Crystal Structures of Two Bacterial 3-Hydroxy-3-methylglutaryl-CoA Lyases Suggest a Common Catalytic Mechanism among a Family of TIM Barrel Metalloenzymes Cleaving Carbon-Carbon Bonds*

Farhad Forouhar†, Munif Hussain†, Ramy Farid‡, Jordi Benach, Mariam Abashidze*, William C. Edstrom*, Sergey M. Vorobiev§, Rong Xiao, Thomas B. Acton*, Zhuji Fu*, Jung-Ja P. Kim†, Henry M. Miziorko**, Gaetano T. Montelione*, and John F. Hunt††

From the †Department of Biological Sciences and Northeast Structural Genomics Consortium, Columbia University, New York, New York 10027, ‡Schrödinger, Inc., New York, New York 10036, §Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers University, Department of Biochemistry, Robert Wood Johnson Medical School, and Northeast Structural Genomics Consortium, Piscataway, New Jersey 08854, the †Department of Biochemistry, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, and the **Division of Molecular Biology and Biochemistry, University of Missouri-Kansas City, Kansas City, Missouri 64110

The enzyme 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) lyase catalyzes the terminal steps in ketone body generation and leucine degradation. Mutations in this enzyme cause a human autosomal recessive disorder called primary metabolic aciduria, which typically kills victims because of an inability to tolerate hypoglycemia. Here we present crystal structures of the HMG-CoA lyases from Brucella melitensis and Brucella melitensis at 2.7 and 2.3 Å resolution, respectively. These enzymes share greater than 45% sequence identity with the human orthologue. Although the enzyme has the anticipated triose-phosphate isomerase (TIM) barrel fold, the catalytic center contains a divergent cation-binding site formed by a cluster of invariant residues that cap the core of the barrel, contrary to the predictions of homology models. Surprisingly, the residues forming this cation-binding site and most of their interaction partners are shared with three other TIM barrel enzymes that catalyze diverse carbon-carbon bond cleavage reactions believed to proceed through enolate intermediates (4-hydroxy-2-ketovalerate aldolase, 2-isopropylmalate synthase, and transcarboxylase S5). We propose the name ‘DRE-TIM metalloenzymes’ for this newly identified enzyme family likely to employ a common catalytic reaction mechanism involving an invariant Asp-Arg-Glu (DRE) triplet. The Asp ligates the divalent cation, while the Arg probably stabilizes charge accumulation in the enolate intermediate. The Glu maintains the precise structural alignment of the Asp and Arg. We propose a detailed model for the catalytic reaction mechanism of HMG-CoA lyase based on the examination of previously reported product complexes of other DRE-TIM metalloenzymes and induced fit substrate docking studies conducted using the crystal structure of human HMG-CoA lyase (reported in the accompanying paper by Fu, et al. (2006) J. Biol. Chem. 281, 7526–7532). Our model is consistent with extensive mutagenesis results and can guide subsequent studies directed at definitive experimental elucidation of this enzyme’s reaction mechanism.

The enzyme HMG-CoA lyase (HL) catalyzes the cleavage of 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) to form acetoacetate and acetyl-CoA (Fig. 1A). This reaction represents the terminal steps in both the metabolic degradation of dietary leucine and the formation of ketone bodies, which represent an essential alternative energy source for animals under conditions of low serum glucose concentration (hypoglycemia). The enzyme is localized to the mitochondrial matrix in eukaryotic cells. It is expressed at some level in a wide variety of human tissues but at the highest levels in the kidneys and especially the liver, which is the primary site of ketone body generation (1).

Primary metabolic aciduria is an autosomal recessive disorder caused by mutations in human HL (Homo sapiens HL or hsHL). This disease is rare but widely distributed among racial and ethnic populations. It is generally detected based on the presence of highly elevated concentrations of hydroxyacetocetic acid and related organic acids in blood and urine. Victims of the disease experience episodes of vomiting, malaise, hypotonia, and apnea that can progress to coma. Progressive neurodevelopmental deterioration can be prevented by careful restriction of dietary leucine intake, which reduces the toxic accumulation of HMG-related organic acids, but victims frequently die in early childhood because of the inability to manage hypoglycemic shock (2).

So-called “ketone bodies” represent a set of the following three related compounds that are used to mobilize fat stores into the blood of mammals: acetoacetate, 3-hydroxybutyric acid, and acetone. This set of compounds represents an alternative metabolic energy source for brain, heart, and kidneys and plays an important role in diabetes and some forms of epilepsy (3, 4). Ketone bodies are essential to maintain physiological viability during hypoglycemia because cells die due to energy deprivation when both glucose and ketone body concentrations fall below critical threshold levels. Acetoacetate, the direct product of HL, is

* This work was supported by National Institutes of Health Grants NIGMS-P50-GM62413 and NIGMS-U54-074958 (to the Northeast Structural Genomics Consortium) and DK21491 (to H. M. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Dept. of Biological Sciences, Columbia University, New York, NY 10027. Tel.: 212-854-5443; Fax: 212-865-8246; E-mail: jfhunt@biology.columbia.edu.

‡1 The atomic coordinates and structure factors (codes 1YDO and 1YDN) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

‡2 The abbreviations used are: HL, HMG-CoA lyase; HMG-CoA, 3-hydroxy-3-methylglutaryl-coenzyme A; HMG, 3-hydroxy-3-methylglutarate; r.m.s.d., root mean square deviation; TIM, triose-phosphate isomerase; hsHL, H. sapiens HL; bsHL, Brucella subtilis HL; PDB, Protein Data Bank; bmHL, B. melitensis HL.
reduced to 3-hydroxybutyric acid in an enzyme-catalyzed equilibrium reaction determined by the NAD/NADH ratio in cells. Acetoacetate also spontaneously decarboxylates to generate acetone, completing the repertoire of compounds comprising ketone bodies. Therefore, HL catalyzes what is effectively the terminal step in the generation of ketone bodies, and patients lacking HL activity can therefore experience a lethal metabolic shock in response to hypoglycemia because of the inability to switch to this alternative circulating energy source.

