



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

理學博士學位論文

**Structural Studies of Csd6 Protein from
Helicobacter pylori and Rv2258c Protein from
*Mycobacterium tuberculosis***

2016 年 8 月

서울대학교 大學院

生物物理 및 化學生物學科 生化學專攻

임 하 나

**Structural Studies of Csd6 Protein from
Helicobacter pylori and Rv2258c Protein from
*Mycobacterium tuberculosis***

Thesis by

Ha Na IM

Professor: Se Won Suh

**A Thesis Submitted to the Graduate Faculty of
Seoul National University in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy**

2016

Structural studies of Csd6 protein from *Helicobacter pylori*
and Rv2258c protein from *Mycobacterium tuberculosis*






指導教授 徐世源

이 論文을 理學博士學位 論文으로 提出함
2016年 8月

서울 大學校 大學院
生物物理 및 化學生物學科 生化學專攻

임 하 나

임하나의 理學博士學位論文을 認准함
2016年 8月

委員長	서영재	
副委員長	서세원	
委員	박승벌	
委員	양진구	
委員	이형호	

Abstract

Helicobacter pylori causes gastrointestinal diseases, including gastric cancer. *Mycobacterium tuberculosis* induces tuberculosis, claiming the lives of millions of people in the world every year. Increasing drug-resistance among bacterial pathogens is a serious global health issue. Therefore, discovery of new antimicrobial agents is needed urgently. Identifying the molecular and biological functions of proteins from pathogenic bacteria would provide the groundwork for the development of new antibacterial drug targets. In this study, I have determined the crystal structures of Csd6 from *H. pylori* and Rv2258c from *M. tuberculosis*.

The *H. pylori* Csd6 protein plays a key role in determining the helical cell shape by trimming of peptidoglycan muropeptides. Csd6 is also involved in deglycosylation of the flagellar protein FlaA. The structure of Csd6 reveals that its middle catalytic domain resembles those of L,D-transpeptidases but its pocket-shaped active-site shows distinct variations from the known L,D-transpeptidases. Mass analyses confirm that Csd6 functions only as an L,D-carboxypeptidase but not as an L,D-transpeptidase. D-Ala-complexed structure of Csd6 reveals binding modes of both the

substrate and product to the catalytic domain. On the basis of structure-function analysis, Csd6 and its homologs are proposed to constitute a new family of L,D-carboxypeptidase.

The *M. tuberculosis* Rv2258c protein is predicted to be an *S*-adenosyl-L-methionine-dependent methyltransferase (MTase). MTases mediate a wide variety of cellular processes, such as cell signaling, metabolite synthesis, and gene regulation in nearly all living organisms. The structure of Rv2258c shows that its monomer consists of two domains linked by a long α -helix. The N-terminal domain is essential for dimerization and the C-terminal catalytic domain has the Class I MTase fold. The overall fold of Rv2258c, as well as its interactions with bound sinefungin (or *S*-adenosyl-L-homocysteine), are similar to small-molecule MTases. Rv2258c has a relatively large hydrophobic cavity for binding the methyl-accepting substrate, suggesting that bulky nonpolar molecules might be targeted for methylation by Rv2258c in *M. tuberculosis*.

Keywords: Csd6 / cell shape / L,D-carboxypeptidase / *Helicobacter pylori* / HP0518 / flagellin / peptidoglycan / cell motility / protein structure /

structure-function / Rv2258c / *Mycobacterium tuberculosis* / small-molecule
methyltransferase / sinefungin / S-adenosyl-L-homocysteine

Student Number: 2012-30719

Table of Contents

Abstract	i
Table of Contents	iv
List of Tables	ix
List of Figures	x
Abbreviations	xii
Curriculum Vitae	xiv
Chapter 1	1
Chapter 2	87

Chapter 1. The cell-shape determining Csd6 protein from *Helicobacter pylori* constitutes a new family of L,D-carboxypeptidase

1.1 Introduction	1
1.2 Material and methods	13
1.2.1 Cloning, expression, and purification	13
1.2.2 Crystallization and X-ray data collection	17
1.2.3 Structure determination and refinement	21
1.2.4 <i>In silico</i> docking	22
1.2.5 Peptidase assay by mass analysis	24
1.2.6 Analytical ultracentrifugation	26
1.2.7 Single-molecule fluorescence resonance energy transfer (FRET)	26
1.2.8 Surface plasmon resonance experiment	28
1.3 Results	30
1.3.1. Csd6 functions as an LD-CPase but not as an LD-TPase	30
1.3.2 The Csd6 monomer is organized into three-domain architecture	34
1.3.3 The NTD of Csd6 plays a dominant role in homo-dimerization	37
1.3.4 Active-site of Csd6 LD-CPase domain is tailored for the LD-CPase activity	47
1.3.5 D-Ala-complexed structure reveals binding modes of both the substrate and product	54

1.3.6 The C-terminal NTF2-like domain has a putative binding pocket for pseudaminic acid	61
1.4 Discussion	68
1.5 References	71

Chapter 2. Crystal structure of Rv2258c from *Mycobacterium tuberculosis* H37Rv, an S-adenosyl-L-methionine-dependent methyltransferase

2.1 Introduction	87
2.2 Material and methods	90
2.2.1 Cloning, expression, and purification	90
2.2.2 Crystallization and X-ray data collection	93
2.2.3 Structure determination and refinement	100
2.2.4 Analytical gel filtration	103
2.3 Results	104
2.3.1 Structure determination and comparisons among monomer models of Rv2258c	105
2.3.2 Overall monomer structure of Rv2258c	107
2.3.3 Rv2258c exists as a homodimer	110
2.3.4 Structural similarity search	114
2.3.5 MTase sequence motifs and SFG (or SAH) binding	119
2.3.6 Rv2258c is a unique SAM-dependent MTase in mycobacteria	123
2.3.7 Relatively large substrate-binding cavity of Rv2258c and functional implications	127
2.4 Conclusions	133
2.5 References	134
Abstract (in Korean)	142

Acknowledgements	144
Appendix: Printouts of the first author publications	146

List of Tables

Chapter 1.

Table 1-1. Statistics for data collection, phasing, and model refinement · · 19

Chapter 2.

Table 2-1. Forward and reverse primers used in PCR · · · · · 92

Table 2-2. Data collection and phasing statistics · · · · · 98

Table 2-3. Refinement statistics · · · · · 102

Table 2-4. Modeled residues of Rv2258c structures · · · · · 106

Table 2-5. Structural homologs of Rv2258c as identified by DALI search · ·
· · · · · 116

List of Figures

Chapter 1.

Figure 1-1. Strategies of <i>H. pylori</i> for colonization and survival	4
Figure 1-2. Schematic illustration of peptidoglycan structure and peptidoglycan-modifying hydrolases	8
Figure 1-3. SDS-PAGE analysis and crystals of Csd6 from <i>H. pylori</i>	16
Figure 1-4. LD-CPase activity of Csd6	32
Figure 1-5. Overall structure of <i>H. pylori</i> Csd6	35
Figure 1-6. Oligomeric state of <i>H. pylori</i> Csd6	40
Figure 1-7. Dimeric structure of <i>H. pylori</i> Csd6	44
Figure 1-8. Comparisons of the Csd6 NTD with the PALS1-L27N domain	46
Figure 1-9. Comparison of Csd6 and other LD-TPases	49
Figure 1-10. D-Ala-complexed Csd6 and key residues for the LD-CPase activity	57
Figure 1-11. Sequence alignment of Csd6 homologs	64
Figure 1-12. The NTF-like domain of Csd6	66

Chapter 2.

Figure 2-1. Crystals of Rv2258c-unbound from <i>M. tuberculosis</i>	94
Figure 2-2. Monomer structure and topology diagram of Rv2258c.	108
Figure 2-3. Dimeric structure and the oligomeric state of Rv2258c	112
Figure 2-4. Sequence alignment of Rv2258c with its close structural homologs	117
Figure 2-5. Interactions of SFG (or SAH) with Rv2258c	121
Figure 2-6. Comparison of monomer structures of Rv2258c-SFG and <i>M. smegmatis</i> EgtD.	125
Figure 2-7. Comparison of homodimeric structures of Rv2258c and its close structural homologs	129

Abbreviations

NAM	<i>N</i> -acetylmuramic acid
NAG	<i>N</i> -acetylglucosamine
DL-CPase	D,L-carboxypeptidase
DD-EPase	D,D-endopeptidase
DD-CPase	D,D-carboxypeptidase
LD-TPases	L,D-transpeptidases
LD-CPase	L,D-carboxypeptidase
Pse	pseudaminic acid
NTD	N-terminal domain
NTF2	nuclear transport factor 2
EMTS	ethyl mercury thiosalicylate
PEG	polyethylene glycol
SAD	single-wavelength anomalous diffraction
IFD	induced fit docking
Neu5Ac2en	2-deoxy-2,3-dideoxy- <i>N</i> -acetylneuraminic acid
MALDI	matrix-assisted laser desorption/ionization
QIT	quadrupole ion trap
TOF MS	time-of-flight mass spectrometer
SPR	surface plasmon resonance
RU	response unit
BSA	bovine serum albumin
ASU	asymmetric unit
CSS	complexation significance score
ALEX	alternating-laser excitation

FRET	fluorescence resonance energy transfer
MTase	methyltransferase
SAM	<i>S</i> -adenosyl- <i>L</i> -methionine
SAH	<i>S</i> -adenosyl- <i>L</i> -homocysteine
SFG	sinefungin
r.m.s.	root mean square
TB	tuberculosis

CURRICULUM VITAE of Ha Na IM

PERSONAL

Date of Birth: February 2, 1986
Gender: Female
Nationality: Korea, Republic of
Place of Birth: Seoul, Korea
Marital Status: Unmarried
Address: Department of Biophysics and Chemical Biology,
College of Natural Sciences
Seoul National University, Seoul 08826, Korea
Tel: 82-2-882-3515, Fax: 82-2-889-1568
E-mail: simplehana@snu.ac.kr

EDUCATION

2016.8 **Ph.D. in Biophysics and Chemical Biology**
Seoul National University, Korea
2011.8 **M.S. in Biophysics and Chemical Biology**
Seoul National University, Korea
2009.2 **B.S. in Chemistry**
Sookmyung Women's University, Korea

RESEARCH EXPERIENCE

2012. 3 – 2016.8 Graduate Student Researcher (Ph.D. course)
Department of Biophysics and Chemical Biology, Seoul
National University, Korea
(Research supervisor: Prof. Se Won Suh)
2009. 9 – 2011.8 Graduate Student Researcher (M.S. course)
Department of Biophysics and Chemical Biology, Seoul
National University, Korea
(Research supervisor: Prof. Se Won Suh)

Many times Photon Factory, Japan

(Synchrotron X-ray data collection)
Pohang Light Source, Korea
(Synchrotron X-ray data collection)

SKILLS AND TECHNIQUES

Molecular Biology Cloning; site-directed mutagenesis; expression and purification of proteins in *E. coli* system
Biochemistry Protein purification; enzyme assay
Structural Biology X-ray diffraction; protein crystallography; structure analysis

DISSERTATIONS

Ph.D. thesis (2016).

Title: Structural studies of Csd6 protein from *Helicobacter pylori* and Rv2258c protein from *Mycobacterium tuberculosis*
Seoul National University, Seoul, Korea
Research supervisor: Professor Se Won Suh

M.S. thesis (2011).

Title: Crystallization and preliminary X-ray crystallographic analyses of the HP0518 protein from *Helicobacter pylori*
Seoul National University, Seoul, Korea
Research supervisor: Professor Se Won Suh

PUBLICATIONS of Ha Na IM

First Author Publications:

- (1) **Im, H. N.**, Kim, H. S., An, D. R., Jang, J. Y., Kim, J., Yoon, H. J., Yang, J. K., Suh, S. W. (2016), "Crystal structure of Rv2258c from *Mycobacterium tuberculosis* H37Rv, an S-adenosyl-l-methionine-dependent methyltransferase" *J. Struct. Biol.* **193(3)**, 172-180.
- (2) Kim, H. S. *, **Im, H. N.** *, An, D. R., Yoon, J. Y., Jang, J. Y., Mobashery, S., Heseck, D., Lee, M., Yoo, J., Cui, M., Choi, S., Kim, C., Lee, N. K., Kim, S. J., Kim, J. Y., Bang, G., Han, B. W., Lee, B. I., Yoon, H. J., Suh, S. W. (2015), "The cell shape-determining Csd6 protein from *Helicobacter pylori* constitutes a new family of L,D-carboxypeptidase" *J. Biol. Chem.* **290(41)**, 25103-25117. (*equally contributed)

Other Publications:

- (1) An, D. R., Kim, H. S., Kim, J., **Im, H. N.**, Yoon, H. J., Yoon, J. Y., Jang, J. Y., Heseck, D., Lee, M., Mobashery, S., Kim, S. J., Lee, B. I., Suh, S. W. (2015), "Structure of Csd3 from *Helicobacter pylori*, a cell shape-determining metallopeptidase" *Acta Crystallogr. sect. D Biol. Crystallogr.* **71(3)**, 675-686.
- (2) Kim, H. S., Kim, J., **Im, H. N.**, An, D. R., Lee, M., Heseck, D., Mobashery, S., Kim, J. Y., Cho, K., Yoon, H. J., Han, B. W., Lee, B. I., Suh, S. W. (2014), "Structural basis for the recognition of muramyltripeptide by *Helicobacter pylori* Csd4, a D,L-carboxypeptidase controlling the helical cell shape" *Acta Crystallogr. sect. D Biol. Crystallogr.* **70(11)**, 2800-2812.
- (3) Yoon, J. Y., An, D. R., Yoon, H. J., Kim, H. S., Lee, S. J., **Im, H. N.**, Jang, J. Y., Suh, S. W. (2013), "High-resolution crystal structure of *Streptococcus pyogenes* β -NAD⁺ glycohydrolase in complex with its endogenous inhibitor IFS reveals a highly water-rich interface" *J. Synchrotron Radiat.* **20(6)**, 962-967.
- (4) Yoon, J. Y., Kim, J., An, D. R., Lee, S. J., Kim, H. S., **Im, H. N.**, Yoon, H. J., Kim, J. Y., Kim, S. J., Han, B. W., Suh, S. W. (2013), "Structural and functional characterization of HP0377, a thioredoxin-fold protein from *Helicobacter pylori*" *Acta Crystallogr. sect. D Biol. Crystallogr.* **69(5)**, 735-746.

- (5) Kim, H. S., Kim, J., **Im, H. N.**, Yoon, J. Y., An, D. R., Yoon, H. J., Kim, J. Y., Min, H. K., Kim, S. J., Lee, J. Y., Han, B. W., Suh, S. W. (2013), "Structural basis for the inhibition of *Mycobacterium tuberculosis* L,D-transpeptidase by meropenem, a drug effective against extensively drug-resistant strains" *Acta Crystallogr. sect. D Biol. Crystallogr.* **69(3)**, 420-431.
- (6) Kim, K. H., An, D. R., Song, J., Yoon, J. Y., Kim, H. S., Yoon, H. J., **Im, H. N.**, Kim, J., Kim, D. J., Lee, S. J., Kim, K. H., Lee, H. M., Kim, H. J., Jo, E. K., Lee, J. Y., Suh, S. W. (2012), "*Mycobacterium tuberculosis* Eis protein initiates suppression of host immune responses by acetylation of DUSP16/MKP-7" *Proc. Natl. Acad. Sci. USA* **109(20)**, 7729-7734.
- (7) Yoon, J. Y., Kim, J., Lee, S. J., Kim, H. S., **Im, H. N.**, Yoon, H. J., Kim, K. H., Kim, S. J., Han, B. W., Suh, S. W. (2011), "Structural and functional characterization of *Helicobacter pylori* DsbG" *FEBS Lett.* **585(24)**, 3862-3867.
- (8) Lee, S. J., Kim, D. J., Kim, H. S., Lee, B. I., Yoon, H. J., Yoon, J. Y., Kim, K. H., Jang, J. Y., **Im, H. N.**, An, D. R., Song, J. S., Kim, H. J., Suh, S. W. (2011), "Crystal structures of *Pseudomonas aeruginosa* guanidinobutyrase and guanidinopropionase, members of the ureohydrolase superfamily" *J. Struct. Biol.* **175(3)**, 329-338.

PRESENTATIONS

2016. 05. Poster presentation in KSBMB (Korean Society for Biochemistry and Molecular Biology) Conference, COEX, Seoul, Korea
2015. 11. Oral presentation in the 27th Synchrotron Radiation User's Workshop & KOSUA Meeting, Pohang, Korea
2015. 11. Poster presentation in Korean Crystallographic Association Meeting, Sungkyunkwan University, Suwon, Korea
2015. 05. Oral presentation in KSBMB (Korean Society for Biochemistry and Molecular Biology) Conference, COEX, Seoul, Korea
2014. 08. Poster presentation in the 23rd Congress and General Assembly of the International Union of Crystallography (IUCr), Montreal, Quebec, Canada
2013. 12. Poster presentation in the 12th Conference of the Asian Crystallographic Association (AsCA), Hong Kong, China

- 2013. 11. Poster presentation in the 25th Synchrotron Radiation User's Workshop & KOSUA Meeting, Pohang, Korea
- 2013. 11. Poster presentation in Korean Crystallographic Association Meeting, KIST, Seoul, Korea
- 2012. 12. Poster presentation in AsCA'12/CRYSTAL 28, A Joint Meeting of the Asian Crystallographic Association (AsCA) and SCANZ, Adelaide, Australia
- 2012. 11. Poster presentation in the 24th Synchrotron Radiation User's Workshop & KOSUA Meeting, Pohang, Korea
- 2012. 05. Poster presentation in KSBMB (Korean Society for Biochemistry and Molecular Biology) Conference, COEX, Seoul, Korea
- 2011. 11. Poster presentation in the 23th Synchrotron Radiation User's Workshop & KOSUA Meeting, Pohang, Korea
- 2011. 11. Poster presentation in Korean Crystallographic Association Meeting, Seoul National University of Science & Technology, Seoul, Korea
- 2011. 05. Poster presentation in KSBMB (Korean Society for Biochemistry and Molecular Biology) Conference, COEX, Seoul, Korea
- 2011. 01. Poster presentation in KSMCB (Korean Society for Molecular and Cellular Biology) Conference, Yongpyong Resort, Gangwon, Korea
- 2010. 10. Poster presentation in KSMCB (Korean Society for Molecular and Cellular Biology) Conference, COEX, Seoul, Korea

Chapter 1

The cell-shape determining Csd6 protein from *Helicobacter pylori* constitutes a new family of L,D-carboxypeptidase

[This chapter was published in *J. Biol. Chem.* (2015) 41:25103–25117.]

1.1. Introduction

Helicobacter pylori is a helical-shaped Gram-negative bacterium that colonizes the upper gastrointestinal tract in approximately half of the world's human population. Its infection of the gastric mucosa is associated with various gastric diseases, including chronic gastritis, peptic ulcer, mucosa-associated lymphoid tissue lymphoma, and gastric adenocarcinoma (Roesler et al., 2014; Kusters et al., 2006). *H. pylori* is regarded as a primary factor for gastric cancer development (Roesler et al., 2014) and was classified as a Group I carcinogen by the International Agency for Research on Cancer of the World Health Organization. In recent years, *H. pylori* infection has also been implicated with some extra-digestive diseases

(Roubaud Baudron et al., 2013). Typical treatment regimens for eradicating *H. pylori* consist of a proton pump inhibitor (for example, omeprazole) and the antibiotics such as clarithromycin and amoxicillin (or metronidazole). However, increasing drug resistance in *H. pylori* requires discovery of new antibiotics (Malfertheiner et al., 2012).

High motility of *H. pylori* is important for its colonization of the human stomach and its survival in the viscous gastric mucosa (Ottemann and Lowenthal, 2002; Schreiber et al., 2004; Lertsethtakarn et al., 2011). The helical cell shape of *H. pylori* is believed to facilitate penetration of the viscous epithelial mucus layer via a cork-screwing mechanism (Berg and Turner, 1979; Hazell et al., 1986; Worku et al., 1999) (Figure 1-1). Several *H. pylori* mutants with altered cell shapes exhibited attenuated colonization (Wyckoff et al., 2012). The peptidoglycan layer, the major component of the bacterial cell wall, plays an essential role in withstanding the turgor pressure and in determining the cell shape (Scheffers and Pinho, 2005; Vollmer and Bertsche, 2008). It is made of linear polysaccharide chains that consist of alternating β -1,4-linked *N*-acetylglucosamine-*N*-acetylmuramic acid (NAG-NAM) disaccharide units, with a pentapeptide linked to NAM (Vollmer et al., 2008). In *H. pylori*, the pentapeptide sequence is L-Ala¹- γ -D-Glu²-mDAP³-D-Ala⁴-D-Ala⁵ (or -Gly⁵), where mDAP refers to meso-2,6-

diaminopimelate, and the neighboring peptides are cross-linked exclusively by the 4→3 linkage between the main-chain of D-Ala⁴ from one strand and the side chain of mDAP³ from another strand (Costa et al., 1999) to form a mesh-like structure termed peptidoglycan sacculus (Meroueh et al., 2006; Sauvage et al., 2008). Colonization and infection by *H. pylori* also rely on the ability of the bacterium to move toward the part of gastric mucosa with more neutral pH. Besides the helical morphology, powerful flagella of *H. pylori* are responsible for its high motility through the viscous gastric mucus layer (Ottemann and Lowenthal, 2002; Schreiber et al., 2004; Lertsethtakarn et al., 2011). The flagella provide a propulsive torque as well as a rotary movement of the cell body; a helical cell shape of *H. pylori* generates a corkscrew-like rotation (Lertsethtakarn et al., 2011; Berg and Turner, 1979; Hazell et al., 1986).

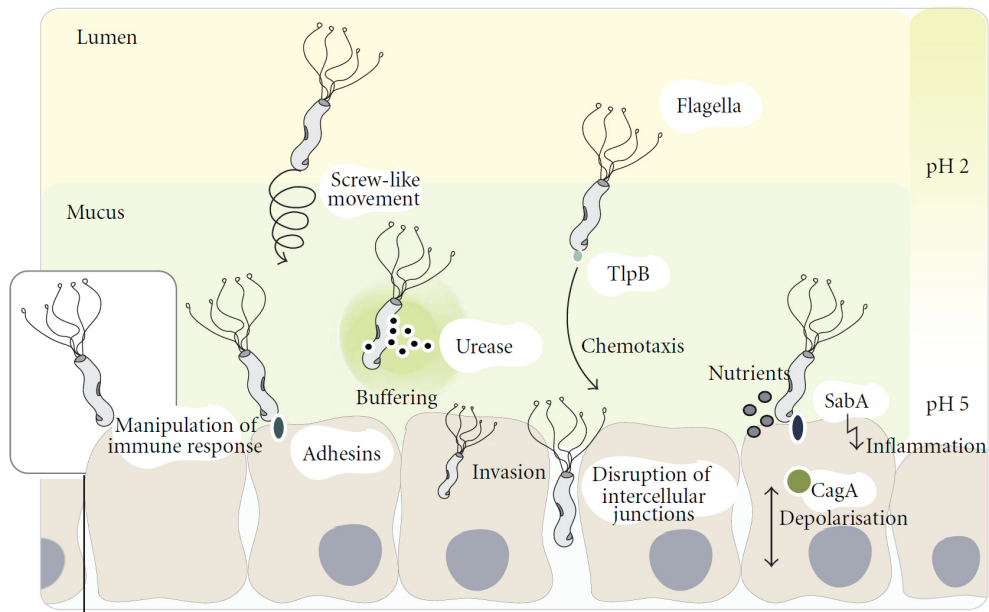


Figure 1-1. Strategies of *H. pylori* for colonization and survival (Bauer and Meyer, 2011)

Using the flagella conferring fast motility, *H. pylori* cells orientate themselves towards the mucosa in close proximity to the epithelial cell layer where the pH is near neutral. Its helical cell shape enables a screw-like movement to facilitate motility through the viscous mucosal layer.

In many bacteria, including *H. pylori*, the peptidoglycan layer is remodeled by a number of cell-wall hydrolases as well as synthetases for the peptidoglycan maturation, regulation of cell-wall growth, cell division, peptidoglycan turnover and recycling, cell lysis, and the release of peptidoglycan fragments for host-pathogen interactions (Vollmer et al., 2008) (Figure 1-2). In *H. pylori*, an amidase AmiA (Chaput et al., 2006), peptidases Csd1–4, potential regulators Csd5 and CcmA (Sycuro et al., 2010; Sycuro et al., 2012; Bonis et al., 2010), and another peptidase Csd6 (Sycuro et al., 2013) are required to tailor the peptidoglycan layer to generate the helical cell shape. I have recently reported the crystal structures of both Csd4 (HP1075 in strain 26695) (Kim et al., 2014) and Csd3 (HP0506 in strain 26695) (An et al., 2015). Csd4 is a Zn²⁺-dependent D,L-carboxypeptidase (DL-CPase) of the M14 metallopeptidase family and cleaves the γ -D-Glu²-mDAP³ bond of the muramyl tripeptide to produce the muramyl dipeptide (Sycuro et al., 2012; Kim et al., 2014). Csd3 (also known as HdpA) belongs to the M23 metallopeptidase family and has both D,D-endopeptidase (DD-EPase) and D,D-carboxypeptidase (DD-CPase) activities (Bonis et al., 2010). Csd6 (encoded by the hp0518 gene in *H. pylori* 26695 strain) was identified as another member of the peptidoglycan trimming pathway and also a cell-shape determinant of *H. pylori* (Sycuro et

al., 2013). The transposon mutant *H. pylori* with the disruption of the *csd6* gene or the *csd6*-deletion mutant displayed a straight rod shape and an increase in tetrapeptide-containing muropeptides (Sycuro et al., 2013). Incubation of the recombinant hexahistidine-tagged Csd6 with tetrapeptide-rich sacculi from the $\Delta csd1csd6$ mutant *H. pylori* resulted in complete conversion of the monomeric tetrapeptides to tripeptides (Sycuro et al., 2013). In *Campylobacter jejuni*, another helical-shaped gastrointestinal pathogen, a homolog of Csd6 has been characterized as an LD-CPase and was named as Pgp2 (Firdich et al., 2014). Loss of *pgp2* resulted in the morphology defect, changes in the peptidoglycan muropeptide profile, reduced motility, and moreover, a decreased interaction with the host (Firdich et al., 2014). Between Csd6 and Pgp2, there exist an overall amino acid sequence identity of 36% and similarity of 58%. Given the significance of mDAP-containing muramyl tripeptide as an agonist for the cytosolic innate immune receptor Nod1 (Girardin et al., 2003; Chamailard et al., 2003), an increased amount of muramyl tripeptide as a result of LD-CPase reaction by Pgp2 (or Csd6 from *H. pylori*) was suggested to affect the Nod1 activation and ultimately the NF- κ B transcriptional activity (Firdich et al., 2014). The complete absence of tripeptide-containing muropeptides in $\Delta pgp2$ peptidoglycan did indeed reflect the reduced Nod1 activation, despite

no effect in either intracellular survival or IL-8 secretion (Firdich et al., 2014).

