A Small, Thermostable, and Monofunctional Chorismate Mutase from the Archeon *Methanococcus jannaschii*†

Gavin MacBeath,‡ Peter Kast,§ and Donald Hilvert*,§

Departments of Chemistry and Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, and Laboratorium für Organische Chemie, Swiss Federal Institute of Technology (ETH), Universitätstrasse 16, CH-8092 Zürich, Switzerland

Received February 25, 1998; Revised Manuscript Received April 22, 1998

ABSTRACT: The gene for chorismate mutase (CM) from the archeon *Methanococcus jannaschii*, an extreme thermophile, was subcloned and expressed in *Escherichia coli*. This gene, which belongs to the *aroQ* class of CMs, encodes a monofunctional enzyme ($AroQ_f$) able to complement the CM deficiency of an *E. coli* mutant strain. The purified protein follows Michaelis−Menten kinetics ($k_{cat} = 5.7\ s^{-1}$ and $K_m = 41\ \mu M$ at 30 °C) and displays pH-independent activity in the range of pH 5−9. Its activation parameters [$\Delta H^{\ddagger} = 16.2$ kcal/mol, $\Delta S^{\ddagger} = -1.7$ cal/(mol·K)] are similar to those of another well characterized AroQ class CM, the mesophilic $AroQ_p$ domain from *E. coli*. Like $AroQ_p$, the thermophilic CM is an α-helical dimer, but approximately 5 kcal/mol more stable than its mesophilic counterpart as judged from equilibrium denaturation studies. The possible origins of the thermostability of *M. jannaschii* $AroQ_f$, the smallest natural CM characterized to date, are discussed in light of available sequence and tertiary structural information.

Chorismate mutase (CM)[1] occupies a central position in the biosynthetic pathway leading to the aromatic amino acids Phe and Tyr in bacteria, fungi, and plants (*1*). CM catalyzes the Claisen rearrangement of chorismate to prephenate. Because this is a rare example of an enzyme-accelerated pericyclic reaction, the enzyme has received much attention during the past 3 decades (*1*−*5*; and references cited therein). With the recent determination of crystal structures of three natural CMs (*6*−*9*), together with a wealth of primary sequence information (Figure 1), it has become clear that at least two entirely different protein folds have evolved to accelerate the chorismate-to-prephenate rearrangement by factors of $>10^6$ (*3*).

The members of one class of CM, represented by the CM domain of the bifunctional *Escherichia coli* CM−prephenate dehydratase, are all-helix-bundle proteins. Most monofunctional prokaryotic CMs, and all CM portions of bifunctional enzymes, show clear primary structural similarity to the monofunctional CM encoded by $aroQ_f$ from *Erwinia herbicola* (see Figure 1). To indicate the relationship among these catalysts, the CMs (or CM domains of bifunctional enzymes) were recently renamed by Jensen and co-workers as $AroQ_f$, $AroQ_p$, $AroQ_t$, or $AroQ_d$, where the subscript indicates whether the CM is monofunctional or fused to a prephenate dehydratase, a prephenate dehydrogenase, or a 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase domain, respectively (*10*). Although CM proteins from eukaryotes share very little primary amino acid sequence similarity with the prokaryotic AroQ proteins (Figure 1), their tertiary structures are clearly related as was shown by comparison of the crystal structures of *E. coli* $AroQ_p$ and the allosteric CM from *Saccharomyces cerevisiae* (*11*). In particular, the helix-bundle topology and crucial catalytic residues are conserved (*11*) (Figure 1). We therefore propose extending the AroQ designation to the known eukaryotic CMs (formerly known as AroR), as well. Since the characterized eukaryotic CMs also contain sequences for a regulatory domain, we will refer to these (entire) polypeptides as $AroQ_r$.

The AroQ class can be contrasted with the AroH class of CM, whose prototypic representative is AroH from *Bacillus subtilis*. AroH has a completely different three-dimensional structure consisting of a trimeric pseudo α/β-barrel (*6*, *7*). Interestingly, however, the active sites of both enzyme classes are similarly functionalized (*3*−*5*), suggesting that they arose by a process of convergent evolution. To date, only one

† This work was supported by the National Institutes of Health through Grant GM37283 and by the Skaggs Institute for Chemical Biology at the Scripps Research Institute. G.M. is the recipient of a Natural Sciences and Engineering Research Council of Canada 1967 Centennial Postgraduate Scholarship and an Eli Lilly Graduate Student Fellowship.

* To whom correspondence should be addressed. Phone: +41-1-632-3176. Fax: +41-1-632-1486. E-mail: hilvert@org.chem.ethz.ch.

‡ Present address: Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford St., MB229, Cambridge, MA 02138.

§ Present address: Swiss Federal Institute of Technology, Universitätstr. 16, CH-8092 Zürich, Switzerland.

[1] Abbreviations: CD, circular dichroism; CM, chorismate mutase; EcCM, CM domain consisting of the 96 N-terminal amino acids of *Escherichia coli* CM−prephenate dehydratase ($AroQ_p$·PheA) followed by a Glu-$(His)_6$ tag; ESI-MS, electrospray ionization mass spectrometry; GdmCl, guanidinium chloride; IPTG, isopropyl 1-thio-β-D-galactopyranoside; MALDI-MS, matrix-assisted laser desorption/ionization mass spectrometry; MjCM, the 99 residue *Methanococcus jannaschii* CM domain ($AroQ_f$) fused to a Leu-Glu-$(His)_6$ tag; MjCM′, the first 93 residues of MjCM, followed by Leu-Glu-$(His)_6$; MW, molecular weight; PCR, polymerase chain reaction; PMSF, phenylmethanesulfonyl fluoride; SDS, sodium dodecyl sulfate.

S0006-2960(98)00449-8 CCC: $15.00
Published on Web 06/20/1998

other sequence has appeared in GenBank that shows strong similarity to AroH; it is encoded by the putative open reading frame sll0109 from the cyanobacterium *Synechocystis* sp. strain PCC6803.

The completion of the first genome sequencing project for an archeon, the hyperthermophilic and prototrophic *Methanococcus jannaschii* (*12*), offered an excellent opportunity to explore how an extreme thermophile that grows optimally at 85 °C accomplishes the chorismate-to-prephenate transformation. From the evolutionary distance of the archeon to either Bacteria or Eukarya, it was intriguing to determine to which class the archeal enzyme would belong. Based on sequence similarity analysis, we find evidence for only a single CM gene in *M. jannaschii*'s genome. The corresponding protein, the smallest natural CM characterized to date, is a thermostable dimer of the AroQ class which is active over a wide temperature range. We report here an extensive biochemical and biophysical characterization of its properties.

MATERIALS AND METHODS

Strains and Plasmids. General cloning was carried out in *E. coli* strain XL1-Blue (Stratagene). Genetic complementation tests were performed using the chorismate mutase-deficient *E. coli* strain KA12 transformed with plasmid pKIMP-UAUC. Both KA12 and pKIMP-UAUC have been described previously (*5*). Protein production was carried out in *E. coli* strain KA13 (*13*), a derivative of KA12 which carries the λDE3 prophage in its chromosome allowing IPTG-inducible expression of genes under the control of the T7 promoter (*14*). The genotype of KA13 is Δ*(srlR-recA) 306*::Tn*10*, Δ*(pheA-tyrA-aroF)*, *thi-1*, *endA-1*, *hsdR17*, Δ*(argF-lac)U169*, *supE44*, λ (DE3) [(*lacUV5*-expressed T7-RNA-pol gene), *imm21*, Δ*nin5*, *Sam7* (*int*$^-$)].

Plasmid clone AMJHG11 contains a putative chorismate mutase-coding region [$aroQ_f$, initially designated MJ0246 (*12*)] from *Methanococcus jannaschii*. The clone was obtained from the American Type Culture Collection (ATCC 625982). Plasmids pKECMT-W and pKECMB-W were described previously (*15*). They both harbor a 3′-truncated version of the *pheA* gene [recently renamed to $aroQ_p$·*pheA* (*10*)] of *E. coli*. While full-length *pheA* encodes the bifunctional chorismate mutase−prephenate dehydratase enzyme, our truncated *pheA*′ ($aroQ_p$) encodes just the chorismate mutase domain ($AroQ_p$), consisting of the 96 natural N-terminal amino acids followed by a Glu-$(His)_6$ tag (EcCM). Expression of *pheA*′ is driven by the *trc* and *bla* promoters in pKECMT-W and pKECMB-W, respectively. Plasmid pET-22b(+) was from Novagen.

DNA Manipulations. All nucleic acid manipulations were according to standard procedures (*16*). Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs. Oligonucleotides were obtained by custom synthesis from GIBCO BRL Life Technologies. Polymerase chain reactions (PCRs) were performed using *pfu* polymerase (Stratagene). All relevant portions of the constructed plasmids were confirmed by DNA sequencing on an Applied Biosystems 373 Automated DNA Sequencer using dye terminator nucleotides and chain termination chemistry (*17*). DNA was prepared for sequencing using a QIAGEN Miniprep kit.

Construction of Plasmids. pKMCMT-W was constructed as follows. First, three overlapping PCR products were generated separately, using AMJHG11 as the template DNA. Unique restriction sites were incorporated into the primer sequences (underlined) without altering the corresponding protein sequence. PCR1 (83 bp) was obtained with primers MCM1S (AAAGGAAGCATATGATCGAGAAACTTGCTGAAATTAGGA) and MCM1N (AGCAATTAGCTTAAGTATCTTATTGTCAATCTCATCAAT), PCR2 (213 bp) was obtained with primers MCM2S (AATAAGATACTTAAGCTAATTGCTGAAAGAAATAGT) and MCM2N (CTTTTGGAGGGCCTTATTATGCTCTATAAGTATTTGAA), and PCR3 (72 bp) was obtained with primers MCM3S (GCATAATAAGGCCCTCCAAAAGCAATATCTTGAGGAA) and MCM3N (TGGTGGTGCTCGAGTTTTTTGTTTTTATTTTGTGTTTCCT). Following gel purification, the three overlapping PCR products were combined in equimolar ratios and assembled into one 322 bp product (PCR4) in a single PCR, using MCM1S and MCM3N as outside primers. The 300 bp *Nde*I−*Xho*I fragment of PCR4 was then ligated into the 2801 bp *Nde*I−*Xho*I fragment of pKECMT-W to yield pKMCMT-W (3101 bp).

