Cloning and Characterization of Rhodotorula glutinis Thymine Hydroxylase

ARTICLE *in* CHEMICAL RESEARCH IN TOXICOLOGY · MAY 2009

Impact Factor: 3.53 · DOI: 10.1021/tx8004482

Cloning and Characterization of *Rhodotorula glutinis* Thymine Hydroxylase

Jonathan W. Neidigh, Agus Darwanto, Adides A. Williams, Nathan R. Wall, and Lawrence C. Sowers*

Department of Basic Sciences, Loma Linda University School of Medicine, Alumni Hall for Basic Science, Room 101, 11021 Campus Street, Loma Linda, California 92350

Received November 24, 2008

Thymine hydroxylase (TH) is a member of the α-ketoglutarate-dependent nonheme iron dioxygenase family that includes a series of DNA repair proteins including alkB. Substantial interest in this family of enzymes derives from their capacity to modify DNA bases and precursors by oxidation. Previously, a sequence has been published for cloned *Rhodotorula glutinis* TH. However, the minimal reported activity of this enzyme, coupled with inconsistencies with previously published mass spectrometry data, compelled us to reexamine TH. The sequence reported here differs from the previously reported sequence at two amino acid positions and is consistent with previously reported mass spectrometry data. The cloned enzyme characterized in this report displayed substantial activity, indicating that the sequence differences are critical for activity. The substrate selectivity of TH against a series of pyrimidine analogues is consistent with that reported for the wild-type enzyme and, in part, explains the mode of selection of uracil analogues. A preliminary model of the active site has been constructed for the purposes of comparing TH with other members of this family. TH and alkB share in common the capacity to oxidize N-methyl groups. However, TH has the added capacity to oxidize the 5-methyl group of thymine, a property that is potentially important for enzymes that could act on DNA and modify DNA−protein interactions.

Introduction

The oxidation of thymine to 5-hydroxymethyluracil (HmU)[1] was initially identified in rat liver and leukocytes (*1, 2*). Subsequent investigation (*3*−*5*) in *Neurospora crassa* identified and characterized an enzyme, thymine hydroxylase (TH), that catalyzes the sequential oxidation of thymine to 5-carboxyuracil (CarboxyU) via the intermediate oxidation products HmU and 5-formyluracil (FoU). TH was found to require iron and α-ketoglutarate, making it a member of the nonheme iron α-ketoglutarate-dependent dioxygenases (*6, 7*). This is an important class of metabolic and catabolic enzymes found in a wide range of organisms including bacteria, yeast, plants, and humans (*8*).

In addition to studies of TH isolated from *N. crassa*, TH isolated from *Aspergillus nidulans* (*9, 10*) and *Rhodotorula glutinis* (*11*−*13*) has been characterized. In 1993, Stubbe and co-workers reported on studies of a partially purified fraction from the red mold *R. glutinis* (*14, 15*). In their studies, they provided biochemical characterization of the TH activity, as well as sequence information of the protein. An N-terminal sequence, VSSGIVPPINFEPFLSGTPEDKLATA, was identified by Edman degradation. An internal sequence was also identified in the protein catalytic center by reaction with a pyrimidine substrate analogue and mass spectrometry. The sequence of the internal trypsin-generated peptide was NSIAF?SNPSLR, where the unknown residue was proposed to be tyrosine (*14*). In subsequent studies, the unknown amino acid was identified as phenylalanine (*16*).

In 2005, Smiley and co-workers created a degenerate DNA primer based upon the N-terminal peptide identified by Stubbe and co-workers and used this primer to identify the TH sequence from a cDNA library derived from *R. glutinis* (*17*). The N-terminal amino acid sequence derived from the cloned TH was MVSSGIVPVINFEPFLSGTP, consistent with the peptide identified by Stubbe and co-workers, except that a valine replaces proline at position 9. An internal predicted peptide, NSIAFFSNPSLR, is also found within the DNA sequence of cloned TH, consistent with the internal peptide identified by Stubbe and co-workers. However, the internal peptide predicted by the sequence of Smiley et al. (*17*) is preceded by cysteine, not an arginine or lysine, and should therefore not give rise to the tryptic fragment observed by Stubbe and co-workers. Additionally, the enzyme cloned by Smiley et al. (*17*) had negligible activity for the conversion of thymine to HmU.

Upon the basis of their cloned sequence, Smiley et al. (*17*) were able to compare TH with the sequences of other known proteins. The DNA sequences of the related species *N. crassa* and *A. nidulans* are known (*18*−*20*), and both are reported to have TH activities (*4*−*7, 9, 10, 21*). Furthermore, the cloned TH sequence was sufficient to establish that the *R. glutinis* TH was a member of the α-ketoglutarate dioxygenase family, consistent with its known substrate properties. Enzymes within this dioxygenase family can utilize either an RxS or an RxN sequence motif for binding α-ketoglutarate. Although the THs from *N. crassa* and *A. nidulans* are found within the RxS motif

* To whom correspondence should be addressed. Tel: 909-558-4480. Fax: 909-558-4035. E-mail: lsowers@llu.edu.

[1] Abbreviations: TH, thymine hydroxylase; ATCC, American Type Culture Collection; MTBSTFA, N-methyl-N-[*tert*-butyldimethylsilyl]trifluoroacetamide; TBDMCS, *tert*-butyldimethylsilylchlorosilane; BCA, bicinchoninic acid; FPLC, fast protein liquid chromatography; PCR, polymerase chain reaction; GST, glutathion-S-transferase; BLAST, Basic Local Alignment Search Tool; NCBI, National Center for Biological Information; HmU, 5-hydroxymethyluracil; FoU, 5-formyluracil; CarboxyU, 5-carboxyuracil; 1MeU, N1-methyluracil; 1MeThy, N1-methylthymine; 1MeHmU, N1-methyl-5-hydroxymethyluracil; 5MeC, 5-methylcytosine; 3MeU, N3-methyluracil; 6MeU, 6-methyluracil.

10.1021/tx8004482 CCC: $40.75
Published on Web 04/03/2009

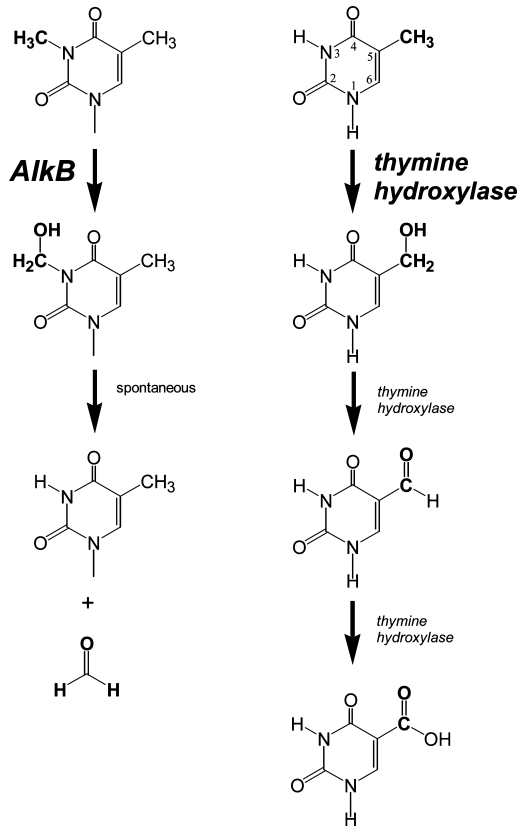


Figure 1. Reactions catalyzed by alkB (left) and TH (right). Each enzyme-catalyzed reaction also consumes one molecule of oxygen and α-ketoglutarate and produces one molecule of succinate and carbon dioxide.

subgroup, the RxN sequence of Smiley et al. (*17*) surprisingly placed the TH from *R. glutinis* within a different subgroup that contained the alkB family of DNA repair enzymes (*17*).

Recently, there has been substantial interest in the *Escherichia coli* protein, alkB, which repairs alkylation damage on DNA (*22*). Bioinformatic analysis of the alkB amino acid sequence by Aravind and Koonin predicted that alkB is a member of the nonheme iron α-ketoglutarate-dependent dioxygenases; subsequent verification of this prediction demonstrated that alkB oxidizes N-methyl groups spontaneously, regenerating the parent DNA base by release of formaldehyde (*23*−*26*).

