Investigating the function and regulation of two CRISPR systems of 
*Streptococcus mutans*

MIHAELA ANCA SERBANESCU

A thesis submitted in conformity with the requirements for the 
Degree of Doctor of Philosophy 
Graduate Department of Dentistry 
University of Toronto

Copyright by Mihaela A. Serbanescu 2015 ©
Investigating the function and regulation of two CRISPR systems of *Streptococcus mutans*

Mihaela Anca Serbanescu  
Doctor of Philosophy, Faculty of Dentistry, University of Toronto  
2015

Abstract

CRISPR-Cas (clustered, regularly interspaced short palindromic repeats–CRISPR-associated proteins) provide adaptive microbial immunity systems against invading viruses and natural transformation via plasmids. Naturally competent *Streptococcus mutans* UA159 harbors two CRISPR-Cas systems: type II-A system (CRISPR1) and type I-C system (CRISPR2) and several spacers matching sequences of phage M102 or genomic sequences of other *S. mutans* strains. In addition, previous transcriptome studies in *S. mutans* linked CRISPR/Cas systems to stress response and virulence. The goal of this work was to determine the role of CRISPR/Cas systems in phage defense and natural transformation in *S. mutans*, and also to investigate if they play additional functions in the cell physiology. Deletion of CRISPR1 and/or CRISPR2 cas genes or removal of the spacers in *S. mutans* UA159 did not affect its M102 phage resistant phenotype, suggesting that CRISPR-independent mechanisms contribute to the phage resistance. Using a plasmid-based interference assay, we identified DNA interference activity in *S. mutans* UA159, which is mediated by its type II-A CRISPR/Cas system (CRISPR1). Spacers 2 and 3 from type II-A, both matching sequences from phage M102, were found to be essential for CRISPR interference against engineered plasmids containing matching proto-spacer sequences. Functional analysis of the cas deletion mutants revealed that CRISPR1-Cas system modulates stress tolerance induced by low pH, high temperature, oxidative and cell membrane stress, as well as DNA damaging conditions, whereas the CRISPR2-Cas participates in the tolerance associated with heat shock. Transcriptional analysis identified that VicR/K two-component signal transduction system differentially regulates the expression of cas genes for both systems in *S. mutans*. Further, structural, biochemical and functional studies found that the Cas5d protein SMU.1763c, a putative endoribonuclease associated with the CRISPR system I-C, acts on structured RNA substrates that is likely to be involved in CRISPR RNA processing but not in sensing cell envelope stress or preserving cell integrity in *S. mutans*. Together our data provide *in vivo* evidence that the CRISPR-Cas systems of *S. mutans* play novel roles in resistance against incoming plasmids that carry matching protospacer sequences and stress response.
Acknowledgements

I would like to extend sincere thanks to both my supervisors, Dr. Dennis Cvitkovitch and Dr. Alexander Yakunin for their patience, mentorship and support over these years. I am particularly grateful to Dr. Cvitkovitch for believing in me with this project. I truly appreciate the opportunities, challenges, advice and friendship you have given me. I am also very grateful to Dr. Yakunin for his scientific advice and knowledge and many insightful discussions and suggestions. He was instrumental in helping me finish up this thesis. I would like to thank Dr. Senadheera for her invaluable critiques and expert guidance on the project. I sincerely thank my committee members, Dr. Celine Levesque and Dr. Morris Manolson for their advice and expert opinions. I am also grateful to my former MSc research advisor Dr. Debora Foster, who set me on this science path. Thank you to all members of Cvitkovitch lab for stimulating discussions over the years and particularly Martha, Kirsten, Timmy, Andrew, Gursonika, Kamna, Iwona, and Marie-Christine. Your friendship and support has been much appreciated. I also thank members of Yakunin lab, Natasha and Robert for providing help and valuable advice on the protein work. Outside the lab, I would like to thank my family and friends for always listening when I needed it, and encouraging through the though times. Special thanks go to my mom for being a wonderful grandparent, and for her constant and unwavering support. The last people I would like to thank are my husband Lucian and our children: Vlad, Tudor and Victor. My little boys' smiles helped me go through the ups and downs of the entire process towards the doctoral dissertation. As for Lucian I am endlessly grateful for the opportunities his love have provided me, and for the peace he brings to my life. He has shared this entire amazing journey with me, and I would like to dedicate this dissertation to him.

M. Anca Serbanescu
2015
Table of Contents

Abstract ................................................................................................................................. ii
Acknowledgements ............................................................................................................... iii
Table of Contents ................................................................................................................ iv
Original Contributions by the Author ................................................................................ vii
List of Tables ....................................................................................................................... viii
List of Figures ...................................................................................................................... ix
Preface ................................................................................................................................... x
Awards ................................................................................................................................. xii
Abbreviations ..................................................................................................................... xiv

Chapter 1: Literature Review ............................................................................................. 1

1.1. Streptococcus mutans - key role in tooth decay ....................................................... 2

1.1.1. Response of S. mutans to environmental changes ............................................. 4
1.1.2. DNA repair in S. mutans .................................................................................... 7
1.1.3. Natural transformation in S. mutans ......................................................................... 9

1.2. Overview of CRISPR-Cas immunity ......................................................................... 18

1.2.1. CRISPR-Cas organization .................................................................................. 18
1.2.2. CRISPR distribution ............................................................................................ 19
1.2.3. Mechanism ............................................................................................................ 19

1.3. History .......................................................................................................................... 21

1.4. Structure of the CRISPR-Cas locus .......................................................................... 22

1.4.1. Repeats .................................................................................................................. 23
1.4.2. Spacers .................................................................................................................. 24
1.4.3. Leader ..................................................................................................................... 24
1.4.4. Cas proteins ........................................................................................................... 25
1.4.5. PAM ....................................................................................................................... 29

1.5. CRISPR-Cas classification ......................................................................................... 30

1.5.1. Type I CRISPR-Cas systems .............................................................................. 30
1.5.2. Type II CRISPR-Cas systems ............................................................................ 34
1.5.3. Type III CRISPR-Cas systems ............................................................................ 35
1.5.4. Unclassified CRISPR-Cas systems ..................................................................... 39

1.6. CRISPR-Cas immunity .............................................................................................. 39
1.6.1. Adaptation ........................................................................................................39
1.6.2. Biogenesis of crRNA .......................................................................................41
1.6.3. Interference ..................................................................................................42
1.7. CRISPR–Cas systems- mechanism(s) of self versus non-self discrimination.................................................................51
1.8. Alternative functions of CRISPR-Cas systems in virulence, gene regulation and physiology..................................................................................................................52
  1.8.1. CRISPR-Cas systems, virulence and gene regulation .......................................52
  1.8.2. CRISPR-Cas systems and cell physiology .......................................................54
1.9. CRISPR/Cas systems in S. mutans ......................................................................55
  1.9.1. Role of CRISPR/Cas systems in bacteriophage resistance ............................55
  1.9.2. CRISPR/Cas link to stress response ................................................................58
1.10. Rationale ........................................................................................................59
1.11. References ......................................................................................................61

Chapter 2: Role of the Streptococcus mutans CRISPR/Cas systems in immunity and cell physiology ..................................................84
2.1. Abstract ...........................................................................................................85
2.2. Introduction ......................................................................................................85
2.3. Materials and Methods ....................................................................................88
2.4. Results ............................................................................................................95
  2.4.1. The S. mutans UA159 CRISPR-Cas systems ..................................................95
  2.4.2. Loss of the M102–specific CRISPR spacers or cas genes of both CRISPR-Cas systems have no effect on the phage resistance phenotype of S. mutans UA159 ..................................................................................................................98
  2.4.3. The type II-A CRISPR-Cas system prevents natural transformation by plasmids in S. mutans UA159 .................................................................99
  2.4.4. Purified recombinant SMU.1763c (Cas5d) has ribonuclease activity .... 103
  2.4.5. Cas5d cleaves cellular RNAs .......................................................................105
  2.4.6. Cas proteins are involved in sensitivity to DNA damage ...............................106
  2.4.7. ∆C1S responds to oxidative, SDS, acid and high temperature stressors ..........................................................................................107
  2.4.8. Loss of VicK modulates the transcription of S. mutans cas genes ...........110
Original Contributions by the Author

Publications resulting from this thesis work:


M. Anca Serbanescu wrote the manuscript and performed all experiments except: transformation assays (Figure 2-2 C and D) which were performed in conjunction with Kirsten Krastel, cleavage of ssRNA substrates by the Cas5 protein from S. mutans UA159 (Figure 2-3) were performed by Dr. Natalya Beloglazova and Robert Flick, and the Cas5d cleavage of RNA transcripts of SMU.995 and SMU.1502c generated by in vitro transcription (Figure 2-4) was performed by Martha Cordova. Dr. Dilani Senadheera helped in the writing of the manuscript. Dr. Cvitkovitch and Dr. Yakunin contributed to data analysis and writing of this manuscript.


M. Anca Serbanescu wrote the manuscript and performed the sequence analysis and structure-based sequence alignment (Figure 3-1 and Figure 3-5), survival of the mutant strain after exposure to heat shock (Figure 3-7). Susceptibility profiles of S. mutans strains exposed to cell envelope-interfering antimicrobials were performed in conjunction with K. Krastel (Table 3-2). Dr. N. Beloglazova performed the following experiments: Cleavage of the 50 nt fragment of S. mutans ssRNA by SMU.1763c (Figure 3-2), enzymatic assays to probe DNase and RNase activity of SMU.1763c protein and SMU.1760c protein were performed in conjunction with R. Flick (Figure 3-3), site-directed mutagenesis of the S. mutans SMU.1763c protein (Figure 3-6) in conjunction with S. Lemak. Dr. Cvitkovitch and Dr. Yakunin contributed to data analysis and writing of this manuscript.
List of Tables

Table 1-1: Two-component signal transduction systems identified in *Streptococcus mutans* UA159 .................................................................................................................................................. 6
Table 1-2: Major Cas proteins (Cas1-Cas10) .................................................................................................................................................................................. 26
Table 1-3: Cas proteins associated with type I system ......................................................................................................................................................... 32
Table 1-4: Cas proteins associated with type II system ..................................................................................................................................................... 35
Table 1-5: Cas proteins associated with type III system ..................................................................................................................................................... 38
Table 1-6: PAM-dependent DNA targeting .................................................................................................................................................................................. 48
Table 2-1: Bacterial strains, plasmids and phages used in this study ................................................. 89
Table S1: Primers used in this study (nucleotide sequence from 5' to 3') ......................... 125
Table S2: Synthetic DNA oligonucleotides used for constructing plasmids for *in vivo* interference assays ............................................................................................................................................... 128
Table S3: CRISPR/Cas systems present in *S. mutans* UA159 and OMZ381 ....................... 130
Table S4: Alignment of potential natural targets for *S. mutans* UA159 spacers to identify possible PAM motifs .................................................................................................................................................................................. 131
Table 3-1: Bacterial strains and primers used in this study......................................................... 144
Table 3-2: Susceptibility profiles of *S. mutans* strains exposed to cell envelope- interfering antimicrobials .................................................................................................................................................................................. 160
List of Figures

Figure 1-1: A schematic diagram describing the key virulence properties of *S. mutans*. .......................... 3

Figure 1-2: Simplified diagram of a typical CRISPR–Cas system. ............................................. 18

Figure 1-3: Stages of the CRISPR/Cas mediated immunity. ....................................................... 20

Figure 1-4: Simplified diagram of a CRISPR locus. ........................................................................... 23

Figure 1-5: Simplified diagram of a CRISPR locus in type I. ....................................................... 31

Figure 1-6: Simplified diagram of a CRISPR locus in type II. ...................................................... 34

Figure 1-7: Simplified diagram of a CRISPR locus in type III. ...................................................... 36

Figure 1-8: Simplified diagram of a CRISPR locus in type U. ....................................................... 39