To investigate the effects of decreased gene expression on genetic complementation (see below), the 300 bp *Nde*I−*Xho*I fragment of pKMCMT-W was inserted into the 2624 bp *Nde*I−*Xho*I fragment of pKECMB-W to yield pKMCMB-W (2924 bp). For production of proteins on a large scale, the relevant genes were expressed from the T7 promoter. pET-EcCM (5652 bp) was constructed by ligating the 288 bp *Nde*I−*Xho*I fragment from pKECMT-W with the 5364 bp *Nde*I−*Xho*I fragment from the T7-promoter vector pET-22b(+). We have subsequently found that pET vectors can cause unwanted translational read-through at the TGA stop codon (*13*). To circumvent the problem of heterogeneous protein production, we constructed pET-22b-pATCH, a derivative of pET-22b(+), by following the protocol outlined elsewhere (*13*). pET-EcCM-pATCH (5668 bp) was obtained in an analogous way from pET-EcCM. pET-MjCM-pATCH (5680 bp) was constructed by ligating the 300 bp *Nde*I−*Xho*I fragment of pKMCMT-W with the 5380 bp *Nde*I−*Xho*I fragment of pET-22b-pATCH. A second version of pET-MjCM-pATCH was prepared in which the last six codons of $aroQ_f$ were deleted. A 245 bp PCR product was obtained using template pET-MjCM-pATCH and primers MCM2S and MCM4N (TGGTGGTGCTCGAGTGTTTCCTCAAGATATTGCTTTTGGA) (*Xho*I site underlined). The 222 bp *Afl*II−*Xho*I fragment of this product was then ligated with the 5440 bp *Afl*II−*Xho*I fragment of pET-MjCM-pATCH to yield pET-MjCM′-pATCH.

Computer-Assisted Sequence Analysis. Gene sequences were retrieved via the Internet from GenBank (http://www.ncbi.nlm.nih.gov/Web/Search/index.html) (*18*) or The Institute for Genomic Research (TIGR), Rockville, MD (http://www.tigr.org/tdb/mdb/mdb.html), unless stated otherwise. Sequence analysis was done using the program collection of the Genetics Computer Group, Inc., Madison, WI. The G+C content was retrieved from the Codon Usage Database (http://www.dna.affrc.go.jp/~nakamura/codon.html).

Genetic Complementation. M. jannaschii $aroQ_f$ expression plasmids were tested for chorismate mutase activity in vivo by complementation of the chorismate mutase deficiency of

the KA12/pKIMP-UAUC system (*5*). Individual, purified transformants were streaked to yield single colonies on agar plates made with selective medium M9c (without L-Phe and L-Tyr) or with M9c supplemented with 20 μg/mL L-Phe and L-Tyr. The plates were incubated at 30 °C for 3 days, and cell growth was recorded and compared with control transformants without a CM gene (pK1F7T-0-transformed) or with the wild-type *Bacillus subtilis* CM gene *aroH* (pKCMT-W-transformed). Media, procedures, and control plasmids used for the complementation tests were as detailed previously (*5*, *15*).

Production and Purification of Proteins. EcCM was produced with plasmid pET-EcCM-pATCH. Plasmids pET-MjCM-pATCH and pET-MjCM′-pATCH were employed for the production of the full-length (MjCM) and truncated (MjCM′) versions of the putative, monofunctional chorismate mutase from *M. jannaschii*. MjCM is composed of all 99 residues specified by the open reading frame MJ0246, followed by Leu-Glu-$(His)_6$. MjCM′ is composed of the first 93 residues of MjCM, followed by Leu-Glu-$(His)_6$. The host strain for protein production was KA13. Cells from a single colony were grown in 500 mL of LB medium supplemented with 150 μg/mL sodium ampicillin at 37 °C up to an OD_{600} of 0.8. The cultures were cooled to room temperature, and isopropyl 1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 0.4 mM. After 16 h induction at room temperature, the cells were harvested and resuspended in 20 mL of PBS (10 mM phosphate, 160 mM NaCl, pH 7.5) supplemented with 100 μM phenylmethanesulfonyl fluoride (PMSF) and 2 μg/mL pepstatin A and aprotinin. Following cell lysis by passage through a French press, insoluble material was removed by centrifugation (28000*g*, 20 min, 4 °C) and the supernatant loaded onto a column packed with 5 mL of Ni-NTA agarose (QIAGEN) that had been preequilibrated with PBS. The column was thoroughly washed with PBS containing 30 mM imidazole, and bound protein was subsequently eluted with PBS containing 250 mM imidazole.

To obtain pure protein devoid of misfolded aggregates (see Results), EcCM was denatured by addition of 4 volumes of PBS containing 8 M guanidinium chloride (GdmCl) and subsequently refolded by 20-fold dilution into PBS (final concentration <6 μM protein). The protein was concentrated by ultrafiltration (Amicon stirred-cell concentrator) and dialyzed extensively against PBS. Pure, correctly folded MjCM and MjCM′ were obtained by preparative gel filtration using a Superdex 75 (26/60) FPLC column from Pharmacia. All three proteins were stored in PBS supplemented with 100 μM PMSF, 2 μg/mL pepstatin A, 2 μg/mL aprotinin, 200 μM 1,10-phenanthroline, and 0.02% sodium azide. Prior to kinetic or biophysical analysis, the protease inhibitors and preservatives were removed by size-exclusion column chromatography using a Superose 12 (10/30) FPLC column from Pharmacia. Protein concentration was determined using the Pierce Micro BCA Protein Assay Reagent (Pierce) with bovine serum albumin as the calibration standard.

Mass Spectrometry. Proteins were prepared for mass spectrometry by desalting on a NAP-5 column (Pharmacia) that had been preequilibrated with 1% aqueous acetic acid. Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry was performed on a PerSeptive Biosystems Voyager-Elite mass spectrometer with delayed extraction, and electrospray ionization (ESI) mass spectrometry was carried out on an API III Perkin-Elmer SCIEX triple quadrupole mass spectrometer. All mass spectrometry experiments were done at the Scripps Mass Spectrometry Laboratory.

Analytical Size-Exclusion Column Chromatography. The oligomeric states of EcCM, MjCM, and MjCM′ were determined by analytical size-exclusion column chromatography using a Superose 12 (10/30) FPLC column from Pharmacia. Chromatography was performed at room temperature with PBS as the running buffer. A standard curve was prepared according to the method outlined in the instruction manual for the LMW Gel Filtration Calibration Kit from Pharmacia. All proteins were used at concentrations of 0.1−0.2 mg/mL. The void volume, V_o, and the total bed volume, V_t, were determined using blue dextran 2000 and L-tyrosine, respectively. Proteins for the standard curve came from Pharmacia (ribonuclease A, MW 13 700; ovalbumin, MW 43 000; bovine serum albumin, MW 67 000) and Sigma (aprotinin, MW 6500; cytochrome *c*, MW 12 400; carbonic anhydrase, MW 29 000). The elution parameter, K_{av}, was calculated for each protein using the equation $K_{av} = (V_e - V_o)/(V_t - V_o)$, where V_e is the elution volume of the protein.

Circular Dichroism Spectroscopy. All circular dichroism (CD) experiments were performed on an Aviv Circular Dichroism Spectropolarimeter, Model 61DS, equipped with a single-position thermoelectric cuvette holder. CD spectra were recorded at 20 °C in PBS, with a protein concentration of 5 μM and a path length of 0.2 cm. Spectra were obtained by averaging five wavelength scans from 260 to 200 nm in 0.5 nm steps, with a signal averaging time of 2 s and a bandwidth of 1.5 nm. Thermal denaturation curves were obtained using 1 μM protein in PBS [deoxygenated with $He_{(g)}$] in a 1.0 cm cuvette with constant stirring. The ellipticity at 222 nm (1.5 nm bandwidth) was measured from 10 to 95 °C in 1 °C steps. For each point, the sample was first equilibrated for 2 min and then the signal was averaged for 1 min before moving to the next temperature.

Equilibrium Denaturation: (A) Circular Dichroism. Urea-induced denaturation of EcCM was followed by measuring the ellipticity of the sample at 222 nm (1.5 nm bandwidth). Measurements were made in a 1 cm × 1 cm quartz cuvette at 20 °C with constant stirring. Initially, 4 μM protein in PBS was present in the cuvette. Each point in the curve was then obtained by adding the appropriate volume of a ∼10 M urea stock solution in PBS. For each point, the sample was allowed to equilibrate for 4 min, and then the ellipticity was recorded every 1 s for 60 s and averaged. The averaged value was then corrected for background signal from the buffer (linear interpolation between 0 M urea and urea concentration in stock solution) and subsequently adjusted to correct for dilution of the protein. The accurate concentration of the ∼10 M stock solution of urea in PBS was determined from ΔN (the difference between its refractive index and that of PBS) using the equation [urea] = $117.66(\Delta N) + 29.753(\Delta N)^2 + 185.56(\Delta N)^3$ (*19*).

Guanidinium chloride (GdmCl)-induced denaturation of MjCM′ was followed by measuring the ellipticity of the sample at 222 nm (1.5 nm bandwidth). Measurements were made in a 0.1 cm quartz cuvette at 20 °C. Each point in the curve was obtained separately by mixing appropriate volumes of PBS, MjCM′ in PBS, and ∼8 M GdmCl in PBS to achieve

a final concentration of 10 μM protein. The samples were allowed to equilibrate at 20 °C for at least 20 min. The ellipticity was then recorded every 1 s for 60 s and averaged. The averaged value was corrected for background signal from the buffer (linear interpolation between 0 M GdmCl and the GdmCl stock solution). The accurate concentration of the ~8 M stock solution of GdmCl in PBS was calculated from ΔN (the difference between its refractive index and that of PBS) using the equation [GdmCl] = $57.147(\Delta N) + 38.68(\Delta N)^2 - 91.60(\Delta N)^3$ (*20*).