This class of proteins is of interest for further study as it might reveal additional proteins involved in the modification of nucleic acids that impact the control of gene expression (Figure 1). A "demethylation" pathway for 5-methylcytosine (5MeC) was suggested based on biological observations, but no clear data have emerged to provide a definitive mechanism (*27*−*30*). One model suggests that a dioxygenase of the alkB family might act on 5MeC; however, no family member with this activity has yet been identified (*31*).

Upon the basis of apparent inconsistencies in data previously reported for *R. glutinis* TH (*14, 17*), in conjunction with mass spectrometry data presented in this manuscript, we sought to independently determine the sequence of TH. The derived protein sequence reported here shows differences with both previous reports but allows reconciliation of most of the data. The cloned enzyme with this new sequence demonstrates significant enzymatic activity, allowing the construction and testing of several site-directed mutants as well as examination of substrate preferences. A preliminary structural model has allowed comparison of the active site residues of *R. glutinis* TH with the alkB family of DNA repair enzymes. A previously unknown activity that converts thymine to HmU in telomere DNA regions of trypanosomatids has been recently identified (*32*) that shares several key active site residues, suggesting that further proteins that oxidize DNA bases are likely to be found.

Materials and Methods

Materials. The same strain (ATCC 2527) examined by Stubbe and co-workers and Smiley et al. (*14, 15, 17*) of *R. glutinis* was obtained from the American Type Culture Collection (ATCC, Manassas, VA) and was grown in bulk by Encore Technologies (Minnetonka, MN). Sodium ascorbate, α-ketoglutarate, ferrous sulfate, and HEPES buffer were obtained from Fisher Scientific (Tustin, CA). The silylating agent, N-methyl-N-[*tert*-butyldimethylsilyl]trifluoroacetamide (MTBSTFA) with 1% *tert*-butyldimethylsilylchlorosilane (TBDMCS), and the bicinchoninic acid (BCA) protein concentration assay kit were obtained from Pierce (Rockford, IL). Deuterium-enriched thymine-α,α,α,6-d4 (>98% enrichment) was obtained from Cambridge Isotope Laboratories (Andover, MA). Synthetic $^2H_2$-5-hydroxymethyluracil was prepared from $^2H_2$-formaldehyde and uracil as previously described (*33, 34*). Sequencing grade trypsin, uracil, thymine, FoU, HmU, CarboxyU, and all other laboratory reagents were obtained from Sigma-Aldrich (St. Louis, MO).

Isolation of TH from *R. glutinis.* *R. glutinis* was grown in media containing thymine as the only nitrogen source, essentially as described previously by Thornburg et al. (*14*) without significant modification. The primary difference is that we substituted a GC/MS method for the measurement of TH activity as described below. The purification steps involved ammonium sulfate precipitation, Sephadex G-25, DE-52, and Sephadex G-100 chromatography. The final monoQ fast protein liquid chromatography (FPLC) purification resulted in a loss of enzyme activity as previously observed (*14*). Protein concentrations were measured with the BCA kit using a Cary 100 spectrophotometer (Varian, Walnut Creek, CA).

Peptide Mass Spectrometry Analysis. Purified TH was reduced and alkylated prior to enzymatic digestion using 9.5 mM dithiothreitol and 32 mM iodoacetamide, respectively. The reduced and alkylated protein was digested with trypsin (1:50 ratio of protease to substrate) overnight at 37 °C in 0.6 M urea and 10 mM Tris buffer at pH 7.8. Tryptic peptides were analyzed with a ThermoFinnigan LCQ Deca XP mass spectrometer (Thermo, Waltham, MA) equipped with a PicoView 500 nanospray apparatus and a 10 cm × 75 μM internal diameter C18 Biobasic bead column (New Objective, Woburn, MA). Mobile phase A consisted of 2% aqueous acetonitrile with 0.1% formic acid, while mobile phase B consisted of 98% acetonitrile, 2% water, and 0.1% formic acid. Injected peptides were eluted with a linear gradient from 20 to 60% mobile phase B following a 5 min delay. Instrumental parameters were previously optimized using a 5 mM solution of angiotensin infused at a flow rate of 250 nL/min. The LC/MS/MS data for the purified TH were analyzed using the de novo sequencing software MassAnalyzer 1.02 (Amgen, Thousand Oaks, CA) and PEAKS (Bioinformatic Solutions Inc., Waterloo, Ontario, Canada). The software program TurboSEQUEST implemented in Bioworks (Thermo) was also used to analyze the LC/MS/MS data using an appropriate database containing a TH protein sequence and the SEQUEST algorithm (*35*).

Cloning and Isolation of Recombinant TH. The RNA from *R. glutinis* was isolated using TRIzol reagent according to the instructions from the manufacturer (Invitrogen, Grand Island, NY). The mRNA obtained was reverse transcribed using the SuperScript first-strand standard protocol (Invitrogen). First-strand cDNA synthesis was performed by priming with 20 pmol of oligo-dT in a 20 μL reaction mix containing 10 mM each dATP, dCTP, dGTP, and dTTP, 40 U/μL RNaseOUT recombinant ribonuclease inhibitor, and 50 U/μL SuperScript II reverse transcriptase. The reverse transcriptase reaction was stopped by cooling to 4 °C for 10 min.

The resulting cDNA was then amplified with Phusion high-fidelity DNA polymerase (New England Biolabs, Beverley, MA) according to the manufacturer's instructions. Thirty-six cycles of

polymerase chain reaction (PCR) (10 s at 98 °C, 30 s at 61 °C, and 32 s at 72 °C) were performed. The following sequences of the oligonucleotides used for PCR amplification were based upon the cDNA sequence of TH [GenBank accession number AY622311, National Center for Biological Information (NCBI)] reported by Smiley et al. (*17*): 5′-CGG CGG *GGA* TCC ATG GTC TCG TCT GGC ATC GTC-3′ (sense, carries a BamH1 restriction site) and 5′-CTT TTC CTT TT*G CG*G CCG CTT GAG GGC ACT GCT GCA TTA C-3′ (antisense, carries a Not1 restriction site).

After resolution of the products by gel electrophoresis on a 1% agarose gel, the expected 1035 base pair product was extracted using the QIAquick gel extraction kit (Qiagen, Valencia, CA) and ligated into the glutathion-*S*-transferase (GST) fusion expression vector pGEX-4T-1 previously digested with BamH1 and Not1. Ligated products were electroporated into *E. coli* BL21 Star DE3 (Invitrogen). The plasmid was isolated and purified using a QIAprep Spin Miniprep kit (Qiagen), and both strands of the insert were sequenced (Davis Sequencing, Inc., Davis, CA) using the primers 5′-TTG GTG GTG GCG ACC ATC CTC CAA-3′ (pGEXmcs5′) and 5′-CTG CAT GTG TCA GAG GTT TTC ACC-3′ (pGEXmcs3′).

Purification of Recombinant TH. To purify TH expressed as a recombinant protein in *E. coli*, 2 L of *E. coli* BL21 Star DE3 carrying the GST-TH construct were grown in LB broth with 50 μg/mL ampicillin at 37 °C until $A_{600} = 0.6-0.7$ and then induced overnight with 0.2 mM isopropyl-β-D-thiogalactopyranoside at 30 °C. The cells were harvested by centrifugation, resuspended in lysis buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, and 0.01% Triton-X) supplemented with 1 mM dithiothreitol and 1 mM phenylmethylsulfonyl fluoride, and incubated at room temperature for 30 min. Lysis was then completed by sonicating the suspension on ice using a Branson Sonifier Cell Disruptor 200 and six bursts of 10 s each, with a 90 s interval between pulses. The lysate was clarified by centrifugation (12000 rpm for 30 min at 4 °C). The supernatant was then mixed with swelled glutathion-agarose beads (Sigma-Aldrich) and incubated at 4 °C overnight with gentle agitation. The suspensions were centrifuged at 3000 rpm for 5 min at 4 °C, and the beads were washed twice with lysis buffer and then washed twice with thrombin buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, and 1 mM dithiothreitol).