Figure 1-9: Three stages of CRISPR-Cas action. ................................................................. 45

Figure 1-10: Putative PAM, SAM and TIM associated with adaptation and interference steps in type I and type II CRISPR systems. ......................................................... 50

Figure 1-11: Organization of CRISPR1 (A) and CRISPR2 (B) arrays and associated genes in *S. mutans* UA159. ............................................................................................................. 57

Figure 2-1: Gene map of the CRISPR-Cas systems in *S. mutans* UA159 at the top is CRISPR1-Cas system (A) and the bottom is CRISPR2-Cas system (B). ....................... 98

Figure 2-2: The CRISPR1-Cas system of *S. mutans* UA159 provides immunity against plasmid transformation. ................................................................................................. 102

Figure 2-3: Cleavage of the synthetic ssRNA substrates by the Cas5d protein from *S. mutans* UA159 ..................................................................................................................... 104

Figure 2-4: Cas5d (SMU.1763c) cleavage of RNA transcripts of SMU.995 and SMU.1502c generated by *in vitro* transcription. ......................................................................................... 106

Figure 2-5: Effects of 0.05 µg/ml MMC (A) or UV irradiation (B) on viability of *S. mutans* UA159 and mutant strains. ...................................................................................................................... 107

Figure 2-6: Growth kinetics of *S. mutans* UA159 and ΔC1S under varying stressors (A) at pH 7.0 and pH 5.5, (B) 25mM paraquat, 0.003% H₂O₂, or (C) 0.004% SDS..... 109
Figure 2-7: Survival of *S. mutans* UA159 and mutant strains after exposure to 50°C temperature stress for 1 h. .................................................................................................................. 110

Figure 2-8: Expression of *cas* genes from the CRISPR1 (A) and CRISPR2 (B) operons. .................................................................................................................................................. 111

Figure S1 (A and B): Growth curves of *S. mutans* OMZ381 and UA159 in the presence of M102. .................................................................................................................................................. 132

Figure S2: Coomassie-blue staining of SDS-PAGE gel showing the purified Cas5d (SMU.1763c) protein. .................................................................................................................................................. 133

Figure S3: Fold expression of genes likely targeted for cleavage by SMU.1763c Cas5d protein. .................................................................................................................................................. 134

Figure S4: Agarose gel electrophoresis of the RT-PCR amplification products of downstream genes SMU.1752 and SMU.1400 respectively, in *S. mutans* wild-type strain and mutants defective in the cas genes. .................................................................................................................................................. 135

Figure S5 (A, B and C): Growth kinetics of *S. mutans* UA159 and ∆C2E under varying stressors (A) at pH 7.0 and pH 5.5, (B) 25mM paraquat, 0.003% H₂O₂, or (C) 0.004% SDS. .................................................................................................................................................. 136

Figure S6 (A and B): Northern blot analysis of tracrRNA expression in *S. mutans* UA159 at different time points of growth and various stressors suggest processing of tracrRNA. .................................................................................................................................................. 137

Figure 3-1: Organization of the *cas* and CRISPR locus for the Type I-C/Dvulg of *S. mutans* UA159. .................................................................................................................................................. 149

Figure 3-2: Cleavage of the 50 nt fragment of *Streptococcus mutans* ssRNA by SMU.1763c. .................................................................................................................................................. 150

Figure 3-3: Enzymatic assays to probe (A) DNase and (B) RNase activity of SMU.1763c protein and SMU.1760c protein. .................................................................................................................................................. 152

Figure 3-4: Crystal structure of the *S. mutans* SMU.1763c. .................................................................................................................................................. 153

Figure 3-5: Structure-based sequence alignment of Cas5d *S. mutans* (SmCas5d) (PDB code 4ROJ) and Cas5d-like proteins from other organisms. .................................................................................................................................................. 155
Figure 3-6: Alanine replacement mutagenesis of *S. mutans* SMU.1763c. ........................ 157

Figure 3-7: Survival of *S. mutans* UA159 and ΔCas5d strain after exposure to 50°C temperature stress for 1h. ................................................................. 158
Preface

Dissertation format

This dissertation is presented in the „Publishable Style“. Chapter 1 presents a general introduction to the subject, and serves to provide context for the following chapters. Chapters 2 and 3 describe experimental data that have either been accepted for publication or in preparation for submission. They are presented in their submitted form, other than minor changes made to improve readability. Chapter 4 serves a brief discussion of all experimental data. Written permission for reproduction of the publication has been obtained, and is held on file.
Awards

2014 Graduate Bursary, University of Toronto
2014 Doctoral Completion Award, University of Toronto
2014 American Society for Microbiology travel award
2014 Bruce and Susan Malloch Fellowship
2014 Eleanor and Gerald Copeland Bursary
2012; 2014 Canadian Institute for Health Research (CIHR) Cell Signals travel award
2009 – 2014 Canadian Institute for Health Research (CIHR) Cell Signals Fellowship
2008 – 2012 Harron Scholarship, Faculty of Dentistry, University of Toronto
2009 – 2011 Harron travel award
2010 Canadian Institute for Health Research (CIHR) travel award
2008 1st place winner (MSc category); Laurence E. Becker Symposium; Hospital for Sick Children
2007 – 2008 Ontario Government Scholarship
2006 – 2007 Ryerson Graduate Scholarship, Ryerson University
1994 – 1997 University of Bucharest Undergraduate Scholarship
Abbreviations

AA: amino acid
Ala: alanine
AP: apurinic-apyrimidinic
ATP: adenosine triphosphate
ATR: acid tolerance response
BIP: bacteriocin-inducing peptide
BlastP: protein homology search
Cas: CRISPR-associated
Cascade: CRISPR-associated complex for antiviral defense
cAMP: cyclic adenosine monophosphate
CDM: chemically-defined medium
CRISPR: clustered regularly interspaced short palindromic repeats
crRNA: CRISPR RNA
CRP: cAMP receptor protein
CSP: competence stimulating peptide
CMR: Cas ribonucleoprotein complex involved in interference type III-B
CSM: Cas interference complex type III-A
dCMP: deoxy-CMP deaminase
ds: double-stranded
EDTA: ethylenediaminetetraacetic acid
EPS: exopolymeric substances
H: hour
His: histidine
HK: histidine kinase
IPTG: isopropyl-beta-D-thiogalactopyranoside
LB: Luria-Bertani
Lys: lysine
MBC: minimum bactericidal concentration
MIC: minimum inhibitory concentration
Min: minute
MMC: mitomycin C
NER: nucleotide excision repair
OppA: oligopeptide permease transporter
qRT-PCR: quantitative real-time PCR
PAM: protospacer-adjacent motif
PCR: polymerase chain reaction
PBS: phosphate buffered saline
RAMP: repetitive associated mysterious proteins
RRM: RNA recognition motif
R-M system: restriction/modification
RR: response regulator
S: second
SAM: spacer acquisition motif
ss: single-stranded
sCSP: synthetic CSP
TCS: two-component signal transduction system
THYE: Todd-Hewitt-Yeast extract
TIM: target interference motif.
TYE: tryptone-yeast extract
Tyr: tyrosine
tracRNA: trans-activating CRISPR RNA
WT: wild-type
XIP: sigX-inducing peptide
Chapter 1: Literature Review
1.1. **Streptococcus mutans- key role in tooth decay**

In humans, numerically predominant microorganisms in the commensal microbiota belong to the genus *Streptococcus* and they inhabit the mouth, skin, intestine and upper respiratory tract. However, there are also pathogenic streptococcal species that can cause a wide range of diseases such as strep throat, meningitis, bacterial pneumonia, endocarditis, necrotizing fasciitis and tooth decay (Mitchell, 2003). Tooth decay is one of the most prevalent infectious dental diseases affecting humans. Of the oral streptococci, mutans streptococci are frequently associated with dental caries with *Streptococcus mutans* and *Streptococcus sobrinus* being the most common members isolated from human sources (Kreth *et al.*., 2009). However, *S. mutans* is the only organism that has been conclusively recognized as the principal aetiological agent responsible for the initiation of dental caries (Loesche, 1986). As a biofilm-dwelling organism *S. mutans* exists almost exclusively in a high density and diversity of biofilm community known also as dental plaque (Cvitkovitch *et al.*, 2003). The structure and composition of the plaque are known to be strongly influenced by such factors as the source and availability of nutrients, the pH in the oral cavity, and the ability of its constituents to adapt to fluctuations in environmental conditions (Quivey Jr *et al.*, 2000). Thus, *S. mutans* uses several key virulence attributes to out-compete other residents of the dental plaque and flourish in cariogenic plaque that include: 1) ability to metabolize a wide range of carbohydrates to produce a highly adhesive dextran that enables it to attach to the tooth surface and form biofilms, 2) is highly acidogenic (produces acid) and aciduric (withstands highly acidic conditions) to eventually become dominant in active carious lesions, 3) produces antimicrobial peptides (bacteriocins) to compete with its neighbors within the dental plaque (Schilling & Bowen, 1992, Van Houte, 1994, Hitch *et al.*, 2004, Arthur *et al.*, 2011). To fine-tune all these virulence factors, *S. mutans* uses the fourteen two-component signal transduction systems 1) to sense and respond to their environment changes and 2) to monitor population density and develop genetic competence, which allows *S. mutans* to take up DNA from the environment and further adapt to its surroundings (Figure 1-1).
When sucrose becomes available, glucan synthesis allows *S. mutans* to attach to the tooth surface. Glucans interact with glucan binding proteins to promote cell-cell aggregation. In addition to sucrose, *S. mutans* can metabolize a wide range of carbohydrates, and the end product of sugar metabolism is mostly lactic acid that acidifies the local environment and is responsible for the damage that occurs on the enamel of the tooth surface and causes dental caries. To withstand these continual cycles of acid shock and persist in the oral biofilm, *S. mutans* uses acid induced mechanisms known as acid tolerance response (ATR). Acid tolerance is maintained by the presence of membrane-bound, proton-translocating F1F0-ATPase that regulates the intracellular pH. Various environmental conditions are detected using two-component signal transduction
systems (TCS). *S. mutans* is also able to monitor its population density using two different signalling peptides designated as CSP and XIP that trigger competence for natural transformation, which allows *S. mutans* to take up DNA from the environment and further adapt to its surroundings.

