Structural studies of the RPFF protein involved in Quorum Sensing and the IGHMBP2 protein involved in Spinal Muscular Atrophy with Respiratory Distress type 1

Lim, Siew Choo

2012

Lim, S. C. (2012). Structural studies of the RPFF protein involved in Quorum Sensing and the IGHMBP2 protein involved in Spinal Muscular Atrophy with Respiratory Distress type 1. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/50220>

<https://doi.org/10.32657/10356/50220>

**STRUCTURAL STUDIES OF THE RPFF PROTEIN
INVOLVED IN QUORUM SENSING
AND
THE IGHMBP2 PROTEIN INVOLVED IN SPINAL
MUSCULAR ATROPHY WITH RESPIRATORY
DISTRESS TYPE 1**

LIM SIEW CHOO

SCHOOL OF BIOLOGICAL SCIENCES

2012

**STRUCTURAL STUDIES OF THE RPFF PROTEIN INVOLVED IN
QUORUM SENSING AND THE IGHMBP2 PROTEIN INVOLVED IN
SPINAL MUSCULAR ATROPHY WITH RESPIRATORY DISTRESS TYPE 1**

LIM SIEW CHOO

2012

**STRUCTURAL STUDIES OF THE RPFF PROTEIN
INVOLVED IN QUORUM SENSING
AND
THE IGHMBP2 PROTEIN INVOLVED IN SPINAL
MUSCULAR ATROPHY WITH RESPIRATORY
DISTRESS TYPE 1**

LIM SIEW CHOO

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2012

Acknowledgements

I will like to thank my supervisor, Dr Song Haiwei, for giving me the great opportunity to work with him. His invaluable guidance, patience and insightful advice has allowed me to grow and increase my ability as a researcher.

I will also like to thank current and past members of SHW lab for their kind help, support and friendship and the many knowledge exchanges over the years. Many thanks also go to collaborators of SHW lab, Dr Zhang Liang Hui and Dr He Yawen of the Institute of Molecular and Cellular Biology (IMCB, Singapore) for their great experimental data and insightful discussions.

I will like to thank the beamline scientists at the European Synchrotron Radiation Facility (ESRF, France) and National Synchrotron Radiation Research Centre (NSRRC, Taiwan) for assistance and access to synchrotron radiation facilities. Thanks also go to the shared facilities in IMCB for experimental supports including DNA sequencing facility and X-ray crystallography facility.

I am grateful to the Agency for Science, Technology and Research (A*STAR, Singapore) for awarding me with a scholarship that has provided the financial support for my graduate study. I will also like to acknowledge that the work presented in the thesis is financially supported by the Biomedical Research Council (BMRC, A*STAR).

Most importantly, I will like to thank my parents and family for their unwavering encouragement and support despite being clueless about my work.

Table of contents

Acknowledgements	i
Table of Contents.....	ii
Summary	vii
Lists of Abbreviations	ix
Lists of Figures.....	xiii
Lists of Tables	xv

Part One:

Structural studies on the RpfF protein involved in quorum sensing

Chapter 1 Introduction	2
1.1 <i>Xanthomonas campestris</i> pv. <i>campestris</i> and black rot of crucifers	2
1.2 Microbial quorum sensing	4
1.3 Quorum sensing and virulence factor production in <i>Xcc</i>	4
1.4 Hydratase/ Isomerase superfamily	5
1.5 Objectives and scope of this study	7
Chapter 2 Materials and Methods	8
2.1 Molecular cloning	8
2.1.1 Polymerase chain reaction and analysis	8
2.1.2 Construct sub-cloning	9
2.1.3 DH5 α cell transformation and colony PCR	10
2.2 Protein expression	11
2.2.1 BL21(DE3)-star cell transformation and protein expression	11
2.2.2 Large scale bacterial cell culture	12
2.3 Protein purification.....	13
2.3.1 Protein purification buffers	13

2.3.2	Cell lysis	14
2.3.3	TALON metal affinity chromatography	14
2.3.4	Desalting and gel filtration chromatography	14
2.4	Crystallization.....	15
2.5	Data collection, structure determination and refinement	15
2.6	Isothermal titration calorimetry binding assay	16
2.7	DSF production assay (performed by Dr He Yawen at IMCB, A*STAR)	17
Chapter 3	Results	18
3.1	Protein expression and purification	18
3.2	The quality of the structure	20
3.3	Description of overall structure	23
3.4	RpfF belongs to the enoyl-coA hydratase/isomerase family	25
3.5	RpfF as an enoyl-CoA hydratase	28
3.6	Active site of RpfF	32
Chapter 4	Discussion	36
4.1	Characterizing RpfF	36
4.2	Implication in quorum sensing	37
Chapter 5	Future Directions	39

Part Two:
Structural studies on the Ighmbp2 protein involved in Spinal Muscular Atrophy with Respiratory Distress Type 1

Chapter 6	Introduction	41
6.1	Spinal muscular atrophy with respiratory distress type 1 (SMARD1)	41
6.2	Helicase superfamily	46
6.3	Superfamily 1 helicases	48
6.3.1.	Structures and mechanisms of SF1 helicases.....	48
6.3.2.	Helicases associated with neurodegenerative diseases	50
6.4	Domain organization of Ighmbp2	54
6.5	Biological significance of Ighmbp2	56
6.6	Objectives and scope of this study	57
Chapter 7	Materials and Methods	58
7.1	Molecular cloning	58
7.1.1	Polymerase chain reaction and analysis	58
7.1.2	Construct sub-cloning	59
7.1.3	DH5 α cell transformation and colony PCR	60
7.2	Protein expression	60
7.2.1	Rosetta(DE3)pLysS cell transformation and protein expression	60
7.2.2	Large scale bacterial cell culture	60
7.3	Protein purification	61
7.3.1	Protein purification buffers for hIghmbp2hd	61
7.3.2	Protein purification buffers for mIghmbp2hd-R3H	61
7.3.3	Cell lysis	61
7.3.4	Glutathione Sepharose 4B affinity chromatography	61
7.3.5	Desalting and second GST affinity chromatography	62
7.3.6	Gel filtration chromatography	62

7.4	Crystallization	63
7.5	Data collection and processing of hIghmbp2hd	64
7.6	Structure determination and refinement of hIghmbp2hd	64
7.7	Data collection and processing of hIghmbp2hd-RNA	65
7.8	Structure determination and refinement of hIghmbp2hd-RNA	65
7.9	Structural analysis and figure preparation	65
7.10	Surface plasmon resonance	66
Chapter 8	Results	67
8.1	Protein expression and purification	67
8.2	Crystallization and structure determination	70
8.3	Structural overview of the helicase domain of Ighmbp2	77
8.4	Structural overview of hIghmbp2hd-RNA complex	80
8.4.1.	Interactions between Ighmbp2 and the bound RNA	80
8.4.2.	Identification of a conserved solvent accessible surface	81
8.5	Conformational changes upon RNA binding	86
8.6	Structural comparison of Ighmbp2 with other SF1 helicases	88
8.6.1.	Overall tertiary structure	88
8.6.2.	Nucleotide binding site	91
8.6.3.	RNA binding site	92
8.7	R3H domain has a regulatory role on Ighmbp2	93
8.8	Structural and placement of SMARD1 missense mutations on hIghmbp2hd structure	96
8.8.1.	SMARD1 missense mutations in Ighmbp2	96
8.8.2.	Mutants at the nucleotide binding site	96
8.8.3.	Mutants at the RNA binding channel	98
8.8.4.	Mutants at neither ligand binding site	99
8.8.5.	Catalytic activities of Ighmbp2 mutants	100

Chapter 9	Discussion	104
Chapter 10	Future directions	108
Accession Numbers		109
Appendix I	Bacterial culture media	110
Appendix II	Publication list and reprint	112
References		128

Summary

This thesis is composed of two separate projects on the proteins RpfF and Ighmbp2, and their functional role in quorum sensing and neurodegenerative disorder, respectively. The details of the RpfF protein and the helicase domain of Ighmbp2 protein are presented in Chapters 1 to 5 and Chapters 6 to 10, respectively.

Quorum sensing is a form of communication in bacterial cells to allow coordination of gene expression in response to fluctuations in cell-population density. The pathogen *Xanthomonas campestris* pathovar *campestris* (*Xcc*) uses a small diffusible molecule named diffusible signal factor (DSF) as its quorum sensing signal to assess the population density and regulate virulence factor production. DSF regulates a cluster of genes designated as *rpfABCDEFG* which are important for the production of virulence factors such as extracellular enzymes and extracellular polysaccharides in *Xcc*. Recent studies have indicated that the regulation of DSF biosynthesis involves the interaction of RpfC and RpfF proteins. However, the mechanism of how RpfC regulates RpfF activity is not well understood.

The crystal structure of the enoyl-CoA hydratase RpfF at a resolution of 1.8 Å is presented in this study. RpfF adopts a self-associating fold such that the C-terminal helix wrap round the N-terminal α/β core domain. Structural homology search in the Protein Data Bank revealed that RpfF is structurally similar to members of the hydratase/isomerase superfamily. Further structural and mutational analyses identified Glu141 and Glu161 as conserved glutamate residues positioned in the catalytic pocket and are critical to RpfF activity in the DSF biosynthesis pathway. Superposition with ligand-bound enoyl-CoA hydratase showed that the C-terminal helix of RpfF is positioned at the entrance of the catalytic pocket and is in steric hinderance with the Co-A moiety of the

substrate. In addition, the structure of RpfF in complex with REC domain of RpfC shows the REC domain interacting with the C-terminal helix of RpfF. These findings suggest that interaction with RpfC may lock RpfF in an inactive conformation, therefore having a negative regulating effect on DSF production.

The second part of the thesis examines the *Ighmbp2* protein which is implicated in the neurodegenerative disease spinal muscular atrophy with respiratory distress type I (SMARD1). Most of the genetic mutations causing SMARD1 occur in the region of the *Ighmbp2* gene coding for the helicase domain. Although several studies have pointed out *Ighmbp2* involvement in DNA replication, transcription, pre-mRNA splicing, and RNA metabolism, its exact role remains elusive. Based on sequence alignment and biochemical studies, *Ighmbp2* protein has been classified as a member of the helicase superfamily I that can unwind RNA and DNA duplexes in the 5' → 3' direction in an ATP-dependent manner.

To have a better understanding of this protein from the structural perspective, crystal structures of the helicase domain of *Ighmbp2* in the free and ssRNA-bound states are obtained. The helicase domain of *Ighmbp2* has high structural similarity with the helicase domain of Upf1, a protein in the Upf1-like subfamily. Upon ssRNA binding, conformation changes were observed in domains 1A and 1B. Data from surface plasmon resonance experiments showed that the presence of a R3H domain, located at the C-terminus of *Ighmbp2*, enhances *Ighmbp2* nucleic acid binding ability. This highlights the importance of accessory domains in regulating helicase enzymatic activities. Based on *Ighmbp2* helicase structure, the implications of SMARD1 causing mutations in *Ighmbp2* are addressed from the structural perspective.

List of Abbreviations

∞	infinity
11-Me-C ₁₂ : Δ^2	cis-11-methyl-2-dodecenoic acid
Å	angstrom
ADP	adenosine diphosphate
AMPPNP	adenosine 5'-(β - γ -imido)triphosphate tetralithium salt hydrate
AHL	N-acyl-L-homoserine lactones
ASU	asymmetric unit
ATP	adenosine triphosphate
Catf1	cardiac transcription factor
CCP4	collaborative computational project number 4
Da	dalton
ddH ₂ O	double-distilled water
DNA	deoxyribonucleic acid
DSF	diffusible signal factor
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylene diamine tetra acetic acid
EG	ethylene glycol
EPS	extracellular polysaccharide
GF1	glial factor 1
GST	glutathione-S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	hexa-histidine tag
hIghmbp2hd	helicase domain of <i>Homo sapiens</i> immunoglobulin μ -binding protein 2

Ighmbp2	Immunoglobulin μ -binding protein 2
IPTG	isopropyl β -D-thiogalactopyranoside
ITC	isothermal titration calorimetry
K	Kelvin
kDa	kilo dalton
L	litres
M	molar
MAH	matrix-associated helicase
mg	milligram
mIghmbp2hd-R3H	helicase and R3H domains of <i>Mus musculus</i> immunoglobulin μ -binding protein 2
MIRAS	multiple isomorphous replacement with anomalous scattering
ml	millilitre
mM	millimolar
mRNA	messenger ribonucleic acid
NCS	non-crystallographic symmetry
nm	nanometre
<i>nmd</i>	neuromuscular degeneration
NLS	nuclear localizing signal
NMR	nuclear magnetic resonance
MIM	Online Mendelian Inheritance in Man
OD ₅₉₅	optical density at 595 nm
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PMSF	phenylmethanesulfonyl fluoride
RNA	ribonucleic acid

Rip1	rat insulin-enhancer-binding protein
rmsd	root mean square difference
rpm	revolutions per minute
rpf	regulation of pathogenic factors
SAD	single wavelength anomalous dispersion
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
<i>S.c.</i>	<i>Saccharomyces cerevisiae</i>
SeMet	seleno-L-methionine
SETX	Senataxin
SF	superfamily
SF1	superfamily 1
SMA	spinal muscular atrophy
SMA1	spinal muscular atrophy type 1
SMN	survival of motor neurons
ssRNA	single stranded ribonucleic acid
SPR	surface plasmon resonance
TAE	tris-acetate EDTA
TLC	thin layer chromatography
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
ul	microlitre
uM	micromolar
WT	wild type
<i>Xcc</i>	<i>Xanthomonas campestris pv. campestris</i>
X-gluc	5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid
v/v	volume by volume

w/v

weight by volume

List of Figures

Figure 1-1.	Black rot of crucifers and quorum sensing in <i>Xcc</i>	3
Figure 3-1.	Purification of recombinant RpfF protein	19
Figure 3-2.	Ramachandran plot of RpfF structure	21
Figure 3-3.	Structure of RpfF in apo-form and in complex with the REC domain of RpfC	24
Figure 3-4.	Structural comparisons of RpfF with hydratase and isomerases	27
Figure 3-5.	Sequence alignment of RpfF and its homologues and with enoyl-CoA hydratase and isomerase	30
Figure 3-6.	Active site of RpfF and enoyl-CoA hydratase	31
Figure 3-7.	Binding isotherm of ITC titrations showing no interaction between RpfF and DSF	35
Figure 6-1.	Schematic flowchart showing the classification of helicases	47
Figure 6-2.	Domain architectures and crystal structures of SF1 helicases.	52
Figure 6-3.	Domains and conserved sequence motifs in the Upf1-like helicases	53
Figure 6-4.	Structure of R3H domain of human Ighmbp2 protein	55
Figure 8-1.	Purification of recombinant hIghmbp2hd protein	68
Figure 8-2.	Purification of recombinant mIghmbp2hd-R3H protein	69
Figure 8-3.	Crystals of hIghmbp2hd and hIghmbp2hd-RNA complex	71
Figure 8-4.	Ramachandran plots of hIghmbp2hd and hIghmbp2hd-RNA structures	74
Figure 8-5.	Crystal structure of hIghmbp2hd-RNA	76
Figure 8-6.	Domain organization and and conserve sequence motif of hIghmbp2hd	78
Figure 8-7.	Overall crystal structures of hIghmbp2hd and hIghmbp2-RNA	79
Figure 8-8.	Interactions of ssRNA with hIghmbp2hd	83
Figure 8-9.	Surface representation of hIghmbp2hd-RNA showing the positioning of conserved residues and the electrostatic charge distribution properties	85
Figure 8-10.	Conformational changes of hIghmbp2 upon RNA binding	87

Figure 8-11.	Sequence alignment of helicase domains of the Upf1 and Ighmbp2	89
Figure 8-12.	Structural comparison of Ighmbp2 and Upf1	90
Figure 8-13.	Sensograms of SPR experiments testing for RNA binding of hIghmbp2hd and mIghmbp2hd-R3H	95
Figure 8-14.	Placement of SMARD1 disease mutation sites on hIghmbp2hd structure	103

List of Tables

Table 3-1.	Data collection and refinement statistics of RpfF	22
Table 3-2.	Top scoring RpfF structural homologues by DALI server	26
Table 3-3.	Site-directed mutagenesis of the residues associated with the putative substrate-binding pocket of RpfF	34
Table 6-1.	Clinical mutations observed in the <i>Ighmbp2</i> gene	43
Table 8-1.	Data collection and refinement statistics of hIghmbp2hd	72
Table 8-2.	Data collection and refinement statistics of hIghmbp2hd-RNA	73
Table 8-3.	Catalytic activities of mutant Ighmbp2	102

Part One

Structural studies on the RpfF protein involved in quorum sensing

Chapter 1

Introduction

1.1 *Xanthomonas campestris* pv. *campestris* and black rot of crucifers

Bacterial infectious disease is one of the most critical problems in commercial agriculture. *Xanthomonas campestris* pathovar *campestris* (*Xcc*), an aerobic gram-negative bacterium, is the causative agent of black rot disease of cruciferous crops (including broccoli, Brussels sprouts, cabbage, cauliflower, kale, mustard, radish, rutabaga, and turnip) worldwide (Onsando, 1992). As with many phytopathogenic bacteria, *Xcc* exhibits its virulence by producing a range of factors that enables it to parasitize the host (Dow and Daniels, 1994). Among these are extracellular enzymes capable of degrading plant cell components and an extracellular polysaccharide (EPS) called xanthan (Tang *et al.*, 1991).

Under favorable conditions (high humidity and temperature), *Xcc* infects vegetable crops by spreading through the plants' vascular tissues, turning the veins in their leaves yellow and black, and causing V-shaped lesions along the margins of the leaves (**Figure 1-1A**).

Figure 1-1. Black rot of crucifers and quorum sensing in *Xcc*. **A)** V-shaped lesions along the margins of the leaf in a *Xcc* infected plant. **B)** Chemical structure of the DSF molecule, cis-11-methyl-2-dodecenoic acid (11-Me-C₁₂:Δ²). **C)** Schematic representation of DSF signaling. At low cell density, there are low levels of extracellular DSF, therefore RpfC is not phosphorylated. The REC domain of RpfC forms a complex with RpfF. At high cell density, there will be an elevated amount of DSF. DSF binds to the extracellular domain of RpfC and triggers autophosphorylation of RpfC and phosphorelay to RpfG. RpfF is released and can synthesise more DSF.

1.2 Microbial quorum sensing

Quorum sensing (QS) enables bacterial cells to communicate effectively and synchronise the expression of QS genes in a population density dependent manner. They regulate their own gene expression by producing, releasing and sensing chemical signals from the environment. These chemical signals are produced and released by the bacteria itself or by its neighbouring bacterial cells. Also, QS is implicated in the production of virulence factors responsible for pathogen–host association.

One of the most well studied QS system in Gram-negative bacteria is that of *Vibrio fischeri* (Milton, 2006; Whitehead *et al.*, 2001). When cell density is high, the QS signal N-acyl-L-homoserine lactones (AHL) bind to the LuxR proteins. The AHL-bound LuxR proteins activate the transcription expression of QS-dependent genes as well as the *luxI* gene, producing more AHL signals (Dong *et al.*, 2007). When cell density is low, there will be low levels of AHL. Hence, there will be no transcriptional activation.

1.3 Quorum sensing and virulence factor production in *Xcc*

In contrast to AHL-type QS system, *Xcc* uses a diffusible signal factor (DSF) as its signaling molecule. Furthermore, autoregulation of DSF biosynthesis is unlikely to occur at the transcriptional level (He *et al.*, 2006). Wang *et al.*, (2004) has structurally characterized the DSF in *Xcc* as a cis-11-methyl-2-dodecenoic acid (11-Me-C₁₂: Δ^2) which is an alpha-beta unsaturated fatty acid (**Figure 1B**). For simplicity, from this point onwards 11-Me-C₁₂: Δ^2 will be referred to as DSF. DSF may act as a stress- or starvation-induced signal for the synthesis of all extracellular enzymes and EPS (Barber *et al.*, 1997). More recent work has implicated the DSF signaling system in the regulation of biofilm dispersal in *Xcc* (Dow *et al.*, 2003).

The autoinduction mechanism in *Xcc* involves a cluster of *Xcc* genes designated as *rpfABCDEFG* (for regulation of pathogenicity factors) which positively regulates the production of extracellular enzymes and EPS (Barber *et al.*, 1997; Dow and Daniels, 1994; Slater *et al.*, 2000; Tang *et al.*, 1991). Several of these *rpf* genes mediate the production of pathogenic factors via a DSF. Of these genes, *rpfF* encode a putative enoyl-CoA hydratase, *rpfB* encodes a putative long chain fatty acyl CoA ligase, and *rpfC* encodes a membrane associated sensor kinase (Barber *et al.*, 1997). The RpfC/ RpfG ‘two-component’ sensor-regulator pair (Dow and Daniels, 1994; Tang *et al.*, 1991) couples the sensing of environmental signal (including DSF) to the modulation of pathogenic factors (Slater *et al.*, 2000) via a conserved phosphorelay mechanism (He, 2006; He *et al.*, 2006) (**Figure 1C**). *rpfA* encodes an aconitase implicated in iron homeostasis (Wilson *et al.*, 1998), whereas the functions of *rpfD* and *rpfE* remains unknown (Dow *et al.*, 2000).

1.4 Hydratase/ Isomerase superfamily

The hydratase/isomerase superfamily, also known as the crotonase superfamily, was first defined to consist of 2-enoyl-CoA hydratases and 3-2-enoyl-CoA isomerases participating in β -oxidation (Müller-Newen and Stoffel, 1993). Since then, this family of proteins have expanded to consist of over 40 members acting in a wide range of metabolic pathways, catalyzing reactions such as halogenation, hydration/dehydration, isomerization, hydrolysis of thioesters, formation/cleavage of carbon-carbon bonds and decarboxylation (Hamed *et al.*, 2008). Although these enzymes have low sequence identity (<25%) but significant similarity in their amino acid sequence, they possess an

amino acid sequence pattern typical of the superfamily with at least one common catalytic amino acid (Müller-Newen *et al.*, 1995; Müller-Newen and Stoffel, 1993).

The hydratase/isomerase superfamily composed of mechanically diverse enzymes that have low overall sequence identity, and their superfamily membership is established through the knowledge of their crystal structures (Holden *et al.*, 2001; Mursula *et al.*, 2004). The architecture can be described as two distinct domains: the N-terminal core domain that is characterized by ~10 strands of β -pleated sheets and the C-terminal helical domain that is composed of three α -helices. The active site is in the N-terminal domain while the C-terminal is more flexible. The flexibility of this C-terminal region have resulted in variable mode of assembly of trimers (Mursula *et al.*, 2004). The C-terminal domain of each subunit of the trimeric/hexameric assembly can be classified into three classes of folds defined by Hubbard and co-workers (2005), namely self-association fold and intra-/inter-trimer association folds (Hubbard *et al.*, 2005).

A common mechanistic feature of this superfamily is the stabilization of an enolate anion intermediate of phosphopantetheine-linked substrates. Stabilization is carried out by two peptidic NH groups forming an oxyanion hole in the N-terminal domain in which the thioester carbonyl group of the substrate is bound (Holden *et al.*, 2001). In the ligand-bound structures available, the carbonyl oxygen atom of the thioester in coenzyme-A (Co-A) is bound to the oxyanion hole in the enzyme, which activates the substrate for the conversion catalyzed by the active-site residues.

Previous functional and structural studies of enoyl-CoA hydratase showed that two highly conserved glutamate residues Glu144 and Glu164 are critical for their hydratase activity; Glu164 acting as a proton donor while Glu144 activating a water molecule to add a hydroxyl group to the substrate (Bahnson *et al.*, 2002; Engel *et al.*,

1998; Engel *et al.*, 1996; Müller-Newen *et al.*, 1995). The two peptidic NH groups form an oxyanion hole in the N-terminal domain to stabilize the kinetically unfavorable enolate anion intermediate (Holden *et al.*, 2001). In addition, Gly138 creates a helix dipole moment that stabilizes the anolate anion intermediate during enzyme reaction (Benning *et al.*, 1996; Holden *et al.*, 2001). These residues act in concert to facilitate the hydration process (Hofstein *et al.*, 1999) such that absence of either one will greatly reduce the hydratase activity (Agnihotri and Liu, 2003).

1.5 Objectives and scope of this study

In *Xcc*, studies have shown that RpfF activity can be regulated by RpfC. However, the mechanism of which RpfC-RpfF interaction mediates RpfF activity and the autoinduction of DSF biosynthesis is not well understood. In addition, several questions remain open with regards to the regulation of cronotase activity. RpfF, being a putative hydratase, provides an opportunity for further study in this area. The first objective is to obtain the structure of RpfF protein. Another objective is to investigate RpfF-RpfC interaction. The crystal structure of RpfF was solved using the single-wavelength anomalous dispersion (SAD) method by X-ray crystallography. The results were analyzed and compared with RpfF/REC complex. In addition, our structure of RpfF which was predicted to be an enoyl-CoA hydratase, provides an insight to the regulation of its enzymatic activity.

Chapter 2

Materials and Methods

2.1 Molecular cloning

2.1.1 Polymerase chain reaction and analysis

Full –length RpfF (amino acids 1 to 289) was amplified by polymerase chain reaction (PCR) using the genomic DNA of *Xcc* (ATCC 33913 strain, gift from Dr Zhang Lianhui). The primers used were:

forward primer 5'- ATATATGAATTCGATGTCTGCAGTTCAACCC -3'

reverse primer 5'- TTTATAAAGCTTTCAGCCCGCGTCGAGCCC -3'

The underlined sequences correspond to the restriction sites *EcoRI* and *HindIII* on the forward and reverse primers, respectively. All primers were ordered from Sigma Genosys. High fidelity KOD DNA polymerase (Novagen) was used to amplify the constructs. 50 µl PCR reaction mix was as follows:

Reaction mixture components	Volume (µl)
10x Buffer for KOD DNA Polymerase	5
25 mM MgCl ₂	2
DMSO	1
dNTPs (2 mM each)	2
Forward primer (100 µM)	1
Reverse primer (100 µM)	1
Template	1
KOD DNA polymerase	1
ddH ₂ O	36
Total	50

PCR reactions were carried out on an iCycler Thermal Cycler (Bio-Rad) with the following cycling parameters:

No. of cycles	Temperature (°C)	Time
1	95	2 min
30	95	30 s
	55	45 s
	68	2 min
1	68	10 min
1	15	∞

To check for the presence and size of the PCR products, 5 µl PCR products were loaded onto a 1% (w/v) agarose gel in 1 X Tris-acetate EDTA (TAE) running buffer (40 mM Tris, 1.14% (w/v) acetic acid, 1 mM EDTA). PCR products were purified with a QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions.

2.1.2 Construct sub-cloning

Both the PCR-amplified fragment and the vector pETDuet-1 (Novagen) were digested with *EcoRI* and *HindIII* (New England BioLabs) for 3 hours at 37 °C. The 40 µl enzyme digestion mix was as follow:

Reaction mixture components	Volume (µl)
Purified PCR products/ Vector	5
Restriction enzyme1	1
Restriction enzyme2	1
10 X BSA	4
10 X Reaction Buffer	4
ddH ₂ O	25
Total	40

The 5' phosphates of the vector were removed by incubating the digested vector with calf intestine alkaline phosphatase (Roche) for 1 hour at 37 °C. The digested products were separated on an agarose gel as described above. The digested DNA fragments were excised from the agarose gel and purified with a QIAquick Gel Extraction kit (Qiagen) according to the manufacturer's instructions.

Ligation was carried out by mixing the digested PCR products and vector in a molar ratio of 3:1 using T4 DNA ligase (Roche) at 15 °C overnight.

2.1.3 DH5 α cell transformation and colony PCR

5 μ l ligation mix was transformed into 100 μ l *E. coli* DH5 α (Invitrogen) high-efficiency competent cells using heat-shock method. Cells were first incubated on ice for 20 minutes, followed by heat shocked at 42 °C for 1 minute, and a final 30 minutes incubated on ice. 500 μ l of LB medium was added to the competent cells and incubated at 37 °C for 45 minutes. After cells were spun down at 7,800 rpm for 3 minutes, 300 μ l of the supernatant was discarded and the pellet was resuspended with the remaining LB. The cells were streaked on LB agar plates (1.5%) containing 100 μ g/ml ampicillin and incubated at 37 °C overnight.

Positive clones were identified by carrying out colony PCR using *Taq* DNA polymerase. A master PCR mix for 10 PCR reactions is first prepared and then aliquoted to the PCR tubes, 15 μ l per reaction.

PCR Mix for 10 reactions	Volume (μ l)
ddH ₂ O	112.5
10 X Polymearse Buffer	15
Forward primer (10 μ M)	4.5
Reverse primer (10 μ M)	4.5
dNTP (10 mM)	3
MgSO ₄ (100 mM)	3
DMSO	6
<i>Taq</i> DNA polymerase	1.5
Total	150

The PCR products were separated using a 1% (w/v) agarose gel. Clones giving the correct DNA fragment size were selected for plasmid preparation. Clones were picked and inoculated overnight at 37 °C in 5 ml LB medium containing 100 μ g/ml

ampicillin. Plasmids were extracted using a QIAprep Spin Miniprep kit (Qiagen) according to the manufacturer's instructions.