The recombinant protein GST-TH was resuspended in 15 mL of thrombin buffer, cleaved with 100 U/mL thrombin (Sigma-Aldrich) at 37 °C for 1 h, and then purified by FPLC using a SuperDex 75 column (GE Healthcare, Waukesha, WI). The protein was concentrated using centricon YM-10 membranes (Millipore, Billerica, MA). The protein concentration was determined using the BCA protein assay reagent kit (Pierce Chemical Co.). The protein was analyzed on NuPAGE 4−12% Bis Tris Gels (Invitrogen, Carlsbad, CA), stained with Simply Blue (Invitrogen) and confirmed by mass spectrometry. TH was digested with trypsin following the protocol of Matsudaira (*36*) and analyzed by MALDI-TOF-MS (Bruker Daltonics, Billerica, MA) and LC/MS/MS as described above.

Bioinformatics Analysis and Comparative Modeling. A Basic Local Alignment Search Tool (BLAST) search using the deduced amino acid sequence of TH was performed to identify TH homologues in other microorganisms with an *E* value of less than 1×10^{-4} using the NCBI Web site with default parameters. A multiple sequence alignment was generated using clustalX as implemented in BioEdit (*37*).

The multiple sequence alignments were manually edited for comparative modeling to ensure alignment of residues critical for catalytic activity (see the Results and Discussion). Potential modeling templates for TH were obtained by using a BLAST search against the database of sequences corresponding to proteins with known structure, which suggested anthocyanin synthase as the best template model. The locations of gaps between the TH and the anthocyanidin synthase amino acid sequences were manually modified to maximize alignment of conserved residues and to ensure that gaps were located in loops on the surface of the structural model. The secondary structure of all sequences was predicted using Profsec (http://cubic.bioc.Columbia.edu/predictprotein/) with the default options (*38*). The predicted secondary structure of TH, when compared with the observed secondary structure of anthocyanidin synthase, was helpful when determining the best sequence alignment.

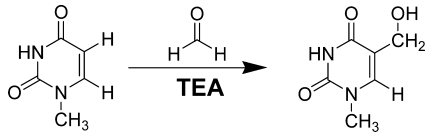
The TH model was built using InsightII (Accelrys, San Diego, CA) to assign coordinates for the structurally conserved residues and to predict the coordinates of loops randomly or based on a database search of loop conformations within the protein databank. An alignment of TH and anthocyanidin synthase detailing the template coordinates used to build our model of the TH active site is included in the Supporting Information.

Site-Directed Mutagenesis. Site-directed mutagenesis was used to mutate Arg-285 to the Cys-285 observed by Smiley et al. (*17*), generating a TH R285C mutant. The pGEX4T-1ADTH plasmid containing the TH sequence was mutagenized using the QuickChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) following the manufacturer's protocol using the primers 5′-GAT GAC GCC CCG CTG CAA CTC GAT CGC-3′ (sense) and 5′-GCG ATC GAG TTG CAG CGG GGC GTC ATC-3′ (antisense) (Sigma-Genosys, Woodland, TX). The expected mutation was confirmed by DNA sequencing (Davis Sequencing). Furthermore, the protein was expressed and purified as above, and the mutant TH R285C was confirmed by MALDI-TOF-MS using peptide mapping analysis after trypsin digestion. A similar strategy was used to generate the mutants N187D, N293D, and N250D.

Steady-State Enzyme Kinetics. The activity and kinetics of TH, also known as thymine dioxygenase (EC 1.14.11.6), isolated from *R. glutinis* or a recombinant protein and its mutants described above were examined. The purity of TH samples was assessed by PAGE stained with a coomassie blue stain and was ≥95% in all cases. The standard enzyme reaction and buffer conditions defined by Stubbe and co-workers (*14*) and used throughout this report, unless explicitly stated otherwise, contained 0.45 mM α-ketoglutarate, 11 μM ferrous sulfate, 2.3 mM sodium ascorbate, and 50 mM HEPES at a pH of 7.5 in a total volume of 200 μL. No enzyme activity was observed when α-ketoglutarate or iron was omitted from the enzyme reaction. The pH was varied from 6.5 to 8.0 to determine that a pH of 7.5 was optimum for TH activity. All steady-state kinetic assays were performed with 5−10 μg of protein in a 200 μL reaction volume. The assays were performed in 2 mL vials placed in an aluminum block heated to 30 °C to control the reaction temperature in all cases. When measuring the specific enzyme activity of TH or comparing the relative activity of TH and its mutants, the reaction was initiated by adding the substrate to obtain a final concentration of 0.9 mM thymine and stopped 1 min later. When testing the TH activity of fractions during purification of the wild-type enzyme, deuterium-enriched thymine-α,α,α,6-d4 was used as the substrate to eliminate interference from unlabeled thymine present in *R. glutinis* extracts. The enzyme reaction was stopped with the addition of hot absolute ethanol (200 μL) followed by heating at 60 °C for 5 min. No residual TH activity was observed following ethanol addition, and ethanol did not interfere in subsequent mass spectrometry measurement of the pyrimidine products.

The kinetics for pyrimidine substrates were determined by substituting thymine for the appropriate pyrimidine substrate and varying its concentration. The concentrations of substrate ranged from 10 μM to 2 mM. The standard reaction conditions defined above and a 1 min reaction time were used. After addition of internal standard, the GC/MS method described below was used to quantitate product formation. Kinetics data were analyzed using Prism 4 (GraphPad, La Jolla, CA) to perform nonlinear least-squares analysis for the determination of K_m and k_{cat} values along with the calculated standard error. We note that steady-state kinetic parameters determined for TH must be interpreted with caution as the pyrimidine concentrations needed to observe maximum enzyme velocity approach the aqueous solubility of the pyrimidines examined here. The high apparent K_m value of FoU (near its solubility limit) prevented accurate observation of a velocity maximum and limited confidence in the nonlinear analysis; the traditional Lineweaver−Burk, Eadie−Hofstee, and Hanes−Woolf plots were also used in this case to derive K_m and k_{cat}.

Scheme 1. Synthesis of 1Me5HmU



Analysis of TH Reaction Products by GC/MS. Thymine and its oxidation damage products can be separated, identified, and quantified by GC/MS methods (*39*). After the enzyme reaction was stopped, 20 μL of an internal standard solution containing 2-hydroxy-4-methylpyrimidine or $^2H_2$-hydroxymethyluracil was added. The internal standard $^2H_2$-hydroxymethyluracil was used when examining the differences in activity for TH mutants, and 2-hydroxy-4-methylpyrimidine was used when determining the kinetics of oxidation for a series of pyrimidine analogues. The internal standard $^{15}N_2$,$^2H_2$-5-hydroxymethyluracil was not used due to interference by α-ketoglutarate, which has the same mass following derivatization (431 *m*/*z*). The resulting mixture was dried under reduced pressure. A silylating mixture of acetonitrile (100 μL) and silylating agent (MTBSTFA + 1% TBDMCS, 100 μL) was added, and the sealed vial was heated at 140 °C for 30 min.

Pyrimidine products from the TH assay were measured with a Hewlett-Packard 6890 gas chromatograph containing an Ultra-2 column interfaced with a Hewlett-Packard 5973 mass spectrometer. An aliquot (1 μL) of the silylation reaction mixture was injected onto the GC/MS with an inlet temperature of 250 °C. The oven temperature was held constant at 100 °C for 2 min, increased at a rate of 30 °C/min for 2.7 min to 180 °C, then increased at a rate of 16 °C/min for 6.3 min to 280 °C, and held at 280 °C from 2 min. The amount of thymine and corresponding oxidation products was determined by comparing the integrated area of the specific analyte at a selected mass with that of the internal standard. The limit of detection for all pyrimidines examined was less than 1 pmol injected onto the column. The retention times for all compounds examined and the ions used for the quantitation of oxidation products are listed in the Supporting Information. Relative peak areas were converted to relative concentrations based upon comparison of the experimental areas with standard curves ($R^2 \geq 0.99$) constructed for each of the analytes vs the internal standard.

Preparation and Characterization of N1-Methyl-5-hydroxymethyluracil (1MeHmU). The 5-hydroxymethyl derivative of N1-methylthymine (1MeThy) was synthesized using a method previously used to convert 2′-deoxyuridine to 5-hydroxymethyl-2′-deoxyuridine (*40*). The synthesis of 1MeHmU is outlined in Scheme 1. To a solution of formaldehyde (CH_2O, 20% w/w, 2.7 mL) and triethylamine (TEA, 4 mL, 2.9×10^{-2} mol) was added 1-methyluracil (1MeU, 481 mg, 3.8×10^{-3} mol). The reaction mixture was then heated at 60 °C for 7 days. Triethylamine was then removed using a rotary evaporator, and the resulting residue was dissolved in methanol (MeOH) and purified by normal phase column chromatography (Silica Gel, Sigma-Aldrich). Unreacted 1MeU was eluted using a solvent mixture of 4% MeOH in dichloromethane (CH_2Cl_2). 1MeHmU was then recovered using a solvent mixture of 9% MeOH in CH_2Cl_2. The final product of 98.5 mg (17% yield) was obtained as a white powder: GC/MS $C_6H_8N_2O_3$ 156. ^{1}H NMR (DMSO-d_6): δ 3.24 (s, 3H, N1-CH_3), δ 4.11 (d, 2H, C5−H), δ 7.51 (s, 1H, C6−H), δ 11.22 (s, 1H, NH).

