ACS chemical biology

Articles

pubs.acs.org/acschemicalbiology

A Distinct MaoC-like Enoyl-CoA Hydratase Architecture Mediates Cholesterol Catabolism in *Mycobacterium tuberculosis*

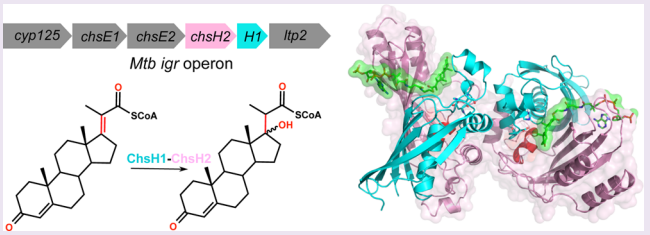
Meng Yang,[†] Kip E. Guja,[‡] Suzanne T. Thomas,[†,§] Miguel Garcia-Diaz,[‡] and Nicole S. Sampson*[,†]

[†]Department of Chemistry, Stony Brook University, Stony Brook, New York 11794-3400, United States

[‡]Department of Pharmacological Sciences, Stony Brook University, Stony Brook, New York 11794-8651, United States

Supporting Information

ABSTRACT: The *Mycobacterium tuberculosis* (*Mtb*) *igr* operon plays an essential role in *Mtb* cholesterol metabolism, which is critical for pathogenesis during the latent stage of *Mtb* infection. Here we report the first structure of a hetero-tetrameric MaoC-like enoyl-CoA hydratase, ChsH1-ChsH2, which is encoded by two adjacent genes from the *igr* operon. We demonstrate that ChsH1-ChsH2 catalyzes the hydration of a steroid enoyl-CoA, 3-oxo-4,17-pregnadiene-20-carboxyl-CoA, in the modified β-oxidation pathway for cholesterol side chain degradation. The ligand-bound and apoenzyme structures of ChsH1-ChsH2^{N} reveal an unusual, modified hot-dog fold with a severely truncated central α-helix that creates an expanded binding site to accommodate the bulkier steroid ring system. The structures show quaternary structure shifts that accommodate the four rings of the steroid substrate and offer an explanation for why the unusual heterotetrameric assembly is utilized for hydration of this steroid. The unique $\alpha\beta$ heterodimer architecture utilized by ChsH1-ChsH2 to bind its distinctive substrate highlights an opportunity for the development of new antimycobacterial drugs that target a pathway specific to *Mtb*.



The resurgence of tuberculosis disease (TB) in the 1990s highlighted that tuberculosis is no longer a disease of the past. HIV/AIDS infection, multidrug resistant TB (MDR-TB), and extensively drug resistant TB (XDR-TB) have made the problem more complex.[1] A quarter of HIV infected deaths are caused by TB, and MDR-TB is present all over the world.[2] The drugs currently used to treat tuberculosis require long treatment times and have limited efficacy and significant toxicities. The first line drugs, including isoniazid, rifampicin, and pyrazinamide, require more than half a year's treatment. Existing drugs only target a small fraction of essential growth pathways, and it is not difficult for *Mycobacterium tuberculosis* (*Mtb*) to overcome their inhibition by activating compensatory pathways. In 2010, 8.8 million people were infected with TB and 1.4 million died from TB.[2] New drugs based on novel pathways or targets that can shorten the therapy and eliminate drug-resistant strains are urgently needed.

In order to develop new strategies against TB, a better understanding of the virulence and pathogenesis of *Mtb*'s infection lifecycle is needed. Macrophages are responsible for defending against invasion by foreign pathogens. However, *Mtb* can reside in macrophages and replicate within them. *Mtb* successfully assimilates nutrition from the macrophage and hides from attack by the immune system. As an intracellular pathogen, the first challenge *Mtb* faces is how to obtain nutrients. *Mtb* preferentially metabolizes host-derived lipids during infection.[3,4] Among a total of 4000 *Mtb* genes identified through genome sequencing, at least 250 genes are potentially involved in lipid metabolism,[5] which is 5 times more lipid-metabolism genes than *Escherichia coli*. Many of the lipid-metabolism genes are induced by intracellular growth and infection.[5] Knowing the functions of genes and their corresponding products involved in lipid-utilization pathways in *Mtb* is critical for understanding *Mtb*'s infection biology and may provide new targets to develop anti-TB drugs.

In the process of defining genes required for survival of *Mtb* in macrophages by transposon library screening, several experiments highlighted the importance of the intracellular growth operon (*igr* operon), comprised of *Rv3540c* to *Rv3545c*, for survival in macrophages.[6,7] Deletion of the operon (H37Rv:Δ*igr*) results in attenuated growth in both macrophages and mice.[6] The genes of the *igr* operon were predicted to encode lipid β-oxidation enzymes.[8] However, the operon is not required for growth on even or odd chain fatty acids.[6] During the course of *Mtb* infection, cholesterol has been demonstrated to serve as a nutrient and is significant for *Mtb* persistence. Many steroid-utilizing genes have been shown to be critical both *in vitro* and *in vivo*.[9] The location of the *igr* operon in the 82-gene cholesterol catabolism cluster suggested a potential function in cholesterol metabolism,[10] and the operon was found to be required for growth of *Mtb in vitro* in the presence of cholesterol.[11] Metabolite profiling of the H37Rv:Δ*igr* strain using isotopically labeled cholesterol resulted in accumulation of a metabolite. The structure of the

Received: March 28, 2014
Accepted: September 9, 2014
Published: September 9, 2014

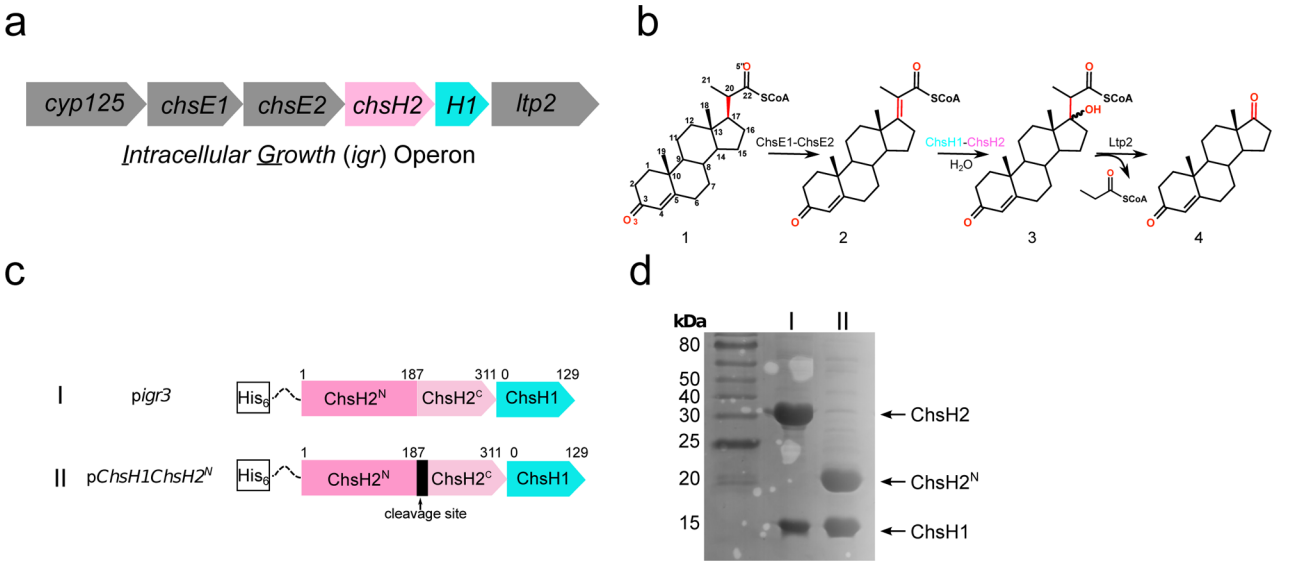


Figure 1. Biochemical function of *igr* operon and gene constructs to express ChsH1-ChsH2, ChsH1-ChsH2^N. (a) The *igr* operon in *Mtb*; *chsH1* (*Rv3541c*) and *chsH2* (*Rv3542c*) are colored in cyan and pink, respectively. Other genes are shown in gray. (b) The biochemical function of the gene products encoded by the *igr* operon. (1) 3-oxo-4-pregnene-20-carboxyl-CoA (3-OPC-CoA); (2) 3-oxo-4,17-pregnadiene-20-carboxyl-CoA (3-OPDC-CoA); (3) 17-hydroxy-3-oxo-4-pregnene-20-carboxyl-CoA (17-HOPC-CoA); (4) androstenedione. (c) Constructs for expression of ChsH1-ChsH2 and ChsH1-ChsH2^N. An N-terminal 6× His tag is encoded at the 5′ end of the first gene of each construct. The *ltp2* gene following *ChsH1* is not shown for clarity. The number of residues encoded by each gene is labeled. The factor Xa cleavage sequence is introduced after the Arg187 codon in the *chsH2* gene. (d) SDS-PAGE analysis of system I isolated by IMAC and of system II isolated by IMAC and then subjected to factor Xa cleavage. Protein identities were confirmed by tryptic digest and MALDI-TOF MS fingerprinting.

metabolite led to the hypothesis that the *igr* operon is required for β-oxidation of C20−C22 in the cholesterol side chain[12] (Figure 1a,b).

The six genes in the *igr* operon are annotated as a lipid transfer protein (Ltp2/Rv3540c), MaoC-like hydratases (Rv3541c, Rv3542c), acyl-CoA dehydrogenases (Rv3543c, Rv3544c), and a cytochrome P450 (cyp125/Rv3545c).[11] Enzymatic and structural studies demonstrated that Cyp125 oxidizes cholest-4-en-3-one or cholesterol at C-26 to cholest-4-en-3-one-26-oic acid.[13] However, in the H37Rv strain employed in the *igr* studies, loss of this activity is compensated by Cyp142.[14] Biophysical characterization of the acyl-CoA dehydrogenases Rv3543c-Rv3544c, named ChsE1-ChsE2, showed that they form an obligate $\alpha_2\beta_2$ heterotetramer with two active sites and two FAD cofactor binding sites.[12,15] ChsE1-ChsE2 catalyzes the C17−C20 dehydrogenation of steroid acyl-CoA thioester 3-oxo-4-pregnene-20-carboxyl-CoA.[15] Earlier pull down experiments demonstrated that Rv3541c and Rv3542c form a protein complex.[12] All of the above information established the importance of the *igr* operon for cholesterol metabolism by *Mtb* and highlighted the possibility that a second heteromeric structural motif may be encoded in the operon.

Based on a protein sequence BLAST search against the nonredundant protein sequence database (blastp), the *igr* operon encodes two potential MaoC-like enoyl-CoA hydratases, Rv3541c and Rv3542c. However, their role in cholesterol side chain degradation was unclear since known fatty acid β-oxidation cycles in bacteria, as well as eukaryotes, utilize a crotonase family member to catalyze hydration of fatty enoyl-CoA metabolites. MaoC-like enoyl-CoA hydratases are structurally distinct from crotonases, and in bacteria, they are employed in polyhydroxyalkanoate energy storage pathways instead of for energy metabolism.[16] In eukaryotes, MaoC-like enoyl-CoA hydratases are employed in peroxisomal β-oxidation of di- and trihydroxy-5β-cholestanic acid to chenodeoxycholic acid and cholic acid, respectively. However, these hydratases are part of a multifunctional enzyme (MFE) that includes a fused 3-hydroxy-acyl-CoA dehydrogenase domain.[17,18] Both the bacterial and peroxisomal MaoC-like enoyl-CoA hydratases characterized to date are all homodimers,[16,19] although the peroxisomal enoyl-CoA domains are larger than their bacterial homologues.