The sequences of those positive clones were confirmed by automated DNA sequencing carried out in an ABI Prism 377 DNA sequencer (Perkin Elmer) by staff from the DNA sequencing unit at Institute of Molecular and Cell Biology (IMCB, A*STAR).

The PCR sequencing reaction was assembled as follows:

Reaction mixture components	Volume (μ l)
BigDye 3.1 sequencing mixture	8
Plasmid	5
sequencing primer (10 μ M)	1
ddH ₂ O	6
Total	20

No. of Cycles	Temperature ($^{\circ}$ C)	Time
1	95	1 min
25	95	10 s
	50	5 s
	60	4 min
1	15	∞

2.2 Protein expression

2.2.1 BL21(DE3)-star cell transformation and protein expression

This plasmid encodes a polypeptide with an N-terminal hexa-histidine tag (His-tag) to facilitate protein purification. The plasmids with correct DNA sequences were selected for transformation into *E. coli* BL21(DE3)-star expression strain for protein over-expression in *E. coli* cells (Invitrogen). The transformation was performed as described above. The cells were plated on LB agar plates (1.5%) containing 100 μ g/ml ampicillin and were incubated at 37 $^{\circ}$ C overnight.

To test the expression of cloned genes, single clones of BL21(DE3)-star cells were picked and inoculated into 2 ml Luria-Bertani (LB) medium containing 100 μ g/ml

ampicillin and grown overnight at 37 °C. 1 ml of cells were stored at -80 °C in 40 % glycerol. 100 µl cells were added into 10 ml fresh LB medium supplemented with 100 µg/ml ampicillin and cultured at 37 °C. When the optical density at 595 nm (OD₅₉₅) reached ~ 0.6, a final concentration of 0.4 mM isopropyl-β-D-thiogalactopyranosid (IPTG) was added to induce protein expression. The cultures were further cultured for 3 hours at 37 °C. 200 µl of uninduced and 100 µl of induced cells were pelleted and lysed by boiling in SDS-PAGE loading buffer. The samples were analyzed for protein expression using denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) on a 12 % SDS-PAGE gel according to the method of Laemmli (Laemmli, 1970). Discontinuous SDS-PAGE with a stacking gel (pH 6.8, 0.125 M Tris-HCl, 0.1 % SDS and 5 % acrylamide/Bis solution) and resolving gel (pH 8.8, 0.375 M Tris-HCl, 0.1 % SDS and 15 % acrylamide/bis solution) was conducted in 1 X SDS running buffer (20 mM Tris Base, 200 mM glycine, 0.1 % SDS) at 25 mA. After SDS-PAGE, the gel was stained in Coomassie staining buffer (45 % (v/v) methanol, 10 % (v/v) acetic acid and 0.25 % (w/v) Coomassie Brilliant Blue R-250), and then destained with 5 % (v/v) methanol and 7.5 % (v/v) acetic acid. The destained gel was washed in ddH₂O to remove the organic solvents before drying.

2.2.2 Large scale bacterial cell culture

For large scale expression, BL21(DE3)-star cells harbouring the protein expression plasmid were inoculated into 100 ml of LB medium supplemented with 100 µg/ml ampicillin for overnight culture at 37 °C. The overnight culture was diluted 100-fold in 10 L of fresh LB medium supplemented with 100 µg/ml ampicillin, and allowed to grow at 37 °C until OD₅₉₅ reached ~0.6. Protein expression was induced by the addition of

IPTG to a concentration of 0.4 mM. Further cell growth was performed at 18 °C for approximately 16 hours. Cells were harvested by centrifugation at 4000 revolutions per minute (rpm) using J6-HC centrifuge (Beckman Coulter) for 30 min at 4 °C, resuspended in lysis buffer (refer to **section 2.3.1**) and cell pellets were stored at -80 °C for later protein purification.

For selenomethionine (SeMet)-substituted protein, cells were inoculated into 50 ml of LB medium supplemented with 100 µg/ml ampicillin for overnight culture at 37 °C. The culture was diluted 200-fold in 5 L MOPS minimal medium (Appendix I) (Neidhardt *et al.*, 1974) supplemented with 100 µg/ml ampicillin and 10 mg seleno-L-methionine. Subsequent steps of cell culture were identical to those of native protein expression.

2.3 Protein purification

2.3.1 Protein purification buffers

Lysis buffer	25 mM Tris, 500 mM NaCl, pH 8.0
Wash buffer	25 mM Tris, 500 mM NaCl, 15 mM imidazole, pH 8.0
Elution buffer	25 mM Tris, 500 mM NaCl, 100 mM imidazole, pH 8.0
Gel filtration buffer	25 mM Tris, 500 mM NaCl, 3 mM DTT, pH 8.0

Protein purification buffers are the same for SeMet protein purification, except that 10 mM dithiothreitol (DTT) (where applicable) was used to reduce selenium oxidation. All purification steps were carried out at 4 °C.

2.3.2 Cell lysis

Cells were thawed and resuspended in 100 ml ice-cold lysis buffer supplemented with 1 mg/ml lysozyme from chicken egg-white (Sigma) and DNaseI. After incubation on ice for 30 min, cells were disrupted by sonication on ice (15 seconds pulses, 60 seconds off, 10 cycles at 15 amplitude microns) using a Soniprep 150 (SANYO). Cell debris was cleared by centrifugation at 18,000 rpm for 1 hour using J6-HC Centrifuge (Beckman Coulter). All purification steps were carried out at 4 °C.

2.3.3 TALON metal affinity chromatography

The clarified supernatant containing soluble His-tagged RpfF was filtered (0.45 µm) and loaded onto a column self-packed with TALON metal affinity resin (Clontech), pre-equilibrated with lysis buffer. The column was then washed three times with 50 ml wash buffer and the protein was eluted with 10 ml elution buffer.

2.3.4 Desalting and gel filtration chromatography

The remaining protein purification steps were carried out using AKTA Xpress system (GE Healthcare). To separate imidazole from the protein, the eluted protein sample was loaded into a HiPrep desalting column (GE Healthcare) equilibrated with gel filtration buffer. The flowthrough were further purified using HiPrep Superdex-200 size-exclusion chromatography column pre-equilibrated with gel filtration buffer. Peak fractions were collected in 2 ml fractions and protein homogeneity was analyzed by SDS-PAGE. Fractions containing pure His-tagged RpfF protein were pooled and concentrated to 8 mg/ml using Vivaspin centrifugal-driven filter concentrators (Sartorius Stedim Biotech) and stored at -80 °C.

Protein concentration was assessed by measuring the absorbance at 280 nm on Nanodrop ND-1000 spectrometer (Thermo Scientific). Protein samples were analyzed by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry to verify protein identity. SeMet-substituted RpfF protein was purified in the same way as the native protein.

2.4 Crystallization

Initial screening for crystallization condition was performed using the sitting-drop vapour-diffusion method at 15°C by mixing 200 nl protein solution with 200 nl reservoir solution (Qiagen protein crystallization kits) in Greiner 96-well plates using Phoenix liquid handling crystallization robot (Art Robbins Instruments). Rod-shaped crystals appear in 1 day from a reservoir solution of 0.2 M potassium thiocyanate and 20% (*w/v*) PEG 3350. This initial condition was then optimized to 0.2 M potassium thiocyanate and 9% (*w/v*) PEG 3350. Crystals suitable for diffraction experiments grew in 48 hours by mixing 2 µl protein solution with 2 µl crystallization buffer, via the hanging-drop vapor-diffusion method. Crystals of SeMet-substituted RpfF were cryoprotected in 25% PEG 400 before flash-frozen in liquid nitrogen.

2.5 Data collection, structure determination and refinement

Crystals belong to $P2_12_12_1$ with cell parameters $a = 96.8 \text{ \AA}$, $b = 112.3 \text{ \AA}$, $c = 119.6 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$. Data sets were collected at 100 K from a single seleno-L-methionine (SeMet)-labelled crystal using synchrotron radiation at beamline ID23-1 at the European Synchrotron Facility (ESRF, France) using an attenuated beam of

dimensions $0.100 \times 0.100 \text{ mm}^2$. Diffraction intensities were recorded on an ADSC Q315R detector.

The structure of Rpff was determined by the Single wavelength anomalous dispersion (SAD) method. The peak data were integrated using MOSFLM (Leslie, 1992) and processed with the CCP4i package (Potterton *et al.*, 2003). Data scaling was done with SCALA. Although it was predicted to have four molecules per asymmetric unit (ASU) with an estimated solvent content of 51.6 % based on a V_m value of 2.54 (Matthews, 1968), the structure was later solved to have three molecules per ASU. Heavy atom search was carried out by SHELXC/D/E (Sheldrick, 2008), all the expected Se sites, were found. Model was built automatically by ARP/wARP (Morris *et al.*, 2002). Crystallographic refinement was performed with the program REFMAC5 (Murshudov *et al.*, 1997) to a final R_{free} of 24.7 %. The final refinement statistics for Rpff are summarized in **Table 1**. Pairwise comparison was done using ClustalW and ESPript (Gouet *et al.*, 1999). Structures were superimposed using Top3D and all structural figures were prepared using PyMol.

2.6 Isothermal titration calorimetry binding assay

All samples were dialysed in fresh buffer (25 mM Tris, 500 mM NaCl, 1.6 % methanol, pH 8.0). Protein concentration was determined by NanoDrop 1000 spectrometer (Thermo Scientific). ITC measurements were performed at 20 °C on a VP-ITC MicroCalorimeter (MicroCal Inc). To analyze the binding of DSF to Rpff, 800 μM DSF was injected into the calorimetric cell containing 50 μM Rpff. Titrations were initiated with one 2 μl injection, followed by twenty-eight 10 μl injections, with 240 s equilibration time between injections. The heat of dilution was measured by additional

injections of syringe sample after saturation and subtracted to obtain the effective heat of binding, which was plotted versus the molar ratio of injectant/ cell. Data were analyzed using the Origin 7.0 program (OriginLab) (Edwards, 2002).

2.7 DSF production assay (performed by Dr He Yawen at IMCB, A*STAR)

Strain XC1 and its derivatives were grown in liquid LB medium till OD₆₀₀ reaching about 2.1. For each strain, the supernatants from 50 ml of cell cultures were collected by centrifugation at 10,000 rpm for 30 min. The supernatants were extracted twice by using equal volume of ethyl acetate. The organic phase was evaporated and the residues containing DSF were dissolved in 50 µl of methanol. For DSF analysis, 5 µl of crude extracts were spotted on thin layer chromatography (TLC) silica gel plate 60F₂₅₄ (Merck), which was separated in a tank containing ethyl acetate and hexane in the ratio 2:8 (v/v) as eluting solvents. The TLC plate was air dried and overlaid with 50 ml of NYG liquid medium supplemented with 0.8% agarose, 30 µg/ml of 5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid (X-gluc) and 1.5 ml of fresh DSF reporter strain FE58 (Wang et al., 2004). The TLC plate was incubated at 28°C in darkness overnight. DSF activity was indicated by the presence of a blue zone. For quantitative comparison, DSF concentration was calculated using the formula: $\text{DSF } (\mu\text{M}) = 0.0099e^{2.2527w}$, where w is the width of blue zone in cm. The formula was derived from a dose-response plot using various dilutions of synthetic DSF, with a correlation coefficient of 0.9657.

Chapter 3

Results

3.1 Protein expression and purification

We initially tried cloning and purifying full-length RpfF (amino acids 1 to 289) with an N-terminus glutathione-S-transferase (GST) tag but resulted in low protein yield. When a smaller 6xHis purification tag was used, there was an increase in RpfF protein yield. The most likely reason could be that the presence of GST tag at the N-terminal may interfere with the homotrimeric binding, resulting in protein insolubility. Our crystal structure of RpfF support this argument as the N-terminal plays a role in the trimer interaction.

The His-tagged RpfF was expressed in *E. coli* and purified to homogeneity. The purity of the protein was estimated using SDS-PAGE at every step and the protein was found to be nearly homogeneous at the end of purification. The gel filtration chromatogram and SDS-PAGE gels of RpfF are shown in **Figure 3-1**. Gel filtration profile showed that the protein is monomeric under high salt condition. Ion exchange was not carried out as the protein aggregates when sodium chloride concentration falls below 500 mM.

Figure 3-1. Purification of recombinant RpfF protein. Gel filtration results from HiPrep Superdex-200 gel filtration column. Protein purities of eluted fractions (F1 to G1) were checked by SDS-PAGE.

3.2 The quality of the structure

The crystal structure of the full-length DSF synthesizing protein (Rpff, molecular weight of 32 kDa) was solved by the single-wavelength anomalous dispersion (SAD) method at a resolution of 1.8 Å. There are three polypeptide chains (Chains A, B and C) per ASU with a three fold non-crystallographic symmetry (NCS) axis. Of all non-glycine and non-proline residues, 91.3% and 8.4% are in the most favoured and allowed regions of the Ramachandran plot (**Figure 3-2**), respectively, as verified with PROCHECK (Laskowski *et al.*, 1993). Statistics of all the structure determinations and refinements are summarized in **Table 3-1**. In all three chains of Rpff, residues of 1-13, 34-41 and 279-289 are disordered. The average root-mean-square deviation (rmsd) when the subunits are superimposed is <0.3Å for all backbone atoms. Chain A is used as the reference molecule for later structure analysis description since there is no substantial difference observed between the structures of the subunits.

Figure 3-2. Ramachandran plot of RpfF structure. The diagram was generated using PROCHECK. The Ramachandran plot shows the phi-psi torsion angles for all residues in the structure (except those at the chain termini). Glycine residues are separately identified by triangles as these are not restricted to the regions of the plot appropriate to the other side chain types. The red coloured region on the plot represents the most favourable combinations of phi-psi values.

Table 3-1. Data collection and refinement statistics of RpffF.

Data collection	RpffF
Derivative	SeMet
Wavelength (Å)	0.9795
Resolution limit (Å)	1.8
Space group	P2 ₁ 2 ₁ 2 ₁
Cell parameters	
a/b/c(Å)	96.8 / 112.3 / 119.6
α/β/γ (°)	90 / 90 / 90
Unique reflections (N)	117,918
I/σ	11.3 (2.9)
Completeness (%)	97.5 (85.2)
R _{merge} ^a	0.094 (0.314)
Number of Se sites	14
Anomalous phasing power	1.46
Figure of merit	0.45 / 0.92
Before/after density modification	
Refinement Statistics	
Data range (Å)	30.0 - 1.8
Used Reflections (N)	111,882
Nonhydrogen atoms	6,142
R _{work} ^b (%)	22.8
R _{free} ^c (%)	24.7
R.m.s deviation	
Bond length (Å)	0.002
Bond angles (°)	0.432
Ramchandran plot (% residues)	
Allowed	99.7
Generously allowed	0.3
Disallowed	0

Values in parentheses indicate the specific values in the highest resolution shell.

^aR_{merge} = $\sum |I_j - \langle I \rangle| / \sum I_j$, where I_j is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

^bR_{work} = $\sum ||F_o| - |F_c|| / \sum |F_c|$, where F_o denotes the observed structure factor amplitude, and F_c denotes the structure factor amplitude calculated from the model.

^cR_{free} is as for R_{work} but calculated with 5.0% of randomly chosen reflections omitted.

3.3 Description of overall structure

RpfF is made up of an N-terminal α/β spiral domain and a C-terminal α helical region. The β -strands in the N-terminal domain are assembled into two β -sheets, A and B, such that β -sheet A lies above and is perpendicular to β -sheet B. The spiral fold in the N-terminal α/β spiral domain starts with $\beta A0$ followed by four turns: $\beta A1/\beta B1/\alpha 1$, $\beta A2/\beta B2/\alpha 2B\alpha 2A$, $\beta A3/\beta B3/\alpha 3$, and $\beta A4/\beta B4/\alpha 4B\alpha 4A$. Each turn consist of two β -strands and an α helical region of one or two helices, similar to all solved structures of hydratase/isomerase family to date. The link region between the two terminals comprises two α -helices ($\alpha 5$ and $\alpha 6$) interspersed with two β -strands ($\beta B5$ and $\beta A5$). $\beta B5$ is anti-parallel to β -sheet B while $\beta A5$ is parallel to β -sheet A. The C-terminal helical domain consists of four α helices ($\alpha 7$ to $\alpha 10$). The nomenclature of the secondary structure elements is defined in **Figure 3-3A**.

Figure 3-3. Structure of RpfF in apo-form and in complex with the REC domain of RpfC. **A)** Overall structure of RpfF showing the self-association fold. **B)** Overall structure of the RpfF/REC complex (Cheng *et al.*, 2010). The N-terminal α/β spiral domain is colored in orange, linker region in green and the C-terminal α -helical region in red. The REC domain of RpfC is shown in magenta with residue Asp512 in stick model.

3.4 RpfF belongs to the enoyl-CoA hydratase/isomerase family

A structural homology search of the Protein Data Bank using the DALI server (Holm and Sander, 1995) revealed more than 50 structural homologues with high Z-score (> 20), all belonging to the crotonase superfamily. The top scoring published structural homologues are tabulated in **Table 3-2**. Similar to other members of the enoyl-CoA hydratase/isomerase family, RpfF has relatively low sequence identity (less than 25%) when compared with other structural homologues. Using TOP3D to superposition of the equivalent C α atoms of RpfF with the members of the subfamily gives pair-wise rmsd values of 1.5-1.8 Å, suggesting that RpfF belongs to the enoyl-CoA hydratase/isomerase family.

Interestingly, RpfF and enoyl-CoA hydratase have a common N-terminal core domain but differ in the C-terminal helices ($\alpha 9$ and $\alpha 10$). Helices $\alpha 9$ and $\alpha 10$ of RpfF join $\alpha 7$ and $\alpha 8$ to wrap around the spiral core domain, forming the self-association fold (Hubbard *et al.*, 2005). Whereas, in all available enoyl-CoA hydratase structures with bound ligands, helices $\alpha 9$ and $\alpha 10$ flip upwards and attach to $\alpha 8$ (**Figure 3-4A**) to form a T2 trimerization domain (Kurimoto *et al.*, 2001), also known as the intra-trimer association fold (Hubbard *et al.*, 2005). In comparison with enoyl-CoA isomerases (PDB code 1SG4), the structures are similar either in ligand bound form or in ligand free form, whereby helices $\alpha 7$ to $\alpha 10$ wrap around the spiral core domain, forming the self-association fold (**Figure 3-4B**) (Hubbard *et al.*, 2005).

Table 3-2. Top scoring Rpff structural homologues by DALI server.

Hydratase/ Isomerase	Z-score	PDB	Reference(s)
Rat mitochondrial enoyl-CoA hydratase	21.2	1DUB/ 2DUB	(Engel <i>et al.</i> , 1998; Engel <i>et al.</i> , 1996)
4-(N, N-dimethylamino) cinnanoyl-CoA hydratase	21.5	1EY3	(Bahnson <i>et al.</i> , 2002)
hexanoyl-CoA hydratase	21.2	1MJ3	(Bell <i>et al.</i> , 2002)
human AU-rich RNA-binding protein/3-hydroxy-3-methylglutaconyl-CoA hydratase	26.7	1HZD	(Kurimoto <i>et al.</i> , 2001)
human mitochondrial Δ^3 - Δ^2 -enoyl-CoA isomerase	26.9	1SG4	(Partanen <i>et al.</i> , 2004)
<i>Saccharomyces cerevisiae</i> Δ^3 - Δ^2 -enoyl-CoA isomerase	22.9	1HNU	(Mursula <i>et al.</i> , 2001)

Figure 3-4. Structural comparisons of RpfF with hydratase and isomerases. A) Stereo view of structure of RpfF (green) superimposed with octanoyl-CoA hydratase (PDB code 2DUB, orange). **B)** Stereo view of structure of RpfF (green) superimposed with Δ^3 - Δ^2 -enoyl-CoA isomerase (PDB code 1SG4, red).

3.5 RpfF as an enoyl-CoA hydratase

It remains unclear whether RpfF plays the role of an isomerase or a hydratase. Previous structural and functional studies have demonstrated that glutamate residues at the active site plays an important role in the catalytic activity. In enoyl-CoA hydratases, two glutamates residues are involved in the catalytic activity, while enoyl-CoA isomerase only contains one catalytic glutamate and the other glutamate is substituted by a Leu residue (Bahnson *et al.*, 2002; Engel *et al.*, 1998; Engel *et al.*, 1996; Müller-Newen *et al.*, 1995). From the sequence alignment of RpfF with enoyl-CoA hydratases and enoyl-CoA isomerase (**Figure 3-5**), RpfF contains two well conserved glutamate residues Glu141 and Glu161, which correspond to the catalytic residues Glu144 and Glu164 in enoyl-CoA hydratase (PDB code 2DUB) (Bahnson *et al.*, 2002), respectively. In contrast, Glu141 and Glu161 of RpfF aligns with Leu114 and Glu136 of enoyl-CoA isomerase (PDB code 1SG4), respectively. In enoyl-CoA isomerase, residue Leu114 is not required for catalysis (Partanen *et al.*, 2004). Hence we hypothesize that RpfF is likely an enoyl-CoA hydratase rather than an enoyl-CoA isomerase. Next we superposition the putative active site of RpfF with that of other enoyl-CoA hydratases/isomerases. The active site of enoyl-CoA hydratases is composed of $\alpha 3$, $\beta B2$ - $\alpha B2$ loop, $\beta B3$ - $\alpha 3$ loop, $\beta B4$ - $\alpha B4$ loop, and two C-terminal α helices ($\alpha 9'$ and $\alpha 10'$) from its neighboring molecule (**Figure 3-6A**). The highly conserved glutamate residues Glu144 and Glu164 are located at $\alpha 3$ and $\beta B4$ - $\alpha B4$ loop respectively. The hydratase activity is carried out by Glu164 acting as a proton donor while Glu144 activating a water molecule to add the hydroxyl group to substrate (Bahnson *et al.*, 2002; Engel *et al.*, 1998; Engel *et al.*, 1996; Müller-Newen *et al.*, 1995). Residues Glu141 and Glu161 in the catalytic site of RpfF is in similar position as those of

enoyl-CoA hydratases. This further supports our notion that RpfF is an enoyl-CoA hydratase rather than an enoyl-CoA isomerase.

To test the role of the glutamate residues, DSF production assay was carried out by expressing wild type or mutant RpfF in DSF-deficient strain $\Delta rpfF\Delta rpfC$, which has lacks DSF synthesis (done in collaboration with Dr He Yawen, Dr Zhang Lianhui's lab, IMCB). Results showed that single point mutation of Glu141 or Glu161 completely abolished DSF production, underscoring that both glutamate residues must be present for catalytic activity of RpfF in DSF biosynthesis. Western blotting was carried out to ensure that the point mutation did not affect the expression of mutant RpfF (**Figure 3-6B**). Altogether, these results suggest that RpfF possess key characteristics similar to the enoyl-CoA hydratase subfamily with two key glutamate residues in its catalytic site.

Figure 3-5. Sequence alignment of RpfF and its homologues and with enoyl-CoA hydratase and isomerase. Homologues of RpfF in *Xanthomonas oryzae* pv. *oryzae* (Xoo), *Stenotrophomonas maltophilia* (Stm), *Xylella fastidiosa* (Xyf) and *Xanthomonas axonopodis* pv. *citri* (Xac). 2DUB and 1SG4 are the PDB codes of enoyl-CoA hydratase and isomerase, respectively. Secondary structural elements of RpfF are indicated on the top. Key catalytic residues are marked with “▲”. Residues located in the hydrophobic pocket are marked with “●”. Identical residues are indicated with a red box and white character. Similar residues are in red character with blue frame.

Figure 3-6. Active site of RpfF and enoyl-CoA hydratase. **A)** Stereo view of catalytic site of RpfF (green) superimposed with ligand bound enoyl-CoA hydratase (PDB code 2DUB, orange). The ligand octanoyl-CoA is shown in stick model. Secondary structures and residues at the catalytic site are labeled, the glycine residues are shown as spheres. **(B)** Data were obtained from (Dr He Yawen, Dr Zhang Lianhui's lab, IMCB). TLC plates were used to quantify DSF activity as indicated by the presence of a blue zone. The amount of DSF production in WT and mutant RpfF are calculated from the three repeats, the error bars indicate standard deviation. **(C)** Cavity analysis of RpfF showed that the cavity in the putative catalytic site is too small to accommodate a substrate with the same carbon chain length as DSF (13 carbon atoms), suggesting that a conformation change will occur when RpfF binds to substrate. The stick models in orange and slant colour are octanoyl-CoA (from PDB code 2DUB) and pre-DSF-CoA (modeled using the structure of octanoyl-CoA as a template), respectively. Figure obtained from (Cheng *et al.*, 2010).

3.6 Active site of RpffF

We identified a hydrophobic pocket at the N-terminal domain of each monomer, and it bears many characteristics suggestive of a putative active site. This hydrophobic pocket is in similar location with rat enoyl-CoA hydratase (Bahnson *et al.*, 2002) and methylmalonyl decarboxylase (Benning *et al.*, 2000) but in contrast to di-acyl-CoA isomerase and 4-chlorobenzoyl-CoA dehalogenase, which have their active sites at the trimer interface. This suggests that the active site is contained solely within one subunit of the trimer.

There are several highly conserved residues in the catalytic site (**Figures 3-5 and 3-6A**). Besides Glu141 and Glu161, Leu136 from β B3, Gly137 and Gly138 from β B3- α 3 loop, Gly85 from β B2- α B2 loop, Met170 from β B4- α B4 loop and Trp258 from α 9 are located at the hydrophobic pocket. Based on the position of octanoyl-CoA in enoyl-CoA hydratase, residues Met170 and Trp258 may form hydrophobic contacts with the long fatty acyl group of the DSF precursor molecule. All of these residues are also highly conserved across different homologues of RpffF (**Figure 3-5**).

The hydrophobic cavity is too small to accommodate the long hydrophobic chain of DSF, a 13 carbon molecule including a branched methyl group. The dimethyl group of this molecule sterically clashes with α 9 of RpffF. In addition, we observed that helix α 10 is in steric hindrance with CoA moiety of the superimposed ligands (**Figure 3-6C**). Putting the observations together, the hydrophobic cavity is probably a DSF precursor docking site, and a conformational change may occur to relieve the steric hindrance for accommodating the incoming substrate.

To examine the role of the residues in the putative substrate binding pocket in RpffF, mutagenesis was carried out and tested for DSF biosynthesis (done in collaboration

with Dr He Yawen, Dr Zhang Lianhui's lab, IMCB). The results were consistent with the structural prediction as single point mutation in these residues resulted in either no DSF production or dramatically reduced DSF levels (**Table 3-3**), suggesting that these residues are important for DSF biosynthesis.

Our attempt to co-crystallize RpfF with either DSF or lauronyl-CoA failed to identify any ligand in the structures we obtained. This could imply that RpfF is unable to bind to its ligand. We then went on to test the binding of RpfF to DSF using ITC experiments. Interestingly, the results showed that RpfF was not able to bind to DSF (**Figure 3-7**). A possible explanation will be that RpfF is in an inactive state (i.e. RpfF in self-association fold), with helices $\alpha 9$ and $\alpha 10$ in steric hindrance with substrate binding. The structure of RpfF in complex with REC domain of RpfC also supports this notion (Cheng *et al.*, 2010). In both free form and in complex with the REC domain of RpfC, RpfF adopts the same self-association fold with rmsd of 0.42Å when all the C α atoms from both forms are superimposed (**Figure 3-3**). The $\alpha 2$ and $\alpha 3$ of REC domains interacts with $\alpha 10$ of RpfF. The $\alpha 10$ of RpfF clashes with the CoA moiety of a superimposed substrate (**Figure 3-6A**). It has been proposed that binding of RpfF to the REC domain locks RpfF in an inactive form (Cheng *et al.*, 2010).

Table 3-3. Site-directed mutagenesis of the residues associated with the putative substrate-binding pocket of RpfF.

RpfF	DSF biosynthesis (μM)
Wild type (WT)	27.4 ± 2.5
G85L	0
G85K	0
G86L	0
G86K	0
L136A	0
G137A	6.4 ± 1.3
G138A	0
E141A	0
P160A	5.2 ± 1.2
P160K	7.6 ± 1.7
E161A	0
M170A	0
W258A	0
L276A	4.7 ± 1.3

Data were obtained from (Dr He Yawen, Dr Zhang Lianhui's lab, IMCB). DSF was quantified after TLC plate assay. The standard deviation was obtained from repeating the assay twice.

Figure 3-7. Binding isotherm of ITC titrations showing no interaction between RpfF and DSF. DSF was injected into the calorimetric cell containing RpfF. The amount of heat absorbed or generated is measured and is proportional to the amount of binding. Upper panel shows the raw data. Bottom panel shows the integrated heat plotted against the molar ratio of DSF added to RpfF.

Chapter 4

Discussion

4.1 Characterizing RpfF

Bacterial pathogens have developed several QS systems to sense and adapt to various environment conditions. The DSF-dependent QS system of *Xcc* plays an important role in synchronising the expression of virulence genes and bacterial cell density. The autoregulation mechanism that controls signal production is a unique feature compared to other QS systems (He and Zhang, 2008). The activity of RpfF has been suggested to be regulated by the RpfC which in turn is regulated by DSF concentration (He *et al.*, 2006). To understand how RpfF, a key DSF synthase (Barber *et al.*, 1997; Wang *et al.*, 2004), catalytic activity can be modulated by its ligand protein RpfC, we solve the crystal structure of the full length protein to a resolution of 1.8Å.