Results

TH from *R. glutinis* Was Isolated and Examined by Mass Spectrometry. Wild-type TH was isolated from *R. glutinis* according to previously published methods (*14*). In the procedure reported here, the use of radiotracers was substituted by a GC/MS method to monitor the conversion of thymine to HmU. To use this nonradioactive procedure to follow enzyme activity during isolation from the host organism that contained endogenous thymine and HmU, we used deuterium-enriched thymine as the substrate. Substrate requirements and the apparent molecular weight of the isolated enzyme were consistent with previously published accounts (*11, 14*).

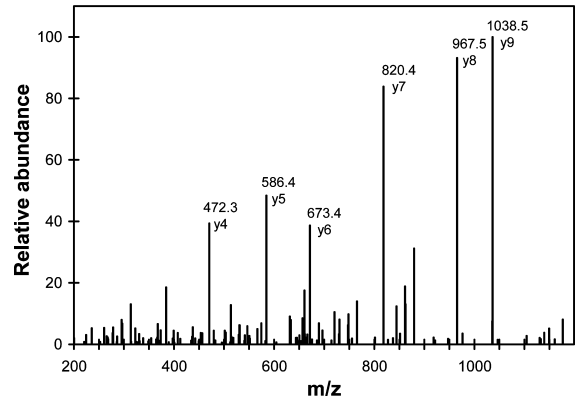


Figure 2. Tryptic peptides of TH purified from *R. glutinis* were examined by LC/MS/MS. One highly significant peptide, NSIAFFSNPSLR, was identified. The mass spectrum shows the fragment ions when the peptide corresponding to the indicated sequence is fragmented by collision-induced dissociation. The masses expected for the y1−y11 ions are 175.12, 288.20, 375.23, 472.29, 586.33, 673.36, 820.43, 967.50, 1038.54, 1151.62, and 1238.65 *m*/*z*. The masses expected for the b2−b11 ions are 202.08, 315.17, 386.20, 533.27, 680.34, 767.37, 881.41, 978.47, 1065.50, and 1178.58 *m*/*z*. The precursor mass and the indicated fragments are consistent with the indicated sequence.

The isolated enzyme was digested with trypsin, and the corresponding fragments were characterized by mass spectrometry. Although insufficient data were obtained to determine the entire protein sequence, one significant peptide was identified as NSIAFFSNPSLR (Figure 2). This peptide is consistent with one identified by Stubbe and co-workers as a critical sequence within the substrate-binding domain (*14*). The identification of this peptide, although consistent with the previous results of Stubbe and co-workers (*14*), is inconsistent with the sequence reported by Smiley et al. (*17*). Although the DNA sequence provided by Smiley et al. (*17*) would encode this peptide sequence, it is preceded by a cysteine residue and therefore would not allow formation of the tryptic fragment described above.

TH Was Cloned and the Sequence Obtained Here Is Different from the Previously Reported Sequence (*17*). TH was cloned from the same *R. glutinis* strain used by the Stubbe laboratory and Smiley et al. (*14, 17*). The PCR primer sequence used here for cloning (Materials and Methods) was also based upon the N-terminal peptide sequence reported by Stubbe and co-workers (*14*) but differed from the primers used by Smiley et al. (*17*). The DNA sequence of TH obtained here is presented in Figures S1−S4 and Table S1 of the Supporting Information and has been deposited in GenBank as accession number FJ648740.

The DNA sequence reported here predicted a protein of 332 amino acids, consistent with Smiley et al. (*17*), and a predicted molecular mass of 37 kDa, consistent with the value reported by Stubbe and co-workers (*14*) for the wild-type enzyme. Substrate requirements and pH optimum for the cloned enzyme were indistinguishable from the isolated enzyme (*11, 14*). In accord with Smiley et al. (*17*), the DNA sequence predicted valine at position 9, which differs from Edman sequencing data from Stubbe and co-workers (*14*), who observed proline at this position.

The DNA sequence reported here differs from the previously reported DNA sequence (*17*) at two positions. At nucleotide position 551, we observe guanine, whereas adenine was previously reported, and at position 853, we observe cytosine, whereas thymine was previously reported. The corresponding

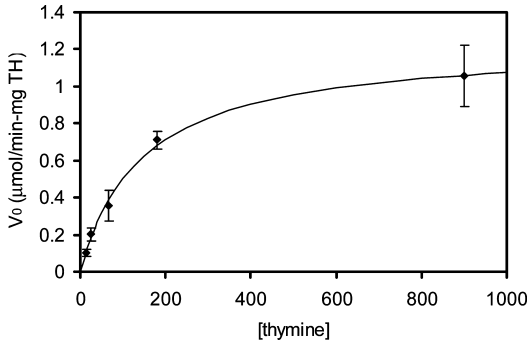


Figure 3. Kinetics of recombinant *R. glutinis* TH oxidation of thymine. All reactions used the standard conditions given in the Materials and Methods and were performed in triplicate. The error bars indicate the standard deviation of observed values. One minute reaction times result in K_m and k_{cat} values of 145 μM and 67 min^{-1}, respectively.

Table 1. Steady-State Kinetic Parameters for the Oxidation of Pyrimidine Substrates by TH[a]

substrate	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (μM^{-1} min^{-1})
1. thymine	67 ± 7	145 ± 15	0.56
2. HmU	25 ± 3	230 ± 43	0.11
3. FoU	68 ± 2	1700 ± 200	0.04
4. CarboxyU	≤0.01		
5. 1MeU	16 ± 2	110 ± 38	0.15
6. uracil	≤0.01		
7. 1MeThy	26 ± 2	1300 ± 220	0.02
8. 1MeHmU	≤0.01		
9. 5MeC	≤0.01		
10. 3MeU	≤0.01		
11. 6MeU	≤0.01		
12. thymidine	≤0.01		

[a] The structures of the substrates examined are found in Figure 4. The sensitivity of the GC/MS assay corresponds to a detection limit for k_{cat} values of less than 0.01 min^{-1} given the quantity of enzyme used. The K_m and k_{cat} values are reported with the standard error.

codons for our protein sequence at positions 184 and 285 would be CGG and CGC, both encoding arginine, whereas the previous study reported CAG and TGC, encoding glutamine and cysteine, respectively. Our translated protein sequence contains arginine at position 285, which would generate the tryptic fragments observed by both our group and Stubbe and co-workers (*14*).

The TH Cloned Here Has Significant Activity against a Series of Pyrimidines and Appears to Interact with the Hydrogen-Bonding Face. Using the GC/MS method reported here, the steady-state kinetic parameters for cloned TH were measured. An example of the kinetic results for the conversion of thymine to HmU is shown in Figure 3, and steady-state kinetic parameters for additional pyrimidine analogues are reported in Table 1. The corresponding structures are shown in Figure 4. We observed that the cloned TH converted thymine to HmU, FoU, and CarboxyU. Previous studies (*14*) were unable to demonstrate that one protein activity could carry out all of the sequential reactions, as the wild-type protein cannot be purified to homogeneity and still retain enzymatic activity. Within this series, K_m varies more among the analogues than does k_{cat}, and the relative reactivity is Thy > HmU > FoU, consistent with previously reported trends (*14*).