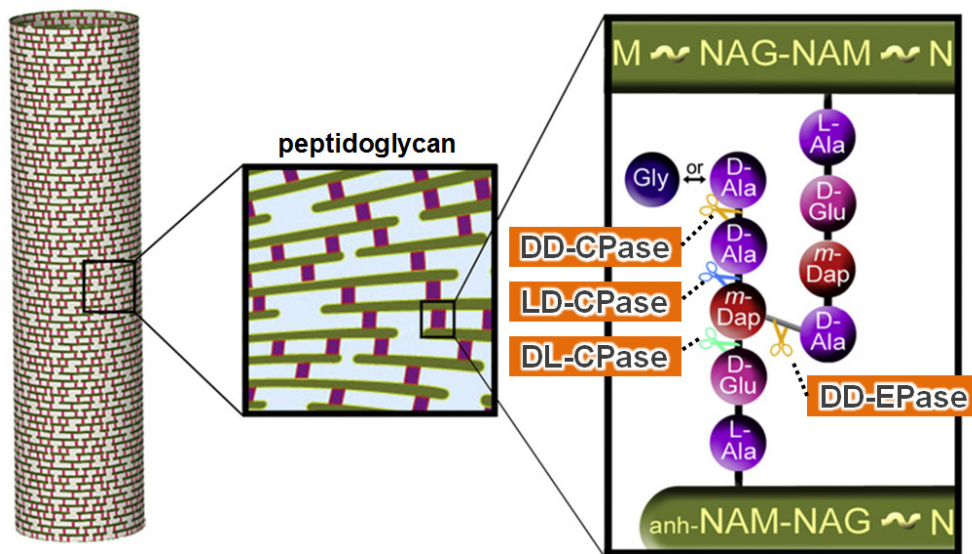


Figure 1-2. Schematic illustration of peptidoglycan structure and peptidoglycan-modifying hydrolases (modified from Sycuro et al., 2010)

Peptidoglycan mesh structure is composed of glycan strands (green) interconnected by peptide cross-links (purple). Peptide bonds marked with orange scissors are hydrolyzed by D,D-endopeptidases (DD-EPases) or D,D-carboxypeptidases (DD-CPases), those marked with blue scissors are hydrolyzed by L,D-carboxypeptidases (LD-CPases), and those marked by green scissors are hydrolyzed by D,L-carboxypeptidases (DL-CPases).

A bioinformatics analysis on the basis of the amino acid sequence predicts that residues 67–200 of *H. pylori* Csd6 form a YkuD domain (formerly called ErfK/YbiS/YcfS/YnhG; ‘L,D-transpeptidase’ catalytic domain; Pfam 03734) that can possibly catalyze non-classical 3→3 cross-linking of peptidoglycan. YkuD domain-containing L,D-transpeptidases (LD-TPases) have been identified in a range of bacteria, including *Enterococcus faecium*, *Bacillus subtilis*, and *Mycobacterium tuberculosis* (Lertsethtakarn et al., 2011; Frirdich et al., 2014; Biarrotte-Sorin et al., 2006; Kim et al., 2013). They generate 3→3 cross-linkages of peptidoglycan, instead of 4→3 cross-linking catalyzed by classical D,D-transpeptidases, resulting in high-level resistance to β -lactam antibiotics (Mainardi et al., 2000) or in peptidoglycan remodeling for dormancy in *M. tuberculosis* (Lavollay et al., 2008). However, the *Helicobacter* peptidoglycan layer is cross-linked exclusively by 4→3 linkages (Costa et al., 1999; Sycuro et al., 2010; Chaput et al., 2007). Consistent with the absence of 3→3 cross-linked muropeptides in the *H. pylori* peptidoglycan sacculus, Csd6 was shown to exhibit the LD-CPase activity only with no “transpeptidase” cross-linking activity against the muropeptides (Sycuro et al., 2013). However, it still remains unanswered why Csd6 has no “transpeptidase” cross-linking activity despite

the presence of a putative ‘LD-TPase’ domain and conservation of the catalytic Cys/His residues.

In many bacterial pathogens, flagella are best known for conferring motility and virulence, as well as serving as an export apparatus for virulence factors (Young et al., 1999) and sensing the viscosity of a medium (McCarter and Silverman, 1990). Flagellar filaments of *H. pylori* are composed of two copolymerized flagellins (FlaA and FlaB) (Kostrzynska et al., 1991), which are heavily *O*-glycosylated with pseudaminic acid (Pse5Ac7Ac; Pse), a sialic acid-resembling sugar (Schirm et al., 2003; Logan, 2006). The regulation of such post-translational modification is critical for the assembly of functional flagella (Schirm et al., 2003; Josenhans et al., 2002). Interestingly, a previous work highlighted the unique role of *H. pylori* G27 Csd6 that involves deglycosylation of FlaA; the Δ csd6 mutant exhibited altered motility to facilitate host-cell interaction and superior colonization (Asakura et al., 2010). It suggests that Csd6 binds Pse molecules or deglycosylates *O*-glycosylated FlaA proteins.

To better understand the molecular function of *H. pylori* Csd6, its biochemical, biophysical, and structural characterizations have been performed in this study. Mass analyses using the synthetic muramyl peptides indicate that Csd6 functions only as the L,D-carboxypeptidase (LD-

CPase) and not as the L,D-transpeptidase (LD-TPase). Analytical ultracentrifugation and single-molecule fluorescence resonance energy transfer (FRET) analyses indicate that Csd6 exists as a dimer in solution. A Csd6 monomer has a three-domain architecture consisting of the N-terminal domain (NTD; Val13–Asn56), the middle LD-CPase domain (Lys67–Glu202), and the C-terminal nuclear transport factor 2-like domain (NTF2-like domain; Thr209–Lys330). The NTD shows very remote structural similarity to other known protein structures and plays a key role in dimerization. The middle catalytic domain has an overall fold of the LD-TPase domain, validating the bioinformatics prediction. However, here this catalytic domain is referred to as the ‘LD-CPase’ domain, because I as well as others show that Csd6 functions only as an LD-CPase. The Csd6 LD-CPase domain resembles those of well-characterized LD-TPases but its pocket-shaped active-site is uniquely defined by the four loops I–IV, among which the loops I and III show the most distinct variations in sequence length and conformation from known LD-TPases. The D-Ala-complexed structure suggests possible binding modes of both the substrate and product to the LD-CPase domain. The NTF2-like domain possesses a deep pocket for possible binding of a hydrophobic ligand such as Pse and an *in silico* docking study supports its role in the control of the glycosylation level of

flagellin. This work provides further insights into the strategy of *H. pylori* for regulating its helical cell shape and motility, which are crucial for its virulence. The reported structural information would serve as the foundation in a search for new drug targets to fight infections by *H. pylori*.

1.2. Material and methods

1.2.1. Cloning, expression, and purification

The *csd6* gene (*hp0518*) from *H. pylori* 26695 strain encoding the N-terminally truncated form (residues 13–330) was amplified by the polymerase chain reaction (PCR) using the genomic DNA as the template. The forward and reverse oligonucleotide used were 5'-G GAA TTC CAT ATG GTG GGA TTG AAT GCT AGT GAT CGT-3' and 5'-CCG CTC GAG TTT TTC CAT TAT AAT AGA CAC TTG ATT G-3', respectively. The underlined sequences are NdeI and XhoI restriction enzyme sites, respectively. The gene was cloned into the expression vector pET-28b(+) (Novagen) to express the recombinant Csd6 protein fused with a His₆-containing tag at its N-terminus. It was overexpressed in *Escherichia coli* Rosetta2(DE3)pLysS cells using a Luria Broth culture medium. Protein expression was induced by 0.5 mM isopropyl β -D-thiogalactopyranoside and the cells were incubated for additional 20 h at 18°C following growth to mid-log phase at 37°C. The cells were lysed by sonication in buffer A (50 mM Tris-HCl at pH 7.9, 500 mM sodium chloride, and 50 mM imidazole) containing 10% (v/v) glycerol and 1 mM phenylmethylsulfonyl fluoride. The crude lysate was centrifuged at 36,000 g for 1 h. The recombinant Csd6

protein showed a tendency to aggregate at 4°C and, therefore, it was purified at room temperature in two column chromatography steps. The supernatant was applied to an affinity chromatography column of HiTrap Chelating HP (GE Healthcare), which was previously equilibrated with buffer A. Upon eluting with a gradient of imidazole in the same buffer, the Csd6 protein was eluted at 120–150 mM imidazole concentration. The eluted protein was applied to a HiLoad XK-16 Superdex 200TM column (GE Healthcare), which was previously equilibrated with 20 mM Tris-HCl at pH 7.9 and 150 mM sodium chloride (Figure 1-3A). Fractions containing the Csd6 protein were pooled and concentrated to 5.8 mg ml⁻¹ for crystallization using an YM10 ultrafiltration membrane (Amicon). This construct gave the best crystals in crystallization experiments. Other constructs were also tried for expression and crystallization. The Val13–Lys330 construct fused with a C-terminal His₆-containing tag was not expressed in *E. coli* Rosetta2(DE3)pLysS cells. The constructs covering residues 4–330 or 17–330 were expressed at very low levels and were insoluble. The construct covering residues 1–330 (full-length) was expressed in a soluble form but its crystals diffracted poorly to 7 Å only.

As a positive control in the LD-TPase assay, Ldt_{Mt2}, an LD-TPase encoded by the *rv2518c* gene of *M. tuberculosis* H37Rv strain, was used. The N-terminal region (Leu20–Ala42) of Ldt_{Mt2} is predicted to form a putative transmembrane helix; the gene covering residues Ala55–Ala408, fused with a His₆-containing tag at its C-terminus, was cloned into pET-21a(+) (Novagen). The recombinant Ldt_{Mt2} was overexpressed and purified essentially as above.

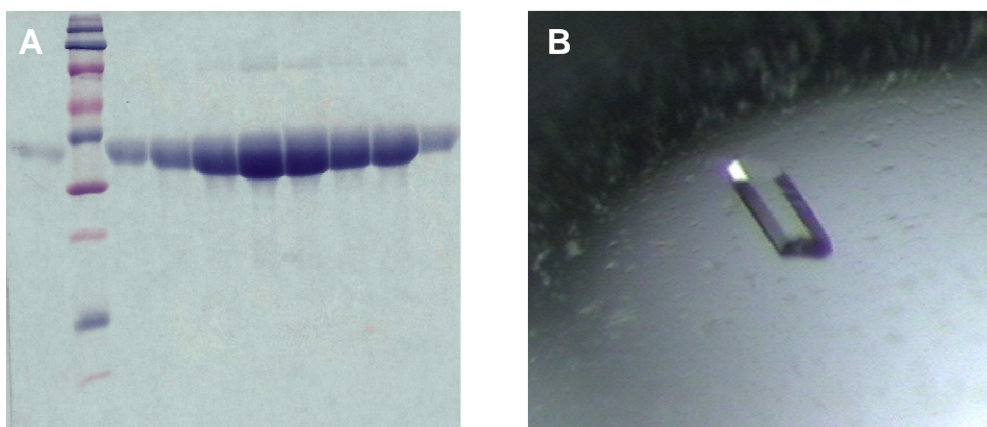


Figure 1-3. SDS-PAGE analysis and crystals of Csd6 from *H. pylori*

(A) SDS-PAGE analysis of Csd6 purified by size exclusion chromatography.

(B) A crystal of Csd6 with approximate dimensions of 0.2 mm × 0.1 mm × 0.05 mm after optimization.

1.2.2. Crystallization and X-ray data collection

To solve the phase problem by anomalous diffraction, ethyl mercury thiosalicylate (EMTS)-derivative crystals of the recombinant Csd6 (residues 13–330) were obtained at 23°C by co-crystallization in the presence of EMTS by the sitting-drop vapor diffusion method. Sitting drops were prepared by mixing 3 μ l of the protein solution and 2 μ l of the reservoir solution [250 mM potassium nitrate, 18% (w/v) polyethylene glycol (PEG) 3,350, and 10 mM EMTS]. EMTS-derivatized crystals grew to approximate dimensions of 0.1 mm \times 0.1 mm \times 0.05 mm within a few days. They were transferred to a cryoprotectant solution, which contained 25% (v/v) glycerol in the reservoir solution. Single-wavelength anomalous diffraction (SAD) data were collected at 100 K at the BL-4A experimental station of Pohang Light Source, Korea. The raw data were processed using the program suite HKL2000 (Otwinowski and Minor, 1997). The crystal of EMTS-derivatized Csd6 belongs to the space group $P2_12_12_1$, with unit cell parameters of $a = 61.8 \text{ \AA}$, $b = 89.6 \text{ \AA}$, $c = 128.7 \text{ \AA}$, and $\alpha = \beta = \gamma = 90^\circ$ (Table 1-1).

Native crystals of ligand-free Csd6 (Csd6-unbound) were grown at 23°C by the sitting-drop vapor diffusion method. Sitting drops were prepared by mixing 0.5 μ l of the protein solution, 0.1 μ l of a ten-fold diluted

microseed crystal solution, and 0.4 μl of the reservoir solution [1.5% (v/v) Tacsimate at pH 7.0, 22% (w/v) PEG 3,350, and 100 mM sodium-HEPES at pH 7.5]. The microseed crystals were originally obtained with Csd6-unbound using the above reservoir solution. The crystals grew to approximate dimensions of 0.2 mm \times 0.1 mm \times 0.05 mm within a few days (Figure 1-3B). They were transferred to a cryoprotectant solution, which contained 20% (v/v) glycerol in the reservoir solution. X-ray diffraction data for Csd6-unbound were collected at 100 K at the BL-1A experimental station of Photon Factory, Japan. The crystal of Csd6-unbound belongs to the space group $P2_12_12_1$, with unit cell parameters of $a = 62.9 \text{ \AA}$, $b = 91.0 \text{ \AA}$, $c = 127.7 \text{ \AA}$, and $\alpha = \beta = \gamma = 90^\circ$ (Table 1-1). To collect the D-Ala-bound Csd6 (Csd6-Ala) diffraction data, a native crystal was pre-incubated for 2 min in a cryoprotectant solution containing both 20% (v/v) glycerol and 96 mM D-Ala, which were added to the reservoir solution. X-ray diffraction data for the Csd6-Ala were collected at 100 K at the BL-5C experimental station of Pohang Light Source, Korea. The crystal of Csd6-Ala belongs to the space group $P2_12_12_1$, with unit cell parameters of $a = 63.4 \text{ \AA}$, $b = 90.6 \text{ \AA}$, $c = 127.9 \text{ \AA}$, and $\alpha = \beta = \gamma = 90^\circ$ (Table 1-1).

Table 1-1. Statistics for data collection, phasing, and model refinement

A. Data collection			
Data set	EMTS-peak	Csd6-unbound	Csd6-Ala
Beamline source ^a	PLS BL-4A	PF BL-1A	PLS BL-5C
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
<i>a</i> , <i>b</i> , <i>c</i> (Å)	61.8, 89.6, 128.7	63.0, 91.0, 127.8	63.4, 90.6, 127.9
X-ray wavelength (Å)	1.0048	1.0000	0.9796
Resolution range ^b (Å)	50.0 – 2.90 (2.95 – 2.90)	50.0 – 2.03 (2.07 – 2.03)	50.0 – 2.04 (2.08 – 2.04)
No. total reflections ^b	373,689 (18,354)	229,193 (11,715)	227,623 (11,074)
No. unique reflections ^b	30,311 (1,515) ^c	47,555 (2,343)	47,575 (2,307)
Completeness (%) ^b	99.9 (100) ^c	98.3 (98.1)	99.9 (100)
$\langle I \rangle / \langle \sigma_I \rangle$ ^b	35.0 (4.7) ^c	25.0 (3.0)	29.3 (3.4)
Wilson <i>B</i> factor (Å ²)	53.1	34.2	30.5
<i>R</i> _{merge} ^{b,d} (%)	11.0 (73.1) ^c	9.5 (47.0)	8.6 (53.7)
B. SAD phasing	Figure of merit (before / after density modification)		0.34 / 0.66
C. Model refinement			
PDB ID code		4XZZ	4Y4V
Resolution range (Å)		30.0 – 2.03	30.0 – 2.04
<i>R</i> _{work} / <i>R</i> _{free} ^e (%)		19.6 / 24.8	19.2 / 24.4
No. of non-hydrogen atoms / average <i>B</i> -factor (Å ²)			
Total		5,589 / 40.6	5,631 / 37.7
Protein		5,248 / 40.4	5,249 / 37.3
Water oxygen		323 / 41.5	334 / 41.4
Glycerol		18 / 63.4	30 / 58.2
D-Ala		- / -	18 / 44.6
R.m.s. deviations from ideal geometry			
Bond lengths (Å) / angles (°)		0.009 / 1.33	0.009 / 1.33
R.m.s. <i>Z</i> -scores			
Bond lengths (Å) / angles (°)		0.46 / 0.61	0.46 / 0.61
Ramachandran (%) ^f			
Favored / Outliers		98.6 / 0.0	97.6 / 0.00
Poor rotamers (%) ^f		0.00	0.00

Footnotes to Table 1-1

^a PF and PLS stand for Photon Factory, Japan and Pohang Light Source, Korea, respectively.

^b Values in parentheses refer to the highest resolution shell.

^c Friedel pairs were treated as separate observations.

^d $R_{\text{merge}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i I(h)_i}$, where $I(h)$ is the intensity of reflection h , \sum_h is the sum over all reflections, and \sum_i is the sum over i measurements of reflection h .

^e $R_{\text{work}} = \frac{\sum (|F_{\text{obs}}| - |F_{\text{calc}}|)}{\sum |F_{\text{obs}}|}$, where R_{free} is calculated for a randomly chosen 5% of reflections, which were not used for structure refinement and R_{work} is calculated for the remaining reflections.

^f Values obtained using MolProbity.

1.2.3. Structure determination and refinement

Two mercury sites in two monomers of EMTS-derivatized Csd6 in the crystallographic asymmetric unit were located using the AutoSol program of the PHENIX software package (Adams et al., 2010). The initial phases were further improved by density modification using the automatic model building program Resolve (Terwilliger, 2003). The initial model was improved through iterative cycles of model building with Coot (Emsley et al., 2010) and refinement with Refmac5 of the CCP4 program suite (Murshudov et al., 1997; Winn et al., 2011). The structures of both Csd6-unbound and Csd6-Ala were determined by molecular replacement with the program MolRep (Vagin and Teplyakov, 2010) using the refined model of EMTS-derivatized Csd6. Stereochemistry of the refined models was evaluated using MolProbity (Chen et al., 2010). Data collection and refinement statistics are given in Table 1-1.

1.2.4. *In silico* docking

Induced Fit Docking (IFD) (Schrödinger Suite Induced Fit Docking protocol; Glide version 5.7., Prime version 3.0, Schrödinger, LLC, New York, NY, USA) (Yoo et al., 2013; Medina-Franco et al., 2014) was employed to predict the binding mode of Pse with consideration of protein flexibility. The Csd6 structure (monomer A model) was prepared for docking calculations using the Protein Preparation Wizard implemented in Maestro (version 9.2, Schrödinger, LLC, New York, NY, USA). The structure of Pse was built in Maestro 9.2, and the possible conformations of the ligand were generated using LigPrep (version 2.5, Schrödinger, LLC, New York, NY, USA). Pse was docked onto Csd6 using the following steps. (i) The receptor grid of Csd6 was defined as an enclosing box at the centroid of the key amino acid residues (i.e., Trp227, Lys258, Tyr297, Lys313, and Glu329) in the binding pocket. (ii) In the initial Glide docking stage, a soften potential docking with van der Waals radii scaling of 0.5 for the protein and ligand was performed retaining a maximum number of 20 poses per ligand. (iii) Residues within 5.0 Å of ligand poses were kept free to move in the Prime refinement step, and the side chains were further optimized. (iv) Poses within 30 kcal mol⁻¹ of the energy cutoff in the previous step were re-docked using Glide XP. (v) The binding energy

(IFDScore) for each output pose was computed as implemented in the IFD protocol. The best docked poses were also reproduced by a different docking program, Surflex-Dock of Sybyl X 2.0 (Tripos Int., St. Louis, MO, USA). All the computations were undertaken on an Intel® Xeon™ Quad-core 2.5 GHz workstation with Linux Cent OS release 5.5.

As a positive control of the docking method, IFD of sialic acid (*N*-acetylneuraminic acid; Neu5Ac) and Pse onto *Micromonospora viridifaciens* sialidase (PDB ID: 1EUS) (Gaskell et al., 1995) and *Pseudomonas aeruginosa* pseudaminidase (PDB ID: 2W38) (Xu et al., 2009) was performed. The docked conformation of sialic acid showed a root-mean-square (r.m.s.) deviation of 1.5 Å from the co-crystallized Neu5Ac2en (2-deoxy-2,3-dideoxy-*N*-acetylneuraminic acid; a sialic acid mimic) in *M. viridifaciens* sialidase (Gaskell et al., 1995). The binding mode of Pse in *P. aeruginosa* pseudaminidase predicted by the present docking calculation was very similar to that of the reported docking result (Xu et al., 2009). These results validate the present docking method.

1.2.5. Peptidase assay by mass analysis

As potential substrates, synthetic muramyl tetrapeptide (β -methyl *N*-acetylmuramic acid-L-Ala¹- γ -D-Glu²-*m*DAP³-D-Ala⁴) and muramyl pentapeptide (β -methyl *N*-acetylmuramic acid-L-Ala¹- γ -D-Glu²-*m*DAP³-D-Ala⁴-D-Ala⁵) were prepared as described previously (Lee et al., 2009) (Figure 1-4, A and B). The analyte, muramyl tetrapeptide (or muramyl pentapeptide) at 5 mM, was incubated for 2 h at 37°C with the recombinant Csd6 protein (5 μ M), which was dissolved in 20 mM sodium phosphate at pH 6.0 and 150 mM sodium chloride. A sample solution (1 μ l) was mixed on the target with a fresh saturated matrix solution of 2,5-dihydroxy benzoic acid dissolved in 0.1% (v/v) trifluoroacetic acid and 50% (v/v) acetonitrile. For sample deposition, a 384-position stainless-steel sample plate was used. Mass spectra were acquired on a matrix-assisted laser desorption/ionization quadrupole ion trap time-of-flight mass spectrometer (MALDI-QIT-TOF MS, AXIMA QIT; Shimadzu/Kratos, Manchester, UK) equipped with a nitrogen laser (337 nm, 3 ns pulse width, maximum pulse rate 10 Hz). Mass spectra were obtained in a positive ion mode. Helium was used for trapping and cooling ions in the ion source. The pressure in the trap was held at 4×10^{-3} torr. Each spectrum constituted an average of 200 profiles. All

spectra were externally calibrated with bradykinin (757.3992 Da), angiotensin II (1046.5418 Da), angiotensin I (1296.6848 Da), Glu-fibrinopeptide B (1570.6768 Da), and *N*-acetyl renin substrate (1800.9432 Da) in TOFMixTM (Shimadzu, Japan). Acquisition and data processing were controlled by the *Lanchpad*TM software. The *m/z* values of multiple sodium adduct ions from the substrates and products are shown in Figure 1-4, A–E. As a positive control, the peaks were detected corresponding to the dimeric crosslinked species between the muramyl tetrapeptide and muramyl tripeptide (Figure 1-4C), a product of the LD-TPase reaction catalyzed by the recombinant Ldt_{M12} (Ala55–Ala408). The LD-TPase domain (residues Asp251–Val378) of *M. tuberculosis* Ldt_{M12} shows sequence identity and similarity of 22% and 38%, respectively, with the Csd6 LD-CPase domain (residues Lys67–Glu202).

To evaluate the importance of specific residues in the LD-CPase activity of Csd6, each of eight residues in the active-site was mutated into alanine: E110A, Y132A, Y133A, H155A, W158A, H160A, and C176A. The mutations were verified by DNA sequencing. The wild-type and these single mutants of Csd6 were incubated with the substrate muramyl tetrapeptide for 30 min at 23°C. Mass spectra were acquired as described above.

