcystosine and enzymes capable of deaminating 5-methyl cytosine to thymine. The experimental conditions are as follows: 400 μM 5-methylcytosine (×); 100 μM arabinose (■); 100 μM arabinose and 100 μM thymine (*); 100 μM arabinose and 100 μM 5-methylcytosine (◆); and 100 μM arabinose and 400 μM 5-methylcytosine (+). (A) Empty pBAD322c vector; (B) CodA from *E. coli*; (C) Rsp0341; and (D) Kpn00632.

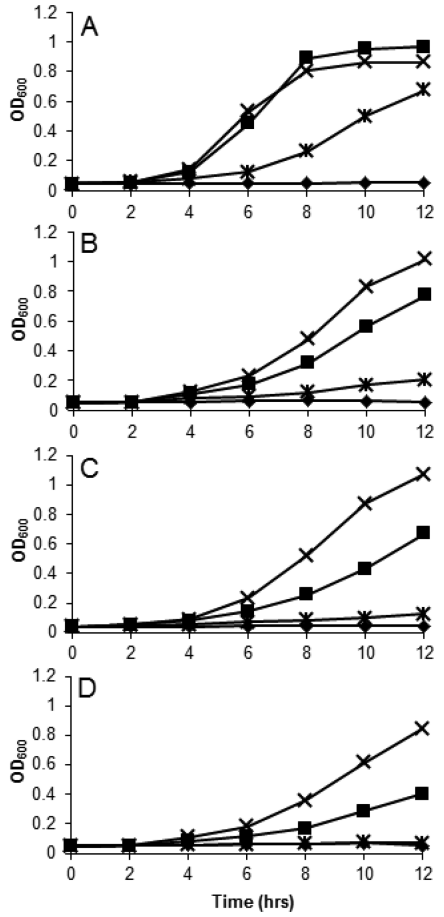


Figure 7. Toxicity of 5-fluorocytosine to *E. coli* in the presence of enzymes capable of deaminating 5-fluorocytosine to 5-fluorouracil. The experiment conditions include the following: induction with 100 μM arabinose (×); 100 μM arabinose and 50 μM 5-fluorocytosine (■); 100 μM arabinose and 500 μM 5-fluorocystosine (*); 100 μM arabinose and 5 μM 5-fluorocystosine (◆). (A) Empty pBAD322c vector; (B) CodA from *E. coli*; (C) CodA-D314A mutant; and (D) Kpn00632.

from 5-methylcytosine and thymidylate can be made from the phosphoribosyltransferase reaction (Figure 6). 5-Fluorocytosine is not normally toxic to *E. coli*, since wild type CodA cannot catalyze the formation of 5-fluorouracil.[27] 5-Fluorouracil is a toxic metabolite that irreversibly inactivates thymidyldate synthase. When Kpn00632 is expressed in *E. coli*, 5-fluorocytosine becomes toxic as exhibited by the substantial retardation of growth (Figure 7).

The discovery of enzymes capable of deaminating 5-methylcytosine and 5-fluorocytosine reveals the difficulty of defining the substrate/sequence boundaries of enzymes based on simple sequence similarity to proteins of known function. Essentially all of the sequences depicted in the SSN of Figure 2 have been annotated as cytosine deaminase. Given the inability of CodA from *E. coli* to deaminate 5-methylcytosine, these enzymes would have been predicted to not catalyze the deamination of either 5-methylcytosine or 5-fluorocytosine. However, a careful examination of the residues that reside in the active site reveals a significant perturbation of a conserved aspartate residue to either a serine or cysteine residue.

Full-length sequence alignments indicate that the subgroups depicted in Figure 2 are highly specific for the amino acids that populate the active site. *E. coli* CodA is found in subroup-*a* and most sequences in this subgroup possess an aspartate residue corresponding to Asp-314. Subgroup-*b* (including Kpn00632) conserves a serine residue at this position and is expected to occupy the same role as Ser-309 in Kpn00632. Subgroup-*c* and Rsp0341 strongly favor a cysteine residue at this position. Finally, subgroup-*d*, previously characterized as creatinine deaminase, possesses a serine residue at this position. While 5-methylcytosine is not known to be a metabolite produced in great quantities for cell proliferation, it seems reasonable to assume there is an advantage for having an enzyme that can catalyze the deamination of the free nucleobase. As demonstrated by the *in vivo* experiment in Figure 6, it is possible to produce thymine by this route. It remains mysterious why the cytosine deaminase from *E. coli* has evolved to exclude 5-methylcytosine from the active site.