Human genetic studies have revealed the existence of at least 22 different disease-causing point mutations in \( \text{hsHL} \) (5), including 4 nonsense mutations and 18 mutations resulting in amino acid substitutions (at the positions shown in Fig. 2). The results of previous biochemical studies performed on recombinant \( \text{hsHL} \) heterologously expressed in \( \text{Escherichia coli} \) (6–9) suggest that the active site includes at least residues Arg-41, Asp-42, Glu-72, His-233, and Cys-266 (using sequence numbers from the full-length human open reading frame). Site-directed mutagenesis results indicate that mutations in any one of these residues reduce the \( V_{max} \) value of the enzyme by 2–5 orders of magnitude (Table 1). A divalent cation in the form of manganese or magnesium is essential for enzyme activity, with the \( K_m \) values for these cofactors being 0.34 and 233 \( \mu \text{M} \), respectively (6). HL can be considered a metalloenzyme based on the requirement for one of these metal cofactors.

Chromatographic and sedimentation studies suggest that \( \text{hsHL} \) forms predominantly a homodimer in solution with a small population of homotetramer (10). Oxidative stress down-regulates \( \text{hsHL} \) activity via formation of a stable intersubunit disulfide linkage. Although the mechanistic basis of the resulting reduction in enzyme activity is not understood, the disulfide bond holding this complex together is believed to be a symmetrical linkage between pairs of Cys-323 residues (11).

Sequence profiling methods (12) suggest that HL has a TIM barrel fold, with the highest scoring homologue being the TIM barrel domain in DmpG or 4-hydroxy-2-ketovalerate aldolase (13). A homology model for the Michaelis complex of the enzyme has been published based on the structure of HisA (PDB code 1QO2) (14, 15). In this model, the lipophilic moiety of the CoA passes through the core of the TIM barrel (i.e. plugging the core of the central \( \beta \)-barrel with its adenine moiety on one face and the HMG moiety in the active site on the opposite face). A similar binding geometry is observed in one TIM barrel protein interacting with CoA (16) but not in others where it binds instead to a groove
FIGURE 2. Sequence structure alignment of HL orthologues and other DRE-TIM metallolyases. The secondary structural elements and G-loop in the crystal structure of bsHL are shown above the multiple sequence alignment with α-helices represented as rectangles and β-strands as arrows. All secondary structural elements are shared between the two bacterial HL structures with the exception of helix α12, which is absent in bmHL (and colored orange to distinguish it in the schematic). Residue numbers in both bmHL (cyan) and hsHL (green) are indicated at the top of each sequence alignment block. Residues in the B. subtilis and H. sapiens enzymes are uniformly numbered 1-2 and 1-28, respectively, compared with equivalent positions in the B. melitensis enzyme because of variations in the lengths of the N termini of these proteins (including the presence of a mitochondrial targeting sequence in the human enzyme). The numbers in parentheses at the start of some rows in the first sequence alignment block give the number of N-terminal residues in each protein that are omitted from this alignment. Strictly conserved and conservatively substituted residues are colored in red and blue, respectively, in the multiple sequence alignment. The violet diamonds below each block indicate residues observed to be mutated in the HMG-CoA-lyase-deficient patients, and the green circles indicate residues involved in binding Ca²⁺ in the bmHL structure. The T-COFFEE program (44) was used for alignment of the HL sequences, and the other DRE-TIM metallolyases were aligned by the program DALI (31) based on optimized superposition of their three-dimensional structures. DmpG corresponds to PDB accession code 1NVM, LeuA to 1SR9, and transcarboxylase 5S to 1RBQ.
on one face of the barrel. This homology model accounted for some mutagenesis results but did not identify a binding site for the essential divalent cation cofactor.

BLAST searches using the sequence of hsHL (Fig. 2) indicate that likely orthologues with high levels of sequence homology (>45% identity) are present in plants and eubacteria but not in yeast or archaebacteria. The Northeast Structural Genomics Consortium targeted this protein family for structure determination based on the absence of reliable homologues of this fold/function family in the PDB. Three likely orthologues were cloned (as reported at www.nesg.org), but only two able homologues of this fold/function family in the PDB. Three likely orthologues with high levels of sequence homology (E. coli) are present in plants and eubacteria but not in yeast or archaebacteria. The Northeast Structural Genomics Consortium targeted this family for structure determination based on the absence of reliable homologues of this fold/function family in the PDB. Three likely orthologues were cloned (as reported at www.nesg.org), but only two were produced in soluble form in E. coli, those from Bacillus subtilis (bsHL) and Brucella melitensis (bmHL) (Northeast Structural Genomics Consortium targets SR181 and LR35, respectively). In this study, we report crystal structures for both of these proteins, which respectively are 46 and 51% identical to hsHL (according to DALI alignment), along with induced fit computational modeling studies of the Michaelis complex of hsHL with HMG-CoA. The crystal structure of this protein was solved by molecular replacement using the coordinates of bmHL and is reported in the accompanying paper (17). Our results demonstrate a previously unknown but striking similarity in active site structure between HL and several other enzymes cleaving carbon-carbon bonds in metabolic reactions, suggesting that these proteins form a new family of TIM barrel enzymes sharing a common catalytic mechanism.