### 1.1.1. Response of *S. mutans* to environmental changes

Bacteria can rely on regulatory systems such as two component signal transduction systems (TCS) to sense and respond to stress and changes in their environment by coordinating appropriate gene expression (Bourret *et al.*, 1991, Parkinson, 1993, Stock *et al.*, 2000). They are typically composed of a membrane-bound sensor kinase that detects environmental changes and a cytoplasmic regulatory protein that binds to DNA when activated to modulate gene expression (Stock *et al.*, 2000). Activation of the TCSs occurs when a external signal triggers the autophosphorylation of the sensor on a conserved histidine residue (referred to as histidine kinases, or HKs), which subsequently transfers the phosphoryl group to a conserved aspartate residue on the regulatory domain of the response regulator (RR) and undergoes a conformational change to activate a response (Stock *et al.*, 2000). To date there have been identified 14 TCSs and a single orphan response regulator in *S. mutans* UA159 (Sato *et al.*, 2000, Ajdic *et al.*, 2002, Idone *et al.*, 2003, Biswas & Biswas, 2006, Levesque *et al.*, 2007, Biswas *et al.*, 2008, Dunning *et al.*, 2008, Mazda *et al.*, 2012) (Table 1-1). Due to their ability to sense and respond to stress, TCSs are viewed as desirable targets for new antimicrobial therapies to reduce ability of *S. mutans* to cause caries (Lemos & Burne, 2008). Therefore, a majority of these TCSs were studied to determine their role in the stress response, ability to form biofilms and to transport and metabolize carbohydrates. The ComDE system, which senses the competence stimulating peptide (CSP) pheromone in a form of intraspecies communication known as quorum sensing (QS) was solidly studied in *S. mutans* and has been shown to regulate several physiological properties, including competence development, biofilm formation, acid tolerance and bacteriocin production all of which facilitate the growth, survival and persistence of *S. mutans* (Bhagwat *et al.*, 2001, Li *et al.*, 2001, Li *et al.*, 2001, Li *et al.*, 2002, Kreth *et al.*, 2005, Ahn *et al.*, 2006). VicRK has been linked to the oxidative stress response
(Senadheera et al., 2007), cell envelope stress (Tremblay et al., 2009) tolerance to acid (Senadheera et al., 2009), bacteriocin production and cell death (Senadheera et al., 2012), in biofilm formation and genetic competence (Senadheera et al., 2005). The CiaHR has been linked to biofilm formation, tolerance to environmental stresses, genetic competence, mutacin production (Qi et al., 2004, Ahn et al., 2006, Levesque et al., 2007, Biswas et al., 2008) as well as resistance to antimicrobial peptides (Mazda et al., 2012). Further work determined a role of the LiaFSR in acid tolerance (Li et al., 2002), oxidative stress response during biofilm growth (Perry et al., 2008), mutacin production (Zhang & Biswas, 2009) and in sensing cell envelope stress and preserving envelope integrity in S. mutans (Suntharalingam et al., 2009). The ScnRK participates in the survival of cells at acidic pH (Levesque et al., 2007), oxidative stress (Chen et al., 2008) and the TCS-9 affects the acid tolerance response and the process of streptococcal competence development (Levesque et al., 2007). Other TCSs have been linked to the stringent response (RelRS) (Lemos et al., 2007), cell envelope stress and DNA damage recognition (BceRS) (Biswas et al., 2008, Ouyang et al., 2010), tolerance to acid (SpaKR) and (HK/RR3) (Li et al., 2002, Kawada - Matsuo et al., 2009), cell wall metabolism (LytST) (Ajdic et al., 2002) and in fructose metabolism and sugar transport (LevSR) (Zeng et al., 2006). Additional reports describe a role for the GcrR in acid tolerance and sucrose dependent adherence (Sato et al., 2000, Idone et al., 2003, Biswas & Biswas, 2006, Dunning et al., 2008). All these investigations provided indirect evidence that cross-regulation between certain histidine kinases and their non-cognate response regulators most likely occur in S. mutans to form global networks that can facilitate the bacterial response to environment. Direct evidence of cross-regulation in S. mutans to modulate stress response pathways has been described only recently (Downey et al., 2014). In addition to activating its cognate response regulator, VicK of the VicKR TCSTS was shown to transphosphorylate the non-cognate stress regulatory response regulator GcrR (Downey et al., 2014).
Table 1-1: Two-component signal transduction systems identified in *Streptococcus mutans* UA159.
(modified from (Levesque *et al.*, 2007) and (Smith & Spatafora, 2012)).

<table>
<thead>
<tr>
<th>TCS</th>
<th>Locus (NCBI)</th>
<th>Known Phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TCS-1 (VicKR)</strong></td>
<td>SMU.1516</td>
<td>oxidative stress, biofilm formation, genetic competence, acid tolerance, cell envelope stress</td>
</tr>
<tr>
<td></td>
<td>SMU.1517</td>
<td></td>
</tr>
<tr>
<td><strong>TCS-2 (CiaHR)</strong></td>
<td>SMU.1128</td>
<td>mutacin production, genetic competence, biofilm formation, oxidative and osmotic stress tolerance, sensitivity to DNA damage, acid tolerance, calcium signalling</td>
</tr>
<tr>
<td></td>
<td>SMU.1129</td>
<td></td>
</tr>
<tr>
<td><strong>TCS-3</strong></td>
<td>SMU.1145</td>
<td>acid stress</td>
</tr>
<tr>
<td></td>
<td>SMU.1146</td>
<td></td>
</tr>
<tr>
<td><strong>TCS-4 (RelRS)</strong></td>
<td>SMU.928</td>
<td>stringent response</td>
</tr>
<tr>
<td></td>
<td>SMU.927</td>
<td></td>
</tr>
<tr>
<td><strong>TCS-5 (ScnRK)</strong></td>
<td>SMU.1814</td>
<td>oxidative stress, acid tolerance</td>
</tr>
<tr>
<td></td>
<td>SMU.1815</td>
<td></td>
</tr>
<tr>
<td><strong>TCS-6 (SpaKR)</strong></td>
<td>SMU.660</td>
<td>acid tolerance</td>
</tr>
<tr>
<td></td>
<td>SMU.659</td>
<td></td>
</tr>
<tr>
<td><strong>TCS-7</strong></td>
<td>SMU.1037</td>
<td>uncharacterized</td>
</tr>
<tr>
<td></td>
<td>SMU.1038</td>
<td></td>
</tr>
<tr>
<td>TCS-8 (BceRS)</td>
<td>SMU.1009</td>
<td>cell envelope stress</td>
</tr>
<tr>
<td>TCS-8 (BceRS)</td>
<td>SMU.1008</td>
<td></td>
</tr>
<tr>
<td>TCS-9 (LevSR)</td>
<td>SMU.1965</td>
<td>fructose metabolism, sugar transport</td>
</tr>
<tr>
<td>TCS-9 (LevSR)</td>
<td>SMU.1964</td>
<td></td>
</tr>
<tr>
<td>TCS-10 (LytST)</td>
<td>SMU.576</td>
<td>cell wall metabolism</td>
</tr>
<tr>
<td>TCS-10 (LytST)</td>
<td>SMU.577</td>
<td></td>
</tr>
<tr>
<td>TCS-11 (LiaRS)</td>
<td>SMU.486</td>
<td>cell envelope stress, acid tolerance, osmotic stress, oxidative stress, biofilm formation, mutacin production</td>
</tr>
<tr>
<td>TCS-11 (LiaRS)</td>
<td>SMU.487</td>
<td></td>
</tr>
<tr>
<td>TCS-12</td>
<td>SMU.1548</td>
<td>uncharacterized</td>
</tr>
<tr>
<td>TCS-12</td>
<td>SMU.1549</td>
<td></td>
</tr>
<tr>
<td>TCS-13 (ComDE)</td>
<td>SMU.1916</td>
<td>genetic competence, biofilm formation, bacteriocin production</td>
</tr>
<tr>
<td>TCS-13 (ComDE)</td>
<td>SMU.1917</td>
<td></td>
</tr>
<tr>
<td>TCS-14</td>
<td>SMU.45</td>
<td>DNA damage recognition</td>
</tr>
<tr>
<td>TCS-14</td>
<td>SMU.46</td>
<td></td>
</tr>
<tr>
<td>GcrR</td>
<td>SMU. 1924</td>
<td>sucrose-dependent adherence, acid tolerance</td>
</tr>
</tbody>
</table>

### 1.1.2. DNA repair in S. mutans

In dental plaque, S. mutans must contend with numerous stresses such as acid (produced and secreted as a result of microbial metabolism), nutritional limitations and oxidative stress (byproducts of metabolism: hydrogen peroxide (H$_2$O$_2$), hydroxyl radical and superoxide) (Hamilton & Buckley, 1991, Carlsson, 1997, Hamilton & Svensäter, 1998). Another negative consequence of living in this environment is the accumulation of exogenous H$_2$O$_2$, produced by oral commensal bacteria including Streptococcus gordonii and Streptococcus sanguinis that are able to compete with
*S. mutans*, a catalase-negative organism (Kreth *et al.*, 2005). These surroundings encountered by *S. mutans* may induce damage to the cell membrane integrity, proteins and chromosomal DNA (Lemos & Burne, 2008). With regards to the latter, damage towards genomic DNA can lead to potential mutations and DNA strand breaks which are detrimental for DNA replication (Lindahl & Nyberg, 1972, Cooke *et al.*, 2003). As part of its ability to adapt to environmental stress, *S. mutans* has evolved multiple DNA repair mechanisms that protect and repair DNA damage from the harmful effects of the environment and enable it to survive and thrive as a pathogenic organism (Quivey Jr *et al.*, 2000). One of the most extensively studied proteins involved in DNA repair is RecA recombinase, which is highly conserved and widely distributed in all living organisms (Walker, 1984, Brendel *et al.*, 1997, Cox, 2007). RecA acts as a moderator of homologous recombination as a form of DNA repair that involves the mechanisms of broken fork repair, double-stranded DNA break repair and recombinational gap-filling repair; (Walker, 1984, Brendel *et al.*, 1997, Cox, 2007). Studies conducted in *S. mutans* demonstrated that a RecA-deficient mutant was found to have increased susceptibility to a killing pH of 2.5 relative to the parent strain (Quivey *et al.*, 1995). However, this sensitization to low pH, ultraviolet or H$_2$O$_2$ due to RecA deficiency could be alleviated when the mutant was allowed to develop an acid-adaptive response, suggesting the presence of a DNA repair system which functions independently of RecA (Quivey *et al.*, 1995). Further analyses demonstrated that this RecA-independent activity involved an apurinic/apyrimidinic (AP) endonuclease and this endonuclease activity was present at greater levels in cells grown at low pH than in cells grown at pH of 7.0 (Hahn *et al.*, 1999). In addition, it was also observed that *uvrA* and the nucleotide excision repair (NER) pathway were involved in the repair of acid-induced DNA damage and were associated with successful adaptation of *S. mutans* to low pH (Hanna *et al.*, 2001). A mutant deficient in *uvrA* revealed increased sensitivity to UV irradiation and higher sensitivity to killing at pH 3.0 after prior exposure to sublethal pH of 5.0. NER locates larger DNA lesions and induces dual incisions of the damaged DNA followed by DNA-repair enzymes with replication proteins that fill the resulting excised site and ligation (Sancar, 1996).
Additionally, minor DNA damage encountered by bacteria often leads to loss of purines and pyrimidines resulting in the formation of an abasic and AP site (Lindahl & Nyberg, 1972). This event involves protonation of the base followed by cleavage of the glycosyl bond (Lindahl & Nyberg, 1972). The repair of AP sites in duplex DNA is initiated by AP endonucleases that cleaves the phosphodiester bond located at 5’ or 3’ proximal to the AP site depending on the class of endonuclease involved (Lindahl & Nyberg, 1972, Hahn et al., 1999). Majority of AP endonucleases including exonucleases III and endonuclease IV act on 5’ or 3’ to the AP site; exonuclease III activity is accounted in 90% of abasic site repair and is induced as a stress-response mechanism (Lindahl, 1982). Studies examining DNA protection and repair in S. mutans demonstrated that under acid conditions, S. mutans can up-regulate the expression of an AP endonuclease that has a similar activity and regulation of exonuclease III of E. coli, therefore implicating its involvement in acid adaptation (Hahn et al., 1999).

1.1.3. Natural transformation in S. mutans

1.1.3.1. Brief overview of competence and regulatory mechanisms

S. mutans is one of the species that has been shown to be naturally transformable. Transformable species enter a transient physiological state of competence in which they bind and take up exogenous DNA as a result of death or lysis of other cells in the environment (Li et al., 2001, Li et al., 2002, Marsh, 2004). Unlike other mechanisms of horizontal gene transfer, such as transduction and conjugation, transformation is entirely directed by the competent cell and all the genes involved in genetic transformation are encoded in its genome. In many naturally transformable bacteria including S. mutans the competence activation occurs under certain conditions and is tightly regulated and triggered by specific inducing signals and detection systems.