1.2.6. Analytical ultracentrifugation

To determine the oligomeric state of the recombinant Csd6 in solution, equilibrium sedimentation and sedimentation velocity experiments were carried out using a Beckman ProteomeLab XL-A analytical ultracentrifuge in 20 mM Tris-HCl buffer at pH 7.9 containing 150 mM sodium chloride and 1.5 mM tris(2-carboxyethyl)phosphine at 20°C. For the equilibrium sedimentation experiment, the absorbance from the Csd6 samples was measured at 235 and 280 nm using a six-sector cell at two speeds (16,000 and 20,000 rpm) and at three different Csd6 monomer concentrations (2.74, 3.65, and 4.56 μM) with a loading volume of 135 μl . For the sedimentation velocity experiment, the Csd6 samples at two different monomer concentrations (0.40 and 5.00 μM) were measured in double-sector cells at 30,000 rpm and the collected data were analyzed by SEDFIT and SEDPHAT programs available from www.analyticalultracentrifugation.com. The concentration of the recombinant Csd6 protein was calculated using $\epsilon_{280\text{nm}} = 54,780 \text{ M}^{-1} \text{ cm}^{-1}$.

1.2.7. Single-molecule fluorescence resonance energy transfer (FRET)

To analyze the dimeric form of Csd6 in solution, a single-molecule FRET technique incorporating alternating-laser excitation (ALEX) was

employed (Kim et al., 2012; Lee et al., 2005). Briefly, in this method the fluorescence signal of a dimeric molecule is analyzed one-by-one and the result is presented in a two-dimensional E-S graph (Figure 1-6D), where E is the FRET efficiency and S denotes the Cy3/Cy5 molar ratio in a dimer. The Csd6 sample was divided into two aliquots, each of which was labeled on the sole cysteine residue (Cys176) with the maleimide-reactive Cy3 dye (donor) or the Cy5 dye (acceptor). Subsequently, Cy3-labeled Csd6 (Cy3-Csd6) and Cy5-labeled Csd6 (Cy5-Csd6) were mixed at 2 μ M final monomer concentration in a buffer (20 mM Tris-HCl at pH 7.9 and 150 mM sodium chloride) for 3 h at 37°C. The protein sample was diluted further to 100 pM using a single-molecule buffer [20 mM Tris-HCl at pH 7.9, 150 mM sodium chloride, 5% (v/v) glycerol, 1 mM 2-mercaptoethylamine, and 0.01% (v/v) bovine serum albumin]. The single-molecule data were acquired from a 10 min measurement.

1.2.8. Surface plasmon resonance experiment

The kinetics and affinity of Csd6 with the synthetic muramyl tetrapeptide and muramyl tripeptide (β -methyl *N*-acetylmuramic acid-L-Ala¹- γ -D-Glu²-*m*DAP³) as a reaction product was assessed using a Reichert SR7500 Surface Plasmon Resonance (SPR) dual channel instrument (Reichert, Depew, NY). Purified Csd6 in 20 mM sodium acetate at pH 5.5 was immobilized using the standard amino coupling at 20 $\mu\text{l min}^{-1}$ on a carboxymethyl dextran hydrogel surface sensor chip (Reichert, Depew, NY) until saturation was achieved. The running buffer B used in all SPR experiments was 20 mM HEPES at pH 7.9 and 150 mM sodium chloride. SPR experiments were performed at 25°C. The muramyl tetrapeptide or muramyl tripeptide at concentrations of 15.6, 31.3, 62.5, 125, 250, and 500 μM was injected over the Csd6-chip at 30 $\mu\text{l min}^{-1}$ for 5 min for association analyses. Subsequently, the running buffer was flowed over the chip for additional 6 min (30 $\mu\text{l min}^{-1}$) for dissociation analyses. Regeneration of the chip was carried out using 20 mM sodium hydroxide. Binding was detected as a change in the refractive index at the surface of the chip as measured by the response unit (RU). A reference flow cell was used to record the response by bovine serum albumin (BSA) as a positive control, and the

response by BSA was subtracted from each sample. SPR data were fit using the Scrubber2 software.

1.3. Results

1.3.1. Csd6 functions as an LD-CPase but not as an LD-TPase

To verify the molecular function of *H. pylori* Csd6 in peptidoglycan modification, the peptidase activity of Csd6 was measured *in vitro* with the synthetic muramyl tetrapeptide and muramyl pentapeptide (Figure 1-4, A and B) as potential substrates using mass analyses. When the muramyl pentapeptide was incubated with the recombinant Csd6 (13–330), no reaction product could be detected (Figure 1-4E). This suggests that Csd6 does not have ‘endo’-, ‘DD-carboxy’- or ‘LD-trans’-peptidase activities with the muramyl pentapeptide. The LD-CPase domain of Csd6 shows limited but significant sequence identity (22%) with the LD-TPase domain of *M. tuberculosis* Ldt_{M12}. *M. tuberculosis* Ldt_{M12} is a functional LD-TPase, catalyzing the formation of 3→3 cross-links between the muramyl tetrapeptide and the muramyl tripeptide (Cordillot et al., 2013). Therefore, as a positive control for measuring the LD-TPase activity, the muramyl tetrapeptide was incubated with the recombinant *M. tuberculosis* Ldt_{M12}. In this control reaction, not only the muramyl tripeptide (‘Tri’ in Figure 1-4C) but also the dimeric cross-linked species (‘Tetra-Tri’ in Figure 1-4C) were

detected as the reaction products. In comparison, when the muramyl tetrapeptide was incubated with the recombinant Csd6, the muramyl tripeptide was detected as the sole reaction product, without any dimeric cross-linked species (Figure 1-4D). This result implies that Csd6 would not be active for cross-linking, if the muramyl tripeptide were used as a substrate. The present data, together with the previous report that Csd6 converts monomeric tetrapeptides of the peptidoglycan sacculus into tripeptides (Sycuro et al., 2013), clearly establish that *H. pylori* Csd6 functions as an LD-‘carboxypeptidase’ devoid of a cross-linking LD-TPase despite the presence of the conserved Cys/His residues, which are characteristics of the ‘LD-TPase’ domains. The active-site features of Csd6, as discussed in more detail below, explain why Csd6 is inactive as LD-TPase.

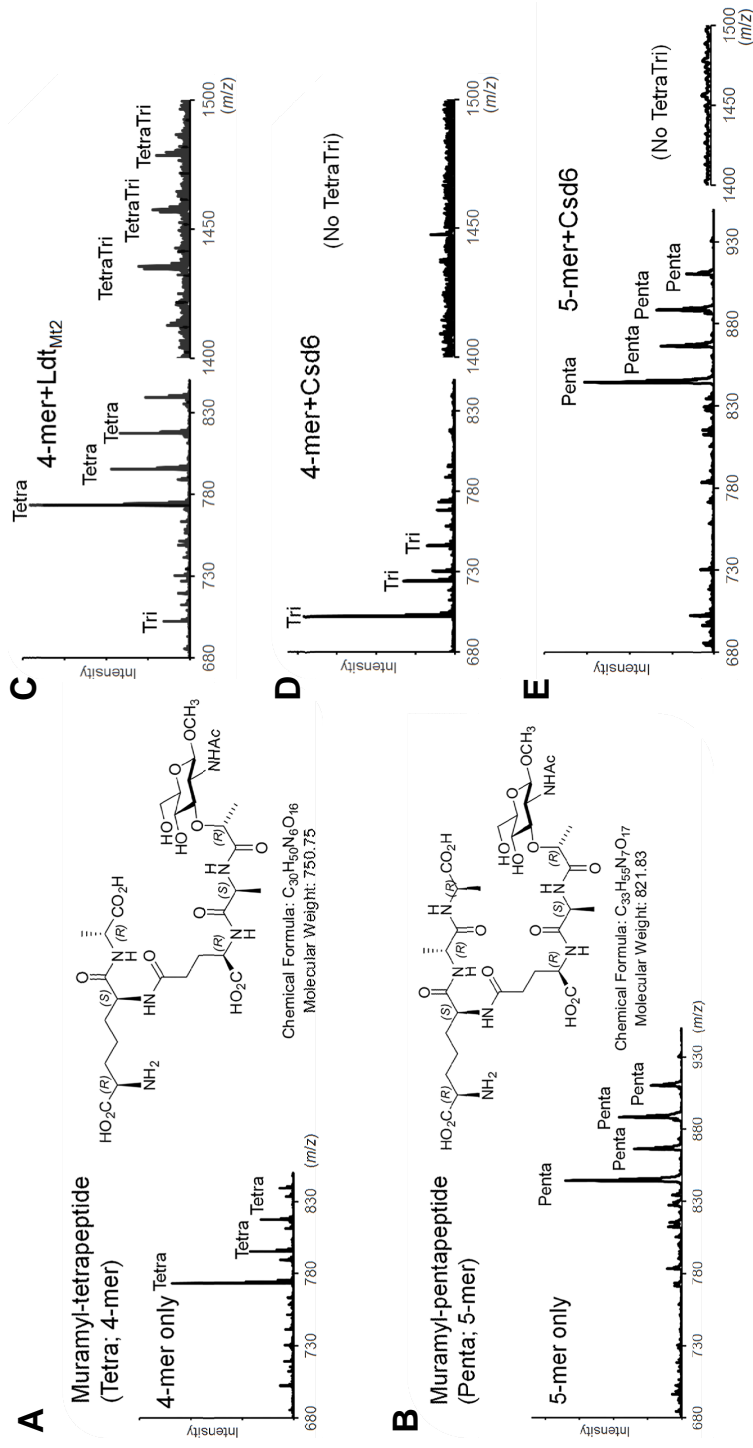


Figure 1-4. LD-CPase activity of Csd6

(A and B) Mass spectra and structural formulas of the synthesized muramyl tetrapeptide (A) and muramyl pentapeptide (B). The observed peaks correspond to a series of sodium adduct ions of the peptides. (C–E) Mass spectra of the peptide samples upon incubation with Ldt_{M12} and Csd6. In the control reaction catalyzed by Ldt_{M12}, the dimeric cross-linked tetra-tripeptide species is produced from the muramyl-tetrapeptide (C). In contrast, in the reaction catalyzed by Csd6, the muramyl-tetrapeptide is converted to the muramyl-tripeptide but not to dimeric cross-linked tetra-tripeptide species (D). Upon incubation with Csd6, the muramyl-pentapeptide undergoes no change (E). The observed m/z values for the charged species with bound sodium ions (up to four: [M+Na]⁺¹, [M+2Na-H]⁺¹, [M+3Na-2H]⁺¹, and [M+4Na-3H]⁺¹) agree with the calculated m/z values for muramyl-tri, muramyl-tetra, muramyl-penta, or dimeric cross-linked tetra-tripeptides (average mass of neutral species = 679.67, 750.75, 821.83, or 1412.41, respectively).

1.3.2. The Csd6 monomer is organized into three-domain architecture

To provide a structural basis of understanding the molecular function, the crystal structures of Csd6 have been determined, both in the ligand-free state ('Csd6-unbound') (Figure 1-5) and in the D-Ala-bound state ('Csd6-Ala') (Table 1). Refined models of both structures are essentially identical to each other with an r.m.s. deviation of 0.17 Å for 632 equivalent C α positions. They account for residues Met15–Lys330 in each of two Csd6 monomers in an asymmetric unit (ASU). In the case of Csd6-unbound structure, two monomers in the ASU are similar to each other with an r.m.s. deviation of 0.64 Å for 316 C α atom pairs, except N-terminal residues (Leu15–Asp41) showing r.m.s. deviations greater than 1.0 Å with a maximum C α deviation (7.6 Å) at Leu15. For the Csd6-Ala model, two monomers in the ASU agree with an r.m.s. deviation of 0.61 Å for 316 C α atom pairs, except N-terminal residues (Leu15–Asp41) showing r.m.s. deviations greater than 1.0 Å with a maximum C α deviation (8.4 Å) at Leu15. The amino-acid residues with the largest deviation encompass the entire helix α A and the N-terminal half of α B.

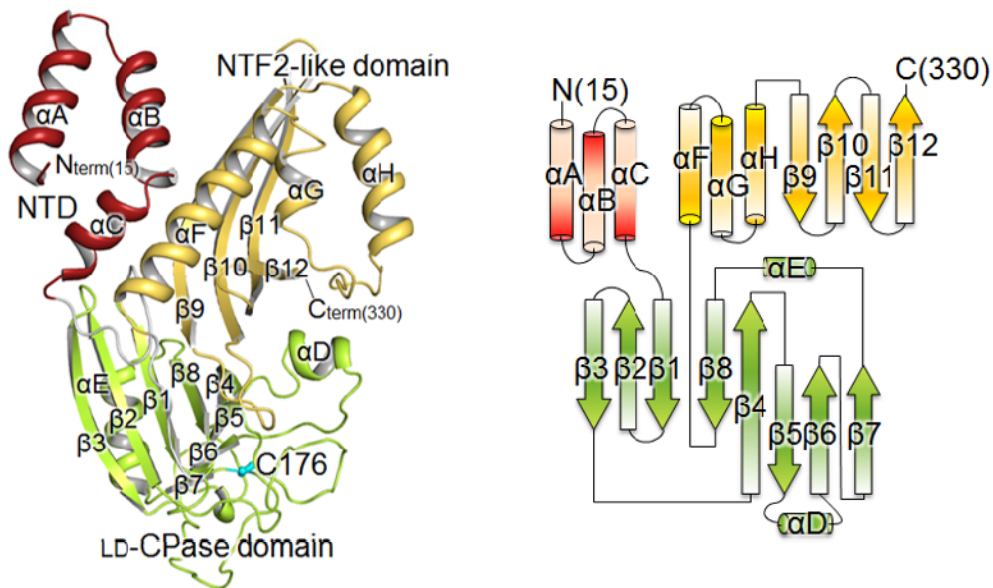


Figure 1-5. Overall structure of *H. pylori* Csd6

The ribbon diagram (left panel) and topology diagram (right panel) of the Csd6-unbound structure. The NTD, LD-CPase domain, and NTF2-like domain are shown in red, green, and yellow, respectively. The secondary structure elements have been defined by the DSSP program (Kabsch and Sander, 1983). The nucleophile Cys176 of the LD-CPase domain is shown in a cyan stick model.

The Csd6 monomer consists of three domains of unequal size: the NTD (Val13–Asn56), the middle LD-CPase domain (Lys67–Glu202), and the C-terminal NTF2-like domain (Thr209–Lys330) (Fig. 1-5). The NTD consists of three alpha helices (αA – αC) and the helix αC is positioned at one end of the antiparallel coiled-coil formed by helices αA and αB . The LD-CPase domain consists of two curved β -sheets ($\beta 3\downarrow$ - $\beta 2\uparrow$ - $\beta 1\downarrow$ - $\beta 8\downarrow$ - $\beta 4\uparrow$ and $\beta 4\uparrow$ - $\beta 5\downarrow$ - $\beta 6\uparrow$ - $\beta 7\uparrow$), which share the $\beta 4$ strand, as well as two α -helices (αD and αE). The NTF2-like domain forms an antiparallel four-stranded β -sheet ($\beta 9\downarrow$ - $\beta 10\uparrow$ - $\beta 11\downarrow$ - $\beta 12\uparrow$), whose concave side is packed by three α -helices (αF – αH). DALI structural similarity searches (Holm and Rosenström, 2010) revealed no significant match with either the entire Csd6 with three domains or any combination of two connected domains in Csd6, suggesting that the three-domain architecture of Csd6 is unique. This is in line with our observation that the organization of the LD-CPase domain and the NTF2-like domain in Csd6 and its homologs is unique among the LD-TPase and LD-CPase families according to the conserved domain searches using the InterPro database (Mitchell et al., 2015).

In the Csd6 monomer structure, three domains are arranged in a tripartite, leaf-like fashion with a single chain connecting two adjacent

domains (Figure 1-5). Virtually no interaction exists between the NTD and the LD-CPase domain; the buried surface area at the interface is only 36 Å², as calculated by the PISA server (Krissinel and Henrick, 2007). Much more extensive interactions exist between the LD-CPase domain and the NTF2-like domain, with a buried surface area of 718 Å². The latter interface is mainly formed between strands β4 and β8 of the LD-CPase domain, and strands β9–β11 of the NTF2-like domain. It is primarily lined with hydrophobic residues of both domains (Val119, Tyr120, Tyr146, Phe197, and Ile199 of the LD-CPase domain and Ile278, Tyr280, Tyr290, Val292, and Tyr316 of the NTF2-like domain). The NTD (helix αC) and the NTF2-like domain (helix αF) interact with each other, burying a surface area of 410 Å² at the interface.

1.3.3. The NTD of Csd6 plays a dominant role in homo-dimerization

Equilibrium sedimentation measurements indicate that the Csd6 protein exists as homogeneous dimers in solution (at the tested Csd6 monomer concentration range of 2.74–4.56 μM). The data fit well to a dimer model and a representative result measured at 16,000 rpm using the 3.65 μM monomer concentration is presented in Figure 1-6A. Further sedimentation velocity experiments at 0.40 and 5.00 μM Csd6 monomer concentrations

also indicate the oligomeric state of Csd6 as a dimer, without any discernable dissociation of a dimer into monomers, down to 0.40 μM (Figure 1-6C). In the crystal of Csd6, two possible models of the dimer could be identified. The dimer model I has a buried surface area of 1,080 \AA^2 per monomer (6.7% of the monomer surface area) at the interface, as analyzed by the PISA server (Krissinel and Henrick, 2007), (Figure 1-6B, the upper panel). In the dimer model I, the two monomers are related by non-crystallographic ‘pseudo’ two-fold symmetry. The two-fold symmetry is broken primarily by the significant difference in the orientation of N-terminal helices αA and αB between the two monomers. The dimer model II has a buried surface area of 726 \AA^2 (4.5% of the monomer surface area) (Figure 1-6B, the lower panel). In the dimer model II, the two monomers are related by an exact crystallographic two-fold symmetry. The $\Delta^i\text{G}$ values (solvation free energy gain upon formation of the interface) are -15.8 and -2.1 kcal mol^{-1} for models I and II, with the Complexation Significance Scores (CSS) of 1.000 and 0.000, respectively. These CSS values imply that the interface in model I, but not in model II, plays an essential role in complex formation. Therefore, I conclude that the dimer model I in the crystal represents the biologically relevant Csd6 dimers in solution. This conclusion experimentally by the distance between two Cys176 residues in

the dimer as estimated using the single-molecule ALEX-FRET technique (Figure 1-6D). The single-molecule ALEX-FRET experiment was performed at 100 pM monomer concentration, which is much lower than those used for analytical ultracentrifugation, to avoid ensemble averaging (i.e., to avoid the case of having two non-interacting species) (Hohlbein et al., 2014). After direct excitation leakage and buffer background corrections, the distance between Cy3 and Cy5 dipoles in the dye-labeled Csd6 dimer is estimated to be 107 Å from the corrected FRET efficiency ($E_{\text{corr}} \approx 0.03$), using R_0 (Förster distance) of 60 Å for the Cy3 and Cy5 pair. In the crystal structure of Csd6-unbound, the distance between sulfur atoms of two Cys176 residues is 96 Å and 66 Å in dimer models I and II, respectively. This result unambiguously favors model I, supporting the assignment of dimer model I in the crystal as the dimer in solution (Figure 1-6B, the upper panel).

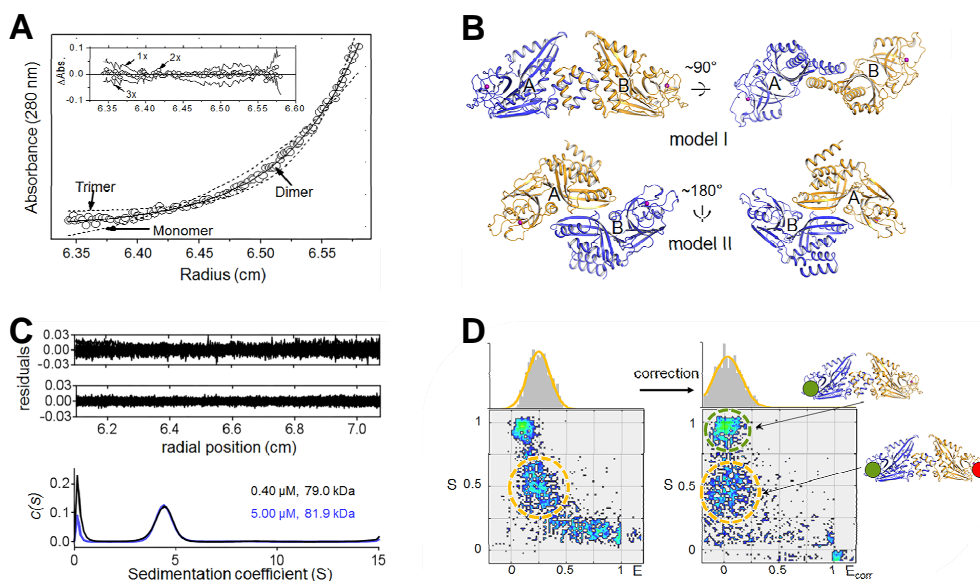


Figure 1-6. Oligomeric state of *H. pylori* Csd6

(A) Equilibrium sedimentation data for Csd6 at an ultracentrifugal speed of 16,000 rpm using 3.65 μ M protein at 20°C. The circles are experimental data and the solid line is a fitting line for an ideal monomer model. The two dotted lines are fitting lines for an ideal monomer and trimer models. Distributions of the residuals for monomer, dimer, and trimer models are shown in the inset panel. These data indicate that Csd6 exists as homogeneous dimers in solution. (B) Two putative models of the Csd6 dimer in the crystal. Dimer model I (upper panel) and model II (bottom panel) are shown in ribbon diagrams. The side chain sulfur atoms of Cys176

are shown as purple spheres. Model I is favored by the single-molecule ALEX-FRET data (shown in Figure 1-6D). (C) Plots of the residuals for dimer species model of Csd6 at 0.40 μM (upper) and 5.00 μM (lower), and the distribution of sedimentation coefficient ($c(s)$ versus s , where s is in Svedberg unit, S) from the sedimentation velocity experiments. (D) Determination of the dimeric model of Csd6 by single-molecule FRET technique. Single-molecule FRET data are presented in a 2-dimensional (2D) E-S graph, where E is the FRET efficiency and S denotes the Cy3 (donor)/Cy5 (acceptor) molar ratio in a dimer (Kim et al., 2012). Three clusters correspond to the dimers formed (i) between two Cy3-Csd6 monomers, (ii) between two Cy5-Csd6 monomers, and (iii) between a Cy3-Csd6 monomer and a Cy5-Csd6 monomer. The cluster appearing at $S \sim 1$ (green dotted ellipse) with $E \approx 0$ corresponds to Cy3-labeled Csd6 dimers. Csd6 dimers labeled with both Cy3 and Cy5 appear at $S \sim 0.5$ (orange dotted ellipse) (Lee et al., 2005).

The Csd6 dimer (i.e., dimer model I) is elongated with approximate dimensions of $100 \text{ \AA} \times 80 \text{ \AA} \times 50 \text{ \AA}$ (Figure 1-7A). In this dimer, two LD-CPase domains are well separated from each other and their active-sites are open toward the bulk solvent (Figure 1-7A). The bulk of the buried surface area at the dimerization interface is contributed by the NTD (82.7%), with the linker between the NTD and the LD-CPase domain contributing 5.7% and the NTF2-like domain contributing 11.6%. The N-terminal helix αA of the NTD takes slightly different conformations in the two monomers, with a maximum $C\alpha$ deviation of 7.6 \AA at Leu15. The two NTDs pack against each other through their antiparallel coiled-coils (helices αA and αB), at an angle of about 60° , to form a four-helix bundle (Figure 1-7A); numerous hydrophobic side chains are present at the interface. The αC -helices from two NTDs cover the sides of the four-helix bundle. Seven hydrogen bonds (involving Arg26, Tyr28, Gln29, Gly37, Asp41, Glu53, Tyr63, Gln65, and Phe272) and a salt bridge (involving Arg26 and Glu53) exist at the interface (Figure 1-7B). Interestingly, the Csd6 dimer has a deep crevice of an approximate size of 20 \AA by 20 \AA , with a widened bottom (indicated by dotted lines in Figure 1-7C). The NTDs provide the bottom of the crevice and the C-terminal NTF2-like domains form the two sides of the crevice.

The sides of the crevice are lined with positively charged residues (Lys267, Lys298, and Lys304) but no negatively charged residues (Figure 1-7C).