**MATERIALS AND METHODS**

**Protein Expression**—The bsHL and bmHL enzymes were produced using identical methods following PCR cloning of the corresponding genes (called yngG) from genomic DNA. The full-length coding sequences were cloned into a PET21d (Novagen) derivative for protein expression using a T7 promoter/polymerase system. The resulting sequences were cloned into a pET21d (Novagen) derivative for protein expression strains were cultured in MJ9 minimal media (18) supplemented with selenomethionine, lysine, phenylalanine, threonine, isoleucine, leucine, and valine for the production of selenomethionine-labeled proteins (19). Initial cell growth was carried out at 37 °C, but the temperature was reduced to 17 °C after induction with 1.0 mM isopropyl 1-thio-β-D-galactopyranoside when the absorbance of the cultures at 600 nm reached 1.0. Cells were harvested by centrifugation after over-night incubation, and pellets were resuspended in lysis buffer containing 300 mM NaCl, 10 mM imidazole, 50 mM β-mercaptoethanol, 50 mM NaH2PO4, pH 8.0.

**Protein Purification**—After cell lysis by sonication, extracts were clarified by centrifugation and loaded onto nickel-nitrilotriacetic acid columns (Qiagen) equilibrated in lysis buffer. Following elution with the same buffer containing 250 mM imidazole, fractions containing HL were pooled and loaded onto a Superdex 75 gel filtration column (GE Healthcare) equilibrated in 5 mM dithiothreitol, 50 mM Tris, pH 8.0. Finally, the protein-containing fractions from the second column were concentrated for crystallization (to final concentrations indicated below). Protein purity and molecular mass were verified by SDS-PAGE and matrix-assisted laser desorption ionization time-of-flight mass spectrometry, respectively. Process yields were ~10 mg of purified protein per liter of bacterial culture. In an effort to control the aggregation tendencies of the purified proteins, 100 mM NaCl was added to the gel filtration buffer and therefore crystallization stock in some preparations, but this addition had little influence on either aggregation or crystallization properties for either enzyme.

**Static Light Scattering Measurements**—Samples of the crystallization stocks were injected onto a Shodex 802.5 silica-based gel filtration column (Showa Denko, Tokyo, Japan) equilibrated at 4 °C in 100 mM NaCl, 0.025% (w/v) NaN3, 100 mM Tris-Cl, pH 7.5. The effluent passed through a UV light detector (also at 4 °C), a 15-channel Dawn static light-scattering detector (Wyatt Technologies, Santa Barbara, CA) at room temperature, and an Optilab refractive index detector (Wyatt Technologies) with its sample chamber heated to 40 °C. Debye analyses of molecular mass were conducted after scaling the light-scattering data collected at 11 different angles according to the protein concentration as deduced from the refractive index signal.

**Protein Crystallization**—Both bacterial HL proteins were crystallized at 20 °C using 1:1 hanging-drop vapor diffusion reactions. For bsHL, the reservoir contained 22.5% (w/v) PEG 3350, 210 mM sodium iodide, 5 mM EDTA, 10 mM dithiothreitol. The bsHL crystals grew to full size in 2 days and were transferred to the mother liquor plus 25% ethylene glycol.

**TABLE 1**

Effects of mutations on human HMG-CoA lyase activity

The entries indicate the ratio of the $V_{max}$ observed for each mutant relative to that observed for the wild-type human enzyme ($V_{max} = 191$ units/mg, $K_m = 48 \mu M$). The values in parentheses indicate the ratio of the $K_m$ of the mutant relative to that observed for the wild-type enzyme. All measurements were performed on recombinant human HMG-CoA lyases produced in E. coli as reported previously (6–9,11). The numbers and residue types are indicated for the homologous positions in H. sapiens, B. subtilis, and B. melitensis enzymes. ND indicates not determined.

|                | H. sapiens | B. subtilis | B. melitensis | Glu | Asn | His | Ala |
|----------------|------------|-------------|---------------|-----|-----|-----|-----|
| Asp-42         | Asp-16     | Asp-14      |               | 0.10(0.67)* | 0.0017 (0.90) | 5.2E-6 (ND) | 5.2E-5 (ND) |
| Asp-204        | Asp-178    | Asp-176     |               | 0.050 (3.46) | 1.00 (0.35)    |               |               |
| Asp-280        | Asp-254    | Ala-252     |               | 0.85 (0.56)  |               |               |               |

|                | H. sapiens | B. subtilis | B. melitensis | Glu | Met |
|----------------|------------|-------------|---------------|-----|-----|
| Arg-41         | Arg-15     | Arg-13      |               | 3.7E-6 (1.29)* | 5.2E-8 (ND) |

|                | H. sapiens | B. subtilis | B. melitensis | Arg | Asp | Ala |
|----------------|------------|-------------|---------------|-----|-----|-----|
| His-233        | His-207    | His-205     |               | 1.6E-4(0.92)* | 1.0E-4 (1.13) | 1.6E-4 (1.54) |

|                | H. sapiens | B. subtilis | B. melitensis | Ser | Ala |
|----------------|------------|-------------|---------------|-----|-----|
| Cys-323        | Lys-297    |             |               | 2.2 (1.67) |               |
| Cys-266        | Cys-240    | Cys-238     |               | 0.0016 (0.73) | 1.0E-4 (1.08) |