In S. mutans, genetic competence is controlled by at least two regulatory systems: the ComDE two-component signal transduction system that responds to
competence-stimulating peptide (CSP) pheromone; and the ComR/S signaling pathway that responds to the SigX-inducing peptide (XIP), respectively. They ultimately modulate the transcription of an alternative sigma factor, SigX encoded by comX, and are dependent on nutrient conditions in the environment (Li et al., 2001, Li et al., 2002, Aspiras et al., 2004, Mashburn - Warren et al., 2010). When grown in a complex medium, S. mutans secretes CSP, and when it reaches a critical threshold, the peptide is detected by the ComD HK resulting in activation of its cognate RR, ComE, which in turn modulates the transcription of the target genes leading to competence activation in a subpopulation of cells that can then take up transforming DNA (Syvitski et al., 2007, Tian et al., 2009, Lemme et al., 2011). ComCDE has also been shown to directly activate transcription of several bacteriocins (Kreth et al., 2005, van der Ploeg, 2005, Yonezawa & Kuramitsu, 2005, Perry et al., 2009). Based on these findings it was proposed that the ComCDE system coordinates natural competence development and mutacin production as a means to acquire transforming DNA either by killing closely related streptococcal species in the vicinity of S. mutans, or through an altruistic suicide mechanism among a subpopulation of competent cells within the S. mutans community (Merritt & Qi, 2012). Recent studies identified ComRS as a second signaling system that is involved in the regulation of competence in S. mutans (Mashburn - Warren et al., 2010). When grown in a chemically-defined medium (CDM), S. mutans secretes XIP encoded by ComS (Desai et al., 2012, Khan et al., 2012, Son et al., 2012). Extracellular processing of the ComS pheromone is followed by its import into the cytoplasm via an oligopeptide permease transporter (OppA). The mature form of ComS (XIP) interacts with the ComR regulator to form a complex that ultimately increases the transcription of comX triggering competence activation (Desai et al., 2012, Khan et al., 2012, Son et al., 2012). ComR has been found to regulate competence downstream of ComE, suggesting that the ComRS signaling cascade has a proximal effect on S. mutans comX expression and competence (Mashburn - Warren et al., 2010).
1.1.3.2. DNA uptake and processing during transformation

Uptake of transforming DNA in Gram-positive bacteria is believed to rely on highly conserved DNA-transport machineries (Johnston et al., 2014). The current model for DNA uptake in streptococci is based on evidence obtained in S. pneumoniae, which is one of the species for which key aspects of transformation have been established. Exogenous double stranded (ds) DNA (both self and foreign) is the substrate for transformation where it may be integrated into the genome (Li et al., 2001). During the first step, dsDNA binds to the surface of competent cells and proteins of the DNA uptake and processing machinery allow DNA to cross the cell wall (Dubnau, 1999, Bergé et al., 2002, Chen et al., 2005). Traversing the cell wall involves the conversion of dsDNA (both chromosomal DNA and plasmids) into single stranded (ss) DNA during the translocation process using a sequence-independent DNA-uptake mechanism (Thomas & Nielsen, 2005). To take up DNA, the cells express pilin-like proteins (ComGC, ComGD, ComGE and ComGG) that allow the DNA to translocate inside the cytoplasm (Dubnau, 1999). Once across the cell wall, dsDNA binds to a specialized membrane-associated DNA binding protein ComEA. An unknown endonuclease is recruited to introduce single stranded breaks into the bound DNA (Mejean & Claverys, 1993). Endonuclease EndA is recruited to degrade one strand of the dsDNA molecule while the opposite strand will be integrated into the recipient’s genome in a 3’ to 5’ direction (Mejean & Claverys, 1993, Bergé et al., 2013). This reaction has been shown to be energy dependent, and to occur through the aqueous pore in the membrane formed by ComEC, ComFA and EndA (Dubnau, 1999). The recombination mediator protein DprA has been proposed to be needed to transport incoming ssDNA across the cytoplasmic membrane and to facilitate the recombinase RecA recruitment onto internalized ssDNA prior to the RecA mediated search for homology (Mortier-Barrière et al., 2007, Claverys et al., 2009). Finally ssDNA can be integrated into the genome (homology-dependent integration) mediated by the recombinase, RecA (Mortier-Barrière et al., 2007, Claverys et al., 2009). Homologous recombination and DNA-repair processes normally limit transformation to DNA from similar species. However, if a gene moves onto a broad-host-range plasmid it might be able to spread without the need for recombination.
Highly similar DNA-uptake and processing systems have also been identified in other related species. *S. gordonii* was found to be dependent upon the presence of a four-gene operon (*comYA-D*) whereas *Bacillus subtilis* rely on a seven-gene operon (*comGA-G*) (Pestova *et al.*, 1996, Lunsford & Roble, 1997). In *S. mutans*, a nine-gene operon (*comYA-I*) was found to facilitate DNA uptake and processing (Merritt *et al.*, 2005). The first seven genes (*comYA-G*) were shown to be essential for natural competence, while *comYH* and *comYI* had reduced and normal natural competence ability, respectively. Analysis of *comY* expression in different background mutations (*comC, ciaH* and *luxS*) revealed that these genetic factors have regulatory role on genetic competence (Merritt *et al.*, 2005). Despite the identification of the main components of the DNA uptake machinery in *S. mutans*, it remains to be elucidated how they interact with DNA and what steps are involved in DNA uptake, processing and recombination.

### 1.1.3.3. The fate of incoming DNA in the cytoplasm

Three non mutually exclusive models have been proposed to explain why bacteria take up DNA: 1) DNA for nutritional needs of the cell - used as a source of carbon, nitrogen and phosphorous; 2) DNA for repair- DNA from closely related species to provide DNA as a template for the repair of DNA damage; 3) DNA for genetic diversity- the acquisition of useful genes such as antibiotic resistance, virulence or metabolic functions.

#### 1.1.3.3.1. The pro and cons of DNA as a nutrient source

It was proposed that *E. coli* uptakes extracellular DNA to use it as source of carbon, and energy for a starving cell (Redfield, 1993, Finkel & Kolter, 2001, Palchevskiy & Finkel, 2009). Deletion of several *E. coli* genes identified as homologs of competence genes in *Haemophilus influenzae* and *Neisseria gonorrhoeae*, affected the ability of the strain to survive when competing with its wild-type parent strain for a nutrient resource, namely extracellular DNA (Finkel & Kolter, 2001, Palchevskiy & Finkel, 2006).The authors concluded that these genes were necessary for the use of extracellular DNA as the sole source of nutrient and energy (Finkel & Kolter, 2001,
Palchevskiy & Finkel, 2006). Further, several exonucleases (ExoI, ExoVII, ExoX and RecJ) in *E. coli* were found to process ssDNA as it enters the cytoplasm, the authors suggesting that the identified genes play a role in the long-term survival and the catabolism of DNA as a nutrient (Palchevskiy & Finkel, 2009). Evidence supporting a role for natural transformation as an alternative carbon-source was also presented in naturally competent bacteria. *Vibrio cholera* was found to promote genetic competence during times of nutrient starvation. In the absence of a preferred carbon source such as glucose, *V. cholera* increased the intracellular concentration of secondary messenger cyclic adenosine monophosphate (cAMP) and in complex with cAMP receptor protein (CRP) interacted with TfoX to activate multiple genes involved in natural transformation to internalize and metabolize alternative carbon sources (Blokesch, 2012, Scrudato & Blokesch, 2012). When pyrimidine levels were low, the nucleotide scavenging cytidine repressor CytR was activated resulting in upregulation of natural competence genes in *V. cholera* (Antonova et al., 2012).

Although DNA can serve as a nutrient source, multiple studies suggest that there are other roles for the DNA uptake machinery. The first example comes from *S. mutans* where comC regulon has been shown to be necessary for transformation in liquid cultures at very low cell density (OD$_{600}$~0.1) where essential nutrients are abundant and scavenging nucleotides from the environment is not necessary suggesting that the uptake of DNA for food is not the main goal of transformation (Li et al., 2001, Senadheera et al., 2009). Another example comes from *B. subtilis* that possesses a powerful nonspecific nuclease as well as uptake systems for the nucleolytic products (Dubnau, 1999). These DNA uptake systems seem to provide a much simpler mechanism for scavenging nucleotides from the environment (Dubnau, 1999). The presence of comEB which encodes a deoxy-CMP deaminase (involved in the salvage of dCMP) in a DNA-uptake operon in *B. subtilis* suggested that this enzyme may mediate utilization of DNA as nutrient (Redfield, 1993, Inamine & Dubnau, 1995). However, comEB in *Streptococcus pneumoniae* or *H. influenza* were not induced in competent cells arguing against a universal role for DNA as a nutrient source (Claverys et al., 2006). Similarly, digestion of unprotected ssDNA has been reported in *H. influenza* and *S. pneumoniae* (Barany et al., 1983, Méjean & Claverys, 1984). However, internalization of self-species DNA in *H. influenza* and *N.*
gonorrhoeae and induction of DNA release only from pneumococci or related species via fratricide in S. pneumoniae are arguing against DNA uptake for catabolism (Claverys et al., 2006, Claverys et al., 2007). The observations that DNA uptake is tightly regulated in S. pneumoniae and N. gonorrhoeae by species-specific DNA uptake sequences (DUS) in the Pasteurellaceae and Neisseriaceae (Aas et al., 2002, Redfield et al., 2006, Frye et al., 2013), strongly support a role of DNA internalization in chromosome repair or genome diversification rather than in nutrition (Johnston et al., 2014).

1.1.3.3.2. DNA for genome repair or genetic diversity

Active protection of internalized ssDNA from endogenous nucleases supports a role for incoming DNA in both repair and genetic diversity. DNA repair hypothesis postulates that transformed bacteria can use incoming DNA as a template for DNA repair (Mongold, 1992, Redfield, 1993). In agreement with this hypothesis, it was found that DNA damaging agents such as aminoglycoside and fluoroquinolone antibiotics as well as mitomycin C promote genetic transformation in S. pneumoniae (Prudhomme et al., 2006). Although there is substantial evidence that DNA repair machinery is induced during competence and uses transforming DNA for the recovery of damaged regions, there is also research to support a role for incoming DNA as a source of new and potentially beneficial genes to increase fitness. DNA for diversity hypothesis postulates that competent bacteria import DNA primarily for the acquisition of heterologous DNA to increase genome diversity. The first evidence in support of this hypothesis was provided in S. pneumoniae (Johnston et al., 2013). The DprnI pneumococcal restriction modification (R-M) system was found to possess a classic restriction enzyme and two methylases: one dsDNA methylase (DpnM) to degrade the newly integrated unmethylated DNA, and another methylase (DpnA) induced during competence to protect transforming foreign ssDNA from restriction. By protecting transforming foreign ssDNA from restriction, the competence-induced DpnA protein was found to be essential for the acquisition of pathogenicity islands (Johnston et al., 2013).
1.1.1.4. Competence and stress response

Bacteria can induce competence in response to environmental stresses that threaten cell survival. Sublethal concentrations of aminoglycoside and fluoroquinolone antibiotics have been shown to induce genetic transformability in *S. pneumoniae* and *Legionella pneumophila* (Prudhomme *et al.*, 2006, Charpentier *et al.*, 2012). Exposure to DNA damaging conditions (mitomycin C and UV radiation) also induced competence in both species (Prudhomme *et al.*, 2006, Charpentier *et al.*, 2012). Similarly, hydroxyurea known to generate stalled replication forks, was shown to induce competence in these species suggesting that there is a link between stress, DNA replication and competence (Charpentier *et al.*, 2012). Despite differences among competence regulatory cascades in these species, it is clear that they use these competence-inducing signals to provide clues to the conditions under which competence is favorable for the uptake of exogenous DNA from the environment (Johnston *et al.*, 2014).

In the case of *S. mutans*, there is also a link between the induction of competence and stressful conditions. The ComCDE competence pathway in *S. mutans* has been shown to be part of a global stress response that affects biofilm formation (Li *et al.*, 2001, Li *et al.*, 2002), bacteriocin production (Kreth *et al.*, 2005, van der Ploeg, 2005, Kreth *et al.*, 2006, Perry *et al.*, 2009), acid tolerance (Li *et al.*, 2001), and cell lysis (Qi *et al.*, 2005, Perry *et al.*, 2009), all of which are major virulence determinants that facilitate the growth, survival and persistence of *S. mutans*. Similarly, the newly identified ComRS, a quorum sensing system used by *S. mutans* for competence, has also been directly and indirectly linked to stress response, facilitating biofilm formation and cell lysis (Mashburn - Warren *et al.*, 2010, Dufour *et al.*, 2011, Lemme *et al.*, 2011, Wenderska *et al.*, 2012).