Sequence similarities indicated that both Rv3541c and Rv3542c have MaoC-like hydratase domains, and in addition Rv3542c includes a DUF35/DUF35_N domain. Here, we demonstrate that Rv3541c and Rv3542c form a heterodimer, two of which further dimerize to form an $\alpha_2\beta_2$ heterotetramer. The unique architecture of this heterotetramer is in stark contrast to previously characterized structural homologues, for example, peroxisomal and bacterial enzymes, which form homodimers. We also demonstrate that Rv3541c-Rv3542c catalyzes the hydration of a steroid substrate, 3-oxo-4,17-pregnadiene-20-carboxyl-CoA (3-OPDC-CoA). We now refer to Rv3541c-Rv3542c as ChsH1-ChsH2 to emphasize its function in cholesterol side chain enoyl-CoA hydration. Our data suggest that *Mtb* metabolizes sterol side chains through a unique assembly of β-oxidation enzymes. The structures described here highlight opportunities for the design of inhibitors specifically targeting cholesterol catabolism in actinobacteria, particularly *Mycobacterium tuberculosis*.

RESULTS AND DISCUSSION

Enzymes that degrade the core structure of steroids like cholesterol are not present in humans. On the other hand, *Mtb* has the ability to metabolize completely steroids of mammalian origin and utilize these catabolic intermediates for both energy and other processes related to virulence and pathogenesis. Despite the fact that many of these steroid-utilizing genes have been shown to be critical both *in vitro* and *in vivo*, no enzyme

dx.doi.org/10.1021/cb500232h | *ACS Chem. Biol.* 2014, 9, 2632−2645

crystal structure to date has been elucidated of a steroid utilizing β-oxidation enzyme.[9] Given the unclear structural and functional homologies of ChsH1 and ChsH2, as well as their association, we undertook investigation of the structure and activity of the ChsH1-ChsH2 complex in order to understand its role in *Mtb* cholesterol metabolism.

ChsH1-ChsH2 Forms an Obligate $\alpha_2\beta_2$ Heterotetramer and Preferentially Catalyzes Hydration of Steroid Enoyl-CoA. Initially, ChsH1 and ChsH2 were expressed individually as N-terminally His_6-tagged proteins in *E. coli*. However, both ChsH1 and ChsH2 precipitated directly after elution from an ion affinity chromatography (IMAC) column indicating that ChsH1 and ChsH2 are not stable under these expression conditions.

Construct *pigr3*[12] (Table 1) was expressed heterologously to provide ChsH1-ChsH2 that was isolated by IMAC and further purified using gel filtration chromatography. SDS-PAGE gel analysis confirmed the coisolation of ChsH1 and ChsH2 (Figure 1c,d). Analytical gel filtration chromatography demonstrated that ChsH1-ChsH2 forms a stable oligomeric complex in solution (Figure 2a) with an approximate 1:1 stoichiometry. Further analysis by analytical ultracentrifugation (AUC) equilibrium experiments established that it forms a complex with a molecular weight of 99 kDa (Figure 2b). Thus, given the relative stoichiometry and the individual molecular weights of ChsH1 (14 kDa) and ChsH2 (36 kDa), the ChsH1-ChsH2 complex is an $\alpha_2\beta_2$ heterotetramer. To confirm the stoichiometry of the complex, the assembly was analyzed by LC/UV/MS (Figure 2d). Integration of the ChsH1 and ChsH2 peaks and calculation of their relative molar concentration provided a relative molar stoichiometry of one ChsH1 to one ChsH2, consistent with formation of an $\alpha_2\beta_2$ heterotetramer.

We assayed the ChsH1-ChsH2 enzyme complex with three different substrates: octenoyl-CoA, decenoyl-CoA, and 3-oxo-4,17-pregnadiene-20-carboxyl-CoA (3-OPDC-CoA). The hydration process was monitored spectrophotometrically, and the formation of the hydrated products was confirmed by matrix-assisted laser desorption ionization (MALDI)-time of flight (TOF) mass spectrometry. Although ChsH1-ChsH2 catalyzed the hydration of all three substrates, the highest turnover activity was observed with 3-OPDC-CoA (Figure 2f). Therefore, ChsH1-ChsH2 has the ability to bind bulky substrates and preferentially catalyzes the hydration of steroid enoyl-CoA compared with aliphatic enoyl-CoAs, thus establishing a role in cholesterol metabolism for this enzyme.

Table 1. Expression Constructs Used in This Work

construct name	genes	restriction sites	purified enzyme	source/ref
pET28b				Novagen
pChsH1	*Rv3541c*	*NdeI*/*XhoI*	N-His_6-ChsH1	this work[a]
pChsH2	*Rv3542c*	*NdeI*/*NotI*	N-His_6-ChsH2	this work[a]
pigr3	*Rv3542c*, *Rv3541c*, *Rv3540c*	*NdeI*/*HindII*	N-His_6-ChsH2-ChsH1	12
pChsH1ChsH2^N	*Rv3542c*, *Rv3541c*, *Rv3540c*	*NdeI*/*HindII*	N-His_6-ChsH2-ChsH1	this work[a]
pigr3H1$_{D29A}$	*Rv3542c*, *Rv3541c*, *Rv3540c*	*NdeI*/*HindII*	N-His_6-ChsH2-ChsH1$_{D29A}$	this work[a]
pigr3H1$_{H4A}$	*Rv3542c*, *Rv3541c*, *Rv3540c*	*NdeI*/*HindII*	N-His_6-ChsH2-ChsH1$_{H34A}$	this work[a]
pCasI	*Ro05822*	*NdeI*/*XhoI*	N-His_6-CasI	this work[b]

[a]Genes were cloned from H37Rv genomic DNA and ligated into pET28b with the indicated restriction sites to include an N-terminal His_6 fusion tag. [b]The *Rhodococcus jostii* RHA1 *casI (Ro05822)* gene was synthesized and ligated into pET15b with the indicated restriction sites to include an N-terminal His_6 fusion tag.

A Truncated Form of ChsH1-ChsH2^N Is Sufficient for Catalytic Activity. In the process of crystallizing full length ChsH1-ChsH2, diffraction quality crystals were obtained. The crystals diffracted to 1.52 Å resolution (Table 2), and the structure was solved using single wavelength anomalous dispersion. The crystals contained four $\alpha\beta$ heterodimers (two $\alpha_2\beta_2$ heterotetramers) in the asymmetric unit (Figure 3a, Supplementary Figure 1 and Note, Supporting Information). The electron density was of excellent quality and allowed a model of full length ChsH1 and the N-terminal domain of ChsH2 (180 amino acids out of 311 amino acids) to be built. The absence of further density for the C-terminal domain of ChsH2 prompted us to analyze the protein content of our crystals. Further analysis by SDS-PAGE and MALDI-TOF mass spectrometry confirmed that around a 130 amino acid C-terminal fragment of ChsH2 was lost through proteolysis prior to crystallization. We henceforth refer to this fragment as ChsH2^C.

In order to reliably obtain the protein complex ChsH1-ChsH2^N and to reproduce the crystallization, an expression plasmid, *pChsH1ChsH2^N* (Table 1), was constructed that introduced a factor Xa cleavage site followed by four glycine residues after Arg187 of ChsH2 (Figure 1c). The ChsH1-ChsH2 protein complex was obtained upon cistronic expression using the construct *pChsH1ChsH2^N*. The full-length protein was purified using IMAC and gel filtration chromatography and then cleaved with factor Xa to provide ChsH1-ChsH2^N as confirmed by SDS-PAGE gel and analytical gel filtration (Figure 1d). The molecular weight of ChsH1-ChsH2^N was determined to be 65 kDa by AUC (Figure 2c) consistent with the predicted molecular weights of ChsH1 and His_6 tagged ChsH2^N (14 kDa and 23 kDa, respectively). The relative molar stoichiometry of the subunits remained 1:1 based on LC/UV/MS analysis (Figure 2e). Therefore, ChsH1-ChsH2^N still forms an $\alpha_2\beta_2$ heterotetramer. The C-terminal fragment of ChsH2, referred to as ChsH2^C, precipitated during factor Xa cleavage.

In order to test whether ChsH1-ChsH2^N is still catalytically active, we analyzed its ability to catalyze hydration of the substrates described above. Importantly, ChsH1-ChsH2^N retained catalytic activity, indicating that ChsH2^C is not required for catalysis of hydration (Figure 2f). Furthermore, we were able to obtain crystals of the new protein complex ChsH1-ChsH2^N. These crystals diffracted to 1.54 Å and yielded a structure identical to that originally obtained.

The Structure of ChsH1-ChsH2^N Reveals a Unique MaoC-like Enoyl-CoA Hydratase Architecture. Consistent with annotations based on BLAST alignments, the crystal structure shown here establishes that ChsH1-ChsH2^N belongs to the MaoC-like enoyl-CoA hydratase family. This family contains homodimeric MaoC-like enoyl-CoA hydratases that are comprised of a β-sheet that wraps around a central α-helix, which is known as a hot-dog fold. A second structural motif in MaoC-like hydratases is a three helix segment, referred to as an active site housing segment, and this motif is usually fused with the hot-dog fold domain in the family.[20]

Surprisingly, rather than the prototypical homodimeric structure, ChsH1-ChsH2^N forms a dimer of a unique $\alpha\beta$