Each monomer of RpfF is structural similar to the members of the enoyl-CoA hydratase/isomerase superfamily. The key difference lies in the positioning of the C-terminal helical domain whereby three different folds can occur. Helices $\alpha 7$ to $\alpha 10$ wrap around the core domain of RpfF forming the self-association fold. Although this fold is very similar to that of isomerase, the presence of two catalytic glutamate residues (Glu141 and Glu161) are needed for activity supports the action of a hydratase/dehydratase. In enoyl-CoA isomerase, a single glutamate residue is sufficient for activity.

The common reaction mechanism of the crontonase superfamily is the stabilization of an enolate anion intermediate by two peptidic NH groups that form an ‘oxyanion hole’. In enoyl-CoA hydratases, this is achieved by two catalytic glutamate residues (Hamed *et al.*, 2008). We compared the structure of RpfF with the enoyl-CoA

hydratase (PDB code 2DUB), which is a well-characterized enzyme (Agnihotri and Liu, 2003; Bahnson *et al.*, 2002). The active sites of both proteins are similar and contained two conserved catalytic glutamate residues (Glu141 and Glu161 in RpfF), whose catalytic role in synthesis of DSF were verified by mutagenesis analysis.

The C-terminal helices $\alpha 9$ and $\alpha 10$ of 2DUB differ from RpfF as they interact with neighboring molecules to form an intra-trimer association fold instead of the self association fold (Hubbard *et al.*, 2005). In addition, superposition of RpfF with the ligand-bound structure of 1EY3 reveals steric hindrance of $\alpha 10$ with CoA moiety. We therefore propose that helices $\alpha 9$ and $\alpha 10$ could act as an ‘arm’ that changes orientation to allow ligand binding. An upward or away movement will allow ligand binding; while its current position will inhibit ligand binding. The activation of RpfF may require substrate aggregation or micelles to displace the helical “arm” that covers the catalytic site, a similar mechanism employed by lipolytic enzymes (Mingarro *et al.*, 1995).

4.2 Implications in quorum sensing

The most critical feature of a QS system is the mechanism that enables bacterial cells to auto-regulate the production of QS signals. *Xcc* has evolved a novel autoregulation mechanism involving the protein-protein interaction between RpfF and RpfC in a cell density dependent manner (He *et al.*, 2006). However, the mechanism by which RpfC modulates the RpfF enzyme activity and hence controls DSF production remains elusive. Structural comparison of the RpfF apo form and the RpfF-REC complex showed that RpfF appears to adopt the same “inactive” form in the absence or in the presence of the REC domain (**Figure 3-3**). In this inactive state, the C-terminal helix $\alpha 10$ of RpfF is orientated downwards that blocks the entrance of the substrate-binding

pocket (**Figure 3-6A**) as discussed above, and the helices $\alpha 2$ and $\alpha 3$ from the REC domain form a helix bundle with the $\alpha 10$ of RpfF to lock it in such a position that prevents interfacial activation. These structural findings suggest a mechanistic explanation as to how DSF autoinduction is mediated by the RpfF/RpfC interaction (**Figure 1-1**). At low cell density, RpfF is associated with the REC domain of RpfC, which might keep the DSF synthase in the inactive state and maintains the DSF production at a basal level. When cell density reaches a threshold level, the diffusible DSF signals accumulate in extracellular environment and may interact with RpfC. The event presumably leads to phosphorylation of the REC domain and consequent release of RpfF, which catalyzes substantial DSF biosynthesis upon interfacial induction.

Chapter 5

Future Directions

The above findings present useful clues for further characterization of the RpfF catalytic mechanisms and the regulation of DSF biosynthesis. A ligand bound structure will help in understanding the conformational change and the functional importance of this enzyme. Assuming RpfF's activation involves micelles formation, it may be possible to obtain ligand bound crystal structure under detergent conditions.

Since the regulation of RpfF's activity is important in DSF biosynthesis, ligands that can inhibit RpfF's activity will be of great interest. DSF auto induction mechanism in *Xcc* is dependent on RpfF and RpfC interaction. Peptides that mimic helices $\alpha 2$, $\alpha 3$ of the REC domain of RpfC can be designed to 'lock' RpfF in the inactive state. This will inhibit DSF biosynthesis and hence reduce virulence factor production.

Part Two

Structural studies on the Ighmbp2 protein involved in Spinal Muscular Atrophy with Respiratory Distress Type 1

Chapter 6

Introduction

6.1 Spinal muscular atrophy with respiratory distress type 1 (SMARD1)

Spinal muscular atrophy (SMA) is a neuromuscular disorder characterized by progressive degeneration of anterior horn cell, leading to α -motor neurons weakness, muscular atrophy and denervation. Various SMA phenotypes have been observed in myriad cases of acute respiratory insufficiency in the first year of life. The major clinical form is spinal muscular atrophy type 1 (SMA1; MIM#253300), an autosomal recessive disease due to deletion or mutation of the survival motor neuron (*SMN*) gene on chromosome 5q13 (Lefebvre *et al.*, 1995; Melki *et al.*, 1994). The manifestation of the SMA1 involves progressive loss of α -motor neurons from the anterior horn leading to atrophy of trunk and limb muscles followed by paralysis of the intercostals muscles.

In 1970s, a different variant of SMA characterized by non-congenital onset of respiratory failure due to diaphragmatic paralysis was reported (Mellins, 1974). The condition is currently known as spinal muscular atrophy with respiratory distress type 1 (SMARD1; MIM#604320) (Grohmann *et al.*, 2001). It is an autosomal recessive disorder resulting from the degeneration of anterior horn α -motor neurons in the spinal cord. The loss of axon leads to muscular atrophy which typically leads to early infant death due to respiratory failure (Grohmann *et al.*, 2001; Grohmann *et al.*, 2003; Kaundl *et al.*, 2008; Mellins, 1974; Pitt *et al.*, 2003). Although the pathological features are very similar to that of SMA1, SMARD1 can be distinguished clinically with paralysis of the diaphragm as the leading symptom and predominantly affecting distal muscle groups (Grohmann *et al.*, 2001; Grohmann *et al.*, 2003; Pitt *et al.*, 2003).

Similarly known as diaphragmatic spinal muscular atrophy or distal hereditary motor neuronopathy type 6, SMARD1 is caused by homozygous or compound heterozygous genetic mutations in the *Ighmbp2* gene located on chromosome 11q13, which encodes immunoglobulin μ -binding protein 2 (Ighmbp2) (Grohmann *et al.*, 2001; Kaindl *et al.*, 2008). The protein has been given different names in several literatures: glial factor 1 (GF1) (Kerr and Khalili, 1991), rat insulin-enhancer-binding protein (Rip1) (Shieh *et al.*, 1995), cardiac transcription factor (Catf1) (Sebastiani *et al.*, 1995) and matrix-associated helicase (MAH) (Molnar *et al.*, 1997). Genetic mutations in the *Ighmbp2* gene were discovered from DNA sequence analysis and gene expression study done on more than 50 SMARD1 patients (**Table 6-1**) (Appleton *et al.*, 2004; Giannini *et al.*, 2006; Grohmann *et al.*, 2001; Grohmann *et al.*, 2003; Grohmann *et al.*, 1999; Maystadt *et al.*, 2004; Mohan *et al.*, 2001; Pierson *et al.*, 2011; Pitt *et al.*, 2003; Wilmshurst *et al.*, 2001; Wong *et al.*, 2006). All except three of the published SMARD1 mutations occur in the helicase core of human Ighmbp2. This gene is homologous to the *Ighmbp2* gene in mouse where a splice-site mutation in this gene accounts for SMA in the neuromuscular degeneration (*nmd*) mouse model (Cox *et al.*, 1998).

There are currently no effective treatments for motor neuron diseases including SMARD1. A potential strategy is to replace the damaged neurons using neural stem cells or motor neurons that have been differentiated from stem cells (Hedlund *et al.*, 2007).

Table 6-1. Clinical mutations observed in the *Ighmbp2* gene.

Mutations			Reference(s)
Nucleotide ^a	Amino Acid Substitutions	Type of Mutations	
638 A→G / 638 A→G	H213R / H213R	Missense/ Missense	(Grohmann <i>et al.</i> , 2001; Grohmann <i>et al.</i> , 2003; Grohmann <i>et al.</i> , 1999)
707 T→G / 707 T→G	L236X / L236X	Nonsense/ Nonsense	(Grohmann <i>et al.</i> , 2001; Grohmann <i>et al.</i> , 2003; Grohmann <i>et al.</i> , 1999; Mohan <i>et al.</i> , 2001; Wilmschurst <i>et al.</i> , 2001)
1540 G→A / 1540 G→A	E514K / E514K	Missense/ Missense	(Grohmann <i>et al.</i> , 2001; Grohmann <i>et al.</i> , 2003; Grohmann <i>et al.</i> , 1999)
1738 G→A / 1738 G→A	V580I / V580I	Missense/ Missense	(Grohmann <i>et al.</i> , 2001; Grohmann <i>et al.</i> , 2003; Grohmann <i>et al.</i> , 1999)
IVS13 + 1G→T / IVS13 + 1G→T	--- / ---	Splice donor/ Splice donor	(Grohmann <i>et al.</i> , 2001; Grohmann <i>et al.</i> , 2003; Grohmann <i>et al.</i> , 1999)
1756 G→T / 1909 C→T	G586C / R637C	Missense/ Missense	(Grohmann <i>et al.</i> , 2003; Mohan <i>et al.</i> , 2001; Wilmschurst <i>et al.</i> , 2001)
1488 C→A / 1488 C→A	C496X / C496X	Nonsense/ Nonsense	(Grohmann <i>et al.</i> , 2003)
114 delA / 114 delA	--- / ---	Frameshift deletion/ Frameshift deletion	(Grohmann <i>et al.</i> , 2003)
983 delAAGAA / 983 delAAGAA	--- / ---	Frameshift deletion/ Frameshift deletion	(Grohmann <i>et al.</i> , 2003)
575 T→C / 1277 T→C	L192P / L426P	Missense/ Missense	(Grohmann <i>et al.</i> , 2003)
338 C→T / 1144 G→A	R130X / E382K	Nonsense/ Missense	(Grohmann <i>et al.</i> , 2003)
1714 delAAG / 2922 T→G	--- / D974E	In-frame deletion/ Missense	(Grohmann <i>et al.</i> , 2003)
2362 C→T / ---	R788X / ---	Nonsense/ ---	(Grohmann <i>et al.</i> , 2003)
1082 T→C / 1730 T→C	L361P / L577P	Missense/ Missense	(Grohmann <i>et al.</i> , 2003)
1488 C→A / 1808 G→A	C496X / R603H	Nonsense/ Missense	(Grohmann <i>et al.</i> , 2003)

Mutations			Reference(s)
Nucleotide ^a	Amino Acid Substitutions	Type of Mutations	
1488 C→A / 1748 A→T	C496X / N583I	Nonsense/ Missense	(Grohmann <i>et al.</i> , 2003)
138 T→A / 1649 ins C	C46X / ---	Nonsense/ Frameshift Insert	(Grohmann <i>et al.</i> , 2003)
439 C→T / 1488 C→A	R147X / C496X	Nonsense/ Nonsense	(Grohmann <i>et al.</i> , 2003)
707 T→G / 1540 G→A	L236X / E514K	Nonsense/ Missense	(Grohmann <i>et al.</i> , 2003)
1488 C→A / 1488 C→A	C496X / C496X	Nonsense/ Nonsense	(Grohmann <i>et al.</i> , 2003)
121 C→T / 675 delT	Q41X / ---	Nonsense/ Frameshift deletion	(Grohmann <i>et al.</i> , 2003)
707 T→G / 721 T→C	L236X / C241R	Nonsense/ Missense	(Grohmann <i>et al.</i> , 2003)
121 delC / 388 C→T	--- / R130X	Frameshift Deletion/ Nonsense	(Grohmann <i>et al.</i> , 2003)
983 delAAGAA / ---	--- / ---	Frameshift deletion/ ---	(Grohmann <i>et al.</i> , 2003)
1000 G→A / 1000 G→A	E334K / E334K	Missense/ Missense	(Grohmann <i>et al.</i> , 2003)
1091 T→C / 2436 delT	L364P / ---	Missense/ Frameshift deletion	(Grohmann <i>et al.</i> , 2003)
1693 G→A / 1730 T→C	D565N / L577P	Missense/ Missense	(Grohmann <i>et al.</i> , 2003)
439 C→T / 2362 C→T	R147X / R788X	Nonsense/ Nonsense	(Grohmann <i>et al.</i> , 2003)
661 A→G / 1813 C→T	T221A / R605X	Missense/ Nonsense	(Grohmann <i>et al.</i> , 2003)
587 A→G / 1909 C→T	Q196R / R637C	Missense/ Missense	(Maystadt <i>et al.</i> , 2004)
647 C→T / 1807 C→T	P216L / R603C	Missense/ Missense	(Maystadt <i>et al.</i> , 2004)
752 T→C / 1730 T→C	L251P / L577P	Missense/ Missense	(Maystadt <i>et al.</i> , 2004)
1693 G→A / 2368 C→T	D565N / R790X	Missense/ Nonsense	(Maystadt <i>et al.</i> , 2004)
1488 C→A / ---	C496X / ---	Nonsense/ ---	(Maystadt <i>et al.</i> , 2004)
50 T→C / 1488 C→A	L17P / C496X	Missense/ nonsense	(Appleton <i>et al.</i> , 2004; Guenther <i>et al.</i> , 2007)

Mutations			Reference(s)
Nucleotide ^a	Amino Acid Substitutions	Type of Mutations	
661delA / 1415T →C	--- / L472P	Frameshift deletion/ nonsense	(Guenther <i>et al.</i> , 2007)
1156T →C / 1813 →T	W386R / R605X	Missense/ nonsense	(Guenther <i>et al.</i> , 2007)
1235 + 3A→G / 1334A →C	--- / H445P	Splice donor site / missense	(Guenther <i>et al.</i> , 2007)
1060 + 1G→T / 1082 T → C	--- / L361P	Splice donor site / missense	(Guenther <i>et al.</i> , 2007)
1281 delC / 1708 C→T	--- / R570X	Frameshift deletion/ Nonsense	(Guenther <i>et al.</i> , 2007)
616 C→T / 1877 delT	Q206X / ---	Nonsense/ frameshift deletion	(Guenther <i>et al.</i> , 2007)
163 C →T / 1969 C →T	Q55X / Q657X	Nonsense/ nonsense	(Guenther <i>et al.</i> , 2007)
904C→T / 1156T →C	Q302X / W386R	Nonsense / missense	(Guenther <i>et al.</i> , 2007)
388C →T / 1743A →C	R130X / R581S	Nonsense/ missense	(Guenther <i>et al.</i> , 2007)
129 delC / 1061 G→A	--- / G354S	Frameshift deletion/ Missense	(Giannini <i>et al.</i> , 2006)
388 C→T / 388 C→T	R130X / R130X	Nonsense/ Nonsense	(Giannini <i>et al.</i> , 2006)
1478 C→T / 2363 C→T	T493I / R788X	Missense/ Nonsense	(Guenther <i>et al.</i> , 2008)
1082 T→C / 1478 C→T	L361P / T493I	Missense/ Missense	(Guenther <i>et al.</i> , 2008)
1082 T→C / 1144 G→A	L361P / E382K	Missense/ Missense	(Pierson <i>et al.</i> , 2011)

^aNucleotide numbering is according to the standard nomenclature starting from the translated part of the Ighmbp2 mRNA with +1 as the A of ATG (GenBank reference NM_002180.1). Mutations in both alleles (where present) are indicated.

6.2 Helicase superfamily

Helicase enzymes can be classified into six superfamilies (SF) based on sequence, structural and biochemical data (Gorbalenya and Koonin, 1993; Singleton *et al.*, 2007). Helicases that form oligomeric (usually hexameric) rings comprise SFs 3 to 6, and those that are non-ring forming comprise SFs 1 and 2 (**Figure 6-1**). Detailed classification of helicase enzymes has been widely reviewed (Fairman-Williams *et al.*, 2010; Jankowsky, 2011; Jankowsky and Fairman, 2007; Singleton *et al.*, 2007).

As the focus of this project is on the Ighmbp2 protein, which has been biochemically characterized to be a Upf1-like helicase (Guenther *et al.*, 2009a); this chapter will focus on the structural and mechanistic aspects of SF1 helicases, in particular the Upf1-like group. The intention of this chapter is not to provide a detailed description of all helicases, which will yield an overwhelmingly amount of information.

Figure 6-1. Schematic flowchart showing the classification of helicases. SF1 and SF2 helicases are further grouped into different families. Branch lengths are not drawn to scale.

6.3.1 Superfamily 1 helicases

6.3.1 Structures and the mechanisms of SF1 helicases

SF1 helicases are further grouped into three families: UvrD/Rep, Pif1-like and Upf1-like families (**Figure 6-1**). The structurally conserved helicase core of SF1 helicases is formed by two highly similar RecA-like domains arranged in tandem and is usually surrounded by C- and N-terminal accessory domains. The domain architectures of these families are shown in **Figure 6-2**. UvrD/Rep and Pif1-like helicases are able to unwind only DNA duplex whereas Upf1-like helicases can unwind both DNA and RNA duplexes. UvrD/Rep helicases translocate along their nucleic acid substrate in the 3'→5' direction and is classified as a SF1A enzyme. In contrast, Pif1-like and Upf1-like helicases translocate in the 5'→3' direction and are classified as SF1B enzymes (Singleton *et al.*, 2007).

There is a wealth of scholarly literature studying the mechanisms of SF1A helicase from members of the UvrD/Rep family such as PcrA, Rep and UvrD both from a structural perspective (Korolev *et al.*, 1997; Subramanya *et al.*, 1996; Velankar *et al.*, 1999) and from biochemical studies (Dillingham *et al.*, 2001; Soultanas *et al.*, 2000; Tomko *et al.*, 2007).

At present only two SF1B helicase structures have been solved: namely the Upf1 protein (Chakrabarti *et al.*, 2011; Cheng *et al.*, 2007) and the RecD protein (Saikrishnan *et al.*, 2008; Saikrishnan *et al.*, 2009; Singleton *et al.*, 2004). The RecD protein is a subunit of the *E. coli* RecBCD enzyme complex involved in recombinational repair of double stranded breaks in DNA, while the Upf1 protein is involved in nonsense-mediated mRNA decay (Bhattacharya *et al.*, 2000; Czapinski *et al.*, 1995; Sun *et al.*, 1998).

The mechanism of SF1B helicases to translocate 5'→3' has been well characterized from the structural and biochemical studies done on the RecD2 protein (RecD homologue from *Deinococcus Radiodurans*), a member of Pif1-like family that unwinds DNA duplex. The translocation process involves closing of domains 1A and 2A upon adenosine triphosphate (ATP) binding, and sliding of the DNA backbone along domains 2A and 2B by one base. ATP hydrolysis induces conformational changes in the protein such that domains 1A and 1B resume the open conformation (Saikrishnan *et al.*, 2009). It was also proposed that a rigid β -hairpin formed by domain 1B of RecD protein is essential for helicase activity as deletion of this hairpin abolishes helicase activity completely (Saikrishnan *et al.*, 2008) (**Figure 6-2B**). In many other helicases, the use of a pin or wedge as a mechanical device to split the nucleic acid duplex have also been observed (Buttner *et al.*, 2007; Singleton *et al.*, 2004; Singleton *et al.*, 2001). Hence, the location of this pin could suggest where the incoming duplex may be situated during unwinding.

In comparison with UvrD/Rep and Pif1-like helicases, Upf1-like helicases have different domain insertions in the helicase core, which may significantly differ its unwinding mechanism from the UvrD/Rep and Pif1-like helicases (**Figure 6-2**). In addition, Upf1-like helicases act on both DNA and RNA duplexes. Therefore, it is difficult to speculate if Upf1-like helicases act in a similar mechanism as described in RecD protein.

Nevertheless, the regulation of Upf1 helicase activity has been well established from the recent crystal structures of the Upf1 protein in various ATP-hydrolysis states and in complex with single stranded RNA (ssRNA) or in complex with the Upf2 protein (Chakrabarti *et al.*, 2011; Cheng *et al.*, 2007; Clerici *et al.*, 2009). Upf1 protein shares the

common theme observed in other helicases, which is the closure of the RecA-like helicase domains upon ATP binding. Comparing RNA-free and RNA-bound Upf1 structures revealed key conformational changes in domain 1B. More importantly, there are evidences showing that interaction of the regulatory CH domain of Upf1 with the Upf2 protein regulates Upf1 activity (Chakrabarti *et al.*, 2011; Chamieh *et al.*, 2008; Clerici *et al.*, 2009). In the absence of Upf2, both the CH and helicase domains of Upf1 bind tightly to the RNA. Upon Upf2 binding, Upf1 undergoes a large conformation change in the CH domain. Consequently, the catalytic helicase domain can now unwind RNA duplexes as it binds the RNA less extensively. As proposed by Chakrabarti *et al.* (2011), Upf2 interaction with Upf1 switches Upf1 from a 'RNA clamping' mode and 'RNA unwinding' mode.

6.3.2 Helicases associated with neurodegenerative diseases

Two members of the SF1 helicases are known to be associated with neurodegenerative diseases. They are the Ighmbp2 protein and the senataxin (SETX) protein. Details of Ighmbp2 will be discussed in **section 6.4 and 6.5**. Senataxin (coded by *SETX* gene) contains an N-terminal putative protein-protein interaction domain and a C-terminal helicase domain, followed by a nuclear localizing signal (NLS) (**Figure 6-3A**) (Chen *et al.*, 2004). Mutations in *SETX* have been reported to cause the autosomal recessive neurodegenerative disorder ataxia-oculomotor apraxia (AOA2) (Moreira *et al.*, 2004) and the autosomal dominant amyotrophic lateral sclerosis (ALS4) (Chen *et al.*, 2004). It has been speculated that the dominant ALS phenotype is due to toxic gain-of-function mutations whereas AOA2 is due to loss-of-function mutations (Chen *et al.*, 2006). The conserved sequence motifs of SF1 helicases are also observed in the helicase

domain of senataxin (**Figure 6-3B**). The helicase domain of senataxin has high sequence similarity to that of human Ighmbp2 and human Upf1 (42% and 46% similarity, respectively; (Chen *et al.*, 2004)) and is postulated to be a RNA/ DNA helicase (Chen *et al.*, 2004; Chen *et al.*, 2006; Moreira *et al.*, 2004).

Similar to the yeast SETX ortholog (Sen1), senataxin plays a role in the regulation of gene expression. This is shown by the ability of senataxin to interact with RNA polymerase II and other RNA processing factors (such as poly(A) binding protein 1 and 2, SMN and heterogeneous nuclear ribonucleoproteins) (Suraweera *et al.*, 2009). Senataxin also acts in Xrn2-mediated transcription termination, to resolve RNA/DNA hybrid structures to allow access of the 5' → 3' exonuclease Xrn2 to the 3' cleavage poly(A) sites (Skourti-Stathaki *et al.*, 2011). Recently, senataxin is shown to regulate cell death and neurite formation by modulating fibroblast growth factor 8 expression levels and its downstream signaling (Vantaggiato *et al.*, 2011).

A

B

Figure 6-2. Domain architectures and crystal structures of SF1 helicases. **A)** The domain organization of representative SF1B helicases is schematically represented. The RecA-like helicase domains (1A and 2A) are coloured green and blue, respectively. Variable termini are coloured in grey. **B)** Representative crystal structures of the helicase core of each family are shown. Protein domains are coloured as in panel (A). Bound RNA is displayed as an orange cartoon tube and nucleotides are shown in black coloured stick model. From left to right: Crystal structures of Upf1 (PDB code 2XZO; (Chakrabarti *et al.*, 2011)), RecD (PDB code 3GPL; (Saikrishnan *et al.*, 2009)) and UvrD (PDB code 2IS4; (Lee and Yang, 2006)).

A

B

Motifs	Q	I	Ia	Ib	Ic	II
Human Upf1	NHSQV	PGT <u>G</u> KT	PSNIAVDQ	VRL	TCVG	<u>D</u> ESTQ
Human Ighmbp2	DTSQK	PGT <u>G</u> KT	PSNIAVDN	LRL	TNTG	<u>D</u> ECAQ
Human Senataxin	NEDQK	PGT <u>G</u> K <u>S</u>	PSNAAVDE	VRL	TLST	<u>D</u> EAGQ

Motifs	III	IIIa	IV	V	Va	Vb	VI
Human Upf1	GDHCQ	QYR	TPYEGQ	EIASVD	FQGRE	VR	ALTRAR
Human Ighmbp2	GDHKQ	QYR	SPYNLQ	EIKSVD	FQGRE	VR	AVTRAR
Human Senataxin	GDPKQ	QYR	THYKAQ	EVDTVD	FQGRQ	VR	TITRAK

Figure 6-3. Domains and conserved sequence motifs in the Upf1-like helicases. A) Schematic representation of the domains in the proteins Upf1, Ighmbp2 and senataxin. Full length Upf1 contains an accessory N-terminal CH domain while full length Ighmbp2 has a R3H domain and an AN-1 type Zinc-finger domain at its C-terminal. Senataxin has an N-terminal protein-protein interaction domain. The domains boundaries of senataxin are estimated based on sequence alignment. Diagram is not drawn to scale. **B)** The protein sequences of the conserved motifs are shown. The residues of Walker A (GKT/S) and Walker B (DE) in motifs I and II, are underlined respectively.

6.4 Domain organization of Ighmbp2

Full-length human Ighmbp2 (residues 1 to 993) is a DNA-binding protein which comprises 3 main domains: an N-terminal helicase domain followed by a R3H domain and an AN1-type zinc-finger motif at the C-terminus (**Figure 6-3A**). Ighmbp2 is an ATP-dependent 5'→3' helicase that can unwind RNA and DNA duplexes (Guenther *et al.*, 2009a).

The helicase domain of Ighmbp2 has high sequence homology with members of the Upf1-like family in the SF1 helicases and shares the conserved sequence motifs in this family (**Figure 6-3B**) (Fairman-Williams *et al.*, 2010; Gorbalenya and Koonin, 1988; Koonin, 1992). These motifs are signatures for ATP-dependent nucleic acid translocation motors (Soultanas and Wigley, 2000). Walker A and B motifs that are involved in ATP binding and hydrolysis (Gorbalenya and Koonin, 1993) are also present in Ighmbp2.

The R3H domain is named after an invariant arginine residue and a highly conserved histidine residue, spaced in the sequence by three residues (Grishin, 1998). It has been suggested that the R3H domain might bind to single stranded DNA or RNA (Anantharaman *et al.*, 2002; Grishin, 1998). Proteins containing the R3H domain have been grouped into eight families on the basis of similarities outside the R3H region (Grishin, 1998). Ighmbp2 contains an SF1 helicase domain unique to family VII. Although the NMR structure of R3H domain from human Ighmbp2 (Liepinsh *et al.*, 2003) and several other proteins have been determined (Biou *et al.*, 1995; Katoh *et al.*, 2000; Wu *et al.*, 2005), the exact function of R3H domain has not been fully determined.

Figure 6-4. Structure of R3H domain of human Ighmbp2 protein. Top left panel shows the conserved basic residues on one face of the structure. Top right panel shows the residues at the other face of the structure which are mostly hydrophobic residues. Bottom panels shows the electrostatic charge distribution (blue positive to red negative) on the surface of R3H domain (PDB code 1MSZ; (Liepinsh *et al.*, 2003)).

6.5 Biological significance of Ighmbp2

The *Ighmbp2* gene has 15 exons (Grohmann *et al.*, 2001) and is ubiquitously expressed (Fukita *et al.*, 1993). As mentioned in **section 6.1**, mutations in *Ighmbp2* cause SMARD1. However, the sequence of events resulting in selective motor neuron vulnerability and its cellular mechanism has not been well understood (Grohmann *et al.*, 2004; Grohmann *et al.*, 2001).

The localization of Ighmbp2 protein was originally reported to be in the nucleus and seems to be involved in replication or transcription of DNA (Shieh *et al.*, 1995; Zhang *et al.*, 1999). Ighmbp2 is also implicated in pre-mRNA processing based on the co-localization with splicing factors (Molnar *et al.*, 1997).

Recent studies carried out on motor neurons found that the Ighmbp2 protein is mainly in the cytosol and in the axons of neuronal cells, hence suggesting a role in the maintenance of muscle fibres (Grohmann *et al.*, 2004). Evidences from enzymatic activity of Ighmbp2 on RNA, and its physical interactions with ribosomes and tRNA revealed its role in RNA processing, regulation or metabolism (de Planell-Saguer *et al.*, 2009; Guenther *et al.*, 2009a).

6.6 Objectives and scope of this study

In recent years, numerous studies on SMARD1 patients have identified several mutations in *Ighmbp2* gene. Although Ighmbp2 protein has been classified as a SF1 helicase, several questions remain open on the exact cellular functions of Ighmbp2. In this study, we aim to address how genetic mutations impaired the molecular functions of Ighmbp2 from a structural perspective. Hence, the first objective of this project is to determine the structure of the helicase domain of Ighmbp2 since it is the mutation hotspot. The crystal structure obtained will provide a foundation for further studies in characterizing Ighmbp2. Another objective is to investigate how RNA binding maybe influenced by the R3H domain of Ighmbp2. Towards these goals, the crystal structures of Ighmbp2 helicase domain in the free and RNA-bound states have been determined by X-ray crystallography. RNA binding experiments were carried out to establish how R3H domain affects Ighmbp2 RNA binding.