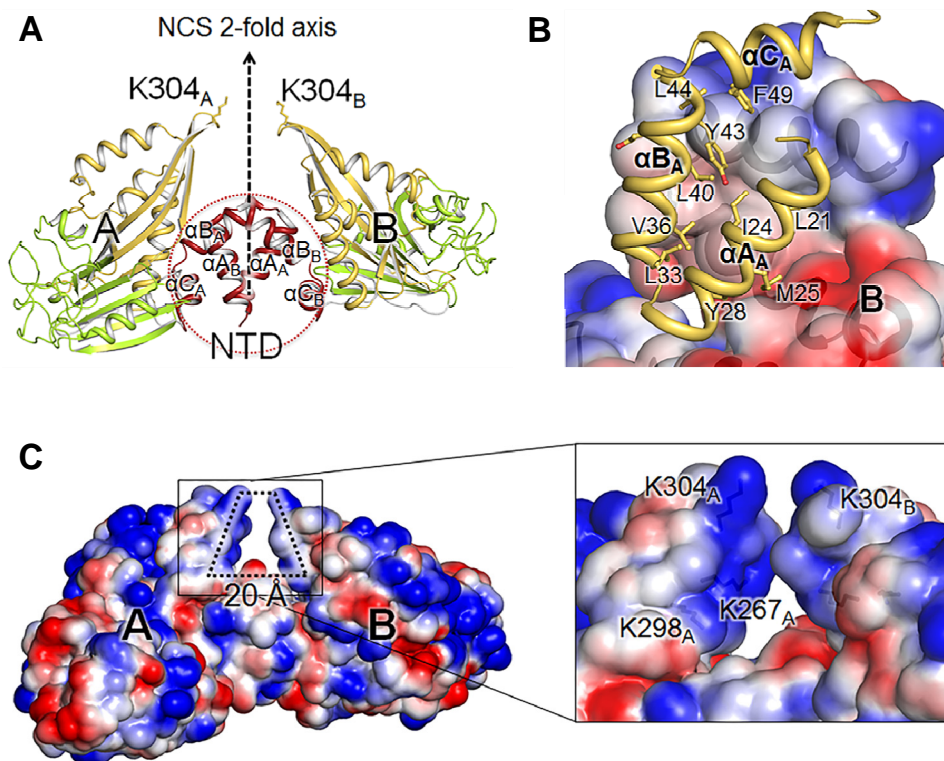


Figure 1-7. Dimeric structure of *H. pylori* Csd6

(A) The ribbon diagram of the Csd6 homodimer (model I) colored as in Figure 1-5. In this dimer, A and B monomers are related by a non-crystallographic pseudo two-fold symmetry. (B) Hydrophobic interactions between the NTDs of A and B monomers, which are shown in the ribbon diagram and the electrostatic surface diagram, respectively. (C) Electrostatic potential surface diagrams of the Csd6 dimer molecule. A detailed view of the specific hole formed by the dimerization is shown in the black lined box.

A DALI search with the Csd6 NTD (Val13–Asn56) alone showed very remote structural similarities to other known protein structures. The highest Z-score was obtained with human phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform (PDB code 4L2Y-A; an r.m.s. deviation of 2.1 Å for 41 equivalent C α positions in residues 16–56, a Z-score of 4.6, and a sequence identity of 5%). The Csd6 NTD is perhaps functionally more related to the L27N domain of *Mus musculus* PALS1-associated tight junction protein (PDB code 1VF6-D; an r.m.s. deviation of 4.2 Å for 41 equivalent C α positions in residues 16–56, a Z-score of 4.3, and a sequence identity of 12%) (Li, et al., 2004). L27 domains, such as PALS1-L27N, have been established as a protein-binding module that brings multiple proteins into complexes for signaling, cell polarity, and epithelial morphogenesis (Doerks et al., 2000; Harris et al., 2002). Like the Csd6 NTD, the L27 domain is composed of three α -helices. However, unlike the Csd6 NTD that plays a key role in homo-dimerization, two different L27 domains (e.g., PALS1-L27N domain and PATJ-L27) hetero-dimerize by building a compact four-helix bundle structure through the first two helices from each L27 domain (Figure 1-8) (Li, et al., 2004).

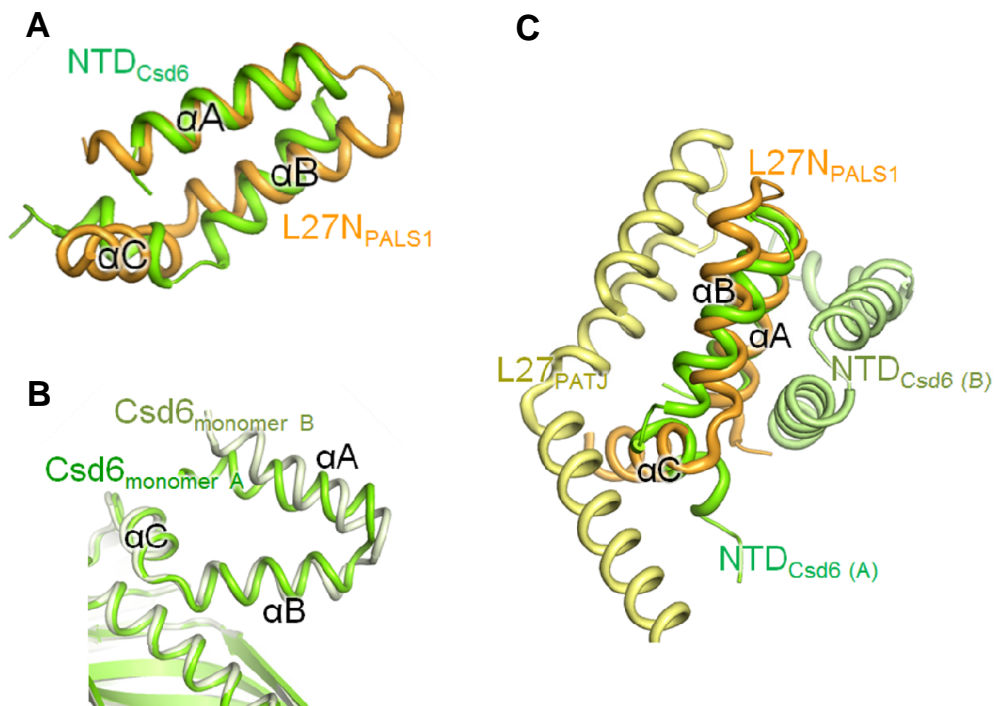


Figure 1-8. Comparisons of the Csd6 NTD with the PALS1-L27N domain

The following pairwise superimpositions are shown in ribbon diagrams: (A) the NTD of the Csd6 monomer A and the PALS1-L27N domain, (B) the NTDs of Csd6 monomers A and B, and (C) the NTDs of the Csd6 dimer and PALS1-L27N domain complexed with PATJ-L27.

1.3.4. Active-site of Csd6 LD-CPase domain is tailored for the LD-CPase activity

A bioinformatics analysis on the basis of amino-acid sequence predicts *H. pylori* Csd6 to have an overall fold of the LD-TPase catalytic domain over residues 67–200. The structure reported here validates this prediction (Figures 1-5 and 1-9A). However, this catalytic domain is referred to as the LD-CPase domain, because Csd6 is shown to function only as an LD-CPase, but not as an LD-TPase against the muramyl tetrapeptide. According to the DALI search, the overall fold of the Csd6 LD-CPase domain (Lys67–Glu202) (Figure 1-9A) resembles those of well-characterized LD-TPases such as *B. subtilis* Ldt_{Bs} (Figure 1-9B) (PDB code 1Y7M; a Z-score of 14.8, an r.m.s. deviation of 1.8 Å, and a sequence identity of 13% for 109 equivalent C α positions) (Bielnicki et al., 2006), *E. faecium* Ldt_{fm} (Figure 1-9C) (PDB code 1ZAT; a Z-score of 12.1, an r.m.s. deviation of 2.7 Å, and a sequence identity of 14% for 111 equivalent C α positions) (Biarrotte-Sorin et al., 2006), and *M. tuberculosis* Ldt_{Mt2} (Figure 1-9D) (PDB code 4GSU; a Z-score of 11.7, an r.m.s. deviation of 2.1 Å, and a sequence identity of 14% for 107 equivalent C α positions) (Kim et al., 2013). Unlike Csd6, all of these LD-TPases are monomeric enzymes (Biarrotte-Sorin et al., 2006; Kim et al., 2013; Bielnicki et al., 2006). Furthermore, the LD-CPase domain of

Csd6 contains the characteristic sequence motif **H(S/G/D)_x14~19C** (Figure 1-9E), where ‘x’ stands for any amino acid and the strictly conserved residues are in boldface. This motif is conserved among proteins containing an LD-TPase catalytic domain and provides the catalytic triad residues, which form a hydrogen-bond network in a wide range of Cys/Ser-based proteolytic enzymes (Dodson and Wlodawer, 1998). The catalytic triad of Csd6 consists of His160, Gly161, and Cys176 (marked by red dots in Figure 1-9E), corresponding to His336, Ser337, and Cys354 of Ldt_{M2} (Kim et al., 2013). However, *H. pylori* Csd6 is unrelated in its sequence and structure to previously-characterized LD-CPases such as L,D-carboxypeptidases A (LdcA) from *Pseudomonas aeruginosa* (Korza and Bochtler, 2005) and *Novosphingobium aromaticivorans* DSM 12444 (Das et al., 2013), and L,D-carboxypeptidases B (LdcB) from *Streptococcus pneumoniae*, *Bacillus anthracis*, and *Bacillus subtilis* (Hoyland et al., 2014). LdcA is a Ser-based LD-CPase. The latter three LdcB proteins belong to the LAS (lysostaphin, D-Ala-D-Ala metallopeptidases, sonic hedgehog) family of Zn²⁺-dependent peptidases (Hoyland et al., 2014). This raises an intriguing question why *H. pylori* Csd6 functions only as an LD-CPase, but not as an LD-TPase, despite the overall ‘LD-TPase’ fold and the presence of the conserved catalytic triad.

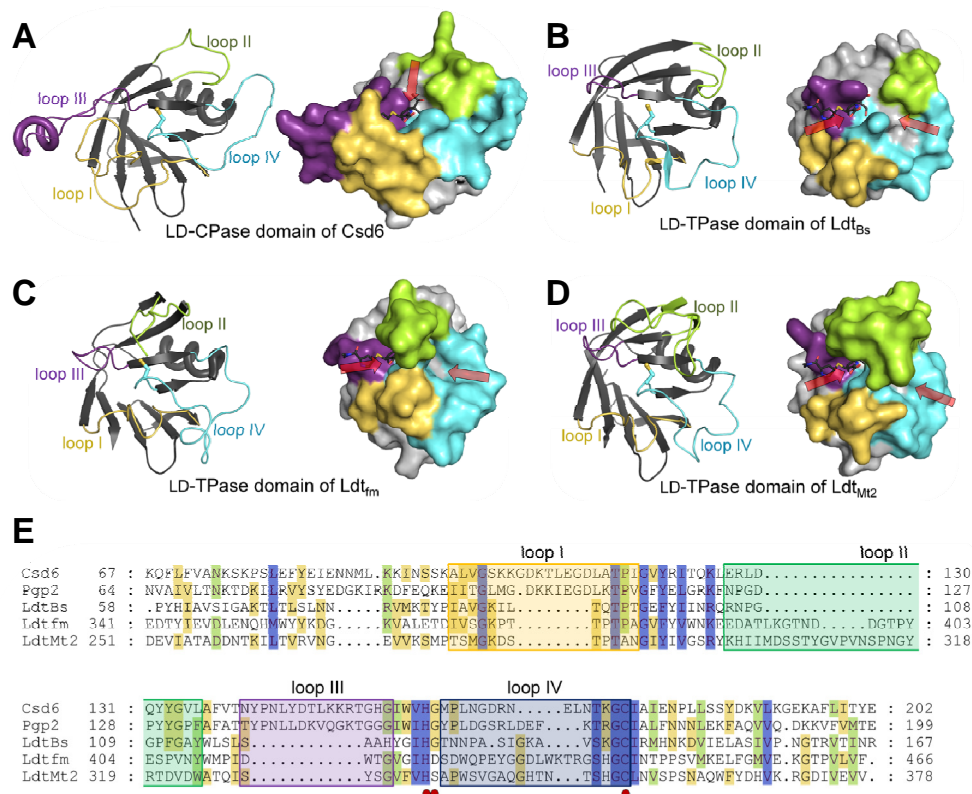


Figure 1-9. Comparison of Csd6 and other LD-TPases

(A–D) the ribbon diagrams (left panels) and surface diagrams (right panels) of the LD-CPase domain of *H. pylori* Csd6 (A) and the LD-TPase domains of *B. subtilis* Ldt_{Bs} (B), *E. faecium* Ldt_{fm} (C), and *M. tuberculosis* Ldt_{Mt2} (D). The four loops I, II, III, and IV are colored in yellow, green, light blue, and purple, respectively. One or two paths that allow access to the catalytic triad are indicated by red arrows. In the meropenem-complexed Ldt_{Mt2} structure (PDB code: 4GSU), the paths A and B correspond to left and right arrows,

respectively (Kim et al., 2013). Meropenem was modeled into each domain structure by structural superimposition of the corresponding domains using the meropenem-complexed Ldt_{Mt2} (PDB code: 4GSU). (E) Structure-based sequence alignment was performed by PROMALS3D (Pei et al., 2008) and the alignment was presented by GeneDoc (<http://www.nrbsc.org/downloads>). Structures of the LD-CPase domain of *H. pylori* Csd6 (SWISS-PROT accession code: O25255; PDB code: 4XZZ) and LD-TPase catalytic domains of *B. subtilis* Ldt_{Bs} (O34816; 1Y7M), *E. faecium* Ldt_{fm} (Q3Y185; 1ZAT), and *M. tuberculosis* Ldt_{Mt2} (O53223; 4GSU) were used in the alignment. In the case of *C. jejuni* Pgp2 (A1VZP0), no structural information is available and thus only the sequence is used. Red dots indicate the catalytic triad (His160, Gly161, and Cys176 in *H. pylori* Csd6). The four loops I, II, III, and IV are shown in colored boxes as in Figure 1-9A.

A structural comparison of Csd6 LD-CPase domain with LD-TPase domains reveals that the shape of the active-site in the Csd6 LD-CPase domain seems to be tailored for the L,D-carboxypeptidation. That is, the active-site is a deep pocket and the catalytic triad positioned at the bottom of this pocket is accessible via a single narrow path (Figure 1-9A), as in typical carboxypeptidases (Covaleda et al., 2012). In contrast, the active-site in the corresponding domains of the above three LD-TPases is an elongated groove and the catalytic triad is accessible via two paths (Figure 1-9, B–D) (Kim et al., 2013). The different shapes of the active-sites are primarily dictated by the length and conformation of the loops around the entrance(s) to the active-sites. Structural comparison and sequence alignment of Csd6 LD-CPase domain with the corresponding domains of three LD-TPases (Figure 4) show that unique features of Csd6 that distinguish it from other LD-TPases reside primarily in four loop regions: loop I (β 3- β 4 loop; Ala98–Ile117), loop II (β 4- β 5 loop; Leu126–Leu136), loop III (β 5- α D- β 6 loop; Asn141–Ile157), and loop IV (β 6- β 7 loop; Met162–Cys176).

The reaction catalyzed by LD-TPases such as Ldt_{M2} occurs in two sequential steps: acylation of the donor substrate (“carboxypeptidation”) and deacylation by the acceptor substrate for cross-linking (“transpeptidation”) (Silva et al., 2014). As a consequence, LD-TPases such as Ldt_{M2} have two

accessible paths for donor and acceptor substrates (indicated by red arrows in Figure 1-9, B–D), thereby allowing an approach of the acceptor substrate upon acylation of the donor substrate by the catalytic Cys residue for transpeptidation. For a better understanding of the substrate-binding site (Figure 1-9, A–D), meropenem (mimicking the substrate) onto Csd6 by superimposing its LD-CPase and the LD-TPase domain of the meropenem-complexed Ldt_{M12} (Kim et al., 2013). The position of the covalently bound meropenem represents the path for the donor substrate (‘path A’). The active-site clefts in three LD-TPases reveal a long groove for binding two substrates, which is contributed by the four loops I–IV (Figure 1-9, B–D). In Ldt_{M12} and Ldt_{fm}, the extended loop II (green colored in Figure 1-9, C and D) forms a lid to cover the active-site cleft, while exposing two discrete paths for the access of two substrates. On the other hand, the loop II of Ldt_{Bs} is much shorter and the active-site cleft is not covered with a lid (green colored in Figure 1-9B). Like Ldt_{Bs}, the loop II is much shorter in Csd6 and the active-site cleft is not covered with a lid. However, in the Csd6 structure, the corresponding paths to the active-site cleft for access of two substrates in three LD-TPases are blocked due to the distinct length and conformations of the loops I–IV and the active-site cleft opens vertically towards the solvent (indicated by a red arrow in Figure 1-9A), allowing access of only

one muramyl peptide for carboxypeptidation. This difference in the shape of the active-site cleft is the likely reason for the enzyme serving as a carboxypeptidase, to the exclusion of the transpeptidase activity.

Among the four loops I–IV, the loops I and III show the most distinct variations in sequence length and conformation. They are considerably longer (eight and twelve residues, respectively) in *H. pylori* Csd6 than in the LD-TPases (Figure 1-9E). Loop III contains the helix α D, which protrudes from the active-site. Together with loop I, it provides one side of the active-site cleft to recognize a substrate. The *C. jejuni* Pgp2, a Csd6 homolog, is predicted to contain an LD-TPase domain on the basis of its amino acid sequence (Figure 1-9E) but it was shown to function as an LD-CPase (Friedrich et al., 2014). I notice that the four loops of Pgp2 show a highly similar pattern with those of Csd6 (Figure 1-9E). Therefore, it is anticipated that the observed unique structural features of these four loops around the active-site cleft are conserved among Csd6 homologs in other ϵ -proteobacteria.

1.3.5. D-Ala-complexed structure reveals binding modes of both the substrate and product

By analogy with the first carboxypeptidation step of Ldt_{M12}, His160 of Csd6 serves as a proton acceptor to assist a nucleophilic attack by Cys176. In present Csd6 structure, the side chains of His160 and Cys176 are suitably predisposed for a cooperative action with the main-chain carbonyl of Gly161 (Figure 1-10A). His160^{N ϵ 2} interacts with Cys176^{S γ} with a distance of 3.96 Å and His160^{N δ 1} interacts with the main-chain oxygen atom of Gly161 with a distance of 2.75 Å (Figure 1-10A). In addition, the LD-CPase domain of Csd6 shows a well-defined oxyanion hole to facilitate catalysis. The backbone amide nitrogen atoms of Lys174, Gly175, and Cys176 (asterisked in Figure 1-10A) could stabilize the transient oxyanion species of the polarized carbonyl bond, as in Ldt_{M12} (Silva et al., 2014).

In order to understand the structural determinants for ligand binding, the structure of Csd6 bound with the product D-Ala ('Csd6-Ala' in Table 1) was determined. In the Csd6-Ala model, two D-Ala molecules are well ensconced within the active-site of monomer A with clear electron density (marked as D-Ala' and D-Ala'' in Figure 1-10A), while only one D-Ala molecule is present in monomer B at the corresponding position as D-Ala' in monomer A. The different binding states of monomers A and B in the

crystal seem to be correlated with the oxidation state of Cys176. Its side chain is oxidized to sulfenic acid only in monomer A, as indicated by the extra electron density (C176_{OH} in Figure 1-10A). An oxygen atom of sulfenic acid possibly stabilizes the weak binding of the product D-Ala^N in monomer A. D-Ala' is located between loops I and III, whereas D-Ala'' is trapped by loops II, III, and IV, right in front of the nucleophile Cys176. D-Ala' is too far from the predicted site of the scissile bond for Cys176 to act, but it interacts with Lys107^{Nζ}, Glu110^{Oε2}, His155^{N,O,Nδ1}, and Gly111^N, as well as Lys174^N and the main-chain oxygen atom of Leu113 via a water molecule. Interestingly, the side-chain portion of mDAP, part of the Csd6 substrate muramyl-L-Ala¹-γ-D-Glu²-mDAP³-D-Ala⁴, is structurally identical to a D-Ala molecule. Therefore, I suggest that D-Ala' and D-Ala'' represent the mDAP portion of the substrate and the end product, D-Ala, respectively, for the link between mDAP³ and D-Ala⁴ (the scissile peptide bond) of the muramyl tetrapeptide. mDAP was modeled into the Csd6 LD-CPase domain on the basis of the positions of D-Ala' and Wat1 bound in the oxyanion hole. The modeled mDAP fits well into the active-site cleft (bottom panel of Fig. 4A). It makes proper interactions with the oxyanion hole and is suitably positioned for a nucleophilic attack by Cys176, thus providing the substrate binding mode. Dissociation constants (K_D) of Csd6 for its complex with D-

Ala (or *m*DAP) have also been measured to be 224 (or 94) μM using SPR measurements. The measured association and dissociation rate constants (k_a and k_d) are $1.05 \text{ M}^{-1}\text{s}^{-1}$ and $2.30 \times 10^{-4} \text{ s}^{-1}$ for D-Ala and $1.49 \text{ M}^{-1}\text{s}^{-1}$ and $1.40 \times 10^{-4} \text{ s}^{-1}$ for *m*DAP, respectively.

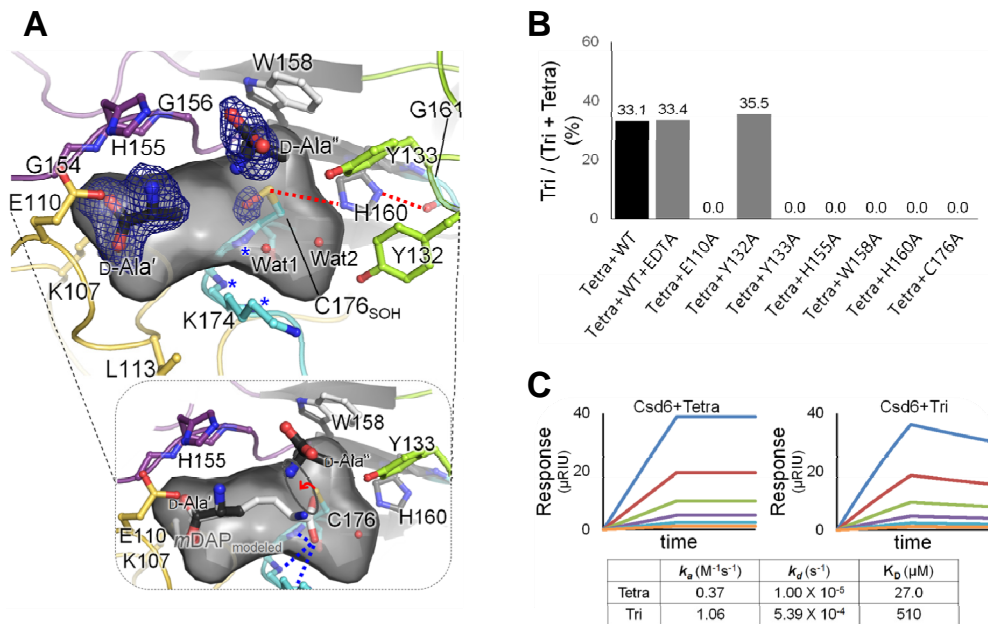


Figure 1-10. D-Ala-complexed Csd6 and key residues for the LD-CPase activity

(A) The ribbon diagram and the accessible inner surface of the active-site in the Csd4-Ala structure. The bound D-Ala molecules (D-Ala' and D-Ala'') (upper panel) and the mDAP (bottom panel) modeled on the basis of D-Ala' and Wat1 are shown in stick models. The omit mFo-DFc maps for D-Ala' and D-Ala'' (contoured at 2.0σ) and the side chain oxygen atom of Cys176 oxidized as the sulfenic acid (contoured at 2.5σ) are colored in blue. The oxyanion hole are asterisked and also shown as blue dotted lines in the

bottom panel. Red dotted lines indicate the interactions among residues of the catalytic triad. (B) The L,D-carboxypeptidation activities of the wild-type Csd6 (with the treatment of EDTA or not) and the mutants (E110A, Y132A, Y133A, H155A, W158A, H160A, and C176A) with the muramyl tetrapeptide. (C) SPR experiments with immobilized Csd6 and the muramyl tetrapeptide (or muramyl tripeptide) as an analyte at different concentrations (15.6, 31.3, 62.5, 125, 250, and 500 μ M) are shown in traces colored as orange, light blue, purple, green, red, and blue, respectively.

To further confirm the role of potential key residues in the active-site cleft, eight single amino-acid mutant proteins (E110A, Y132A, Y133A, H155A, W158A, H160A, and C176A) were prepared. When the L,D-carboxypeptidation activities of the wild-type Csd6 and the mutants against the muramyl tetrapeptide for a limited time (30 min) were compared, mutation of Glu110, Tyr133, His155, and Trp158, as well as His160 and Cys176 that belong to the catalytic triad, resulted in a complete loss of the L,D-carboxypeptidation activity (Figure 1-10B). Glu110, Tyr133, and Trp158 are strictly conserved in *C. jejuni* Pgp2 as well (Figure 1-9E). Lack of L,D-carboxypeptidation activity in E110A and H155A mutants supports that Glu110 and His155 recognize the side chain of mDAP, as shown in D-Ala' bound to Csd6 and mDAP modeled into the position of D-Ala'' (Figure 1-10A). Trp158 holds the aliphatic side chain of D-Ala'', whereas Tyr133 likely interact with the main chain of mDAP, possibly via a water molecule such as Wat2 in Figure 1-10A. Meanwhile, the retained activity of the Y132A mutant indicates that Tyr132 is not important for the L,D-carboxypeptidation activity of Csd6 (Figure 1-10B). The L,D-carboxypeptidation activity of Csd6 is independent of metal ions, as there is no change upon EDTA treatment (Figure 1-10B).

To estimate the kinetics and affinity of Csd6 with the muramyl tetrapeptide (as a substrate) and muramyl tripeptide (as a product), SPR experiments were conducted. The association rate constant (k_a) and the dissociation rate constant (k_d) for the muramyl tripeptide are roughly three-times and fifty-times larger than those for the muramyl tetrapeptide, respectively (Figure 1-10C). As a consequence, K_D values for the muramyl tetrapeptide and the muramyl tripeptide are 27.0 μM and 510 μM , respectively, indicating that the muramyl tripeptide dissociates readily from Csd6 as the product of Csd6-catalyzed LD-CPase reaction. This result is also consistent with the muramyl tripeptide not being used as the substrate for the possible LD-TPase cross-linking reaction in the subsequent step.