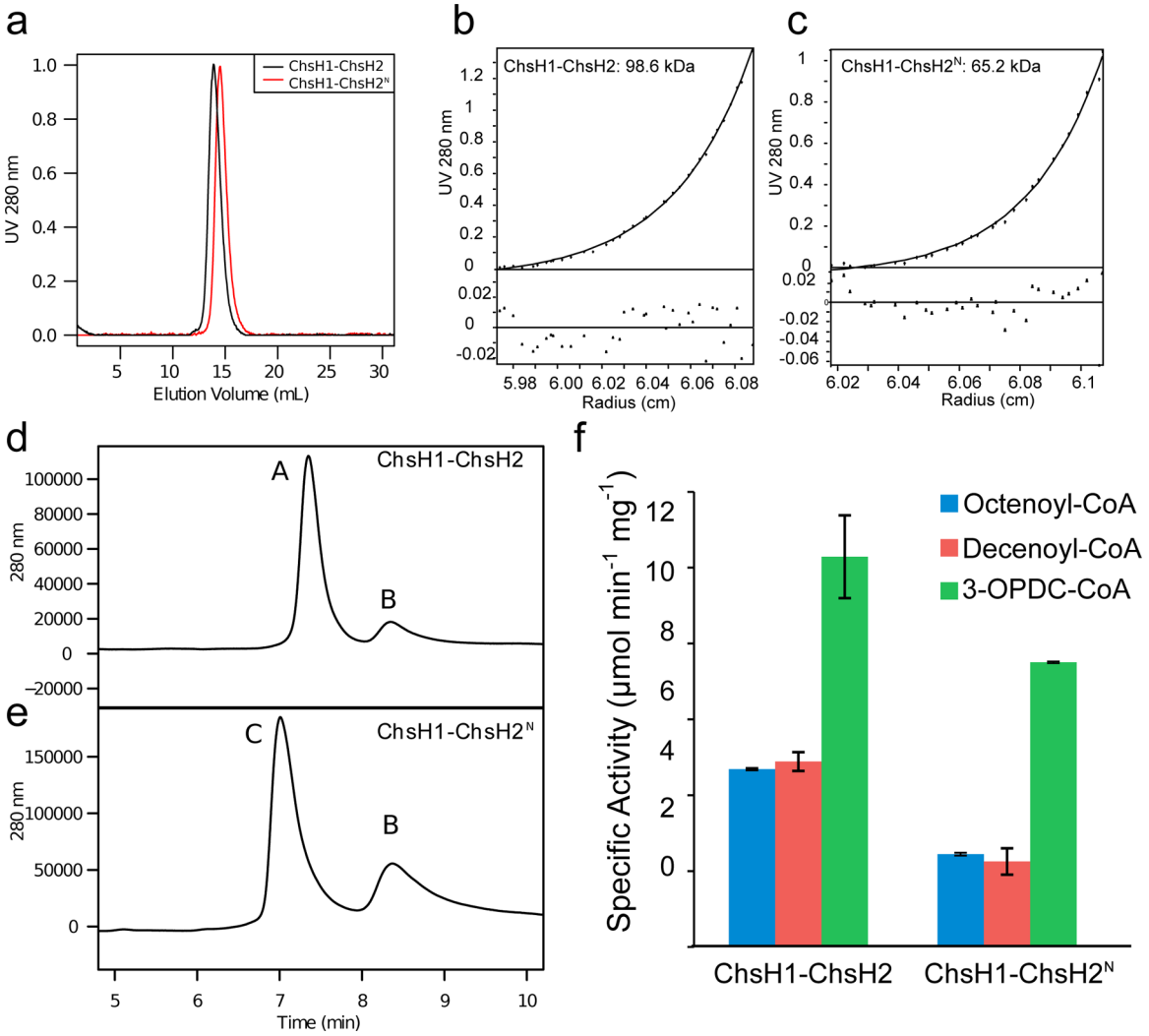


Figure 2. ChsH1-ChsH2 and ChsH1-ChsH2^N are heterotetramers in solution and prefer to catalyze hydration of a steroid enoyl-CoA substrate. (a) ChsH1-ChsH2 and ChsH1-ChsH2^N were analyzed by analytical size exclusion chromatography on a Superdex 75 column. (b) ChsH1-ChsH2 and (c) ChsH1-ChsH2^N were centrifuged at speeds of 10000, 20000, and 30000 rpm at 20 °C in an analytical ultracentrifuge. A representative fit for each sample is shown. The solid line shows the fit of the data to the ideal species model, and the residuals of the fit are graphed below the fit. The best global fit for each protein provided molecular weights of 98.6 and 65.2 kDa for ChsH1-ChsH2 and ChsH1-ChsH2^N, respectively. (d, e) Reverse phase LC/UV/MS chromatograms of ChsH1-ChsH2 and ChsH1-ChsH2^N. Peaks A, B, and C were identified as ChsH2, ChsH1, and ChsH2^N, respectively, by deconvolution of multiple charged states in the corresponding ESI+ MS spectra. The absorbance of peaks were integrated, and relative concentrations were determined from the calculated extinction coefficients of ChsH2, ChsH1, and ChsH2^N [ε_{280}(ChsH2) = 59993 M^{-1} cm^{-1}, ε_{280}(ChsH1) = 16960 M^{-1} cm^{-1}, and ε_{280}(ChsH2^N) = 33523 $M^{-1}cm^{-1}$]. (f) Specific activities of ChsH1-ChsH2 and ChsH1-ChsH2^N. ChsH1-ChsH2 and ChsH1-ChsH2^N were assayed with octenoyl-CoA, decenoyl-CoA, and 3-oxo-4,17-pregnadiene-20-carboxyl-CoA (3-OPDC-CoA).

heterodimeric assembly comprising one standard and one nonstandard MaoC-like hydratase. In the ChsH1-ChsH2^N tetramer, one ChsH1 chain and one ChsH2^N chain form an intimate heterodimer. Two of these heterodimers combine to form the heterotetramer (Figure 3b,c,d). As far as we know, ChsH1-ChsH2 is the first example of an $\alpha_2\beta_2$ heteromeric MaoC-like enoyl-CoA hydratase.

The interaction between the two heterodimers involves helices $\alpha1'$ from ChsH1 and $\alpha2$ from ChsH2^N. The helices from one heterodimer form a compact interface with the other heterodimer through salt bridges and hydrogen bonding interactions (Supplementary Figure 2a,b,c, Supporting Information). Arg37 from ChsH1 in one heterodimer forms a double salt bridge with Asp38 from the second ChsH1. Because of the 2-fold rotational symmetry in the dimer−dimer interface, four salt bridges are formed. Six hydrogen bonds are contributed from helices $\alpha2$ of the ChsH2^N chains (Supplementary Figure 2e,f, Supporting Information).

Importantly, this heterodimer−heterodimer interface is clearly divergent with respect to other MaoC-like hydratase homologues. A DALI[21] search revealed homologues from several species, ranging from bacteria (*Aeromonas caviae*; AcRH) and yeast (*Candida tropicalis*; CtRH) to mammals (*Homo sapiens*; HuRH). Interestingly, the ChsH1-ChsH2^N heterotetramer is comparable to a CtRH homodimer and a HuRH homodimer. The two monomers in CtRH or HuRH dimerize through a similar four-helix bundle (Supplementary Figure 2d, Supporting Information). However, in those cases, the helices from one monomer interact with the helices from the other monomer orthogonally[22] in a very different structural arrangement from ChsH1-ChsH2 (Supplementary Figure 2b,d, Supporting Information).

dx.doi.org/10.1021/cb500232h | *ACS Chem. Biol.* 2014, 9, 2632−2645

Table 2. Data Collection and Refinement Statistics

	ChsH1-ChsH2^N	ChsH1-ChsH2^N/3-OPC-CoA
	Data Collection	
space group	$P2_1$	$P6_122$
cell dimensions		
a, *b*, *c* (Å)	51.8, 134.1, 90.4	51.9, 51.9, 436.2
α, *β*, *γ* (deg)	90, 91.6, 90	90, 90, 120
resolution (Å)	37.6−1.52 (1.50−1.54)[a]	72.70−1.76 (1.768−1.762)[a]
R_{merge}	0.033 (0.434)	0.068 (0.605)
R_{meas}	0.038 (0.585)	0.073 (0.642)
R_{pim}	0.026 (0.408)	0.024 (0.208)
$CC_{1/2}$	0.999 (0.886)	0.999 (0.852)
$I/\sigma I$	19.2 (2.2)	18.8 (8.8)
completeness (%)	99.2 (95.7)	100 (100)
multiplicity	13.4 (11.2)	8.0 (8.8)
Wilson B-factor	17.04	21.19
	Refinement	
resolution (Å)	37.6−1.52	72.70−1.76
no. reflns	180 565	40 369
R_{work}/R_{free}	0.1768/0.2046	0.1895/0.2322
CC*	1.0 (0.976)	1.0 (0.963)
no. atoms		
total	10,103	2,607
protein	8,920	2,286
ligand		76
water	1163	263
B-factors		
protein	24.0	34.8
ligand		41.0
water	33.7	40.6
rms deviations		
bond lengths (Å)	0.009	0.016
bond angles (deg)	1.15	1.57
PDB ID	4W78	4W7B

[a]Values in parentheses are for the highest-resolution shell.

The ChsH1-ChsH2^N Heterodimer Is Structurally Distinct from Its MaoC-like Hydratase Homologues. In the ChsH1-ChsH2^N heterodimer, five *β*-strands (*β*1′−*β*3′−*β*4′−*β*5′−*β*2′) from ChsH1 and four *β*-strands (*β*1−*β*4−*β*3−*β*2) from ChsH2^N form an antiparallel nine-stranded *β*-sheet. Helix *α*3′ from ChsH1 and *α*4 from ChsH2^N are above the antiparallel *β*-sheet in the viewpoint shown in Figure 3d. Helices *α*2−*η*1−*α*3 from ChsH2^N and *α*1′−*η*1′−*α*2′ from ChsH1 form a third layer on top of *α*3′ and *α*4 and are almost 2-fold symmetric (Figure 3d,e). This architecture and the extensive interface between the monomers imply that ChsH1 and ChsH2^N form an obligate heterodimer.

The overall structure of the ChsH1-ChsH2^N heterodimer from *Mtb* is disparate from its homologues across different species. The ChsH1-ChsH2^N heterodimer has two hot-dog folds in total. Interestingly, a similar architecture is present in other structural homologues, but the two hot-dog folds are present as a single monomer (CtRH or HuRH) or a homodimer (AcRH) (Supplementary Table 1, Supporting Information). The fold of ChsH1 is most similar to the C-terminal domain of the CtRH or HuRH monomer and corresponds to the structure of a single chain of AcRH (Supplementary Table 1, Supporting Information). Surprisingly, ChsH2^N, although similar to the N-terminal domains of CtRH and HuRH monomers, deviates from the standard hot-dog fold of AcRH. In addition, both the structure and sequence have low similarity in the active site housing segment (Figure 5c, Supplementary Figure 3, Supporting Information).

The ChsH1-ChsH2^N Heterotetramer Forms a Complex with 3-Oxo-4-pregnene-20-carboxyl CoA. In order to investigate the structural basis of preferential hydration of a steroid enoyl-CoA instead of aliphatic enoyl-CoA's, we cocrystallized ChsH1-ChsH2^N with a substrate analog, 3-oxo-4-pregnene-20-carboxyl CoA (3-OPC-CoA) (Figure 4). The crystals diffracted to a resolution of 1.76 Å (Table 2), and molecular replacement was used to solve the ChsH1-ChsH2^N/3-OPC-CoA structure using the apoprotein structure as the template.

The atomic picture of the ChsH1-ChsH2^N/3-OPC-CoA complex established that one binding site exists per ChsH1-ChsH2^N heterodimer. In each heterodimer, the binding pocket is located at the ChsH1-ChsH2^N interface and intrudes into ChsH2^N (Figure 4a). Binding of 3-OPC-CoA results in a large quaternary structural change compared with the apo-ChsH1-ChsH2^N complex (Figure 4b,c). The major changes occur in the ChsH2^N chain. Movements of *α*1, *α*5, loop I, and loop II (6−7 Å) and the smaller movements of *β*1−*β*4−*β*3−*β*2 generate enough space to accommodate the 3-OPC-CoA ligand in the tunnel (Figure 4b,c). We posit that the heterodimer interface serves to enable the large structural change required for binding the polycyclic steroid system,