Chapter 7

Materials and Methods

7.1 Molecular cloning

7.1.1 Polymerase chain reaction and analysis

Constructs containing the following gene fragments were made:

A. Human *Ighmbp2* helicase domain (hIghmbp2hd, residues 1 to 652)

forward primer 5'-CGGGTCGACTCATGGCCTCGGCAGCTGTGGAGAGC – 3'

reverse primer 5'-GATGCGGCCGCTTAGGAATAGTTTTCTGGGACAATATC
GTCAAG-3'

B. Mouse *Ighmbp2* helicase and R3H domains (mIghmbp2hd-R3H, residues 1 to 786)

forward primer 5'-CGGGTCGACTCATGGCCTCGTCCACCGTGGAGAGTT– 3'

reverse primer 5'-ATAAGAATGCGGCCGCTTAGGCAGGGCTCCTCCTGCT
CAC- 3'

Gene fragments were amplified by PCR using the cDNA of human and mouse *Ighmbp2*. The underlined sequences correspond to *SalI* and *NotI* restriction sites on the forward and reverse primers, respectively. All primers were ordered from 1st Base. High fidelity KOD DNA polymerase (Novagen) was used to amplify the constructs.

Reaction mixture components	Volume (μl)
10 X Buffer for KOD DNA Polymerase	5
25 mM MgCl ₂	2
DMSO	1
dNTPs (2 mM each)	2
Forward primer (100 μM)	1
Reverse primer (100 μM)	1
Template	1
KOD DNA polymerase	1
ddH ₂ O	36
Total	50

PCR reactions were carried out on an iCycler Thermal Cycler (Bio-Rad) with the following cycling parameters:

No. of Cycles	Temperature (°C)	Time
1	95	2 min
30	95	30 s
	55	45 s
	68	2 min
1	68	10 min
1	15	∞

To check the presence and size of the PCR products, 5 µl PCR products were loaded onto a 1 % (w/v) agarose gel in 1 X TAE running buffer. PCR products were purified with the QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions.

7.1.2 Construct sub-cloning

Both the PCR-amplified fragment and the vector pGEX6p-1 (GE Healthcare) were digested with *Sall* and *NotI* for 3 hours at 37 °C. The 40 µl enzyme digestion mix was as follow:

Reaction mixture components	Volume (µl)
Purified PCR products/ Vector	5
Restriction enzyme 1	1
Restriction enzyme 2	1
10 X BSA	4
10 X Reaction Buffer	4
ddH ₂ O	25
Total	40

The 5' phosphates of the vector were removed by incubating the digested vector with calf intestine alkaline phosphatase (Roche) for 1 hour at 37 °C. The digested products were separated on an agarose gel as described above. The digested DNA

fragments were excised from the agarose gel and purified with QIAquick Gel Extraction kit (Qiagen) according to the manufacturer's instructions.

Ligation was carried out by mixing the digested PCR product and vector in the molar ratio of 3:1 using T4 DNA ligase (Roche) at room temperature for 3 hours.

7.1.3 DH5 α cell transformation and colony PCR

Same as the procedure described in **section 2.1.3**. This generates a plasmid coding for a polypeptide with an N-terminal GST tag to facilitate protein purification.

7.2 Protein expression

7.2.1 Rosetta(DE3)pLysS cell transformation and protein expression

GST-fusion proteins (hIghmbp2hd and mIghmbp2hd-R3H) were separately expressed in Rosetta(DE3)pLysS expression strain cells (Novagen). Cell transformation was performed as described in **section 2.1.3**. The cells were plated on LB agar plates (1.5%) containing 100 μ g/ml ampicillin and were incubated at 37°C overnight.

The testing of the expression of cloned genes is the same as that described in **section 2.2.1**.

7.2.2 Large scale bacterial cell culture

Same as the procedure described in **section 2.2.2**. Cells were harvested and resuspended in lysis buffer (refer to **section 7.3.1** and **7.3.2** for buffer conditions). The cell pellets were stored at -80 °C for later protein purification.

7.3 Protein purification

7.3.1 Protein purification buffers for hIghmbp2hd

Lysis buffer 25 mM HEPES, 500 mM NaCl, 3 mM DTT, 3 mM MgCl₂, pH
7.3

Elution Buffer 25 mM HEPES, 500 mM NaCl, 3 mM DTT, 3 mM MgCl₂, 1 %
w/v reduced glutathione, pH 7.3

7.3.2 Protein purification buffers for mIghmbp2hd-R3H

Lysis buffer 25 mM Tris, 500 mM NaCl, 3 mM DTT, 3 mM MgCl₂, pH 8.0

Elution Buffer 25 mM Tris, 500 mM NaCl, 3 mM DTT, 3 mM MgCl₂, 1 % w/v
reduced glutathione, pH 8.0

7.3.3 Cell lysis

Same as the procedure described in **section 2.3.2**.

7.3.4 Glutathione Sepharose 4B affinity chromatography

The clarified supernatant containing soluble GST-fusion protein was loaded onto a column self-packed with Glutathione Sepharose 4B beads (GE Healthcare), pre-equilibrated with lysis buffer. The column was washed with three times with 50 ml lysis buffer and the GST-fusion protein was eluted with 15 ml elution buffer. The eluted fraction was then incubated with 3C PreScission protease (GE Healthcare) (in the ratio 1 mg 3C protease to 50 mg eluted protein) for GST tag cleavage at 4 °C overnight.

7.3.5 Desalting and second GST affinity chromatography

Using the AKTA Xpress system (GE Healthcare), GST tag cleaved protein sample was loaded onto a HiPrep Desalting column (GE Healthcare) pre-equilibrated with lysis buffer. After desalting, the sample was passed through a regenerated Glutathione Sepharose 4B column to remove cleaved GST tags. The flowthrough was further purified by size exclusion chromatography.

7.3.6 Gel filtration chromatography

The proteins hIghmbp2hd and mIghmbp2hd-R3H were loaded into pre-equilibrated HiPrep Superdex-75 and Superdex-200 size exclusion chromatography column (GE Healthcare), respectively. Peaks were collected in 2 ml fractions and analyzed on SDS-PAGE gels. Fractions containing pure homogenous protein were pooled and concentrated (12 mg/ml for hIghmbp2hd and 20 mg/ml for mIghmbp2hd-R3H) in lysis buffer using Vivaspin centrifugal-driven filter concentrators (Sartorius Stedim Biotech).

Protein concentration was assessed by measuring the absorbance at 280 nm using Nanodrop ND-1000 spectrometer (Thermo Scientific). Purified proteins were run on an SDS-PAGE gel and stained with Coomassie blue for visualization. The required bands were excised from the gel and submitted for protein identification by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (Mass Spectrometry Facility, School of Biological Sciences, NTU).

7.4 Crystallization

For crystallization, 5 mM potassium phosphate was added to the protein prior to screening. Initial screening for crystallization condition was performed using the sitting-drop vapour-diffusion method using protein crystallization kits from Qiagen at 15 °C.

Typically, 200 nl protein solution was mixed with 200 nl reservoir solution in Greiner 96-well plates using Phoenix liquid handling crystallization robot (Art Robbins Instruments). Plate-like crystals appear in 1 day from a reservoir solution of 0.2 M sodium malonate pH 7.0 and 20 % (*w/v*) PEG 3350 (Qiagen Screening Kit JCSG+, well G6). Crystals suitable for diffraction experiments grew in 4 days by mixing 2 µl protein solution with 2 µl crystallization buffer, via the hanging-drop vapor-diffusion method. As SeMet-substituted proteins were insoluble, phases were obtained by soaking crystals in heavy atoms. Mercury-derivative crystals were obtained by soaking the crystals in 0.1 mM ethyl mercury thiosalicylate (EMTS) for 6 hours. Iodide-derivative crystals were obtained by quick soaking the crystals in 500 mM potassium iodide for 5 seconds. Crystals were cryoprotected with 25 % ethylene glycol before flash-frozen in liquid nitrogen.

hIghmbp2hd-RNA complex was obtained by adding ssRNA of 10 adenine bases long (A_{10}) to the protein in molar ratio of 2:1. Crystals were obtained by hanging drop method with equal volumes of protein solution and well solution of 0.1 M imidazole pH 8.0, 1.0 M sodium citrate at 15 °C. Crystals were cryoprotected with 25 % ethylene glycol before flash-frozen in liquid nitrogen.

7.5 Data collection and processing of hIghmbp2hd

Native diffraction data set for hIghmbp2hd was collected at 100K using synchrotron radiation at beamline ID14-4 at the European Synchrotron Facility (ESRF, Grenoble, France). Mercury and iodide derivative data sets were collected at 100K using synchrotron radiation at beamline 13B1 at the National Synchrotron Radiation Research Centre (NSRRC, Taiwan R.O.C). The datasets were integrated using MOSFLM (Leslie, 2006) then scaled and reduced using the CCP4i suite (Potterton *et al.*, 2003).

7.6 Structure determination and refinement of hIghmbp2hd

The structure of hIghmbp2hd was determined by the multiple isomorphous replacement with anomalous scattering (MIRAS) method using mercury and iodide derived crystals. Heavy atom search was carried out by Autosol in the PHENIX software suite (Adams *et al.*, 2010). Three mercury and 17 iodide sites, corresponding to one molecule of hIghmbp2hd in the ASU, were found. Automatic partial model building was carried out using AutoBuild in the PHENIX software suit. Subsequent manual model building was done using COOT (Emsley and Cowtan, 2004). Crystallographic refinement was performed using PHENIX software suit (Adams *et al.*, 2010). TLS refinement was introduced in the last refinement steps. Phosphate group and solvent molecules were included at the final stage of refinement. The quality of the models was analyzed using PROCHECK (Laskowski, 1993).

7.7 Data collection and processing of hIghmbp2hd-RNA

Diffraction data set for hIghmbp2hd-RNA complex was collected at 100K using synchrotron radiation at beamline ID14-4 at the European Synchrotron Facility (ESRF, Grenoble, France). Data were integrated with XDS (Kabsch, 1988, 1993) then scaled and reduced using CCP4i suite (Potterton *et al.*, 2003).

7.8 Structure determination and refinement of hIghmbp2hd-RNA

The structure was determined by molecular replacement using PHASER (McCoy *et al.*, 2007). The structure of hIghmbp2hd was used as the search model. Two molecules in the ASU in the space group P3₁21 were found with Z-score of 37.6. Domains 1A and 2A matches the electron density map but domains 1B and 1C do not. Domains 1A and 2A were first refined before domains 1B and 1C were manually built using COOT (Emsley and Cowtan, 2004). Model refinement was carried out with PHENIX software suit to a final R_{free} of 29.28%. RNA molecules were included in the final stages of refinement. The quality of the model was analyzed using PROCHECK (Laskowski, 1993).

7.9 Structural analysis and figure preparation

Structure figures were prepared using the program PyMol (<http://www.pymol.org>). Sequence alignment was done by ClusterW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) and displayed by ESPript (<http://esprict.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>). Mapping of sequence conservation on the molecular surface was carried out with ConSurf. Molecular surface electrostatic

properties were evaluated by PyMol. The extent to which domain 1B rotates upon RNA binding was calculated using Superimpose from the CCP4i suite by fixing domains 1A and 1B. Superpositions of structures were carried out using program Top3D.

7.10 Surface plasmon resonance

Surface plasmon resonance (SPR) was carried out on the Biacore 3000 instrument. The 5' end biotin labeled ssRNA containing 15 adenine bases was attached to a streptavidine-coated sensor chip (SA chip, Biacore) with an immobilization level of 3000 RU. Procedures were as previously described (Cheng *et al.*, 2007). Binding buffer of 20 mM HEPES, pH 7.4, 100 mM NaCl, 2 mM MgCl₂, 2 mM DTT and 0.002% Tween-20 was used. Typically 75 µl of protein sample was injected into flow cells 1 and 2 at a rate of 30 µl/min. When required, 2 mM nucleotide was added. The data were evaluated using BIAevaluation 3.1

Chapter 8

Results

8.1 Protein expression and purification

The GST-fusion proteins (hIghmbp2hd and mIghmbp2hd-R3H) were expressed separately in *E. coli* and purified to homogeneity. The purity of the protein was estimated using SDS-PAGE gels at every step and the protein was found to be nearly homogeneous at the end of purification. The gel filtration chromatograms and SDS-PAGE gels of hIghmbp2hd and mIghmbp2hd-R3H are shown in **Figures 8-1 and 8-2**, respectively. Gel filtration profile showed that both proteins are monomeric under high salt condition. The molecular weights of hIghmbp2hd and mIghmbp2hd-R3H are 72 kDa and 83 kDa, respectively. Ion exchange was not carried out as both the proteins aggregate when sodium chloride concentration falls below 500 mM.

Figure 8-1. Purification of recombinant hghmbp2hd protein. Protein is eluted at 130 ml on the size exclusion chromatography column HiPrep Superdex-75. This elution volume corresponds to protein size of approximately 75 kDa. Protein purities of eluted fractions (C11, C12, D5 to D9) were checked by SDS-PAGE. Lane 1 contains the protein molecular weight marker.

Figure 8-2. Purification of recombinant mIghmbp2hd-R3H protein. Protein is eluted at 210 ml on the size exclusion chromatography column HiPrep Superdex-200. This elution volume corresponds to protein size of approximately 85 kDa. Protein purities of eluted fractions (F7 to G3) were checked by SDS-PAGE. Lane 1 contains the protein molecular weight marker.

8.2 Crystallization and structure determination

Purified hIghmbp2hd protein yielded crystals of good diffraction to a resolution of 2.5 Å. Crystals belong to the space group C2 with one molecule per ASU. The structure was solved by MIRAS using mercury and iodide derivatives. A phosphate ion is observed at the nucleotide-binding site (**Figure 8-5B**). Residues 73-75, 120-126, 310-315, 503-507 and 649-652 were disordered in the structure. The final model has an R_{free} of 25.23 % with good stereochemistry.

The structure of hIghmbp2hd with bound ssRNA (hIghmbp2hd-RNA) was solved by molecular replacement using hIghmbp2hd as a search model. The search identifies two copies of hIghmbp2hd-RNA per ASU. However domains 1B and 1C do not fit well with the electron density map. These two domains were later manually built. The final model was refined to a resolution of 2.85Å with R_{free} 29.28%. Residues 32-36, 94-97, 121-124, 308-320, 502-506 and 649-652 for both molecules were disordered. Electron density for 9 RNA bases in molecule A and for 8 RNA bases in molecule B were observed (**Figure 8-5**). As there were no substantial difference between the two molecules (rmsd of 0.85 Å for all equivalent C α atoms of the protein), all subsequent analyzes are based on molecule A as it is more complete with 9 RNA bases.

Similar to the apo form, a phosphate ion occupies the nucleotide-binding pocket in both molecules in the ASU. The presence of phosphate ion in the nucleotide pocket is due to the addition of potassium phosphate prior to crystallization in both structures. Details of all the structure determinations and refinements for hIghmbp2hd and hIghmbp2hd-R3H are summarized in **Tables 8-1 and 8-2**, respectively. The quality of

the models were analyzed using PROCHECK (Laskowski, 1993). The Ramachandran plots are shown in **Figure 8-4**.

Figure 8-3. Crystals of hIghmbp2hd and the hIghmbp2hd-RNA complex. Plate-like crystals of hIghmbp2hd (left) and diamond-shaped crystal of hIghmbp2hd-RNA (right).

Table 8-1. Data collection and refinement statistics of hIghmbp2hd.

Data collection	hIghmbp2hd	hIghmbp2hd - Hg	hIghmbp2hd - I
Derivative	--	Mercury	Iodide
Wavelength (Å)	0.9795	0.99315	1.54
Resolution limit (Å)	2.5	3.0	3.0
Space group	C2	C2	C2
Cell parameters			
a/b/c(Å)	116.57/76.72 88.54	115.83/76.65/87.29	117.67/76.67/87.97
α/β/γ (°)	90/107.32/90	90/107.40/90	90/107.72/90
Unique reflections (N)	25840	14640	15006
Total Reflections (N)	114847	95240	95220
I/σ	16.0 (1.4)	13.0 (2.2)	8.4 (1.8)
Mean (I/σ)	18.9 (2.3)	24.4 (3.6)	26.0 (3.5)
Completeness (%)	91.8 (99.7)	99.9 (99.7)	99.8 (99.7)
R _{merge} ^a	0.04 (0.512)	0.044 (0.341)	0.065 (0.418)
Number of derivative sites	--	3	17
Refinement Statistics			
Data range (Å)	63.16-2.50		
Used Reflections (N)	23987		
Protein residues	625		
Protein atoms	4853		
Solvent molecules	99		
Phosphate molecules	1		
RNA bases	--		
R _{work} ^b (%)	19.02		
R _{free} ^c (%)	25.23		
R.m.s deviation			
Bond length (Å)	0.008		
Bond angles (°)	1.181		
Ramchandran plot			
(% residues)			
Allowed	98.5		
Generously allowed	1.2		
Disallowed	0.2		

Values in parentheses indicate the specific values in the highest resolution shell.

^aR_{merge} = $\sum |I_j - \langle I \rangle| / \sum I_j$, where I_j is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

^bR_{work} = $\sum ||F_o| - |F_c|| / \sum |F_c|$, where F_o denotes the observed structure factor amplitude, and F_c denotes the structure factor amplitude calculated from the model.

^cR_{free} is as for R_{work} but calculated with 5.0% of randomly chosen reflections omitted from the refinement.

Table 8-2. Data collection and refinement statistics of hIghmbp2hd-RNA

Data collection	hIghmbp2hd-RNA
Derivative	--
Wavelength (Å)	0.9760
Resolution limit (Å)	2.85
Space group	P3 ₁ 21
Cell parameters	
a/b/c(Å)	87.29/87.29/372.69
α/β/γ (°)	90/90/120
Unique reflections (N)	39,419
Total Reflections (N)	222,197
I/σ	21.6 (1.5)
Mean (I/σ)	12.5 (1.9)
Completeness (%)	90.2 (99.5)
R _{merge} ^a	0.026 (0.551)
Number of derivative sites	--
Refinement Statistics	
Data range (Å)	19.97-2.85
Used Reflections (N)	38,694
Protein residues/Protein atoms	1,222
Protein atoms	18,460
Solvent molecules	140
Phosphate molecules	2
RNA bases	17
R _{work} ^b (%)	20.58
R _{free} ^c (%)	29.28
R.m.s deviation	
Bond length (Å)	0.009
Bond angles (°)	1.338
Ramchandran plot (% residues)	
Allowed	97.2
Generously allowed	2.3
Disallowed	0.5

Values in parentheses indicate the specific values in the highest resolution shell.

^aR_{merge} = $\sum |I_j - \langle I \rangle| / \sum I_j$, where I_j is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

^bR_{work} = $\sum ||F_o| - |F_c|| / \sum |F_c|$, where F_o denotes the observed structure factor amplitude, and F_c denotes the structure factor amplitude calculated from the model.

^cR_{free} is as for R_{work} but calculated with 5.0% of randomly chosen reflections omitted from the refinement.

A

(continues on next page)

B

Figure 8-4. Ramachandran plots of hIghmbp2hd and hIghmbp2hd-RNA structures. The Ramachandran plots of hIghmbp2hd and hIghmbp2hd-RNA structures shown in **A** and **B**, respectively, were generated using PROCHECK. The plot shows the phi-psi torsion angles for all residues in the structure (except those at the chain termini). Glycine residues are separately identified by triangles as these are not restricted to the regions of the plot appropriate to the other side chain types. The red coloured region on the plot represents the most favourable combinations of phi-psi values.

A

B

Figure 8-5. Crystal structure of hIghmbp2hd-RNA. **A)** Overall view of the structure of RNA-bound hIghmbp2hd. Molecule A is in magenta while molecule B is in pink. The RNA bases are shown as green sticks. The phosphate ion in the nucleotide binding site is shown as green sphere. **B)** A partial electron density map of molecule A of hIghmbp2hd-RNA complex showing the electron density for RNA bases. The $2|F_o|-|F_c|$ map is contoured at 1.0σ .

8.3 Structural overview of the helicase domain of Ighmbp2

The conserved helicase domain of Ighmbp2 belongs to SF1B and has essentially the same fold as that of the Upf1 protein (Chakrabarti *et al.*, 2011; Cheng *et al.*, 2007). The helicase core of Ighmbp2 comprised four domains: two RecA-like domains (domains 1A and 2A) with domains 1B and 1C inserted into domain 1A (**Figure 8-6A**). Domains 1A (residues 159-270 and 347-440) and 2A (residues 441- 648) share the α - β fold with a central seven-stranded parallel β -sheet surrounded by α helices (**Figures 8-6 and 8-7**), resembling the RecA-like fold seen in all SF1 and SF2 helicases (Singleton *et al.*, 2007; Singleton and Wigley, 2002). RecA-like domains have been proposed to act as ‘motor’ to translocate along the bound nucleic acid using the energy of ATP hydrolysis, with domains 1B and 1C acting as a ‘wedge’ to destabilize the double stranded nucleic acid (Soultanas and Wigley, 2001).

Domain 1B (residues 34-141) consisting of a short α -helix and a β -barrel, is connected to domain 1A by two α -helices (residues 3-33 and 142-158) also referred to as the ‘stalk’ (Chakrabarti *et al.*, 2011). Domain 1C (residues 271-346) is inserted between the Walker A and Walker B motifs, such that it forms 4 helices situated above domain 1A (**Figures 8-6 and 8-7**). The conical motifs I, II , III, V, and VI involved in ATP binding (reviewed in (Singleton *et al.*, 2007)) and the motifs III and Va which played a role in the coordination between NTP and nucleic acid binding site (Fairman-Williams *et al.*, 2010), are present in the equivalent positions of other SF1 helicases at the domains 1A and 2A interface in hIghmbp2hd (**Figure 8-6B**).

A **B**

Figure 8-6. Domain organization and conserve sequence motif of hIghmbp2hd. **A)** Schematic showing the domain organization in hIghmbp2. Domains 1B and 1C are inserted into domain 1A. Missense mutations are labeled. **B)** The positions of the canonical motifs are shown by coloured bars and location mapped onto hIghmbp2hd crystal structure. The phosphate ion in the nucleotide binding site is shown as sphere.

A

B

Figure 8-7. Overall crystal structures of hIghmbp2hd and hIghmbp2-RNA. The colours of the domains are as per Figure 8-6A. Phosphate ions are shown as sphere, RNA is shown as orange tube. **A)** Structure of hIghmbp2hd. **B)** Structure of hIghmbp2hd-RNA.

8.4 Structural overview of hIghmbp2hd-RNA complex

8.4.1 Interactions between Ighmbp2 and the bound RNA

From the crystal structure of hIghmbp2-RNA, the RNA binding channel is gated by domains 1B and 2A at the 5'-end, and domains 1B and 1C at the 3'-end (**Figure 8-7B**). Although the crystal was obtained with a ssRNA composed of 10 adenine bases, we only observed at most nine adenine bases. For simplicity, we have numbered the bases A₁ to A₉ in a 5'→3' direction. The RNA bases observed are buried, preventing access to other RNA-processing machinery. This RNA binding site is the same as observed in other SF1 helicases.

The resolution of the structure was sufficient to visualize the interactions between the protein and the RNA. Interactions with the ssRNA were mediated mainly by the motifs (Ia and Ic) and domain 1B, via the phosphodiester backbone rather than the bases, as depicted in **Figure 8-8**. Briefly, A₁ at the extreme end is well ordered by stacking with the side chain of His411. The bases A₁ to A₃ are stacked against each other, similar to that found in an RNA duplex. As the base A₃ is perpendicular to A₄, it protrudes outwards causing a bend in the RNA backbone between A₃ and A₄. The side chain of the residue Gln507 coordinates the ribose sugar of the base A₃. The position of A₄ is determined by the interaction of Asn542 side chain with the RNA phosphodiester backbone. The residue Thr407 coordinates the nitrogenous base of A₅ which forms base stacking interactions with A₆ to A₈. Domain Ia interacts with the phosphate backbone of A₇ via the side chain of residue Asn245. The Thr351 and Thr353 residues from motif Ic interact with the phosphate backbone of A₇ and A₈, respectively. A second bend of the RNA backbone is observed when A₉, located ~10 Å from A₈, flips away from the base

stacking configuration. The side chain of Arg270 interacts with the phosphodiester backbone at this position. Except for Gln507, there are no other residues coordinating the 2'-hydroxyl of the sugar ribose ring. We speculate the lack of recognition of 2'-hydroxyl of the sugar ribose ring results in lack of specificity of Ighmbp2 for DNA or RNA. This explains the ability of the Ighmbp2 protein to unwind both DNA and RNA duplexes (Guenther *et al.*, 2009a).

It is noteworthy to mention that the position of domain 1C makes it an appropriate wedge to mechanically split the incoming duplex nucleic acid, similar to those observed in other helicases which use a pin or wedge to split the duplex nucleic acid (Buttner *et al.*, 2007; Saikrishnan *et al.*, 2008; Singleton *et al.*, 2004; Singleton *et al.*, 2001). A previous study done on Upf1 have shown the importance of domain 1C in ssRNA binding and RNA-dependent ATPase activity (Cheng *et al.*, 2007).

8.4.2 Identification of a conserved solvent accessible surface

The locations of the exposed conserved molecular surface and the electrostatic charge distribution were mapped onto the surface of the hIghmbp2hd-RNA structure. Clusters of conserved residues were found in the RNA- and ATP-binding sites, and at the ends of the RNA binding channel. Besides the conserved helicase motifs, there are 3 other conserved regions (residues 37-41, 145-152 and 412-426) and all coincides with the positively charged patches of the proteins (**Figure 8-9**).

Mapping the electrostatic potential on the molecular surface of hIghmbp2hd-RNA revealed a negatively charged patch (comprising of residues Asp119 and Glu334) on one side of the 3' end of the RNA-binding channel rim (**Figure 8-9B**). This could provide

negative charge repulsion with the bound ssRNA and prevents it from moving out of the channel. Two positively charged patches were observed: one lining the surface of domains 1A and 1B; the other on the surface of domain 1C facing the 3' end of the channel (**Figure 8-9**). These positively charged surfaces are away from the symmetry-related molecules in the ASU unit and may act either as nucleic acid or protein interacting sites.

A

(continues on next page)

B

Figure 8-8. Interactions of ssRNA with hIghmbp2hd. **A)** Stereo view of the nucleic acid binding pocket. Residues interacting with the ssRNA are shown and labeled. **B)** Schematic representation of the contacts between the hIghmbp2hd protein and RNA. Blue dotted lines represent side-chain interactions while black dotted lines represent main chain interactions. Residues are coloured by its domain (same as Figure 8-7).

Figure 8-9. Surface representation of hIghmbp2hd-RNA showing the positioning of conserved residues and the electrostatic charge distribution properties. A) Sequence conservation shown on the surface of hIghmbp2hd-RNA. Conserved residues are shown on the molecular surface from deep blue (identical) to light blue (highly similar) to cyan (similar). Other conserved regions not in the motifs are indicated in the diagram. **B)** Electrostatic surface of hIghmbp2hd-RNA (blue positive to red negative). Two highly electropositive patches are circled.

8.5 Conformational changes upon RNA binding

Superposition of RNA-bound and free structures of hIghmbp2hd showed little positional changes in the C α atoms of domains 1A and 2A (rmsd value of 0.87Å for all equivalent C α atoms) (**Figure 8-10A**), suggesting that the relative orientations of domains 1A and 2A are largely conserved in the presence and absence of RNA. Similar residues were disordered in both structures (residues 310-315 and 503-507 of RNA-free form; 308-320 and 502-506 of RNA-bound form) indicating the intrinsic flexibility of these regions.

Comparing the hIghmbp2hd structures offer insights on conformational changes that occur upon RNA binding. These changes are observed in domains 1B and 1C. In the RNA-bound form, the β -barrel of domain 1B rotates approximately 30° away from domain 1C such that it bends over domains 1A and 2A to form a channel (diameter of ~15 Å, length ~50 Å) that could accommodate a single stranded but not a duplex nucleic acid (**Figure 8-10A**). This marked conformational change widens the opening of the 3' end of the channel. In addition, a ~9 Å outward translation of a helix (residues 37-41) in domain 1B is observed upon RNA binding (**Figure 8-10A**). Interestingly, the above mentioned residues are also highly conserved in Upf1 (**Figure 8-11**). The opening of the channel requires the outwards tilting of domain 1C by 10° (**Figure 8-10A**). The importance of domain 1C in ATP hydrolysis and ssRNA binding has been previously demonstrated in hUpf1 (Cheng *et al.*, 2007). Lastly, there is a reorganization of a loop (residues 264-273, located just before domain 1C), where residue Arg270 reorientates its side chain from the nucleotide binding pocket to the RNA binding pocket (between bases 8 and 9) upon ssRNA binding (**Figure 8-10B**). Hence, Arg270 could provide a direct link between RNA binding and ATP recognition.