* Human disease-causing mutations.
glycol for cryoprotection and then flash-frozen in liquid propane. Although crystals were consistently obtained from nine different \textit{bs}HL protein preparations with stock concentrations ranging from 2 to 16 mg/ml, only two crystals were obtained that were large enough for diffraction data to be collected at sufficient resolution for structure determination. One of these decayed during data collection, while the other, which yielded the structure reported in this paper, had a satellite crystal and visible ice rings. The \textit{bs}HL crystals contain two dimers forming a pseudotetramer per asymmetric unit. The \textit{bm}HL protein was crystallized using a reservoir containing 19\% (w/v) PEG 3350, 200 mM CaCl$_2$, 10 mM dithiothreitol. The crystals grew to full size in 6 days and were cryoprotected by transfer to paratone and then flash-frozen in liquid propane. Crystals were obtained from three different \textit{bm}HL protein preparations with stock concentrations ranging from 2.5 to 10 mg/ml. Although many crystals were screened for their diffraction properties, all but one of them had severe morphological twinning problems that prevented collection of high quality diffraction data. The \textit{bm}HL crystals contain four protomers per asymmetric unit. For both bacterial HL crystallizations, the well solution used in vapor diffusion reactions was developed starting from a condition in the unbuffered PEG-Ion Screen from Hampton Research (Laguna Hills, CA), so the pH of the crystallization reactions was controlled by the buffers in the protein crystallization stocks. In both cases, extensive efforts were made to improve the quality of the crystals, including broad additive screens sampling all commonly used monovalent and divalent ions. Over the course of 6 months, it was not possible to find any conditions yielding better crystals or any crystals growing in the presence of Mg$^{2+}$ or Mn$^{2+}$. Entire crystallization screens were conducted for both proteins in the presence of Mg$^{2+}$ plus either HMG-CoA or a substrate analogue, but no viable lead conditions were identified. Mg$^{2+}$ was observed to reduce significantly the solubility of both bacterial HL proteins, a factor that substantially increases the difficulty of obtaining crystals.

\textbf{Structure Solution and Refinement}—For both proteins, single-wavelength anomalous diffraction data sets were collected with crystals maintained at 100 K on beamline X4A at the National Synchrotron Light Source. Data collected at the peak absorption wavelength of selenium were integrated and scaled (Table 2) using the HKL package (20). For the \textit{bs}HL crystals, the program SnB (21) located 21 of 28 selenium sites in the asymmetric unit. These sites were used to initiate iterative phasing and automated model building in SOLVE/RESOLVE (22), which produced a trace for 50\% of the residues in the final model with side chains for 30\%. Most of the residues in the $\beta$-sheet and some in the $\alpha$-helices were inaccurately modeled by this procedure, but completion of the structure was straightforward using iterative cycles of model building in XtalView (23) and computational refinement in CNS (24). For \textit{bm}HL, 24 of 28 possible selenium sites were found by SnB, and 70\% of the final model was built by SOLVE/RESOLVE, including side chains for 45\%. Strong 4-fold non-crystallographic symmetry restraints were maintained using default parameters throughout the models for both crystal structures.

\textbf{Computational Docking of HMG-CoA}—Candidate models for the stereocentrism of the Michaelis complex were produced using ‘induced fit docking’ (25), an automated ligand docking method that accounts for both ligand and protein flexibility. These models were subsequently screened for consistency with published mutagenesis results and qualitative chemical features prioritized through consideration of the structure and chemical reactivity of a series of homologous enzymes (as explained in detail under ‘Results’). Induced fit docking calculations were performed with the program GLIDE 3.5 (Schrodinger, Inc., Portland, OR) (26, 27) using the following parameters aimed at softening the potential: 0.5 van der Waals scaling for both ligand and protein atoms, Coulomb-van der Waals energy cutoff of 100 kcal/mol, and hydrogen bond energy cutoff of $-0.05$ kcal/mol. The calculations sampled $\sim$13,000 conformations and orientations for the HMG-CoA substrate. All atoms of the substrate were constrained to remain in a 32 Å box that completely encompassed the active site, whereas the geometric center of the substrate was constrained to remain in a concentric 12 Å box (centered at orthogonal coordinates of 25.7, 40.6, and 75.9 Å in \textit{hs}HL). Initially, the substrate was docked into a rigid protein model using either the Ca$^{2+}$-bound structure of \textit{bm}HL or the Mg$^{2+}$-bound structure of \textit{hs}HL (reported in the accompanying paper; see Ref. 17). Next, induced fit conformational changes in the receptor were modeled in each of the top 30 ligand-protein complexes ranked according to GLIDE score. PRIME 1.2 (Schrodinger, Inc.) (28, 29) was used for side chain rotamer prediction and energy minimization of the conformation of residues located within 5 Å of the substrate in any of the poses (26). The Ca$^{2+}$- and Mg$^{2+}$-ligating side chains were held fixed during this step. Finally, GLIDE (with default settings) was used to re-dock the substrate into the low energy structures produced by PRIME lying within 60 kcal/mol of the minimum obtained. The resulting models were ranked by a composite score that combines the GLIDE score and 5\% of the PRIME energy, and the top 10 of these were evaluated for consistency with the enzymological criteria explained in detail below.

\textbf{RESULTS AND DISCUSSION}
remain stably associated during gel filtration chromatography. Thus, the distribution of hydrodynamic species formed by \textit{bs}HL is similar to that previously reported for \textit{hs}HL. Equivalent studies on \textit{bm}HL show a mixture of monomers and high order protein aggregates (Fig. 3B). Although only a very small population of large aggregates was detected in hydrodynamic studies of \textit{bs}HL, both proteins had a tendency to spontaneously form amorphous precipitates in their crystallization stocks, a tendency that was significantly exacerbated by freeze/thaw or by the addition of millimolar concentrations of \textit{Mg}^{2+} and/or substrate analogue (30). In both cases, crystallization properties deteriorated following the appearance of amorphous precipitate in the protein stock solutions.