(B) Fluorescence. Urea-induced denaturation of EcCM and GdmCl-induced denaturation of EcCM and MjCM′ were also followed by fluorescence using an Aminco Bowman Series 2 Luminescence Spectrometer. Tyrosine fluorescence was measured by excitation at 278 nm (4 nm bandwidth) and emission at 320 nm (8 nm bandwidth). Measurements were made in a 1 cm × 1 cm quartz cuvette at 20 °C with constant stirring. Initially, 10 μM protein in PBS was present in the cuvette (4 μM for the urea-induced denaturation of EcCM). Each point on the curve was obtained by adding the appropriate volume of either a ~10 M urea stock solution in PBS or ~8 M GdmCl in PBS. For each point, the sample was allowed to equilibrate for 4 min, and then the fluorescence was recorded every 1 s for 60 s and averaged. The averaged value was first corrected for the background fluorescence of the buffer and then adjusted to correct for dilution of the protein.

(C) Gel Filtration. Denaturation of MjCM′ was investigated by GdmCl gradient size-exclusion column chromatography (*21*). Experiments were performed at room temperature using a Superose 12 (10/30) FPLC column from Pharmacia (1 × 30 cm; volume ~23 mL). A linear gradient of GdmCl in PBS was formed by mixing 60 mL of PBS containing 2 M GdmCl with 60 mL of PBS containing 6 M GdmCl (flow rate of 0.4 mL/min). This resulted in a gradient of ~26 mM/cm column length and an increment in [GdmCl] of 0.4 M every 30 min. Successive 100 μL samples of MjCM′ (50 μM) were injected every 30 min, and the elution profile was monitored by recording the absorbance at 280 nm (path length of 1 cm).

(D) Data Analysis. To derive $\Delta G_U(H_2O)$, the free energy of unfolding in PBS at 20 °C, the data were fitted to the following two-state model:

$$N_2 \rightleftharpoons 2D \tag{1}$$

where N_2 is the native dimeric protein and D is the denatured monomer. Pre- and posttransition regions were fitted to straight lines, and the fraction of unfolded protein, f_U, was calculated as described (*22*). The equilibrium constant for the unfolding reaction at any given denaturant concentration is

$$K_U = [D]^2/[N_2] = 2P_t[f_U^2/(1 - f_U)] \tag{2}$$

where P_t is the total concentration of polypeptide (*23*). The free energy of unfolding, ΔG_U, was then calculated at each concentration of denaturant in the transition region of the data using the relation:

$$\Delta G_U = -RT \ln (K_U) \tag{3}$$

Linear extrapolation of ΔG_U to 0 M denaturant yields $\Delta G_U(H_2O)$. Some denaturation data were collected by diluting the protein with concentrated stock solutions of denaturant. The different protein concentrations at each point in the resulting curves were accounted for in the calculation of K_U. For presentation purposes, f'_U, the fraction of unfolded protein at a single, constant polypeptide concentration, P'_t, was calculated from the observed value of f_U at concentration P_t using the relation:

$$2P'_t[f'^2_U/(1 - f'_U)] = 2P_t[f_U^2/(1 - f_U)] \tag{4}$$

and solving for f'_U (equation not shown). Though formally necessary, this correction produced an almost imperceptible change in the appearance of the data since P_t varied by no more than 2-fold in any given experiment.

For both urea and GdmCl, ΔG_U is often found to vary linearly with the concentration of denaturant:

$$\Delta G_U = mx + \Delta G_U(H_2O) \tag{5}$$

where x is the concentration of denaturant and m is the slope of the line. Combining eqs 2, 3, and 5, we can express f_U as a function of m, x, and $\Delta G_U(H_2O)$:

$$f_U = \{-e^{[-mx-\Delta G_U(H_2O)]/RT} + \sqrt{e^{[-2mx-2\Delta G_U(H_2O)]/RT} + 8P_t e^{[-mx-\Delta G_U(H_2O)]/RT}}\}/4P_t \tag{6}$$

This equation was used for the curve fits shown in Figures 7 and 8 (KaleidaGraph, Abelbeck Software).

Kinetics. All kinetic measurements were performed in PBS unless otherwise indicated. Initial rates were determined by monitoring the disappearance of chorismate spectrophotometrically at 274 nm ($\epsilon_{274\ nm} = 2630\ M^{-1}\ cm^{-1}$) using a Cary 3 Bio UV−Visible Spectrophotometer equipped with a thermoelectric cuvette holder. All initial rates were corrected for the corresponding temperature-specific background reaction. Steady-state kinetic parameters k_{cat} and K_m were calculated from the initial rates as described (*24*), using a minimum of five substrate concentrations ranging from at least 4-fold below K_m to at least 4-fold above K_m. Thermodynamic activation parameters $\Delta H^‡$ and $\Delta S^‡$ for MjCM were obtained from least-squares fitting of data collected from 10 to 70 °C using the equation

$$k_{cat} = (\mathbf{k}T/h)e^{-[(\Delta H^‡/RT)-(\Delta S^‡/R)]} \tag{7}$$

(*25*). $\Delta G^‡$ at 25 °C was calculated using the equation

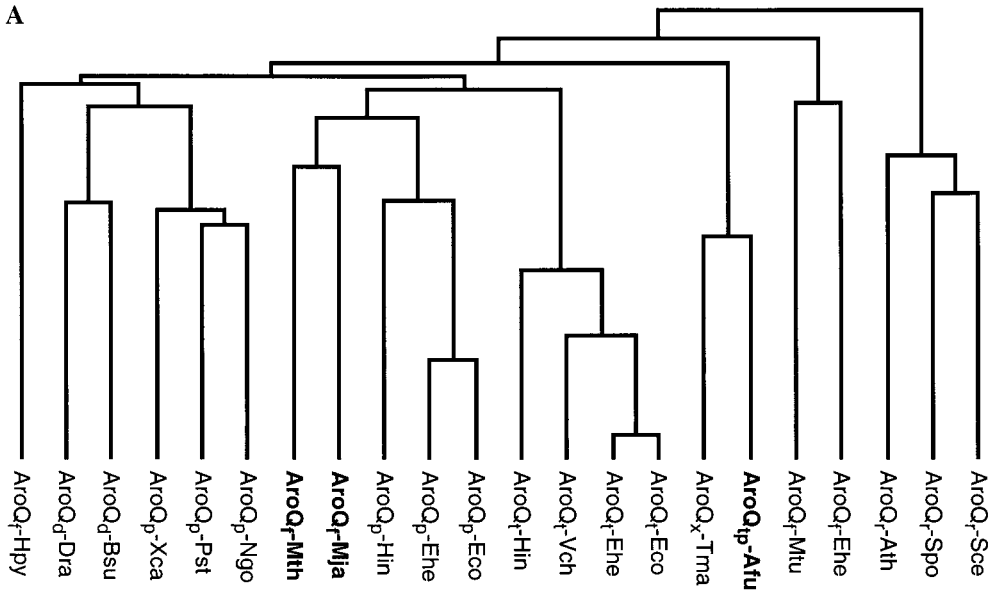
$$\Delta G^‡ = \Delta H^‡ - T\Delta S^‡ \tag{8}$$

with $T = 298.15$ K.

To investigate the pH dependence of the rate of the MjCM-catalyzed reaction, initial rates were determined at 20 °C, using 20 nM MjCM and 12 μM chorismate in Universal Buffer [33 mM sodium acetate, 33 mM 2-(*N*-morpholino)-ethanesulfonic acid, 33 mM 1,3-bis[tris(hydroxymethyl)-methylamino]propane, 100 mM sodium chloride, adjusted to the desired pH with either aqueous hydrochloric acid or aqueous sodium hydroxide].

RESULTS

Sequence and Structural Comparisons. The recently published genomic sequence of the methanogenic archeon