dx.doi.org/10.1021/cb500232h | ACS Chem. Biol. 2014, 9, 2632−2645

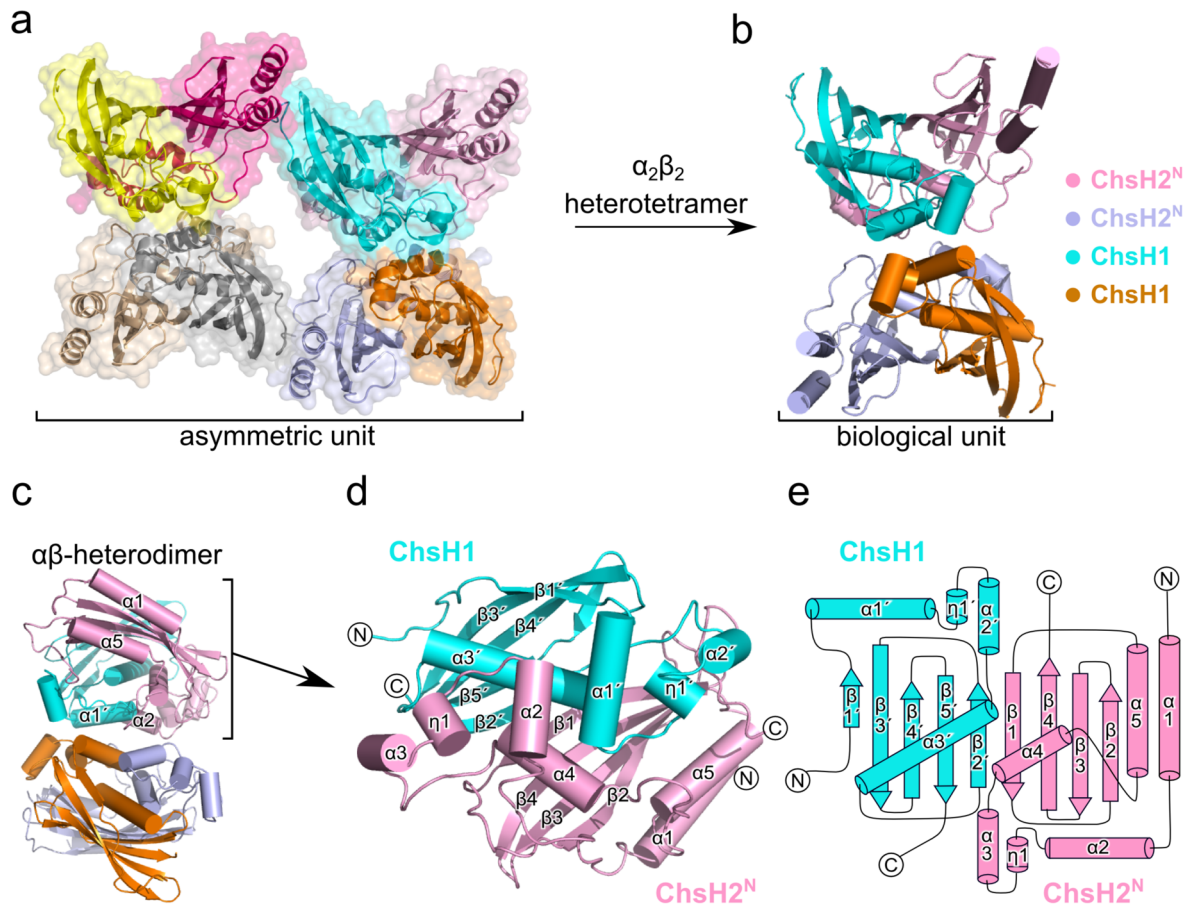


Figure 3. Overall architecture and topology of ChsH1-ChsH2^N. (a) Two $\alpha_2\beta_2$ heterotetramers of ChsH1-ChsH2^N in one asymmetric unit (ASU). Surface representation of the structure corresponds to chain color. (b, c) ChsH1-ChsH2^N adopts a heterotetrameric assembly with two ChsH1 chains and two ChsH2^N chains. One ChsH1 chain and one ChsH2^N chain fold into a heterodimer and a pair of heterodimers assembles into a tetramer. The identities of the chains are labeled. The structure in panel c is obtained by a 90° rotation of the structure in panel b around the *Y*-axis. (d) Schematic representation of one heterodimer from the tetramer. Each chain possesses a conserved "hot-dog" fold. α-Helices and β-strands from ChsH1 and ChsH2^N are shown as ribbons and arrows, respectively; C-termini and N-termini are labeled. 3_{10}-Helices are labeled η. (e) Topology representation of one heterodimer. The color-coding is the same as in panel b. Secondary structure was assigned using DSSP.

which could not occur in a monomeric binding site like that seen in CtRH or HuRH.

The ligand is in a boomerang conformation, which is similar to the orientation of 3-hydroxy-octanoyl-CoA in CtRH.[19] The "boomerang" sits between the ChsH1 chain and the ChsH2^N chain with its elbow, cysteamine, positioned across β1 and β2′ (Figure 4e). The whole boomerang bends toward ChsH2^N with the coenzyme A moiety exposed to solvent and the four-ring steroid system buried inside ChsH2^N (Figure 4a,c; Supplementary Figure 4, Supporting Information).

Binding interactions with ChsH2^N anchor the coenzyme A moiety to the protein. The amine group (N1″) is hydrogen bonded to Asn106 in β1, N2″ is hydrogen bonded to Arg163 in β5, and O3″ is stabilized by Ala137 from loop II. The elbow of the boomerang is stabilized through a hydrogen bond between N4″ nitrogen and Gly81 from β2′ in ChsH1. The steroid part of the "boomerang" has fewer specific interactions with the enzyme compared with the coenzyme A moiety, consistent with the steroid's hydrophobicity. The thioester O5″ is stabilized through a hydrogen bonding interaction with Ala104-ChsH2^N. No polar interactions were found that stabilize the O3−C3 ketone (Figure 4d, 4e).

ChsH1 possesses a canonical MaoC-like hydratase hot-dog fold providing the active site, but ChsH2^N possesses a modified hot-dog fold that confers the ability to bind bulky substrates. ChsH1 includes a complete hot-dog fold and an active site housing segment (Figure 5a). Five antiparallel β strands (β1′-β3′-β4′-β5′-β2′) act as a bun, wrapping around the sausage-like central helix α3′ to form a standard hot-dog fold (Figure 5a), characterized by the four-turns in the central α-helix.[19] The region from α1′, η1′ to α2′ comprises the active site housing segment. The strictly conserved active site residues, Asp29 and His34 (Supplementary Figure 5a), located in the active site housing segment (α1′−η1′−α2′) in ChsH1 (Figure 5a), are almost in identical positions with respect to the active sites in AcRH, CtRH and HuRH (Figure 6b). From the ChsH1-ChsH2^N:3-OPC−COA structure, the C17−C20 bond, which is the site of hydration, sits in apposition to the active site residues Asp29/His34 from ChsH1 (Figure 4d, 4e).

To establish that Asp29 and His34 correspond to the active site of the ChsH1-ChsH2 complex, Asp29 and His34 were mutated to alanine in independent constructs (Table 1). The purified protein complexes were analyzed by SDS-PAGE and by analytical size exclusion chromatography. They both had the same elution profile as wild-type ChsH1-ChsH2. Therefore, ChsH1$_{D29A}$-ChsH2 and ChsH1$_{H34A}$-ChsH2 still form $\alpha_2\beta_2$ heterotetramers. The activities of these two mutant protein

dx.doi.org/10.1021/cb500232h | *ACS Chem. Biol.* 2014, 9, 2632−2645

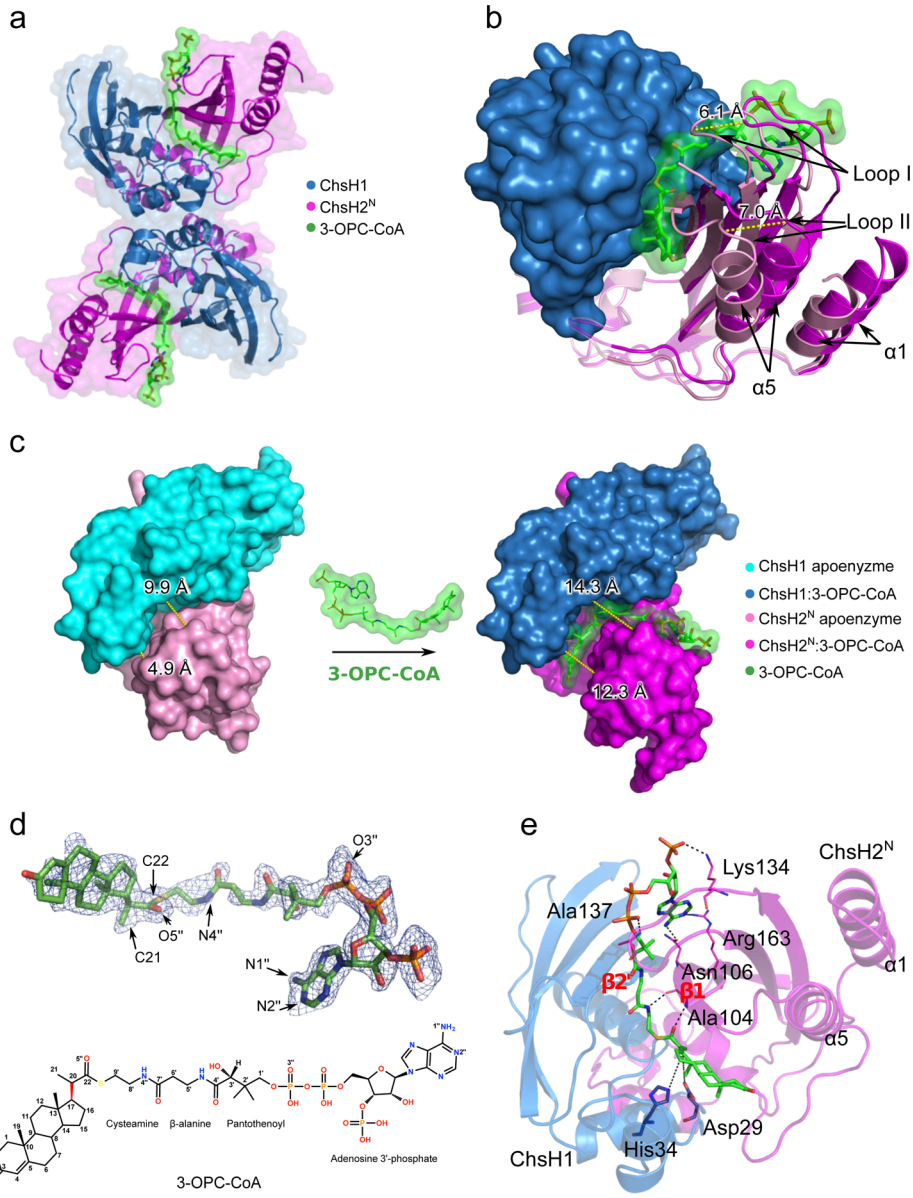


Figure 4. ChsH1-ChsH2^N complexed with 3-oxo-4-pregnene-20-carboxyl-CoA (3-OPC-CoA). (a) Overall atomic picture of ChsH1-ChsH2^N/3-OPC-CoA. The holo-heterotetramer is in the same orientation as Figure 3b. A single heterotetramer is in one asymmetric unit (ASU). ChsH1 chains are colored blue, and ChsH2^N chains are colored magenta. Ligand 3-OPC-CoA is rendered as sticks and colored by atom. Molecular surfaces are rendered transparent. (b) Superimposed heterodimer apo-ChsH1-ChsH2^N structure and ChsH1-ChsH2^N/3-OPC-CoA structure. Superimposed ChsH1 chains are colored in blue, and their surface is shown. The ChsH2^N chains from Apo-ChsH1-ChsH2^N and ChsH1-ChsH2^N/3-OPC-CoA are shown in pink and magenta, respectively. The variations occur in the segments of the ChsH2^N chains that are labeled. (c) The difference in the opening of the binding pocket before and after 3-OPC-CoA binding. (d) A simulated annealing $F_o - F_c$ omit map of 3-OPC-CoA contoured at 2.5σ was calculated to reduce the effects of model bias. A chemical structure of 3-OPC-CoA is shown and colored by atom. (e) Ligand-binding interactions in the ChsH1-ChsH2^N heterodimer. The hydrogen bonds are shown as black dashes. Residues that interact with 3-OPC-CoA are labeled. 3-OPC-CoA fits between β1 from ChsH2^N and β2′from ChsH1.

complexes were tested with octenoyl-CoA, decenoyl-CoA, and 3-OPDC-CoA. ChsH1$_{D29A}$-ChsH2 retained 2%−4% of the wild-type activity. Moreover mutation of His34 to alanine abolished the catalytic activity within detection limits (less than 0.0062% activity) (Supplementary Figure 5b, Supporting Information).