Crystal Structures of Bacterial HL Orthologues—The \textit{hs}HL crystal structure was refined to a free \textit{R}-factor of 30.3% at 2.7 Å, whereas that of \textit{bm}HL was refined to a free \textit{R}-factor of 30.4% at 2.3 Å (Table 2 and Fig. 4). Both structures were solved using single-wavelength anomalous diffraction phasing of selenomethionine-labeled protein crystals. The free \textit{R}-factor of the latter structure is slightly high, which is probably attributable to modest problems in the diffraction data caused by the presence of a satellite crystal (see "Materials and Methods"). However, the \textit{bm}HL coordinates were used successfully to solve the crystal structure of \textit{hs}HL (17) by molecular replacement, and the refined structure of \textit{hs}HL at 2.1 Å resolution (with a free \textit{R}-factor of 26.5%) agrees in almost all stereochemical details with those of the bacterial homologues except for small conformational differences in the flexible G-loop (see discussion of Fig. 6B, below) and a significant backbone dihedral angle change at just one position elsewhere in the protein (data not shown).

The \textit{bm}HL protomer comprises 287 residues of which residues 2–284 are observed in its crystal structure, while the \textit{hs}HL protomer comprises 299 residues of which residues 1–298 are observed in its crystal structure. The binding site for the essential divalent cation is unoccupied in the \textit{bs}HL structure but occupied by \textit{Ca}^{2+} from the mother liquor in the \textit{bm}HL structure. The active site geometry in the latter is therefore closer to that in the active \textit{Mn}^{2+}- or \textit{Mg}^{2+}-bound form of the enzyme.

\textit{HL} Protomer Structure—The enzyme has a TIM barrel fold formed by 8 \textit{β}-\textit{α} supersecondary structural units. These units form an inner core comprising an \(n = 8, s = 8\)-β-barrel surrounded by an outer layer of \(α\)-helices (Fig. 4A). As in the canonical TIM barrel, the strands forming the β-barrel make exclusively nearest neighbor inter-strand hydrogen-bonding (H-bonding) interactions, with the exception of strand 8 that H-bonds to strand 1 to close the barrel.

Both \textit{bs}HL and \textit{bm}HL have three \(α\)-helices in addition to those in the canonical TIM barrel fold (Fig. 2 and Fig. 4, \textit{A} and \textit{B}). The first of these \((α1)\) is a short helix inserted between the strand and helix forming the first \textit{β}-\textit{α} unit; the second \((α5)\) is a short helix inserted at an equivalent position in the fourth \textit{β}-\textit{α} unit, and the third \((α11)\) is a longer helix located immediately after the eighth \textit{β}-\textit{α} unit at the end of the canonical TIM barrel fold. Helix \(α1\) contains several residues previously implicated in catalysis (Fig. 2 and Table 1) and projects upward from the core of the TIM barrel on one face (Fig. 4A). Helix \(α5\) lies flat on the same face but on the opposite rim of the barrel. An ~12-Å deep cavity lined by all of the residues previously implicated in catalysis is found over the center of the TIM barrel between helices \(α1\) and \(α5\), and this surface cavity presumably forms the active site of the enzyme (Fig. 5). (See also the discussion of Fig. 8A below.) Helix \(α11\) packs between two of the helices forming the canonical TIM barrel fold and runs up the outer wall of the barrel to contact helix \(α1\).

Both \textit{bs}HL and \textit{bm}HL also share a relatively long loop inserted between the strand and helix in the eighth and final \textit{β}-\textit{α} unit. This glycine-rich ‘G-loop’ contains five invariant glycines and is specific to HL. Its sequence is conserved as strongly as any segment of the protein (Fig. 2), and it contains a Cys residue previously implicated in catalysis (Table 1). It lies on the putative active site where it directly contacts helices \(α1\) and \(α11\) (Fig. 4A). The \textit{β}-factors of the G-loop are modestly elevated in \textit{bm}HL and much more significantly elevated in \textit{bs}HL and \textit{hs}HL (17) (Fig. 6A), and it differs slightly in conformation in the three crystal structures (Fig. 6B). These observations indicate that the G-loop is flexible and may change conformation when binding the HMG-CoA substrate.

The core of the central β-barrel is occluded by bulky side chains in the HL crystal structures. Therefore, it is unlikely that the substrate inserts into the core of the barrel as observed in one TIM barrel enzyme and previously proposed for HL (14, 15).
A fourth α-helix beyond those in the canonical TIM barrel fold (α12) is found at the C terminus of both bsHL (Fig. 2 and Fig. 4B) and hsHL (17) where it stabilizes the dimers formed by these enzymes. (See “Oligomeric Interactions of bsHL and Implications for Oxidative Down-regulation of hsHL Activity” and Fig. 8.) Notably, this helix is not present in the bmHL protein that does not form a dimer in solution.

**HL Active Site Architecture**—In the crystal structure of bmHL, the residues previously implicated in catalysis by mutagenesis studies (Table 1) surround the bound Ca$^{2+}$ cation in the surface cavity alluded to above that forms the active site of the enzyme (Fig. 4A, Fig. 5, and Fig. 6B). The protein ligands of the Ca$^{2+}$ are Asp-14, His-205, His-207, and Asn-247, the first two of which are the sites of disease-causing mutations that severely impair enzyme activity (Table 1 and Fig. 2). The Ca$^{2+}$ cation is coordinated in an octahedral geometry, although one ligand atom site is unoccupied and a second is occupied by a crystallographic water molecule (Fig. 5). (At least one of these sites is likely to be occupied by an oxygen atom on the substrate molecule in the Michaelis complex, as discussed below.) Thus, enzymological and structural data suggest that the Ca$^{2+}$ ion in the structure of bmHL occupies the site of the Mn$^{2+}$ or Mg$^{2+}$ ion cofactor that is required for enzyme activity (6), and indeed a Mg$^{2+}$ ion is bound at the equivalent site in the crystal structure of hsHL (17). However, the distances between the Ca$^{2+}$ ion and its ligand atoms range from 2.5 to 2.8 Å compared with typical distances of ~2 Å for a Mn$^{2+}$ or Mg$^{2+}$ cofactor. Therefore, the active site geometry in the crystal structure of bmHL is likely expanded slightly compared with that in the physiological Mn$^{2+}$ complex.