B

```
AroQr-Sce  -LISKRDGDD KNNFGSVATR DIECLQSLSR RIHFGKFVAE AKFQSDIPLY TKLIKSKDVE .GIMKNITNS AVEEKILERL TKKAEVYG.. .......... .VDPTNESGE RRITPEYLV-
AroQr-Spo  -KISSP.GDD FDNYGSTVVC DIRCLQSLSR RIHYGKFVAE AKYLANPEKY KKLILARDIK .GIENEIVDA AQEERVLKRL HYKALNYGR. ..........D AADPTKPSD. .RINADCVA-
AroQr-Ath  -RLVKKGDDG ..NYGSTAVC DAICLQCLSR RIHYGKFVAE AKFQASPEAY ESAIKAQDKD .RLMDMLTFP TVEDAIKKRV EMKTRTYGQE VKVGMEEKEE EEEEGNESHV YKISPILVG-

AroQf-Ehe  .......... ...GSVSSVS LGSLSSALNE RMQVMKAVAG YKAL.....H HLPIEDLPRE QVVLDHM..L QNAQQAGLEP HSVEPFVHAL MNASKTIQYR YRADWLSSPD SAVP.VRDL-
AroQf-Mtu  .......... .....DGTSQ LAELVDAAAE RLEVADPVAA FKWR.....A QLPIEDSGRV EQQLAKL..G EDARSQHIDP DYVTRVFDDQ IRATEAIEYS RFSDWKLNPA SAPPEPPDL-

AroQtp-Afu -KATKKAESI EELRGLIKSI DSLILRLIER RIDAARQIAR IKME.....R GEPIELKDVE EEKLWEVMSK TTLNPVKLK. EIFEGIMSLA KEEEYKVAGV KYTIAVLGPQ GSFSEEMAL-
AroQx-Tma  -AALRDPSDI TEMRHKIRLI DFILLRFIEE RMKVARDVAK EKKK.....T DSPIVIRDIE EQKIKEILDR TDLNPVRMR. RIFEEIMKLA KEEEYRLLGI SKTVAVLGPM GSFSDEMAM-
AroQt-Eco  .....MVAEL TALRDQIDEV DKALLNLLAK RLELVAEVGE VKSRF..... GLPIYVPERE ASMLASRRA. .EAEALGVPP DLIEDVLRRV MRESYSSEND KGFKTLCPSL RPVVIVGGG-
AroQt-Ehe  .....MVAEL TALRDQIDSV DKALLDLLAK RLELVAEVGE VKSRY..... GLPIYVPERE ASMLASRRK. .EAEALGVPP DLIEDVLRRV MRESYTSEND KGFKTLCPEL RPVVIVGGK-
AroQt-Vch  .....MAVEL NQLRDQIDEV DKQMVELLAR RLALVEQVGQ VKSRY..... GLPIYAPDRE AAMLASRRT. .EAESKGVPP QLIEDILRRT MRESYAS*KD SGFKCLNPEL RSVVIVGGN-
AroQt-Hin  ...MSFMEAL KDLRSEIDSL DRELIQLFAK RLELVSQVGK VKHQH..... GLPIYAPERE IAMLQARRL. .EAEKAGISA DLIEDVLRRF MRESYANENQ FGFKTINSDI HKIVIVGGY-
AroQp-Eco  ...MTSENPL LALREKISAL DEKLLALLAE RRELAVEVGK AKLLS..... HRPVRDIDRE RDLLERLITL GKAHH..LDA HYITRLFQLI IEDSVLTQQA LLQQHLNKIN PHSARIAFL-
AroQp-Ehe  ...MNPDNPL LALRDKISAV DKKLLTLLAE RRLLAVEVAQ AKLAT..... HRPIRDVERE RALLENLIVL GKAHN..LDA HYITRLFQLV IEDSVLTQQA LLQKNLNHPH AHAARIAFL-
AroQp-Hin  .....MALEL SDIRQQITQI DRSLLKLLSE RHRLAFDVVR SKEIS..... QKSLRDVERE QQLLQELVQF AENENYQLEA QYITSIFQKI IEDSVLTQQV YLQNKLNEQR NQNLHIAFL-
AroQf-Mja  .....MIEKL AEIRKKIDEI DNKILKLIAE RNSLAKDVAE IKNQL..... GIPINDPERE KYIYDRIRKL CKEHN..VDE NIGIKIFQIL IEHNKALQKQ YLEETQNKNK K.........
AroQf-Mth  MDEYRAREVL RRSRQKIDGI DRDILDLITS RIALAREIAE AKEVL..... GMEILDPERE LQIIERTRKI ARENG..IDE NKLTELMKIL MDLSKTEQKE MLRRQ..... ..........
AroQp-Ngo  .MSQTIDELL IPHRNAIDTI DAEILRLLNE RAQHAHAIGE LKGT..... .GAVYRPERE VAVL...RRI QDLNKGPLPD ESVARLFREV MSECLAVERP .LTIAYLGPQ GTFTQQAAI-
AroQp-Pst  ...MSEADQL KALRVRIDSL DERILDDLISE RARCAQEVAR VKTASWPKAE EAVFYRPERE AWVL...KHI MELNKGPLDN EEMARLFREI MSSCLALEQP .LRVAYLGPE GTFSQAAAL-
AroQp-Xca  -PLATAAPVL ADVRAKIDEI DRGIQALIAE RANFAHQVGK AKGK...LAA AVDYYRPERE AQVL...RMV VDRNEGPLSD EVLVHVFREI MSACLAQQEP .LKIGYLGPE GTFSQQAVL-
AroQd-Bsu  ....MSNTEL ELLRQKADEL NLQILKLINE RGNVVKEIGK AKEA....QG .VNRFDPVRE RTML...NNI IENNDGPFEN STIQHIFKEI FKAGLELQEE DHSKALLVSR KKKPEDTIV-
AroQd-Dra  ...MTQQRSI DDLRAEVDQI NRDLLALISR RGEVVAQIGH AKSA....EG RPNHYDPARE EQQL...KEL EKINPGPFPT ATVKGIFKEI FRASLALEES NDKKQLLVSR KVKREDTVL-
AroQf-Hpy  .MQKNLDSLL ENLRAEIDAL DNELSDLLDK RLEIALKIAL IK.......Q ESPIYCPKRE QEIL...KRL SQRDFKHLNG EILTGFYTEV FKISRKFQEN ALKELKK... ..........

Cat. site (AroQp-Eco)     R              R          K            RE                    Q
```

FIGURE 1: Phylogenetic tree (A) and multiple sequence alignment (B) of (putative) CMs. The tree and the alignment were created with the GCG program PileUp (gap weight 3.0; gap length weight 0.1). The final comparison only includes the sequence segments shown in (B). A dash indicates that the sequence was truncated. Incomplete *aroQ* sequences found in the databases and those for PchB and PapB (see Discussion) were not considered. The alignment of the three eukaryotic sequences [on the right in (A)] was aided by superposition of crystal structure data (*11*). Only the segments of the $AroQ_r$ proteins corresponding to the region of $AroQ_p$-Eco encompassing residues Arg28 and Lys39 are aligned for the purpose of illustration; the large number of insertions in the eukaryotic genes requires separate alignments of other homologous sections, as reported in (*11*). The subscript after AroQ indicates whether the (putative) CM domain is monofunctional ($AroQ_f$), provided with a regulatory domain ($AroQ_r$), or fused to prephenate dehydratase ($AroQ_p$), prephenate dehydrogenase ($AroQ_t$), or 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase ($AroQ_d$) domains. Missing sequence information did not allow us to make an assignment for $AroQ_x$-Tma. In $AroQ_{tp}$-Afu, the AroQ domain is located between an N-terminal prephenate dehydrogenase and a C-terminal prephenate dehydratase domain. The three archaeal proteins in (A) are emphasized in boldface type. The source of each sequence is as follows: Sce, *Saccharomyces cerevisiae* (Acc. No. M24517); Spo, *Schizosaccharomyces pombe* (Acc. No. Z98529); Ath, *Arabidopsis thaliana* (Acc. No. Z26519); Ehe, *Erwinia herbicola* (Acc. No. M95628, X60420); Mtu, *Mycobacterium tuberculosis* (Acc. No. U38939); Afu, *Archaeoglobus fulgidus* (TIGR, coding region AF0227); Tma, *Thermotoga maritima* (TIGR, Contig BTMDD43R); Eco, *Escherichia coli* (Acc. No. M10431); Vch, *Vibrio cholerae* (TIGR, sequence GVCDG49F); Hin, *Haemophilus influenzae* (Acc. No. U32809, U32794); Mja, *Methanococcus jannaschii* (Acc. No. U67480); Mth, *Methanobacterium thermoautotrophicum* (Ohio State University, coding region mt0804); Ngo, *Neisseria gonorrhoeae* (University of Oklahoma, Contig 314); Pst, *Pseudomonas stutzeri* (Acc. No. M73971); Xca, *Xanthomonas campestris* (*10*); Bsu, *Bacillus subtilis* (Acc. No. X65945); Dra, *Deinococcus radiodurans* (TIGR, sequence gdr_119); Hpy, *Helicobacter pylori* [TIGR, coding region HP0291, (*49*)]. Some of the sequences were retrieved from ongoing genome sequencing projects and may still contain errors. For example, the stop codon (*) in AroQt-Vch is probably either Gln or Glu (*10*). Residues shown in boldface type are Arg11, Arg28, Lys39, Arg51, Glu52, and Gln88 of $AroQ_p$-Eco whose side chains are involved in prominent interactions with a bound transition state analogue (see Figure 2B). Only the mature sequences resulting from (putative) cleavage of a signal peptide are shown for $AroQ_f$-Ehe (*50*) and $AroQ_f$-Mtu (PSORT; http://psort.nibb.ac.jp/). These proteins have a topology different from that of the prototypic AroQ fold (unpublished).

M. jannaschii annotates two coding regions as chorismate mutase subunit A (MJ0246) and subunit B (MJ0612) (*12*). The 99 amino acid protein encoded by MJ0246 exhibits strong similarity to known CMs from the AroQ class. As shown in Figure 1A, it clusters with a group of CMs from the γ subdivision of proteobacteria. Whereas these bacterial CMs constitute the N-terminal domains of bifunctional proteins also carrying prephenate dehydratase activity, the archeal protein appears to be a monofunctional CM as judged by the size of its coding region and by the absence of

appropriate protein coding sequences immediately 3′ to the gene. In accordance with the nomenclature by Jensen (*10*), we have thus renamed MJ0246 as *aroQ*$_f$ to indicate its monofunctionality and relationship to the AroQ class.

On reanalysis, the translated second coding region MJ0612, in contrast to MJ0246, shows no homology to any known CM. Instead, the translated N-terminal 300-residue sequence shows high similarity to bacterial prephenate dehydrogenase domains, but the remaining C-terminal 146 amino acids show no convincing homology to any known protein. If the latter segment encodes a CM at all, it must belong to a structurally unique class of enzyme.

The multiple sequence alignment in Figure 1B illustrates that the gene product of *M. jannaschii aroQ*$_f$ (AroQ$_f$) contains all of the strongly conserved amino acids characteristic of this family of enzymes. They correspond to residues that line the active site as identified in the crystal structure of a typical AroQ domain, the CM portion of *E. coli* AroQ$_p$·PheA [(*8*), Figure 2]. Based on our sequence analysis, we therefore expected the *M. jannaschii* protein to be a functional, dimeric CM with a 6-helix-bundle structure similar to the *E. coli* AroQ$_p$ domain.

Subcloning of M. jannaschii aroQ$_f$ *and Genetic Complementation*. We obtained the plasmid clone AMJHG11 from the American Type Culture Collection. It contains the putative chorismate mutase gene *aroQ*$_f$ (MJ0246) in a 2101 bp fragment of chromosomal DNA from *M. jannaschii* inserted into the *Sma*I site of cloning vector pUC18 (*12*). The gene's orientation in AMJHG11 is opposite to the *lac* promoter of pUC18, and is thus lacking bacterial transcription and translation initiation signals. To test for in vivo activity of *aroQ*$_f$ in *E. coli*, we constructed plasmids pKMCMB-W and pKMCMT-W. These plasmids provide the weak *bla* and the roughly 20-fold stronger *trc* promoter upstream of *aroQ*$_f$, respectively, allowing for different levels of gene expression in vivo (*26*).

The two *aroQ*$_f$-containing plasmids were introduced into *E. coli* KA12/pKIMP-UAUC cells to test for complementation of the strain's CM deficiency, as described under Materials and Methods. The growth of the pKMCMB-W and pKMCMT-W transformants was compared with CM-negative (pK1F7T-0-transformed) and CM-positive (pKCMT-W-transformed) control cells. The tests showed that *aroQ*$_f$ of *M. jannaschii* provides sufficient chorismate mutase activity in vivo at 30 °C to complement the genetic defect of the assay strain, but growth was retarded compared to the positive control. Both the *bla* and the *trc* expressed genes were of comparable efficiency on selective media, but the construct with the *trc* promoter yielded much smaller colonies than either the controls or pKMCMB-W on nonselective minimal agar plates (data not shown). This indicates that MjCM (the His-tagged form of *M. jannaschii* AroQ$_f$) interferes with the growth of *E. coli* cells when produced at elevated levels.