In striking contrast to ChsH1, the structure of ChsH2^N varies substantially with respect to the standard MaoC-like hydratase structure, for example, ChsH1, AcRH, or the C-terminal domains of CtRH or HuRH (Figures 5b,c,d and 6). Only four antiparallel β-strands wrap around the central helix α4. Furthermore, helix α4 contains two turns, and one turn is a 3_{10} helix (Figure 5b,d). The short central helix (α4) renders the hot-dog fold incomplete (Figure 5c). This short, modified helix results in an open binding pocket formed by the four-stranded β-sheet. Helix α1 and α5 are two new flexible elements (Figures 3d and 5c) that are not present in other known MaoC-like hydratase structures. Even though the active site housing segment (α2− η1−α3) is structurally conserved, the active site histidine is replaced with a tyrosine (Figure 5b,d).

The distinctively short central α-helix (α4) is a source of another major difference (Figures 5c and 6c) and has the most

dx.doi.org/10.1021/cb500232h | ACS Chem. Biol. 2014, 9, 2632−2645

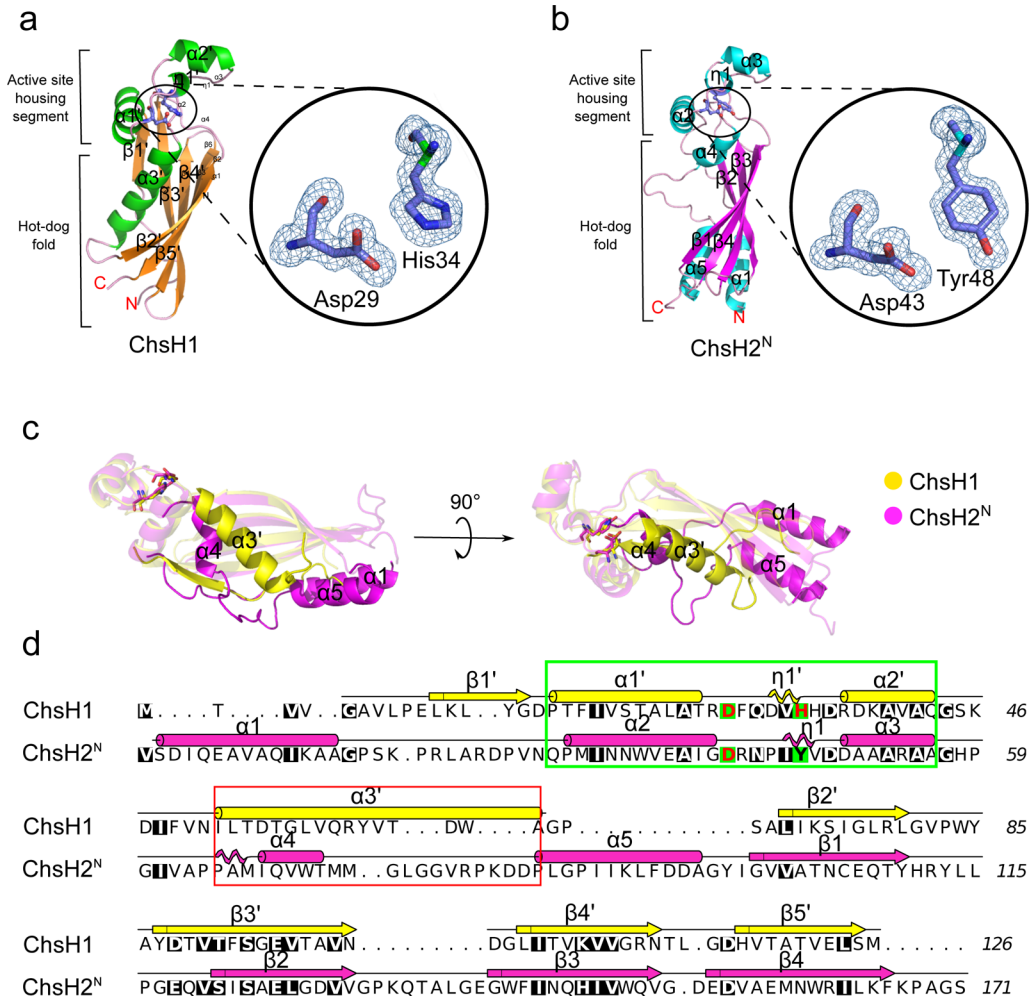


Figure 5. Structure comparison and secondary structure sequence alignment between ChsH1 and ChsH2^N. (a, b) ChsH1 possesses a standard hot-dog fold, and ChsH2^N adopts a nonstandard hot-dog fold. The cartoon representations are colored by secondary structure. The secondary structures are labeled as in Figure 3d,e. N-termini and C-termini are labeled. In ChsH1, the five-stranded β-sheet ($\beta1'$-$\beta3'$-$\beta4'$-$\beta5'$-$\beta2'$) wraps around $\alpha3'$, forming a hot-dog fold, and the proposed active site housing segment is comprised of $\alpha1'$, $\eta1'$, and $\alpha2'$. In ChsH2^N, a four-stranded β-sheet ($\beta1$-$\beta4$-$\beta3$-$\beta2$) wraps around $\alpha4$, assembling into a nonstandard hot-dog fold; $\alpha2$, $\eta1$, and $\alpha3$ fold to form a potential active site housing segment. The potential active site residues are colored blue, and the simulated annealing omit maps of those residues are calculated to reduce the effects of model bias. Electron density is shown as mesh (σ = 3). (c) Cartoon representation of superimposed ChsH1 (yellow) and ChsH2^N (magenta). Structurally similar parts are rendered transparent, and the variations are marked. The structure on the right is obtained by rotating the structure on the left 90° about the *X*-axis. Potential active site residues are shown as sticks. (d) Secondary structural sequence alignment between ChsH1 and ChsH2^N. Identical residues are highlighted in black. The secondary structure elements from ChsH1 (yellow) and from ChsH2^N (magenta) correspond to panels a and b. Residues that constitute active site housing segments are highlighted in the green box. The central α-helices from ChsH1 and ChsH2^N are highlighted in the red box.

important functional implication. The short helix $\alpha4$ and the following long loop generate a bigger space in the hot-dog fold compared with other MaoC-like hydratases (Figure 6c), and the presence of this space is critical to allow the ChsH1-ChsH2^N heterodimer to bind the steroid substrate. Consistently, comparison of this structural feature in the structures of ChsH1-ChsH2^N and its homologues (Figure 6c) reveals that as the central α-helix of the hot-dog fold becomes shorter, the binding cavity becomes wider and sufficiently spacious to accommodate increasingly larger and bulkier substrates. It appears that upon shifting from linear alkyl chain and branched alkyl chain substrates to the rigid rings of a steroid system, the protein architecture transitions from small monomer to extended monomer, and finally to a heterodimer (Figure 6c).

For example, in the structure of N-CtRH complexed with its physiologic product, (3*R*)-hydroxy-octanoyl-CoA, the ligand is accommodated in the interface between the nonstandard N-terminal hot-dog fold and the complete C-terminal hot-dog fold.[19] The pantothenate and adenosine triphosphate of the ligand are exposed to solvent, and the ten-carbon acyl chain points toward the shorter central α-helix in the N-terminal nonstandard hot-dog fold.[19] Interestingly, the ChsH1-ChsH2^N dimer has a similar, yet distinct, substrate binding mode; ChsH1 provides the active site and ChsH2^N provides the binding pocket (Figure 4e). The available space in the suggested binding tunnel of the hot-dog fold is restricted by the rigid central helix, so the length of the helix influences substrate preferences (Figure 6c; Supplementary Table 1, Supporting Information). The existence of a nonstandard hot-dog fold with shorter central helices in the N-terminal domains of CtRH, HuRH, and ChsH2^N chain suggests that one active site has evolved into a binding site to accommodate bulky

dx.doi.org/10.1021/cb500232h | *ACS Chem. Biol.* 2014, 9, 2632−2645

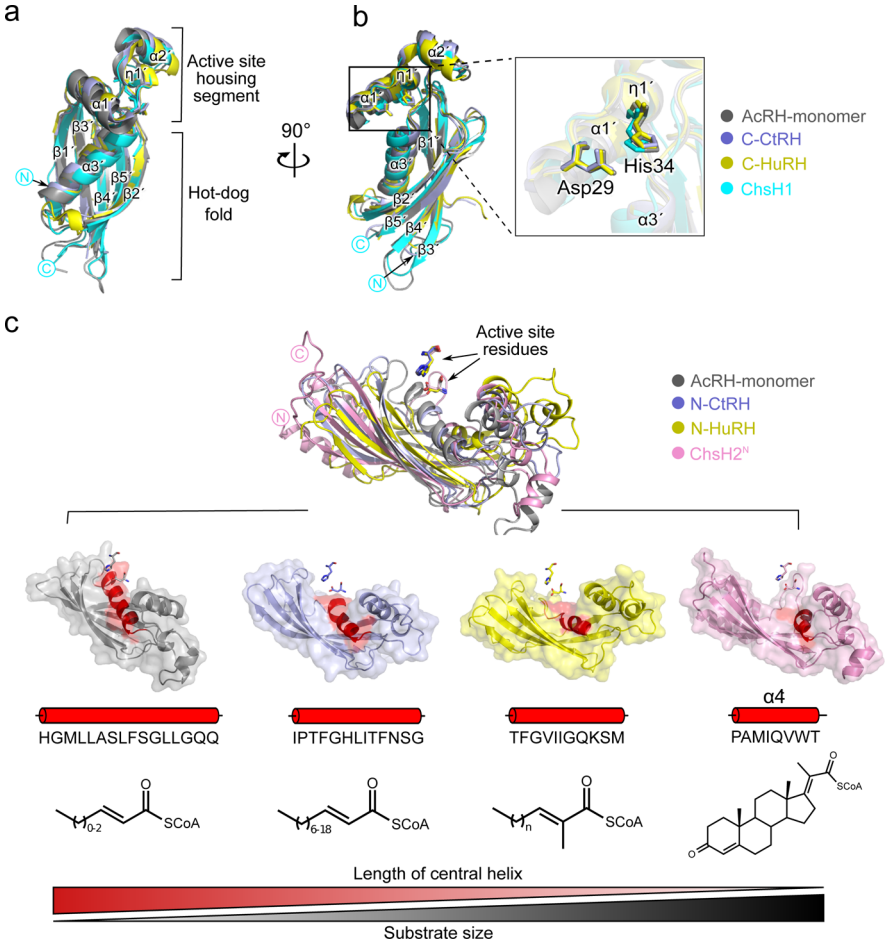


Figure 6. Comparisons of MaoC-like enoyl-CoA hydratases across species. (a, b) Three-dimensional structural conservation between AcRH monomer (gray), C-terminal domain of CtRH (C-CtRH) (residue numbers 163−280; purple), C-terminal domain of HuRH (C-HuRH) (residue numbers 174−298; yellow), and ChsH1 chain (cyan). The structural elements are labeled following Figures 3d and 5a. The conserved active sites are enlarged in the box. (c) Three-dimensional structural comparison between AcRH monomer (gray), N-terminal domain of CtRH (N-CtRH) (residue numbers 1−135; purple), N-terminal domain of HuRH (NHuRH) (residue numbers 1−146; yellow), and $ChsH2^N$ chain (pink). Active site residues from the other chain of AcRH, C-terminal domains of CtRH and HuRH, or ChsH1 are shown as sticks. The central α-helices are colored red. Molecular surfaces are rendered transparent. The length of the central α-helices correspond to the 3-D structures above them. For simplicity, 3_{10} helices and α-helices are both shown as cylinders. Protein sequences shown underneath are the amino acids comprising the central helices. The sizes of the central helices dictate the substrate preferences. The preferred substrates correspond to the structure above them. AcRH (PDB code 1iq6); CtRH (PDB code 1pn2); HuRH (PDB code 1s9c).

substrates concomitant with the sacrifice of one set of active site catalytic residues (Figure 6c; Supplementary Table 1, Supporting Information).