We also examine the activity of TH toward 1MeU and 1MeThy. The methyl group of 1MeU can be oxidized and spontaneously lost, generating uracil and formaldehyde (Figure 4). Therefore, a methyl group in either the C5-position (Thy) or the N1-position (1MeU) is oxidized by TH. The methyl groups of 1MeThy at either the C5- or the N1-position could be oxidized, depending upon the orientation with which it binds to the active site. Oxidation of the N1-methyl group would generate thymine, whereas oxidation at the 5-position would generate 1MeHmU. Previously, 1MeHmU was proposed as a product from TH-mediated oxidation of 1MeThy (*14*). We report here that 1MeHmU was prepared by chemical synthesis and characterized by NMR spectroscopy and high-resolution mass spectrometry, allowing confirmation of the reactivity of 1MeThy. As shown in Figure 4, 1MeThy could generate HmU via two pathways. We also observed that TH does not convert 1MeHmU to HmU, indicating that HmU must arise from 1MeThy by oxidation first at the N1-position, generating thymine, followed by oxidation in the 5-position, generating HmU. We also observed that TH does not act on the nucleoside thymidine.

TH from *R. glutinis* does not act on 3MeU or 5MeC. A methyl group in the N3-position would interfere with interaction of the pyrimidine with the hydrogen-bonding face (O2−N3H−O4), whereas a methyl group in the N1-position would not. The apparent discrimination for thymine over 5MeC suggests that hydrogen-bonding interactions with the O4 and N3H groups of thymine are likely important for recognition.

We considered the possibility that the activity of TH might be modulated by self-oxidation of one of the active site aromatic amino acid residues, as has been reported for other enzymes with similar substrate requirements (*41*−*43*). We also considered the possibility that self-oxidation could account for earlier ambiguity with respect to either phenylalanine or tyrosine at position 291 (*15, 16*). Cloned TH was purified and studied by MS prior to and after exposure to iron and α-ketoglutarate with or without thymine. The tryptic peptide with phenylalanine at position 285 was observed; however, no peptides containing tyrosine or other oxidized phenylalanine derivatives at position 291 were observed, although the enzyme was incubated under conditions that resulted in significant conversion of thymine to HmU. This result indicates that the chemical oxidation of phenylalanine to tyrosine is not involved in modulating the activity of *R. glutinis* TH.

TH from *R. glutinis* Shows Substantial Homology with TH from Related Yeast Species. The identity of the amino acid at position 285 distinguishes our sequence from the previously reported sequence and is significant with respect to comparison among the yeast TH homologues. As shown in Figure 5, the sequence reported here for this region of the protein shows considerable homology with protein sequences derived from the genome sequences of *N. crassa* and A. nidulans (*19, 20*), whereas substantially less overlap is observed with the previously reported sequence (*17*).

TH from *R. glutinis* Is a Member of the Nonheme Iron α-Ketoglutarate Dependent Dioxygenase Family. Upon the basis of substrate requirements, TH could be placed within the family of dioxygenases. TH can now be placed within this family upon the basis of sequence as well. Members of the dioxygenase family contain sequence motifs to bind iron and α-ketoglutarate. Iron is bound by the highly conserved HxD-x_n-H sequence motif, where x is any amino acid and x_n is a variable number of amino acids from 40 to 153 (*8*). The C5 carboxylate of α-ketoglutarate is stabilized by either the RxN or the RxS sequence motif. The previously reported sequence (*17*) contained RCNS, which allowed TH to be placed within the overall dioxygenase family; however, as it had the RxN motif, it was placed within the alkB subfamily as opposed to the subfamily containing the other yeast TH homologues. In contrast, we report that the corresponding sequence is RRNS, allowing *R. glutinis* TH to be placed within the same subfamily as the other yeast homologues.

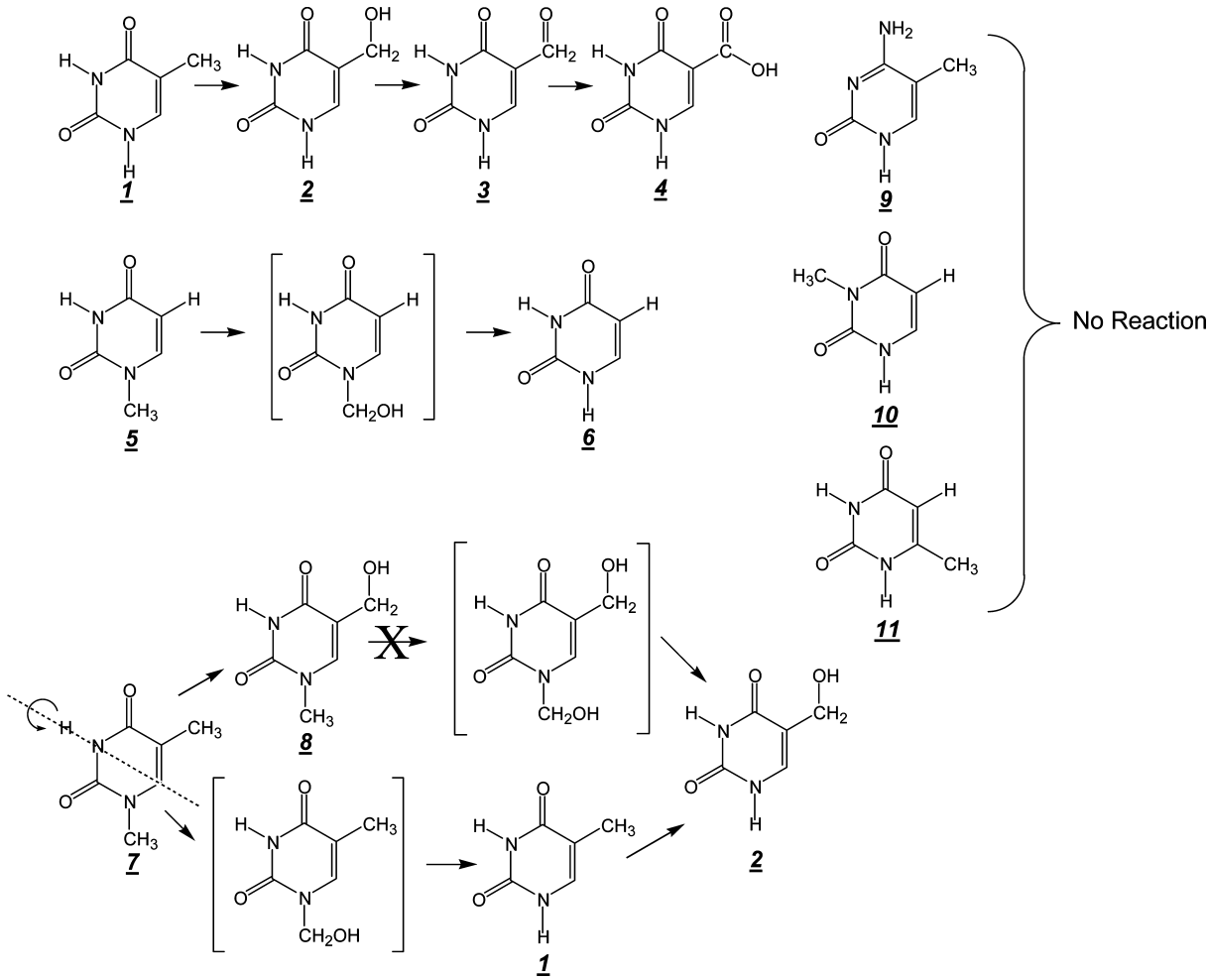


Figure 4. Proposed reaction scheme for the pyrimidines examined in this study. The compounds shown are (1) thymine, (2) HmU, (3) FoU, (4) CarboxyU, (5) N1-methyluracil (1MeU), (6) uracil, (7) 1MeThy, (8) 1MeHmU, (9) 5MeC, (10) N3-methyluracil (3MeU), and (11) 6-methyluracil (6MeU).

```
Smiley et al.   -PRCNSIAFFSNPSLRTVISALPGT
RGTH            PRRNSIAFFSNPSLRTVISALPGT-
NSTH            PPRYSIAYFCNPNHKSYIEAIPGT-
ANTH            PERYSVAYFCNPNMDKFIEAIPGTY
```

Figure 5. Sequence alignment of TH enzymes from *R. glutinis* (RGTH), *N. crassa* (NSTH), and *A. nidulans* (ANTH). The conserved arginine that binds α-ketoglutarate is highlighted. The alignment of the Smiley et al. (*17*) sequence tests the proposition that the conserved arginine that binds the α-ketoglutarate is Arg-284 instead of Arg-285.