Production and Purification of Proteins. Efficient production of EcCM (the His-tagged form of the *E. coli* AroQ$_p$ domain) and MjCM was achieved using the T7 expression system (*14*). The genes for EcCM and MjCM were each subcloned into pET-22b-pATCH, a derivative of pET-22b(+) (Novagen) that has been modified to prevent unwanted translational readthrough at the TGA stop codon (*13*). The resulting plasmids (pET-EcCM-pATCH and pET-MjCM-pATCH, respectively) were introduced into host strain KA13, a *recA*-deleted K-12 strain of *E. coli* which is devoid of endogenous *E. coli* chorismate mutase activity. Upon induction of the chromosomally integrated T7 RNA polymerase gene with IPTG (see Materials and Methods), efficient production of soluble protein was achieved for both EcCM and MjCM. The yield of recombinant protein was about 5-fold higher for EcCM than for MjCM, possibly due to differences in codon usage between the bacterium *E. coli* and the archeon *M. jannaschii* (*27*).

The C-terminal (His)$_6$-tag attached to the produced proteins allowed efficient purification by affinity chromatography on a matrix containing chelated Ni^{2+} ions (*28*). Both EcCM and MjCM were purified in this way under native conditions,

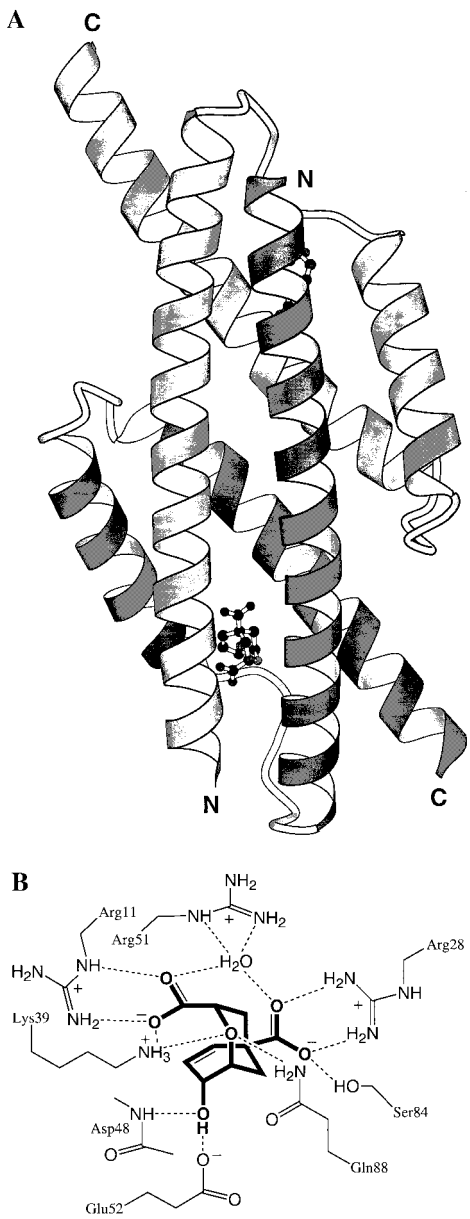


FIGURE 2: Structure of EcCM, the prototype of the AroQ class of CMs (*3*, *8*). The active site is occupied by an endo-oxabicyclic dicarboxylic acid inhibitor that mimics the transition state of the CM reaction. (A) Three-dimensional structural representation of the dimeric CM. The figure was generated with the program MOLSCRIPT (*51*). Coordinates for the EcCM crystal structure were from the Protein Data Bank, Brookhaven National Laboratory, PDB accession code 1ECM (*8*). (B) Simplified schematic view of the active site with catalytically important amino acid residues.

yielding samples of >95% purity as judged by SDS−polyacrylamide gel electrophoresis and Coomassie blue staining. Analysis by electrospray ionization mass spectrometry (ESI-MS) and by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) confirmed the identity of each protein and indicated that the N-terminal methionine residue (Met1) had been removed in most of the EcCM sample but was intact in the MjCM sample. The experimentally determined molecular masses correlated well with the expected masses, both for EcCM lacking Met1 (expected [M−H$^+$], 11 886.7 Da; observed$_{ESI-MS}$, 11 883 Da; observed$_{MALDI-MS}$, 11 881 Da) and for full-length MjCM (expected [M−H$^+$], 12 847.9 Da; observed$_{ESI-MS}$, 12 846 Da; observed$_{MALDI-MS}$, 12 843 Da).

Upon prolonged storage of MjCM, a discrete truncated fragment of the protein was observed by SDS−polyacrylamide gel electrophoresis (data not shown). Analysis of the partially degraded sample by MALDI-MS suggested cleavage of MjCM following Asn95 (predicted [M−H$^+$], 11 284.1 Da; observed 11 277 Da). This results in loss of the four C-terminal residues of the protein (Lys-Asn-Lys-Lys) as well as the appended Leu-Glu-(His)$_6$ sequence. In contrast to the full-length protein, the proteolytic fragment failed to bind to Ni-NTA agarose resin, consistent with our assignment of the cleavage site. Sequence alignment of MjCM with other CMs (Figure 1B) indicates that the last six residues of MjCM extend beyond the C-terminus of our version of EcCM. Since EcCM is fully active (see below), these residues are unlikely to be structurally important. In fact, the observed susceptibility of MjCM to proteolysis between residues 95 and 96 suggests a lack of compact structure in this region. This is supported by the crystal structure of an extended version of EcCM (Figure 2) which shows that residues beyond position 95 are located in a solvent-exposed α-helix (*8*). To avoid heterogeneity in the sample, the expression vector pET-MjCM-pATCH was modified to remove the bases encoding the six C-terminal residues of AroQ$_f$, yielding pET-MjCM′-pATCH. The resulting truncated protein, MjCM′, was produced and purified as before. Analysis of the purified protein by ESI-MS and MALDI-MS confirmed its integrity (expected [M−H$^+$], 12 106.0 Da; observed$_{ESI-MS}$, 12 106 Da; observed$_{MALDI-MS}$, 12 104 Da). Even after prolonged storage, MjCM′ showed no indication of proteolytic degradation as judged by SDS−polyacrylamide gel electrophoresis. MjCM′ was similar to MjCM in terms of its far-UV circular dichroism (CD) spectrum and indistinguishable in terms of its aggregation state, thermodynamic stability, and kinetic parameters (k_{cat} and K_m). All data presented below, with the exception of kinetic measurements and gel filtration data, were obtained using MjCM′ rather than MjCM.

Analytical size-exclusion column chromatography (see below) showed that EcCM, MjCM, and MjCM′ preparations were composed of a mixture of correctly folded, dimeric protein (25−50%) and various misfolded, higher-order aggregates (remaining 50−75%). In the case of EcCM, the higher-order aggregates were converted into fully active, dimeric protein by denaturation with guanidinium chloride and subsequent renaturation upon dilution into native buffer (PBS, see Materials and Methods). The efficiency of refolding was >95% as judged by analytical gel filtration, native polyacrylamide gel electrophoresis, and kinetic analysis.

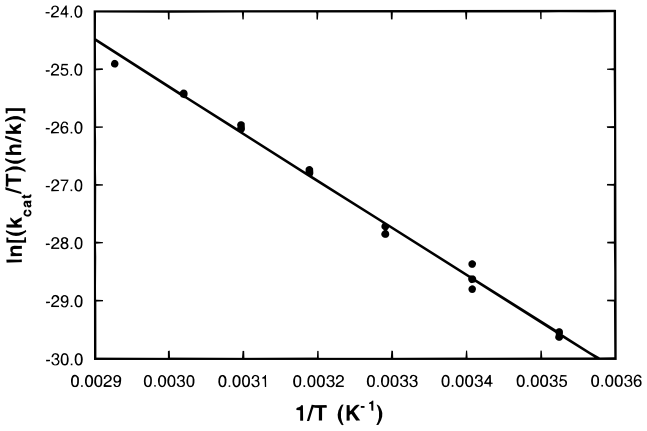


FIGURE 3: Temperature dependence of k_{cat} for the MjCM-catalyzed transformation as determined from individual Michaelis−Menten plots between 10 and 70 °C.

In contrast, we chose to isolate the compact, well-behaved, and catalytically active dimeric MjCM by preparative gel filtration to avoid potential ambiguities resulting from the refolding experiments. Although this approach reduced the final yield of protein by up to 75%, we were able to obtain sufficient pure, homogeneous MjCM and MjCM′ for kinetic and biophysical analysis (∼3 mg/L of culture).

Kinetic Properties of MjCM. Enzyme assays with purified MjCM (and MjCM′) confirmed expectations that *M. jannaschii aroQ$_f$* specifies a chorismate mutase. The enzyme follows classical Michaelis−Menten kinetics. We determined the activation parameters for the MjCM-catalyzed reaction by measuring the temperature dependence of k_{cat} over a wide temperature range (10−70 °C). From the data in Figure 3, an enthalpy of activation $\Delta H^{\ddagger}$ of 16.2 (±0.4) kcal/mol and an activation entropy $\Delta S^{\ddagger}$ of −1.7 (±1.2) cal/(mol·K) were obtained. At 30 °C, a k_{cat} value of 5.7 s^{-1} and a K_m of 41 μM were determined from interpolation of all Michaelis−Menten data sets gathered over the whole temperature range. We found that k_{cat}/K_m for the MjCM-catalyzed chorismate rearrangement is virtually independent of pH in the enzyme reaction mixture over the range of pH 5−9 [k_{cat}/K_m = 6.0 (±0.9) × 10^4 M^{-1} s^{-1} at 20 °C].

Analytical Size-Exclusion Column Chromatography. The oligomeric state of EcCM and MjCM was determined by analytical size-exclusion column chromatography as described under Materials and Methods. Figure 4 summarizes the results obtained on a calibrated Superose column. A plot of the elution parameter K_{av} versus log(MW) for the standard proteins gave a straight line, yielding apparent molecular weights for EcCM and MjCM of 25 790 and 29 390, respectively (2.2 and 2.3 times the molecular weights of the corresponding polypeptides). We conclude that both EcCM and MjCM exist as homodimers in solution (as predicted from the crystal structure of EcCM). Interestingly, the apparent molecular weight of each protein is slightly higher than that expected for the dimeric species. This probably results from the elongated shape of the dimer (Figure 2) and from the addition of a noncompact (His)$_6$-tag to the C-terminus.