The standard central helix of 16 amino acids in the complete hot-dog fold confers specificity for four to six carbon enoyl-CoAs.[23] In CtRH, the bent central helix of 13 amino acids in the N-terminal hot-dog fold enables accommodation of long-chain enoyl-CoAs (C_{10}−C_{22}) in the binding site.[19] The shorter central α-helix of 11 amino acids and bigger pocket in the N-terminal hot-dog fold of HuRH extends the accepted enoyl-CoA ester chain length up to C_{26} and methyl-branched enoyl-CoA's. The central helix α4 of eight amino acids and adjacent flexible loop in $ChsH2^N$ generates an even more flexible active site with a larger pocket. The enlarged pocket in ChsH1-$ChsH2^N$ makes it possible to accommodate steroid CoA thioesters as substrates, consistent with its function in cholesterol degradation in *Mtb*[12] (Figure 6c; Supplementary Table 1, Supporting Information).

Phylogenetic Relationships of ChsH1-ChsH2. Our characterization of the unusual ChsH1-ChsH2 architecture prompted us to assess whether the heteromeric assembly might also exist in other organisms, and if so, whether those organisms were known to metabolize cholesterol. The presence of a hot-dog motif is insufficient to identify MaoC-like hydratase candidates because it is present in other enzyme families.[20] However, sequences in the active site housing segment are highly conserved and, in combination with the hot-dog fold, clearly define MaoC-like hydratases (Supplementary Figure 5a, Supporting Information). By comparing the MaoC-like hydratase sequences from human, fungi, and bacteria, we identified a generalized motif [VIL]-[AVI]-[SA]-X-[AY]-[ILRA]-[AL]-[ST]-[RGEN]-D-[FYWR]-[NQEF]-[PDN]-[VLG]-H-[ILH]-[PRK]-[ANDE]-X-A, which is located in the active site housing segments (Figure 6a; Supplementary Figure 5a, Supporting Information).

We used this motif to identify additional family members through individual BLAST queries with the protein sequences of ChsH1 and ChsH2 against the nonredundant protein sequence database. Then we examined their genomic environment to assess the prevalence of the heterotetrameric MaoC-

dx.doi.org/10.1021/cb500232h | *ACS Chem. Biol.* 2014, 9, 2632−2645

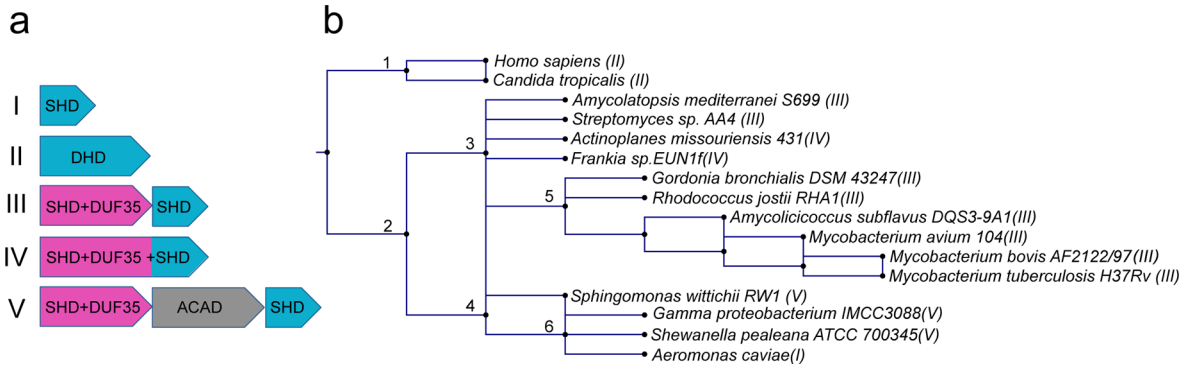


Figure 7. Phylogenetic relationships of ChsH1-ChsH2. (a) Homologues of ChsH1-ChsH2 were identified in five different genomic contexts by BLAST. The arrow lengths are scaled to gene lengths. SHD (cyan), single hot-dog fold; DHD (cyan), double hot-dog; SHD+DUF35 (magenta), single hot-dog fold fused with DUF35/DUF35_N domain; ACAD (gray), acyl-CoA dehydrogenase. (b) Phylogenetic tree for five different genomic contexts of *chsH1-chsH2* gene homologues. (1) Domain Opisthoknot; (2) Domain Bacteria; (3) Phylum Actinobacteria; (4) Phylum Proteobacteria; (5) Suborder Corynebacterineae; (6) Class Gammaproteobacteria. Representative organisms were selected for each category. The Roman numerals in parentheses after each organism correspond to the gene organization in panel a.

like hydratase motif.[24] We found five different genomic contexts for ChsH1 and ChsH2 homologues (Figure 7a). Importantly, inclusion of genomic proximity in the search for homologues allowed us to identify three different motifs for encoding structural homologues of ChsH1 and ChsH2. All of these homologues are found in bacterial families known to catabolize steroids.[25]

The two most common classes of homologues are single hot-dog fold and double hot-dog fold MaoC-like hydratases encoded without any neighboring MaoC-like enoyl-CoA hydratase genes. Typically, they form homodimers like AcRH or CtRH as described above. These are evolutionarily most distant from *chsH1* and *chsH2* and represent the previously identified MaoC-like hydratases. The third category contains two homologous genes in the same adjacent context as *chsH1*, a single hot-dog fold, and *chsH2*, a single hot-dog fold fused to a DUF35/DUF35_N domain, which is hypothesized to bind and deliver acyl-CoA moieties in acyl-CoA-utilization processes[26,27] (Supplementary Note). This arrangement is predominantly in Corynebacteria, which includes Mycobacteria (Figure 7b).

In less closely related Actinobacteria, the *chsH1-chsH2* motif exists as a single fused gene. The sequence alignments suggest that these proteins will form a homodimer that is structurally similar to the ChsH1-ChsH2 heterotetramer, that is, a Rosetta fusion protein.[28,29] Whether these homologues catalyze the hydration of a steroid enoyl-CoA or a structurally simpler steroid metabolite remains to be determined. The fifth category is found in Proteobacteria and is comprised of operons that encode a *chsH2* homologue separated from a *chsH1* homologue by an acyl-CoA dehydrogenase (ACAD/FadE) homologue. Interestingly, in *Agrobacterium tumefaciens* strain C58, these three proteins are encoded as a single Rosetta fusion protein suggesting that the enoyl-CoA hydratase uses the acyl-CoA dehydrogenase product as a substrate, as is the case for the *igr*-encoded enzymes (Figure 1a,b). However, the organization of the *igr* operon is different; the acyl-CoA dehydrogenase genes (*chsE1* and *chsE2*) are adjacent to *chsH2* (Figure 1a). In addition, we have found no evidence for a tight association between the ChsH1-ChsH2 enzyme and the ChsE1-ChsE2 enzyme in our work.[15]

Within *Mtb*, 14 proteins are annotated as possessing hot-dogs folds and belong to the MaoC-like-hydratase family.[30−32] Analysis of their sequences suggests that they either assemble as single hot-dog folds or double hot-dog folds and form homodimers like AcRH and CtRH, respectively. We found no other *Mtb* homologues like ChsH1 and ChsH2, which would be predicted to form a dimer of heterodimers. This unique utilization of a heterotetrameric MaoC-like hydratase is in contrast to the presence of six cholesterol-regulated heterotetrameric acyl-CoA dehydrogenases in *Mtb*.[25] We conclude that in *Mtb*, this heterotetrameric enoyl-CoA hydratase is only employed in the final step of cholesterol side chain β-oxidation.

Conclusion. The enzymes encoded by the *igr* operon diverge structurally from the β-oxidation enzymes typically found in bacteria or eukaryotes. Oxidation of 3-OPC-CoA is catalyzed by a heterotetrameric acyl-CoA dehydrogenase (ChsE1-ChsE2) in place of the typical homotetrameric or homodimeric ACAD assembly. Hydration of the product 3-OPDC-CoA is catalyzed by an $\alpha_2\beta_2$ heterotetrameric MaoC-like enoyl-CoA hydratase instead of the usual mitochondrial or bacterial crotonase family member or homodimeric peroxisomal multifunctional enzyme-2 (MFE-2) family member. The heterotetrameric architectures of these two *Mtb* enzymes appear to have evolved in order to accommodate the large steroid-CoA substrates. The structures of ChsH1-ChsH2 elucidated in this work provide insight into binding site cavities that differ substantially from its homologues and will allow discrimination between host and pathogen enzymes by inhibitors in the future.

METHODS

Materials, Strains, Media, and General Methods. Ferricenium hexafluorophosphate was purchased from Sigma-Aldrich. Coenzyme A was purchased from MP Biomedicals. Isopropyl β-D-1-thiogalactopyranoside was from Denville Scientific. Tryptone, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), and tris(hydroxymethyl)-aminomethane (TRIS) were purchased from Fisher Scientific. Kanamycin is from IBI Scientific. Yeast extract was purchased from Research Products International Co. iProof DNA polymerase was from Bio-Rad. Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, factor Xa, and protein ladder were from New England Biolabs. HisTrap FF columns and Superdex 200 HiLoad 16/60 and 10/300 GL columns were from GE Healthcare Biosciences Corp. Oligonucleotides were from IDT Inc. Total genomic DNA of *M. tuberculosis* H37Rv was obtained from the TB Research Materials Facility at Colorado State University (NIAD NO1-AI40091). MALDI mass spectra were acquired on a Bruker Autoflex II TOF/TOF. Big

dx.doi.org/10.1021/cb500232h | *ACS Chem. Biol.* 2014, 9, 2632−2645

Dye DNA sequencing (Applied Biosystems; performed by the Stony Brook University Sequencing Facility) was used to verify the coding sequence of the expression plasmids. BL21(DE3) *E. coli* was obtained from BioRad. The 2× YT media is composed of 16 g of tryptone, 10 g of yeast extract, and 5 g of NaCl per liter. Buffer A consists of 20 mM Tris-HCl buffer, pH 8.0, supplemented with 300 mM NaCl and 10 mM imidazole. Buffer B consists of 20 mM Tris-HCl buffer, pH 8.0, supplemented with 300 mM NaCl and 500 mM imidazole. Buffer C consists of 50 mM Tris-HCl buffer, pH 8.0, supplemented with 200 mM NaCl. The *casI* (*Ro05822*) gene sequence was synthesized by GenScript USA Inc. Primary crystallization screens were from Hampton research.

Expression Plasmid Construction. *ChsH1* (*Rv3541c*) and *ChsH2* (*Rv3542c*) were amplified from *M. tuberculosis* H37Rv total genomic DNA by PCR using forward and reverse primers. The PCR product was digested with the appropriate restriction endonuclease and ligated into similarly digested pET28b (Table 1). DNA sequencing of the plasmids confirmed that the sequence was correct and that no mutations were introduced during PCR. Active site mutants were prepared using Quick Change site-directed mutagenesis.[33] The mutations were confirmed by DNA sequencing.

To produce the C-terminal truncation of ChsH2, a factor Xa cleavage site, followed by four glycine residues, was introduced at Arg187 of ChsH2. Insertion was conducted by PCR with primers 5′-ATC GAC GGG CGT GGT GGT GGT CGT CCC TCG TCG TCG CGG GAC-3′ and 5′-CAT CAT AGC GTC GGG ATC CAA ATC GTC AGG-3′ to give construct *pChsH1ChsH2*^{N}.