The Proposed Active Site of *R. glutinis* TH Can Be Modeled Based upon Homology with Similar Enzymes. The nonheme iron α-ketoglutarate-dependent dioxgenases comprise a large and varied group of metabolic enzymes. A BLAST search against a database of proteins with known structure yielded isopenicillin N synthase and anthocyanidin synthase, whose sequences are significantly (BLAST scores of 63 and 52, respectively) similar to that of *R. glutinis* TH and could therefore serve as potential template structures. Anthocyanidin synthase (PDB file 1GP5, 1GP6) was chosen as a template structure because structures with bound α-ketoglutarate (or succinate and oxygen) and iron were available (*44*). The structural model was generated as described in the Materials and Methods and is shown in Figure 6. When Arg-285 is modeled as the residue that binds α-ketoglutarate, the TH sequence RRNSIAFFSN forms a *β*-strand with the underlined residues oriented toward the active site.

The Homology Model Led to a Series of Mutant Proteins That Were Expressed and Examined. To test the importance of specific amino acids to the activity of *R. glutinis* TH, a series of mutants were constructed. The first mutant was R285C, which recreated the sequence of Smiley et al. (*17*) at one of the two positions where our respective sequences differed.

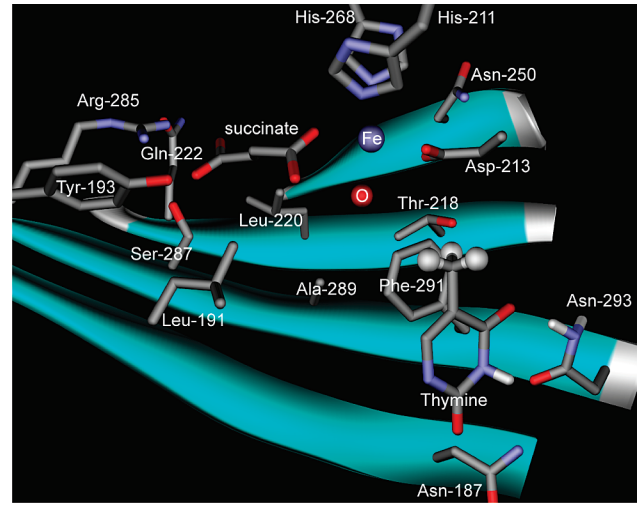


Figure 6. Active site model of recombinant *R. glutinis* TH. The active site was modeled using the published coordinates of anthocyanadin synthase; the alignment used during modeling is included in the Supporting Information. The active site residues discussed in the text are shown, and the labels indicate the residue numbering of *R. glutinis* TH.

Arg-285 is within the active site as proposed by Stubbe and co-workers (*14*), and the assignment of arginine at this position allows close alignment among the yeast homologues. In our hands, the R285C mutant had no detectable activity for the conversion of thymine to HmU. This result suggests that the sequence reported here is likely correct and confirms the importance of Arg-285 as a critical active site residue.

We also prepared a series of mutants in which asparagine was converted to aspartate to provide additional information on the mechanism by which TH selects thymine analogues over 5MeC analogues. We note that in previous studies with uracil DNA glycosylase and thymidylate synthase, discrimination between thymine and cytosine analogues could be reversed by substituting asparagine for aspartate (*45*−*47*). We therefore mutated several asparagine residues near the active site to aspartate residues. The k_{cat} values of the mutants N187D, N250D, and N293D, respectively, for thymine were found to be 14.9 ± 0.3, 0.67 ± 0.07, and 3.2 ± 0.3 min^{-1}, whereas the corresponding K_m values were observed to be 67 ± 8, 1150 ± 280, and 2100 ± 450 μM. While the first substitution had only a modest impact on the enzyme activity, the latter two significantly diminished enzyme activity through increasing K_m. Neither wild-type TH nor any of the mutants examined here had any measurable activity against 5MeC. The data reported here suggest that discrimination between thymine and 5MeC cannot be attributed to a single amino acid.

Discussion

The DNA bases in all living organisms are persistently damaged by oxidation and require continuous repair (*48, 49*). Damage can occur at multiple sites, including pyrimidine methyl groups found on both thymine and 5MeC (*50, 51*). Although the corresponding 5-hydroxymethyl pyrimidines are not miscoding, repair glycosylases have been identified in mammals (*52, 53*). The 5-hydroxymethyl analogues of both thymine and 5MeC are known to interfere with sequence-specific DNA protein interactions that could influence epigenetic programming of gene activity (*54*−*57*). It has been suggested that cytosine methylation patterns could be modulated by a "demethylase" activity that acts by oxidative demethylation; however, no such activity has yet been identified (*27*−*30*).

TH is an enzymatic activity that similarly oxidizes thymine to HmU. TH could therefore serve as a model for other enzymes that oxidize pyrimidine methyl groups and further assist in the identification of other as yet unidentified enzymes that catalyze similar chemistry. Interestingly, a related enzyme has recently been identified in trypanosomatids that converts thymine residues to HmU in telomeric DNA (*32*). TH is also related by substrate requirements to the alkB family of α-ketoglutarate-dependent dioxygenases that remove N-methyl groups from DNA bases damaged by alkylating agents.

Although TH activity has been known for some time, challenges with protein purification have limited the amount of protein available for study. The absence of DNA sequence information in the yeasts with known TH activity prevented the use of genome-wide homology searches to identify candidate gene sequences. We therefore began our studies by attempting to determine the protein sequence of *R. glutinis* TH. While this work was in progress, another group (*17*) reported a sequence for TH while searching for a related gene involved in pyrimidine metabolism in yeast using partial sequence information obtained from tryptic peptides generated by Thornburg et al. (*14*). The reported sequence (*17*) would not generate a tryptic peptide that had been identified (*14*) as part of the catalytic pocket, and the cloned TH had minimal activity.

We therefore recloned *R. glutinis* TH from the same strain used by previous investigators. The DNA sequence obtained here predicted a protein of similar molecular weight and amino acid sequence differing in only two nucleotides from the previously published DNA sequence (*17*). Both nucleotide changes would result in amino acid substitutions. Importantly, our sequence would generate a tryptic peptide from the catalytic center previously observed (*14*) from the isolated enzyme.

Upon the basis of the sequence reported here, TH was cloned and purified. The cloned enzyme showed substantial activity and allowed us to demonstrate that one protein sequentially converted thymine to HmU to FoU and CarboxyU. TH was able to oxidize methyl groups in either the N1- or the C5-positions of uracil analogues, indicating that the pyrimidine can enter and bind in the active site in two potential orientations. Oxidation of the N-methyl group resulted in spontaneous demethylation, as has been reported for the alkB series of DNA proteins. In contrast to alkB, which oxidizes methyl groups at the N3-position of pyrimidines bases, TH oxidizes methyl groups at the N1-position.

The alkB family of DNA repair proteins has been known for some time to protect organisms from alkyating agent-induced toxicity (*23*). Its mechanism of action and substrate targets were identified only after results of homology comparisons were published (*25, 26*). Avarind and Koonin (*24*) conducted homology searches, which placed alkB in the larger family of nonheme iron α-ketoglutarate-dependent dioxygenases, revealing its likely mechanism of action as well as its likely DNA damage targets. Subsequent studies have revealed additional homologues in other organisms including human highlighting the potential strength of genomic approaches, which rely upon knowledge of the DNA coding sequence. We report here a sequence for *R. glutinis* TH that generates an active protein and places TH by sequence within the larger family of dioxygenases and correctly places it within the subfamily of TH enzymes from other yeast species.

The dioxygenase family comprises a large group of enzymes involved in numerous biochemical metabolic pathways, generally involving oxidative demethylation reactions. The active site structure of nonheme, α-ketoglutarate-dependent dioxygenases with known structures is highly conserved. The distances between the active site iron, the Cα atom of the two histidine and one aspartate residue that bind the iron atom, and the Cα atom of the α-ketoglutarate binding arginine residue vary no more than 0.5 Å when comparing the crystal structures for a biochemically disparate group of enzymes including anthocyanidin synthase (1GP5), alkB (2FD8, 3BIE, 3BKZ, and 3BI3), ABH2 (3BU0 and 3BUC), and ABH3 (2IUW). Homology modeling within this family has been used to generate a proposed structure of the active site of TH. This preliminary structure indicated the probable importance of Arg-285 in the binding of α-ketoglutarate and suggested that the R285C mutation in the previously reported sequence might explain its apparent lack of activity (*17*). This mutant was constructed, tested, and found to be without activity, as suggested by the model.