Circular Dichroism Spectroscopy. Far-UV circular dichroism (CD) spectra were obtained for EcCM and MjCM′ (Figure 5). The curves are similar in shape and match the spectrum expected for a highly α-helical protein (*29*, *30*).

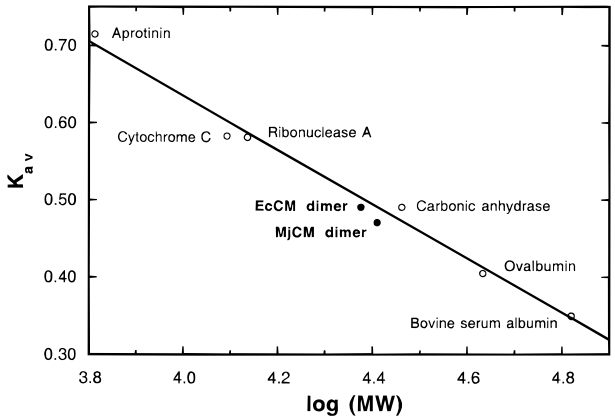


FIGURE 4: Determination of the quaternary structure of EcCM and MjCM by analytical size-exclusion chromatography. The elution parameter K_{av} was calculated for each protein and is plotted against the logarithm of the molecular weight (MW) of the standard proteins (○), or of the MW predicted from the amino acid sequence and assuming a dimeric structure for EcCM and MjCM (●).

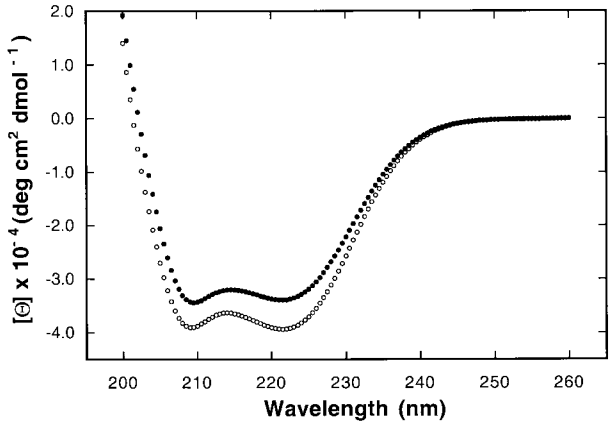


FIGURE 5: Far-UV circular dichroism (CD) spectra of EcCM (○) and MjCM′ (●) at 20 °C. The *y*-axis shows the mean residue ellipticity.

The curves differ in intensity, which might reflect slight differences in structure (degree of α-helicity). However, these intensity differences more likely arise from inaccuracies in the determination of protein concentration (Micro BCA Protein Assay; see Materials and Methods). It should be noted that the CD spectrum of a slightly larger version of EcCM has been reported previously (*31*). Our spectrum differs from that of Zhang et al. both in intensity and in shape. The lower intensity of their spectrum presumably arises from differences in concentration determination and from the presence of additional C-terminal residues in their version of EcCM. The shape of their spectrum is similar to that of our spectrum between 250 and 205 nm, but follows an unexpected downward curve at lower wavelengths that is most likely an artifact resulting from strong buffer absorbance in their sample.

Thermal Denaturation. Temperature-dependent unfolding curves were obtained for EcCM and MjCM′ by measuring the ellipticity at 222 nm (indicative of α-helical structure) as the samples were heated from 5 to 95 °C (Figure 6). As the midpoint of the unfolding transitions, we determined T_m's of 63 and 88 °C for 1 μM EcCM and MjCM′, respectively.

Equilibrium Denaturation and Free Energy of Unfolding. The free energy of unfolding, $\Delta G_U(H_2O)$, was determined

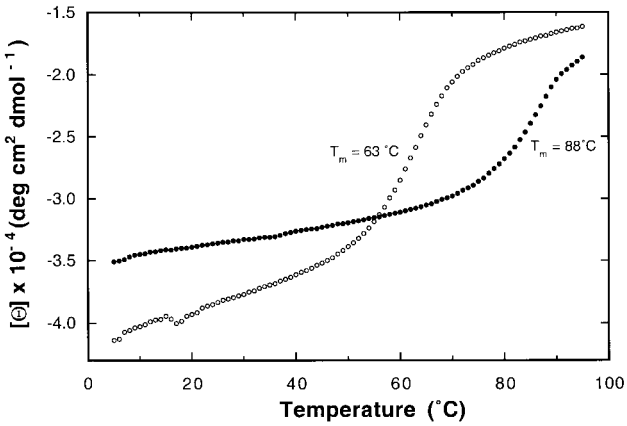


FIGURE 6: Temperature-dependent unfolding (denaturation) curves for EcCM (○) and MjCM′ (●). The change in (α-helical) structure was assessed by measuring the CD at 222 nm. The same protein concentration (1 μM polypeptide) was used for both curves since denaturation is concentration-dependent for multimeric (in this case dimeric) proteins. The midpoint of the unfolding transition, T_m, was obtained for each curve from the maximum of the first derivative with respect to temperature.

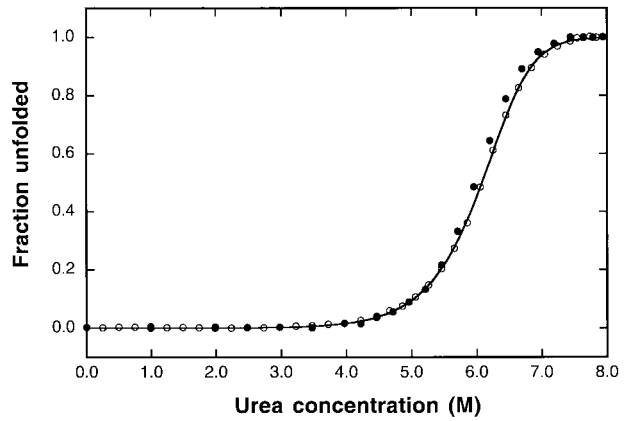


FIGURE 7: Urea-induced denaturation of EcCM at 20 °C. The fraction of unfolded protein was determined by CD (ellipticity at 222 nm; ○) and by fluorescence spectroscopy (tyrosine fluorescence; ●). The curve fit includes all CD data. The concentration of the polypeptide was 4 μM.

for EcCM by equilibrium unfolding experiments using urea as the denaturant. Tyrosine fluorescence was employed as a probe for tertiary structure while CD at 222 nm was employed as a probe for secondary structure. To analyze the data, the following two-state model was assumed:

$$N_2 \rightleftharpoons 2D \qquad (1)$$

where N_2 is the native dimer and D is the denatured monomer. The fraction of unfolded protein, f_U, was calculated as described under Materials and Methods. Figure 7 shows the fraction of unfolded protein as a function of urea concentration, determined independently from both CD and fluorescence data. The coincidence of the two data sets supports our assumption of a two-state model. Given this scheme, the equilibrium constant for the unfolding reaction, K_U, was calculated from eq 2, and the free energy of the unfolding reaction, ΔG_U, was calculated from eq 3 for both data sets. Plots of ΔG_U as a function of urea concentration yielded nearly coincident straight lines in the transition regions of the data (not shown). Based on this observed

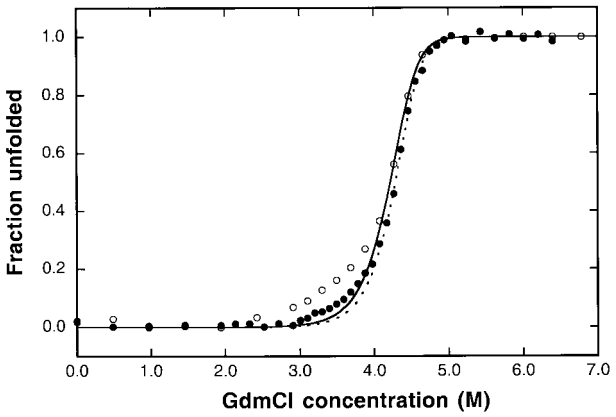


FIGURE 8: Guanidinium chloride-induced denaturation of MjCM′ at 20 °C. The fraction of unfolded protein was determined by CD (ellipticity at 222 nm; ○) and by fluorescence spectroscopy (tyrosine fluorescence; ●). Curve fittings to the late transition regions (see text) are shown as solid (CD) and dotted (fluorescence) lines. The concentration of the polypeptide was 10 μM.

linearity, f_U was fitted to eq 6 for both data sets, yielding values for $\Delta G_U(H_2O)$ of 20.1 kcal/mol (CD) and 20.9 kcal/mol (fluorescence) (average of 20.5 ± 0.4 kcal/mol). A similar value was obtained from fluorescence detection of GdmCl-induced denaturation of EcCM (data not shown), supporting the validity of the linear extrapolation of ΔG_U. From the averaged $\Delta G_U(H_2O)$, and by substituting in eq 6, we can calculate the fraction of unfolded protein, $f_U(P_t)$, in the absence of denaturant at a given protein concentration. At 20 °C and 1 nM polypeptide, only 1.6×10^{-3}% of the protein is unfolded and even at 1 mM the fraction of folded protein is still >99.9%. This suggests that EcCM exhibits substantial stability at high dilutions, a supposition supported by the independence of k_{cat} from protein concentration at nanomolar concentrations (data not shown).