Protein Expression and Purification. To express ChsH1 or ChsH2 individually, construct *pChsH1* or *pChsH2* was transformed into BL21(DE3) *E. coli*. Single colonies were selected on LB plates containing 30 μg/mL kanamycin and cultured in 2× YT media at 37 °C. Expression was induced at $OD_{600} \approx 0.6-0.8$ by the addition of 50 μM to 1 mM IPTG, and cells were grown 20 h at 16−25 °C. Purified proteins were analyzed by reducing SDS-PAGE. CasI was expressed as described for ChsH1 or ChsH2 using construct *pCasI* and 100 μg/mL ampicillin in the media.

ChsH1-ChsH2 Complex Was Obtained Using Construct *pigr3*.[12] Expression was induced at $OD_{600} \approx 0.6-0.8$ by the addition of 1 mM IPTG, and cells were grown 20 h at 25 °C. Cells were lysed by French press or sonication in buffer A, and cellular debris was removed by centrifugation at 125 000*g* for 1 h. ChsH1-ChsH2 was purified by IMAC, with HisTrap FF column and buffers A and B. Protein was further purified by size exclusion chromatography on a Superdex 200 HiLoad 16/60 column equilibrated with buffer C. Active site mutants were expressed and purified following the method for *pigr3* using constructs $pigr3H1_{D29A}$ and $pigr3H1_{H34A}$.

ChsH1-ChsH2^{N} was expressed as described for *pigr3* and isolated by IMAC. ChsH2 was cleaved by factorXa at 23 °C in buffer C supplemented with 2 mM $CaCl_2$. Factor Xa protease reaction was monitored by SDS-PAGE, and cleaved ChsH1-ChsH2^{N} was purified by size exclusion chromatography on a Superdex 200 HiLoad 16/60 column equilibrated with buffer C.

Solution-State Biophysical Analysis of ChsH1-ChsH2 and ChsH1-ChsH2^{N}. ChsH1-ChsH2 (10 mg mL^{-1}) and ChsH1-ChsH2^{N} (7 mg mL^{-1}) were analyzed by analytical gel-filtration on a Superdex 200 (10/300 GL) column (GE Healthcare). The column was equilibrated in buffer C. Samples were eluted isocratically in buffer C, monitoring at 220 and 280 nm. Several standard proteins were analyzed under the same conditions to generate standard curves to estimate molecular weights of analyzed proteins.

Molecular weights were determined using analytical ultracentrifugation sedimentation equilibrium (Beckman Optima XL-A). ChsH1-ChsH2 (5.7, 2.8, and 1.4 μM) and ChsH1-ChsH2^{N} (7.3, 3.7, and 1.8 μM) were centrifuged at speeds of 10000, 20000, and 30000 rpm at 20 °C. Scans were acquired after 18 and 20 h of centrifugation at each speed monitoring at 280 nm. The protein partial-specific volume of 0.7359 for ChsH1-ChsH2 and 0.7379 for ChsH1-ChsH2^{N} and a solvent density 1.0079 for buffer C were calculated using SEDNTERP. Data were fit globally to the ideal, single species model using Heteroanalysis to determine the molecular weight.

Protein complex stoichiometries of ChsH1-ChsH2 and ChsH1-ChsH2^{N} were confirmed by LC/UV/MS. Samples were separated on a XBridge BEH 300 C4 3.5 μm column (2.1 mm × 100 mm) at 40 °C with a linear gradient from 95% A to 95% B over 15 min, where A is 5% isopropanol/0.1% trifluoroacetic acid and B is 99.9% isopropanol/0.1% trifluoroacetic acid. MS spectra were collected in ESI positive ion mode with a cone voltage of 40 V, a capillary voltage of 4.5 kV, and source temperature of 150 °C. MS spectra were deconvoluted using ESIprot 1.0,[34] and peaks in the UV 280 nm chromatograms were integrated using R. The integrated peak areas of each protein were divided by the corresponding molar extinction coefficient for the protein to yield the molar concentrations. Protein stoichiometries were determined from the ratio of the molar concentrations.

Synthesis of Enoyl-CoAs. *trans*-2-Decenoyl-CoA and *trans*-2-octenoyl-CoA were synthesized using the mixed anhydride method.[35] Briefly, the mixed anhydride was prepared by mixing the corresponding acid (0.2 mmol) with ethyl chloroformate (0.4 mmol) in 4 mL of dry THF in the presence of TEA (56 μL). After approximately 25 min, the acid was completely converted into the corresponding anhydride according to TLC. The mixed anhydride was filtered though glass wool in a disposable Pasteur pipet into newly prepared lithium CoA solution (55 mg) dissolved in 5 mL of H_2O and THF (3:2 v/v) (pH, 8). The reaction was stirred at rt for 2 days. The pH of the reaction mixture was adjusted to 3, and the unreacted acid was removed by extraction with ether. The enoyl-CoAs were purified by HPLC using 20 mM ammonium bicarbonate, pH 8, with a linear gradient from 0 to 90% methanol in 20 mM ammonium bicarbonate.

3-Oxo-4-pregnadiene-20-carboxylic acid. 3-Oxo-4-pregnadiene-20-carboxylic acid was synthesized following a previously reported method starting from stigmasta-4,22-diene-3-one.[12,36]

3-Oxo-4-pregnene-20-carboxyl CoA, 3-OPC-CoA. 3-Oxo-4-pregnene-20-carboxyl CoA was synthesized following the previously reported CoA synthetase assay. 3-Oxo-4-pregnene-20-carboxylic acid (3-OPC) was dissolved to 50 mM in 94% ethanol containing 60 mM NaOH. The thioesterification was performed for 4 h at 22 °C in a 1 mL reaction volume containing 5 mM 3-OPC, 10 μM CasI (Ro05822), 100 mM HEPES (pH 7.4), 5 mM $MgCl_2$, 2.5 mM ATP, and 1 mM CoASH.[37,38] The reaction was quenched by adding 1 mL of MeOH to precipitate the enzyme. The supernatant was isolated by centrifugation and filtration. 3-Oxo-4-pregnadiene-20-carboxyl CoA was purified by HPLC using 100 mM ammonium acetate, pH 4.5, with a linear gradient from 0 to 90% MeOH in 100 mM ammonium acetate to yield 3-OPC-CoA.

3-Oxo-pregna-4,17-diene-20-carboxyl-CoA, 3-OPDC-CoA. 3-Oxo-4-pregnene-20-carboxyl CoA was redissolved to 2.8 mM in H_2O. The oxidation was performed for 3 h at 20 °C in 1 mL of 0.5 mM 3-OPC-CoA, 3 μM ChsE1-ChsE2 (FadE28-FadE29), 100 mM *N*-tris[hydroxymethyl]methyl-3-aminopropanesulfonic acid (TAPS) (pH 8.5), and 1 mM ferricenium hexafluorophosphate.[15] The product was purified by HPLC in 100 mM ammonium acetate, pH 4.5, with a linear gradient of 90% MeOH. MeOH was removed under reduced pressure, and H_2O was removed by lyophilization to yield 3-OPDC-CoA.

Hydratase Assay. ChsH1-ChsH2, ChsH1-ChsH2^{N}, $ChsH1_{D29A}$-ChsH2, and $ChsH1_{H34A}$-ChsH2 were assayed for hydratase activity with substrates octenoyl-CoA, decenoyl-CoA, and 3-oxo-pregna-4,17-diene-20-carboxyl CoA in 100 mM HEPES, pH 7.4, buffer and 100 nM (50 nM for 3-oxo-pregna-4,17-diene-20-carboxyl CoA) enzyme at 25 °C. Reactions were monitored at 263 nm, and initial velocities were determined. Product formation was quantified using ϵ_{263nm} = 6700 M^{-1} cm^{-1}, which corresponds to the α,β-unsaturation of the enoyl-CoA substrates. Hydration product formation was confirmed by analysis of the assay mixture by MALDI-TOF mass spectrometry.

Crystallization. The ChsH1-ChsH2^{N} apoenzyme crystals were obtained by hanging drop vapor diffusion at rt. Briefly, 1 μL of 15 mg mL^{-1} protein was mixed 1:1 with a reservoir solution of 20%−24% PEG 3350, 20 mM $CaCl_2$, 20 mM $CdCl_2$, and 20 mM $CoCl_2$ at pH 7 or pH 6.5 and equilibrated against 500 μL of the reservoir solution. The crystals were then harvested and transferred to a cryoprotectant solution containing the mother liquor and 17% glycerol. For co-

crystallization, ChsH1-ChsH2^N was mixed with 3-OPC-CoA. Crystals were obtained at 4 °C by hanging drop vapor diffusion, in mother liquor containing 1 mg mL^{-1} 3-OPC-CoA, 20 mM $CaCl_2$, 20 mM $CdCl_2$, 20 mM $CoCl_2$, and 200 mM NaCl at pH 7.0 and 25% PEG 3350. The crystals were harvested and transferred to a cryoprotectant solution containing the mother liquor supplemented with 1 mg mL^{-1} 3-OPC-CoA and 17% ethylene glycol. All crystals were flash cooled in liquid N_2 before data collection.

X-ray Data Collection and Structure Determination. Diffraction data for ChsH1-ChsH2^N were collected on beamline X25 of the National Synchrotron Light Source at Brookhaven National Laboratory (Upton, NY) at a wavelength of 1.7 Å, and for ChsH1-ChsH2^N/3-OPC-CoA, data were collected on beamline 23ID-D of the Advanced Photon Source at Argonne National Laboratory (Argonne, IL) at a wavelength of 1.075 Å. All data sets were processed using XDS[39] and Aimless[40] as implemented in autoPROC.[41] Four cadmium sites in ChsH1-ChsH2^N were located with SHELX,[42] phases were calculated to 1.54 Å via single-wavelength anomalous dispersion[43] using SHARP,[44] and automated modeling building with ARP/wARP[45] produced a starting model consisting of 85% of ChsH1 and 90% of ChsH2^N. For the ChsH1-ChsH2^N/3-OPC-CoA structure, phases were calculated via molecular replacement with Phaser, using ChsH1-ChsH2^N as a search model. Six cadmium sites were identified in the ChsH1-ChsH2^N/3OPC-CoA structure; four of them are located in the tetramer−tetramer packing interface, and two of them coordinate with Asp29/His34 in ChsH2^N, blocking the active site (Supplementary Figure 5b and Note, Supporting Information). For both structures, manual model building was carried out in Coot[46] followed by refinement with Refmac[47] and Phenix.[48] Final model quality was assessed using MolProbity.[49] Data collection and refinement statistics are shown in Table 2.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Additional information on the asymmetric unit of the ChsH1-ChsH2^N, putative role of the DUF35/DUF35_N domain, effects of metals identified in the structures for crystallization, dimer−dimer interface in the heterotetramer, comparisons of MaoC-like enoyl-CoA hydratases across species, electrostatic surface potential map of ChsH1-ChsH2^N/3-OPC-CoA, protein sequence alignment of ChsH1 and its homologs across species, and specific activity data for active site mutants. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

Atomic coordinates and structure factors for ChsH1-ChsH2^N apoenzyme and ChsH1-ChsH2^N/3-OPC-CoA have been deposited in the Protein Data Bank with accession codes 4W78 and 4W7B, respectively.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: nicole.sampson@stonybrook.edu.