Another important structural question is how TH can selectively act on thymine but not 5MeC. In other enzymes that act on pyrimidines, including thymidylate synthase and uracil DNA glycosylase, specificity for thymine over cytosine analogues is afforded by a specific asparagine residue that hydrogen bonds to the N3H and O4 carbonyl of thymine analogues. Conversion of asparagine to aspartate alters these hydrogen-bonding interac-

tions and changes the specificity from thymine to cytosine (*45*−*47*). A series of TH mutants were constructed here in which asparagine residues near the catalytic site were substituted with aspartate. Activity toward thymine was diminished for two of the three mutants, while neither the wild-type enzyme nor any of the mutants demonstrated any activity in the conversion of 5MeC to 5-hydroxymethylcytosine. The selectivity toward thymine therefore likely results from interactions with specifically oriented side chain residues or multiple amino acids.

The active sites of TH and the alkB family share considerable similarity within the active site with respect to the binding of α-ketoglutarate, iron, and oxygen. Whereas TH acts on a free base, alkB acts on alkylated thymine residues within the context of DNA. The capacity of alkB to bind duplex nucleic acids is found on the amino terminal side of the oxidation center. The active site in TH that allows selective binding of thymine analogs is as yet unknown but likely exists on the carboxyl terminal side of the catalytic center. Further structural studies will be needed to define the substrate interactions with TH; however, the apparent large number of members of the dioxygenase family that can act on diverse substrates suggests that additional enzymes might be found that could modify DNA bases and thus have important biological functions.

Supporting Information Available: Table of GC/MS data for derivatized products of TH reactions; DNA sequence for the TH gene from *R. glutinis*; sequence alignment of TH enzymes from *R. glutinis* (RGTH), *N. crassa* (NSTH), and *A. nidulans* (ANTH); model of the TH active site; and sequence alignment of nonheme dioxygenases that oxidize methylated pyrimidines. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Fink, K., Cline, R. E., Henderson, R. B., and Fink, R. M. (1956) Metabolism of thymine (methyl-C14 or -2-C14) by rat liver in vitro. *J. Biol. Chem. 221*, 425–433.

(2) Schandl, E. K. (1973) Thymine 7-hydroxylase activity in normal and leukemic leukocytes. *Biochem. Biophys. Res. Commun. 52*, 524–529.

(3) Abbott, M. T., Kadner, R. J., and Fink, R. M. (1964) Conversion of thymine to 5-hydroxymethyluracil in a cell-free system. *J. Biol. Chem. 239*, 156–159.

(4) Liu, C. K., Shaffer, P. M., Slaughter, R. S., McCroskey, R. P., and Abbott, M. T. (1972) Stoichiometry of the pyrimidine deoxyribonucleoside 2′-hydroxylase reaction and of the conversions of 5-hydroxymethyluracil to 5-formyluracil and of the latter to uracil-5-carboxylic acid. *Biochemistry 11*, 2172–2176.

(5) Liu, C. K., Hsu, C. A., and Abbott, M. T. (1973) Catalysis of three sequential dioxygenase reactions by thymine 7-hydroxylase. *Arch. Biochem. Biophys. 159*, 180–187.

(6) Abbott, M. T., Schandl, E. K., Lee, R. F., Parker, T. S., and Midgett, R. J. (1967) Cofactor requirements of thymine 7-hydroxylase. *Biochim. Biophys. Acta 132*, 525–528.

(7) McCroskey, R. P., Griswold, W. R., Sokoloff, R. L., Sevier, E. D., Lin, S., Liu, K., Shaffer, P. M., Palmatier, R. D., Parker, T. S., and Abbott, M. T. (1971) Studies pertaining to the purification and properties of thymine 7-hydroxylase. *Biochim. Biophys. Acta 227*, 264–277.

(8) Hausinger, R. P. (2004) FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol. 39*, 21–68.

(9) Shaffer, P. M., and Arst, H. N., Jr. (1984) Regulation of pyrimidine salvage in *Aspergillus nidulans*: A role for the major regulatory gene are A mediating nitrogen metabolite repression. *Mol. Gen. Genet. 198*, 139–145.

(10) Shaffer, P. M., Cairns, J. R., Dorman, D. C., and Lott, A. M. (1984) Separation and characterization of pyrimidine deoxyribonucleoside 2′-hydroxylase and thymine 7-hydroxylase from *Aspergillus nidulans*. *Int. J. Biochem. 16*, 429–434.

(11) Wondrack, L. M., Hsu, C. A., and Abbott, M. T. (1978) Thymine 7-hydroxylase and pyrimidine deoxyribonucleoside 2′-hydroxylase activities in *Rhodotorula glutinis*. *J. Biol. Chem. 253*, 6511–6515.

(12) Wondrack, L. M., Warn, B. J., Saewert, M. D., and Abbott, M. T. (1979) Substitution of nucleoside triphosphates for ascorbate in the thymine 7-hydroxylase reaction of *Rhodotorula glutinis*. *J. Biol. Chem. 254*, 26–29.

(13) Warn-Cramer, B. J., Macrander, L. A., and Abbott, M. T. (1983) Markedly different ascorbate dependencies of the sequential alpha-ketoglutarate dioxygenase reactions catalyzed by an essentially homogeneous thymine 7-hydroxylase from *Rhodotorula glutinis*. *J. Biol. Chem. 258*, 10551–10557.

(14) Thornburg, L. D., Lai, M. T., Wishnok, J. S., and Stubbe, J. (1993) A non-heme iron protein with heme tendencies: An investigation of the substrate specificity of thymine hydroxylase. *Biochemistry 32*, 14023–14033.

(15) Thornburg, L. D., and Stubbe, J. (1993) Mechanism-based inactivation of thymine hydroxylase, an alpha-ketoglutarate-dependent dioxygenase, by 5-ethynyluracil. *Biochemistry 32*, 14034–14042.

(16) Lai, M., Wu, W., and Stubbe, J. (1995) Characterization of a novel, stable norcaradiene adduct resulting from the inactivation of thymine hydroxylase by 5-ethynyluracil. *J. Am. Chem. Soc. 117*, 5023–5030.

(17) Smiley, J. A., Kundracik, M., Landfried, D. A., Barnes, V. R., Sr., and and Axhemi, A. A. (2005) Genes of the thymidine salvage pathway: Thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa*. *Biochim. Biophys. Acta 1723*, 256–264.

(18) Braun, E. L., Halpern, A. L., Nelson, M. A., and Natvig, D. O. (2000) Large-scale comparison of fungal sequence information: mechanisms of innovation in *Neurospora crassa* and gene loss in *Saccharomyces cerevisiae*. *Genome Res. 10*, 416–430.

(19) Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., FitzHugh, W., Ma, L. J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S., Nielsen, C. B., Butler, J., Endrizzi, M., Qui, D., Ianakiev, P., Bell-Pedersen, D., Nelson, M. A., Werner-Washburne, M., Selitrennikoff, C. P., Kinsey, J. A., Braun, E. L., Zelter, A., Schulte, U., Kothe, G. O., Jedd, G., Mewes, W., Staben, C., Marcotte, E., Greenberg, D., Roy, A., Foley, K., Naylor, J., Stange-Thomann, N., Barrett, R., Gnerre, S., Kamal, M., Kamvysselis, M., Mauceli, E., Bielke, C., Rudd, S., Frishman, D., Krystofova, S., Rasmussen, C., Metzenberg, R. L., Perkins, D. D., Kroken, S., Cogoni, C., Macino, G., Catcheside, D., Li, W., Pratt, R. J., Osmani, S. A., DeSouza, C. P., Glass, L., Orbach, M. J., Berglund, J. A., Voelker, R., Yarden, O., Plamann, M., Seiler, S., Dunlap, J., Radford, A., Aramayo, R., Natvig, D. O., Alex, L. A., Mannhaupt, G., Ebbole, D. J., Freitag, M., Paulsen, I., Sachs, M. S., Lander, E. S., Nusbaum, C., and Birren, B. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature (London) 422*, 859–868.