1.1.1.5. Limiting DNA import

The process of natural transformation has the potential to facilitate the acquisition of beneficial genes such as antibiotic resistance genes or fitness-enhancing genes contributing to rapid evolutionary changes by shaping bacterial genomes (Senadheera & Cvitkovitch, 2008, Perry *et al.*, 2009, Perry *et al.*, 2009). However,
natural transformation can also become costly depending on the composition of DNA fragments in the environment (Redfield, 1988, Moradigaravand & Engelstädter, 2013). Transformation with defective DNA that originates from bacteria that are selected against by natural selection can impede adaptation (Johnston et al., 2014). These circumstances most likely resulted in the evolution of certain cellular mechanisms limiting horizontal gene transfer (Johnston et al., 2014). DNA transformation is an energy-expensive process, with possible deleterious effects on genome integrity and recipient bacteria, including S. mutans have also evolved regulatory circuits that shut-off natural transformation (Tian et al., 2013, Dong et al., 2014). In streptococci, competence development is under the control of a master regulator (SigX) that binds to the promoters of late com genes encoding DNA uptake, protection homologous recombination machineries (Johnston et al., 2014). To avoid its activation under inappropriate conditions a strict control of the production of the master regulator is essential (Johnston et al., 2014). In S. pneumoniae, a ComCDE encoding organism, it was found that the abundance of SigX is regulated by the ClpEP and ClpCP proteolytic machineries, which degrade SigX (Piotrowski et al., 2009). Further, it was identified that DprA (a DNA protecting protein) in S. pneumoniae plays a dual role: in the processing of internalized ssDNA via interaction with RecA ensuring production of transformants as well as a crucial role in the shut-off of competence via interaction with ComE∼P to repress ComE-driven transcription, impacting SigX (ComX) production (Mirouze et al., 2013). Further, phylogenetic analyses revealed that the acquisition of this new function by DprA had affected its evolution in streptococci relying on ComE to regulate SigX expression suggesting that DprA has a similar role in competence shut-down in other streptococcal species (Mirouze et al., 2013). However, in the case of S. mutans, it was found that a different mechanism to regulate the cellular level and stability of SigX under inappropriate conditions. Adaptor protein MecA was found to target SigX for degradation by the protease ClpC/ClpP accelerating the escape of S. mutans cells grown in a complex medium from competence (Tian et al., 2013, Dong et al., 2014). Interestingly, MecA-regulated proteolysis of SigX appeared to be ineffective when S. mutans was grown in CDM medium suggesting the possibility that an unknown mechanism may control the cellular levels of SigX to secure the
timely escape of the cells from competence (Dong et al., 2014) (Tian et al., 2013). A similar mechanism involved in the regulation of the stability of SigX and genetic competence was also found in *S. thermophilus* (Wahl et al., 2014).

Recently, another distinct mechanism limiting DNA uptake has been identified in some bacteria and shown to protect microbial cells against uptake of exogenous DNA. It was discovered that naturally transformable *Neisseria* spp. can limit competence employing a sophisticated adaptive immune system that has been termed CRISPR/Cas (will be discussed in great detail in the following section) (Zhang et al., 2013). In short, the mechanism of CRISPR/cas activity was based on the integration of small fragments of invading DNAs into microbial genome, which were subsequently transcribed into short RNAs that directed the degradation of foreign invading DNA elements. In this way, the host organism acquired immunity toward mobile elements carrying matching sequences (Zhang et al., 2013). Similarly, it was shown that the native meningoccal CRISPR/Cas locus was able to prevent natural transformation of spacer-matched sequences, suggesting that it can limit the horizontal spread of virulence genes (Zhang et al., 2013). Artificial means of transformation (e.g. electroporation) was also shown to be blocked by CRISPR interference (Chen et al., 2005, Marraffini & Sontheimer, 2008, Deltcheva et al., 2011, Sapranauskas et al., 2011, Semenova et al., 2011), though natural transformation uses a very different DNA uptake process (Chen et al., 2005). Similarly, an engineered, heterologous CRISPR/Cas system introduced into *S. pneumoniae* was shown to block natural transformation during active infection in mice (Bikard et al., 2012). Thus it has been proposed that CRISPR/Cas systems which originally were shown to prevent phage infection (phage transduction) and conjugation, play a broader role in preventing also the third main route of HGT, natural transformation.
1.2. Overview of CRISPR-Cas immunity

1.2.1. CRISPR-Cas organization

To evade infective foreign DNA bacteria have evolved a plethora of defense systems. Among them, two systems act by specifically cleaving the incoming DNA as it enters the host cells, namely the restriction modification systems (R-M) and CRISPR-Cas systems. In R-M systems, a methyltransferase protects host DNA by modifying specific bases while a restriction endonuclease cleaves any foreign DNA which has not been protected by this modification (Tock & Dryden, 2005). On the other hand, CRISPR-Cas is a sequence-specific adaptive immunity system used by bacteria and archaea to destroy foreign DNA (Sorek et al., 2008, Deveau et al., 2010, Westra et al., 2010). The three major components of a CRISPR-Cas system are: cas genes (CRISPR-associated genes), a leader sequence and a CRISPR repeat-spacer array. CRISPRs (clustered regularly interspaced short palindromic repeats) contain short (30-50 nt) DNA repeats separated by unique spacers that sometimes originate from phages and plasmids (Horvath & Barrangou, 2010) (Figure 1-2).

Figure 1-2: Simplified diagram of a typical CRISPR –Cas system.

The three major components of a CRISPR locus are shown: cas genes (colored bars), a leader sequence (dark blue), and a repeat-spacer array. Repeats are shown as grey boxes and spacers are colored bars.

The CRISPR-Cas mechanism includes three main steps: adaptation (spacer acquisition), processing (maturation), and interference. In the adaptation step, a new repeat- spacer unit is added to the CRISPR locus, in which the spacer is acquired from a part of the invading DNA (named the proto-spacer). A CRISPR RNA (crRNA) transcript is then produced and cleaved at repeats by Cas protein(s), with or without
the help of other host proteins, to produce mature crRNAs. In the interference stage, the mature crRNAs guide Cas proteins to bind and cleave in a sequence-specific manner the invading DNA.

1.2.2. CRISPR distribution

Genomes may include one or multiple CRISPR loci. Approximately 50% of CRISPR-containing bacterial and archaenal species have been found to possess more than one locus (Godde & Bickerton, 2006). For example in Methanococcus jannaschii have been identified 20 CRISPR loci (Grissa et al., 2007, Horvath et al., 2008). The repeat sequences vary among CRISPR loci in the same genome however sometimes several CRISPR loci share the same repeat sequences but in these cases only one locus is associated with a set of cas genes (Grissa et al., 2007). Since CRISPRs with the same repeat sequences and the same set of cas genes have been identified in two distinct bacterial species, it has been suggested that CRISPR loci can be acquired through horizontal gene transfer (Chakraborty et al., 2010, Millen et al., 2012). This hypothesis was supported further by other observations including 1) GC content variability and codon bias between CRISPR locus and the rest of the chromosome, 2) CRISPR loci encoded by conjugative plasmids and 3) mobile genetic elements in the vicinity of CRISPR locus (Godde & Bickerton, 2006, Horvath et al., 2009, Portillo & Gonzalez, 2009).

1.2.3. Mechanism

CRISPR immune systems function in three stages: adaptation, maturation, and interference (Makarova et al., 2011). Only a brief overview of these distinct stages is presented here, however a detailed description will be covered later in this chapter.

During infection by a genetic element, new spacers with sequences identical to those of an invading genomic element are incorporated into the genomic CRISPR array of the challenged cell resulting in immunity against the invader (adaptation stage) (Barrangou et al., 2007, Garneau et al., 2010). The CRISPR locus is transcribed and small processed CRISPR RNAs (crRNA) carrying the spacer sequence form a complex with Cas proteins (maturation stage) (Brouns et al., 2008,
Carte et al., 2008, Hale et al., 2009, Haurwitz et al., 2010, Deltcheva et al., 2011, Wang et al., 2011). Finally, these ribonucleoprotein complexes target a complementary sequence in the invading nucleic acid and cleave within the matching sequence (interference stage) (Barrangou et al., 2007, Deveau et al., 2007, Brouns et al., 2008, Van Der Oost et al., 2009, Garneau et al., 2010). Though CRISPR immunity is divided into three stages according to the time course of the immune response, each stage has more than one step, each differing between types and organisms (Makarova et al., 2006, Wiedenheft et al., 2012) (Figure 1-3).

Figure 1-3: Stages of the CRISPR/Cas mediated immunity.

(a) Stage I: Adaptation. Entry of foreign DNA into a cell through transformation, conjugation, or transduction can lead to the acquisition of new spacer sequences
within the CRISPR locus of the host genome by the adaptation Cas proteins. The newly acquired spacer identical to the corresponding phage genomic sequence (proto-spacer) will provide immunity against the invader (discussed further in Figure 1-9). Nucleotide sequence flanking the proto-spacer named PAM (proto-spacer adjacent motif) also appears to be important for the acquisition of new spacers and immunity. (b) Stage II: Interference. CRISPR loci transcribed from a long precursor and subsequently processed into smaller units recruit the interference Cas complex for interference (see Figure 1-9 for details). A cell carrying a crRNA targeting a region (by perfect pairing) of foreign nucleic acid can target and destroy the invasive genetic element via an interference Cas complex. If there is no matching between the spacer and the proto-spacer, the CRISPR/Cas system is counteracted and the phage lytic cycle or plasmid replication can proceed. Figure modified from (Deveau et al., 2010).

1.3. History

CRISPR sequences were first observed in 1987 in the bacterium Escherichia coli K12 (Ishino et al., 1987). In 2002, they were observed in more bacterial species and in archaea (Jansen et al., 2002). In 2005, it was discovered that CRISPR spacers had homology to phage and plasmid DNA and this was an indication that CRISPR sequences could be part of an immune system in bacteria (Bolotin et al., 2005, Mojica et al., 2005, Pourcel et al., 2005). The first experimental evidence of acquired immunity mediated by CRISPR/Cas to prokaryotes was provided in 2007; integration of a short phage-specific sequence conferred resistance of bacteria to the phage with spacer DNA. Resistance to the phage was altered by any mismatch between the CRISPR spacer and the target phage sequence (Barrangou et al., 2007). In 2010, Cas proteins were found to be necessary for CRISPR-Cas mediated immunity against invasive DNA. Disabling some cas genes disrupted the cell’s ability to incorporate new spacers or caused a loss of resistance against phages, even when the relevant spacers were present (Horvath & Barrangou, 2010). Since their discovery, a number of CRISPR-Cas systems have been found in various bacteria.
They have been classified into three main types (Types I-III) each containing different subtypes based on the set of cas genes and their mode of action (Makarova et al., 2011). Due to their valuable feature (to be able to target specific DNA sequences), these systems have been the subject of intense investigation. In 2012, in a study conducted by two research teams, CRISPR2-Cas system from *Streptococcus pyogenes* that uses Cas9 as a nuclease specialized for cutting dsDNA was found to respond to an invading phage by transcribing a CRISPR spacer-repeat array into a long RNA molecule that is next processed by trans-activating crRNA (tracrRNA) and Cas9 into smaller pieces of crRNAs (Jinek et al., 2012). Cas9 has two nuclease active sites cleaving different DNA strands of the double helix. By disabling one or both sites Cas9 preserved its ability to find its target DNA (Deltcheva et al., 2011, Jinek et al., 2012, Wiedenheft et al., 2012). Soon after that discovery, it was shown that Cas9 combined with tracrRNA and spacer RNA into a single-guide RNA molecule can find and cut the correct DNA targets leading to the use of CRISPR/Cas9 as a genome engineering/editing tool (Jinek et al., 2012, Cong et al., 2013, Jiang et al., 2013). Various groups have used it to delete, add specific genes in various organisms from bacteria, fungi, plants, nematodes, animals and human cells demonstrating broad utility for the technique (Jinek et al., 2012, DiCarlo et al., 2013, Jiang et al., 2013, Wang et al., 2013). This system was recently shown to be programmable to also target and activate or suppress targeted genes (Larson et al., 2013, Yin et al., 2013). Last year there was evidence that CRISPR/Cas technology can reverse disease symptoms of a rare liver disorder in mice (Yin et al., 2014).