1.3.6. The C-terminal NTF2-like domain has a putative binding pocket for pseudaminic acid

H. pylori Csd6 is distinct from the known LD-TPases by having a C-terminal NTF2-like domain (residues Thr209–Lys330), following its LD-CPase domain. The combination of these two domains is conserved in the so-called ErfK/YbiS/YcfS/YnhG family proteins from both *Helicobacter* and *Campylobacter* species (Figure 1-11). It was reported that *csd6* gene-deficient transposon mutants of *H. pylori* G27 exhibited altered motility in comparison to its parental strain and the altered pattern of motility resulted from elevated levels of *O*-linked FlaA Pse-glycosylation of flagellin (Asakura et al., 2010). *H. pylori* flagellin filaments are post-translationally modified by glycosylation with Pse, a nine carbon sugar derivative that resembles sialic acid, which is typically found on mammalian cell surfaces (Figure 1-12B) (Schirm et al., 2003; Logan, 2006). The transposon insertions causing the altered motility were located at either positions +765 bp or +803 bp from the 5' terminus of the *csd6* gene (Asakura et al., 2010), corresponding to the region encoding the NTF2-like domain. Furthermore, the recombinant *H. pylori* G27 Csd6 protein expressed in *E. coli* was capable of deglycosylating the purified FlaA protein (Asakura et al., 2010).

These results suggest a possible function of the Csd6 NTF2-like domain as a deglycosylase, i.e., pseudaminidase.

The present structure shows that the overall-fold of the Csd6 NTF2-like domain differs from other non-viral sialidase or pseudaminidase of the six-bladed β -propeller fold (Xu et al., 2009; Kim et al., 2011). Instead, it is shared with other NTF2-like superfamily members such as the *Magnaporthe grisea* scytalone dehydratase (Motoyama et al., 2002) (PDB code 1IDP; a Z-score of 14.3, and a sequence identity of 8% for 118 equivalent C α positions), *Pseudomonas putida* Δ^5 -3-ketosteroid isomerase (Kim et al., 1997) (PDB code 1OPY; a Z-score of 12.0, and a sequence identity of 8% for 103 equivalent C α positions), and *Rhodococcus sp.* naphthalene dioxygenase (Gakhar et al., 2005) (PDB code 2B1X; a Z-score of 14.1, and a sequence identity of 7% for 116 equivalent C α positions) (Figure 1-12A). None of these enzymes bears any significant sequence or functional homology to each other, except that they share common hydrophobic substrates. This suggests that the fold of NTF2-like proteins may be a suitable scaffold for building a pocket to bind hydrophobic ligands. Although the Csd6 NTF2-like domain differs considerably from these enzymes in the sequence and the potential function, it shares a deep, largely hydrophobic pocket that is open to the solvent (Figure 1-12, A and C). This

potential ligand-binding pocket is lined with numerous hydrophobic residues (**Trp223**, **Trp227**, **Tyr235**, **Tyr239**, **Phe243**, Tyr246, Tyr257, **Tyr297**, and Tyr307). In addition, four positively charged residues (Arg245, **Lys258**, Lys268, and **Lys313**), and a negatively charged residue (**Glu329**) are located within the pocket. The bold-faced residues are strictly conserved in Csd6 homologs in ϵ -proteobacteria (Figure 1-11). The pocket is somewhat elongated in its shape, unlike hemispherical pockets in scytalone dehydratase and Δ^5 -3-ketosteroid isomerase (Figure 1-12, A and C). The Csd6 NTF2-like domain is highly basic with calculated pI of 9.4; the electrostatic potential surface shows a highly positively charged surface around its deep pocket (Figure 1-12A), compared to the above three NTF2-like superfamily enzymes.

In light of the previous report that *H. pylori* Csd6 is involved in the deglycosylation of flagellin (Asakura et al., 2010) and the presence of a deep pocket in the NTF2-like domain of Csd6 for possible binding of a hydrophobic ligand with an acidic group(s) such as Pse, an *in silico* docking study has been performed. It shows that Pse can fit well into the pocket of the NTF-like domain (Figure 1-12C). However, it remains to be determined whether Csd6 has a pseudaminidase activity.

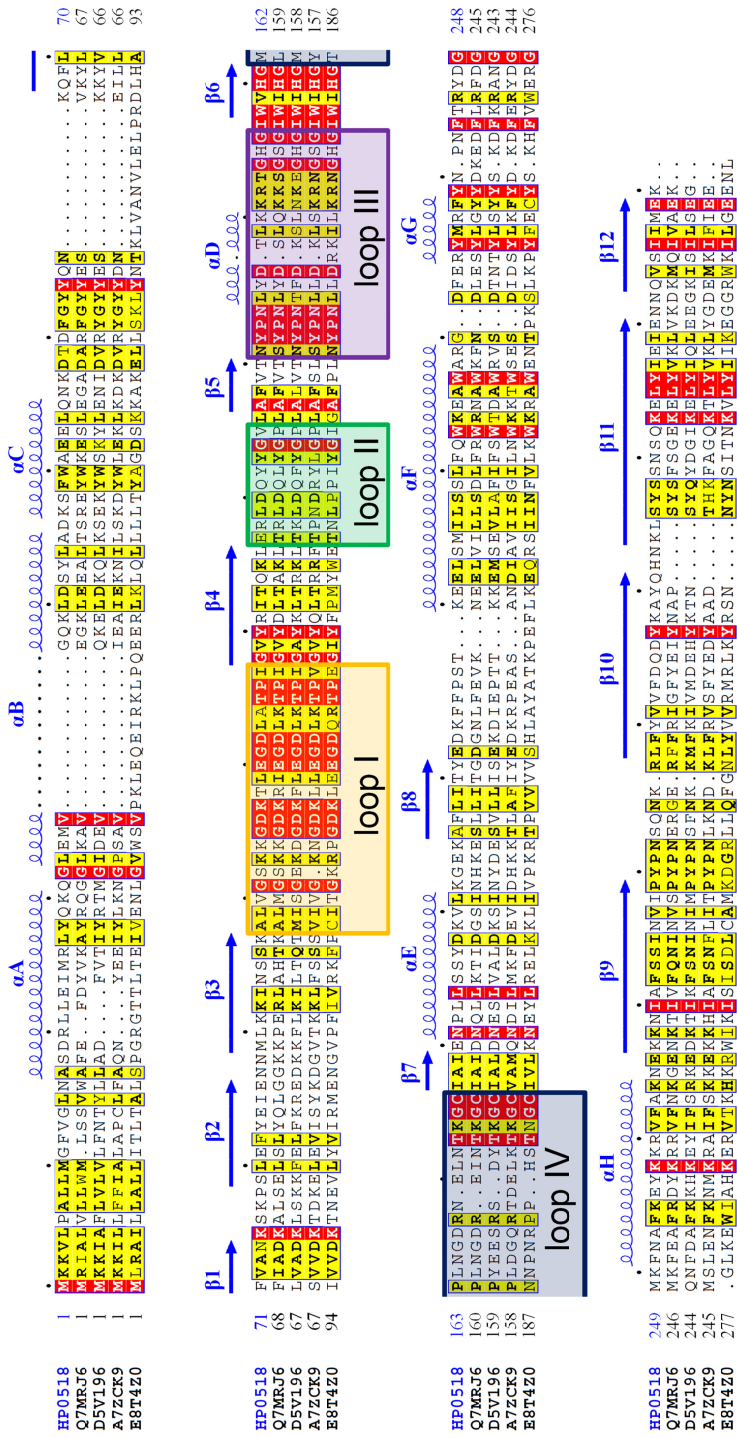


Figure 1-11. Sequence alignment of Csd6 homologs

Sequence alignment of Csd6 homologs from *H. pylori* strain 26695 (HP0518; SWISS-PROT accession code: O25255), from *Wolinella succinogenes* (Q7MRJ6), from *Arcobacter nitrofigilis* (D5V196), from *Campylobacter concisus* strain 13826 (A7ZCK9), and from *Thermovibrio ammonificans* (E8T4Z0) was performed and presented by ClustalX (Larkin et al., 2007) and ESPript (<http://esprict.ibcp.fr>) (Gouet et al., 2003). The secondary structure of Csd6 is presented above the aligned sequences. The four loops I–IV around the active site of the Csd6 LD-CPase domain are indicated with boxes colored as in Figure 1-9E.

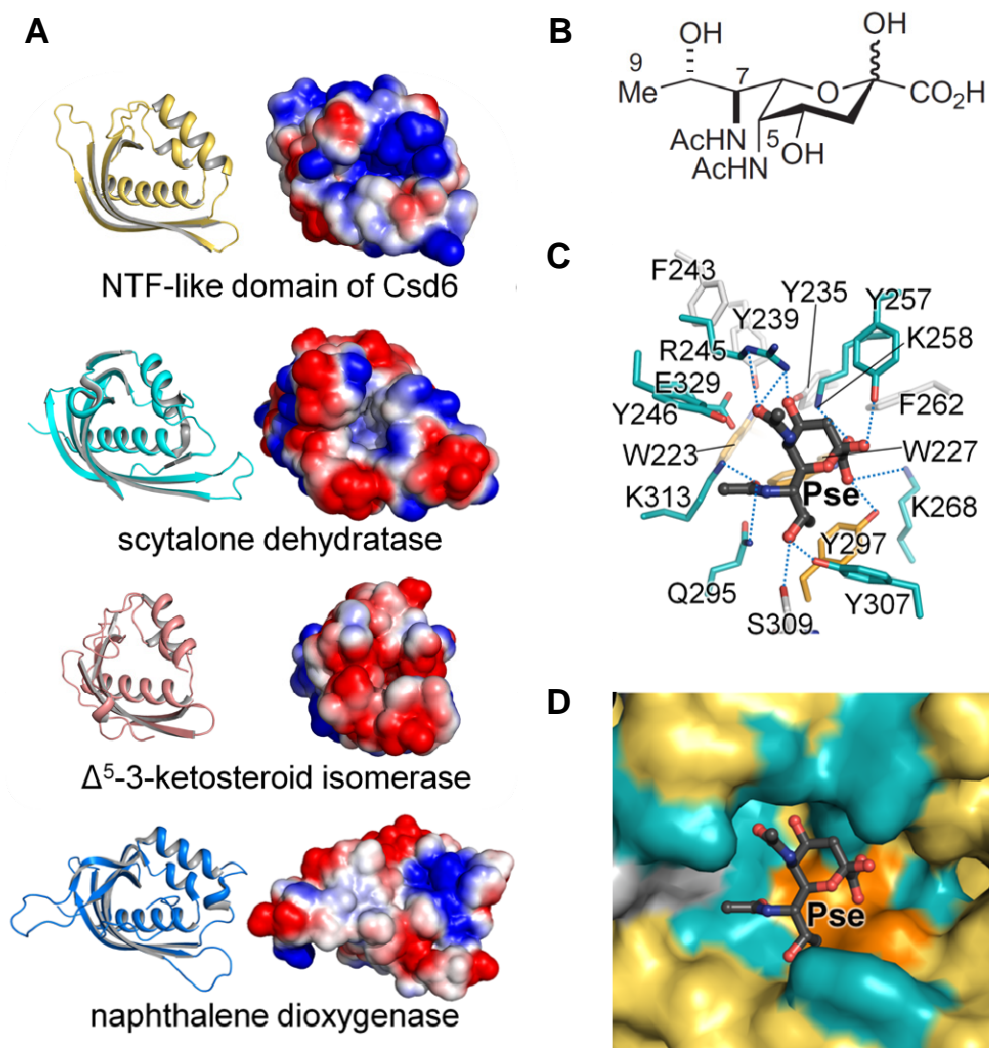


Figure 1-12. The NTF-like domain of Csd6

(A) Ribbon diagrams and electrostatic potential surface diagrams of the NTF-like domains of Csd6, scytalone dehydratase, Δ^5 -3-ketosteroid isomerase, and naphthalene dioxygenase. (B) Chemical structure of Pse. (C)

and (D), *In silico* docking of Pse onto the pocket in the NTF-like domain of Csd6. The potential binding pocket is shown in surface representation (C) and the residues around the docked ligand are shown in stick-models (D).

1.4. Discussion

I have determined the first crystal structure of Csd6, one of the cell-shape determining proteins in *H. pylori*. Each of its three domains is associated with a distinct function. The NTD plays a dominant role in homo-dimerization. The C-terminal NTF2-like domain has a putative binding pocket for Pse. The central LD-CPase domain has the ‘LD-TPase’ fold but the shape of its active-site pocket explains why Csd6 possesses only the L,D-carboxypeptidation activity despite the presence of the conserved catalytic residues for the potential transpeptidase reaction.

H. pylori Csd6, as well as its close homolog Pgp2 from *C. jejuni*, contribute to the helical morphology of the cells. Interestingly, *H. pylori* Csd6 and *C. jejuni* Pgp2 are unrelated in their sequence and structure to previously-characterized LD-CPases such as LdcA and LdcB from other microorganisms. They are thus presumed to form a third family of LD-CPases. However, it is currently impossible to assign and distinguish genuine LD-TPases from the functional LD-CPases via bioinformatics analyses of amino-acid sequences alone. The finding that Csd6-specific active-site loops I and III are much longer than the corresponding loops in

known LD-TPases (Figure 1-9E) enables one to distinguish LD-CPases possessing the ‘LD-TPase’ fold from *bona fide* LD-TPases. Indeed, when amino-acid sequences of Csd6 homologs in ϵ -proteobacteria are aligned, Csd6-specific loops I and III display high levels of sequence conservation (Figure 1-11). In light of this observation, the current annotation of *H. pylori* Csd6, as well as its homologs in ϵ -proteobacteria, as belonging to the protein family containing an ‘LD-TPase’ catalytic domain needs to be revised. I propose that they should be classified as a third distinct family of LD-CPase, as initially noted by others (Cadby and Lovering, 2014).

Enhanced cell attachment of *H. pylori* leads to an increased immune response of the host (Guruge et al., 1998). Too much attachment may thus be detrimental to the bacterium, suggesting a role for motility in the proper balance of bacterial levels during infection. This is intriguing, given that *H. pylori* Csd6 is known to be involved in regulating motility through trimming of the peptidoglycan-derived peptides as an LD-CPase or through affecting the Pse-glycosylation level of flagellin. This study shows that the active-site of Csd6 ‘LD-TPase’ domain is tailored to function as the LD-CPase, which converts muramyl tetrapeptides into muramyl tripeptides. As Csd6 is nonfunctional as LD-TPase, muramyl tripeptides in the peptidoglycan cannot be 3→3 cross-linked. This is consistent with the previous reports that the

Helicobacter peptidoglycan layer lacks 3→3 cross-links (Costa et al., 1999; Sycuro et al., 2010; Lavollay et al., 2008). An *in silico* docking shows that the Csd6 NTF-like domain presents a potential binding site for Pse, supporting its role in the control of the glycosylation level of flagellin. Relaxation of peptidoglycan 4→3 cross-linking was reported to promote *H. pylori*'s helical shape and stomach colonization (Sycuro et al., 2010). Csd6 plays a critical role in determining the helical cell shape of *H. pylori*, because it can limit the supply of substrates for peptidoglycan 4→3 cross-linking. In conclusion, our study on the multi-domain, multi-functional Csd6 provides further insights into the strategy of *H. pylori* for regulating its helical cell shape and motility, which are crucial for its virulence.

1.5. References

- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. sect. D Biol. Crystallogr.* **66**:213–221.
- An, D. R., Kim, H. S., Kim, J., Im, H. N., Yoon, H. J., Yoon, J. Y., Jang, J. Y., Heseck, D., Lee, M., Mobashery, S., Kim, S. J., Lee, B. I., and Suh, S. W. (2015) Structure of Csd3 from *Helicobacter pylori*, a cell shape-determining metallopeptidase. *Acta Crystallogr. sect. D Biol. Crystallogr.* **71**:675–686.
- Asakura, H., Churin, Y., Bauer, B., Boettcher, J. P., Bartfeld, S., Hashii, N., Kawasaki, N., Mollenkopf, H. J., Jungblut, P. R., Brinkmann, V., and Meyer, T. F. (2010) *Helicobacter pylori* HP0518 affects flagellin glycosylation to alter bacterial motility. *Mol. Microbiol.* **78**:1130–1144.

- Berg, H. C., and Turner, L. (1979) Movement of microorganisms in viscous environments. *Nature* **278**:349–351.
- Biarrotte-Sorin, S., Hugonnet, J. E., Delfosse, V., Mainardi, J. L., Gutmann, L., Arthur, M., and Mayer, C. (2006) Crystal structure of a novel beta-lactam-insensitive peptidoglycan transpeptidase. *J. Mol. Biol.* **359**:533–538.
- Bielnicki, J., Devedjiev, Y., Derewenda, U., Dauter, Z., Joachimiak, A., and Derewenda, Z. S. (2006) *B. subtilis* ykuD protein at 2.0 Å resolution: insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. *Proteins* **62**:144–151.
- Bonis, M., Ecobichon, C., Guadagnini, S., Prevost, M. C., and Boneca, I. G. (2010) A M23B family metallopeptidase of *Helicobacter pylori* required for cell shape, pole formation and virulence. *Mol. Microbiol.* **78**:809–819.
- Cadby, I. T., and Lovering, A. L. (2014) Life in the “old bag” yet: structure of peptidoglycan L,D-carboxypeptidases. *Structure* **22**:932–934.
- Chamaillard, M., Hashimoto, M., Horie, Y., Masumoto, J., Qiu, S., Saab, L., Ogura, Y., Kawasaki, A., Fukase, K., Kusumoto, S., Valvano, M. A., Foster, S. J., Mak, T. W., Nunez, G., and Inohara, N. (2003) An

- essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nature immunol.* **4**:702–707.
- Chaput, C., Ecobichon, C., Cayet, N., Girardin, S. E., Werts, C., Guadagnini, S., Prevost, M. C., Mengin-Lecreulx, D., Labigne, A., and Boneca, I. G. (2006) Role of AmiA in the morphological transition of *Helicobacter pylori* and in immune escape. *PLoS Pathog.* **2**:e97
- Chaput, C., Labigne, A., and Boneca, I. G. (2007) Characterization of *Helicobacter pylori* lytic transglycosylases Slt and MltD. *J. Bacteriol.* **189**:422–429.
- Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. sect. D Biol. Crystallogr.* **66**:12–21.
- Cordillot, M., Dubee, V., Triboulet, S., Dubost, L., Marie, A., Hugonnet, J. E., Arthur, M., and Mainardi, J. L. (2013) *In vitro* cross-linking of *Mycobacterium tuberculosis* peptidoglycan by L,D-transpeptidases and inactivation of these enzymes by carbapenems. *Antimicrob. Agents Chemother.* **57**:5940–5945.

- Costa, K., Bacher, G., Allmaier, G., Dominguez-Bello, M. G., Engstrand, L., Falk, P., de Pedro, M. A., and Garcia-del Portillo, F. (1999) The morphological transition of *Helicobacter pylori* cells from spiral to coccoid is preceded by a substantial modification of the cell wall. *J. Bacteriol.* **181**:3710–3715.
- Covaleda, G., del Rivero, M. A., Chávez, M. A., Avilés, F. X., and Reverter, D. (2012) Crystal structure of novel metallo-carboxypeptidase inhibitor from marine mollusk *Nerita versicolor* in complex with human carboxypeptidase A4. *J. Biol. Chem.* **287**:9250–9258.
- Das, D., Herve, M., Elsliger, M. A., Kadam, R. U., Grant, J. C., Chiu, H. J., Knuth, M. W., Klock, H. E., Miller, M. D., Godzik, A., Lesley, S. A., Deacon, A. M., Mengin-Lecreulx, D., and Wilson, I. A. (2013) Structure and function of a novel LD-carboxypeptidase a involved in peptidoglycan recycling. *J. Bacteriol.* **195**:5555–5566.
- Dodson, G., and Wlodawer, A. (1998) Catalytic triads and their relatives. *Trends Biochem. Sci.* **23**:347–352.
- Doerks, T., Bork, P., Kamberov, E., Makarova, O., Muecke, S., and Margolis, B. (2000) L27, a novel heterodimerization domain in receptor targeting proteins Lin-2 and Lin-7. *Trends Biochem. Sci.* **25**:317–318.

- Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. sect. D Biol. Crystallogr.* **66**:486–501.
- Firdich, E., Vermeulen, J., Biboy, J., Soares, F., Taveirne, M. E., Johnson, J. G., DiRita, V. J., Girardin, S. E., Vollmer, W., and Gaynor, E. C. (2014) Peptidoglycan LD-carboxypeptidase Pgp2 influences *Campylobacter jejuni* helical cell shape and pathogenic properties and provides the substrate for the DL-carboxypeptidase Pgp1. *J. Biol. Chem.* **289**:8007–8018.
- Gakhar, L., Malik, Z. A., Allen, C. C., Lipscomb, D. A., Larkin, M. J., and Ramaswamy, S. (2005) Structure and increased thermostability of *Rhodococcus sp.* naphthalene 1,2-dioxygenase. *J. Bacteriol.* **187**:7222–7231.
- Gaskell, A., Crennell, S., and Taylor, G. (1995) The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* **3**:1197–1205.
- Girardin, S. E., Boneca, I. G., Carneiro, L. A., Antignac, A., Jehanno, M., Viala, J., Tedin, K., Taha, M. K., Labigne, A., Zahringer, U., Coyle, A. J., DiStefano, P. S., Bertin, J., Sansonetti, P. J., and Philpott, D. J.

- (2003) Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. *Science* **300**:1584–1587.
- Gouet, P., Robert, X., and Courcelle, E. (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**:3320–3323.
- Guruge, J. L., Falk, P. G., Lorenz, R. G., Dans, M., Wirth, H. P., Blaser, M. J., Berg, D. E., and Gordon, J. I. (1998) Epithelial attachment alters the outcome of *Helicobacter pylori* infection. *Proc. Natl. Acad. Sci. U.S.A.* **95**:3925–3930.
- Harris, B. Z., Venkatasubrahmanyam, S., and Lim, W. A. (2002) Coordinated folding and association of the LIN-2, -7 (L27) domain. An obligate heterodimerization involved in assembly of signaling and cell polarity complexes. *J. Biol. Chem.* **277**:34902–34908.
- Hazell, S. L., Lee, A., Brady, L., and Hennessy, W. (1986) *Campylobacter pyloridis* and gastritis: association with intercellular spaces and adaptation to an environment of mucus as important factors in colonization of the gastric epithelium. *J. Infect. Dis.* **153**:658–663.
- Hohlbein, J., Craggs, T. D., and Cordes, T. (2014) Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.* **43**:1156–1171.

- Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**:W545–549.
- Hoyland, C. N., Aldridge, C., Cleverley, R. M., Duchene, M. C., Minasov, G., Onopriyenko, O., Sidiq, K., Stogios, P. J., Anderson, W. F., Daniel, R. A., Savchenko, A., Vollmer, W., and Lewis, R. J. (2014) Structure of the LdcB LD-carboxypeptidase reveals the molecular basis of peptidoglycan recognition. *Structure* **22**:949–960.
- Josenhans, C., Vossebein, L., Friedrich, S., and Suerbaum, S. (2002) The *neuA/flmD* gene cluster of *Helicobacter pylori* is involved in flagellar biosynthesis and flagellin glycosylation. *FEMS Microbiol. Lett.* **210**:165–172.
- Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**:2577–2637.
- Kim, C., Kim, J. Y., Kim, S. H., Lee, B. I., and Lee, N. K. (2012) Direct characterization of protein oligomers and their quaternary structures by single-molecule FRET. *Chem. Commun. (Camb)* **48**:1138–1140.
- Kim, H. S., Kim, J., Im, H. N., An, D. R., Lee, M., Heseck, D., Mobashery, S., Kim, J. Y., Cho, K., Yoon, H. J., Han, B. W., Lee, B. I., and Suh, S. W. (2014) Structural basis for the recognition of muramyltripeptide by

- Helicobacter pylori* Csd4, a D,L-carboxypeptidase controlling the helical cell shape. *Acta Crystallogr. sect. D Biol. Crystallogr.* **70**:2800–2812.
- Kim, H. S., Kim, J., Im, H. N., Yoon, J. Y., An, D. R., Yoon, H. J., Kim, J. Y., Min, H. K., Kim, S. J., Lee, J. Y., Han, B. W., and Suh, S. W. (2013) Structural basis for the inhibition of *Mycobacterium tuberculosis* L,D-transpeptidase by meropenem, a drug effective against extensively drug-resistant strains. *Acta Crystallogr. sect. D Biol. Crystallogr.* **69**:420–431.
- Kim, S. W., Cha, S. S., Cho, H. S., Kim, J. S., Ha, N. C., Cho, M. J., Joo, S., Kim, K. K., Choi, K. Y., and Oh, B. H. (1997) High-resolution crystal structures of Δ^5 -3-ketosteroid isomerase with and without a reaction intermediate analogue. *Biochemistry* **36**:14030–14036.
- Kim, S., Oh, D. B., Kang, H. A., and Kwon, O. (2011) Features and applications of bacterial sialidases. *Appl. Microbiol. Biotechnol.* **91**:1–15.
- Korza, H. J., and Bochtler, M. (2005) *Pseudomonas aeruginosa* LD-carboxypeptidase, a serine peptidase with a Ser-His-Glu triad and a nucleophilic elbow. *J. Biol. Chem.* **280**:40802–40812.

- Kostrzynska, M., Betts, J. D., Austin, J. W., and Trust, T. J. (1991) Identification, characterization, and spatial localization of two flagellin species in *Helicobacter pylori* flagella. *J. Bacteriol.* **173**:937–946.
- Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**:774–797.
- Kusters, J. G., van Vliet, A. H., and Kuipers, E. J. (2006) Pathogenesis of *Helicobacter pylori* infection. *Clin. Microbiol. Rev.* **19**:449–490.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947–2948.
- Lavollay, M., Arthur, M., Fourgeaud, M., Dubost, L., Marie, A., Veziris, N., Blanot, D., Gutmann, L., and Mainardi, J. L. (2008) The peptidoglycan of stationary-phase *Mycobacterium tuberculosis* predominantly contains cross-links generated by L,D-transpeptidation. *J. Bacteriol.* **190**:4360–4366.
- Lee, M., Zhang, W., Heseck, D., Noll, B. C., Boggess, B., and Mobashery, S. (2009) Bacterial AmpD at the crossroads of peptidoglycan recycling and manifestation of antibiotic resistance. *J. Am. Chem. Soc.* **131**:8742–8743.

