Published in final edited form as:

Biochem Biophys Res Commun. 2010 April 16; 394(4): 940–946. doi:10.1016/j.bbrc.2010.03.087.

Crystal structure and functional insight of HP0420-homolog from *Helicobacter felis*

Shunfu Piao^a, Xiao Ling Jin^a, Bo-Young Yun^a, Nahee Kim^b, Hyun-Soo Cho^b, Minoru Fukuda^c, Heeseob Lee^d, and Nam-Chul Ha^{a,*}

^a College of Pharmacy and Research Institute for Drug Development, Pusan National University, Jangjeon-dong, Geumjeong-gu, Busan 609-735, Republic of Korea

^b Department of Biology, College of Science, Yonsei University, Seoul 120-749, Republic of Korea

^c Tumor Microenvironment Program, Cancer Research Center, Burnham Institute for Medical Research, La Jolla, CA 92037, USA

^d Department of Food Science and Nutrition, Pusan National University, Jangjeon-dong, Geumjeong-gu, Busan 609-735, Republic of Korea

Abstract

Helicobacter pylori infect more than half of the world's population and are considered a cause of peptic ulcer disease and gastric cancer. Recently, hypothetical gene HP0421 was identified in *H. pylori* as a cholesterol α -glucosyltransferase, which is required to synthesize cholesteryl glucosides, essential cell wall components of the bacteria. In the same gene-cluster, HP0420 was co-identified, whose function remains unknown. Here we report the crystal structure of HP0420-homolog of *H. felis* (HF0420) to gain insight into the function of HP0420. The crystal structure, combined with size-exclusion chromatography, reveals that HF0420 adopts a homodimeric hot-dog fold. The crystal structure suggests that HF0420 has enzymatic activity that involves a conserved histidine residue at the end of the central α -helix. Subsequent biochemical studies provide clues to the function of HP0420 and HF0420.

Keywords

Crystal structure; *Helicobacter*; Hot-dog fold; Structural genomics; HP0420

1. Introduction

Helicobacter pylori is a Gram-negative bacteria that lives in human stomach and may cause gastric chronic inflammation and even stomach cancer [1]. *Helicobacter* species have cholesteryl glucosides (CGs) as unique and essential cell membrane components [2,3]. CGs are synthesized from cholesterol, and thus cholesterol must be taken up from the host cell because the bacteria lack the genes responsible for its biosynthesis [4]. Recently, it was reported that the hypothetical gene HP0421 from *H. pylori* exhibits cholesterol α -glucosyltransferase activity that converts cholesterol into CGs and can be inhibited by the natural antibiotic mucin O-glycan in deeper portions of the gastric mucosa [5]. Bacterial growth is severely inhibited

*Corresponding author. hnc@pusan.ac.kr (N.-C. Ha).

Conflict of interest

The authors declare that they have no conflicts of interest.

in the absence of the *HP0421* gene [6,7], suggesting a critical role for HP0421 in the survival of *H. pylori*.

During expression cloning of cholesterol α -glucosyltransferase from *H. pylori* and *H. felis*, Lee et al. identified HP0420 and its homolog in a single plasmid harboring genomic sequences for two open reading frames (HP0420 and HP0421) [6]. *H. felis*, which lives in cat stomach [8] and shows genomic similarity to *H. pylori*, has been used to study colonization, pathogenesis, and eradication of *H. pylori* [9,10]. The gene structures of HP0420 and HP0421 were conserved in both *Helicobacter* species. Unlike HP0421, HP0420 was not essential for the survival of the bacteria [6]. Although the gene structure suggested that the function of HP0420 might be associated with HP0421, such association remains to be elucidated [6]. The HP0420-homolog from *H. felis* (herein referred to as HF0420) shares high sequence identity (40%) with HP0420, indicating that HF0420 is a functional and structural homolog of HP0420 [6].

To gain insight into the function of the hypothetical protein HP0420 and its homologs from *Helicobacter* species, we determined the crystal structure of HF0420 from *H. felis* and performed subsequent biochemical studies.