1.4. **Structure of the CRISPR ‐ Cas locus**

CRISPR loci consist of a series of short repeat sequences separated by unique spacer sequences of similar length (Grissa et al., 2007) (Figure 1-4).
There are four types (I-III and U) which have different architecture, cas gene content, their specific target and mechanism of action. Major components of a CRISPR locus are cas genes (colored bars), a leader sequence (dark blue), and a repeat-spacer array. Repeats are shown as grey boxes and spacers are colored bars. Arrows represent the direction of transcription. While most CRISPR loci contain each of the three components, the arrangement is not always as shown here.

1.4.1. Repeats

The repeat sequences (21 to 48 bp) are typically identical in length and sequence but often the repeat at the end of an array deviates from the consensus sequence (Mojica et al., 2000, Lillestø et al., 2006, Grissa et al., 2007, Horvath et al., 2008). Repeat sequences are highly conserved within a species, but between species they can vary greatly (Grissa et al., 2007). Many repeats are partially palindromic, allowing the two halves of the sequence to base pair. Based on this, after these sequences are transcribed, they are able to form hairpin-like secondary structures.
(Horvath et al., 2008). However, there are repeats, which are not palindromic and are predicted to be unstructured (Kunin et al., 2007). The difference in the ability to form secondary structures has implications for the mechanism of processing these arrays into short crRNAs (Kunin et al., 2007, Makarova et al., 2011).

1.4.2. Spacers

In contrast, the spacer sequences are usually unique within the genome. Spacer length can vary between 21 and 72 bp, but usually the size is similar to that of the repeats in the same array (Grissa et al., 2007). Many spacers originate from extrachromosomal DNA sources illustrating the genetic memory of previous infections (Mojica et al., 2005, Pourcel et al., 2005). This captured foreign DNA segments may correspond to sequences of coding or non-coding DNA from phage, plasmid, or chromosomal sequences (Bolotin et al., 2005, Mojica et al., 2005, Pourcel et al., 2005, Brodt et al., 2011). Some spacers match sequences elsewhere within their own genome and their function has been proposed that some CRISPR systems might be involved in gene regulation. The existence of such self-targeting spacers is surprising considering that self-targeting is expected to lead to cell death (Bikard et al., 2012, Jiang et al., 2013, Paez-Espino et al., 2013). Since their function is not yet resolved, another possibility is that some genome sequences might be accidentally incorporated (Stern et al., 2010). CRISPR loci are primarily located on the genome, but can also be found on plasmids (Horvath & Barrangou, 2010).

1.4.3. Leader

The third component of the CRISPR array is the leader sequence, which is located at the proximal edge of the first repeat (Mojica et al., 2000, Jansen et al., 2002). This noncoding, AT-rich sequence is typically located at the 5’ flank of CRISPR array and depending on the species ranges between 20-534 bp long (Jansen et al., 2002, Lillestø et al., 2006). Leader sequences preceding CRISPR loci within the same genome are 80% identical but can vary among species (Bult et al., 1996). The role of leader region has been proposed to involve insertion of a repeat-spacer unit at the leader proximal edge of the repeat cassette (Lillestø et al., 2006, Brouns et al., 2008, Pougach et al., 2010, Pul et al., 2010, Westra et al., 2010, Wurtzel et al., 2010).
Novel spacers are integrated into loci from the leader end between the leader and the first spacer while older spacers are localized to the distal end. The idea that insertion of new spacers occurs from the leader end is supported by the observation that repeat sequences in this region are highly homologous. It was determined that degenerative repeat sequences are found at the distal end and are associated with conserved spacers that are generally inactive while less degenerate repeats are found closer to the leader region and are associated with spacers that are usually specific for a strain (Díez-Villaseñor et al., 2010). CRISPR loci appear to be transcribed from the leader region that functions as a promoter region (Lillestø et al., 2006, Brouns et al., 2008, Marraffini & Sontheimer, 2008, Hale et al., 2009). This hypothesis was subsequently supported by experiments showing that CRISPR loci lacking a leader region do not incorporate new spacers (Lillestø et al., 2006).

### 1.4.4. Cas proteins

Several CRISPR subtypes exist and each subtype has its own particular set of conserved genes referred to as CRISPR-associated (Cas) genes (Mojica et al., 2005, Kunin et al., 2007). These cas genes are usually found adjacent to CRISPR sequences, often in the form of operons. Importantly, these genes are absent from genomes lacking CRISPR and this physical link between arrays and cas genes led to the proposal that they are co-functional. It was also noted that when multiple CRISPR arrays with the same repeat sequence are present in a genome, cas genes are associated with only one of them, whereas when multiple CRISPR arrays with different repeat sequences are present, each CRISPR locus has its own set of cas genes (Jansen et al., 2002). Types (I-III) of CRISPR-Cas systems include two universal cas genes: cas1 and cas2 (Makarova et al., 2006, Makarova et al., 2011). Otherwise, however all these types differ in their sets of constituent genes and each is characterized by a unique signature gene. Within the three types of CRISPR-Cas systems have been identified signature genes including cas3 for the type I cas9 for the type II and cas10 for the type III (Table 1-2).
Table 1-2: Major Cas proteins (Cas1-Cas10).

(adapted from (Barrangou & Van der Oost, 2012)).

<table>
<thead>
<tr>
<th>Family</th>
<th>Distribution</th>
<th>Characteristic(s)</th>
<th>Function (proposed or demonstrated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas1</td>
<td>universal</td>
<td>Metal-dependent deoxyribonuclease</td>
<td>Integration of spacer into CRISPR repeat</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensitivity to DNA damage</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Impaired chromosomal segregation</td>
<td></td>
</tr>
<tr>
<td>Cas2</td>
<td>universal</td>
<td>RNase or DNase</td>
<td>Spacer selection and integration</td>
</tr>
<tr>
<td>Cas3</td>
<td>Type I</td>
<td>ssDNase; interference</td>
<td>Target DNA cleavage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATP-dependent helicase</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metal-dependent dsDNase</td>
<td></td>
</tr>
<tr>
<td>Cas4</td>
<td>Type I, II</td>
<td>RecB like nuclease</td>
<td>Spacer acquisition</td>
</tr>
<tr>
<td>Cas5</td>
<td>Type I</td>
<td>Metal-independent endoribonuclease</td>
<td>crRNA expression and interference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Generates crRNAs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cascade complex to cleave target DNA</td>
<td></td>
</tr>
<tr>
<td>Cas6</td>
<td>Type I, III</td>
<td>Metal-independent endoribonuclease</td>
<td>crRNA expression and interference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to generate crRNAs</td>
<td></td>
</tr>
<tr>
<td>Cas7</td>
<td>Type I</td>
<td>Cascade complex</td>
<td>Interference</td>
</tr>
<tr>
<td>Cas8</td>
<td>Type I</td>
<td>Cascade complex, PAM recognition</td>
<td>Interference and spacer selection</td>
</tr>
<tr>
<td>Cas9</td>
<td>Type II</td>
<td>Generate crRNA, cleave target DNA</td>
<td>Interference</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>-----------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Cas10</td>
<td>Type III</td>
<td>Cascade complex</td>
<td>Interference and spacer selection</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Strand separation</td>
</tr>
</tbody>
</table>

All these CRISPR/Cas systems include the core genes, *cas1* and *cas2* as well as at least one of a set of type-specific genes (Horvath & Barrangou, 2010). CRISPR loci typically consist of fewer than 50 repeat-spacer alterations, although some include several hundred (Horvath & Barrangou, 2010). Genomes may include one or several different CRISPR loci. In extreme cases, these loci can constitute over 1% of the genome. CRISPR loci are found in roughly 40% of bacterial genomes and 90% of archaea genomes. It is currently unknown why they are more common in archaea. Moreover, CRISPR sequences can be transmitted both horizontally and vertically (Horvath & Barrangou, 2010).

### 1.4.4.1. Cas1 and Cas2

Cas1 and Cas2 are the only proteins conserved among all types of CRISPR systems (Makarova *et al.*, 2011). Due to this valuable feature, these proteins are widely used as a marker for detection of CRISPR-Cas systems in bacterial genomes and for construction of phylogenetic trees. Genetic experiments as well as spacer acquisition assays demonstrated that Cas1 and Cas2 are required for the acquisition of new spacer sequences into the CRISPR locus (Datsenko *et al.*, 2012, Yosef *et al.*, 2012). Based on biochemical studies identifying Cas1 and Cas2 as metal-dependent nucleases (Beloglazova *et al.*, 2008, Wiedenheft *et al.*, 2009, Babu *et al.*, 2010, Nam *et al.*, 2012, Kim *et al.*, 2013) and crystal structure studies indicating that Cas1 assembles with Cas2 in a complex, it was proposed that Cas1-Cas2 integration complex has a marked preference for binding to the leader sequence within CRISPR locus, which serves as the target site for spacer integration (Nuñez *et al.*, 2014).
1.4.4.2. Major families of RAMPs

The Cas proteins known as RAMP (Repeat Associated Mysterious Proteins) form a distinct class of Cas proteins and are present in both type I and type III. These Cas proteins have been shown to exhibit weak conservation to each other but share glycine-rich regions (G-loop) and contain distinct forms of the RNA recognition motif (RRM) domain (Haft et al., 2005, Makarova et al., 2006). Based on all RAMP crystal structures, secondary structure predictions, and sequence comparisons, the RAMP family was classified into three major families: Cas5, Cas6 and Cas7 family (Makarova et al., 2011). The Cas5 family includes Csm4, Csx10, Cmr3, Cas5, Csy2, Csc1. Cas5 proteins are classified in two subfamilies, one that contains two RRM domains and the other with only one RRM domain (Barrangou & Van der Oost, 2012). The Cas6 family consists of Cas6, Cas6e and Cas6f. They all contain two RRM domains and a glycine-rich loop upstream of the second RRM domain (Barrangou & Van der Oost, 2012). The Cas7 group includes Cas7, Csy3, Csm3, Csm5, Cmr1, Cmr4, Cmr6, Csf2 and Csc2. They all have a single RRM domain (Barrangou & Van der Oost, 2012). RAMP proteins exhibit a common pattern of organization on the operons of Type I and Type III CRISPR systems. While Type I system have a single copy of the Cas5, Cas6 and Cas7, type III CRISPR systems typically have multiple Cas7 copies. Cas5 and Cas7 have been found to form Cascade complex (Van Der Oost et al., 2009, Jore et al., 2011, Lintner et al., 2011).