Figure 8-10. Conformational changes of hIghmbp2 upon RNA binding. **A)** Stereo view of hIghmbp2hd-RNA (magenta) superimposed with hIghmbp2hd (green). Phosphate ions are shown as spheres, bound ssRNA is shown in orange cartoon. The arrows indicate the movement of domain 1B and 1C in the RNA-bound state compared to the RNA-free state. **B)** Diagram shows the reorganization of a region of the loop (residues 264 to 273) and the reorientation of residue Arg270. The position of the phosphate ions indicates the location of the nucleotide binding site.

8.6 Structural comparison of Ighmbp2 with other SF1 helicases

8.6.1 Overall tertiary structure

Most helicases have functionally important accessory domains located on top of nucleic acid binding site. This distinct domain architecture and positioning could give specific functional and polarity properties. For instance, structure of the RecD2 protein has its domain 2B located on top of the cleft between the RecA-like domains (1A and 2A), whereas the Upf1 and Ighmbp2 proteins have their domain 1B at this position (**Figure 6-2**). Although both domains 2B and 1B adopt the β -barrel fold, domain 2B of RecD2 connected to domain 2A instead of domain 1A. Since Ighmbp2 is more similar to Upf1 than RecD2, the rest of the chapter will focus on the comparison between Ighmbp2 and Upf1.

Sequence and structural alignment of helicase domains of Ighmbp2 with Upf1 shows well conserved secondary structure and helicase motifs (**Figures 6-3, 8-11 and 8-12**). They have structurally similar domains 1B and 1C inserted into the helicase domain 1A. These observations validate that members of the Upf1-like family have similar domain fold. Of note, the RecA-like domains are highly conserved at both sequence and secondary structure levels. On the other hand, there is low sequence similarity in domains 1B and 1C. Hence it is not possible to predict the secondary structures of these domains of other Upf1-like proteins based on sequence alignment alone.

Figure 8-11. Sequence alignment of helicase domains of Upf1 and Ighmbp2. The sequence of human (*H.s.*), mouse (*M.m.*) and hamster (*M.a.*) Ighmbp2 were aligned with human and yeast (*S.c.*) Upf1. Residues of Upf1 interacting with the nucleotide and with the bound RNA are indicated by Δ and ●, respectively. Residues of Ighmbp2 interacting with the phosphate and with the bound RNA are indicated by Δ and ●, respectively.

Figure 8-12. Structural comparison of Ighmbp2 and Upf1. The ribbon diagrams are shown in the same orientation and are generated by superposition of domain 1A. hIghmbp2hd is shown in multi colours according the domains. Nucleotides and phosphates are shown in sticks. ssRNA is shown in tube where present. **A)** Helicase domains of RNA free structures of Ighmbp2 and Upf1. hUpf1 Δ CH-ADP:PO₄⁻ (PDB code 2GK6; (Cheng *et al.*, 2007)) is shown in grey colour. **B)** Helicase domains of RNA bound structures of Ighmbp2 and Upf1. hUpf1 Δ CH-RNA-ADP:AlF₄⁻ (PDB code 2XZO; (Chakrabarti *et al.*, 2011)) is shown in yellow colour. **C)** Comparison of RNA binding channel. The structure of *S. cerevisiae* Upf1 with RNA (scUpf1-RNA-ADP:AlF₄⁻) is shown in beige colour.

8.6.2 Nucleotide binding site

In both our structures, a phosphate ion is bound in the nucleotide binding pocket. Close inspection of the nucleotide binding site showed that the phosphate ion in the RNA-free structure is located in a position similar to that observed in the hUpf1 Δ CH-ADP:PO₄⁻ structure (PDB code 2GK6, (Cheng *et al.*, 2007)) (**Figure 8-12A**). The conserved residues Gln403, Arg443 and Arg603 from motifs III, IIIa and VI, respectively in hIghmbp2hd make direct contacts with the phosphate ion (**Figures 8-11 and 8-12A**). These residues correspond to residues Gln665, Arg703 and Arg865 in hUpf1, which have been observed to interact with the free phosphate in the hUpf1 Δ CH-ADP:PO₄⁻ structure (**Figure 8-11 and 8-12A**).

In the PcrA structure, Arg610 has been proposed to act as an ‘arginine finger’ to stabilize the transition state, thus facilitating ATP hydrolysis (Caruthers and McKay, 2002; Velankar *et al.*, 1999). Arg603 in hIghmbp2hd corresponds to residue Arg610 of PcrA. Hence the interaction of Arg603 with the phosphate ion may have a similar role in the stabilization of the transition state. In hIghmbp2, residue Gln403 is equivalent to Gln665 in hUpf1, which has been proposed to be a ‘ γ -phosphate sensor’ to detect the presence or absence of γ -phosphate in ATP and relay the information to other sites

through conformational changes (Cheng *et al.*, 2007; Story and Steitz, 1992). In both hIghmbp2hd and hIghmbp2hd-RNA structures, the relative positions of Gln403 and Arg603 remain unchanged upon RNA binding, probably due to the lack of nucleotide binding.

In the hIghmbp2hd-RNA structure, the position of the phosphate ion corresponds to that occupied by the β -phosphate of adenosine diphosphate (ADP) in the hUpf1 Δ CH-RNA-ADP:AlF₄⁻ structure (Chakrabarti *et al.*, 2011) (**Figure 8-12B**). The side chains of residues Thr221 and Arg443 in hIghmbp2hd-RNA form hydrogen bonds with the phosphate ion (**Figure 8-12B**). Similar interactions are present in the hUpf1 Δ CH-RNA-ADP:AlF₄⁻ structure between residues Thr499 and Arg703 in hUpf1 and the β -phosphate of bound ADP (**Figures 8-11 and 8-12B**).

8.6.3 RNA binding site

Although RNA binding in Ighmbp2 and Upf1 involves the similar domains, there is an obvious difference in the orientation of domain 1B. Domain 1B of hIghmbp2hd-RNA was found to be perpendicular to that of RNA-bound Upf1 structures (hUpf1 Δ CH-RNA-ADP:AlF₄⁻ and scUpf1-RNA-ADP:AlF₄⁻) (**Figure 8-12B**). A possible explanation is that the presence of AlF₄ in the γ -phosphate site brings domain 2A closer to domains 1A and 1B. This results in steric hindrance between domains 2A and 1B at the 5' end of RNA binding channel. In order for RNA to bind, a drastic rotation of domain 1B away from domain 2A is required. In the case of hIghmbp2hd with an empty γ -phosphate site, domains 2A and 1B are observed to be further apart. Hence, a 30° rotation of domain 1B is sufficient to allow RNA binding.

There was also significant difference in the number of bases the RNA binding pocket can accommodate. Upf1 lacking the CH domain (hUpf1 Δ CH-RNA-ADP:AlF₄⁻) could only bind to nucleic acid of 6 bases long compare to 9 bases observed in the hIghmbp2hd-RNA structure. In the presence of CH domain, domain 1B is pushed towards domain 1C and the RNA binding channel of Upf1 is extended to accommodate 8 bases (**Figure 8-12C**). In this conformation, Upf1 binds more tightly to the 3' end of the RNA, 'clamping' it and reduces helicase activity. Ighmbp2 lacks the CH domain but has a R3H domain located downstream of the helicase core domain. Given that the R3H domain has been proposed to bind nucleic acids, it is envisaged that the R3H domain might have a similar role in regulating Ighmbp2 catalytic activities.

8.7 R3H domain has a regulatory role on Ighmbp2

Based on a previous study done by Fukita and co-workers (1993), a segment (residues 638 to 786) in Ighmbp2 was identified to binds guanine-rich, single stranded DNA (Fukita *et al.*, 1993). In our structure, the ssRNA lies in the channel formed by the N-terminal helicase domain. This raised the possibility that a second RNA binding site exists in Ighmbp2 and involves the residues 638 to 786. This could explain why R3H domain alone is not sufficient for high affinity binding, and that the presence of both binding sites would enhance RNA binding (Fukita *et al.*, 1993). Although the region containing residues 638 to 648 could be observed in our structures (**Figure 8-9**), it is difficult to accurately predict where the putative second nucleic acid binding site (or R3H domain) might lie due to the presence of a long flexible random coil (~30 residues) between the helicase and the R3H domains.

To gain a better understanding on the role of R3H domain, we expressed and purified the protein fragment composed of the helicase core domain and the R3H domain from mouse mIghmbp2 (mIghmbp2hd-R3H) since the human protein containing the similar regions degraded severely during purification.

Data from SPR experiments showed that both hIghmbp2hd and mIghmbp2hd-R3H bind RNA better in the absence of nucleotides (**Figure 8-13B and C**), similar to that observed for Upf1 (Chakrabarti *et al.*, 2011; Chamieh *et al.*, 2008; Weng *et al.*, 1998). Consistent with this observation, our RNA-bound structure provides structural evidence that Ighmbp2 is able to bind RNA in the absence of nucleotides.

Importantly, SPR data also showed that mIghmbp2hd-R3H binds more strongly to ssRNA compared to hIghmbp2hd, in the absence or presence of nucleotides (**Figure 8-13**). The presence of nucleotides reduce the binding of hIghmbp2hd to RNA by half whereas mIghmbp2hd-R3H is influenced to a lesser extent. These observations support the notion that Ighmbp2 bears two nucleic acid interacting sites, one lies in the helicase region of the protein and is largely influenced by nucleotide binding while the other one is located in the R3H domain and is probably not affected by nucleotide binding. Both helicase and R3H domains might work in a cooperative manner to enhance RNA binding.

Figure 8-13. Sensograms of SPR experiments testing for RNA binding in hIghmbp2hd and mIghmbp2hd-R3H. (A) SPR data comparing RNA binding of hIghmbp2hd and mIghmbp2hd-R3H in the absence of nucleotides. Insert shows the binding profile of hIghmbp2hd. (B) SPR data shows the changes of RNA binding of hIghmbp2hd in the presence of nucleotide. (C) SPR data shows the changes of RNA binding of mIghmbp2hd-R3H in the presence of nucleotide. 100 nM of protein was used in all experiments. Note the difference in the scale of the y-axis. The experiments were repeated with 300 nM of protein and similar binding trend was observed (data not shown).

8.8 Structural and functional placement of SMARD1 missense mutations on Ighmbp2

8.8.1 SMARD1 missense mutations in Ighmbp2

All published clinically observed mutations (including nonsense and deletion mutations) associated with SMARD1 are summarized in **Table 6-1**. From the list of mutations, only nine were missense mutations namely: Gln196Arg, Thr221Ala, Cys241Arg, Glu382Lys, His445Pro, Thr493Ile, Asp565Asn, Asn583Ile and Arg603His. We mapped these missense mutants onto the hIghmbp2hd-RNA structure and analyzed each of them in detail (**Table 8-3, Figure 8-14**). Based on the locations of the missense mutant in the structure, they are further classified as either at the nucleotide binding pocket or at RNA binding channel, or at neither ligand binding sites.

8.8.2 Mutants at the nucleotide binding site

Three of the above mutants located at the nucleotide binding site are: Gln196Arg, Thr221Ala and Arg603His (**Figure 8-14E**).

Residue Gln196 is the invariant glutamine residue in the Q motif, a motif previously proposed to be unique to the DEAD box family of helicases to control ATP binding and hydrolysis (Cordin *et al.*, 2004; Tanner *et al.*, 2003). When the non-

hydrolyzable ATP analog (AMPPNP) from the Upf1 structure is superimposed onto Ighmbp2 nucleotide binding pocket, residue Gln196 is shown to form hydrogen bonding with the adenine base of AMPPNP (**Figure 8-14E**). Mutation of Gln196 to an arginine residue would result in a longer side chain extending into the nucleotide binding pocket. This will cause steric hindrance for ATP binding and may lead to loss of ATPase activity.

Residue Thr221 is part of the Walker A sequence in motif I and could form hydrogen bonding with the α - and β -phosphate groups of modeled AMPPNP (**Figure 8-14E**). Motif I is well characterized to be important for nucleotide binding and catalyze its hydrolysis (Caruthers and McKay, 2002; Cordin *et al.*, 2006). Hence mutation of Thr221 to alanine will inevitably impact ATPase activity.

The third missense mutations located at the nucleotide binding pocket is mutant Arg603His in motif VI (**Figure 8-14E**). This motif was shown to function in ATP hydrolysis and RNA binding (Caruthers and McKay, 2002; Pause *et al.*, 1993). From hIghmbp2hd-RNA structure, residue Arg603 forms electrostatic interactions with the γ -phosphate of AMPPNP. As the side chain of histidine is shorter, the mutant Arg603His is not be able to interact with the nucleotide. As previously discussed, this residue could act as an ‘arginine finger’ to stabilize the transition state as observed in PcrA (Caruthers and McKay, 2002; Velankar *et al.*, 1999). It was reported that mutating the corresponding conserved arginine residue in eIF4A will abolish its RNA binding and ATPase and helicase activity (Pause *et al.*, 1993). Therefore, mutating Arg603 to a histidine residue should lead to the loss of ATPase activity.

8.8.3 Mutants at the RNA binding channel

Two of the nine missense mutants, Asp565Asn and Asn583Ile, are located in close proximity to the RNA binding channel. It is conceived that these two mutations would affect Ighmbp2 ability to bind RNA and/or unwinding activity.

Residue Asp565 lies in motif V which is a flexible loop in domain 2A (**Figure 8-14B**). Although the function of motif V is less well studied, there is evidence that it is involved in binding to substrate RNA (Cordin *et al.*, 2006). The hIghmbp2hd-RNA structure shows that residue Asp565 is hydrogen bonded to the ribose 2' hydroxyl group of A₄ and forms salt bridge with the basic residue Arg596 (**Figure 8-14B**). Residues Asp565 and Arg596 are conserved in Upf1 (residues Asp827 and Arg858 of human Upf1) and show similar interactions in the RNA-bound Upf1 structure. Substitution of Asp565 with asparagine would disrupt electrostatic interactions between Asp565 and Arg596 but should not interfere with the ability to form hydrogen bond with the RNA backbone. Thus, it is highly possible that the mutation Asp565Asn would not affect RNA binding.

The mutant Asn583Ile is located near the 5' end of the RNA binding channel but Asn583 does not directly interact with the bound RNA (**Figure 8-14A**). We observed that Asn583 is surrounded by several basic residues Arg581, Lys585 and Arg584. Mutation of Asn583 to isoleucine would place a hydrophobic residue into a positively-charged groove formed by these three residues, thus destabilizing the protein.

8.8.4 Mutants at neither ligand binding site

The remaining four missense mutants (Cys241Arg, Glu382Lys, His445Pro and Thr493Ile) are located neither at the nucleotide binding site nor the RNA binding channel.

Cys241 is located at the central β -sheet of domain 1A and surrounded by several hydrophobic residues (Leu239, Val348, Phe369, Val372, Ile374 and Pro388) (**Figure 8-14C**). Therefore, mutation of Cys241 to a positively-charged arginine residue could with high possibility disrupt the β -sheet, hence affecting the structural stability of the RecA-like domain 1A.

Residue Glu382 lies on the helix just before motif II. It forms hydrogen bonding with Trp386 in the same helix and forms salt-bridge interaction with Arg425 in a neighbouring helix, all within domain 1A (**Figure 8-14D**). Mutation of Glu382 to lysine (which has a longer and positively charged side chain) would result in charge-charge repulsion with Arg425 and also could disrupt the hydrogen bonds formed between Glu382 and Trp386. Hence it is likely that this mutation would reduce the stability of domain 1A.

Residue His445 is well conserved in mouse, human and hamster Ighmbp2 (**Figure 8-11**). It is located in motif IIIa which is the loop linking domains 1A and 2A. Since the motor domains 1A and 2A move relative to each other in response to ATP binding/hydrolysis critical for the SF1 helicases (Ye *et al.*, 2004), this loop might act as a hinge to regulate the conformational changes between domains 1A and 2A. Mutation of His445 to proline would reduce the flexibility of this loop, thus restraining the movement of motor domains 1A and 2A and affecting the ATPase activity.

Residue Thr493 is located on a loop region away from the nucleotide binding site and is not located in proximity to any helicase motifs. In addition, the side chain of Thr493 does not make significant contacts with neighbouring residues. From the structural point of view, it is difficult to predict how Thr493Ile mutation would affect the catalytic activities of Ighmbp2. Collectively, the mutants Cys241arg, Glu382Lys and His445Pro could affect Ighmbp2 catalytic activities probably through reducing protein stability.

8.8.5 Catalytic activities of Ighmbp2 mutants

The mutational effects predicted based on structural analysis are consistent with the functional data of mutant full-length Ighmbp2 carried out by Guenther *et al.* (2009) except for mutant Asp565Asn (**Table 8-3**).

The mutants at the nucleotide binding site (Gln196Arg, Thr221Ala and Arg603His), at the RNA binding site (Asn583Ile) and those that affect protein stability (Cys241Arg, Glu382Lys and His445Pro) all displayed deficient ATP hydrolysis. Consistent with the observation that Ighmbp2 is an ATP-dependent helicase, these mutants also showed defective RNA and DNA duplex unwinding activity (**Table 8-3**).

Interestingly, mutant Asp565Asn which was predicted to reduce RNA binding, retained the ability to bind nucleic acid. It has an ATPase activity similar to the wild type despite its helicase activity being abolished. As unwinding of a nucleic acid duplex always requires translocation, the inability of mutant Asp565Asn to translocate provides an explanation for the observed catalytic defects described above.

The only mutant with minimal effect on Ighmbp2 catalytic activities is Thr493Ile. It was reported that the Thr493Ile mutant has lower protein level in the cells due to reduced intracellular steady state of the protein (Guenther *et al.*, 2008). Therefore the pathogenicity of this mutant is caused by the lowering of Ighmbp2 protein level instead of its enzymatic activity.

Table 8-3. Catalytic activities of mutant Ighmbp2.

Mutations ^a	Amino Acid substitutions	Helicase motif	RNA binding ^b	DNA binding ^b	ATPase activity ^b	Unwinding activity ^b	References
587 A→G	Gln196Arg	Q	+	-	-	-	(Guenther <i>et al.</i> , 2009a; Maystadt <i>et al.</i> , 2004)
661 A→G	Thr221Ala	I	++	++	-	-	(Grohmann <i>et al.</i> , 2003; Guenther <i>et al.</i> , 2009a)
721 T→C	Cys241Arg	Ia	+	-	-	-	(Grohmann <i>et al.</i> , 2003; Guenther <i>et al.</i> , 2009a)
1144 G→A	Glu382Lys	Near II	+	+	-	-	(Grohmann <i>et al.</i> , 2003; Guenther <i>et al.</i> , 2009a; Pierson <i>et al.</i> , 2011)
1334 A→C	His445Pro	IIIa	+	+	-	-	(Guenther <i>et al.</i> , 2009a; Guenther <i>et al.</i> , 2007)
1478 C→T	Thr493Ile	-	+	+	++	++	(Guenther <i>et al.</i> , 2009a; Guenther <i>et al.</i> , 2009b; Joseph <i>et al.</i> , 2009)
1693 G→A	Asp565Asn	V	++	++	++	-	(Grohmann <i>et al.</i> , 2003; Guenther <i>et al.</i> , 2009a; Maystadt <i>et al.</i> , 2004)
1748 A→T	Asn583Ile	Near Vb	+	-	-	-	(Grohmann <i>et al.</i> , 2003; Guenther <i>et al.</i> , 2009a)
1808 G→A	Arg603His	VI	-	-	-	-	(Grohmann <i>et al.</i> , 2003; Guenther <i>et al.</i> , 2009a)

^aNucleotide numbering is according to the standard nomenclature starting from the translated part of the Ighmbp2 mRNA with +1 as the A of ATG. All mutations are missense mutations.

^bFunctional studies of full-length Ighmbp2 protein (Guenther *et al.*, 2009a), complete loss of activity (-); moderate reduction of activity (+); full activity (++); not analyzed (N.A).

Chapter 9

Discussion

Previous studies have showed that the Ighmbp2 protein belongs to SF1B based on its sequence alignment and the ability to unwind both RNA and DNA duplex in the 5' → 3' direction (Guenther *et al.*, 2009a). From sequence alignment, the closest related proteins are the Upf1 and the senataxin proteins. Indeed, our crystal structure of hIghmbp2hd showed that Ighmbp2 is structurally similar to Upf1 which belongs to the Upf1-like subfamily. Unlike other members of SF1B, Upf1-like proteins have a β -barrel domain 1B and an additional domain 1C inserted into domain 1A. To obtain a better understanding of how nucleic acid binds to Ighmbp2, the structure of hIghmbp2-RNA was determined.

By comparing the RNA-free and the RNA-bound states of hIghmbp2hd, nucleic acid binding involves conformational changes in domains 1B and 1C. This is in line with the Upf1 helicase whereby conformation changes are observed in domain 1B upon nucleic acid binding (Chakrabarti *et al.*, 2011) and deletion of domain 1C abolished nucleic acid binding (Cheng *et al.*, 2007). Our attempt to crystallize hIghmbp2hd-RNA in complex with AMPPNP or ADP was unsuccessful. Given the similarity with Upf1, we took advantage of the available Upf1 structures and postulate that nucleotide binding will bring the motor domains (1A and 2A) closer. In order for RNA to bind, a large rotation ($\sim 90^\circ$) of domain 1B similar to that observed in Upf1 is required. The exact effect of nucleotide binding awaits the structure of Ighmbp2 in complex with RNA and nucleotide.

Many helicases have C-terminal or N-terminal accessory domains. These domains usually have defined folds and could influence the function of a helicase. These terminal

domains could also add physiological specificity to the helicase. In comparison to full length Upf1 which has an N-terminal CH domain, full length Ighmbp2 has an accessory C-terminal R3H domain and an AN1-type zinc-finger motif. In Upf1, the CH domain has been shown to play a cis-inhibitory role on the helicase activity of Upf1 by ‘clamping’ onto the RNA and inhibiting unwinding activity (Chakrabarti *et al.*, 2011).

To establish a role for R3H domain, we tested RNA binding of hIghmbp2hd and mIghmbp2hd-R3H. Both hIghmbp2hd and mIghmbp2hd-R3H bind RNA best in the absence of nucleotides (**Figure 8-13**), similar to that observed in Upf1 which binds RNA with highest affinity in the absence of nucleotides (Chakrabarti *et al.*, 2011). The difference in RNA binding could be due to conformational changes associated with nucleotide binding. Interestingly, SPR data showed that mIghmbp2hd-R3H binds nucleic acid better than hIghmbp2hd, suggesting that the R3H domain enhances RNA binding of Ighmbp2. This further corroborates the importance and the functional flexibility of the accessory domains.

The results described above imply that the R3H domain of Ighmbp2 protein might act either as an allosteric activator by enhancing RNA binding of the helicase domain or as a second nucleic acid interacting site. The notion that R3H domain can directly interact with nucleic acid is supported by its NMR structure of the R3H domain (Liepinsh *et al.*, 2003). The structure containing residues 725 to 786 is made up of three anti-parallel β -strands, sandwiched between two helices (**Figure 6-4**). As discussed by Liepinsh and co-workers, there are three conserved basic residues (Arg755, His759 and Arg778) clustered on one face of the R3H domain which could be a binding site for phosphodiester backbones of DNA or RNA (**Figure 6-4**) (Liepinsh *et al.*, 2003). Assuming that the R3H

domain does bind to DNA or RNA, it will be interesting to find out whether it binds to the same strand as seen in the hIghmbp2hd-RNA structure or to the orphan strand of the unwound nucleic acid duplex. The structure of Ighmbp2hd-R3H in complex with RNA is required to address this issue. It would also give insights as to how R3H domain modulates Ighmbp2 catalytic activity. We also cannot rule out the possibility that R3H may affect translocation activity independent of helicase activity.

Thus far, there are two ways to regulate helicase activity, i.e. via nucleotide binding and influence of accessory domains, as observed in the Upf1 protein. The enzymatic activities of Upf1 are modulated by CH domain, which is in turn regulated by interaction with the Upf2 protein. As for Ighmbp2, its R3H domain has a hydrophobic area opposite the electropositive surface (**Figure 6-4**). This could act as a protein-binding which in turn regulate Ighmbp2 activities through interactions with other proteins. Further work needs to be done to establish the interacting partners and how they influence helicase activity and prognosis of SMARD1.

The structure of Ighmbp2 helicase provides an opportunity to address how genetic defects in patients with SMARD1 impact the molecular function of the protein. Available information indicates specific mutations do not correlate with the severity of the clinical features (Grohmann *et al.*, 2003; Joseph *et al.*, 2009). From our structure, not all clinical mutations in Ighmbp2 are located in the nucleotide or nucleic acid binding sites (**Figure 8-14**). Seven out of the nine missense mutations affects ATPase activity either through disrupting ATP binding/hydrolysis or through reducing the structural stability of the helicase motors (**Table 8-3**). Of particular interest are the mutants Thr493Ile and Asp565Asn which did not affect ATPase activity of the enzyme. Instead of causing

enzymatic dysfunction of Ighmbp2, mutant Thr493Ile causes neuronal degeneration through reduced intracellular Ighmbp2 protein levels. As for mutant Asp565Asn, its enzymatic defect is most likely caused by the inability of Ighmbp2 to translocate along the bound RNA. Based on our current level of understanding, identification of mutation sites in Ighmbp2 cannot deterministically predict SMARD1 risk and phenotype.

There have been increasing interests and studies done on the senataxin protein known to be involved in the neurodegenerative diseases ALS4 and AOA2 (Chen *et al.*, 2004; Chen *et al.*, 2006; Moreira *et al.*, 2004; Suraweera *et al.*, 2009). Sequence alignment shows that mutant Cys2006Tyr in senataxin (Anheim *et al.*, 2009) corresponds to Cys241Arg mutation in Ighmbp2. As mutations in senataxin and Ighmbp2 result in different neuronal related diseases, this will suggest that the two proteins regulate different downstream events with different specificity. The molecular function of Ighmbp2 is an open question for future studies.

Chapter 10

Future directions

The challenge of finding a cure for SMARD1 will first require a better understanding of how Ighmbp2 gene expression is regulated and how downstream events are regulated by the Ighmbp2 protein. The discovery of Ighmbp2 as an ATP-dependent helicase brings speculation to its cellular function. Further studies need to be done to decipher its biological role. This includes identifying the signaling pathway that regulates Ighmbp2 gene expression, and identifying the biological targets of Ighmbp2 protein. The availability of the structure of Ighmbp2 helicase domain allows designs of site-directed mutagenesis experiments for a more detailed functional characterization.

As we have seen, accessory domains at either ends of the helicase core have unique roles in regulating the protein enzymatic activity and specificity. The R3H domain of Ighmbp2 may give the protein its unique cellular location and function through various protein-protein interactions which can be tested by *in vivo* experiments.

Accession Numbers

The coordinates and structure-factor amplitudes of RpfF have been deposited in the Protein Data Bank with accession codes 3M6N.

The coordinates of hIghmbp2hd and hIghmbp2hd-RNA are yet to be deposited.

Appendix I

Bacterial culture media

LB Culture Medium

0.5% (w/v) Bacto-yeast extract, 1% (w/v) Bacto-tryptone, 1% (w/v) NaCl

LB Agar

Liquid LB medium supplemented with 1.5% agar

MOPS Minimal Medium (Neidhardt *et al.*, 1974)

(1) 10x MOPS buffer (1L)

Reagents	Concentration (mM)	Quantity (g)
MOPS	400	83.72
Tricine	40	7.17
FeSO ₄ •7H ₂ O	0.1	0.028
NH ₄ Cl	95	5.08
K ₂ SO ₄	2.76	0.48
CaCl ₂	0.005	0.000555
MgCl ₂	5.3	0.0011
NaCl	500	29.22

Adjust to pH 7.4 with KOH, sterilize with a 0.22 µm filter.

(2) 5X Amino Acids stock (1L)

Reagents	Concentration (mM)	Quantity (mg)
Alanine	4	356.4
Arginine	2.5	526.75
Asparagine	2	264.2
Aspartic	2	266.2
Cysteine	0.5	60.6
Glutamic acid	3	441.3
Glutamine	3	438.3
Glycine	4	300.4
Histidine	1	155.2
Isoleucine	4	524.8
Leucine	8	1049.6
Lysine	4	730.3
Phenylalanine	2	330.4
Proline	2	230.2
Serine	18.3	1922.6
Threonine	2	238.2
Tryptophan	0.5	102.1
Tyrosine	2	362.4

Valine	6	703.2
Adenine	1	171.6
Guanine	1	151.1
Cytosine	1	111.1
Uracil	1	112.1
Thymine	1	126.1

Sterilize with a 0.22 μ m filter.

(3) 5000X Micro nutrients (500 ml)

Reagents	Quantity (g)
Na ₂ MoO ₄ •2H ₂ O	0.463
H ₃ BO ₃	0.618
CoCl ₂	0.178
CuSO ₄	0.043
MnCl ₂	0.396
Zn(OAc) ₂	0.072

Sterilize with a 0.22 μ m filter.

(4) 2000x Vitamin mix (10 ml)

Reagents	Quantity (g)
Thiamine	67.4
Pantothenic acid	47.6
<i>p</i> -hydroxybenzoic acid	32
<i>p</i> -aminobenzoic acid	27.4
2,3-dihydroxybenzoic acid	30.8

Add a little NaOH to help in solubility. Sterilize with a 0.22 μ m filter.