- Lee, N. K., Kapanidis, A. N., Wang, Y., Michalet, X., Mukhopadhyay, J., Ebright, R. H., and Weiss, S. (2005) Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys. J.* **88**:2939–2953.
- Lertsethtakarn, P., Ottemann, K. M., and Hendrixson, D. R. (2011) Motility and chemotaxis in *Campylobacter* and *Helicobacter*. *Annu. Rev. Microbiol.* **65**:389–410.
- Li, Y., Karnak, D., Demeler, B., Margolis, B., and Lavie, A. (2004) Structural basis for L27 domain-mediated assembly of signaling and cell polarity complexes. *EMBO J.* **23**:2723–2733.
- Logan, S. M. (2006) Flagellar glycosylation – a new component of the motility repertoire? *Microbiology* **152**:1249–1262.
- Mainardi, J. L., Legrand, R., Arthur, M., Schoot, B., van Heijenoort, J., and Gutmann, L. (2000) Novel mechanism of beta-lactam resistance due to bypass of DD-transpeptidation in *Enterococcus faecium*. *J. Biol. Chem.* **275**:16490–16496.
- Malferteiner, P., Selgrad, M., and Bornschein, J. (2012) *Helicobacter pylori*: clinical management. *Curr. Opin. Gastroenterol.* **28**:608–614.

- McCarter, L., and Silverman, M. (1990) Surface-induced swarmer cell differentiation of *Vibrio parahaemolyticus*. *Mol. Microbiol.* **4**:1057–1062.
- Medina-Franco, J. L., Mendez-Lucio, O., and Yoo, J. (2014) Rationalization of activity cliffs of a sulfonamide inhibitor of DNA methyltransferases with induced-fit docking. *Int. J. Mol. Sci.* **15**:3253–3261.
- Meroueh, S. O., Bencze, K. Z., Heseck, D., Lee, M., Fisher, J. F., Stemmler, T. L., and Mobashery, S. (2006) Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proc. Natl. Acad. Sci. U.S.A.* **103**:4404–4409.
- Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S. Y., Bateman, A., Punta, M., Attwood, T. K., Sigrist, C. J., Redaschi, N., Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H., Natale, D. A., Wu, C. H., Orengo, C., Sillitoe, I., Mi, H., Thomas, P. D., and Finn, R. D. (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43**:D213–221.

- Motoyama, T., Nakasako, M., and Yamaguchi, I. (2002) Crystallization of scytalone dehydratase F162A mutant in the unligated state and a preliminary X-ray diffraction study at 37 K. *Acta Crystallogr. sect. D Biol. Crystallogr.* **58**:148–150.
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. sect. D Biol. Crystallogr.* **53**:240–255.
- Ottemann, K. M., and Lowenthal, A. C. (2002) *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. *Infect. Immun.* **70**:1984–1990.
- Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**:307–326.
- Pei, J., Kim, B. H., and Grishin, N. V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**:2295–2300.
- Roesler, B. M., Rabelo-Goncalves, E. M., and Zeitune, J. M. (2014) Virulence Factors of *Helicobacter pylori*: A Review. *Clin. Med. Insights Gastroenterol.* **7**:9–17.
- Roubaud Baudron, C., Franceschi, F., Salles, N., and Gasbarrini, A. (2013) Extragastric diseases and *Helicobacter pylori*. *Helicobacter* **18**:44–51.

- Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A., and Charlier, P. (2008) The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol. Rev.* **32**:234–258.
- Scheffers, D. J., and Pinho, M. G. (2005) Bacterial cell wall synthesis: new insights from localization studies. *Microbiol. Mol. Biol. Rev.* **69**:585–607.
- Schirm, M., Soo, E. C., Aubry, A. J., Austin, J., Thibault, P., and Logan, S. M. (2003) Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol. Microbiol.* **48**:1579–1592.
- Schreiber, S., Konradt, M., Groll, C., Scheid, P., Hanauer, G., Werling, H. O., Josenhans, C., and Suerbaum, S. (2004) The spatial orientation of *Helicobacter pylori* in the gastric mucus. *Proc. Natl. Acad. Sci. U.S.A.* **101**:5024–5029.
- Silva, J. R., Roitberg, A. E., and Alves, C. N. (2014) Catalytic mechanism of L,D-transpeptidase 2 from *Mycobacterium tuberculosis* described by a computational approach: insights for the design of new antibiotics drugs. *J. Chem. Inf. Model.* **54**:2402–2410.
- Sycuro, L. K., Pincus, Z., Gutierrez, K. D., Biboy, J., Stern, C. A., Vollmer, W., and Salama, N. R. (2010) Peptidoglycan crosslinking relaxation

- promotes *Helicobacter pylori*'s helical shape and stomach colonization. *Cell* **141**:822–833.
- Sycuro, L. K., Rule, C. S., Petersen, T. W., Wyckoff, T. J., Sessler, T., Nagarkar, D. B., Khalid, F., Pincus, Z., Biboy, J., Vollmer, W., and Salama, N. R. (2013) Flow cytometry-based enrichment for cell shape mutants identifies multiple genes that influence *Helicobacter pylori* morphology. *Mol. Microbiol.* **90**:869–883.
- Sycuro, L. K., Wyckoff, T. J., Biboy, J., Born, P., Pincus, Z., Vollmer, W., and Salama, N. R. (2012) Multiple peptidoglycan modification networks modulate *Helicobacter pylori*'s cell shape, motility, and colonization potential. *PLoS Pathog.* **8**:e1002603.
- Terwilliger, T. C. (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr. sect. D Biol. Crystallogr.* **59**:38–44.
- Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. sect. D Biol. Crystallogr.* **66**:22–25.
- Vollmer, W., and Bertsche, U. (2008) Murein (peptidoglycan) structure, architecture and biosynthesis in *Escherichia coli*. *Biochim. Biophys. Acta* **1778**:1714–1734.

- Vollmer, W., Blanot, D., and de Pedro, M. A. (2008) Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* **32**:149–167.
- Vollmer, W., Joris, B., Charlier, P., and Foster, S. (2008) Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol. Rev.* **32**, 259–286
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. sect. D Biol. Crystallogr.* **67**:235–242.
- Worku, M. L., Sidebotham, R. L., Walker, M. M., Keshavarz, T., and Karim, Q. N. (1999) The relationship between *Helicobacter pylori* motility, morphology and phase of growth: implications for gastric colonization and pathology. *Microbiology* **145**:2803–2811.
- Wyckoff, T. J., Taylor, J. A., and Salama, N. R. (2012) Beyond growth: novel functions for bacterial cell wall hydrolases. *Trends Microbiol.* **20**:540–547.
- Xu, G., Ryan, C., Kiefel, M. J., Wilson, J. C., and Taylor, G. L. (2009) Structural studies on the *Pseudomonas aeruginosa* sialidase-like

enzyme PA2794 suggest substrate and mechanistic variations. *J. Mol. Biol.* **386**:828–840.

Yoo, J., Choi, S., and Medina-Franco, J. L. (2013) Molecular modeling studies of the novel inhibitors of DNA methyltransferases SGI-1027 and CBC12: implications for the mechanism of inhibition of DNMTs. *PLoS One* **8**:e62152.

Young, G. M., Schmiel, D. H., and Miller, V. L. (1999) A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. *Proc. Natl. Acad. Sci. U.S.A.* **96**:6456–6461.

Chapter 2

Crystal structure of Rv2258c from *Mycobacterium tuberculosis* H37Rv, an *S*-adenosyl-L-methionine-dependent methyltransferase

[This chapter was published in *J. Struct. Biol.* (2016) 3:172–180.]

2.1. Introduction

Mycobacterium tuberculosis is a highly successful intracellular pathogen, infecting nearly one-third of the world's population. It causes tuberculosis (TB), claiming the lives of millions of people in the world every year. The advent of multidrug-resistant TB cases, the HIV epidemic, imperfect diagnostic assays, limited vaccine efficacy, and non-availability of new anti-TB drugs pose a global health problem (Lin and Flynn, 2010). Therefore, world-wide efforts are being made to develop new anti-TB drugs and more effective vaccines to combat TB. As an important step, the genome sequence of *M. tuberculosis* H37Rv strain was reported in 1998

(Cole et al., 1998). However, I still have little or limited functional information on a significant portion of approximately 4,090 genes in *M. tuberculosis* (Lew et al., 2011). Identifying the molecular and biological functions of the proteins that are encoded by the *M. tuberculosis* genome would provide the groundwork for the development of new anti-TB drug targets.

Methyltransferases (MTases) mediate a wide variety of cellular processes, such as cell signaling, metabolite synthesis, and gene regulation in nearly all living organisms. They comprise a large family of over 300 members and transfer a methyl group most frequently from *S*-adenosyl-L-methionine (SAM) to various acceptor substrates, which include small molecules, lipids, proteins, and nucleic acids, yielding a methylated product with *S*-adenosyl-L-homocysteine (SAH) as a by-product. The Rv2258c protein from *M. tuberculosis* H37Rv and its close orthologs in mycobacteria are annotated as SAM-dependent methyltransferases and possible transcriptional regulatory proteins. To gain structural insights into the function of Rv2258c, specifically, to provide a structural basis to decipher substrate binding and specificity, I have determined its crystal structure in three forms: ligand-unbound Rv2258c, a binary complex with sinefungin (SFG), and a binary complex with SAH. A monomer of Rv2258c consists of

two domains, which are linked by a long connecting α -helix. The N-terminal domain is essential for dimerization and the C-terminal domain has the Class I MTase fold. The structure of Rv2258c is distinct from eleven other mycobacterial SAM-dependent MTases that have been structurally characterized but the overall fold of Rv2258c resembles those of small-molecule *O*-MTases. Rv2258c forms a homodimer in the crystal, with the N-terminal domains facing each other. Size-exclusion chromatography confirms that Rv2258c also exists as a homodimer in solution. Our structure reveals that the Rv2258c dimer has a large cavity for binding a methyl-accepting substrate adjacent to the SFG (or SAH) binding site. It also shows that dimerization is essential to form such a cavity, as it is contributed not only by two domains and the connecting α -helix of one monomer but also by an α -helix in the N-terminal domain of the other monomer. Due to high variability of the amino acid residues defining the substrate-binding site, further experiments are required to establish the methyl-accepting substrate and the biological role of Rv2258c.

2.2. Material and methods

2.2.1. Cloning, expression, and purification

To obtain well-diffracting crystals, six constructs encompassing residues 1–353 (full-length), 1–320, 4–353, 6–353, 9–353, and 28–353 of the Rv2258c protein from *M. tuberculosis* H37Rv were generated. The genes encoding these constructs were amplified by PCR using the genomic DNA of *M. tuberculosis* H37Rv as the template and were cloned into the expression vector pET-28b(+) (Novagen). The sequences of both the forward and reverse primers used in PCR are listed in Table 2-1. The resulting recombinant proteins are fused with hexa-histidine containing tags at both N- and C-termini (MGSSHHHHHHSSGLVPRGSH and LEHHHHHH, respectively). They were overexpressed in *Escherichia coli* Rosetta 2(DE3) cells. The cells were grown at 37°C in Luria Broth culture medium containing 30 µg ml⁻¹ kanamycin. Protein expression was induced by 0.5 mM isopropyl β-D-thiogalactopyranoside and the cells were incubated for additional 20 h at 30°C. The cells were harvested by centrifugation at 5,600 g for 15 min at 4°C and subsequently lysed by sonication in ice-cold buffer A (20 mM Tris-HCl at pH 7.9, 500 mM sodium

chloride, and 50 mM imidazole), which was supplemented with 10% (v/v) glycerol and 1 mM phenylmethylsulfonyl fluoride. The crude lysate was centrifuged at 36,000 g for 1 h at 4°C to discard the cell debris. The supernatant was applied to an affinity chromatography column of HiTrap Chelating HP (GE Healthcare), which was previously equilibrated with buffer A. The column was eluted with buffer B (20 mM Tris-HCl at pH 7.9, 500 mM sodium chloride, and 500 mM imidazole), with the Rv2258c protein being eluted at 120–180 mM imidazole concentration. The eluted protein was further purified by gel filtration on a HiLoad 16/60 Superdex 200 prep-grade column (GE Healthcare), which was previously equilibrated with buffer C (20 mM Tris-HCl at pH 7.0 and 200 mM sodium chloride). The protein purity was analyzed by SDS-PAGE. Fractions containing the recombinant Rv2258c were pooled and concentrated to 17 mg ml⁻¹ (0.42 mM monomer concentration) for crystallization using an Amicon Ultra-15 Centrifugal Filter Unit (Millipore). Four of the above constructs, 1–353 (full-length), 4–353, 6–353, and 9–353, were expressed in a soluble form in *E. coli* and were purified for crystallization trials. The 1-320 and 28-353 constructs were expressed in an insoluble form.

Table 2-1. Forward and reverse primers used in PCR

Construct	Description	5' to 3' sequence
1-320	Forward primer	GGAATTCCATATGAGCGGAGCACTCGAGACCA
	Reverse primer	ATAGTTTAGCGGGCCCTGCCCCACACCCGTC CCC
1-353	Forward primer	GGAATTCCATATGAGCGGAGCACTCGAGACCA
	Reverse primer	ATAGTTTAGCGGGCCGCTCACTTCCGGGGGATGTAGTAGT
4-353	Forward primer	GGAATTCCATATGGCACTCGAGACCACCGAAGAATT
	Reverse primer	ATAGTTTAGCGGGCCGCTCACTTCCGGGGGATGTAGTAGT
6-353	Forward primer	GGAATTCCATATGGAGACCACCGAAGAATTCCGGCAA
	Reverse primer	ATAGTTTAGCGGGCCGCTCACTTCCGGGGGATGTAGTAGT
9-353	Forward primer	GGAATTCCATATGGAAGAATTCCGGCAAACCGTTTGT
	Reverse primer	ATAGTTTAGCGGGCCGCTCACTTCCGGGGGATGTAGTAGT
28-353	Forward primer	GGAATTCCATATGGTGAGCGTTGGACACCAGACC
	Reverse primer	ATAGTTTAGCGGGCCGCTCACTTCCGGGGGATGTAGTAGT

*The underlined sequences are NdeI and NotI restriction enzyme sites, respectively

2.2.2. Crystallization and X-ray data collection

Crystals were grown at 23°C by the sitting drop vapor diffusion method using a Mosquito robotic system (TTP Labtech). The 1–353 (full-length) and 4–353 constructs produced crystals diffracting poorly to 4 Å only, while the 9–353 construct did not yield crystals. Only the 6–353 construct gave well-diffracting crystals. To obtain ligand-free crystals of the 6–353 construct ('Rv2258c-unbound'), a sitting drop was prepared by mixing 0.2 µl of the protein solution in buffer C and 0.2 µl of the reservoir solution [200 mM sodium malonate at pH 6.0 and 20% (v/v) PEG 3350]. The sitting drop was equilibrated against 100 µl of the reservoir solution. Crystals grew up to approximate dimensions of 0.2 mm × 0.3 mm × 0.2 mm within a few days (Figure 2-1).

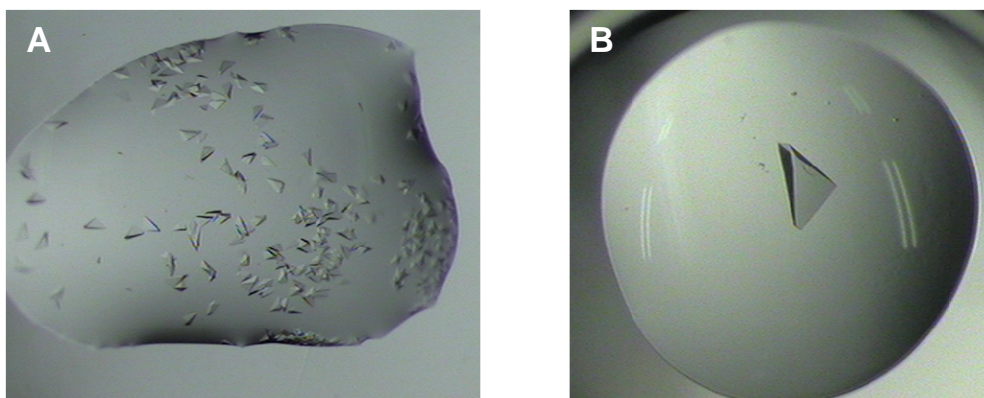


Figure 2-1. Crystals of Rv2258c-unbound from *M. tuberculosis*

(A) Crystals with approximate dimensions of $0.05 \text{ mm} \times 0.1 \text{ mm} \times 0.05 \text{ mm}$ before optimization. (B) Best crystal with approximate dimensions of $0.2 \text{ mm} \times 0.3 \text{ mm} \times 0.2 \text{ mm}$ after optimization.

In an effort to achieve phasing by single-wavelength anomalous diffraction (SAD), the selenomethionine-substituted Rv2258c protein (6–353 construct) was expressed in *E. coli* but it did not produce crystals. Instead, I prepared a platinum derivative of Rv2258c-unbound crystals by soaking them for 90 min in 5 μ l of a heavy atom-containing cryoprotectant solution, which was prepared by supplementing the reservoir solution with 30% (v/v) glycerol and 5 mM K_2PtCl_4 .

Attempts to co-crystallize the ligand-bound Rv2258c protein were not successful due to the tendency of the protein to aggregate in the presence of the ligands. Therefore, crystals of Rv2258c complexed with either sinefungin (SFG), an analog of the co-substrate *S*-adenosyl-L-methionine (SAM), or the by-product *S*-adenosyl-L-homocysteine (SAH) ('Rv2258c-SFG' and 'Rv2258c-SAH', respectively) were obtained by soaking crystals of Rv2258c-unbound for 1 min in 5 μ l of a cryoprotectant solution, which was prepared by supplementing the reservoir solution with 30% (v/v) glycerol and 19.9 mM SFG (or 12.5 mM SAH).

A set of X-ray diffraction data from a ligand-free crystal of Rv2258c ('Rv2258c-unbound') was collected to 1.83 Å on an Area Detector Systems Corporation Q270 CCD detector at the beamline BL-7A of Pohang Light Source, Korea. All the raw data were processed and scaled using the

program suite HKL2000 (Otwinowski and Minor, 1997). Crystals were flash-frozen in a nitrogen gas stream at 100 K. The crystal of Rv2258c-unbound belongs to the space group $C2$, with unit cell parameters of $a = 109.1 \text{ \AA}$, $b = 140.6 \text{ \AA}$, $c = 97.1 \text{ \AA}$, and $\beta = 98.5^\circ$. Assuming the presence of three monomers of the recombinant Rv2258c in the asymmetric unit, the Matthew's parameter and the solvent content are $3.26 \text{ \AA}^3 \text{ Da}^{-1}$ and 62.3%, respectively. Data collection statistics are given in Table 2-2.

Several sets of SAD data were collected at 100 K from different platinum-derivatized crystals at a wavelength of 1.0717 \AA using the Area Detector Systems Corporation Q315r CCD detector at the beamline BL-5A of Photon Factory, Japan. Data collection statistics are given in Table 2-2.

X-ray diffraction data from crystals of Rv2258c-SFG and Rv2258c-SAH were collected at 100 K using an Area Detector Systems Corporation Q270 CCD detector at the beamline BL-7A experimental station of Pohang Light Source, Republic of Korea and on an Area Detector Systems Corporation Q315r CCD detector at the beamline BL-5A of Photon Factory, Japan, respectively. The crystal of the Rv2258c-SFG diffracted to 1.90 \AA and belongs to the space group $C2$, with unit cell parameters of $a = 108.9 \text{ \AA}$, $b = 140.9 \text{ \AA}$, $c = 96.6 \text{ \AA}$, and $\beta = 98.3^\circ$. The crystal of Rv2258c-SAH diffracted to 2.90 \AA and belongs to the space group $C2$, with unit cell

parameters of $a = 109.6 \text{ \AA}$, $b = 140.5 \text{ \AA}$, $c = 96.1 \text{ \AA}$, and $\beta = 97.9^\circ$. Both crystals contain three monomers per asymmetric unit, giving a Matthew's parameter and solvent fraction of $3.31 \text{ \AA}^3 \text{ Da}^{-1}$ and 62.9% for Rv2258c-SFG, and $3.27 \text{ \AA}^3 \text{ Da}^{-1}$ and 62.4% for Rv2258c-SAH, respectively. Data collection statistics are given in Table 2-2.

Table 2-2. Data collection and phasing statistics

<i>Data collection</i>		Rv2258c-unbound	Rv2258c-SFG	Rv2258c-SAH	Rv2258c-Pt (peak)
Data set					
Space group		C2	C2	C2	C2
Unit cell lengths, <i>a</i> , <i>b</i> , <i>c</i> (Å)		109.1, 140.6, 97.1	108.9, 140.9, 96.9	109.6, 140.5, 96.1	113.1, 140.0, 97.2
Unit cell angles, α , β , γ (°)		90, 98.5, 90	90, 98.3, 90	90, 97.9, 90	90, 97.8, 90
X-ray wavelength (Å)		0.9201	0.9793	1.0000	1.07178
Resolution range (Å)		50.0–1.83 (1.86–1.83) ^a	50.0–1.90 (1.93–1.90) ^a	50.0–2.90 (2.95–2.90) ^a	50.0–2.60 (2.64–2.60) ^a
No. total reflections		534,964	464,116	132,373	431,911
No. unique reflections		127,941	113,690	32,186	43,315 ^b
Completeness (%)		99.9 (99.9) ^a	99.9 (100) ^a	99.9 (99.9) ^a	93.7 (68.7) ^{a,b}
$\langle I \rangle / \langle \sigma \rangle$		28.7 (2.9) ^a	28.8 (2.8) ^a	10.9 (2.8) ^a	29.1 (2.4) ^{a,b}
R_{rim}^c (%)		8.6 (70.1) ^a	8.3 (70.6) ^a	20.1 (58.9) ^a	11.1 (56.2) ^{a,b}
R_{pim}^d (%)		4.2 (34.4) ^a	4.1 (35.0) ^a	9.8 (28.9) ^a	3.5 (18.3) ^{a,b}
$CC_{1/2}^e$ (%)		99.7 (86.8) ^a	99.7 (86.3) ^a	98.6 (89.2) ^a	99.8 (95.7) ^{a,b}
MR-SAD phasing			Figure of merit (before / after density modification)	0.29 / 0.71	

Footnotes to Table 2-2

^a Values in parentheses refer to the highest resolution shell.

^b Friedel pairs were treated as separate observations.

$$^c R_{\text{rim}} = \sum_{\text{hkl}} \{N(\text{hkl}) / [N(\text{hkl}) - 1]\}^{1/2} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl}).$$

The redundancy-independent merging R factor gives the precision of individual intensity (Diederichs and Karplus, 1997).

^d $R_{\text{pim}} = \sum_{\text{hkl}} \{1 / [N(\text{hkl}) - 1]\}^{1/2} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$. The precision indicating merging R factor describes the precision of the averaged intensity (Weiss, 2001).

^e $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data (Diederichs and Karplus, 2013).

2.2.3. Structure determination and refinement

Initially, I tried to solve the structure by molecular replacement using a monomer structure of RebM from *Lechevalieria aerocolonigenes* (PDB: 3BUS) as a search model. The sequence identity between the residues 59–177 of RebM and the residues 154–276 of Rv2258c is 33%. There was no search model that covered the N-terminal half of Rv2258c. The molecular replacement solution gave an interpretable electron density only for the β -strands of the C-terminal domain. Therefore, I attempted SAD phasing. The best set of the platinum derivative data located 21 platinum atoms per asymmetric unit but the electron density map calculated using the SAD phases was poor and largely uninterpretable. Therefore, I improved the phases by combining SAD and molecular replacement phases using the program AUTOSOL of the PHENIX software package (Adams et al., 2010). When the combined phases were further improved by density modification using the program Resolve (Terwilliger, 2003), the electron density map became interpretable with an overall figure of merit of 0.70. The initial model obtained by autobuilding by Resolve was improved through iterative cycles of manual model building with Coot (Emsley et al., 2010) and refinement with Refmac5 of the CCP4 program suite (Murshudov et al., 1997). A total of 5% of the data was randomly set aside as test data for the

calculation of R_{free} (Brünger, 1992). The model quality was assessed using the program MolProbity (Chen et al., 2010). Phasing and model refinement statistics are given in Tables 2-2 and 2-3.

Structures of the two binary complexes were determined by molecular replacement with the program *Phaser* within the PHENIX software package (Adams et al., 2010) using the refined model of Rv2258c-unbound as a search model.