Present Address

[§]S.T.T.: Jack H Skirball Center for Chemical Biology and Proteomics, Salk Institute for Biological Studies, La Jolla, CA 92037.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by NIH AI092455 (N.S.S.), NIH AI085349 (N.S.S.), NIH AI065251 (N.S.S.), NIH HL53306 (N.S.S.), NIH RR021008 (N.S.S.), NIH F30-ES022930 (K.E.G), NIH R01-GM100021 (M.G-D.), NSF BIO1039771, and a DOE-GAANN fellowship (S.T.T.). Use of beamlines x25 and x29 at the National Synchrotron Light Source Brookhaven National Laboratory was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-98CH10886. Use of the GM/CA beamlines (23ID-B and 23ID-D) at Advanced Photon Source, Argonne National Laboratory was funded in whole or in part with Federal funds from the National Cancer Institute (Grant Y1-CO-1020) and the National Institute of General Medical Sciences (Grant Y1-GM-1104). Use of the Advanced Photon Source facilities was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. DE-AC02-06CH11357. The authors wish to thank all members of the Sampson and Garcia-Diaz laboratory for insightful discussion and support, as well as Nukri Sanishvili and the CCP4 school at APS for assistance with X-ray data collection.

■ REFERENCES

(1) Lienhardt, C., Glaziou, P., Uplekar, M., Lönnroth, K., Getahun, H., and Raviglione, M. (2012) Global tuberculosis control: Lessons learnt and future prospects. *Nat. Rev. Microbiol. 10*, 407−416.

(2) The World Health Organization. (2012) Global Tuberculosis Report.

(3) Russell, D. G., Cardona, P. J., Kim, M. J., Allain, S., and Altare, F. (2009) Foamy macrophages and the progression of the human tuberculosis granuloma. *Nat. Immunol. 10*, 943−948.

(4) McKinney, J. D., Honer zu Bentrup, K., Munoz-Elias, E. J., Miczak, A., Chen, B., Chan, W. T., Swenson, D., Sacchettini, J. C., Jacobs, W. R., Jr., and Russell, D. G. (2000) Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature 406*, 735−738.

(5) Cole, S. T., Brosch, R., Parkhil, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature 393*, 537−544.

(6) Chang, J. C., Harik, N. S., Liao, R. P., and Sherman, D. R. (2007) Identification of mycobacterial genes that alter growth and pathology in macrophages and in mice. *J. Infect. Dis. 196*, 788−795.

(7) Rengarajan, J., Bloom, B. R., and Rubin, E. J. (2005) Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc. Natl. Acad. Sci. U.S.A. 102*, 8327−8332.

(8) Agnihotri, G., and Liu, H.-w. (2003) Enoyl-CoA hydratase. Reaction, mechanism, and inhibition. *Bioorg. Med. Chem. 11*, 9−20.

(9) Wipperman, M. F., Sampson, N. S., and Thomas, S. T. (2014) Pathogen roid rage: Cholesterol utilization by *Mycobacterium tuberculosis*. *Crit. Rev. Biochem. Mol. Biol. 49*, 269−293.

(10) Van der Geize, R., Yam, K., Heuser, T., Wilbrink, M. H., Hara, H., Anderton, M. C., Sim, E., Dijkhuizen, L., Davies, J. E., Mohn, W. W., and Eltis, L. D. (2007) A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proc. Natl. Acad. Sci. U.S.A. 104*, 1947−1952.

(11) Chang, J. C., Miner, M. D., Pandey, A. K., Gill, W. P., Harik, N. S., Sassetti, C. M., and Sherman, D. R. (2009) *igr* Genes and *Mycobacterium tuberculosis* cholesterol metabolism. *J. Bacteriol. 191*, 5232−5239.

(12) Thomas, S. T., VanderVen, B. C., Sherman, D. R., Russell, D. G., and Sampson, N. S. (2011) Pathway profiling in *Mycobacterium tuberculosis*: Elucidation of cholesterol-derived catabolite and enzymes that catalyze its metabolism. *J. Biol. Chem. 286*, 43668−43678.

(13) Ouellet, H., Guan, S., Johnston, J. B., Chow, E. D., Kells, P. M., Burlingame, A. L., Cox, J. S., Podust, L. M., and De Montellano, P. R. O. (2010) *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one. *Mol. Microbiol. 77*, 730−742.
(14) Johnston, J. B., Ouellet, H., and Ortiz de Montellano, P. R. (2010) Functional redundancy of steroid C26-monooxygenase activity in *Mycobacterium tuberculosis* revealed by biochemical and genetic analyses. *J. Biol. Chem. 285*, 36352−36360.
(15) Thomas, S. T., and Sampson, N. S. (2013) *Mycobacterium tuberculosis* utilizes a unique heterotetrameric structure for dehydrogenation of the cholesterol side chain. *Biochemistry 52*, 2895−2904.
(16) Hisano, T., Tsuge, T., Fukui, T., Iwata, T., Miki, K., and Doi, Y. (2003) Crystal structure of the (*R*)-specific enoyl-CoA hydratase from *Aeromonas caviae* involved in polyhydroxyalkanoate biosynthesis. *J. Biol. Chem. 278*, 617−624.
(17) Jiang, L. L., Kurosawa, T., Sato, M., Suzuki, Y., and Hashimoto, T. (1997) Physiological role of D-3-hydroxyacyl-CoA dehydratase/D-3-hydroxyacyl-CoA dehydrogenase bifunctional protein. *J. Biochem. 121*, 506−513.
(18) Russell, D. W. (2003) The enzymes, regulation, and genetics of bile acid synthesis. *Annu. Rev. Biochem. 72*, 137−174.
(19) Koski, M. K., Haapalainen, A. M., Hiltunen, J. K., and Glumoff, T. (2004) A two-domain structure of one subunit explains unique features of eukaryotic hydratase 2. *J. Biol. Chem. 279*, 24666−24672.
(20) Dillon, S. C., and Bateman, A. (2004) The Hotdog fold: Wrapping up a superfamily of thioesterases and dehydratases. *BMC Bioinf. 5*, No. 109.
(21) Holm, L., and Rosenström, P. (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res. 38*, W545−W549.
(22) Koski, K. M., Haapalainen, A. M., Hiltunen, J. K., and Glumoff, T. (2005) Crystal structure of 2-enoyl-CoA hydratase 2 from human peroxisomal multifunctional enzyme type 2. *J. Mol. Biol. 345*, 1157−1169.
(23) Doi, Y., Kitamura, S., and Abe, H. (1995) Microbial synthesis and characterization of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate). *Macromolecules 28*, 4822−4828.
(24) Zhao, S., Kumar, R., Sakai, A., Vetting, M. W., Wood, B. M., Brown, S., Bonanno, J. B., Hillerich, B. S., Seidel, R. D., Babbitt, P. C., Almo, S. C., Sweedler, J. V., Gerlt, J. A., Cronan, J. E., and Jacobson, M. P. (2013) Discovery of new enzymes and metabolic pathways by using structure and genome context. *Nature 502*, 698−702.
(25) Wipperman, M. F., Yang, M., Thomas, S. T., and Sampson, N. S. (2013) Shrinking the FadE proteome of *Mycobacterium tuberculosis*: Insights into cholesterol metabolism through identification of an $\alpha_2\beta_2$ heterotetrameric acyl-Coenzyme A dehydrogenase family. *J. Bacteriol. 195*, 4331−4341.
(26) Krishna, S. S., Aravind, L., Bakolitsa, C., Caruthers, J., Carlton, D., Miller, M. D., Abdubek, P., Astakhova, T., Axelrod, H. L., Chiu, H. J., Clayton, T., Deller, M. C., Duan, L., Feuerhelm, J., Grant, J. C., Han, G. W., Jaroszewski, L., Jin, K. K., Klock, H. E., Knuth, M. W., Kumar, A., Marciano, D., McMullan, D., Morse, A. T., Nigoghossian, E., Okach, L., Reyes, R., Rife, C. L., van den Bedem, H., Weekes, D., Xu, Q., Hodgson, K. O., Wooley, J., Elsliger, M. A., Deacon, A. M., Godzik, A., Lesley, S. A., and Wilson, I. A. (2010) The structure of SSO2064, the first representative of Pfam family PF01796, reveals a novel two-domain zinc-ribbon OB-fold architecture with a potential acyl-CoA-binding role. *Acta Crystallogr., Sect. F: Struct. Biol. Commun. 66*, 1160−1166.
(27) Krishna, S. S., Weekes, D., Bakolitsa, C., Elsliger, M. A., Wilson, I. A., Godzik, A., and Wooley, J. (2010) TOPSAN: Use of a collaborative environment for annotating, analyzing and disseminating data on JCSG and PSI structures. *Acta Crystallogr., Sect. F: Struct. Biol. Commun. 66*, 1143−1147.
(28) Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., and Yeates, T. O. (1999) Detecting protein function and protein-protein interactions from genome sequences. *Science 285*, 751−753.
(29) Enright, A. J., Iliopoulos, I., Kyrpides, N. C., and Ouzounis, C. A. (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature 402*, 86−90.
(30) Castell, A., Johansson, P., Unge, T., Jones, T. A., and Bäckbro, K. (2005) Rv0216, a conserved hypothetical protein from *Mycobacterium tuberculosis* that is essential for bacterial survival during infection, has a double hotdog fold. *Protein Sci. 14*, 1850−1862.
(31) Sacco, E., Legendre, V., Laval, F., Zerbib, D., Montrozier, H., Eynard, N., Guilhot, C., Daffé, M., and Quémard, A. (2007) Rv3389C from *Mycobacterium tuberculosis,* a member of the (*R*)-specific hydratase/dehydratase family. *Biochim. Biophys. Acta, Proteins Proteomics 1774*, 303−311.
(32) Johansson, P., Castell, A., Jones, T. A., and Bäckbro, K. (2006) Structure and function of Rv0130, a conserved hypothetical protein from *Mycobacterium tuberculosis*. *Protein Sci. 15*, 2300−2309.
(33) Kim, Y.-G., and Maas, S. (2001) Multiple site-directed mutagenesis *in vitro*, in *In Vitro Mutagenesis Protocols* (Braman, J., Ed.), pp 29−36, Springer/Humana Press, Totowa, NJ.
(34) Winkler, R. (2010) ESIprot: A universal tool for charge state determination and molecular weight calculation of proteins from electrospray ionization mass spectrometry data. *Rapid Commun. Mass Spectrom. 24*, 285−294.
(35) Tserng, K. Y., Jin, S. J., and Hoppel, C. L. (1991) Spiropentaneacetic acid as a specific inhibitor of medium-chain acyl-CoA dehydrogenase. *Biochemistry 30*, 10755−10760.
(36) Slomp, G., Jr., and Johnson, J. L. (1958) Ozonolysis. II. The effect of pyridine on the ozonolysis of 4, 22-stigmastadien-3-one. *J. Am. Chem. Soc. 80*, 915−921.
(37) Mohn, W. W., Wilbrink, M. H., Casabon, I., Stewart, G. R., Liu, J., van der Geize, R., and Eltis, L. D. (2012) Gene cluster encoding cholate catabolism in *Rhodococcus spp*. *J. Bacteriol. 194*, 6712−6719.
(38) Casabon, I., Crowe, A. M., Liu, J., and Eltis, L. D. (2013) FadD3 is an acyl-CoA synthetase that initiates catabolism of cholesterol rings C and D in actinobacteria. *Mol. Microbiol. 87*, 269−283.
(39) Kabsch, W. (2010) XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 125−132.
(40) Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr. 69*, 1204−1214.
(41) Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67*, 293−302.
(42) Sheldrick, G. (2008) A short history of SHELX. *Acta Crystallogr., Sect. A: Found. Crystallogr. 64*, 112−122.
(43) Dauter, Z., Dauter, M., and Dodson, E. (2002) Jolly SAD. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 494−506.
(44) Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP, in *Methods in Molecular Biology* (Doublié, S., Ed.), pp 215−230, Humana Press, Totowa, NJ.
(45) Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc. 3*, 1171−1179.
(46) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2126−2132.
(47) Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67*, 355−367.
(48) Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 1948−1954.
(49) Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: All-atom contacts and

structure validation for proteins and nucleic acids. *Nucleic Acids Res. 35*, W375−W383.