The original evidence that homologues of CodA from *E. coli* may have promiscuous activity for the deamination of 5-methylcytosine is based on mutagenesis studies that produced a more efficient 5-fluorocytosine deaminase, with the ultimate goal of transfecting cancer cells with this enzyme.[27] In suicide gene therapy, the gene for an enzyme capable of deaminating the nontoxic prodrug 5-fluorocytosine is delivered

dx.doi.org/10.1021/bi5012767 | *Biochemistry* 2014, 53, 7426−7435

to cancer cells.[42] The expressed deaminase subsequently converts 5-fluorocytosine to 5-fluorouracil, which is ultimately transformed to 5-fluorouridine monophosphate, an irreversible inhibitor of thymidylate synthase. DNA replication is ultimately blocked due to the inhibition of deoxythymidine triphosphate synthesis.

The D314A/S/G mutants created by Mahan et al. each showed cytotoxicity in the presence of 5-fluorocytosine in an experiment similar to that presented in Figure 7.[27] However, Kpn0062 deaminates 5-fluorocytosine with a value of k_{cat}/K_m that is greater than that of any of the D314 mutants of CodA reported previously.[27] Kpn00632 also deaminates 5-fluorocytosine more efficiently than it does cytosine. This enzyme may therefore provide a novel starting point for the creation of even better enzymes for the deamination of 5-fluorocytosine to 5-fluorouracil.

AUTHOR INFORMATION

Corresponding Authors

*Telephone: 979-845-3373. E-mail: raushel@tamu.edu.

*Telephone: 718-430-2746. E-mail: steve.almo@einstein.yu.edu.

Funding

This work was supported in part by the Robert A. Welch Foundation (A-840) and the National Institutes of Health (GM 71790).

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

5-FC, 5-fluorocytosine; 5-FU, 5-fluorouracil; COG, cluster of orthologous groups; PDB, Protein Data Bank; CGSC, Coli Genetic Stock Center at Yale University; PMSF, phenylmethylsulfonyl fluoride; BLAST, Basic Local Alignment Search Tool; SSN, sequence similarity network

REFERENCES

(1) Suzuki, M. M., and Bird, A. (2008) DNA methylation landscapes: Provocative insights from epigenomics. *Nat. Rev. Genet. 9*, 465−476.

(2) Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol. 203*, 971−983.

(3) Pavlopoulou, A., and Kossida, S. (2007) Plant cytosine-5 DNA methyltransferases: Structure, function, and molecular evolution. *Genomics 90*, 530−541.

(4) Jabbari, K., and Bernardi, G. (2004) Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene 333*, 143−149.

(5) Gomez-Eichelmann, M. C., Levy-Mustri, A., and Ramirez-Santos, J. (1991) Presence of 5-methylcytosine in CC(A/T)GG sequences (Dcm methylation) in DNAs from different bacteria. *J. Bacteriol. 173*, 7692−7694.

(6) Kahramanoglou, C., Prieto, A. I., Khedkar, S., Haase, B., Gupta, A., Benes, V., Fraser, G. M., Luscombe, N. M., and Seshasayee, A. S. N. (2012) Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription. *Nat. Commun. 3*, 886.

(7) He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., and Xu, G.-L. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science 333*, 1303−1307.

(8) Wu, S. C., and Zhang, Y. (2010) Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol. 11*, 607−620.

(9) Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science 324*, 930−935.

(10) Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., and Zhang, Y. (2011) TET proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science 333*, 1300−1303.

(11) Katsuragi, T., Sakai, T., Matsumoto, K. y., and Tonomura, K. (1986) Cytosine deaminase from *Escherichia coli*-production, purification, and some characteristics. *Agric. Biol. Chem. 50*, 1721−1730.

(12) Kream, J., and Chargaff, E. (1952) On the cytosine deaminase of yeast1. *J. Am. Chem. Soc. 74*, 5157−5160.

(13) Cohen, R. M., and Wolfenden, R. (1971) Cytidine Deaminase from Escherichia coli: Purification, properties, and inhibition by the potential transition state analog 3,4,5,6-tetrahydrouridine. *J. Biol. Chem. 246*, 7561−7565.

(14) Weiner, K. X., Weiner, R. S., Maley, F., and Maley, G. F. (1993) Primary structure of human deoxycytidylate deaminase and overexpression of its functional protein in *Escherichia coli*. *J. Biol. Chem. 268*, 12983−12989.

(15) Hou, H.-F., Liang, Y.-H., Li, L.-F., Su, X.-D., and Dong, Y.-H. (2008) Crystal structures of *Streptococcus mutans* 2′-deoxycytidylate deaminase and its complex with substrate analog and allosteric regulator dCTP·Mg^{2+}. *J. Mol. Biol. 377*, 220−231.