Table 2-3. Refinement statistics

<i>Data set</i>	<i>Rv2258c-unbound</i>	<i>Rv2258c-SFG</i>	<i>Rv2258c-SAH</i>
PDB code	5F8C	5F8F	5F8E
Resolution range (Å)	50.0–1.83	50.0–1.90	50.0–2.90
$R_{\text{work}} / R_{\text{free}}^a$ (%)	19.1 / 22.6	19.2 / 22.1	20.1 / 27.0
No. of non-hydrogen atoms / average <i>B</i> -factor (Å ²)			
Protein	8,030 / 33.1	8,000 / 37.9	7,933 / 47.2
Water oxygen	716 / 38.5	697 / 44.1	132 / 26.8
SFG	–	81 / 44.4	–
SAH	–	–	52 / 51.9
Glycerol	24 / 41.0	18 / 43.3	–
R.m.s. deviations from ideal geometry			
Bond lengths (Å)	0.007	0.009	0.010
Bond angles (°)	1.15	1.39	1.44
R.m.s. Z-scores ^b			
Bond lengths	0.438	0.490	0.486
Bond angles	0.589	0.638	0.657
Ramachandran plot (%) ^c			
Favored / Outliers	96.2 / 0.0	95.3 / 0.0	95.6 / 0.1
Poor rotamers (%) ^c	0.49	0.85	0.62

Footnotes to Table 2-3

^a $R_{\text{work}} = \Sigma | |F_{\text{obs}}| - |F_{\text{calc}}| | / \Sigma |F_{\text{obs}}|$, where R_{free} is calculated for a randomly chosen 5% of reflections, which were not used for structure refinement and R_{work} is calculated for the remaining reflections.

^b Values obtained using Refmac.

^c Values obtained using MolProbity.

2.2.4. Analytical gel filtration

The recombinant Rv2258c (6–353 construct) in the ligand-unbound state was subjected to analytical gel filtration chromatography on a Superdex 200 (10/300 GL) column, eluting with buffer C (20 mM Tris–HCl at pH 7.0 and 200 mM sodium chloride) at a flow rate of 0.5 ml min⁻¹. The standard curve was obtained using molecular-weight markers (Sigma MWGF200-1KT). Stokes radii of β -amylase, alcohol dehydrogenase, carbonic anhydrase, and cytochrome *c* were calculated from their crystal structures (PDB: 1FA2, 2HCY, 1V9E, and 1HRC, respectively) using the HYDROPRO program (García de la Torre et al., 2000).

2.3. Results

2.3.1. Structure determination and comparisons among monomer models of Rv2258c

In this study, I have determined the crystal structures of Rv2258c in three forms: (i) Rv2258c-unbound at 1.83 Å, (ii) Rv2258c-SFG at 1.90 Å, and (iii) Rv2258c-SAH at 2.90 Å (Table 2-2), using the (6–353) construct of Rv2258c from *M. tuberculosis* H37Rv, which is fused with hexahistidine-containing tags at both N- and C-termini. Three monomers (chains A–C) are present in each asymmetric unit of these crystals. Nine monomer models of Rv2258c in these structures contain all 348 residues of the Rv2258c construct (residues 6–353) but some residues of the N- and C-terminal hexahistidine-containing tags are disordered (Table 2-4). No ligand electron density is found in the active site of Rv2258c-unbound. One molecule of SFG is bound to each of chains A–C in Rv2258c-SFG. In Rv2258c-SAH, one molecule of SAH is bound to the active site in each of chains A and B, while no ligand is bound to chain C. Rv2258c does not show any significant conformational changes upon binding of SFG (or SAH). Therefore, chain A of Rv2258c-SFG is used to describe the overall structure, unless stated

otherwise.

When I make structural comparisons between any pair of Rv2258c monomer models, the root mean square (r.m.s.) deviations are in the range of 0.20–1.29 Å for 348 equivalent C α positions over the 36 pairwise comparisons. The minimal deviation is observed between chain C of Rv2258c-unbound and chain A of Rv2258c-SFG. The maximal deviation is between chain A of Rv2258c-unbound and chain B of Rv2258c-SAH. Structural variations of C α atoms larger than 2.0 Å occur within the two solvent-exposed loops (residues Gly99–Pro100 and Glu341–Val343), which do not form the co-substrate-binding pocket, and the solvent-exposed regions surrounding SFG (or SAH) [α 7 (Phe138–Ala152), a loop connecting β 3 and α 9 (Cys180–Arg184), and the α 10– β 5 segment (Ser204–Asp234)]. A similar pattern of variations is observed between chains A and B of Rv2258c-unbound and between chains A and B of Rv2258c-SFG. Therefore, the observed structural variations among the Rv2258c models are not correlated with binding of the co-substrate analogs.

Table 2-4. Modeled residues of Rv2258c structures

<i>Structure</i>	<i>Chain A (Mean B-factor)</i>	<i>Chain B (Mean B-factor)</i>	<i>Chain C (Mean B-factor)</i>
Rv2258c-unbound	M- Glu(6) -Lys(353)- LEHHH (34.2 Å ²)	HSSGLVPRGSHM- Glu(6) -Lys(353)-LEH (35.0 Å ²)	M- Glu(6) -Lys(353)- LEHHH (30.2 Å ²)
Rv2258c-SFG	M- Glu(6) -Lys(353)- LEHH (33.9 Å ²)	SSGLVPRGSHM- Glu(6) -Lys(353)-LE (41.0 Å ²)	M- Glu(6) -Lys(353)- LEHHH (38.9 Å ²)
Rv2258c-SAH	M- Glu(6) -Lys(353)- LEHH (50.5 Å ²)	Glu(6) -Lys(353)-LE (47.3 Å ²)	HM- Glu(6) -Lys(353)- LEHHH (43.6 Å ²)

The amino acid sequences of Rv2258c (residues 6–353) are denoted by bold-faced three-letter codes and the rest (histidine tags at N- and C-termini and methionine from restriction enzyme site) are denoted by single letter codes. Mean B-factors for all atoms are also given.

2.3.2. Overall monomer structure of Rv2258c

The monomer structure of Rv2258c comprises two distinct domains: (i) the N-terminal domain (residues Glu6–Tyr135 and Ser291–Ala308) and (ii) the C-terminal MTase domain (residues Ile155–Leu290 and Leu309–Lys353) (Figure 2-2A), which are linked by a long connecting α -helix ($\alpha 7$, residues Pro136–Leu154). The N-terminal domain of Rv2258c consists of seven α -helices ($\alpha 1$ – $\alpha 6$ and $\alpha 12$) and two β -strands ($\beta 1$ – $\beta 2$). $\alpha 12$ is inserted between $\beta 7$ and $\alpha 13$ of the C-terminal domain. The connecting $\alpha 7$ helix between the N-terminal and C-terminal domains participates in the formation of a putative binding site for a methyl-accepting substrate. The C-terminal domain of Rv2258c displays a typical Class I MTase fold. It is made up of a central seven-stranded β -sheet ($\beta 8 \uparrow$ - $\beta 9 \downarrow$ - $\beta 7 \uparrow$ - $\beta 6 \uparrow$ - $\beta 3 \uparrow$ - $\beta 4 \uparrow$ - $\beta 5 \uparrow$) that is flanked by five α -helices ($\alpha 8$ – $\alpha 11$ and $\alpha 13$) on both sides of the sheet (Figure 2-2B). Among different Classes of SAM-dependent MTases, a vast majority of known MTases belong to Class I, which has the Rossmann-like fold with a seven-stranded β -sheet adjoined by α -helices (Schubert et al., 2003).

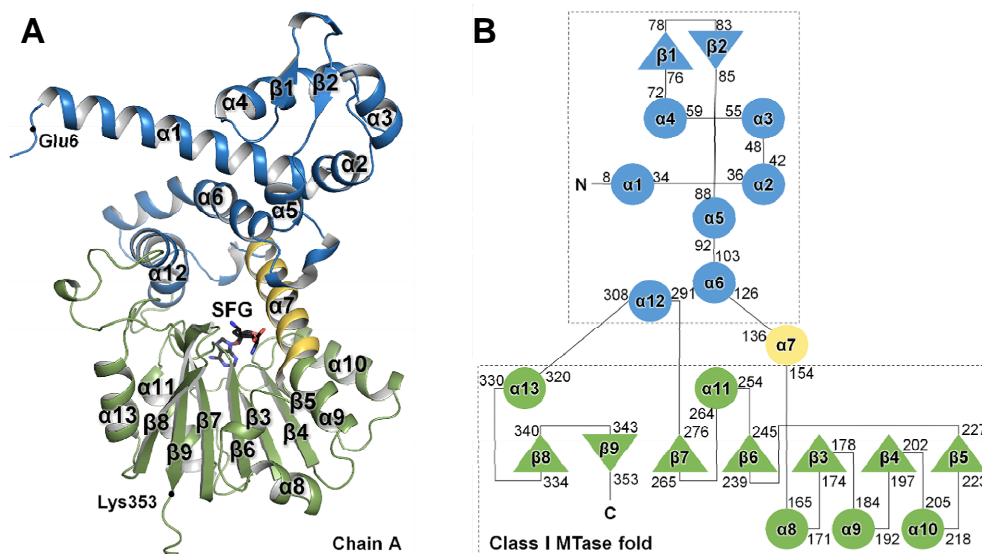


Figure 2-2. Monomer structure and topology diagram of Rv2258c

(A) Ribbon diagram of Rv2258c-SFG monomer (chain A). The secondary structure elements have been defined by the STRIDE program (Heinig and Frishman, 2004). The monomer is colored by structural domains: N-terminal domain (residues 6–135 and 291–308) in blue and C-terminal domain (residues 155–290 and 309–353) in green. The connecting α -helix 7 (residues 136–154) is colored in yellow. The bound SFG is shown in ball-and-stick representation and colored according to atom types (carbon, black; nitrogen, blue; and oxygen, red). (B) Topology diagram of the Rv2258c monomer fold, with α -helices and β -strands denoted by circles and triangles,

respectively. Secondary structures of the N-terminal domain, the connecting α -helix 7, and the Class I MTase fold are colored in blue, yellow, and green, respectively. Residue numbers are given for the secondary structure elements.

2.3.3. Rv2258c exists as a homodimer

Two chains A and B in the asymmetric unit of Rv2258c-unbound, Rv2258c-SFG, and Rv2258c-SAH are related by a non-crystallographic two-fold symmetry and they form an intertwined dimer around their N-terminal domains (Figure 2-3A) in a head-to-head fashion. Chain C and a neighboring chain in the other asymmetric unit (designated as C'), which are related to each other by a crystallographic two-fold symmetry, form a crystallographic dimer. These two kinds of dimer are structurally highly similar to each other with an r.m.s. deviation of 0.83 Å for 696 equivalent C α positions. The buried surface area at the interface between chains A and B are 3730, 3750, and 3710 Å² per monomer for Rv2258c-unbound, Rv2258c-SFG, and Rv2258c-SAH, respectively, as calculated by the PISA server (Krissinel and Henrick, 2007). The interface encompasses about 20% of the monomer surface area. Approximately 60% of these interface areas is contributed by the N-terminal domain. The extensive dimer interface is primarily composed of hydrophobic interactions involving alanine, isoleucine, leucine, phenylalanine, and valine (Figure 2-3B). The Stokes radius of the ligand-free Rv2258c in solution was estimated to be 3.93 nm by performing gel filtration experiments (Figure 2-3C). It agrees well with the calculated Stokes radii: 3.68 nm for the dimer between chains A and B

of Rv2258c-unbound or 3.70 nm for the dimer between chains C and C' of Rv2258c-unbound. This result supports that Rv2258c exists as a homodimer in solution. In the dimeric structure of Rv2258c, the putative methyl-accepting substrate-binding pocket is contributed by both monomers (helices $\alpha 6$ – $\alpha 7$ and $\alpha 12$ of one chain; helix $\alpha 1$ of the other chain). Therefore, the observed dimerization of Rv2258c is essential for determining the substrate specificity and for proper positioning of the methyl-accepting substrate.

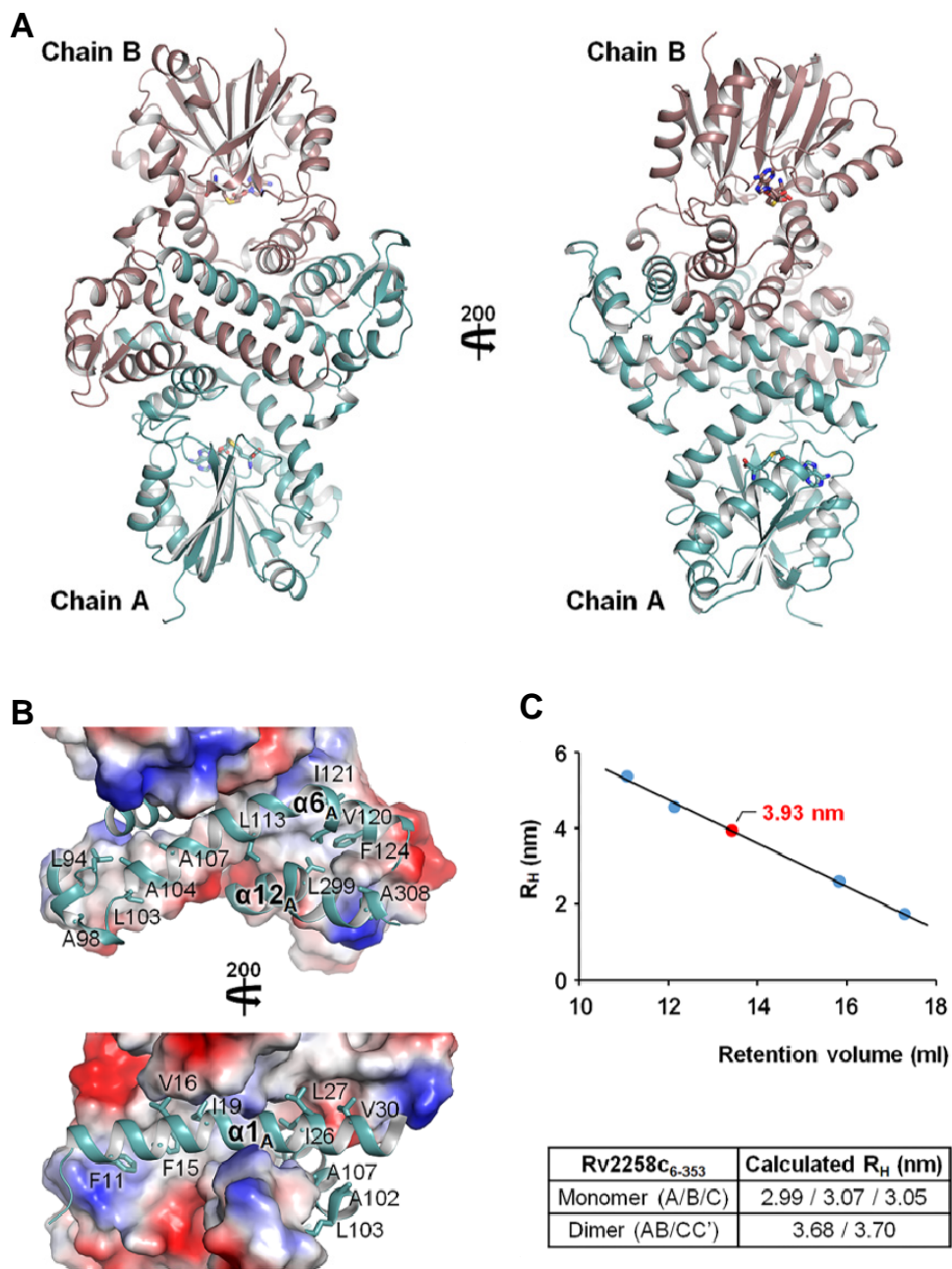


Figure 2-3. Dimeric structure and the oligomeric state of Rv2258c

(A) Ribbon diagram of an Rv2258c-SAH homodimer in two different orientations. Chains A and B are in deep teal and dark salmon, respectively, and the bound SAH molecules are shown as stick models. (B) Electrostatic potential surface diagram of chain B in the Rv2258c-SAH homodimer is colored in blue and red according to positive and negative potentials, respectively. Residues of chain A interacting with the other monomer (chain B) are shown as sticks. To show the detailed interactions more clearly, this view has a slightly different orientation from the right panel of (A). (C) Plot of the calculated Stokes radii (RH) against the retention volumes in size exclusion chromatography. Filled sky blue circles are molecular-weight markers: β -amylase, 200 kDa; alcohol dehydrogenase, 150 kDa; carbonic anhydrase, 29 kDa; cytochrome C, 12 kDa. The red filled circle is Rv2258c.

2.3.4. Structural similarity search

A DALI structural similarity search (Holm and Rosenström, 2010) with a monomer model of Rv2258c-unbound (chain A, residues 6-353) indicates that the overall fold of Rv2258c closely resembles those of Class I, small-molecule O-, N-, or C-MTases (Table 2-5; Figure 2-4). MTases displaying the highest *Z*-scores include: (i) caffeic acid *O*-MTase from *Lolium perenne* (LpOMT1; PDB: 3P9K, chain D; an r.m.s. deviation of 3.2 Å for 312 equivalent C α positions, a *Z*-score of 25.3, and a sequence identity of 19%), (ii) mitomycin 7-*O*-MTase from *Streptomyces lavendulae* (MmcR; PDB: 3GXO, chain A; an r.m.s. deviation of 3.6 Å for 319 equivalent C α positions, a *Z*-score of 25.5, and a sequence identity of 17%), (iii) carminomycin 4-*O*-MTase from *Streptomyces peucetius* (DnrK; PDB: 1TW3, chain B; an r.m.s. deviation of 3.6 Å for 317 equivalent C α positions, a *Z*-score of 25.0, and a sequence identity of 15%), and (iv) human N-acetyl serotonin MTase (ASMT; PDB: 4A6E, chain A; an r.m.s. deviation of 3.9 Å for 316 equivalent C α positions, a *Z*-score of 24.9, and a sequence identity of 15%). Similar results are obtained when I make a DALI structural similarity search with the dimer model of Rv2258c. As in Rv2258c, these MTases consist of a C-terminal SAM-dependent MTase domain and an N-terminal domain,

which is indispensable for the formation of an intertwined dimer and the cavity for binding the substrates.

The above structural homologs of Rv2258c are involved in the biosynthesis of various metabolites such as lignin, mitomycin, daunorubicin, and melatonin. Lignin is a heterogeneous polymeric macromolecule and a major component of the cell wall in vascular plants. It provides a mechanical support for plant tissues and protects the plant from pathogen invasion (Louie et al., 2010). Mitomycin is a quinone-containing antibiotic isolated from *S. caespitosus* or *S. lavendulae* and is used in antitumor chemotherapy (Singh et al., 2011). Daunorubicin isolated from *S. peucetius* is also used in the chemotherapy of some cancers (Jansson et al., 2004). Melatonin is a multi-tasking molecule found in animals, plants, fungi and bacteria, and is involved in various physiological functions such as sleep induction and circadian rhythm regulation (Botros et al., 2013). Interestingly, many substrates of these structurally similar MTases contain aromatic rings. It is plausible that the natural substrate of Rv2258c might be a small molecule with aromatic rings. However, the primary substrate of Rv2258c cannot be inferred from the structural similarity alone, as the sequence motifs involved in binding the methyl-accepting substrate are not highly conserved among Class I MTases.

Table 2-5. Structural homologs of Rv2258c as identified by DALI search

Abbreviated Name	PDB_Chain	Z-score	Sequence identity (%)	Metabolic pathway	Acceptor atom type	MTase fold
LpOMT1	3P9K_D	25.5	19	Lignin	O	Class I
MmcR	3GXO_A	25.5	17	Mitomycin	O	Class I
DnrK	1TW3_B	25.0	15	Daunorubicin	O	Class I
ASMT	4A6E_A	24.9	15	Melatonin	N	Class I
CalO1	3LST_A	23.2	17	Calicheamicin	O	Class I
NesB1	3I5U_B	23.2	17	Neocarzinostatin	O	Class I
MppJ	4M6X_A	22.7	17	Mannoheptimycin	C	Class I
PhzM	2IP2_A	22.2	16	Pyocyanin	N	Class I

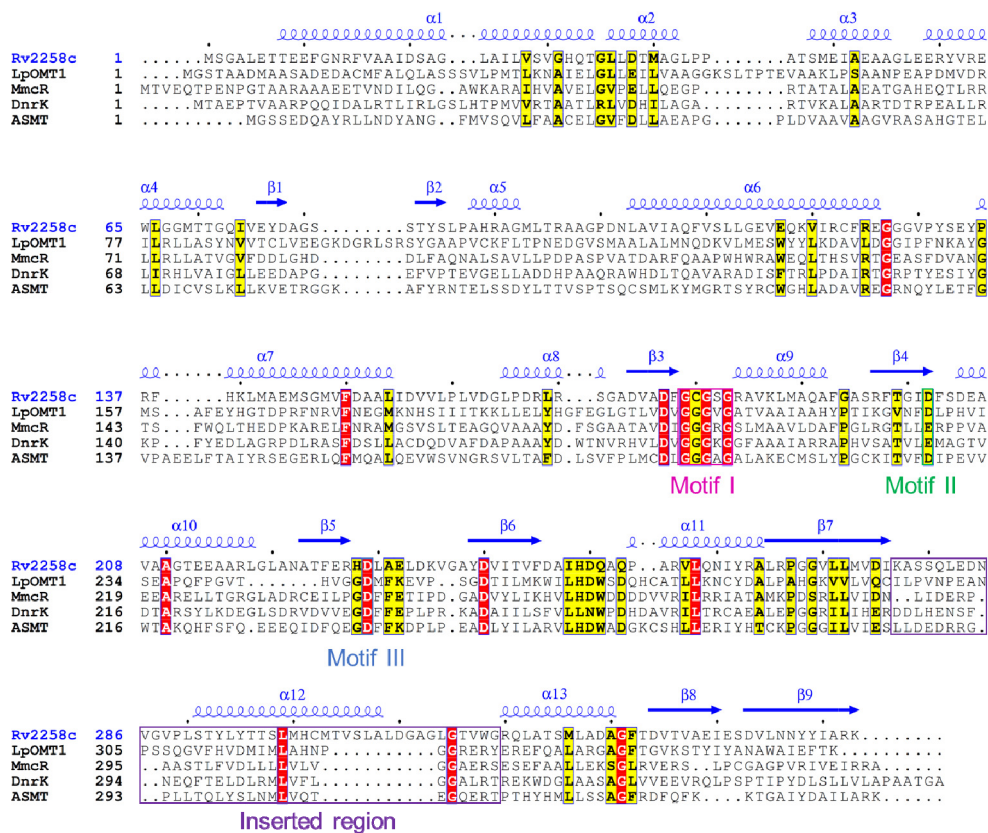


Figure 2-4. Sequence alignment of Rv2258c with its close structural homologs

Sequence alignment of Rv2258c (SWISS-PROT accession code: O53532) and structurally similar MTases: LpOMT1 (caffeic acid *O*-MTase; SWISS-PROT accession code: Q9ZTU2), MmcR (mitomycin 7-*O*-MTase; SWISS-PROT accession code: Q9X5T6), DnrK (carminomycin 4-*O*-MTase; SWISS-PROT accession code: Q06528), and ASMT (human *N*-acetyl

serotonin MTase; SWISS-PROT accession code: P46597). Every tenth residue is marked with a black dot above the sequence of Rv2258c. The secondary structures of Rv2258c are colored in blue. Strictly conserved residues are highlighted with a red background. Semi-conserved residues are highlighted with a yellow background. Conserved sequence motifs (Motifs I–III) and the inserted region are boxed. Sequences were aligned using the program ClustalX (Larkin et al., 2007) and the figure was generated with the program ESPript (Robert and Gouet, 2014).

2.3.5. MTase sequence motifs and SFG (or SAH) binding

Up to six sequence motifs (I–VI) are used to identify MTases (Kozbial and Mushegian, 2005; Liscombe et al., 2012; Martin and McMillan, 2002). One or more residues in “GxGxG” motif at the end of the first β -strand of the Class I MTase core fold, the highly-conserved Motif I, bind the adenosyl part of SAM. Another strongly conserved acidic residue in Motif II at the end of the second β -strand forms hydrogen bonds with the ribose part of SAM. Motif III spans the third β -strand, followed by Motif IV spanning the fourth strand and the adjoining loops. Both Motifs III and IV include a partially-conserved acidic residue, and interact with the adenosyl part and the sulfonium part of SAM, respectively. In some Class I MTases, hydrophobic residues located on a helix between the fourth and fifth β -strands (Motif V) make hydrophobic interactions with the adenosyl part of SAM (Kozbial and Mushegian, 2005; Liscombe et al., 2012; Martin and McMillan, 2002).

As in other MTase-SAM complexes, Rv2258c interacts with SFG (or SAH) in an extended conformation maintained by a network of hydrogen bonds and stacking interactions. The SFG (or SAH) binding site of Rv2258c is formed by the four loops: $\beta 3$ – $\alpha 9$, $\beta 4$ – $\alpha 10$, $\beta 5$ – $\beta 6$, and $\beta 6$ – $\alpha 11$ loops. Motifs I, II, and III are conserved in Rv2258c and are responsible for the

interaction with SFG (or SAH). The main chain atoms of Gly179 and Gly181 of the GxGxG motif (residues 179–183; Motif I), the most important SAM-binding motif, are located in the β 3– α 9 loop and are hydrogen-bonded to the carboxylate moiety of SFG (or SAH) (Figure 2-5, A and B). The carboxylate moiety of SFG (or SAH) makes additional hydrogen bonds with the side chain of Ser146 and the main chain of Phe245. Motif II of Rv2258c is a conserved acidic residue Asp202, which lies at the C-terminus of the β -strand β 4. It forms hydrogen bonds with the ribosyl moiety of SFG (or SAH) (Figure 2-5, A and B). The main chain atoms of Phe203 and Leu230 and the side chain of Gln251 make hydrogen bonds with the adenine ring of SFG (or SAH) as well as the side chain of Asp229 (Motif III). The side chain of Phe203 forms a stacking interaction with the adenine ring. Tyr132 (side chain), Phe150 (side chain), and His228 (main chain) interact with the ribosyl moiety, the carboxylate moiety, and the adenine ring of SFG (or SAH), respectively.