(20) Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R., Batzoglou, S., Lee, S. I., Basturkmen, M., Spevak, C. C., Clutterbuck, J., Kapitonov, V., Jurka, J., Scazzocchio, C., Farman, M., Butler, J., Purcell, S., Harris, S., Braus, G. H., Draht, O., Busch, S., D'Enfert, C., Bouchier, C., Goldman, G. H., Bell-Pedersen, D., Griffiths-Jones, S., Doonan, J. H., Yu, J., Vienken, K., Pain, A., Freitag, M., Selker, E. U., Archer, D. B., Penalva, M. A., Oakley, B. R., Momany, M., Tanaka, T., Kumagai, T., Asai, K., Machida, M., Nierman, W. C., Denning, D. W., Caddick, M., Hynes, M., Paoletti, M., Fischer, R., Miller, B., Dyer, P., Sachs, M. S., Osmani, S. A., and Birren, B. W. (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature 438*, 1105–1115.

(21) Bankel, L., Lindstedt, G., and Lindstedt, S. (1977) Thymine 7-hydroxylase from *Neurospora crassa*. Substrate specificity studies. *Biochim. Biophys. Acta 481*, 431–437.

(22) Sedgwick, B. (2004) Repairing DNA-methylation damage. *Nat. Rev. Mol. Cell Biol. 5*, 148–157.

(23) Kondo, H., Nakabeppu, Y., Kataoka, H., Kuhara, S., Kawabata, S., and Sekiguchi, M. (1986) Structure and expression of the alkB gene of *Escherichia coli* related to the repair of alkylated DNA. *J. Biol. Chem. 261*, 15772–15777.

(24) Aravind, L., and Koonin, E. V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol. 2*, research0007.10007.8.

(25) Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., and Sedgwick, B. (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature (London) 419*, 174–178.

(26) Falnes, P. O., Johansen, R. F., and Seeberg, E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature (London) 419*, 178–182.

(27) Bhattacharya, S. K., Ramchandani, S., Cervoni, N., and Szyf, M. (1999) A mammalian protein with specific demethylase activity for mCpG DNA. *Nature (London) 397*, 579–583.

(28) Cedar, H., and Verdine, G. L. (1999) Gene expression. The amazing demethylase. *Nature (London) 397*, 568–569.

(29) Wolffe, A. P., Jones, P. L., and Wade, P. A. (1999) DNA demethylation. *Proc. Natl. Acad. Sci. U.S.A. 96*, 5894–5896.

(30) Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000) Demethylation of the zygotic paternal genome. *Nature (London) 403*, 501–502.

(31) Koivisto, P., Robins, P., Lindahl, T., and Sedgwick, B. (2004) Demethylation of 3-methylthymine in DNA by bacterial and human DNA dioxygenases. *J. Biol. Chem. 279*, 40470–40474.

(32) Yu, Z., Genest, P. A., ter Riet, B., Sweeney, K., DiPaolo, C., Kieft, R., Christodoulou, E., Perrakis, A., Simmons, J. M., Hausinger, R. P., van Luenen, H. G., Rigden, D. J., Sabatini, R., and Borst, P. (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res. 35*, 2107–2115.

(33) LaFrancois, C. J., Fujimoto, J., and Sowers, L. C. (1998) Synthesis and characterization of isotopically enriched pyrimidine deoxynucleoside oxidation damage products. *Chem. Res. Toxicol. 11*, 75–83.

(34) LaFrancois, C. J., Yu, K., and Sowers, L. C. (1998) Quantification of 5-(hydroxymethyl)uracil in DNA by gas chromatography/mass spectrometry: Problems and solutions. *Chem. Res. Toxicol. 11*, 786–793.

(35) Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectrom. 5*, 976–989.

(36) Matsudaira, P. (1990) Limited N-terminal sequence analysis. *Methods Enzymol. 182*, 602–613.

(37) Sauder, J. M., Arthur, J. W., and Dunbrack, R. L., Jr. (2000) Large-scale comparison of protein sequence alignment algorithms with structure alignments. *Proteins 40*, 6–22.

(38) Ward, J. J., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2003) Secondary structure prediction with support vector machines. *Bioinformatics 19*, 1650–1655.

(39) Privat, E., and Sowers, L. C. (1996) Photochemical deamination and demethylation of 5-methylcytosine. *Chem. Res. Toxicol. 9*, 745–750.

(40) Burdzy, A., Noyes, K. T., Valinluck, V., and Sowers, L. C. (2002) Synthesis of stable-isotope enriched 5-methylpyrimidines and their use as probes of base reactivity in DNA. *Nucleic Acids Res. 30*, 4068–4074.

(41) Kinzie, S. D., Thevis, M., Ngo, K., Whitelegge, J., Loo, J. A., and Abu-Omar, M. M. (2003) Posttranslational hydroxylation of human phenylalanine hydroxylase is a novel example of enzyme self-repair within the second coordination sphere of catalytic iron. *J. Am. Chem. Soc. 125*, 4710–4711.

(42) Muller, I., Stuckl, C., Wakeley, J., Kertesz, M., and Uson, I. (2005) Succinate complex crystal structures of the alpha-ketoglutarate-dependent dioxygenase AtsK: Steric aspects of enzyme self-hydroxylation. *J. Biol. Chem. 280*, 5716–5723.

(43) Ryle, M. J., Koehntop, K. D., Liu, A., Que, L., Jr., and and Hausinger, R. P. (2003) Interconversion of two oxidized forms of taurine/alpha-ketoglutarate dioxygenase, a non-heme iron hydroxylase: Evidence for bicarbonate binding. *Proc. Natl. Acad. Sci. U.S.A. 100*, 3790–3795.

(44) Wilmouth, R. C., Turnbull, J. J., Welford, R. W., Clifton, I. J., Prescott, A. G., and Schofield, C. J. (2002) Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*. *Structure 10*, 93–103.

(45) Kavli, B., Slupphaug, G., Mol, C. D., Arvai, A. S., Peterson, S. B., Tainer, J. A., and Krokan, H. E. (1996) Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *EMBO J. 15*, 3442–3447.

(46) Agarwalla, S., LaPorte, S., Liu, L., Finer-Moore, J., Stroud, R. M., and Santi, D. V. (1997) A novel dCMP methylase by engineering thymidylate synthase. *Biochemistry 36*, 15909–15917.

(47) Handa, P., Acharya, N., and Varshney, U. (2002) Effects of mutations at tyrosine 66 and asparagine 123 in the active site pocket of *Escherichia coli* uracil DNA glycosylase on uracil excision from synthetic DNA oligomers: Evidence for the occurrence of long-range interactions between the enzyme and substrate. *Nucleic Acids Res. 30*, 3086–3095.

(48) Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature (London) 362*, 709–715.

(49) Mullaart, E., Lohman, P. H., Berends, F., and Vijg, J. (1990) DNA damage metabolism and aging. *Mutat. Res. 237*, 189–210.

(50) Zuo, S., Boorstein, R. J., and Teebor, G. W. (1995) Oxidative damage to 5-methylcytosine in DNA. *Nucleic Acids Res. 23*, 3239–3243.

(51) Burdzy, A., Noyes, K. T., Valinluck, V., and Sowers, L. C. (2002) Synthesis of stable-isotope enriched 5-methylpyrimidines and their use as probes of base reactivity in DNA. *Nucleic Acids Res. 30*, 4068–4074.

(52) Haushalter, K. A., Todd Stukenberg, M. W., Kirschner, M. W., and Verdine, G. L. (1999) Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Curr. Biol. 9*, 174–185.

(53) Baker, D., Liu, P., Burdzy, A., and Sowers, L. C. (2002) Characterization of the substrate specificity of a human 5-hydroxymethyluracil glycosylase activity. *Chem. Res. Toxicol. 15*, 33–39.

(54) Cannon-Carlson, S. V., Gokhale, H., and Teebor, G. W. (1989) Purification and characterization of 5-hydroxymethyluracil-DNA glycosylase from calf thymus. Its possible role in the maintenance of methylated cytosine residues. *J. Biol. Chem. 264*, 13306–13312.

(55) Boorstein, R. J., Chiu, L. N., and Teebor, G. W. (1989) Phylogenetic evidence of a role for 5-hydroxymethyluracil-DNA glycosylase in the maintenance of 5-methylcytosine in DNA. *Nucleic Acids Res. 17*, 7653–7661.

(56) Rogstad, D. K., Liu, P., Burdzy, A., Lin, S. S., and Sowers, L. C. (2002) Endogenous DNA lesions can inhibit the binding of the AP-1 (c-Jun) transcription factor. *Biochemistry 41*, 8093–8102.

(57) Valinluck, V., Tsai, H. H., Rogstad, D. K., Burdzy, A., Bird, A., and Sowers, L. C. (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res. 32*, 4100–4108.

TX8004482