2. Materials and methods

2.1. DNA construction and protein expression

DNA fragments encoding HF0420 (residues 1–141) and HP0420 (residues 1–142) were amplified from *H. felis* and *H. pylori* genomic DNA, respectively, using the polymerase chain reaction. The DNA fragments were inserted into the *Nco*I and *Xho*I sites of pProEXHTa (Invitrogen, USA) containing the hexa-histidine tag and the TEV protease cleavage sites at the N-terminus. The recombinant HF0420 and HP0420 proteins were expressed in *Escherichia coli* strain BL21 (DE3) RIL in Luria–Bertani (LB) medium at 37 °C until the OD_{600 nm} reached 0.5. The protein was expressed by adding 0.5 mM isopropyl- β -D-1-thiogalactopyranoside (IPTG). The cells were harvested by centrifugation at 5000 rpm for 15 min at 4 °C.

2.2. Protein purification of HF0420 and HP0420

To produce HF0420 and HP0420 proteins, harvested cells were suspended in lysis buffer containing 20 mM Tris (pH 8.0), 150 mM NaCl and 2 mM β -mercaptoethanol, and were disrupted by sonication. The lysate was centrifuged at 13,000 rpm for 30 min at 4 °C. The resulting supernatant was loaded onto Ni–NTA agarose resin that was pre-equilibrated with lysis buffer. The resin was washed with lysis buffer supplemented with 20 mM imidazole and then eluted with lysis buffer supplemented with 200 mM imidazole. The fractions containing the HF0420 protein were pooled and treated with recombinant TEV protease overnight at room temperature to remove the hexa-histidine tag after addition of 10 mM β -mercaptoethanol. The reaction mixture was subsequently loaded onto a Q-anion exchange column (Hitrap-Q; GE Healthcare, USA) for further purification, and eluted with a gradient from 0 to 1 M NaCl. The collected fractions containing the target protein were pooled, concentrated, and separated on a HiLoad Superdex 200 gel-filtration column (GE Healthcare, USA) pre-equilibrated with lysis buffer. During the purification, the presence of the protein was confirmed by SDS–PAGE. The purified HF0420 and HP0420 proteins were concentrated to 60 mg/ml and 40 mg/ml, respectively, using Centriprep (Millipore, USA) and stored frozen at –80 °C until use.

2.3. Crystallization, data collection and structure determination

Crystals of the wild-type HF0420 protein were obtained by the vapor-diffusion technique at 14 °C. The initial crystallization screening was performed by the sitting-drop method with Crystal Screen HT, a high-throughput sparse-matrix screening kit (Hampton Research, USA). Crystals grown in the solution containing 0.1 M Tris–HCl (pH 8.5) and 2 M ammonium sulfate

were directly chosen from the initial crystallization screening plate for data collection. For cryoprotection, the HF0420 crystals were soaked with the sticky oil Paratone-N. The data sets were collected on BL6C at Pohang Accelerator Laboratory with a CCD detector Quantum 210 (ADSC) at -173°C . The diffraction data sets were processed and scaled with the HKL2000 package [11]. The crystal belongs to the space group $P2_12_12_1$ with cell dimensions of $a = 70.5$, $b = 70.3$, and $c = 58.8$ Å. Initial phases were determined by the molecular replacement package MOLREP [12] using the coordinates of Cj0977 (Protein Data Bank code 3BNV) as a search model. Model building was performed using the program COOT [13], and model refinement was conducted using the program CNS 1.2 [14]. PHENIX. REFINER [15] was applied at the final round of the model refinement.

Crystals of mutant HF0420 (C46A) were obtained by the same method as wild-type HF0420, and the crystallization conditions were optimized to produce high quality crystals in droplets containing 1 μl of protein solution and 1 μl of a precipitant solution consisting of 0.01 M cobalt (II) chloride hexahydrate, 0.1 M MES, pH 6.5, and 1.8 M ammonium sulfate. Data sets were collected as described above, and the crystal belonged to the space group $P2_12_12_1$ with cell dimensions of $a = 53.3$, $b = 67.1$, and $c = 70.2$ Å. Initial phases were determined by the molecular replacement package MOLREP [12] using coordinates of the wild-type HF0420 structure as a search model.

Crystallographic data statistics are summarized in Table 1. All figures were prepared with PYMOL [16].

2.4. Size-exclusion chromatography

Size-exclusion chromatography was performed at room temperature at a flow rate of 0.5 ml/min on Superdex S-200 HR 10/30 (GE Healthcare) equilibrated with 20 mM Tris buffer (pH 8.0) containing 150 mM NaCl and 2 mM β -mercaptoethanol. Five hundred microliters of each protein (100 μg) were injected onto the column.