1.4.4.3. Cascade associated proteins

Both type I and III systems involve a CRISPR-specific endoribonuclease (Cas6 family of endoribonucleases) to process the long primary transcript CRISPR RNA into short crRNAs (Carte et al., 2008, Hale et al., 2009, Nam et al., 2012). DNA target recognition in both type systems is carried out by a multi-subunit ribonucleoprotein complex (Brouns et al., 2008). This complex was named Cascade in the type I-E from *E. coli* (CRISPR-associated complex for antiviral defense, type I-E) (Brouns et al., 2008, Jore et al., 2011), Csy complex in the type I-F system from *P. aeruginosa* (Wiedenheft et al., 2011), Cascade-like interference complex in the type I-C system from *B. halodurans* (Nam et al., 2012). For the type III systems Cascade complexes were called CSM in the type III-A system from *S. solfataricus*.
(Lintner et al., 2011, Zhang et al., 2012) or CMR in the type III-B complex from *Pyrococcus furiosus* (Hale et al., 2009). Cascade complexes of Type I systems typically include multiple subunits of Cas7, Cas5, Cas6, large subunits (Cse1/CasA/Csd1) and small subunits (Cse2/CasB)/Csd2) if present (Barrangou & Van der Oost, 2012). *E. coli* Cascade complex was reported to be composed of five proteins (11 subunits): one Cse1 (CasA), two Cse2 (CasB), six Cas7 (CasC), one Cas5 (CasD), one Cas6e (CasE) and the crRNA (Brouns et al., 2008, Jore et al., 2011). In subtype I-C system from *B. halodurans*, two subunits of Cas5d protein were found to assemble to mature crRNA one Csd1 and six Csd2 Cas proteins to form a multi-subunit interference complex similar to *E. coli* cascade (Nam et al., 2012). The *P. aeruginosa* Csy ribonucleoprotein complex was found to be composed of a set of distinct Csy1-4 proteins that resemble the stoichiometry and the morphology of the Cascade complex from *E. coli* (Wiedenheft et al., 2011). CSM complex of the Type III-A of *Thermus thermophiles* is composed of five different protein subunits (Csm1–Csm5) with an uneven stoichiometry and a single crRNA (Staals et al., 2014). For the subtype III-B system, the Cascade complex contains subunits of the CMR complex (Cmr1, Cmr3, Cmr4, Cmr6) (Zhang et al., 2012). The Cascade complex bound to crRNA is recruited in the interference stage by promoting R-loop formation (formation of a RNA-DNA hybrid) to match a spacer within crRNA to the target ssDNA, leading to DNA degradation and invader neutralization (Jore et al., 2011).

1.4.5. PAM

Proto-spacers are sequences present in phage or plasmid genomes that are excised by Cas1-Cas2 proteins and inserted into CRISPR loci as new spacers. A short motif is often present in these genomes consisting of 2-5 bp immediately downstream (or upstream) of the proto-spacer sequence. These motifs are called proto-spacer adjacent motifs (PAMs). PAMs were first identified in sequence alignments of putative protospacers of phages that matched CRISPR spacers (Bolotin et al., 2005). Subsequently, other motifs were identified for a variety of organisms and their CRISPR systems (Deveau et al., 2007, Horvath et al., 2008, Lillestøl et al., 2009, Mojica et al., 2009, Semenova et al., 2011, Shah & Garrett, 2011). It was also shown
that the conservation of PAM is a common theme for the most diverse CRISPR-Cas
types (Mojica et al., 2009). Similar CRISPR variants shared identical PAM
sequences, suggesting that there is a CRISPR-motif directed choice of the spacers,
which subsequently are implicated in the interference of the target (Mojica et al.,
2009). PAMs also appear to direct the orientation of spacers in the CRISPR array
(Mojica et al., 2009).

1.5. CRISPR–Cas classification

CRISPR–Cas systems have been recently reclassified into types I, II and III based
primarily on their cas composition, the organization of the cas operon and the
structure of repeats in the CRISPR arrays (Makarova et al., 2011). Detailed
comparative analysis of sequences and structures led to delineation of
approximately 30 Cas protein families that were further unified into distinct major
superfamilies (Makarova et al., 2011). All these distinct CRISPR types contain two
universal proteins: Cas1, a metal-dependent DNase involved in the integration of
alien DNA into CRISPR array and Cas2, a metal-dependent endoribonuclease that
also has been shown to be involved in spacer integration during the adaptation stage
(Beloglazova et al., 2008, Wiedenheft et al., 2009, Marraffini & Sontheimer, 2010).
The signature proteins for the three types are Cas3 (a helicase), Cas9 (Ruv-like and
HNH nuclease domains) and Cas10 (Palm domain of polymerases) (Makarova et al.,
2011). These three types of CRISPR systems can be classified into subtypes based
on the distinct signature genes and phylogeny of Cas1 (Makarova et al., 2011). The
Cas proteins known as RAMP proteins are present in multiple copies in both type I
and III and have multiple functional roles in endoribonucleolitic cleavage and
ribonucleoprotein assembly (Brouns et al., 2008, Carte et al., 2008, Haurwitz et al.,
2010).

1.5.1. Type I CRISPR–Cas systems

Type I CRISPR/Cas system is the most wide-spread and is found in both bacteria
and archaea. It is comprised of six different subtypes (subtype I-A to I-F) based on
the presence of distinct subtype-specific genes (Makarova et al., 2011). Cas3 is the signature protein of the type I system and is responsible for cleaving and degrading target DNA (Figure 1-5).

![Figure 1-5: Simplified diagram of a CRISPR locus in type I.](image)

It consists of CRISPR locus, core cas genes (cas1, cas2, cas3 and cas4), subtype-specific genes (cas5, cas6, cas7).

cas3 encodes a large protein with two domains: a helicase that unwinds dsDNA (helicase domain) and a HD phosphohydrolase that cleaves ssDNA (HD nuclease domain) (Beloglazova et al., 2011, Makarova et al., 2011, Mulepati & Bailey, 2011). cas3 has been shown to interact with genes encoding proteins that processes pre-crRNA and assembles into a cascade-like interference complex with different compositions (Brouns et al., 2008, Jore et al., 2011). These cascade complexes may contain proteins within the RAMP superfamily such as Cas5 (Cas5d), Cas6 (Cas6e and Cas6f), and Cas7 as well as Cse1, Cse2 or other less conserved subunits (Brouns et al., 2008, Lintner et al., 2011, Mulepati et al., 2012, Nam et al., 2012, Plagens et al., 2012). The long precursor CRISPR RNA (pre-crRNA) is processed into mature crRNA by one of the RAMP proteins with RNase. For some CRISPR-Cas subtypes the cRNA is further processed at the 3' end. The cascade complex then invades, pairs with the complementary dsDNA and recruits Cas3 (a nuclease-helicase) to cleave the target (29, 30) (Table 1-3 adapted from (Makarova et al., 2011)).
Table 1-3: Cas proteins associated with type I system.

<table>
<thead>
<tr>
<th>Name</th>
<th>System type</th>
<th>Structure (PDB code)</th>
<th>Family of encoded protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas 1</td>
<td>Type I</td>
<td>3GOD, 3LFX, 2YZS</td>
<td>COG1518</td>
</tr>
<tr>
<td>Cas 2</td>
<td>Type I</td>
<td>2IVY, 218E, 3EXC</td>
<td>COG1343; COG3512</td>
</tr>
<tr>
<td>Cas 3</td>
<td>Type I</td>
<td>N/A</td>
<td>COG1203</td>
</tr>
<tr>
<td>Cas 3</td>
<td>Subtype I-A</td>
<td>N/A</td>
<td>COG2254</td>
</tr>
<tr>
<td></td>
<td>Subtype I-B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cas 4</td>
<td>Subtype I-A</td>
<td>N/A</td>
<td>COG1468</td>
</tr>
<tr>
<td></td>
<td>Subtype I-B</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtype I-C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtype I-E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cas 5</td>
<td>Subtype I-A</td>
<td>3KG4</td>
<td>RAMP</td>
</tr>
<tr>
<td></td>
<td>Subtype I-B</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtype I-C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtype I-D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cas 6</td>
<td>Subtype I-A</td>
<td>3I4H</td>
<td>COG1583; RAMP</td>
</tr>
<tr>
<td></td>
<td>Subtype I-B</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtype I-D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cas6e</td>
<td>Subtype I-E</td>
<td>1WJ9</td>
<td>RAMP</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Cas6f</td>
<td>Subtype I-F</td>
<td>2XLJ</td>
<td>RAMP</td>
</tr>
<tr>
<td>Cas7</td>
<td>Subtype I-A</td>
<td>N/A</td>
<td>COG1857; RAMP</td>
</tr>
<tr>
<td></td>
<td>Subtype I-B</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtype I-C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtype I-E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cas8a1</td>
<td>Subtype I-A</td>
<td>N/A</td>
<td>BH0338-like</td>
</tr>
<tr>
<td>Cas8a2</td>
<td>Subtype I-A</td>
<td>N/A</td>
<td>PH0918</td>
</tr>
<tr>
<td>Cas8b</td>
<td>Subtype I-B</td>
<td>N/A</td>
<td>BH0338-like</td>
</tr>
<tr>
<td>Cas8c</td>
<td>Subtype I-C</td>
<td>N/A</td>
<td>BH0338-like</td>
</tr>
<tr>
<td>Cas10d</td>
<td>Subtype I-D</td>
<td>N/A</td>
<td>COG1353</td>
</tr>
<tr>
<td>Csy1</td>
<td>Subtype I-F</td>
<td>N/A</td>
<td>Y1724-like</td>
</tr>
<tr>
<td>Csy2</td>
<td>Subtype I-F</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Csy3</td>
<td>Subtype I-F</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Cse1</td>
<td>Subtype I-E</td>
<td>N/A</td>
<td>YgcL-like</td>
</tr>
<tr>
<td>Cse2</td>
<td>Subtype I-E</td>
<td>2ZCA</td>
<td>YgcL-like</td>
</tr>
<tr>
<td>Csc1</td>
<td>Subtype I-D</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Csc2</td>
<td>Subtype I-D</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Csa5</td>
<td>Subtype I-A</td>
<td>N/A</td>
<td>AF1870</td>
</tr>
<tr>
<td>Csb1</td>
<td>Subtype I-U</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Csb2</td>
<td>Subtype I-U</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Csb3</td>
<td>Subtype I-U</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Csx17</td>
<td>Subtype I-U</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Csx14</td>
<td>Subtype I-U</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Csx10</td>
<td>Subtype I-U</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
</tbody>
</table>
1.5.2. Type II CRISPR–Cas systems

The simplest among the CRISPR-Cas systems is the type II CRISPR/Cas system (subtype II-A and II-B). These systems found in bacteria but not archaea, consist of a minimal set of four cas genes (cas1, cas2, cas9 and either csn2 (subtype II-A) or cas4 (subtype II-B) (Makarova et al., 2011) (Figure 1-6).

![Figure 1-6: Simplified diagram of a CRISPR locus in type II.](image)

It consists of CRISPR locus and four cas genes (cas1, cas2, cas9 and csn2 (subtype-II-A) or cas4 (subtype-II-B).

In type II systems, the large multifunctional protein, Cas9 is the signature protein required for processing pre-crRNA into crRNA and cleaving the target, whereas the other proteins are thought to be involved in the acquisition of new CRISPR spacers (Sapranaukas et al., 2011, Gasiunas et al., 2012, Jinek et al., 2012). Cas9 was found to contain a RuvC-like nuclease domain and HNH (McrA-like) nuclease domain (Kleanthous et al., 1999, Jakubauskas et al., 2007) and this observation was an indication that this protein is likely to be responsible for target cleavage. In S. thermophilus, inactivation of Cas9 abolished the interference function against phage DNA and plasmid DNA (Barrangou et al., 2007, Garneau et al., 2010). In the type II systems, Cas9 uses the host enzyme RNase III (encoded by the rnc gene) and a noncoding RNA called trans-activating RNA (tracrRNA) to process the precursor pre-crRNA into mature cr-RNA with 20 nt spacer-derived 5’-tags and 19-22 nt repeat-derived 3’-tags (Deltcheva et al., 2011, Jinek et al., 2012). tracrRNA is encoded in the vicinity of CRISPR loci and contains a 25 nt long stretch that is complementary to the crRNA repeat sequence. Comparison of several tracrRNA molecules showed no conserved sequence, localization within type II CRISPR-Cas loci or structure elements but revealed a conserved feature to base pair to cognate pre-crRNA repeats (Chylinski et al., 2013). Following co-processing of the tracrRNA and pre-crRNA by RNase III, duplex tracrRNA:crRNA guides Cas9 to cleave site specifically.
cognate target DNA. DNA targeting in both type I and type II systems require beside crRNA/target complementarity and also a PAM, located immediately upstream/downstream of the protospacer (Deveau et al., 2007, Horvath et al., 2008, Mojica et al., 2009, Semenova et al., 2011, Sashital et al., 2012). The precise DNA cleavage site was identified 3 nt upstream of the PAM for the complementary strand, whereas the non-complementary DNA strand was cleaved at additional sites within three to eight base pairs upstream of the PAM producing blunt-ended cleavage products (Gasiunas et al., 2012, Jinek et al., 2012) (Table 1-4 adapted from Makarova et al., 2011)).