(5) MOPS minimal medium (10 L)

Reagents or Buffers	Volume (ml)
10X MOPS buffer	1000
20% Glucose	300
5X Amino acid stock	2000
Water	6600
0.132M K ₂ PO ₄ *	100
2000X Vitamin stock	5
5000X Micronutrients	0.2

*0.132M K₂PO₄ is to be added after adding water. Add 1 g of selen-methionine to 10 L of culture. Sterilize with a 0.22 μ m filter.

Appendix II

Publication list and reprint

Structural basis of the sensor-synthase interaction in autoinduction of the quorum sensing signal DSF biosynthesis. *Structure* 18(9): 1199-1209 (2010)
Cheng Z, He YW, Lim SC, Qamra R, Walsh MA, Zhang LH and Song H

Structural Basis of the Sensor-Synthase Interaction in Autoinduction of the Quorum Sensing Signal DSF Biosynthesis

Zhihong Cheng,^{1,5} Ya-Wen He,^{1,5} Siew Choo Lim,^{1,2,5} Rohini Qamra,^{1,5} Martin A. Walsh,³ Lian-Hui Zhang,^{1,4,*} and Haiwei Song^{1,2,4,*}

¹Institute of Molecular and Cell Biology, 61 Biopolis Drive, Proteos, Singapore 138673

²School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551

³MRC France, CRG BM14, ESRF, B.P.220, F-38043 Grenoble CEDEX, France

⁴Department of Biological Sciences, National University of Singapore, 14 Science Drive, Singapore 117543

⁵These authors contributed equally to the work

*Correspondence: lianhui@imcb.a-star.edu.sg (L.-H.Z.), haiwei@imcb.a-star.edu.sg (H.S.)

DOI 10.1016/j.str.2010.06.011

SUMMARY

The diffusible signal factor (DSF)-dependent quorum sensing (QS) system adopts a novel protein-protein interaction mechanism to autoregulate the production of signal DSF. Here, we present the crystal structures of DSF synthase RpfF and its complex with the REC domain of sensor protein RpfC. RpfF is structurally similar to the members of the crotonase superfamily and contains an N-terminal α/β spiral core domain and a C-terminal α -helical region. Further structural and mutational analysis identified two catalytic glutamate residues, which is the conserved feature of the enoyl-CoA hydratases/dehydratases. A putative substrate-binding pocket was unveiled and the key roles of the residues implicated in substrate binding were verified by mutational analysis. The binding of the REC domain may lock RpfF in an inactive conformation by blocking the entrance of substrate binding pocket, thereby negatively regulating DSF production. These findings provide a structural model for the RpfC-RpfF interaction-mediated QS autoinduction mechanism.

INTRODUCTION

A quorum sensing mechanism allows bacteria to sense their population density and synchronize individual behavior into cooperative community behavior (Bassler and Losick, 2006; Von Bodman et al., 2003; Zhang and Dong, 2004) which provides bacterial pathogens an obvious competitive advantage over their hosts in pathogen-host interaction. The most critical feature of quorum sensing (QS) could be the molecular mechanism that enables bacterial cells to autoregulate the production of QS signals. In the QS system of *Vibrio fischeri*, which is considered as the paradigm of QS in Gram-negative bacteria (Milton, 2006; Whitehead et al., 2001), the QS signal AHL functions as a ligand to its cognate transcription factor LuxR. At low population density, each cell in the bacterial population produces a basal

level of AHL signals via the AHL-synthase LuxI. When a “quorum” population density is reached, the accumulated AHL signals interact with LuxR proteins; the resultant AHL-LuxR complex activates the transcriptional expression of the QS-dependent genes as well as the *luxI* gene, leading to the boosted production of AHL signals (Dong et al., 2007). This elegant signal autoinduction mechanism enables bacterial cells to sense their population density, to synchronize the expression of QS-regulon within the community in an effective way, and allows resetting of the QS circuit when a portion of bacterial cells are transferred to a new environment.

In addition to the well-characterized AHL-type QS system, a diffusible signal factor (DSF)-dependent QS system has recently been identified in a range of plant and human bacterial pathogens (Barber et al., 1997; Boon et al., 2008; Colnaghi Simionato et al., 2007; Fouhy et al., 2007; Huang and Wong, 2007; Wang et al., 2004). The DSF-type QS system, which plays a key role in regulation of bacterial virulence in various pathogens, was initially discovered from bacterial pathogen *Xanthomonas campestris* pv. *campestris* (Xcc) (Barber et al., 1997). Genetic and biochemical analyses show that the pathogen has evolved a novel form of autoregulation mechanism that allows DSF, which has been characterized as cis-11-methyl-2-dodecenoic acid, to autoregulate its biosynthesis (He et al., 2006a). This autoinduction mechanism involves two proteins, i.e., the DSF synthase RpfF and the membrane-associated DSF sensor RpfC. Mutation of *rpfF* abolishes DSF production and results in reduced virulence factor production (Barber et al., 1997; He et al., 2006b), whereas disruption of *rpfC* resulted in contrasting phenotypes. The *rpfC* mutant synthesizes about 16-fold higher DSF signal than the wild-type Xcc but produces significantly reduced virulence factors in a level similar to the *rpfF* mutant (Wang et al., 2004). Similar to the AHL-type QS signal production, the DSF level in Xcc increases proportionally following the increment of bacterial population density (Barber et al., 1997; Wang et al., 2004). However, the transcript level of *rpfF* remains more or less constant throughout growth and is not influenced by exogenous addition of DSF (Barber et al., 1997; He et al., 2006b), suggesting that autoregulation of DSF biosynthesis unlikely occurs at the transcriptional level.

RpfC is a hybrid sensor consisting of multidomains including a transmembrane domain, a histidine kinase (HK) domain,

a CheY-like receiver (REC) domain, and a histidine phosphotransferase (HPT) domain. Knocking out *rpfC* results in decreased virulence factor production but enhanced DSF biosynthesis (Slater et al., 2000; Wang et al., 2004). Our recent study has shown that mutation of the key residues implicated in phosphorelay in RpfC decreases virulence factor but has no effect on DSF production, whereas overexpression of its REC domain alone abolishes DSF biosynthesis (He et al., 2006a). In addition, coimmunoprecipitation and far western analysis showed that RpfF binds specifically to the REC domain of RpfC (He et al., 2006a). These findings suggest a model in which RpfC employs two different mechanisms to regulate virulence factor and DSF production simultaneously; i.e., the hybrid sensor relies on the conserved phosphorelay mechanism and its cognate response regulator RpfG to induce the expression of virulence genes, and suppresses DSF biosynthesis by its REC domain via a novel mechanism involving the RpfC/RpfF interaction.

In order to investigate the mechanism of how DSF autoinduction is mediated by the RpfF/RpfC interaction, we determined the crystal structures of full-length RpfF alone and in complex with the REC domain of RpfC. RpfF in complex with the REC domain adopts a crotonase-like fold, which consists of four C-terminal helices, that is essentially same as that observed in its apo form. The binding of the RpfC REC domain appears to block ligand entrance to the active site of RpfF, thereby negatively regulating DSF production. Structural comparison combined with mutagenesis sheds light on the mechanism that governs DSF autoinduction and further enriches our understanding on the diversity of bacterial QS systems.

RESULTS

Structure Determination

The crystal structure of full-length RpfF has been solved by the single-wavelength anomalous dispersion (SAD) method at a resolution of 1.8 Å. There are three polypeptide chains per asymmetric unit (AU) with a three fold noncrystallographic symmetry (NCS) axis. Residues of 1–13, 34–41, and 279–289 are disordered in all three chains of RpfF. The three subunits are virtually identical and can be superimposed upon one another with an average root mean square deviation (rmsd) less than 0.3 Å.

The crystal structure of full-length RpfF in complex with the REC domain of RpfC (designated as RpfF/REC) has been determined at a resolution of 2.5 Å by molecular replacement using the structure of RpfF as the search model. A representative portion of the initial electron density map in the region of the RpfC REC domain is shown in Figure S1 (available online). The final model of the RpfF/REC complex contains three complexes in the AU (RpfF: chains A, B, and C; REC: chains D, E, and F), which are related by a 3-fold NCS axis. Residues of 1–13, 34–41, and 279–289 in each subunit of the RpfF molecules are disordered. Residues of 449–461, 484–486, and 582–590 in chain D, residues of 449–461, 481–486, and 581–590 in chain E, and residues of 449–461, 482–486, and 582–590 in chain F are disordered. Since no substantial differences are observed between the three complexes in the AU (pairwise rmsd values of 0.42 and 0.44 Å for RpfF and REC, respectively, when all the equivalent C α atoms are superimposed), all the subsequent analysis uses the coordi-

Table 1. Data Collection and Refinement Statistics

	RpfF	RpfF/REC
Data Collection		
Derivative	SeMet	—
Wavelength (Å)	0.9795	0.9795
Resolution limit (Å)	1.8	2.5
Space group	P2 ₁ 2 ₁ 2 ₁	P6 ₅
Cell parameters		
a/b/c (Å)	96.8/112.3/119.6	130.9/130.9/156.5
$\alpha/\beta/\gamma$ (°)	90/90/90	90/90/120
Unique reflections (N)	117918	35459
I/ σ	11.3 (2.9)	11.5 (2.5)
Completeness (%)	97.5 (85.2)	99.4 (99.8)
R _{merge} ^a	0.094 (0.314)	0.068 (0.418)
Number of Se sites	14	—
Anomalous phasing power	1.46	—
Figure of merit	0.45/0.92	—
Before/after density modification	—	—
Refinement Statistics		
Data range (Å)	30.0–1.8	20.0–2.5
Used reflections (N)	111,882	56,289
Nonhydrogen atoms	6142	8680
R _{work} ^b (%)	22.8	24.7
R _{free} ^c (%)	24.7	27.8
Rmsd		
Bond length (Å)	0.002	0.004
Bond angles (°)	0.432	0.72
Ramachandran plot (% residues)		
Allowed	99.7	96.0
Generously allowed	0.3	3.7
Disallowed	0	0.3

Values in parentheses indicate the specific values in the highest resolution shell.

^a $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i$, where I_i is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

^b $R_{\text{work}} = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o denotes the observed structure factor amplitude, and F_c denotes the structure factor amplitude calculated from the model.

^c R_{free} is as for R_{work} but calculated with 5.0% of randomly chosen reflections omitted.

ates of chain A of RpfF and chain D of the REC domain. Statistics of structure determination and refinement are summarized in Table 1 (see Experimental Procedures).

Overall Architectures of RpfF and the REC Domain of RpfC

RpfF, in both free form and in complex with the REC domain of RpfC, adopts essentially the same fold with the rmsd of 0.42 Å when all the C α atoms from both forms are superimposed. As shown in Figures 1A and 1B, RpfF contains an N-terminal spiral core domain and a C-terminal α -helical region. The spiral fold in RpfF contains four turns, each of which consists of three continuous secondary structure elements, namely, two β strands

Figure 1. Structure of RpfF Alone and in Complex with the REC Domain of RpfC

Ribbon diagrams showing the self-association fold of RpfF (A) and the overall structure of the RpfF/REC complex (B). The N-terminal α/β spiral domain is colored in orange, linker region in green, and the C-terminal α -helical region in red. The REC domain of RpfC is shown in magenta with residue Asp512 in stick model. See Figure S1, which shows a representative portion of the Fo-Fc electron density map for the RpfC REC domain that is missing in the search model.

and one α helix. Two β strands of each turn are categorized into two parallel β sheets separately, named β sheet A and B. Thus, after the very beginning $\beta A0$, which antiparallels with β sheet A, the following spiral core domain contains four turns: $\beta A1/\beta B1/\alpha 1$, $\beta A2/\beta B2/\alpha 2B/\alpha 2A$, $\beta A3/\beta B3/\alpha 3$, and $\beta A4/\beta B4/\alpha 4B/\alpha 4A$. The link region between N-terminal spiral fold and C-terminal α helix consists of two α helices ($\alpha 5$ and $\alpha 6$) interspersed with two β strands ($\beta B5$, antiparalleled with β sheet B, and $\beta A5$ lined up with β sheet A in parallel). The C-terminal α -helical domain is composed of four α helices wrapping around the spiral fold successively in such a way that $\alpha 7$ and $\alpha 10$ are facing the outer faces of β sheets A and B, respectively, with $\alpha 8$ and $\alpha 9$ in between.

The REC domains are the dominant molecular switches in bacterial two components signaling pathways. Several structures of nonphosphorylated REC domains (Baikalov et al., 1996; Djordjevic and Stock, 1998; Feher et al., 1997; Sola et al., 1999; Stock et al., 1989; Volkman et al., 1995), and two of phosphorylated REC domains have been determined (Casino et al., 2009; Kern et al., 1999). These structures indicated that the REC domains share a common fold and phosphorylation induced a large conformational change. As expected, the REC domain of RpfC in the RpfF-REC complex adopts a similar 5 α /5 β fold: parallel β sheet ($\beta 2/\beta 1/\beta 3/\beta 4/\beta 5$) surrounded by three helices ($\alpha 2/\alpha 3/\alpha 4$) on one side and two helices ($\alpha 1/\alpha 5$) on the other side (Figure 1B). The phosphoacceptor Asp512 of the REC domain is located on the loop $\beta 3$ - $\alpha 3$. Structural comparison

showed that a Mg^{2+} from the crystallization buffer occupies the position of the phospho-group and is coordinated with residues Asp468, Asp469, Lys566, and His514.

RpfF Belongs to the enoyl-CoA Hydratase/Isomerase Family

A search of Protein Data Bank using the Dali server revealed more than 50 structural homologs with high Z-score (>20), all belonging to the crotonase superfamily. The homologs with solved crystal structures include rat mitochondrial enoyl-CoA hydratase (Engel et al., 1996, 1998) (PDB: 1DUB/2DUB, Z-score 21.2), 4-(N, N-dimethylamino) cinnanoyl-CoA hydratase (Bahnsen et al., 2002) (PDB: 1EY3, Z-score 21.5), hexanoyl-CoA hydratase (Bell et al., 2002) (PDB: 1MJ3, Z-score 21.2), human AU-rich RNA-binding protein/3-hydroxy-3-methylglutaconyl-CoA hydratase (Kurimoto et al., 2001) (PDB: 1HZD, Z-score 26.7), human mitochondrial $\Delta 3$ - $\Delta 2$ -enoyl-CoA isomerase (Partanen et al., 2004) (PDB: 1SG4, Z-score 26.9), and *Saccharomyces cerevisiae* $\Delta 3$ - $\Delta 2$ -enoyl-CoA isomerase (Mursula et al., 2001) (PDB: 1HNU, Z-score 22.9). Although RpfF shares relatively low sequence identity ($<25\%$) with the members of enoyl-CoA hydratase/isomerase subfamily, superposition of the equivalent C α atoms of RpfF with the members of the subfamily gives pairwise rmsd values of 1.5–1.8 Å, suggesting that RpfF belongs to the enoyl-CoA hydratase/isomerase family.

Interestingly, the last two helices ($\alpha 9$ and $\alpha 10$) of all available enoyl-CoA hydratase structures with bound ligands do not

wrap around the spiral core domain (Figure S2) but flip upward and attach to $\alpha 8$ to form a T2 trimerization domain (Kurimoto et al., 2001). RpfF and enoyl-CoA hydratase have a common N-terminal core domain but differ in the C-terminal domain fold in that helices $\alpha 9$ and $\alpha 10$ of RpfF join $\alpha 7$ and $\alpha 8$ to wrap around the spiral core domain (Figure S2). RpfF is also similar to enoyl-CoA isomerases (PDB code: 1SG4) either in ligand bound form or in free form, in which the last four helices in the C-terminal region wrap around the spiral core domain, forming a so-called self-association fold (Figure S2) (Hubbard et al., 2005).

RpfF Resembles an enoyl-CoA Hydratase Containing Two Conserved Catalytic Glutamate Residues

The exact role of RpfF in DSF biosynthesis remains unclear. Previous structural and functional studies have demonstrated that enoyl-CoA hydratases contain two catalytic glutamates while enoyl-CoA isomerase only contains one catalytic glutamate and the other glutamate is substituted by a Leu residue (Bahnsen et al., 2002; Engel et al., 1996, 1998; Muller-Newen et al., 1995). Sequence alignment of RpfF with the enoyl-CoA hydratases and enoyl-CoA isomerase (Figure S3) indicated that, like enoyl-CoA hydratase, RpfF harbors two well-conserved glutamate residues Glu141 and Glu161, which correspond to the catalytic residues Glu144 and Glu164 in enoyl-CoA hydratase, respectively (PDB code: 2DUB) (Bahnsen et al., 2002). In contrast, only Glu161 of RpfF aligns with Glu136 of enoyl-CoA isomerase, whereas Glu141 of RpfF corresponds to Leu114, which is not required for catalysis in enoyl-CoA isomerase (Partanen et al., 2004). This result suggests that RpfF is likely an enoyl-CoA hydratase rather than an enoyl-CoA isomerase. Further support to this notion comes from structural superposition of the putative active site of RpfF with those of enoyl-CoA hydratases/isomerase. As shown in Figure 2A, the active site of enoyl-CoA hydratases is composed of $\alpha 3$, $\beta B2$ - $\alpha 2B$ loop, $\beta B3$ - $\alpha 3$ loop, $\beta B4$ - $\alpha 4B$ loop, and $\alpha 9'$ and $\alpha 10'$ from its neighboring molecule. Previous functional and structural studies showed that two highly conserved glutamate residues Glu144 from $\alpha 3$ and Glu164 from the $\beta B4$ - $\alpha 4B$ loop are critical for their hydratase activity with Glu164 acting as a proton donor while Glu144 activating a water molecule to add the hydroxyl group to substrate (Bahnsen et al., 2002; Engel et al., 1996, 1998; Muller-Newen et al., 1995). Close inspection of the catalytic site shows that residues Glu141 and Glu161 in RpfF spatially align well, respectively, with the catalytic residues Glu144 and Glu164 from the enoyl-CoA hydratases. Consistent with these findings, point mutation of Glu141 or Glu161 completely abolished DSF production, underscoring the critical roles of these two residues in DSF biosynthesis (Figure 2B). Taken together, these results suggest that RpfF likely belongs to the enoyl-CoA hydratase subfamily with two key glutamate residues in its catalytic site.

RpfF Contains a Hydrophobic Pocket which Is Probably a DSF Precursor Docking Site

In addition to these two critical catalytic residues Glu141 and Glu161, other important residues in the catalytic site are also highly conserved between RpfF and enoyl-CoA hydratases (Figure 2A; Figure S3). For example, Leu136 from $\beta B3$, Gly137 and Gly138 from $\beta B3$ - $\alpha 3$ loop, Gly85 from $\beta B2$ - $\alpha 2B$ loop, and

Leu276 from $\alpha 10$ may coordinate the substrate binding as well as catalytic activity. Residue Met170 from $\beta B4$ - $\alpha 4B$ loop and Trp258 from $\alpha 9$ are hydrophobic and may contact the long fatty acyl group of the DSF precursor molecule. All of these residues are highly conserved across different homologs of RpfF (Figure S3) and form a hydrophobic pocket (Figures 2C). This pocket is similar to the substrate binding pockets in rat enoyl-CoA hydratase (Bahnsen et al., 2002) and methylmalonyl decarboxylase (Benning et al., 2000) but too small to accommodate the 13 carbon fatty acyl chain of a DSF molecule since the dimethyl group of this molecule sterically clashes with $\alpha 9$ of RpfF. Moreover, we observed that helix $\alpha 10$ also clashes with the CoA moiety of the superimposed ligands (Figure 2C; see below). Altogether, these observations suggest that this cavity is probably a DSF precursor docking site and that RpfF may undergo a conformational change to relieve the steric hindrance for accommodating the incoming substrate.

To examine the role of the residues in the putative substrate binding pocket in RpfF, mutagenesis was performed on these residues and the resultant variant RpfF were expressed in the DSF-deficient strain $\Delta rpfF \Delta rpfC$. Consistent with the structural prediction, single point mutation in these residues resulted in either no DSF production or dramatically reduced DSF levels (Table S1), suggesting that these residues are important for DSF biosynthesis. Western blotting analysis confirmed that single point mutation did not affect RpfF protein expression level (data not shown).

The RpfF/REC Interface

The interface between RpfF and REC domain involves the C-terminal helix α -10, $\alpha 4A$, the β sheet B of RpfF and helices $\alpha 2$ and $\alpha 3$ of the REC domain (Figure 1B). The interaction between these two proteins buries a pairwise accessible surface area of 1368 Å². The interaction involves both hydrophilic and hydrophobic contacts. The C-terminal $\alpha 10$ of RpfF contacts the surface groove of $\alpha 2$ and $\alpha 3$ of the REC domain to form a helix bundle mainly through hydrogen bond and polar interactions (Figure 3A). For example, Arg278 and Arg271 from $\alpha 10$ of RpfF form hydrogen bonds with Asp499 and Glu495 of $\alpha 2$ in the REC domain, respectively. Arg275 and Thr272 from $\alpha 10$ of RpfF is hydrogen bonded to Gln526 in the $\alpha 3$ of the REC domain, which in turn makes a polar contact with Asp522 in the same α helix. The interface between $\alpha 2$ and $\alpha 3$ is composed of four hydrophobic residues Met518 and Leu498 from $\alpha 2$, and Met535 and Met530 from $\alpha 3$. On the other site, the $\alpha 3$ of REC domain contacts the β sheet B mainly through hydrophobic interaction. As shown in Figure 3A, the side chain of Val529 and the methylene group of Arg528 contact the hydrophobic surface of RpfF composed of Leu136 from loop $\beta B3$ - $\alpha 3$, Pro160 and Leu163 from $\alpha 4A$, and Leu194 from $\beta B5$. The residues involved in the interface are highly conserved across various bacterial species (Figures 3B and 3C), suggesting that the interaction of RpfF and the REC domain of RpfC is likely a conserved feature.

To verify the key structural features of RpfF that governs the RpfF-RpfC interaction, we first generated a truncated RpfF without C-terminal helix α -10. The bacterial two-hybrid assay showed that the *Escherichia coli* strain coexpressing the truncated RpfF and REC domain could not grow in the selection

Figure 2. Comparison of the Catalytic Site of RpfF and Octanoyl-CoA Hydratase

(A) Stereo view of catalytic site of RpfF (green) superimposed with enoyl-CoA hydratase (PDB code: 2DUB, orange) in the presence of its ligand octanoyl-CoA shown in stick model. Glycine residues near active site are shown in spheres. Secondary structures and residues involved in catalytic site are labeled. See Figure S2 for structural comparison of RpfF with hydratase/isomerases, and Figure S3 for sequence alignment of RpfF and its homologs with enoyl-CoA hydratase (PDB: 2DUB) and isomerase (PDB: 1SG4).

(B) Mutagenesis reveals the critical residues involved in catalytic activity. Upper panel: TLC plate was used to quantify DSF activity as indicated by the presence of a blue zone. Middle panel: the amount of DSF production in WT and mutant RpfF. Lower panel: Western blot analysis to check expression of WT and mutant RpfF. The data are means of three repeats and error bars indicate SD.

See Table S1 for more information on mutational analysis of the residues associated with the putative substrate-binding pocket of RpfF.

(C) Cavity analysis of RpfF showed that the cavity in the putative catalytic site is too small to accommodate a substrate with the same carbon chain length as DSF (13 carbon atoms). The stick model in orange color is octanoyl-CoA from PDB code: 2DUB. The stick model in slate color is pre-DSF-CoA, which is modeled using the structure of octanoyl-CoA as a template, and is obviously larger than the cavity, suggesting that a conformation change will occur when RpfF binds to substrate.

screening medium (Figure 4A), suggesting that the C-terminal helix of RpfF plays a critical role in the interaction between RpfF and REC. In addition to the C-terminal helix α -10, we have also tested the roles of residues Leu136, Pro160, Leu163, and Leu194 in the interaction. Our results showed that single point mutation in any of these residues had no effect on the interaction. However, double point mutations in Leu136 and Leu194

resulted in no binding of RpfF to REC domain (Figure 4A), which is consistent with our prediction that residues Leu136 and Leu194 are involved in the interface.

The involvement of the residues of the REC domain in binding to RpfF was also confirmed by using bacterial two hybrid assay. The *E. coli* strain coexpressing RpfF and the REC domain grow well on the selection medium; however, the *E. coli* strains

Figure 3. Interaction between RpfF and the REC Domain of RpfC

(A) Stereo view of the RpfF/REC interface. RpfF and the REC domain are shown in pink and lemon cartoon, respectively. Residues involved in the interactions are shown in stick models. Secondary elements and residues involved in the interface are labeled and indicated with “†” in (B) and (C). (B) Sequence alignment of RpfF and its homologs, *Xanthomonas oryzae* pv. *oryzae* (Xoo), *Stenotrophomonas maltophilia* (Stm), *Xylella fastidiosa* (Xyf), and *Xanthomonas axonopodis* pv. *citri* (Xac). Secondary structural elements are indicated. Mutated residues are marked with “*.” (C) Sequence alignment of the REC domain of RpfC from different bacterial species. Secondary structural elements are indicated. Mutated residues are marked with “#.”

coexpressing RpfF and the REC domain with triple mutations in Glu495, Leu498, and Asp499 or in Arg528, Val529, and Met530 failed to grow on the same selection screen medium (Figure 4A). The results of western blotting analysis precluded the possibility that the differences in growth were due to variation in protein expression level (Figure 4B).

To examine the functional role of the key residues of the REC domain in the interaction with RpfF, we created single or multiple point mutants in the REC domain and evaluated the mutational effects on DSF production. As shown in Figure 4C, deletion of

rpfC in *Xcc* resulted in overproduction of DSF and overexpression of REC domain in the mutant Δ rpfC decreased the DSF biosynthesis by more than 97%. This is consistent with the notion that the binding between REC and RpfF blocks DSF biosynthesis. While a single point mutation in Leu498, Asp499, Arg528, or Met530 of REC did not affect DSF biosynthesis, substitution of Glu495 and Val529 with alanine substantially attenuated the inhibitory activity of REC on DSF production (Figure 4C). Moreover, double point mutations in Glu495 and Leu498, and triple point mutations in Glu495, Leu498, and

Figure 4. Mutational Analysis of the Interface between RpfF and the REC Domain

(A) Bacterial two-hybrid assay to confirm the roles of the predicted residues in the binding.

(B) Western blotting analysis to show the expression of variant REC or RpfF.

(C) DSF biosynthesis assay to verify the roles of the residues of REC domain in interaction with RpfF. The data are means of three repeats and error bars indicate SD.

(D) Western blotting to show that point mutation did not affect the expression of the REC domain.

Asp499 further compromised the REC inhibitory activity by up to 62% and 80%, respectively (Figure 4C). Similarly, the REC domain with double point mutations in Arg528 and Val529, triple point mutations in Arg528, Val529, and Met530 also resulted in loss of its DSF biosynthesis inhibitory activity by about 56% and 75%, respectively (Figure 4C). Western blotting analysis showed that the point mutations within REC did not affect the expression level of this domain (Figure 4D). Taken together, these data indicate that these residues are involved in the interface and work in a concerted way to mediate the interaction between RpfF and the REC domain.

The REC Domain Appears to Lock RpfF in an Inactive State and Physically Blocks the Putative Substrate Binding Pocket

Based on the structure of RpfF apo form or in complex with RpfC REC domain, cavity analysis reveals that there is a closed cavity located in the catalytic site, which is too small to accommodate a CoA substrate with the length of DSF carbon chain (13 carbon atoms; Figure 2C) compared with the ligand octanoyl-CoA (PDB: 2DUB), suggesting that a conformation change may occur when the substrate binds to RpfF. Superimposition of RpfF in REC-bound form with those of hydratases and isomerases with bound

Figure 5. Possible Conformational Changes of RpfF and the Inhibitory Effect of the REC Domain on Substrate Binding

The stereo figure depicts the conformational change of helices of RpfF that are involved in catalytic site. Cylindrical cartoons of helix α B2, α A2, α 9, and α 10 of RpfF and other hydratases/isomerase (PDB code: 2DUB and 1SG4) are shown in green, magenta and orange, respectively. Ligands are shown in stick model with colors corresponding to their respective protein colors. The rest of RpfF is shown in gray solid surface while helix α 3 of the REC domain is shown in red.

ligands reveals that conformational change occurs mainly in four helices: α 9, α 10, α 2B, and α 2A, which were highlighted in cylindrical cartoon in Figure 5. Helices α 9 and α 2A coordinate with the carbon chain of substrates, and different length of carbon chain may induce diversified conformational changes in these two helices. Thus, the binding of ligands with longer carbon chain would result in wide split between helices α 2A and α 9. Furthermore, α 10 of RpfF in both apo- and REC-bound forms would clash with the CoA moiety if RpfF binds to a substrate. Such a steric hindrance would force α 10 to rotate about 30° in the presence of a CoA-conjugated ligand (Figure 5). Besides α 10 of RpfF, helix α 3 of the REC domain also clashes with the CoA moiety of a substrate, suggesting that the REC domain itself would directly interfere the substrate binding to RpfF. Since the binding of RpfF to the REC domain is partially mediated by the interaction between α 10 of RpfF and α 3 of the REC domain, and α 10 adopts an essentially same conformation in both apo- and REC-bound forms of RpfF, the binding of the REC domain may lock the position of α 10 and force RpfF to stay an inactive form, thereby blocking substrate binding to the catalytic site.