The metal-binding center in the active site participates in an intricate network of H-bonding interactions with residues that are invariant in all HL orthologues (i.e. all of the residues shown in Fig. 5 except Asn-18, which is still highly conserved). An extensive cooperative H-bonding...
Crystal Structures of Bacterial HMG-CoA Lyases

The crystal structure of bacterial HMG-CoA lyase (HL) provides insight into the catalytic mechanism and evolutionary relationships with other enzymes. The active site is delineated by a metal-binding pocket that accommodates the catalytic metal ion, usually calcium or magnesium, which facilitates the chemical reactions. The active site is characterized by a tripartite strategy involving a catalytic base, a general base, and a nucleophile.

**Active site stereochemistry in Ca\(^{2+}\)-bound bHL.** Stereo pair (45, 46) showing the protein backbone as a yellow coil with the G-loop colored orange. Side chains are shown in ball-and-stick representation for a set of invariant residues in all HL orthologues (with the exception of Asn-18 which is changed to Ala in a single orthologue in Fig. 2). The bonds and carbon atoms in these side chains are shown in cyan, and their oxygen, nitrogen, and sulfur atoms are shown in red, blue, and dark yellow, respectively. All of these residues except Asn-18, Arg-137, Tyr-139, Ser-173, and Cys-238 are also conserved in most other DRE-TIM metallolyases (Fig. 2). The bound Ca\(^{2+}\) ion is shown as a magenta sphere, and the green and gray dashed lines represent, respectively, metal-ligand interactions and H-bonds (with distances \(\leq 3.5\) Å). W1 and W2 represent ordered water molecules.

Network can be broken into two halves, the first involving residues Gln-17, Arg-13, and Glu-44 and the second involving residues Arg-137, Tyr-139, Asp-229, and Ser-173. The two sub-networks are connected by a single H-bond between residues Glu-44 and Tyr-139. Mutations in residues Arg-13 and Glu-44 in the first sub-network cause human disease.

Relationship of HL to Previously Reported Protein Structures—A systematic search for related structures using the program DALI (31) shows significant similarity between HL and a wide variety of TIM barrel proteins, which give Z-scores ranging from 29 to 5. The top three of these hits are statistically separated from the remainder (Table 3), giving Z-scores of 29, 24, and 20 compared with a continuum of Z-scores \(\leq 15\) for all the rest. Most strikingly, the three top-scoring DALI hits are all divergent cation-dependent enzymes that cleave carbon-carbon bonds via likely enolate intermediates (as shown in Fig. 1, B–D). A structural superposition of these enzymes with bHL is shown in Fig. 7. These enzymes are the catalytic domains of the following enzymes: DmpG or 4-hydroxy-2-ketovalerate aldolase (PDB code 1NVM (13), Z-score of 29, and C-\(\alpha\) r.m.s.d. of 2.0 Å for alignment of 255 residues with 18% sequence identity); LeuA or 2-isopropylmalylate synthetase (PDB code 1SR9 (32), Z-score of 24, and C-\(\alpha\) r.m.s.d. of 2.7 Å for alignment of 257 residues with 16% sequence identity); and transcarboxylase 5S (PDB code 1QB (33), Z-score of 20, and C-\(\alpha\) r.m.s.d. of 2.7 Å for alignment of 230 residues with 22% sequence identity).

All three of these structurally homologous enzymes possess helices \(\alpha\)1 and \(\alpha\)11, which represent two of the distinguishing features of the HL fold, but lack the G-loop and helix \(\beta\)5. Most importantly, the correspondence of the residues forming the active site of HL (as shown in Fig. 5) is identical in all three of the top-scoring DALI hits (Fig. 7) but not the lower scoring TIM barrel homologues (not shown). Strictly identical residues include Asp-14, His-205, and His-207, which are three of the four metal-ligating residues in bHL, as well as Arg-13 and Glu-44 which complete the DRE triplet with Asp-14. The fourth metal-ligating residue Asn-247 is identical in two of the three structural homologues and replaced by histidine in the third, and residues Gln-17, Thr-177, and Asp-229 are either identical or conservatively substituted in at least two of the structural homologues (Fig. 2). Divalent cations are bound at an equivalent location in the active sites of all of these enzymes: Ca\(^{2+}\) as a surrogate for Mn\(^{2+}\) in bHL, Mn\(^{2+}\) in DmpG, Zn\(^{2+}\) in LeuA, and Ca\(^{2+}\) in transcarboxylase 5S. In all cases, the metal ions are bound by an identical ligand sphere with the exception of the substitution of one asparagine ligand with histidine in transcarboxylase 5S.

Given their overall structural similarity (Fig. 7A) combined with the striking similarity in their active site stereochemistry (Fig. 7B) and the likely similarity in the mechanism of the chemical reactions that they catalyze (Fig. 1), we propose that HL and the catalytic domains of DmpG, LeuA, and transcarboxylase 5S form a family of enzyme domains that catalyze carbon-carbon bond cleavage reactions using a common catalytic reaction mechanism involving an enolate intermediate. Considerations outlined below suggest that this intermediate is stabilized primarily by Arg-13 (in bHL) in the DRE triplet found in the active sites of all of these enzymes (Fig. 7B). On this basis, we propose the name ‘DRE-TIM metallolyase’ for this family of metal ion-dependent TIM barrel enzymes (Table 3).