When similar experiments were attempted with MjCM′ using urea as the denaturant, it was found that the protein did not fully unfold, even at concentrations of up to 10 M urea. However, it was possible to obtain denaturation curves using GdmCl, a stronger denaturant than urea. As before, data sets were collected using both tyrosine fluorescence and far-UV CD. As with EcCM, MjCM′ exhibited a clear unfolding transition, and so the data were fitted to the two-state model of eq 1. However, when f_U was plotted as a function of GdmCl concentration, it was evident that the data sets were not entirely coincident (Figure 8). Since eq 6 is based on the two-state model, significant deviations from the fit indicate that the model is inappropriate. While the late transition regions superimpose, the early regions do not. When the CD data, excluding the early transition region, were fitted to eq 6, the excluded data did not lie on the corresponding curve (Figure 8). The early transition may be due, in part, to impurities in the MjCM′ sample. However, as judged from SDS−polyacrylamide gel electrophoresis, MALDI-MS, and size-exclusion chromatography of MjCM′, we have no indication of significant sample heterogeneity caused, for instance, by errors of translation of *M. jannaschii aroQ*$_f$ mRNA in *E. coli* (*27*). Thus, the deviation from the two-state model suggests the presence of an unfolding intermediate. Since the protein is never predominantly in this intermediate state, we cannot determine its spectroscopic properties and hence cannot calculate the fraction of each of the three (or more) species present during the unfolding transition. Nevertheless, it appears that the fluorescence properties of the intermediate are fairly similar to those of the native state. If we assume that the fluorescence of the intermediate is the same as that of the native protein, the intermediate can be ignored as a rough approximation. Fitting all the fluorescence data to eq 6 then yields $\Delta G_U(H_2O) = 24.0 \pm 0.6$ kcal/mol (correlation coefficient $R = 0.99905$, curve fit not shown). This approximation becomes better in the late transition when the observed fluorescence approaches that of the unfolded protein. When the early transition data are excluded from the fit, we obtain $\Delta G_U(H_2O) = 26.9 \pm 0.9$ kcal/mol ($R = 0.99967$; Figure 8). With the caveat that these values are only approximate, it appears that MjCM′ is about 5 kcal/mol more stable than the homologous *E. coli* protein.

Two alternative three-state models can be proposed for the unfolding of MjCM′:

$$N_2 \rightleftharpoons I_2 \rightleftharpoons 2D \quad (9)$$

$$N_2 \rightleftharpoons 2I \rightleftharpoons 2D \quad (10)$$

The oligomeric state of the intermediate was investigated by GdmCl gradient size-exclusion column chromatography (*21*). On a gel filtration column, the fully guanidinium-unfolded monomer (D) elutes earlier (behaves larger) than the compact, folded dimer (N_2). However, a compact, structured, monomeric intermediate (I) should elute later (behave smaller) than the folded dimer. This behavior was observed, for example, in the unfolding of aspartate aminotransferase (*32*). When MjCM′ was analyzed by this technique (see Materials and Methods), however, no evidence was seen for a shift to longer retention times in the early portion of the transition (data not shown). Instead, the overall unfolding was accompanied by a shift to shorter retention times. On the basis of this result, it is likely that the intermediate observed by CD (and more subtly by fluorescence) is dimeric (i.e., eq 9 may describe the transition more appropriately than eq 10).

DISCUSSION

Archeal Members of the AroQ Protein Family. We have demonstrated that the protein specified by the coding region MJ0246 of *M. jannaschii* is a monofunctional chorismate mutase of the AroQ class. It shows higher sequence similarity to a defined subgroup within the γ subdivision of proteobacterial CMs, which normally occur as N-terminal domains of a bifunctional CM−prephenate dehydratase, than to monofuntional bacterial or eukaryotic AroQ domains (Figure 1). In contrast to the assignment of genomic reading frames by Bult et al. (*12*), we find no evidence for any other CMs in *M. jannaschii*, even when the translated genome is searched with representatives of all known CM sequences. Similarly, we find only one putative CM sequence in the very recently published genome of another archeon, *Methanobacterium thermoautotrophicum* (*33*). [NB: The genes mt1220 and mt1640 of *M. thermoautotrophicum* which were assigned as CM genes (*33*) actually encode prephenate dehydratase and prephenate dehydrogenase sequences, respectively.] The encoded protein, AroQ$_f$-Mth (Figure 1), exhibits 43.5% sequence identity to AroQ$_f$-Mja.

Further sequence analysis revealed that homologues of prephenate dehydratase and prephenate dehydrogenase, which are frequently fused to CM domains in bacteria (*10*), also exist as monofunctional proteins in *M. jannaschii* (and in *M. thermoautotrophicum*). Thus, *M. jannaschii* and *E. coli* organize the biosynthetic pathway from chorismate to Phe and Tyr differently (monofunctional versus bifunctional enzymes). Nevertheless, the AroQ domains of these evolutionarily very distantly related organisms share a high degree of primary structure similarity. A third archeal sequence (for a putative CM from *Archaeoglobus fulgidus*; $AroQ_{tp}$-Afu) exhibits much lower similarity to the cluster containing $AroQ_f$-Mja and $AroQ_p$-Eco (Figure 1). The wide scattering of archeal AroQ sequences within the phylogenetic tree may be explained by lateral transfer of *aroQ* genes between specific families of organisms. Alternatively, the degree of evolutionary drift in the *aroQ* family may have been strongly influenced by the location of the protein in the cell or by the nature of its fusion to attached functional domains. AroQ domains with significantly diverged primary structures include the extracellular proteins $AroQ_f$-Mtu and $AroQ_f$-Ehe, where topological rearrangements appear to have occurred (unpublished observation), and the allosterically controlled yeast CM, whose catalytic domain shows clear deviations (insertion of two helices) from the simple α-helical structure of the *E. coli* enzyme (*9*, *11*). Similarly, $AroQ_{tp}$-Afu appears to have diverged significantly from the prototypical AroQ domain. Not only does it lack an otherwise conserved Arg residue (Figure 1), but also it seems to be part of a unique trifunctional protein with a prephenate dehydrogenase fused to its N-terminus and a prephenate dehydratase joined to its C-terminus. It is also conceivable that some of the AroQ proteins shown in Figure 1 have diverged in function during evolution like the recently described AroQ homologues PchB and PapB (sequences not shown) which have proposed isochorismate−pyruvate lyase (*34*) and 4-amino-4-deoxychorismate mutase (*35*) activities, respectively.

Properties of M. jannaschii Chorismate Mutase. Active site residues identified in the crystal structure of the *E. coli* $AroQ_p$ domain (Figure 2B) are strictly conserved in $AroQ_f$-Mja, with the exception of the replacement of the weakly conserved Ser84 by Asn (Figure 1). The pH independence between pH 5 and 9 of k_{cat}/K_m of $AroQ_f$-Mja is consistent with the absence of an ionizable residue at the position corresponding to Gln88 of $AroQ_p$-Eco. The presence of Gln or Glu at this site has been found to correlate, respectively, with a wide or a narrow (and acidic) pH optimum of the corresponding CM catalyst (*31*, *36*, *37*). Our data support the notion (*11*) that the homologue of position 88 must be protonated for efficient AroQ activity.

Although the activity of $AroQ_f$-Mja is expected to be optimized for the growth temperature of the source organism (85 °C), the catalytic parameters (k_{cat}, K_m, k_{cat}/K_m) at 30 °C for MjCM (5.7 s^{-1}, 41 μM, 1.4×10^5 s^{-1} M^{-1}; this work) are quite comparable to those of other well-characterized CMs such as $AroQ_p$-Eco [23 s^{-1}, 300 μM, 7.8×10^4 s^{-1} M^{-1}; (*31*, *38*)] and *B. subtilis* AroH [46 s^{-1}, 67 μM, 6.9×10^5 s^{-1} M^{-1}; (*39*)]. The enthalpy and entropy of activation for the MjCM-catalyzed chorismate rearrangement were determined by measuring the temperature dependence of k_{cat} over a temperature span of 60 °C. The broader temperature range accessible to the thermostable enzyme is a significant advantage for the accuracy of the determination, compared to the very narrow range that can be probed in the case of mesophilic enzymes [for instance, $AroQ_p$-Eco was measured between 10 and 37 °C (*38*)]. Nevertheless, the values for the activation parameters of MjCM [$\Delta H^{\ddagger} = 16.2$ kcal/mol, $\Delta S^{\ddagger} = -1.7$ cal/(mol·K)] are similar to the ones reported for $AroQ_p$-Eco, the other well-characterized AroQ class CM [$\Delta H^{\ddagger} = 16.3$ kcal/mol, $\Delta S^{\ddagger} = 1.5$ cal/(mol·K); (*38*)]. Both parameter sets are very different from the one measured for the AroH class CM from *B. subtilis* [$\Delta H^{\ddagger} = 12.7$ kcal/mol, $\Delta S^{\ddagger} = -9.1$ cal/(mol·K); (*39*)]. While it is tempting to interpret these activation parameters in mechanistic terms (*3*, *24*, *38*), such speculation is not warranted until chemistry is established as rate limiting (*39*). Kinetic isotope effect studies on a closely related AroQ CM ($AroQ_t$-Eco) suggest that the rate-determining transition state actually precedes the chemical step in these enzymes (*40*). Additional data are clearly needed, but given the coincidence of the activation parameters for at least two AroQ proteins, the reaction mechanism is likely to be conserved across the AroQ class.

The spontaneous rearrangement of chorismate to prephenate is strongly temperature dependent. Using the activation parameters for the uncatalyzed reaction (*41*), an increase in k_{uncat} of 230-fold can be calculated for a rise in temperature from 30 to 85 °C, yielding a half-life for chorismate of 4.5 min at 85 °C. Given the efficiency of the uncatalyzed process, why does *M. jannaschii*, which grows optimally at 85 °C, employ an enzyme for this transformation? By making a series of estimations [intracellular chorismate concentration ≈ K_m ≈ 100 μM (*42*); CM concentration ≈ 1 μM, based on unpublished results in *E. coli*], we deduce that the total flux of chorismate in the cell in the presence of the enzyme would be roughly 3 orders of magnitude more efficient than the spontaneous background reaction at 85 °C. Catalysis of the conversion of chorismate to prephenate is particularly important at higher temperatures in light of the ready decomposition of chorismate into undesired side products (*41*, *43*). An efficient CM also gives the thermophilic organism the flexibility to survive at lower temperatures [*M. jannaschii* has been found to grow at 48 °C (*12*)] where the spontaneous reaction is less efficient.