2.5. Thermal stability

Thermal stability studies of HF0420 (wild-type and C46A) and HP0420 were performed by circular dichroism in a JASCO-J750 spectropolarimeter. Samples were prepared in 20 mM Tris, 0.15 M NaCl, pH 8.0, and thermal unfolding experiments were performed by monitoring the circular dichroism signal at 220 nm between 25°C and 90°C using a heating rate of $2^{\circ}\text{C}/\text{min}$ at a concentration of 0.25 mg/ml.

2.6. Isothermal titration calorimeter (ITC)

Isothermal titration calorimeter (ITC) measurement was done using a VP-ITC from MicroCal, Inc. (Northampton, MA). HF0420 protein and CoA were dissolved in 100 mM phosphate buffer (pH 8.0) at a final concentration of 20 μM and 150 μM , respectively. The titration of HF0420 with CoA involved 25 injections of CoA solution, of 10 μl each. The duration of each injection was 20 s, and the time delay between successive injections was 220 s. The contents of the sample cell were stirred at 300 rpm throughout the experiment to ensure thorough mixing. The temperature of the titration cell was set at 25°C .

3. Results

3.1. Structural determination of HF0420

To gain insight into the functions of HP0420 and HF0420, we attempted to determine the three-dimensional structures of HP0420 and HF0420. Full-length proteins were successfully overexpressed in *E. coli* and were purified to homogeneity. The purified proteins were highly soluble and were concentrated for crystallization. Fortunately, we obtained crystals of HF0420

under several conditions, and these crystals were suitable for data collection without optimization. However, we failed to obtain crystals of HP0420 despite extensive screening of the crystallization conditions.

The crystal structure of HF0420 was solved by molecular replacement using Cj0977 from *Campylobacter jejuni*, which shows the highest sequence homology, as a search model. The asymmetric unit of the crystal contains two molecules, consists of 248 residues, and comprises 89% of the total number of amino acid residues in the crystallized protein (Fig. 2A). The Matthews coefficient (V_M) of the crystal [17] was $2.1 \text{ \AA}^3/\text{Da}$ and the solvent content was calculated to be 42%.

3.2. Overall structure and dimeric nature of HF0420

The crystal structure of HF0420 revealed that HF0420 adopts a “hot-dog” fold that consists of a central long α -helix and a surrounding four-stranded β -sheet (Fig. 2B). The hot-dog fold was so named because the α -helix resembles a sausage of hot-dog, and the β -sheet wraps around this helix like a bun [18]. In addition to the hot-dog fold elements, loops and short α -helices are found in the N- and C-terminal regions of the HF0420 structure (Fig. 2B).

The structural homologs of HF0420 were searched using the DALI server [19]. A nonflagella virulence protein from *C. jejuni* (Cj0977) that has a putative acyl-CoA-binding site [20] was indicated as the top match (Z-score 21.9) (Fig. 2C). A hypothetical protein PH1136 from *Pyrococcus horikoshii* [21] and malonyl-CoA-binding transcription factor FapR from *Bacillus subtilis* [22] were also indicated as matches (Z-scores 21.5 and 18.7). Phenylacetyl-CoA thioesterase from *Thermus thermophilus* (PaaI) and other acyl-CoA thioesterases were matched, as listed in Supplementary Table S1.

All hot-dog fold proteins form dimers, and some hot-dog fold proteins form tetramers or hexamers by dimerization or trimerization of the dimeric unit [20]. Given the extensive dimeric interface between two molecules in the asymmetric unit (approximately 30% of HF0420 surface area), the crystal structure suggests that HF0420 forms a dimer. Subsequent size-exclusion chromatography showed that both HF0420 and HP0420 proteins are present in dimeric form in solution (Fig. 2D). These results indicate that HF0420 and HP0420 function as dimeric units.

3.3. Putative substrate binding region

The hot-dog fold proteins were first reported in the crystal structure of FabA [18], and this fold was found in diverse thioesterases and coenzyme A (CoA)-derived binding proteins. To obtain insight into the function of the hypothetical proteins HP0420 and its homologs, we superposed the structure of HF0420 with PaaI in complex with phenylacetyl-CoA (PDB code: 1WN3) (Z-score 18.7). Structures of HF0420 with PaaI were well fitted, and the superposition identified a putative substrate binding site of HF0420 from the location of the superposed phenylacetyl-CoA (Fig. 3A and B). The putative active site is located at the interface between two protomers of HF0420. The nucleotide moiety of CoA is positioned in a shallow groove at the surface of one subunit. The thioester bond of acyl-CoA is positioned at the N-terminal end of the central α -helix, where the His52 and Gly54 residues are present (Fig. 3C). The hydrocarbon moiety of CoA is in a deep crevice formed at the dimeric interface (Fig. 3B). Based on this structural superposition, we hypothesize that the real substrate of HF0420 may be acyl-CoA or its derivatives.