DISCUSSION

Evidence is accumulating that the DSF-dependent QS system of *Xcc* represents another family of widely conserved bacterial QS systems implicated in regulation of multiple biological functions. This system differs from other known QS systems in various aspects, in particular, the autoregulation mechanism that controls signal production (He and Zhang, 2008). Characterization of the protein structure of DSF synthase holds the key to understand how its catalytic activity can be modulated by its ligand protein RpfC. In this study, the crystal structure of the full-length RpfF, which is a key DSF synthase (Barber et al., 1997; Wang et al., 2004), was determined at a resolution of 1.8 Å. Structural comparison showed that RpfF has similar fold to the members of the enoyl-CoA hydratase/isomerase superfamily.

The common reaction mechanism of enoyl-CoA hydratases involves two catalytic glutamate residues (Hamed et al., 2008). We compared RpfF with the enoyl-CoA hydratase (PDB code: 2DUB), which is a well-characterized enzyme (Agnihotri and Liu, 2003; Bahnson et al., 2002). RpfF and 2DUB share two conserved catalytic glutamate residues (Glu141 and Glu161 in

RpfF), whose catalytic role in synthesis of DSF were verified by mutagenesis analysis. The C-terminal helices α 9 and α 10 of 2DUB do not form the self-association fold but interact with neighboring molecule and form an intratrimer association fold (Hubbard et al., 2005). In contrast, superposition of RpfF with the ligand-bound structures of 1EY3 reveals that the helices α 9 and α 10 of RpfF might act as an “arm” that can be flipped upward or away to allow ligand binding, or flipped downward to inhibit ligand binding. Presumably, RpfF may share a similar interfacial activation mechanism as the lipolytic enzymes, whose activation needs substrate aggregation or micelles to displace the helical “lid” that covers the catalytic site (Mingarro et al., 1995). These findings present useful clues for further characterization of the RpfF catalytic mechanisms and the precursor for DSF biosynthesis.

The most critical feature of a QS system is the mechanism that enables bacterial cells to autoregulate the production of QS signals. Our previous data show that *Xcc* has evolved a novel autoregulation mechanism involving the protein-protein interaction between RpfF and RpfC (He et al., 2006a). However, the mechanism by which RpfC modulates the RpfF enzyme activity and hence controls DSF production remains elusive. Structural comparison of the RpfF apo form and the RpfF-REC complex showed that RpfF appears to adopt the same “inactive” form in the absence or in the presence of the REC domain. In this inactive state, the C-terminal helix α 10 of RpfF is orientated downward that blocks the entrance of the substrate-binding pocket as discussed above, and the helices α 2 and α 3 from the REC domain form a helix bundle with the α -10 of RpfF to lock it in such a position that prevents interfacial activation. These structural findings not only provide plausible mechanistic explanation for how DSF autoinduction is mediated by the RpfF/RpfC interaction but also present further refinement for our previous proposed model of DSF autoinduction (He et al., 2006a) (Figure 6). At low cell density, RpfF is associated with the REC domain of RpfC, which might keep the DSF synthase in the inactive state and maintains the DSF production at a basal level (Figure 6A). When cell density reaches a threshold level, the diffusible DSF signals accumulate in extracellular environment and may interact with RpfC. The event presumably leads to phosphorylation of the REC domain and consequent release of RpfF, which catalyzes substantial DSF biosynthesis upon interfacial induction (Figure 6B).

Figure 6. Mechanism of DSF Autoinduction

(A) In the case of low cell density, concentration of DSF is also very low, most of RpfF are sequestered by the RpfC REC domain and adopt an inactive form. A small group of RpfF is free from RpfC and is capable of DSF biosynthesis. (B) When cell grows and DSF accumulates to a certain threshold, DSF diffuses out of cell and binds to extracellular domain of RpfC, which triggers phosphorylation relay. Once the REC domain is phosphorylated, the interaction between RpfF and RpfC REC domain will be disrupted transiently. Thus, RpfF is released and able to bind to substrate and synthesize DSF. As the binding of substrate and the REC domain to RpfF seem to be mutually exclusive, RpfC can no longer sequester RpfF, which results in more DSF biosynthesis and downstream virulence regulation.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Growth Conditions

Xcc strain XC1, and its derivatives FE 58 and Δ rpfC, has been described previously (He et al., 2006a, 2006b; Wang et al., 2004). The double deletion mutant Δ rpfF Δ rpfC was generated using Δ rpfC as the parental strain following the method described previously (He et al., 2006a). Xcc strains were grown at 30°C in LB medium unless otherwise stated. *Escherichia coli* strains were maintained at 37°C in LB medium. Antibiotics were added at the following concentrations when required: kanamycin, 100 μ g/ml; ampicillin, 200 μ g/ml; rifampicin, 50 μ g/ml; tetracycline, 10 μ g/ml; chloramphenicol, 10 μ g/ml. X-gluc (5-bromo-4-chloro-3-indolyl β -D-glucopyranoside) was included in NYG medium at a final concentration of 30 μ g/ml for detection of GUS (β -glucuronidase) activity.

Protein Expression and Purification

Full-length *rpfF* from Xcc was cloned into the vector pETDuet-1 with a N-terminal 6xHis tag. RpfF was expressed in *E. coli* strain BL21 (DE3) grown at 37°C in LB medium containing ampicillin. At OD₆₀₀ = 0.6, cells were induced with 0.4 mM isopropylthio- β -galactoside (IPTG) and grown at 18°C for an additional 16 hr prior to harvest. Cells were lysed in 25 mM Tris buffer (pH 8.0) with 500 mM NaCl and lysozyme. Cell debris was removed by centrifugation at 18,000 rpm at 4°C. The supernatant containing His-tagged RpfF was incubated with TALON resin (BD Biosciences) pre-equilibrated with the lysis buffer (25 mM Tris buffer [pH 8.0] and 500 mM NaCl). The column was washed with 15 mM imidazole, and protein was eluted with 100 mM imidazole. The protein was further purified by Superdex-200 gel filtration column (Amersham Biosciences). SeMet-substituted RpfF was expressed in a minimal medium containing 20 mg/l seleno-L-methionine (SeMet) and purified in the same way as the native protein.

To get the RpfF/REC complex, full-length RpfF and the REC domain (residues 449–590) of RpfC were cloned into the vector pETDuet-1 of multiple cloning sites-1 and -2, respectively, with a N-terminal 6xHis-Tag fused to RpfF. The complex was purified using TALON resin and Superdex 200 gel filtration columns. The eluted protein complexes were concentrated to ~15 mg/ml in a buffer containing 50 mM Tris-HCl (pH 8.0) and 100 mM NaCl for crystallization.

Crystallization and Data Collection

For crystallization, RpfF was concentrated to 8 mg/ml in 25 mM Tris (pH 8.0), 500 mM NaCl, and 10 mM DTT. Crystallization screening was performed using the sitting-drop vapor-diffusion method at 15°C by mixing 200 nl protein solution with 200 nl reagent solution in 96-well plates using Phoenix liquid handling

robot. Rod-shaped crystals appear in 1 day from a reservoir solution of 0.2 M potassium thiocyanate and 20% (w/v) PEG 3350. This initial condition was then optimized to 0.2 M potassium thiocyanate and 9% (w/v) PEG 3350. Crystals suitable for data collection were grown in 48 hr by mixing 2 μ l protein solution with 2 μ l crystallization buffer, via the hanging-drop vapor-diffusion method. Crystals of SeMet-substituted RpfF were cryoprotected by 25% PEG 400 before flash-frozen in liquid nitrogen.

Crystals belong to P2₁2₁2₁ with cell parameters $a = 96.8$ Å, $b = 112.3$ Å, $c = 119.6$ Å, $\alpha = \beta = \gamma = 90^\circ$ and contains three molecules per asymmetric unit. SAD data sets were collected at 100 K from a single SeMet-labeled crystal using synchrotron radiation at beamline ID23-1 at the European Synchrotron Facility (ESRF), Grenoble, France.

Crystals of the RpfF/REC complex were grown at 15°C by hanging drop vapor diffusion. Equal volume of protein sample was mixed with the crystallization solution (100 mM MES-Na [pH 6.5], 20%–22% PEG4000, 200 mM MgCl₂). Single crystals were transferred to the crystallization buffer with 20% ethylene glycol. Before being fast frozen in liquid nitrogen, crystals were quick soaked for less than 10 s in the cryo-buffer with 300 mM KI. Crystals belong to the space group P6₅ with cell parameters $a = b = 130.9$ Å, $c = 156.5$ Å, $\alpha = \beta = 90^\circ$, and $\gamma = 120^\circ$ and contain three complexes per asymmetric unit. X-ray data were collected at ESRF, ID29 and processed with MOSFLM (Leslie, 1992).

Structural Determination

The structure of RpfF was determined by the SAD method. The peak data were integrated using MOSFLM (Leslie, 1992) and processed with the CCP4i package (Potterton et al., 2003). Heavy atom search was carried out by SHELXC/D/E (Sheldrick, 2008). Model was built automatically by ARP/wARP (Morris et al., 2002). Crystallographic refinement was performed with the programs REFMAC5 (Murshudov et al., 1997) to a final R_{free} of 24.7%. The final refinement statistics for RpfF are summarized in Table 1.

The structure of RpfF/REC was determined by the molecular replacement method using program PHASER (McCoy et al., 2007) with the structure of RpfF as a search model. Three copies of RpfF in the AU were found sequentially with Z-scores 24.2, 46.4, and 58.5, respectively. Further molecular replacement trials searching for the REC domains of RpfC failed, but it was clearly shown in the difference Fourier maps. The partial mode of the REC domain was manually built using COOT (Emsley and Cowtan, 2004). Crystallographic refinement was carried out with REFMAC5 (Murshudov et al., 1997) with TLS and NCS restraints. The final refinement statistics for the complex are summarized in Table 1.

Site-Directed Mutagenesis to Identify Key Residues Associated with DSF Biosynthesis and RpfF-RpfC Interaction

The coding sequences of *rpfF* and the REC domain of RpfC were amplified by PCR and cloned into the vector pGEM-T-easy (Promega), respectively. Point mutations were conducted using the QuickChange site-directed mutagenesis

kit following the manufacturer's instructions (Stratagene). The mutations were verified by DNA sequencing, and the primers used in PCR and mutagenesis are listed in Table S2. For identification of key residues associated with DSF biosynthesis, *rpfF* and its variants were cloned in the expression vector pLAFR3 and mobilized into double deletion mutant Δ rpfF Δ rpfC for DSF production analysis using the method described below. Similarly, the coding sequence of REC domain and its variants were cloned in the same vector and introduced into the *rpfC* deletion mutant Δ rpfC to verify the putative residues implicated in binding of RpfF. To identify the key residues of RpfF involved in binding to the REC domain, the construct pLAFR3 containing the coding region of REC domain and the construct pDSK519 containing RpfF or its variants were mobilized into the double deletion strain Δ rpfF Δ rpfC for analysis of DSF production as described below.

Quantification of DSF Production

Strain XC1 and its derivatives were grown in liquid LB medium until OD₆₀₀ reaching about 2.1. For each strain, the supernatants from 50 ml of cell cultures were collected by centrifugation at 10,000 rpm for 30 min. The supernatants were extracted twice by using equal volume of ethyl acetate. The organic phase was evaporated and the residues containing DSF were dissolved in 50 μ l of methanol. For DSF analysis, 5 μ l of crude extracts was spotted on TLC silica gel plate 60F₂₅₄ (Merck), which was separated in a tank containing ethyl acetate and Hexane (2:8, v:v) as eluting solvents. The TLC plate was air-dried and overlaid with 50 ml of NYG liquid medium supplemented with 0.8% agrose, 30 μ g/ml of X-gluc, and 1.5 ml of fresh DSF reporter strain FE58 (Wang et al., 2004). The TLC plate was incubated at 28°C in darkness overnight. DSF activity was indicated by the presence of a blue zone. For quantitative comparison, DSF concentration was calculated using the formula: DSF (μ M) = $0.0099e^{2.2527w}$, based on the width (w) of blue zone in centimeter. The formula was derived from a dose-response plot of the biosensor using various dilutions of synthetic DSF signal, with a correlation coefficient (R^2) of 0.9657.

BacterioMatch II Two-Hybrid System to Evaluate RpfF and REC Binding

The RpfF and REC interaction was also detected using BacterioMatch II Two-hybrid system kit (Stratagene) following the manufacturer's instruction. In brief, *rpfF* and its variants were amplified using the primers listed in Table S2 and were fused separately with the gene λ cl encoding the full-length bacteriophage λ repressor protein (237 amino acids), containing the N-terminal DNA-binding domain and the C-terminal dimerization domain in the vector pBT. The coding sequences of REC domain and its derivatives were fused separately to the N-terminal domain of the α subunit of RNA polymerase (248 amino acids) in the vector pTRG (Stratagene). The resultant constructs were cotransformed into the XL1-Blue RF' Kan strain. At the same time, the pTRG-Gal11P and the pBT-LGF2 positive control plasmids included in the kit were also cotransformed. The Nonselective Screening Medium (without 3-AT) was used for screening the *E. coli* colonies containing the cotransformed constructs. The growth of the selected strains was further verified on the Selection Screening Medium (5 mM 3-AT) by stripping. Normal growth on the Selection Screening Medium indicates a strong binding between two proteins. Western blotting analysis was used to confirm the expression of RpfF or REC.

ACCESSION NUMBERS

The coordinates and structure-factor amplitudes of RpfF and the RpfF/REC complex have been deposited in the Protein Data Bank with accession codes 3M6N and 3M6M, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and two tables and can be found with this article online at doi:10.1016/j.str.2010.06.011.

ACKNOWLEDGMENTS

We thank the beamline scientists at ID-29 and ID23-1 (ESRF, France) for assistance and access to synchrotron radiation facilities. This work is financially

supported by the Biomedical Research Council of A*STAR (Agency for Science, Technology and Research, Singapore) (H.S. and L.H.-Z.).

Received: April 13, 2010

Revised: June 17, 2010

Accepted: June 20, 2010

Published: September 7, 2010

REFERENCES

- Agnihotri, G., and Liu, H.W. (2003). Enoyl-CoA hydratase: reaction, mechanism, and inhibition. *Bioorg. Med. Chem.* 11, 9–20.
- Bahnsen, B.J., Anderson, V.E., and Petsko, G.A. (2002). Structural mechanism of enoyl-CoA hydratase: three atoms from a single water are added in either an E1c stepwise or concerted fashion. *Biochemistry* 41, 2621–2629.
- Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R.P., and Dickerson, R.E. (1996). Structure of the Escherichia coli response regulator NarL. *Biochemistry* 35, 11053–11061.
- Barber, C.E., Tang, J.L., Feng, J.X., Pan, M.Q., Wilson, T.J., Slater, H., Dow, J.M., Williams, P., and Daniels, M.J. (1997). A novel regulatory system required for pathogenicity of Xanthomonas campestris is mediated by a small diffusible signal molecule. *Mol. Microbiol.* 24, 555–566.
- Bassler, B.L., and Losick, R. (2006). Bacterially speaking. *Cell* 125, 237–246.
- Bell, A.F., Feng, Y., Hofstein, H.A., Parikh, S., Wu, J., Rudolph, M.J., Kisker, C., Whitty, A., and Tonge, P.J. (2002). Stereoselectivity of enoyl-CoA hydratase results from preferential activation of one of two bound substrate conformers. *Chem. Biol.* 9, 1247–1255.
- Benning, M.M., Haller, T., Gerit, J.A., and Holden, H.M. (2000). New reactions in the crotonase superfamily: structure of methylmalonyl CoA decarboxylase from Escherichia coli. *Biochemistry* 39, 4630–4639.
- Boon, C., Deng, Y., Wang, L.H., He, Y., Xu, J.L., Fan, Y., Pan, S.Q., and Zhang, L.H. (2008). A novel DSF-like signal from Burkholderia cenocepacia interferes with Candida albicans morphological transition. *ISME J.* 2, 27–36.
- Casino, P., Rubio, V., and Marina, A. (2009). Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell* 139, 325–336.
- Colnaghi Simionato, A.V., da Silva, D.S., Lambais, M.R., and Carrilho, E. (2007). Characterization of a putative Xylella fastidiosa diffusible signal factor by HRGC-EI-MS. *J. Mass Spectrom.* 42, 490–496.
- Djordjevic, S., and Stock, A.M. (1998). Structural analysis of bacterial chemotaxis proteins: components of a dynamic signaling system. *J. Struct. Biol.* 124, 189–200.
- Dong, Y.H., Wang, L.Y., and Zhang, L.H. (2007). Quorum-quenching microbial infections: mechanisms and implications. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 1201–1211.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Engel, C.K., Mathieu, M., Zeelen, J.P., Hiltunen, J.K., and Wierenga, R.K. (1996). Crystal structure of enoyl-coenzyme A (CoA) hydratase at 2.5 angstroms resolution: a spiral fold defines the CoA-binding pocket. *EMBO J.* 15, 5135–5145.
- Engel, C.K., Kiema, T.R., Hiltunen, J.K., and Wierenga, R.K. (1998). The crystal structure of enoyl-CoA hydratase complexed with octanoyl-CoA reveals the structural adaptations required for binding of a long chain fatty acid-CoA molecule. *J. Mol. Biol.* 275, 847–859.
- Fehér, V.A., Zapf, J.W., Hoch, J.A., Whiteley, J.M., McIntosh, L.P., Rance, M., Skelton, N.J., Dahlquist, F.W., and Cavanagh, J. (1997). High-resolution NMR structure and backbone dynamics of the Bacillus subtilis response regulator, Spo0F: implications for phosphorylation and molecular recognition. *Biochemistry* 36, 10015–10025.
- Fouhy, Y., Scanlon, K., Schouest, K., Spillane, C., Crossman, L., Avison, M.B., Ryan, R.P., and Dow, J.M. (2007). Diffusible signal factor-dependent cell-cell signaling and virulence in the nosocomial pathogen Stenotrophomonas maltophilia. *J. Bacteriol.* 189, 4964–4968.

- Hamed, R.B., Batchelar, E.T., Clifton, I.J., and Schofield, C.J. (2008). Mechanisms and structures of crotonase superfamily enzymes—how nature controls enolate and oxyanion reactivity. *Cell. Mol. Life Sci.* **65**, 2507–2527.
- He, Y.W., and Zhang, L.H. (2008). Quorum sensing and virulence regulation in *Xanthomonas campestris*. *FEMS Microbiol. Rev.* **32**, 842–857.
- He, Y.W., Wang, C., Zhou, L., Song, H., Dow, J.M., and Zhang, L.H. (2006a). Dual signaling functions of the hybrid sensor kinase RpfC of *Xanthomonas campestris* involve either phosphorelay or receiver domain-protein interaction. *J. Biol. Chem.* **281**, 33414–33421.
- He, Y.W., Xu, M., Lin, K., Ng, Y.J., Wen, C.M., Wang, L.H., Liu, Z.D., Zhang, H.B., Dong, Y.H., Dow, J.M., et al. (2006b). Genome scale analysis of diffusible signal factor regulon in *Xanthomonas campestris* pv. *campestris*: identification of novel cell-cell communication-dependent genes and functions. *Mol. Microbiol.* **59**, 610–622.
- Huang, T.P., and Wong, A.C. (2007). A cyclic AMP receptor protein-regulated cell-cell communication system mediates expression of a FecA homologue in *Stenotrophomonas maltophilia*. *Appl. Environ. Microbiol.* **73**, 5034–5040.
- Hubbard, P.A., Yu, W., Schulz, H., and Kim, J.J. (2005). Domain swapping in the low-similarity isomerase/hydratase superfamily: the crystal structure of rat mitochondrial Delta3, Delta2-enoyl-CoA isomerase. *Protein Sci.* **14**, 1545–1555.
- Kern, D., Volkman, B.F., Luginbuhl, P., Nohaile, M.J., Kustu, S., and Wemmer, D.E. (1999). Structure of a transiently phosphorylated switch in bacterial signal transduction. *Nature* **402**, 894–898.
- Kurimoto, K., Fukai, S., Nureki, O., Muto, Y., and Yokoyama, S. (2001). Crystal structure of human AUH protein, a single-stranded RNA binding homolog of enoyl-CoA hydratase. *Structure* **9**, 1253–1263.
- Leslie, A.G.W. (1992). Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography No. 26, 27–33.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
- Milton, D.L. (2006). Quorum sensing in vibrios: complexity for diversification. *Int. J. Med. Microbiol.* **296**, 61–71.
- Mingarro, I., Abad, C., and Braco, L. (1995). Interfacial activation-based molecular bioimprinting of lipolytic enzymes. *Proc. Natl. Acad. Sci. USA* **92**, 3308–3312.
- Morris, R.J., Perrakis, A., and Lamzin, V.S. (2002). ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 968–975.
- Muller-Newen, G., Janssen, U., and Stoffel, W. (1995). Enoyl-CoA hydratase and isomerase form a superfamily with a common active-site glutamate residue. *Eur. J. Biochem.* **228**, 68–73.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255.
- Mursula, A.M., van Aalten, D.M., Hiltunen, J.K., and Wierenga, R.K. (2001). The crystal structure of delta(3)-delta(2)-enoyl-CoA isomerase. *J. Mol. Biol.* **309**, 845–853.
- Partanen, S.T., Novikov, D.K., Popov, A.N., Mursula, A.M., Hiltunen, J.K., and Wierenga, R.K. (2004). The 1.3 Å crystal structure of human mitochondrial Delta3-Delta2-enoyl-CoA isomerase shows a novel mode of binding for the fatty acyl group. *J. Mol. Biol.* **342**, 1197–1208.
- Potterton, E., Briggs, P., Turkmen, M., and Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131–1137.
- Sheldrick, G.M. (2008). A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122.
- Slater, H., Alvarez-Morales, A., Barber, C.E., Daniels, M.J., and Dow, J.M. (2000). A two-component system involving an HD-GYP domain protein links cell-cell signalling to pathogenicity gene expression in *Xanthomonas campestris*. *Mol. Microbiol.* **38**, 986–1003.
- Sola, M., Gomis-Ruth, F.X., Serrano, L., Gonzalez, A., and Coll, M. (1999). Three-dimensional crystal structure of the transcription factor PhoB receiver domain. *J. Mol. Biol.* **285**, 675–687.
- Stock, A.M., Mottonen, J.M., Stock, J.B., and Schutt, C.E. (1989). Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis. *Nature* **337**, 745–749.
- Volkman, B.F., Nohaile, M.J., Amy, N.K., Kustu, S., and Wemmer, D.E. (1995). Three-dimensional solution structure of the N-terminal receiver domain of NTRC. *Biochemistry* **34**, 1413–1424.
- Von Bodman, S.B., Bauer, W.D., and Coplin, D.L. (2003). Quorum sensing in plant-pathogenic bacteria. *Annu. Rev. Phytopathol.* **41**, 455–482.
- Wang, L.H., He, Y., Gao, Y., Wu, J.E., Dong, Y.H., He, C., Wang, S.X., Weng, L.X., Xu, J.L., Tay, L., et al. (2004). A bacterial cell-cell communication signal with cross-kingdom structural analogues. *Mol. Microbiol.* **51**, 903–912.
- Whitehead, N.A., Barnard, A.M., Slater, H., Simpson, N.J., and Salmond, G.P. (2001). Quorum-sensing in Gram-negative bacteria. *FEMS Microbiol. Rev.* **25**, 365–404.
- Zhang, L.H., and Dong, Y.H. (2004). Quorum sensing and signal interference: diverse implications. *Mol. Microbiol.* **53**, 1563–1571.

Supplemental Information

Structural Basis of the Sensor-Synthase

Interaction in Autoinduction of the Quorum

Sensing Signal DSF Biosynthesis

Zhihong Cheng, Ya-Wen He, Siew Choo Lim, Rohini Qamra, Martin A. Walsh, Lian-Hui Zhang, and Haiwei Song

Figure S1, related to Figure 1.

Stereo view of a representative portion of the Fo-Fc electron density map (contoured at 1.5σ) covering the region around $\beta1/\beta2/\beta3/\beta4$ and $\alpha3$ of the RpfC REC domain in the RpfF/REC complex. The map was calculated at a resolution of $2.5\text{\AA}$ from the initial molecular replacement phases contributed by the RpfF molecules only.

Figure S2, related to Figure 2. Structural comparisons of RpfF with hydratase/isomerases

(A) Stereo view of structure of RpfF (green) superimposed with octanoyl-CoA hydratase (PDB code:2DUB, orange).

(B) Stereo view of structure of RpfF (green) superimposed with Δ^3 - Δ^2 -enoyl-CoA isomerase (PDB code:1SG4, magenta).

Figure S3, related to Figure 2.

Sequence alignment of RpfF and its homologs with enoyl-CoA hydratase (PDB:2DUB) and isomerase (PDB:1SG4). Secondary structural elements of RpfF are indicated on the top. Key catalytic residues are marked with "▲". Residues located in the hydrophobic pocket are marked with "•". Identical residues are indicated with a red box and white character. Similar residues are in red character with blue frame.

Table S1, related to Figure 2. Site-directed mutagenesis of the residues associated with the putative substrate-binding pocket of RpfF

Wild type RpfF and its variants	DSF biosynthesis (μM) *
wild type	27.4 ± 2.5
G85L	0
G85K	0
G86L	0
G86K	0
L136A	0
G137A	6.4 ± 1.3
G138A	0
E141A	0
P160A	5.2 ± 1.2
P160K	7.6 ± 1.7
E161A	0
M170A	0
W258A	0
L276A	4.7 ± 1.3

* DSF was quantified after TLC plate assay as described in Materials and methods. The assay was repeated twice.