The only member of this group currently in the SCOP data base is the catalytic domain of DmpG, which forms its own family in the aldolase superfamily under the TIM barrel fold. The other DRE-TIM metallolyases are likely to be classified in the same SCOP structural family.

**Tripartite Strategy to Develop a Model for the Catalytic Reaction Mechanism of HL.**—Having characterized the stereochemistry of the active site in HL, we sought to develop hypotheses concerning the catalytic reaction mechanism of the enzyme and especially to explain the chemical functions of the conserved active site residues in binding the substrate and activating it for hydrolysis. The large body of published mutagenesis results on HL (6–9,11) provides fundamental constraints, but these results do not directly provide insight into the geometry of substrate binding in the Michaelis complex. Therefore, we turned to two other sources of information to develop a model for the stereochemistry of this complex that would be constrained to account for the available mutagenesis data.

One source of information is the conserved active site architecture of the DRE-TIM metallolyases. Because all of these enzymes catalyze carbon-carbon bond cleavage reactions proceeding through an enolate intermediate (Fig. 1), we hypothesized that one of the conserved chemical constituents in the active site would stabilize the negative charge...
accumulation that makes this a high energy structure. On this basis, we concluded that one of the two conserved positively charged groups in the DRE-TIM metallolyase active site (Fig. 7B) should contact the enolate-forming carbonyl group in the Michaelis complex, i.e. that either the bound metal cation or the arginine in the DRE triplet should function as the $X$ group in Fig. 1.

Previously published product complex structures of two of the other DRE-TIM metallolyses strongly support this hypothesis and provide consistent guidance as to which of these two possibilities is correct. In product complexes of both LeuA and transcarboxylase 5S, a carboxylate or carbamide group on one of the products interacts with the central metal cation in the active site (Fig. 1 and Fig. 8B), and neither of these
An additional complication is that enzyme active sites themselves can be conformationally flexible, so that it is common to observe ‘induced fit’ conformational changes in rotamer state and even local backbone conformation upon substrate binding. To deal with this well known conformational flexibility in the substrate/ligand-binding sites of proteins, an induced fit computational docking protocol has been developed recently that efficiently samples the allowed conformations of the protein at the binding site in addition to a multitude of ligand conformations and interaction geometries (25). We have applied this algorithm to generate an energetically ranked list of candidate structures for the Michaelis complex. These candidates were filtered for consistency with the qualitative chemical principles established from the homology analysis described in the previous paragraph and then checked for consistency with the published mutagenesis literature.

This tripartite analysis produces a stereochemical model for the catalytic reaction mechanism of HL that implies specific chemical functionalities for the conserved active site residues in binding HMG-CoA and activating its cleavage. This model can guide new experimental studies designed to definitively elucidate the catalytic reaction mechanism of HL.

**Computational Modeling of the Michaelis Complex with HMG-CoA and Implications for the Catalytic Reaction Mechanism of HL**—A recently developed induced fit docking method was used to explore possible substrate binding geometries in the active site of HL. This procedure combines rigid receptor docking using the program GLIDE (26, 27) with ab initio prediction of protein side chain and backbone conformations using the program PRIME (28, 29), as described in detail under “Methods and Methods.” Calculations were performed using both the Mg2⁺-bound structure of bmHL and the Mg2⁺-bound structure of hsHL (17), but the results obtained using the latter structure were prioritized for interpretation based on the fact that optimal enzyme activity is observed when Mg2⁺ or Mn2⁺ but not Ca2⁺ is bound to the enzyme (6).

Using either HL structure, the vast majority of high scoring substrate poses from either rigid receptor or induced fit docking calculations have the substrate bound in a deep surface groove (Fig. 8B) with the negatively charged carboxylate group of its HMG moiety binding to the divalent cation in the active site. In different poses, either one or both oxygen atoms of the carboxylate occupy the sites in the octahedral coordination sphere of the metal ion that are not occupied by protein side chain atoms. Thus, docking calculations with either bmHL or hsHL consistently recapitulate the interaction between the bound metal ion and a carboxylate group that is observed in the crystal structures of the product complexes of LeuA and transcarboxylase 5S (Fig. 8A).

Among the 10 top scoring poses obtained in induced fit substrate docking calculations on hsHL, two were identified that have the enolate-forming carbonyl on HMG H-bonding to the arginine residue in the DRE triplet in the manner observed in the product complex of transcarboxylase 5S with pyruvate (Fig. 8A). These poses involve similar conformations of the substrate that interact in an equivalent manner with the conserved residues in the active site of HL. Most strikingly, in both poses, the 3-hydroxy group that is deprotonated to initiate carbon-carbon (C–C) bond cleavage is H-bonded to residue Cys-238 in the G-loop in the bmHL residue numbers). This residue is invariant in identity in all HL orthologues (Fig. 2) and is essential for catalysis (Table 1) (7). An ~1 Å movement of the G-loop occurs during induced fit docking to enable formation of this H-bond (Fig. 6B and Fig. 8C).