3.4. Putative catalytic residue

The position of the His52 residue in HF0420 is fixed by the interaction with the conserved Gly54 residue, and the residues are buried in the main body of the protein (Fig. 3B). However,

His52 is accessible from the solvent through a narrow channel that is formed by a long loop from the neighboring subunit (Fig. 3B). Interestingly, the histidine and glycine residues are conserved in HP0420 and its homologs of *Helicobacter* species, whereas they are replaced in Paal and acyl-CoA-binding hot-dog proteins such as Cj0977 (Fig. 1).

The central α -helix of hot-dog fold proteins is highly polarizing because the long α -helix is surrounded by the largely hydrophobic β -sheet (Fig. 2B). Since the conserved His52 residue is located at the positive (N-terminal) end of the polarizing α -helix, where the negative charge is stabilized and positive charge is destabilized, the electron density of the unshared electron pair of the His52 imidazole ring may become richer. Thus, the nucleophilic reactivity of His52 is likely to be enhanced due to the electron-rich unshared electron pair. Moreover, the imidazole ring of His52 makes a hydrogen bond to the NH group of the conserved Gly54 residue, which may also augment the negative charge of His52 (Fig. 3C). Given that no residues interact with the side chain of His52 except for Gly54, His52 might directly attack the substrate in the enzymatic reaction. Taken together, the structural features of HF0420 indicate that the HP0420 family proteins may be hydrolases in which the conserved histidine is a catalytic residue that directly attacks the substrate.

3.5. N-terminal cysteines of HF0420 form inter-subunit disulfide bonds

The HF0420 crystal structure shows that two cysteine residues (Cys15 and Cys46) form inter-subunit disulfide bonds, which is the first observation in the hot-dog fold proteins (Fig. 2A). This disulfide bridge is located around His52, connecting the loops that form the narrow channel to the histidine residue (Fig. 3C). Since there is no signal sequence for secretion in HF0420, HF0420 is thought to be present in the cytosol, where most cysteines are present in reduced form, when expressed in bacteria. However, the possibility that HF0420 is secreted to the extracellular medium, where cysteines are oxidized to form a disulfide bond as observed in the crystal structure, by some unknown mechanism cannot be excluded. To examine whether this inter-subunit disulfide bond is involved in the oxidation/reduction-induced conformational change, we generated and crystallized a C46A mutant HF0420 protein in which the disulfide bond is not formed. Comparison of the crystal structures and the molecular sizes in solution revealed that the disulfide bridge did not affect the conformation of the loop. As shown in Fig. 4A and B, the structure of the C46A mutant was very similar to that of wild-type HF0420, except for the mutated residue. However, the mutant protein showed reduced thermostability compared to wild-type HF0420, as measured by circular dichroism spectroscopy (Fig. 4C). This result indicates that this disulfide bond contributes to the structural integrity of HF0420, although the disulfide bond is not associated with the oxidation/reduction-induced conformational change of HF0420.

In HP0420 from *H. pylori*, the inter-subunit disulfide bridge is not expected to be present because of the lack of the corresponding Cys46 residue. Nonetheless, HP0420 was intrinsically thermostable, as the recombinant HP0420 protein showed high thermostability ($T_m = 95^\circ\text{C}$) (Fig. 4C).

4. Discussion

In this study, we demonstrated that HP0420 and HF0420 are dimeric proteins and adopt the hot-dog fold. The crystal structure of HF0420 suggested that HP0420 family proteins may be hydrolases. However, the substrate of HP0420 family proteins remains unknown. Sequence homology using CLUSTALX [23] and BLAST [24] suggests that HP0420, HF0420, and HAC_0608 from *H. acinonychis* are most similar, and phylogenetic analysis [25] suggests that the three proteins comprise a subgroup distinct from other hot-dog fold proteins (Supplementary Fig. S1). Moreover, they are all followed by cholesterol α -glucotransferase in the gene structure. Thus, we examined whether HF0420 has cholesterol esterase activity, since

HP0420 might be functionally involved with cholesteryl α -glucotransferase (HP0421). This idea is supported by the facts that cholesterol esters are the major storage form in serum rather than cholesterol, cholesterol esters are similar to acyl-CoA in size, and cholesteryl α -glucotransferase requires cholesterol, not cholesterol ester, as a substrate. Thus, cholesterol esterase activity would be needed to efficiently uptake the substrate of cholesteryl α -glucotransferase. However, cholesterol esterase activity was not detected from HF0420 protein (data not shown).