MjCM differs from EcCM primarily in its stability, as assessed by thermal and chemical denaturation experiments. MjCM has a 25 °C higher midpoint for its thermal unfolding transition (at 1 μM concentration), and we estimate from the results of guanidinium-induced denaturation that the thermophilic protein is ca. 5 kcal/mol more stable than its mesophilic counterpart. Whereas EcCM denatures via a single-step process, an intermediate is detected in the unfolding studies with MjCM′. Based on the size-exclusion chromatography experiments, we assume this species is a partially denatured dimer. The presence of such an intermediate would indicate that the subunit interactions, which form much of the core of $AroQ_f$-Mja, are particularly strong and remain intact even when peripheral portions of the protein begin to unfold.

The Origins of Thermostability. A comparison of the relevant amino acid sequence of MjCM′ with the corresponding 93 amino acids of EcCM reveals eight Ile for Leu exchanges as the most pronounced difference between the thermophilic and mesophilic enzymes ($AroQ_p$-Eco residue

positions 10, 17, 21, 25, 55, 59, 76, and 79). All of the affected residues are located at or near the interface between the two CM subunits, which would be consistent with the experimental denaturation data that suggest stronger subunit interactions in MjCM′. However, interactions away from the subunit interface almost certainly contribute to the additional stability. Nor is it clear why the very conservative replacements of Leu by Ile should result in a more thermostable protein. This specific Leu to Ile substitution has not emerged in statistical analyses of amino acid differences between other mesophilic and thermophilic proteins (*44*, *45*). The preference for Ile (with its A+T-rich codons) over Leu is not just a consequence of the remarkably low G+C content of the *M. jannaschii* genome (31.9% in the coding regions compared with 51.6% for *E. coli*); the relevant sequence for a putative, thermostable CM ($AroQ_f$-Mth) from *M. thermoautotrophicum* (which has an optimum growth temperature of 65 °C and a G+C content of 50.4% in the coding regions) also exhibits eight Leu to Ile substitutions as the most favored amino acid change relative to $AroQ_p$-Eco. Most of these $AroQ_f$-Mth-specific isoleucines are identical with those in $AroQ_f$-Mja.

Leu to Ile replacements are clearly favored for the two thermophiles in the cluster of five fairly related sequences $AroQ_f$-Mth, $AroQ_f$-Mja, $AroQ_p$-Hin, $AroQ_p$-Ehe, and $AroQ_p$-Eco (Figure 1). This preference becomes less obvious, however, if all prokaryotic AroQ domains compiled in Figure 1 are compared. Among the sequences listed are four from thermophilic organisms, including the archea *M. jannaschii*, *M. thermoautotrophicum*, and *A. fulgidus*, as well as the bacterium *Thermotoga maritima*. A bias for Ile in the thermophiles is still evident, but a significant number of mesophiles also prefer this residue at the corresponding positions (Figure 1). An analysis of other enzyme pairs for which mesophilic and thermophilic representatives are known has suggested that individual proteins can exploit distinctive sets of amino acid replacements to achieve adaptation to warmer environments (*45*). Moreover, there are often no obvious clues as to why certain enzymes are more thermostable than others. Relatively minor changes distributed over the whole protein which may cause better packing of the hydrophobic core, extended secondary structures, decreased structural flexibility, less conformational strain, and more or stronger hydrogen bonds and salt bridges probably all contribute to thermostability (*46*−*48*). Consequently, there appears to be no consensus or "set of special rules" on how to stabilize a given mesophilic protein (*48*). Our sequence analysis of AroQ domains is in line with these observations, and the bias for Ile over Leu in thermophilic CMs of the AroQ class may point to a mechanism specific to this protein family for stabilizing its α-helical scaffold.

In summary, $AroQ_f$-Mja represents the smallest natural chorismate mutase characterized to date. Its size, high intrinsic stability, and the absence of attached domains for other, nonrelated functions should make this protein useful for studies of how structure, stability, and dynamics influence the catalytic efficiency of CM (*52*).

ACKNOWLEDGMENT

We thank Dr. Bruce Ganem for providing us with the gene for EcCM and Dr. Reza Ghadiri for the use of the CD and luminescence spectrometers.

REFERENCES

1. Haslam, E. (1993) *Shikimic Acid: Metabolism and Metabolites*, John Wiley & Sons, New York.
2. Ganem, B. (1996) *Angew. Chem., Int. Ed. Engl. 35*, 936−945.
3. Lee, A. Y., Stewart, J. D., Clardy, J., and Ganem, B. (1995) *Chem. Biol. 2*, 195−203.
4. Kast, P., Hartgerink, J. D., Asif-Ullah, M., and Hilvert, D. (1996) *J. Am. Chem. Soc. 118*, 3069−3070.
5. Kast, P., Asif-Ullah, M., Jiang, N., and Hilvert, D. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 5043−5048.
6. Chook, Y. M., Ke, H., and Lipscomb, W. N. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8600−8603.
7. Chook, Y. M., Gray, J. V., Ke, H., and Lipscomb, W. N. (1994) *J. Mol. Biol. 240*, 476−500.
8. Lee, A. Y., Karplus, P. A., Ganem, B., and Clardy, J. (1995) *J. Am. Chem. Soc. 117*, 3627−3628.
9. Xue, Y., Lipscomb, W. N., Graf, R., Schnappauf, G., and Braus, G. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10814−10818.
10. Gu, W., Williams, D. S., Aldrich, H. C., Xie, G., Gabriel, D. W., and Jensen, R. A. (1997) *Microb. Comp. Genomics 2*, 141−158.
11. Xue, Y., and Lipscomb, W. N. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 10595−10598.
12. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J.-F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, B. P., Borodovsky, M., Klenk, H.-P., Fraser, C. M., Smith, H. O., Woese, C. R., and Venter, J. C. (1996) *Science 273*, 1058−1073.
13. MacBeath, G., and Kast, P. (1998) *BioTechniques 24*, 789−794.
14. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
15. MacBeath, G., Kast, P., and Hilvert, D. (1998) *Protein Sci. 7*, 325−335.
16. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.
17. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
18. Benson, D. A., Boguski, M. S., Lipman, D. J., Ostell, J., and Oullette, B. F. (1998) *Nucleic Acids Res. 26*, 1−7.
19. Warren, J. R., and Gordon, J. A. (1966) *J. Phys. Chem. 70*, 297−300.
20. Nozaki, Y. (1972) in *Methods in Enzymology* (Hirs, C. H. W., and Timasheff, S. N., Eds.) pp 43−50, Academic Press, New York.
21. Endo, S., Saito, Y., and Wada, A. (1983) *Anal. Biochem. 131*, 108−120.
22. Pace, C. N., Shirley, B. A., and Thomson, J. A. (1989) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 311−330, IRL Press, Oxford.
23. Neet, K. E., and Timm, D. E. (1994) *Protein Sci. 3*, 2167−2174.
24. Görisch, H. (1978) *Biochemistry 17*, 3700−3705.
25. Connors, K. A. (1990) *Chemical Kinetics. The Study of Reaction Rates in Solution*, VCH Publishers, New York.
26. Deuschle, U., Kammerer, W., Gentz, R., and Bujard, H. (1986) *EMBO J. 5*, 2987−2994.
27. Kane, J. F. (1995) *Curr. Opin. Struct. Biol. 6*, 494−500.
28. Van Dyke, M. W., Sirito, M., and Sawadogo, M. (1992) *Gene 111*, 99−104.
29. Greenfield, N., and Fasman, G. D. (1969) *Biochemistry 8*, 4108−4116.
30. Chen, Y.-H., Yang, J. T., and Chau, K. H. (1974) *Biochemistry 13*, 3350−3359.

31. Zhang, S., Kongsaeree, P., Clardy, J., Wilson, D. B., and Ganem, B. (1996) *Bioorg. Med. Chem. 4*, 1015−1020.
32. Herold, M., and Kirschner, K. (1990) *Biochemistry 29*, 1907−1913.
33. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrokovski, S., Church, G. M., Daniels, C. J., Mao, J.-I., Rice, P., Nölling, J., and Reeve, J. N. (1997) *J. Bacteriol. 179*, 7135−7155.
34. Serino, L., Reimmann, C., Baur, H., Beyeler, M., Visca, P., and Haas, D. (1995) *Mol. Gen. Genet. 249*, 217−228.
35. Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., Zagorec, M., Schleuniger, J., Bisch, D., Blanche, F., Debussche, L., Crouzet, J., and Thibaut, D. (1997) *Mol. Microbiol. 23*, 191−202.
36. Schnappauf, G., Sträter, N., Lipscomb, W. N., and Braus, G. H. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 8491−8496.
37. Liu, D. R., Cload, S. T., Pastor, R. M., and Schultz, P. G. (1996) *J. Am. Chem. Soc. 118*, 1789−1790.
38. Galopin, C. C., Zhang, S., Wilson, D. B., and Ganem, B. (1996) *Tetrahedron Lett. 37*, 8675−8678. Correction: *Tetrahedron Lett. 38*, 1467 (1997).
39. Kast, P., Asif-Ullah, M., and Hilvert, D. (1996) *Tetrahedron Lett. 37*, 2691−2694.
40. Addadi, L., Jaffe, E. K., and Knowles, J. R. (1983) *Biochemistry 22*, 4494−4501.
41. Andrews, P. R., Smith, G. D., and Young, I. G. (1973) *Biochemistry 12*, 3492−3498.
42. Fersht, A. (1985) *Enzyme Structure and Mechanism*, 2nd ed., W. H. Freeman, New York.
43. Young, I. G., Gibson, F., and MacDonald, C. G. (1969) *Biochim. Biophys. Acta 192*, 62−72.
44. Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frank, G., and Tratschin, J. D. (1979) *Biochemistry 18*, 5698−5703.
45. Menéndez-Arias, L., and Argos, P. (1989) *J. Mol. Biol. 206*, 397−406.
46. Adams, M. W. W. (1993) *Annu. Rev. Microbiol. 47*, 627−658.
47. Jiang, X., Bishop, E. J., and Farid, R. S. (1997) *J. Am. Chem. Soc. 119*, 838−839.
48. Adams, M. W. W., and Kelly, R. M. (1998) *Trends Biotechnol.* (in press).
49. Tomb, J.-F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M., and Venter, J. C. (1997) *Nature 388*, 539−547.
50. Xia, T., Song, J., Zhao, G., Aldrich, H., and Jensen, R. A. (1993) *J. Bacteriol. 175*, 4729−4737.
51. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
52. MacBeath, G., Kast, P., and Hilvert, D. (1998) *Science 279*, 1958−1961.

BI980449T