Table S2, related to experimental procedures. Oligo primers used in this study

Application		Oligos and sequences (5' to 3')
Protein expression	RpIF	pDUET-RpIF-FOR: GGAATTC ATG TCT GCA GTT CAA CCC TTC ATT C
		pDUET-RpIF-REV: CCCAAGCTT TCA GCC CGC GTC GAG CCC TGA G
	REC domain	pDUET-REC-FOR: GGAATTCATATG ATG TCC AAC CCA TTC CTG CGT CAT CG
		pDUET-REC-REV: CCGCTCGAG CTA CAG CAC GCT GGA ATC GAG CAC
Overexpression	rpIF	RpIF-FOR: CGGGATCC ATG TCT GCA GTT CAA CCC TTC ATT C
		RpIF-REV: CCCAAGCTT TCA GCC CGC GTC GAG CCC TGA G
	REC domain	REC-FOR: CGGGATCC ATG TCC AAC CCA TTC CTG CGT CAT CG
		REC-REV: CCCAAGCTT CTA CAG CAC GCT GGA ATC GAG CAC
Site-directed mutagenesis	G85L	85L-FOR: GTG TTC AAT CTG CTC GGT GAT CTG GCC CTG
		85L-REV: CAG GGC CAG ATC ACC GAG CAG ATT GAA CAC
	G85K	85K-FOR: GTG TTC AAT CTG AAG GGT GAT CTG GCC CTG
		85K-REV: CAG GGC CAG ATC ACC CTT CAG ATT GAA CAC
	G86L	86L-FOR: TTC AAT CTG GGC CTT GAT CTG GCC CTG
		86L-REV: CAG GGC CAG ATC AAG GCC CAG ATT GAA
	G86K	86K-FOR: TTC AAT CTG GGC AAA GAT CTG GCC CTG
		86K-REV: CAG GGC CAG ATC TTT GCC CAG ATT GAA
	L136A	136A-FOR: CAG GGC AAT GCG GCT GGC GGC GGG TTC
		136A-REV: GAA CCC GCC GCC AGC CGC ATT GCC CTG
	G137A	137A-FOR: GGC AAT GCG CTT GCC GGC GGG TTC GAA
		137A-REV: TTC GAA CCC GCC GGC AAG CGC ATT GCC
	G138A	138A-FOR: AAT GCG CTT GGC GCC GGG TTC GAA GCG
		138A-REV: CGC TTC GAA CCC GGC GCC AAG CGC ATT
	E141A	141A-FOR: GGC GGC GGG TTC GCA GCG GCA CTA AGC
		141A-REV: GCT TAG TGC CGC TGC GAA CCC GCC GCC
	P160A	160A-FOR: ATG ATG GGG CTG GCC GAA GTG CTG TTC
		160A-REV: GAA CAG CAC TTC GGC CAG CCC CAT CAT
	P160K	160K-FOR: ATG ATG GGG CTG AAG GAA GTG CTG TTC
		160K-REV: GAA CAG CAC TTC CTT CAG CCC CAT CAT
	E161A	161A-FOR: ATG GGG CTG CCC GCA GTG CTG TTC GAC
		161A-REV: GTC GAA CAG CAC TGC GGG CAG CCC CAT
	M170A	170A-FOR: CTA TTT CCG GGG GCG GGC GCC TAC TCC
		170A-REV: GGA GTA GGC GCC CGC CCC CGG AAA TAG
	W258A	258A-FOR: CAT CAC CGA AAT GCG GGT AGA TAC CGC
		258A-REV: GCG GTA TCT ACC CGC ATT TCG GTG ATG
	L276A	276A-FOR: ACC ATG GAC CGC GCG GTG CGC GCG CAG
		276A-REV: CTG CGC GCG CAC CGC GCG GTC CAT GGT
	REC-E495A	495A-FOR: GTC AAC GGC GCC GCG CAG GTG CTC GAT
		495A-REV: ATC GAG CAC CTG CGC GGC GCC GTT GAC
	REC-L498A	498A-FOR: GCC GAG CAG GTG GCC GAT GCG ATG GCC
		498A-REV: GGC CAT CGC ATC GGC CAC CTG CTC GGC
	REC-D499A	499A-FOR: GAG CAG GTG CTC GCT GCG ATG GCC GAA
		499A-REV: TTC GGC CAT CGC AGC GAG CAC CTG CTC
	REC-R528A	528A-FOR: CTC AAG CAA TTG GCG GTG ATG CAG GCC
		528A-REV: GGC CTG CAT CAC CGC CAA TTG CTT GAG
	REC-V529A	529A-FOR: AAG CAA TTG CGG GCG ATG CAG GCC AGT
		529A-REV: ACT GGC CTG CAT CGC CCG CAA TTG CTT
	REC-M530A	530A-FOR: CAA TTG CGG GTG GCG CAG GCC AGT GGC
		530A-REV: GCC ACT GGC CTG CGC CAC CCG CAA TTG
Bacterial two-hybrid analysis	RpIF	pBT-RpIF-FOR: GGAATTC ATG TCT GCA GTT CAA CCC TTC
		pBT-RpIF-REV: CCG CTCGAG TCA GCC CGC GTC GAG CCC TGA
	REC domain	pTRG-REC-FOR: CGGGATCC CGG TTG CTA GAG AAG GCT GG
		pTRG-REC-REV: GGAATTC CCG GGT GCT CAC TGC CAG ATC

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., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* *66*, 213-221.
- Agnihotri, G., and Liu, H.W. (2003). Enoyl-CoA hydratase. reaction, mechanism, and inhibition. *Bioorg Med Chem.* *11*, 9-20.
- Anantharaman, V., Koonin, E.V., and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* *30*, 1427-1464.
- Anheim, M., Monga, B., Fleury, M., Charles, P., Barbot, C., Salih, M., Delaunoy, J.P., Fritsch, M., Arning, L., Synofzik, M., *et al.* (2009). Ataxia with oculomotor apraxia type 2: clinical, biological and genotype/phenotype correlation study of a cohort of 90 patients. *Brain.* *132*, 2688-2698.
- Appleton, R.E., Hubner, C., Grohmann, K., and Varon, R. (2004). 'Congenital peripheral neuropathy presenting as apnoea and respiratory insufficiency: spinal muscular atrophy with respiratory distress type 1 (SMARD1)'. *Dev Med Child Neurol.* *46*, 576.
- Bahnson, B.J., Anderson, V.E., and Petsko, G.A. (2002). Structural mechanism of enoyl-CoA hydratase: three atoms from a single water are added in either an E1cb stepwise or concerted fashion. *Biochemistry.* *41*, 2621-2629.
- Barber, C.E., Tang, J.L., Feng, J.X., Pan, M.Q., Wilson, T.J., Slater, H., Dow, J.M., Williams, P., and Daniels, M.J. (1997). A novel regulatory system required for pathogenicity of *Xanthomonas campestris* is mediated by a small diffusible signal molecule. *Mol Microbiol.* *24*, 555-566.
- Bell, A.F., Feng, Y., Hofstein, H.A., Parikh, S., Wu, J., Rudolph, M.J., Kisker, C., Whitty, A., and Tonge, P.J. (2002). Stereoselectivity of enoyl-CoA hydratase results from preferential activation of one of two bound substrate conformers. *Chem Biol.* *9*, 1247-1255.
- Benning, M.M., Haller, T., Gerlt, J.A., and Holden, H.M. (2000). New reactions in the crotonase superfamily: structure of methylmalonyl CoA decarboxylase from *Escherichia coli*. *Biochemistry.* *39*, 4630-4639.
- Benning, M.M., Taylor, K.L., Liu, R.Q., Yang, G., Xiang, H., Wesenberg, G., Dunaway-Mariano, D., and Holden, H.M. (1996). Structure of 4-chlorobenzoyl coenzyme A dehalogenase determined to 1.8 Å resolution: an enzyme catalyst generated via adaptive mutation. *Biochemistry.* *35*, 8103-8109.
- Bhattacharya, A., Czaplinski, K., Trifillis, P., He, F., Jacobson, A., and Peltz, S.W. (2000). Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA.* *6*, 1226-1235.

- Biou, V., Shu, F., and Ramakrishnan, V. (1995). X-ray crystallography shows that translational initiation factor IF3 consists of two compact alpha/beta domains linked by an alpha-helix. *EMBO J.* *14*, 4056-4064.
- Buttner, K., Nehring, S., and Hopfner, K.P. (2007). Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat Struct Mol Biol.* *14*, 647-652.
- Caruthers, J.M., and McKay, D.B. (2002). Helicase structure and mechanism. *Curr Opin Struct Biol.* *12*, 123-133.
- Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., and Conti, E. (2011). Molecular Mechanisms for the RNA-Dependent ATPase Activity of Upf1 and Its Regulation by Upf2. *Mol Cell.* *41*, 693-703.
- Chamieh, H., Ballut, L., Bonneau, F., and Le Hir, H. (2008). NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat Struct Mol Biol.* *15*, 85-93.
- Chen, Y.Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A., *et al.* (2004). DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am J Hum Genet.* *74*, 1128-1135.
- Chen, Y.Z., Hashemi, S.H., Anderson, S.K., Huang, Y., Moreira, M.C., Lynch, D.R., Glass, I.A., Chance, P.F., and Bennett, C.L. (2006). Senataxin, the yeast Sen1p orthologue: characterization of a unique protein in which recessive mutations cause ataxia and dominant mutations cause motor neuron disease. *Neurobiol Dis.* *23*, 97-108.
- Cheng, Z., He, Y.W., Lim, S.C., Qamra, R., Walsh, M.A., Zhang, L.H., and Song, H. (2010). Structural basis of the sensor-synthase interaction in autoinduction of the quorum sensing signal DSF biosynthesis. *Structure.* *18*, 1199-1209.
- Cheng, Z., Muhlrads, D., Lim, M.K., Parker, R., and Song, H. (2007). Structural and functional insights into the human Upf1 helicase core. *EMBO J.* *26*, 253-264.
- Clerici, M., Mourao, A., Gutsche, I., Gehring, N.H., Hentze, M.W., Kulozik, A., Kadlec, J., Sattler, M., and Cusack, S. (2009). Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J.* *28*, 2293-2306.
- Cordin, O., Banroques, J., Tanner, N.K., and Linder, P. (2006). The DEAD-box protein family of RNA helicases. *Gene.* *367*, 17-37.
- Cordin, O., Tanner, N.K., Doere, M., Linder, P., and Banroques, J. (2004). The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J.* *23*, 2478-2487.
- Cox, G.A., Mahaffey, C.L., and Frankel, W.N. (1998). Identification of the mouse neuromuscular degeneration gene and mapping of a second site suppressor allele. *Neuron.* *21*, 1327-1337.

- Czaplinski, K., Weng, Y., Hagan, K.W., and Peltz, S.W. (1995). Purification and characterization of the Upf1 protein: a factor involved in translation and mRNA degradation. *RNA*. *1*, 610-623.
- de Planell-Saguer, M., Schroeder, D.G., Rodicio, M.C., Cox, G.A., and Mourelatos, Z. (2009). Biochemical and genetic evidence for a role of IGHMBP2 in the translational machinery. *Hum Mol Genet*. *18*, 2115-2126.
- Dillingham, M.S., Soultanas, P., Wiley, P., Webb, M.R., and Wigley, D.B. (2001). Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. *Proc Natl Acad Sci U S A*. *98*, 8381-8387.
- Dong, Y.H., Wang, L.Y., and Zhang, L.H. (2007). Quorum-quenching microbial infections: mechanisms and implications. *Philos Trans R Soc Lond B Biol Sci*. *362*, 1201-1211.
- Dow, J.M., Crossman, L., Findlay, K., He, Y.Q., Feng, J.X., and Tang, J.L. (2003). Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. *Proc Natl Acad Sci U S A*. *100*, 10995-11000.
- Dow, J.M., and Daniels, M.J. (1994). Pathogenicity determinants and global regulation of pathogenicity of *Xanthomonas campestris* pv. *campestris*. *Curr Top Microbiol Immunol*. *192*, 29-41.
- Dow, J.M., Feng, J.X., Barber, C.E., Tang, J.L., and Daniels, M.J. (2000). Novel genes involved in the regulation of pathogenicity factor production within the *rpf* gene cluster of *Xanthomonas campestris*. *Microbiology*. *146* (Pt 4), 885-891.
- Edwards, P.M. (2002). Origin 7.0: Scientific Graphing and Data Analysis Software. *Journal of Chemical Information and Computer Sciences*. *42*, 1270-1271.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. *60*, 2126-2132.
- Engel, C.K., Kiema, T.R., Hiltunen, J.K., and Wierenga, R.K. (1998). The crystal structure of enoyl-CoA hydratase complexed with octanoyl-CoA reveals the structural adaptations required for binding of a long chain fatty acid-CoA molecule. *J Mol Biol*. *275*, 847-859.
- Engel, C.K., Mathieu, M., Zeelen, J.P., Hiltunen, J.K., and Wierenga, R.K. (1996). Crystal structure of enoyl-coenzyme A (CoA) hydratase at 2.5 angstroms resolution: a spiral fold defines the CoA-binding pocket. *EMBO J*. *15*, 5135-5145.
- Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol*. *20*, 313-324.
- Fukita, Y., Mizuta, T.R., Shirozu, M., Ozawa, K., Shimizu, A., and Honjo, T. (1993). The human S mu bp-2, a DNA-binding protein specific to the single-stranded guanine-rich

sequence related to the immunoglobulin mu chain switch region. *J Biol Chem.* 268, 17463-17470.

Giannini, A., Pinto, A.M., Rossetti, G., Prandi, E., Tiziano, D., Brahe, C., and Nardocci, N. (2006). Respiratory failure in infants due to spinal muscular atrophy with respiratory distress type 1. *Intensive Care Med.* 32, 1851-1855.

Gorbalenya, A.E., and Koonin, E.V. (1988). One more conserved sequence motif in helicases. *Nucleic Acids Res.* 16, 7734.

Gorbalenya, A.E., and Koonin, E.V. (1993). Helicases: amino acid comparisons and structure-function relationships. *Current Opinion Structural Biology.* 419-429.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics.* 15, 305-308.

Grishin, N.V. (1998). The R3H motif: a domain that binds single-stranded nucleic acids. *Trends Biochem Sci.* 23, 329-330.

Grohmann, K., Rossoll, W., Kobsar, I., Holtmann, B., Jablonka, S., Wessig, C., Stoltenburg-Didinger, G., Fischer, U., Hubner, C., Martini, R., *et al.* (2004). Characterization of Ighmbp2 in motor neurons and implications for the pathomechanism in a mouse model of human spinal muscular atrophy with respiratory distress type 1 (SMARD1). *Hum Mol Genet.* 13, 2031-2042.

Grohmann, K., Schuelke, M., Diers, A., Hoffmann, K., Lucke, B., Adams, C., Bertini, E., Leonhardt-Horti, H., Muntoni, F., Ouvrier, R., *et al.* (2001). Mutations in the gene encoding immunoglobulin mu-binding protein 2 cause spinal muscular atrophy with respiratory distress type 1. *Nat Genet.* 29, 75-77.

Grohmann, K., Varon, R., Stolz, P., Schuelke, M., Janetzki, C., Bertini, E., Bushby, K., Muntoni, F., Ouvrier, R., Van Maldergem, L., *et al.* (2003). Infantile spinal muscular atrophy with respiratory distress type 1 (SMARD1). *Ann Neurol.* 54, 719-724.

Grohmann, K., Wienker, T.F., Saar, K., Rudnik-Schoneborn, S., Stoltenburg-Didinger, G., Rossi, R., Novelli, G., Nurnberg, G., Pfeufer, A., Wirth, B., *et al.* (1999). Diaphragmatic spinal muscular atrophy with respiratory distress is heterogeneous, and one form is linked to chromosome 11q13-q21. *Am J Hum Genet.* 65, 1459-1462.

Guenther, U.P., Handoko, L., Laggenbauer, B., Jablonka, S., Chari, A., Alzheimer, M., Ohmer, J., Plottner, O., Gehring, N., Sickmann, A., *et al.* (2009a). IGHMBP2 is a ribosome-associated helicase inactive in the neuromuscular disorder distal SMA type 1 (DSMA1). *Hum Mol Genet.* 18, 1288-1300.

Guenther, U.P., Handoko, L., Varon, R., Stephani, U., Tsao, C.Y., Mendell, J.R., Lutzkendorf, S., Hubner, C., von Au, K., Jablonka, S., *et al.* (2008). Clinical variability in distal spinal muscular atrophy type 1 (DSMA1): determination of steady-state IGHMBP2 protein levels in five patients with infantile and juvenile disease. *J Mol Med.* 87, 31-41.

- Guenther, U.P., Handoko, L., Varon, R., Stephani, U., Tsao, C.Y., Mendell, J.R., Lutzkendorf, S., Hubner, C., von Au, K., Jablonka, S., *et al.* (2009b). Clinical variability in distal spinal muscular atrophy type 1 (DSMA1): determination of steady-state IGHMBP2 protein levels in five patients with infantile and juvenile disease. *J Mol Med.* 87, 31-41.
- Guenther, U.P., Varon, R., Schlicke, M., Dutrannoy, V., Volk, A., Hubner, C., von Au, K., and Schuelke, M. (2007). Clinical and mutational profile in spinal muscular atrophy with respiratory distress (SMARD): defining novel phenotypes through hierarchical cluster analysis. *Hum Mutat.* 28, 808-815.
- Hamed, R.B., Batchelar, E.T., Clifton, I.J., and Schofield, C.J. (2008). Mechanisms and structures of crotonase superfamily enzymes--how nature controls enolate and oxyanion reactivity. *Cell Mol Life Sci.* 65, 2507-2527.
- He, Y.W. (2006). Cell-cell communication in *Xanthomonas campestris* pv. *campestris*. In Library (National University of Singapore).
- He, Y.W., Wang, C., Zhou, L., Song, H., Dow, J.M., and Zhang, L.H. (2006). Dual signaling functions of the hybrid sensor kinase RpfC of *Xanthomonas campestris* involve either phosphorelay or receiver domain-protein interaction. *J Biol Chem.* 281, 33414-33421.
- He, Y.W., and Zhang, L.H. (2008). Quorum sensing and virulence regulation in *Xanthomonas campestris*. *FEMS Microbiol Rev.* 32, 842-857.
- Hedlund, E., Hefferan, M.P., Marsala, M., and Isacson, O. (2007). Cell therapy and stem cells in animal models of motor neuron disorders. *Eur J Neurosci.* 26, 1721-1737.
- Hofstein, H.A., Feng, Y., Anderson, V.E., and Tonge, P.J. (1999). Role of glutamate 144 and glutamate 164 in the catalytic mechanism of enoyl-CoA hydratase. *Biochemistry.* 38, 9508-9516.
- Holden, H.M., Benning, M.M., Haller, T., and Gerlt, J.A. (2001). The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme a thioesters. *Acc Chem Res.* 34, 145-157.
- Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem Sci.* 20, 478-480.
- Hubbard, P.A., Yu, W., Schulz, H., and Kim, J.J. (2005). Domain swapping in the low-similarity isomerase/hydratase superfamily: the crystal structure of rat mitochondrial Delta3, Delta2-enoyl-CoA isomerase. *Protein Sci.* 14, 1545-1555.
- Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. *Trends Biochem Sci.* 36, 19-29.

Jankowsky, E., and Fairman, M.E. (2007). RNA helicases--one fold for many functions. *Curr Opin Struct Biol.* *17*, 316-324.

Joseph, S., Robb, S.A., Mohammed, S., Lillis, S., Simonds, A., Manzur, A.Y., Walter, S., and Wraige, E. (2009). Interfamilial phenotypic heterogeneity in SMARD1. *Neuromuscul Disord.* *19*, 193-195.

Kabsch, W. (1988). Automatic indexing of rotation diffraction patterns. *Journal of Applied Crystallography* *21*, 67-72.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography.* *26*, 795-800.

Kaindl, A.M., Guenther, U.P., Rudnik-Schoneborn, S., Varon, R., Zerres, K., Schuelke, M., Hubner, C., and von Au, K. (2008). Spinal muscular atrophy with respiratory distress type 1 (SMARD1). *J Child Neurol.* *23*, 199-204.

Katoh, E., Hatta, T., Shindo, H., Ishii, Y., Yamada, H., Mizuno, T., and Yamazaki, T. (2000). High precision NMR structure of YhhP, a novel *Escherichia coli* protein implicated in cell division. *J Mol Biol.* *304*, 219-229.

Kerr, D., and Khalili, K. (1991). A recombinant cDNA derived from human brain encodes a DNA binding protein that stimulates transcription of the human neurotropic virus JCV. *J Biol Chem.* *266*, 15876-15881.

Koonin, E.V. (1992). A new group of putative RNA helicases. *Trends Biochem Sci.* *17*, 495-497.

Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M., and Waksman, G. (1997). Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell.* *90*, 635-647.

Kurimoto, K., Fukai, S., Nureki, O., Muto, Y., and Yokoyama, S. (2001). Crystal structure of human AUH protein, a single-stranded RNA binding homolog of enoyl-CoA hydratase. *Structure.* *9*, 1253-1263.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography.* *26*, 283-291.

Laskowski, R.A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. . *J. Appl. Crystallogr.* *26*, 283.

Lee, J.Y., and Yang, W. (2006). UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell.* *127*, 1349-1360.

- Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., *et al.* (1995). Identification and characterization of a spinal muscular atrophy-determining gene. *Cell*. *80*, 155-165.
- Leslie, A.G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr*. *62*, 48-57.
- Leslie, A.G.W. (1992). Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography No. 26 27-33.
- Liepinsh, E., Leonchiks, A., Sharipo, A., Guignard, L., and Otting, G. (2003). Solution structure of the R3H domain from human Smubp-2. *J Mol Biol*. *326*, 217-223.
- Matthews, B.W. (1968). Solvent content of protein crystals. *J Mol Biol*. *33*, 491-497.
- Maystadt, I., Zarhrate, M., Landrieu, P., Boespflug-Tanguy, O., Sukno, S., Collignon, P., Melki, J., Verellen-Dumoulin, C., Munnich, A., and Viollet, L. (2004). Allelic heterogeneity of SMARD1 at the IGHMBP2 locus. *Hum Mutat*. *23*, 525-526.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr*. *40*, 658-674.
- Melki, J., Lefebvre, S., Burglen, L., Burlet, P., Clermont, O., Millasseau, P., Reboullet, S., Benichou, B., Zeviani, M., Le Paslier, D., *et al.* (1994). De novo and inherited deletions of the 5q13 region in spinal muscular atrophies. *Science*. *264*, 1474-1477.
- Mellins, R.B. (1974). Pulmonary physiotherapy in the pediatric age group. *Am Rev Respir Dis*. *110*, 137-142.
- Milton, D.L. (2006). Quorum sensing in vibrios: complexity for diversification. *Int J Med Microbiol*. *296*, 61-71.
- Mingarro, I., Abad, C., and Braco, L. (1995). Interfacial activation-based molecular bioimprinting of lipolytic enzymes. *Proc Natl Acad Sci U S A*. *92*, 3308-3312.
- Mohan, U., Misra, V.P., Britto, J., Muntoni, F., King, R.H., and Thomas, P.K. (2001). Inherited early onset severe axonal polyneuropathy with respiratory failure and autonomic involvement. *Neuromuscul Disord*. *11*, 395-399.
- Molnar, G.M., Crozat, A., Kraeft, S.K., Dou, Q.P., Chen, L.B., and Pardee, A.B. (1997). Association of the mammalian helicase MAH with the pre-mRNA splicing complex. *Proc Natl Acad Sci U S A*. *94*, 7831-7836.
- Moreira, M.C., Klur, S., Watanabe, M., Nemeth, A.H., Le Ber, I., Moniz, J.C., Tranchant, C., Aubourg, P., Tazir, M., Schols, L., *et al.* (2004). Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nat Genet*. *36*, 225-227.

- Morris, R.J., Perrakis, A., and Lamzin, V.S. (2002). ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr D Biol Crystallogr.* 58, 968-975.
- Müller-Newen, G., Janssen, U., and Stoffel, W. (1995). Enoyl-CoA hydratase and isomerase form a superfamily with a common active-site glutamate residue. *Eur J Biochem.* 228, 68-73.
- Müller-Newen, G., and Stoffel, W. (1993). Site-directed mutagenesis of putative active-site amino acid residues of 3,2-trans-enoyl-CoA isomerase, conserved within the low-homology isomerase/hydratase enzyme family. *Biochemistry.* 32, 11405-11412.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* 53, 240-255.
- Mursula, A.M., Hiltunen, J.K., and Wierenga, R.K. (2004). Structural studies on delta(3)-delta(2)-enoyl-CoA isomerase: the variable mode of assembly of the trimeric disks of the crotonase superfamily. *FEBS Lett.* 557, 81-87.
- Mursula, A.M., van Aalten, D.M., Hiltunen, J.K., and Wierenga, R.K. (2001). The crystal structure of delta(3)-delta(2)-enoyl-CoA isomerase. *J Mol Biol.* 309, 845-853.
- Neidhardt, F.C., Bloch, P.L., and Smith, D.F. (1974). Culture medium for enterobacteria. *J Bacteriol.* 119, 736-747.
- Onsando, J.M. (1992). Black rot of crucifers. In *Diseases of Vegetables and Oil Seed Crops*, K.J. Chaube H.S., Mukhopadhyay A.N., Singh U.S., ed. (New Jersey, Prentice Hall), pp. 243-252.
- Partanen, S.T., Novikov, D.K., Popov, A.N., Mursula, A.M., Hiltunen, J.K., and Wierenga, R.K. (2004). The 1.3 Å crystal structure of human mitochondrial Delta3-Delta2-enoyl-CoA isomerase shows a novel mode of binding for the fatty acyl group. *J Mol Biol.* 342, 1197-1208.
- Pause, A., Methot, N., and Sonenberg, N. (1993). The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol Cell Biol.* 13, 6789-6798.
- Pierson, T.M., Tart, G., Adams, D., Toro, C., Golas, G., Tifft, C., and Gahl, W. (2011). Infantile-onset spinal muscular atrophy with respiratory distress-1 diagnosed in a 20-year-old man. *Neuromuscul Disord.* 21, 353-355.
- Pitt, M., Houlden, H., Jacobs, J., Mok, Q., Harding, B., Reilly, M., and Surtees, R. (2003). Severe infantile neuropathy with diaphragmatic weakness and its relationship to SMARD1. *Brain.* 126, 2682-2692.
- Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr.* 59, 1131-1137.

Saikrishnan, K., Griffiths, S.P., Cook, N., Court, R., and Wigley, D.B. (2008). DNA binding to RecD: role of the 1B domain in SF1B helicase activity. *EMBO J.* 27, 2222-2229.

Saikrishnan, K., Powell, B., Cook, N.J., Webb, M.R., and Wigley, D.B. (2009). Mechanistic basis of 5'-3' translocation in SF1B helicases. *Cell.* 137, 849-859.

Sebastiani, G., Durocher, D., Gros, P., Nemer, M., and Malo, D. (1995). Localization of the Catf1 transcription factor gene to mouse chromosome 19. *Mamm Genome.* 6, 147-148.

Sheldrick, G.M. (2008). A short history of SHELX. *Acta Crystallogr A.* 64, 112-122.

Shieh, S.Y., Stellrecht, C.M., and Tsai, M.J. (1995). Molecular characterization of the rat insulin enhancer-binding complex 3b2. Cloning of a binding factor with putative helicase motifs. *J Biol Chem.* 270, 21503-21508.

Singleton, M.R., Dillingham, M.S., Gaudier, M., Kowalczykowski, S.C., and Wigley, D.B. (2004). Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature.* 432, 187-193.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem.* 76, 23-50.

Singleton, M.R., Scaife, S., and Wigley, D.B. (2001). Structural analysis of DNA replication fork reversal by RecG. *Cell.* 107, 79-89.

Singleton, M.R., and Wigley, D.B. (2002). Modularity and specialization in superfamily 1 and 2 helicases. *J Bacteriol.* 184, 1819-1826.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell.* 42, 794-805.

Slater, H., Alvarez-Morales, A., Barber, C.E., Daniels, M.J., and Dow, J.M. (2000). A two-component system involving an HD-GYP domain protein links cell-cell signalling to pathogenicity gene expression in *Xanthomonas campestris*. *Mol Microbiol.* 38, 986-1003.

Soultanas, P., Dillingham, M.S., Wiley, P., Webb, M.R., and Wigley, D.B. (2000). Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism. *EMBO J.* 19, 3799-3810.

Soultanas, P., and Wigley, D.B. (2000). DNA helicases: 'inching forward'. *Curr Opin Struct Biol.* 10, 124-128.

Soultanas, P., and Wigley, D.B. (2001). Unwinding the 'Gordian knot' of helicase action. *Trends Biochem Sci.* 26, 47-54.

- Story, R.M., and Steitz, T.A. (1992). Structure of the recA protein-ADP complex. *Nature*. 355, 374-376.
- Subramanya, H.S., Bird, L.E., Brannigan, J.A., and Wigley, D.B. (1996). Crystal structure of a DExx box DNA helicase. *Nature*. 384, 379-383.
- Sun, X., Perlick, H.A., Dietz, H.C., and Maquat, L.E. (1998). A mutated human homologue to yeast Upf1 protein has a dominant-negative effect on the decay of nonsense-containing mRNAs in mammalian cells. *Proc Natl Acad Sci U S A*. 95, 10009-10014.
- Suraweera, A., Lim, Y., Woods, R., Birrell, G.W., Nasim, T., Becherel, O.J., and Lavin, M.F. (2009). Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Hum Mol Genet*. 18, 3384-3396.
- Tang, J.L., Liu, Y.N., Barber, C.E., Dow, J.M., Wootton, J.C., and Daniels, M.J. (1991). Genetic and molecular analysis of a cluster of rpf genes involved in positive regulation of synthesis of extracellular enzymes and polysaccharide in *Xanthomonas campestris* pathovar *campestris*. *Mol Gen Genet*. 226, 409-417.
- Tanner, N.K., Cordin, O., Banroques, J., Doere, M., and Linder, P. (2003). The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell*. 11, 127-138.
- Tomko, E.J., Fischer, C.J., Niedziela-Majka, A., and Lohman, T.M. (2007). A nonuniform stepping mechanism for *E. coli* UvrD monomer translocation along single-stranded DNA. *Mol Cell*. 26, 335-347.
- Vantaggiato, C., Bondioni, S., Airoldi, G., Bozzato, A., Borsani, G., Rugarli, E.I., Bresolin, N., Clementi, E., and Bassi, M.T. (2011). Senataxin modulates neurite growth through fibroblast growth factor 8 signalling. *Brain*. 134, 1808-1828.
- Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S., and Wigley, D.B. (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell*. 97, 75-84.
- Wang, L.H., He, Y., Gao, Y., Wu, J.E., Dong, Y.H., He, C., Wang, S.X., Weng, L.X., Xu, J.L., Tay, L., *et al.* (2004). A bacterial cell-cell communication signal with cross-kingdom structural analogues. *Mol Microbiol*. 51, 903-912.
- Weng, Y., Czaplinski, K., and Peltz, S.W. (1998). ATP is a cofactor of the Upf1 protein that modulates its translation termination and RNA binding activities. *RNA*. 4, 205-214.
- Whitehead, N.A., Barnard, A.M., Slater, H., Simpson, N.J., and Salmond, G.P. (2001). Quorum-sensing in Gram-negative bacteria. *FEMS Microbiol Rev*. 25, 365-404.

Wilmshurst, J.M., Bye, A., Rittey, C., Adams, C., Hahn, A.F., Ramsay, D., Pamphlett, R., Pollard, J.D., and Ouvrier, R. (2001). Severe infantile axonal neuropathy with respiratory failure. *Muscle Nerve*. 24, 760-768.

Wilson, T.J., Bertrand, N., Tang, J.L., Feng, J.X., Pan, M.Q., Barber, C.E., Dow, J.M., and Daniels, M.J. (1998). The *rpfA* gene of *Xanthomonas campestris* pathovar *campestris*, which is involved in the regulation of pathogenicity factor production, encodes an aconitase. *Mol Microbiol*. 28, 961-970.

Wong, V.C., Chung, B.H., Li, S., Goh, W., and Lee, S.L. (2006). Mutation of gene in spinal muscular atrophy respiratory distress type I. *Pediatr Neurol*. 34, 474-477.

Wu, M., Reuter, M., Lilie, H., Liu, Y., Wahle, E., and Song, H. (2005). Structural insight into poly(A) binding and catalytic mechanism of human PARN. *EMBO J*. 24, 4082-4093.

Ye, J., Osborne, A.R., Groll, M., and Rapoport, T.A. (2004). RecA-like motor ATPases--lessons from structures. *Biochim Biophys Acta*. 1659, 1-18.

Zhang, Q., Wang, Y.C., and Montalvo, E.A. (1999). Smubp-2 represses the Epstein-Barr virus lytic switch promoter. *Virology*. 255, 160-170.