Hot-dog fold proteins have been characterized as acyl-CoA thioesterases or acyl-CoA-binding proteins, and acyl-CoA or its derivatives are suggested as the substrate of HF0420 by the structural superposition in this study. Thus we next examined whether HF0420 exhibits acyl-CoA thioesterase activity. The recombinant HF0420 did not show any acyl-CoA thioesterase activity when palmitoyl-CoA was used as the substrate in an enzymatic assay described previously [26] (Supplementary Fig. S2). Moreover, no CoA-binding ability of HF0420 was detected by titration calorimetry experiments (Supplementary Fig. S3). These results seem to exclude the possibility that HP0420 family proteins are associated with CoA, and suggest that an acyl-CoA analog, not acyl-CoA, is the substrate.

Since the closest structural homolog, Cj0977, is known as a virulence factor that is transferred into the cytosol of host cells, we next examined the virulence of HF0420. HF0420 did not show any cytotoxic effect on human cells when human embryonic kidney cells were directly treated with HF0420 protein or when 293 cells were transfected with the HF0420 gene for expression in the cytosol of human cells (data not shown). Taken together, HP0420 family proteins may not be virulence factors on their own and do not change the essential processes of eukaryotic cells.

In this study, the crystal structure of HF0420 was determined to gain insight into the function of HP0420 family proteins, and subsequent diverse experiments were carried out to reveal the function of the proteins. Although no conclusive results were produced, the experiments suggested that HP0420 family proteins may not be acyl-CoA thioesterases, CoA-binding proteins, cholesterol esterases, or virulence factors that affect mammalian cell survival. This study provides clues to the function of HP0420 and HF0420, which may be different from other hot-dog fold proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study made use of beamline 6C at Pohang Accelerator Laboratory (Pohang, Korea). This study was supported by a grant from the 21C Frontier Microbial Genomics and Applications Center Program, Ministry of Education, Science & Technology, Republic of Korea. The coordinates of the *HF0420* structures have been deposited into the Protein Data Bank (PDB codes: 3LW3 and 3LWG).

References

1. Brown LM. *Helicobacter pylori*: epidemiology and routes of transmission. *Epidemiol Rev* 2000;22:283–297. [PubMed: 11218379]
2. Hirai Y, Haque M, Yoshida T, Yokota K, Yasuda T, Oguma K. Unique cholesteryl glucosides in *Helicobacter pylori*: composition and structural analysis. *J Bacteriol* 1995;177:5327–5333. [PubMed: 7665522]
3. Haque M, Hirai Y, Yokota K, Mori N, Jahan I, Ito H, Hotta H, Yano I, Kanemasa Y, Oguma K. Lipid profile of *Helicobacter* spp.: presence of cholesteryl glucoside as a characteristic feature. *J Bacteriol* 1996;178:2065–2070. [PubMed: 8606185]

4. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzgerald LM, Lee N, Adams MD, Hickey EK, Berg DE, Gocayne JD, Utterback TR, Peterson JD, Kelley JM, Cotton MD, Weidman JM, Fujii C, Bowman C, Watthey L, Wallin E, Hayes WS, Borodovsky M, Karp PD, Smith HO, Fraser CM, Venter JC. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 1997;388:539–547. [PubMed: 9252185]
5. Kawakubo M, Ito Y, Okimura Y, Kobayashi M, Sakura K, Kasama S, Fukuda MN, Fukuda M, Katsuyama T, Nakayama J. Natural antibiotic function of a human gastric mucin against *Helicobacter pylori* infection. *Science* 2004;305:1003–1006. [PubMed: 15310903]
6. Lee H, Kobayashi M, Wang P, Nakayama J, Seeberger PH, Fukuda M. Expression cloning of cholesterol alpha-glucosyltransferase, a unique enzyme that can be inhibited by natural antibiotic gastric mucin O-glycans, from *Helicobacter pylori*. *Biochem Biophys Res Commun* 2006;349:1235–1241. [PubMed: 16978585]
7. Lebrun AH, Wunder C, Hildebrand J, Churin Y, Zahringer U, Lindner B, Meyer TF, Heinz E, Warnecke D. Cloning of a cholesterol-alpha-glucosyltransferase from *Helicobacter pylori*. *J Biol Chem* 2006;281:27765–27772. [PubMed: 16844692]
8. Paster BJ, Lee A, Fox JG, Dewhirst FE, Tordoff LA, Fraser GJ, O'Rourke JL, Taylor NS, Ferrero R. Phylogeny of *Helicobacter felis* sp. nov., *Helicobacter mustelae*, and related bacteria. *Int J Syst Bacteriol* 1991;41:31–38. [PubMed: 1704791]
9. Lavelle JP, Landas S, Mitros FA, Conklin JL. Acute gastritis associated with spiral organisms from cats. *Dig Dis Sci* 1994;39:744–750. [PubMed: 8149839]
10. Wegmann W, Aschwanden M, Schaub N, Aenishanslin W, Gyr K. Gastritis associated with *Gastrosprillum hominis*—a zoonosis? *Schweiz Med Wochenschr* 1991;121:245–254. [PubMed: 2011719]
11. Otwinosky Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997;276:307–326.
12. CCP4i. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D: Biol Crystallogr* 1994;50:760–763. [PubMed: 15299374]
13. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D: Biol Crystallogr* 2004;60:2126–2132. [PubMed: 15572765]
14. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D: Biol Crystallogr* 1998;54(Pt. 5):905–921. [PubMed: 9757107]
15. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D: Biol Crystallogr* 2002;58:1948–1954. [PubMed: 12393927]
16. DeLano, W. The PyMOL User's Manual. DeLano Scientific; Palo Alto: 2002.
17. Kantardjieff KA, Rupp B. Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci* 2003;12:1865–1871. [PubMed: 12930986]
18. Leesong M, Henderson BS, Gillig JR, Schwab JM, Smith JL. Structure of a dehydratase–isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* 1996;4:253–264. [PubMed: 8805534]
19. Holm L, Sander C. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res* 1998;26:316–319. [PubMed: 9399863]
20. Yokoyama T, Paek S, Ewing CP, Guerry P, Yeo HJ. Structure of a sigma28-regulated nonflagellar virulence protein from *Campylobacter jejuni*. *J Mol Biol* 2008;384:364–376. [PubMed: 18835274]
21. Tajika Y, Sakai N, Tanaka Y, Yao M, Watanabe N, Tanaka I. Crystal structure of conserved protein PH1136 from *Pyrococcus horikoshii*. *Proteins* 2004;55:210–213. [PubMed: 14997554]

22. Schujman GE, Guerin M, Buschiazzi A, Schaeffer F, Llarrull LI, Reh G, Vila AJ, Alzari PM, de Mendoza D. Structural basis of lipid biosynthesis regulation in Gram-positive bacteria. *EMBO J* 2006;25:4074–4083. [PubMed: 16932747]
23. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 1998;23:403–405. [PubMed: 9810230]
24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410. [PubMed: 2231712]
25. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25:1189–1191. [PubMed: 19151095]
26. Yokoyama T, Choi KJ, Bosch AM, Yeo HJ. Structure and function of a *Campylobacter jejuni* thioesterase Cj0915, a hexameric hot dog fold enzyme. *Biochim Biophys Acta* 2009;1794:1073–1081. [PubMed: 19303060]

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.bbrc.2010.03.087.

Fig. 1.

Secondary structure and sequence alignment of hot-dog fold proteins. Secondary structure elements of HF0420 are shown at the top of the sequences, with β -strands and helices denoted as arrows and helical symbols, respectively. Conserved residues only in the HP0420 family proteins are highlighted by the red box and may be involved in catalytic activity. Black shade indicates identical residues in all seven sequences. Gray shades indicate identical residues in 4–6 sequences. The top three sequences show the *H. pylori* family proteins and the bottom three sequences show hot-dog fold proteins. HP0420, *H. pylori* 0420; Hac0608, *H. acinonychis* str. *Sheeba*; HP0496, *H. pylori* 0496; Cj0977, *C. jejuni* 0977; TtPaaI, *T. thermophilus* PaaI and PH1136, *P. horikoshii*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 2.