(16) Johansson, E., Fanø, M., Bynck, J. H., Neuhard, J., Larsen, S., Sigurskjold, B. W., Christensen, U., and Willemoës, M. (2005) Structures of dCTP deaminase from *Escherichia coli* with bound substrate and product: Reaction mechanism and determinants of mono-and bifunctionality for a family of enzymes. *J. Biol. Chem. 280*, 3051−3059.

(17) Wang, L., and Weiss, B. (1992) dcd (dCTP deaminase) gene of *Escherichia coli*: Mapping, cloning, sequencing, and identification as a locus of suppressors of lethal dut (dUTPase) mutations. *J. Bacteriol. 174*, 5647−5653.

(18) Vita, A., Amici, A., Cacciamani, T., Lanciotti, M., and Magni, G. (1985) Cytidine deaminase from *Escherichia coli* B. Purification and enzymic and molecular properties. *Biochemistry 24*, 6020−6024.

(19) Xie, K., Sowden, M. P., Dance, G. S. C., Torelli, A. T., Smith, H. C., and Wedekind, J. E. (2004) The structure of a yeast RNA-editing deaminase provides insight into the fold and function of activation-induced deaminase and APOBEC-1. *Proc. Natl. Acad. Sci. U. S. A. 101*, 8114−8119.

(20) Ireton, G. C., Black, M. E., and Stoddard, B. L. (2003) The 1.14 Å crystal structure of yeast cytosine deaminase: evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. *Structure 11*, 961−972.

(21) Hall, R. S., Fedorov, A. A., Xu, C., Fedorov, E. V., Almo, S. C., and Raushel, F. M. (2011) Three-dimensional structure and catalytic mechanism of cytosine deaminase. *Biochemistry 50*, 5077−5085.

(22) Holm, L., and Sander, C. (1997) An evolutionary treasure: Unification of a broad set of amidohydrolases related to urease. *Proteins: Struct., Funct., Bioinf. 28*, 72−82.

(23) Seibert, C. M., and Raushel, F. M. (2005) Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry 44*, 6383−6391.

(24) Gupta, N. K., and Glantz, M. D. (1985) Isolation and characterization of human liver guanine deaminase. *Arch. Biochem. Biophys. 236*, 266−276.

(25) Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K., and Raushel, F. M. (2007) Structure-based activity prediction for an enzyme of unknown function. *Nature 448*, 775−779.

(26) 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., and Raushel, F. M. (2010) The hunt for 8-oxoguanine deaminase. *J. Am. Chem. Soc. 132*, 1762−1763.

(27) Mahan, S. D., Ireton, G. C., Knoeber, C., Stoddard, B. L., and Black, M. E. (2004) Random mutagenesis and selection of *Escherichia coli* cytosine deaminase for cancer gene therapy. *Protein Eng., Des. Sel. 17*, 625−633.

(28) Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol. 2*, 0008.
(29) Cronan, J. E. (2006) A family of arabinose-inducible *Escherichia coli* expression vectors having pBR322 copy control. *Plasmid 55*, 152−157.
(30) Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene 77*, 51−59.
(31) Otwinowski, Z., and Minor, W. (1997) *Processing of X-ray diffraction data collected in oscillation mode*, pp 307−326, Elsevier, New York.
(32) Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2008) BALBES: A molecular-replacement pipeline. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 64*, 125−132.
(33) Emsley, P., and Cowtan, K. (2004) COOT: Model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr. D60*, 2126−2132.
(34) Adams, P. D., Afonine, P. V., Bunkoczi, 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., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 213−221.
(35) Lamzin, V. S., and Wilson, K. S. (1993) Automated refinement of protein models. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 49*, 129−147.
(36) Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) Errors in protein structures. *Nature 381*, 272.
(37) Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 12−21.
(38) Collaborative Computation Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 50*, 760−763.
(39) *The PyMOL Molecular Graphics System*, Version 1.5, Schrödinger, LLC, http://www.pymol.org.
(40) Atkinson, H. J., Morris, J. H., Ferrin, T. E., and Babbitt, P. C. (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One 4*, e4345.
(41) Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A. R., Vailaya, A., Wang, P.-L., Adler, A., Conklin, B. R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G. J., Ideker, T., and Bader, G. D. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc. 2*, 2366−2382.
(42) Freytag, S. O., Khil, M., Stricker, H., Peabody, J., Menon, M., DePeralta-Venturina, M., Nafziger, D., Pegg, J., Paielli, D., Brown, S., Barton, K., Lu, M., Aguilar-Cordova, A., and Kim, J. H. (2002) Phase I study of replication-competent adenovirus-mediated double suicide gene therapy for the treatment of locally recurrent prostate cancer. *Cancer Res. 62*, 4968−4976.