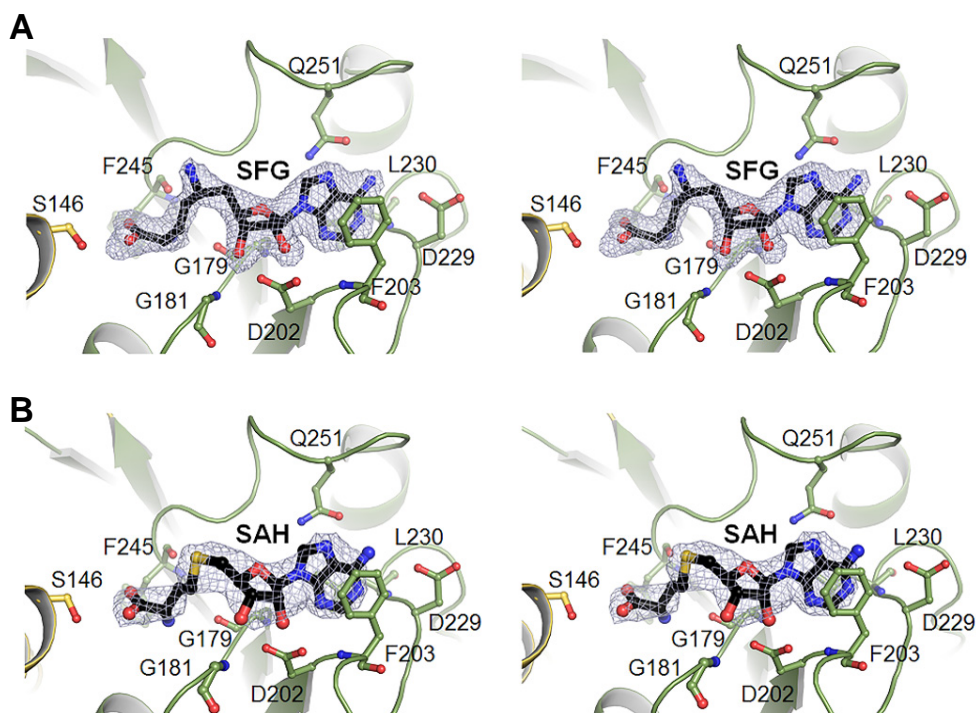


Figure 2-5. Interactions of SFG (or SAH) with Rv2258c

Stereo views of the active site region in Rv2258c-SFG (A) and Rv2258c-SAH (B), with SFG and SAH in black. Most of the interacting residues belong to the C-terminal MTase domain in chain A, except Ser146 from the connecting α -helix 7 of the same chain. The omit *mFo*-*DFc* maps for bound ligands in Rv2258c-SFG (A) or Rv2258c-SAH (B) are shown in blue white colored mesh contoured at a level of 2.5σ . Hydrogen bonds are indicated by black dotted lines. For clarity, only selected side chains and main chains of

interacting residues are shown. Tyr132 (side chain), Phe150 (side chain), and His228 (main chain) are in close proximity to SFG (or SAH) but are not shown, because they block the view of SFG (or SAH).

2.3.6. Rv2258c is a unique SAM-dependent MTase in mycobacteria

Until now, three-dimensional structures of eleven mycobacterial SAM-dependent MTases have been determined. They fall into different functional groups: (i) lipid MTases [Rv0470c (PDB: 1L1E; Huang et al., 2002), Rv0503c (PDB: 1KPI; Huang et al., 2002), Rv0642c (PDB: 2FK8; Boissier et al., 2006), Rv0644c (PDB: 1TPY), Rv3392c (PDB:1KP9; Huang et al., 2002)], (ii) nucleic acid MTases [Rv2118c (PDB: 1I9G; Gupta et al., 2001), Rv2372c (PDB: 4L69; Kumar et al., 2014), Rv2966c (PDB: 3P9N; Sharma et al., 2015), and MAB_3226c (PDB: 3QUV)], (iii) a small-molecule MTase [*M. smegmatis* EgtD (MSMEG_6247; PDB: 4UY6; Jeong et al., 2014)], and (iv) a MTase of unknown substrate [*M. leprae* ML2640c (PDB: 2UYO; Graña et al., 2007)]. Whereas Rv2372c and MAB_3226c are Class IV MTases, the remainders are Class I MTases. Class I and Class IV MTases have unrelated folds. Therefore, Rv2258c seems to be distinct from these structurally-characterized SAM-dependent MTases in mycobacteria.

As mentioned above, the overall monomer fold of Rv2258c closely resembles those of small-molecule *O*-, *N*-, or *C*-MTases of Class I. Therefore, I make some detailed comparisons between Rv2258c and *M. smegmatis* EgtD, which trimethylates the nitrogen atom of histidine to form hercynine in ergothioneine biosynthesis. Ergothioneine is an antioxidant

secreted by mycobacteria presumably for protection from oxidative stresses (Jeong et al., 2014). Although catalytic domains of both Rv2258c and EgtD share the Class I MTase fold, they exhibit significantly different structural features (Figure 2-6). The N-terminal domain of Rv2258c ($\alpha 1$ – $\alpha 6$ and $\beta 1$ – $\beta 2$; residues 1–135) is much larger than that of EgtD ($\alpha 1$ – $\alpha 2$; residues 11–55). In contrast, the insertion of Rv2258c [helix $\alpha 12$ (residues 291–308) between $\beta 7$ and $\alpha 13$] is much smaller than that of EgtD ($\alpha 8$ – $\alpha 9$ and $\beta 7$ – $\beta 11$ between $\beta 6$ and $\alpha 10$; residues 2–41 and 198–286). Furthermore, Rv2258c and EgtD differ in their oligomeric states. I have shown that Rv2258c exists as a homodimer, whereas EgtD is a monomeric enzyme (Jeong et al., 2014).

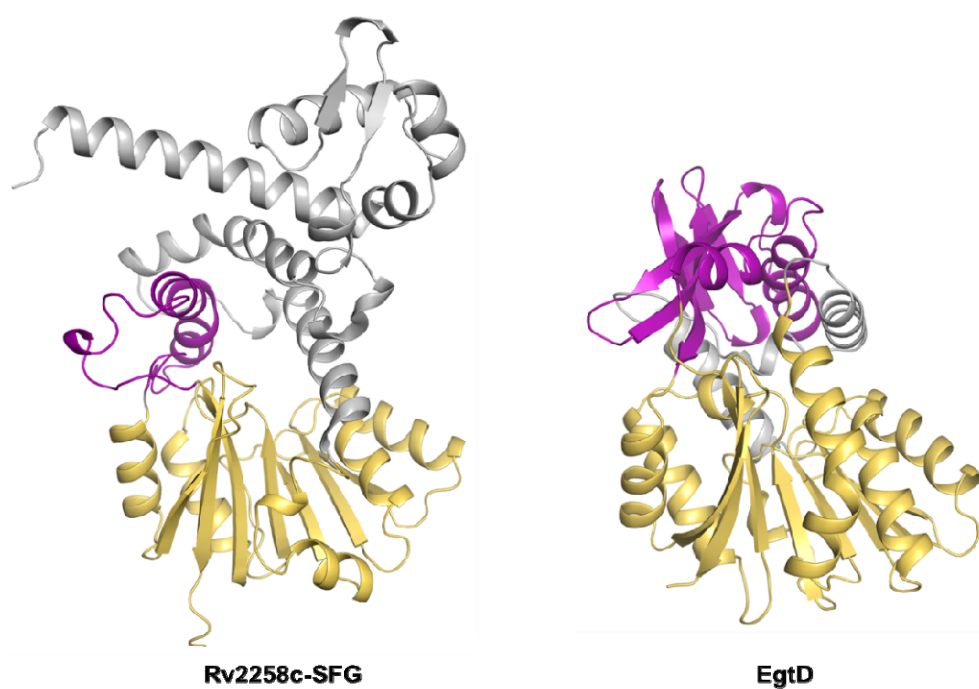


Figure 2-6. Comparison of monomer structures of Rv2258c-SFG and *M. smegmatis* EgtD

Ribbon diagram of an Rv2258c-SFG monomer (chain A) is colored as follows: the N-terminal domain (residues 6–154) and the MTase domain (residues 155–276 and 320–353) are colored in grey and yellow, respectively. The inserted region (residues 277–319) in MTase domain of Rv2258c is colored in purple. Ribbon diagram of an *M. smegmatis* EgtD monomer (chain A; PDB code 4UY6) is colored as in the Rv2258c-SFG monomer: the N-terminal addition (residues 11–55), the MTase domain

(residues 2–10, 56–197, and 287–321), and the inserted region in MTase domain (residues 198–286).

2.3.7. Relatively large substrate-binding cavity of Rv2258c and functional implications

Rv2258c functions as a homodimer, which possesses two large cavities of approximately equal volume. The cavity volume in the dimer models of ligand-free Rv2258c, i.e., dimers between chains A and B of Rv2258c-unbound, between chains C and C' of Rv2258c-unbound, and chains C and C' of Rv2258c-SAH, ranges between 1,560 and 1,860 Å³ with an average of 1,770 Å³, as calculated by the KVFinder program (Oliveira et al., 2014). The cavity volume for ligand-bound Rv2258c dimers between chains A and B of Rv2258c-SFG, between chains C and C' of Rv2258c-SFG, and between chains A and B of Rv2258c-SAH, ranges between 530 and 740 Å³ with an average of 680 Å³, which is available for binding the methyl-accepting substrate. The inside of this cavity for the methyl-accepting substrate in Rv2258c is lined with mainly hydrophobic residues largely from one monomer such as proline, leucine, valine, isoleucine, phenylalanine, and methionine. Val16 in helix α 1 from another monomer of the dimer also contributes to this cavity surface. The cavity volumes for binding both donor- and acceptor-substrates are 730 Å³ for LpOMT1, 900 Å³ for MmcR, 1260 Å³ for DnrK, and 850 Å³ for ASMT, respectively. The cavity volumes for binding a methyl-accepting substrate are 360 Å³ for

LpOMT1, 560 Å³ for MmcR, 910 Å³ for DnrK, and 420 Å³ for ASMT, respectively (Fig. 4), which correlate well with the molecular volume of their methyl-accepting substrates (220, 410, 700, and 270 Å³ for coniferaldehyde bound to LpOMT1, mitomycin A in MmcR, 4-methoxy- ϵ -rhodomycin T in DnrK, and N-acetylserotonin in ASMT, respectively). Except DnrK, the structural homologs of Rv2258c have smaller cavity volumes for binding a methyl-accepting substrate. Most of these substrates fall into aromatic ring-containing compounds (Figure 2-7). If Rv2258c functions as a small-molecule MTase, I could infer that a natural substrate of Rv2258c might be akin to bulky and nonpolar compounds with aromatic rings.

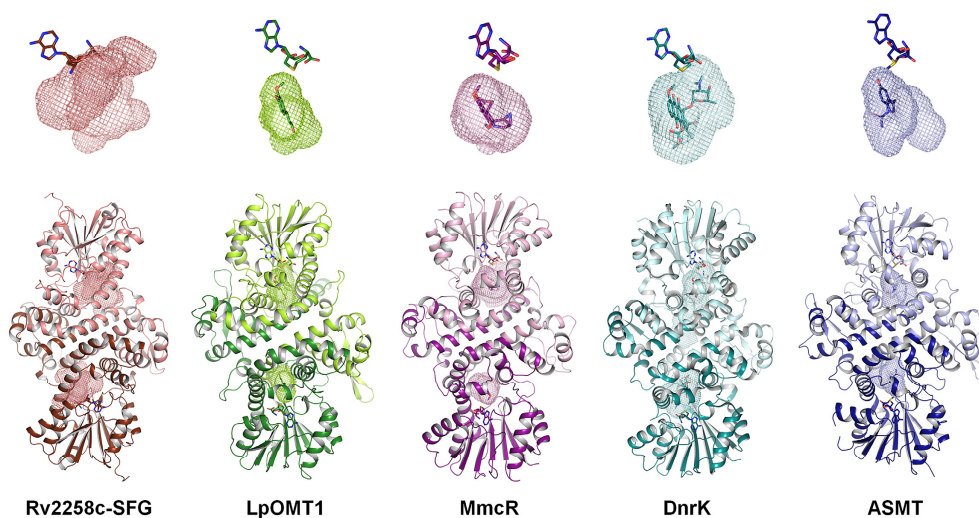


Figure 2-7. Comparison of homodimeric structures of Rv2258c and its close structural homologs

Comparison of dimeric structures of Rv2258c-SFG, LpOMT1, MmcR, DnrK, and ASMT (lower panel). A methyl donor (or its analog) and a methyl-accepting substrate (when present) are shown in sticks, together with the cavity for binding a methyl-accepting substrate (upper panel). Ribbon diagrams of the dimeric structure are colored as follows: Rv2258c-SFG (chain C in chocolate and chain C' in salmon), LpOMT1 (chain A in forest and chain B in limon), MmcR (chain A in deep purple and chain B in light pink), DnrK (chain A in deep teal and chain B in pale cyan), and ASMT (chain A in deep blue and chain A' in light blue) (bottom panel). The

cavities are shown as mesh and are colored in lighter color of the two monomers, respectively. The depicted cavities correspond to the volume for binding the methyl-accepting substrates. The cavities are calculated by the KVFinder program (Oliveira et al. 2014) using the default values of step size (0.6 Å), probe in size (1.4 Å), and probe out size (4.0 Å).

Another potential clue to the function of Rv2258c may be the location (2,530,836 – 2,531,897 bp) of its gene in the genome of *M. tuberculosis* H37Rv. Interestingly, *Rv2257c*, the gene encoding a potential β -lactamase, is located adjacent to the Rv2258c with only a 14-bp interval (at 2,530,004 – 2,530,822 bp). Indeed, these two genes are predicted to lie within the same operon by the ProOpDB site (<http://operons.ibt.unam.mx/OperonPredictor/>) (Taboada et al., 2012) and the TB database site (<http://www.broadinstitute.org/annotation/genome/tbdb/>) (Galagan et al., 2010). Mb2281c from *M. bovis* is identical in its amino acid sequence to that of Rv2257c and the crystal structure of Mb2281c (PDB: 3I7J, unpublished deposition) confirms that it has a β -lactamase fold. Additionally, the subcellular location prediction servers, including CELLO (Yu et al., 2006), SubLoc (Hua and Sun, 2001), and PSORTb (Yu et al., 2010), indicate that Rv2258c may exist in the cytoplasm of mycobacteria. Taken together, it seems that Rv2258c may be expressed together with the putative β -lactamase Rv2257c and it may methylate unknown small molecules in the cytoplasm, when *M. tuberculosis* is subject to harsh environments like the presence of antibiotics. One conceivable scenario is that Rv2258c might be involved in conferring drug resistance to *M. tuberculosis*. The present structural work provides useful, although limited,

insights into the function of Rv2258c, while further studies are required to establish its precise biological role.

2.4. Conclusions

To date, over 300 members of the MTase family (EC 2.1.1.-) have been classified on the basis of their substrate specificity (small molecule, lipid, protein, DNA, and RNA) and on the atom type targeted for methylation (oxygen, carbon, nitrogen, sulfur) (Liscombe et al., 2012). In addition, most of them are divided into five Classes (Classes I–V) based on the overall fold of their catalytic domain (Schubert et al., 2003). In this study, I have determined the crystal structure of Rv2258c, a SAM-dependent MTase in *M. tuberculosis*. I have found that Rv2258c exists as a homodimer both in the crystal and in the solution, indicating that Rv2258c functions as a dimer. Rv2258c is structurally similar to small-molecule *O*-MTases with nonpolar substrates with aromatic rings. However, further experiments are required to establish the natural substrate and the biological role of Rv2258c, because small-molecule MTases do not possess widely conserved structural determinants for methyl-accepting substrate recognition.

2.5. References

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. sect. D Biol. Crystallogr.* **66**:213–221.
- Boissier, F., Bardou, F., Guillet, V., Uttenweiler-Joseph, S., Daffé, M., Quémar, A., Mourey, L. (2006) Further insight into S-adenosylmethionine-dependent methyltransferases: structural characterization of Hma, an enzyme essential for the biosynthesis of oxygenated mycolic acids in *Mycobacterium tuberculosis*. *J. Biol. Chem.* **281**:4434–4445.
- Botros, H.G., Legrand, P., Pagan, C., Bondet, V., Weber, P., Ben-Abdallah, M., Lemièrre, N., Huguet, G., Bellalou, J., Maronde, E., Beguin, P., Haouz, A., Shepard, W., Bourgeron, T. (2013) Crystal structure and functional mapping of human ASMT, the last enzyme of the melatonin synthesis pathway. *J. Pineal Res.* **54**:46–57.

- Brünger, A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**:472–475.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. sect. D Biol. Crystallogr.* **66**:12–21.
- Cole, S.T., Brosch, R., Parkhill, 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., Barrell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544.
- Diederichs, K., Karplus, P.A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**:269–275.

- Diederichs, K., Karplus, P.A. (2013) Better models by discarding data? *Acta Crystallogr. sect. D Biol. Crystallogr.* **69**:1215–1222.
- Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. sect. D Biol. Crystallogr.* **66**:486–501.
- Galagan, J.E., Sisk, P., Stolte, C., Weiner, B., Koehrsen, M., Wymore, F., Reddy, T.B., Zucker, J.D., Engels, J.S. (2010) TB database 2010: overview and update. *Tuberculosis (Edinb)* **90**:225–235.
- García, de la Torre, J., Huertas, M. L., Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**:719–730.
- Graña, M., Haouz, A., Buschiazzo, A., Miras, I., Wehenkel, A., Bondet, V., Shepard, W., Schaeffer, F., Cole, S.T., Alzari, P.M. (2007) The crystal structure of *M. leprae* ML2640c defines a large family of putative S-adenosylmethionine-dependent methyltransferases in mycobacteria. *Protein Sci.* **16**:1896–1904.
- Gupta, A., Kumar, P.H., Dineshkumar, T.K., Varshney, U., Subramanya, H.S. (2001) Crystal structure of Rv2118c: an AdoMet-dependent methyltransferase from *Mycobacterium tuberculosis* H37Rv. *J. Mol. Biol.* **312**:381–391.

- Heinig, M., Frishman, D. (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**:W500–W502.
- Holm, L., Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**:W545–W549.
- Hua, S., Sun, Z. (2001) Support vector machine approach for protein subcellular localization prediction. *Bioinformatics* **17**:721–728.
- Huang, C.C., Smith, C.V., Glickman, M.S., Jacobs, W.R. Jr., Sacchettini, J.C. (2002) Crystal structures of mycolic acid cyclopropane synthases from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **277**:11559–11569.
- Jansson, A., Koskiniemi, H., Mäntsälä, P., Niemi, J., Schneider, G. (2004) Crystal structure of a ternary complex of DnrK, a methyltransferase in daunorubicin biosynthesis, with bound products. *J. Biol. Chem.* **279**:41149–41156.
- Jeong, J.H., Cha, H.J., Ha, S.C., Rojviriyaya, C., Kim, Y.G. (2014) Structural insights into the histidine trimethylation activity of EgtD from *Mycobacterium smegmatis*. *Biochem. Biophys. Res. Commun.* **452**:1098–1103.
- Kozbial, P.Z., Mushegian, A.R. (2005) Natural history of S-adenosylmethionine-binding proteins. *BMC Struct. Biol.* **5**:19.

- Krissinel, E., Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**:774–797.
- Kumar, A., Kumar, S., Taneja, B. (2014) The structure of Rv2372c identifies an RsmE-like methyltransferase from *Mycobacterium tuberculosis*. *Acta Crystallogr. sect. D Biol. Crystallogr.* **70**:821–832.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947–2948.
- Lew, J.M., Kapopoulou, A., Jones, L.M., Cole, S.T. (2011) TubercuList - 10 years after. *Tuberculosis (Edinb)* **91**:1–7.
- Lin, P.L., Flynn, J.L. (2010) Understanding latent tuberculosis: a moving target. *J. Immunol.* **185**:15–22.
- Liscombe, D.K., Louie, G.V., Noel, J.P. (2012) Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat. Prod. Rep.* **29**:1238–1250.
- Louie, G.V., Bowman, M.E., Tu, Y., Mouradov, A., Spangenberg, G., Noel, J.P. (2010) Structure-function analyses of a caffeic acid O-methyltransferase from perennial ryegrass reveal the molecular basis for substrate preference. *Plant Cell* **22**:4114–4127.

- Martin, J.L., McMillan, F.M. (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12**:783–793.
- Murshudov, G.N., Vagin, A.A., Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. sect. D Biol. Crystallogr.* **53**:240–255.
- Oliveira, S.H., Ferraz, F.A., Honorato, R.V., Xavier-Neto, J., Sobreira, T.J., de Oliveira, P.S. (2014) KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC Bioinformatics* **15**:197.
- Otwinowski, Z., Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**:307–326.
- Robert, X., Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**:W320–W324.
- Schubert, H.L., Blumenthal, R.M., Cheng, X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* **28**:329–335.
- Sharma, G., Upadhyay, S., Srilalitha, M., Nandicoori, V.K., Khosla, S. (2015) The interaction of mycobacterial protein Rv2966c with host

- chromatin is mediated through non-CpG methylation and histone H3/H4 binding. *Nucleic Acids Res.* **43**:3922–3937.
- Singh, S., Chang, A., Goff, R.D., Bingman, C.A., Grüşchow, S., Sherman, D.H., Phillips, G.N., Jr., Thorson, J.S. (2011) Structural characterization of the mitomycin 7-O-methyltransferase. *Proteins* **79**:2181–2188.
- Taboada, B., Ciria, R., Martinez-Guerrero, C.E., Merino, E. (2012) ProOpDB: Prokaryotic Operon DataBase. *Nucleic Acids Res.* **40**:D627–D631.
- Terwilliger, T.C. (2003) Automated main-chain model-building by template-matching and iterative fragment extension. *Acta Crystallogr. sect. D Biol. Crystallogr.* **59**:38–44.
- Weiss, M.S. (2001) Global indicators of X-ray data quality. *J. Appl. Cryst.* **34**:130–135.
- Yu, C.S., Chen, Y.C., Lu, C.H., Hwang, J.K. (2006) Prediction of protein subcellular localization. *Proteins* **64**:643–651.
- Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J., Brinkman, F.S. (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined

localization subcategories and predictive capabilities for all prokaryotes.

Bioinformatics **26**:1608–1615.

국문초록

Structural Studies of Csd6 Protein from *Helicobacter pylori* and Rv2258c Protein from *Mycobacterium tuberculosis*

헬리코박터 파일로리균 Csd6 와 결핵균 Rv2258c 단백질의 구조 연구

헬리코박터 파일로리균은 위암을 비롯한 위장관 질병을 일으킨다. 결핵균은 결핵을 일으키는데, 전 세계적으로 매년 수백만 명의 목숨을 앗아가고 있다. 약제 내성을 갖는 병원성 박테리아의 증가는 세계적으로 심각한 문제이다. 따라서 새로운 항미생물제의 발견이 시급하다. 병원성 박테리아 단백질의 분자적, 생물학적 기능을 밝히는 것은 새로운 항미생물제 개발에 초석을 제공한다. 본 연구에서는 헬리코박터 파일로리균 Csd6 단백질과 결핵균 Rv2258c 단백질의 삼차원 구조를 규명하였다.

헬리코박터 파일로리균 Csd6 단백질은 펩티도글리칸 펩타이드를 자름으로써 나선형의 세포 모양 결정하는데 중요한 역할을 한다. Csd6 는 또한 flagellar FlaA 단백질의 탈당화 반응에 관여한다고 알려져 있다. Csd6 의 구조는 Csd6 의 가운데에 위치한 활성 도메인이 L,D-transpeptidases 와 유사하나 그 활성 부위의 형태가 지금까지 알려진 L,D-transpeptidases 과는 다른 특이성을 갖는다는 것을 보였다. 질량 분석도 Csd6 가 L,D-transpeptidase 가 아닌 단지 L,D-carboxypeptidase 로서 기능한다는 것을 확인했다. D-Ala 가 결합된 구조를 통해 활성 부위에 기질과 생성물의 결합 모드를 보였다. 구조-기능 분석을 기반으로 하여,

Csd6 와 Csd6 유사단백질들이 새로운 L,D-carboxypeptidase 로 분류되어야 한다고 제안했다.

결핵균 Rv2258c 단백질은 S-adenosyl-L-methionine-dependent methyltransferase (MTase)로 예상되었다. MTase 는 거의 모든 생물들의 세포 신호, 대사물질 합성, 유전자 조절 등과 같은 다양한 세포 과정에 관여하고 있다. Rv2258c 의 구조는 Rv2258c 단량체가 두 개의 도메인이 하나의 긴 α -helix 에 의해 연결되어 있음을 보였다. Rv2258c 의 아미노 말단 도메인은 이량체를 이루는데 필요하며 카르복실기 말단 활성 도메인은 Class I 의 MTase fold 를 가지고 있다. 결합된 sinefungin (or S-adenosyl-L-homocysteine)과의 상호작용뿐만 아니라 Rv2258c 의 전체 fold 는 small-molecule MTases 과 유사하다. Rv2258c 는 메틸기를 받는 기질이 결합할 수 있는 상대적으로 큰 소수성의 공간을 가지고 있으며, 이를 통해 부피가 크고 극성이 없는 분자가 결핵균의 Rv2258c 에 의해 메틸화될 것임을 제시하였다.

주요어: Csd6 / cell shape / L,D-carboxypeptidase / *Helicobacter pylori* / HP0518 / flagellin / peptidoglycan / cell motility / protein structure / structure-function / Rv2258c / *Mycobacterium tuberculosis* / small-molecule methyltransferase / sinefungin / S-adenosyl-L-homocysteine

학번: 2012-30719