Overall structure of HF0420. (A) Ribbon representations of the dimeric structure of HF0420 in the asymmetric unit. Two protomers are shown in green and cyan; the non-crystallographic twofold axis is indicated and the inter-subunit disulfide bonds are highlighted in the stick representation. The N- and C-terminal are labeled with "N" and "C", respectively. (B) Ribbon representation of the monomer structure of HF0420. The structure is colored according to secondary structure: α -helix in red, β -strand in yellow, and loop in green. The N- and C-terminal ends are labeled with "N" and "C", respectively. The α -helix and β -strand are labeled. (C) Structural superposition of HF0420 with Cj0977 and PH1136, colored in green, orange, and gray, respectively. The N- and C-terminal ends are labeled. (D) Size-exclusion chromatography of HF0420 and HP0420 proteins. The elution profiles of the HF0420 and HP0420 proteins are shown with standard molecular weight markers. The estimated molecular weight of each protein is 41 kDa and 42 kDa, respectively, which indicates dimers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 3.

Proposed substrate binding region by structural superposition. (A) Superposition of HF0420 with *T. thermophilus* PaaI in complex with phenylacetyl-CoA. HF0420 (green and cyan) and *T. thermophilus* PaaI (gray) are shown in Ca. The ligand (phenylacetyl-CoA) is highlighted with red sticks. (B) CoA fits the narrow channel to the conserved His52 residue. Surface view of CoA fit in the active channel of HF0420. (Left) Substrate is highlighted in stick representation. The protomers of HF0420 are colored in green and cyan, respectively. Conserved His52 residue in the pocket is highlighted in red. (Right) A section view focusing on putative substrate binding groove. Surface representation of the long groove connecting to the active channel. The conserved His52 residue is positioned in the bottom of the narrow channel, and the CoA hydrocarbon and thioester groups are inserted into the narrow channel. (C) Close-up view focusing on thioester bond of acyl-CoA and His52 of HF0420. The thioester bond is highlighted in red. Key amino acid residues are labeled. Central long α -helix is shown in the cartoon (green and cyan for each protomer). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4.

Superposition of wild-type HF0420 and C46A mutant. (A) The structures of wild-type and C46A mutant are in green and magenta, respectively. (B) The inter-subunit disulfide bonds are highlighted with sticks. (C) Measurement of melting temperature (T_m) of wild-type and C46A HF0420, and of HP0420 using circular dichroism spectroscopy. The arrows indicate the T_m value of each protein. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

X-ray data collection and refinement statistics.

Data set	Wild-type	C46A mutant
Source	Beamline 6C at PLS	Beamline 6C at PLS
Wavelength (Å)	1.0000	1.0000
Resolution limit (Å)	30–1.6 (1.66–1.60) ^a	50–1.8 (1.83–1.80) ^a
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Unit cell (Å)	<i>a</i> = 70.6, <i>b</i> = 70.3, <i>c</i> = 53.9 Å, $\alpha = \gamma = \beta = 90^\circ$	<i>a</i> = 53.3, <i>b</i> = 67.1, <i>c</i> = 70.2 Å, $\alpha = \gamma = \beta = 90^\circ$
Redundancy	8.5 (3.5)	9.5 (6.7)
<i>R</i> _{sym} ^b (%)	5.3 (40.0) ^a	8.2 (31.8) ^a
<i>I</i> / σ	49.4 (2.8) ^a	30.5 (3.5) ^a
Completeness (%)	98.1 (84.6) ^a	99.7 (99.4) ^a
<i>Refinement</i>		
Resolution range (Å)	30–1.6 (1.66–1.60) ^a	50–1.8 (1.83–1.80) ^a
<i>R</i> -factor (%)	20.8	18.0
<i>R</i> _{free} ^c (%)	23.3	21.6
Number of refined atoms		
Protein	1941	2012
Water	173	118
Average <i>B</i> value ^b (Å)	37.1	17.4
Rmsd for bonds (Å)	0.017	0.017
Rmsd for angles (°)	1.602	1.65
Ramachandran plot		
Most favored (%)	89.2	93.2
Additionally favored (%)	8.2	5.1

^a The numbers in parentheses are statistics for the highest resolution shell.^b $R_{\text{sym}} = \sum_{hkl} |\sum_i I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ and $\langle I(hkl) \rangle$ are the observed intensity and the mean intensity of related reflections, respectively.^c *R*_{free} was calculated with 10% of the data set.