Fig. 8 shows one of these poses that is our leading candidate for the stereochemistry of the Michaelis complex of HL based on its consistency with all published enzymological data as well as the product structures observed for the other DRE-TIM metallolysases. This model
implies that there are three important chemical interactions that contribute to efficient catalysis, the first two of which are shared with the other DRE-TIM metallolyses. The first of these is the binding of the central divalent cation in the active site to the carbamylated lysine residue, which functions as the catalytic base removing a proton from this oxygen atom to induce it to form a C–O bond coupled to C–C bond cleavage (Fig. 1A). Residues that are not shared by HL have been proposed to function as the catalytic base in the other DRE-TIM metallolyses. Most interestingly, in the case of DmpG and LeuA, these candidates come from other protein subunits that interact with their active sites.

Docking calculations performed on the Ca$^{2+}$-bound bmHL yield a similar proportion of poses in which the enolate-forming carbonyl H-bonds to the arginine in the DRE triplet. However, these poses show a greater diversity of orientations for the 3-hydroxy group that is deprotonated to induce C–C bond cleavage in HMG (data not shown). Most of these poses have the 3-hydroxy group H-bonding to the carbamylated residue Asp-14 in the DRE triplet (using bmHL numbering), where it is also in close proximity to the hydroxy group of Tyr-139, another residue that is invariant in all HL orthologues. In this binding geometry, either Asp-14 or Tyr-139 could function as the catalytic base. However, several considerations indicate that these candidates are less likely to be correct than Cys-238. First, the presence of the Mg$^{2+}$ cation in the hsHL structure compared with Ca$^{2+}$ in the bmHL structure produces an angstrom-scale contraction of the protein atoms surrounding the cation in the active site. Given the fact that Mg$^{2+}$ fully activates the enzyme whereas Ca$^{2+}$ does not, this structure is likely to be more physiological, and it consistently yields induced docking structures in which Cys-238 H-bonds to the 3-hydroxy of HMG when the enolate-forming carbonyl H-bonds to Arg-13 in the DRE triplet. Moreover, function of Cys-238 as the catalytic base provides a simple and straightforward explanation for the fact that substitution of this residue with serine produces a 625-fold reduction in enzyme activity (Table 1) (7), whereas this dramatic reduction in activity is more difficult to rationalize if Asp-14 or Tyr-139 functions as the catalytic base.

Therefore, the preponderance of computational and enzymological data consistently favor the function of Cys-238 as the catalytic base and a stereochemistry for the Michaelis complex similar to that shown in Fig. 8. However, further enzymological or structural studies will be required to evaluate critically the validity of this model for the catalytic reaction mechanism of HL.

Although most induced fit docking models place the nonreactive CoA moiety of the substrate in a similar position and conformation to that shown in Fig. 8B, there is substantial variation in the detailed protein-substrate interactions occurring in the different poses. However, in the absence of experimental or phylogenetic constraints, it is not possible to discriminate among these different stereochemical models for the binding of the CoA moiety.

Oligomeric Interactions of hsHL and Implications for Oxidative Down-regulation of hsHL Activity—There are four protomers in the asymmetric units in both bacterial HL crystal structures. However,
there is at most 813 Å² of solvent-accessible surface area per subunit buried in the interface of any pair of protomers in the crystal structure of the monomeric bmHL protein, a level within the range typically displayed by nonphysiological crystal packing interactions (41–43). In contrast, two of the protomers in the crystal structure of bsHL bury 1300 Å² of solvent-accessible surface area per subunit in their mutual interface (Fig. 9A), a level typical of stable physiological dimers formed by protein domains of this size (41–43). Two of these dimers make a packing interaction burying 1000 Å² of solvent-accessible surface area per subunit in their mutual interface (Fig. 9B), an amount that is at the upper end of that observed for nonphysiological crystal packing interactions. This dimer of dimers could potentially represent the stable tetrameric species constituting ~10% of the population of bsHL in solution (Fig. 3A). A very similar packing interaction between four protomers is observed in the crystal structure of hsHL (17), which also has a tendency to tetramerize in solution (10). The dimer structure is preserved almost exactly between bsHL and hsHL, whereas the inter-dimer packing interaction in the tetramer differs by only ~15°.

The C-terminal 12 residues in bsHL form helix α12, which makes extensive polar and nonpolar interactions with the G-loop, strand β7, and helix α12 across the interface of the likely physiological dimer (Fig. 9A). Thus, helix α12 plays a critical role in stabilizing the dimer of bsHL. This helix is conserved in the sequence of hsHL, which is also predominantly dimeric in solution and forms an equivalent dimer in its crystal structure, but it is notably absent from the sequence of bmHL, which is monomeric in solution (Fig. 2).

Most interestingly, residue Lys-297 from two subunits of bsHL are located in close proximity to each other in the tetramer found in the asymmetric unit of its crystal structure (Fig. 9, B and C). The residue occurring at the equivalent position in hsHL is Cys-323 (Fig. 2), and this residue is known to form a disulfide bond with the same residue in another subunit under oxidizing conditions, a reaction that strongly attenuates enzyme activity (11). The S-γ atoms from this residue in subunits from two different dimers come within 5.5 Å of one another in the tetrameric assembly found in the crystal structure of bsHL (Fig. 9C). Therefore, this assembly provides an approximate model for the disulfide-linked complex formed by hsHL. The G-loop is located close to the interface between the dimers in this assembly and directly contacts one of the subunits in the other dimer pair. An inter-dimer contact of this kind in a disulfide-linked tetramer of hsHL could either constrain the G-loop in a catalytically inefficient conformation or propagate an activity-attenuating allosteric conformational change into the active site of the enzyme. Thus, oxidative cross-linking in the common tetrameric assembly observed in the crystal structures of bsHL and hsHL could play a role in the attenuation of hsHL activity under oxidative stress conditions in human cells (11).

Acknowledgments—We thank Randy Abramowitz and John Schwanof for assistance with synchrotron data collection and Roland Satterwhite for performance of hydrodynamic studies.