Table 1-4: Cas proteins associated with type II system.

<table>
<thead>
<tr>
<th>Name</th>
<th>System type</th>
<th>Structure (PDB code)</th>
<th>Family of encoded protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas 1</td>
<td>Type II</td>
<td>3GOD, 3LFX, 2YZS</td>
<td>COG1518</td>
</tr>
<tr>
<td>Cas 2</td>
<td>Type II</td>
<td>2IVY, 218E, 3EXC</td>
<td>COG1343, COG3512</td>
</tr>
<tr>
<td>Cas 4</td>
<td>Subtype II-B</td>
<td>N/A</td>
<td>COG1468</td>
</tr>
<tr>
<td>Cas9</td>
<td>Type II</td>
<td>N/A</td>
<td>COG3553</td>
</tr>
<tr>
<td>Csn2</td>
<td>Subtype II-A</td>
<td>N/A</td>
<td>Spy1049-like</td>
</tr>
</tbody>
</table>

1.5.3. Type III CRISPR–Cas systems

Type III CRISPR-Cas systems (subtype III-A and III-B) constitute nearly a quarter of all known CRISPR systems and are found in both archaea and bacteria but are predominantly found in archaeal genomes. Some subtype III-B systems can be
found in combination with one or more other subtype (Makarova et al., 2011, Makarova et al., 2011). The hallmark of type III system is the presence of the signature protein Cas10 (Figure 1-7).

Figure 1-7: Simplified diagram of a CRISPR locus in type III.

It consists of an array of spacers (blue and green boxes) and direct repeats (grey boxes) and a set of cas genes universal genes (cas1 and cas2), signature gene cas10 and RAMP genes (csm/cmr).

When type III systems are found to lack Cas1 and Cas2 proteins they co-exist with type I systems within the same genome allowing them to borrow these two proteins from the latter (Makarova et al., 2011). In type III systems, Cas6 nuclease is responsible as a stand-alone nuclease for the primary processing of the pre-crRNA into crRNAs that is further transferred to the cascade complex (Csm in subtype III-A and Cmr in subtype III-B systems). After transfer of the crRNAs to these complexes, the crRNA undergoes further processing at the 3’ end (Hale et al., 2009, Zhang et al., 2012). In type III-A system of Staphylococcus epidermis, the crRNA guides the cascade complex (Cas-Csm) to target DNA for cleavage with the involvement of Cas10-like protein Csm1 (Marraffini & Sontheimer, 2008). DNA targeting by this system seems to lack the PAM based system, but it uses a self inactivation strategy that involves base pairing between the 5’ repeat fragment of the mature crRNA and the repeat sequence in the CRISPR locus on the host chromosome which trigger the switching off of the interference process (Marraffini & Sontheimer, 2010).

In the type III-B system of Pyrococcus furiosus, Cas 6 protein is not part of the complex but a 5’ repeat fragment of the crRNA serves as an anchor for the assembly of the cascade complex (Cmr1-Cmr6) to cleave RNA. Similarly, Cmr complex (Cmr1–Cmr7) of Sulfolobus solfataricus targets invading RNA however the reported catalytic mechanism differs substantially from that which has been described for the
P. furiosus [48]. RNA targeting by these two interference complexes is also PAM-independent (Hale et al., 2009, Deng et al., 2013, Osawa et al., 2013, Shao & Li, 2013). Unexpectedly, the Cmr complex (Cmr1-Cmr6) of Sulfolobus islandicus has been reported to target plasmid DNA in a PAM-independent manner using a Csx1-dependent mechanism that requires the transcription of its target (Deng et al., 2013). These studies suggest that there is further mechanistic and/or functional diversity among type III systems (Deng et al., 2013) (Table 1-5 adapted from Makarova et al., 2011).
Table 1-5: Cas proteins associated with type III system.

<table>
<thead>
<tr>
<th>Name</th>
<th>System type</th>
<th>Structure (PDB code)</th>
<th>Family of encoded protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas 1</td>
<td>Type III</td>
<td>3GOD, 3LFX, 2YZS</td>
<td>COG1518</td>
</tr>
<tr>
<td>Cas 2</td>
<td>Type III</td>
<td>2IVY, 218E, 3EXC</td>
<td>COG1343, COG3512</td>
</tr>
<tr>
<td>Cas 6</td>
<td>Subtype III-A</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td></td>
<td>Subtype III-B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cas10</td>
<td>Type III</td>
<td>N/A</td>
<td>COG1353</td>
</tr>
<tr>
<td>Csm2</td>
<td>Subtype III-A</td>
<td>N/A</td>
<td>COG1421</td>
</tr>
<tr>
<td>Csm3</td>
<td>Subtype III-B</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Csm4</td>
<td>Subtype III-A</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Csm6</td>
<td>Subtype III-A</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Cmr1</td>
<td>Subtype III-B</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Cmr3</td>
<td>Subtype III-B</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Cmr4</td>
<td>Subtype III-B</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Cmr5</td>
<td>Subtype III-B</td>
<td>2ZOP; 2OEB</td>
<td>COG3337</td>
</tr>
<tr>
<td>Cmr6</td>
<td>Subtype III-B</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
<tr>
<td>Csx16</td>
<td>Subtype III-U</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>CsaX</td>
<td>Subtype III-U</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Csx3</td>
<td>Subtype III-U</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Csx1</td>
<td>Subtype III-U</td>
<td>N/A</td>
<td>RAMP</td>
</tr>
</tbody>
</table>
1.5.4. Unclassified CRISPR–Cas systems

For the CRISPR/Cas loci that could not be classified into one of the three distinct type of systems (I, II, III) according to the presence of type specific and subtype specific signature genes, they have been re-named type U (Figure 1-8). One example of unclassified CRISPR-Cas system that could not be classified at the type level has been found in Acidithiobacillus ferrooxidans (Makarova et al., 2011).

Figure 1-8: Simplified diagram of a CRISPR locus in type U.

It consists of a CRISPR array and a set of cas genes.

1.6. CRISPR–Cas immunity

1.6.1. Adaptation

CRISPR–Cas systems act in three stages: adaptation, maturation and interference.

Insertion of new spacers within a CRISPR array, a process known as adaptation, occurs during infection by invaders such as plasmids or phages (Datsenko et al., 2012, Swarts et al., 2012). The recognition and fragmentation of invading DNA is likely the first step in the process. For type I and type II CRISPR–Cas systems, but not type III systems, suitable proto-spacers from the invading DNA are selected by the detection of a specific PAM sequence and processed into pre-spacers of defined size (Swarts et al., 2012). Recently, it has been proposed that in addition to the participation of Cas 1 and Cas 2, there are additional Cas proteins such as Csn2, Cas4, Csa1 and Cas3 required for spacer uptake (Yosef et al., 2012). The integration machinery likely carries the pre-spacer and recognizes the leader strand to incorporate it into the CRISPR locus to form spacers. After the opening of the leader-end repeat by the nicking of both strands, the new spacer is integrated in a specific, PAM-dependent orientation between the leader-proximal repeat and the newly synthesized repeat, so that spacers remain flanked by repeated CRISPR...
sequences (Erdmann & Garrett, 2012, Swarts et al., 2012, Yosef et al., 2012). The leader is involved in the generation of the new repeat and requires the presence of a CRISPR spacer matching the foreign DNA to promote the addition of the new spacer (process called primed acquisition) (Datsenko et al., 2012, Swarts et al., 2012, Yosef et al., 2012) (Figure 1-9).

1.6.1.1. Putative role of PAM in spacer acquisition

The acquisition of new spacers requires recognition and excision of protospacers, and their integration within CRISPR repeats. In addition to Cas1, Cas2 and Cas4 proteins, the protospacer motif has been suggested to be involved in acquisition or adaptation (Lillestøl et al., 2009, Shah & Garrett, 2011). The first evidence implicating PAM sequences in the acquisition step was provided by Shah et al where distance trees were prepared for different loci of Sulfolobales based on three sequences (repeats, leaders and Cas1 protein) (Shah et al., 2009). Each tree exhibited similar sequences for the same CRISPR loci. In addition to Cas1, repeat and leader sequences, it was also revealed that the protospacer motif was implicated in the acquisition process. Despite being of high interest, the detailed mechanism of spacer acquisition still remains to be elucidated. Recent studies suggested that the PAM sequence likely generates a recognition site for protospacer cleavage from invading elements while at the other end of the protospacer there is no sequence specificity for excision (Erdmann & Garrett, 2012). They proposed that protospacer DNAs were cleaved from the invading element immediately after the PAM motif, on either strand with the secondary cut produced by a ruler mechanism (Erdmann & Garrett, 2012).

Diez-Villasenor et al., also provided a second ruler mechanism, with the first cut occurring at the leader-repeat end of a CRISPR locus and a second cut at the leader distal end of the first repeat (Díez Villaseñor et al., 2013). Uptake of spacer in a reverse direction has been observed for type I-A system of S. solfataricus, a type II-A system of S. agalactiae and type I-E system of E. coli (Erdmann & Garrett, 2012, Lopez - Sanchez et al., 2012, Díez Villaseñor et al., 2013). Type I-A of S. solfataricus and type II-A of S. agalactiae revealed that PAM located at opposite ends of the protospacer can generate the same sequences when inverted 5'-CCN
and NGG-3’. For the type I-E system of *E. coli* it was found a relatively low conservation at position -3 and -2 at the AWG PAM sequence while position -1 was highly conserved (Yosef *et al.*, 2012). The authors also provided evidence that during spacer acquisition the first repeat is copied to form the newly synthesized repeat. Further, Swarts *et al.*, demonstrated a co-occurrence of PAM deviations and repeat mutations, inferring that one nucleotide from the PAM is taken up in the newly synthesized repeat, a finding that was also confirmed by Datsenko *et al.*, (Datsenko *et al.*, 2012, Swarts *et al.*, 2012). It was also proposed that -1 position of the PAM is the terminal nucleotide of the protospacer sequence and not a part of the repeat, a conclusion that was supported by Mojica *et al.*, for some type I-C, I-B and I-F systems where the -1 PAM position matched the first nucleotide of the repeat (Mojica *et al.*, 2009). In summary, despite the recent results for the type I-E system of *E. coli* that give insights into spacer acquisition using the -1 position of the PAM sequence, this mechanism cannot be applied for other type I and type II since the -1 position of the PAM is not conserved (Mojica *et al.*, 2009, Erdmann & Garrett, 2012).

### 1.6.2. Biogenesis of crRNA

The second stage is termed maturation, during which the long primary transcript of a CRISPR locus (pre-crRNA) is generated and processed into short crRNAs. Primary processing of the pre-crRNA involves endoribonucleolytic cleavage within the repeat sequences, either by Cas6 homologues or by RNase III. In Type I, the processing step is mediated by Cas6-like nucleases that bind the repeat stem-loop in pre-crRNA (Haurwitz *et al.*, 2010, Gesner *et al.*, 2011, Sashital *et al.*, 2011). These include Cse3 from Subtype I-E, Csy4 from Subtype I-F (Ypest), Cas6 from Subtype I-A (Apern), I-B (Tneap) and Cas5d from Subtype C that process the pre-crRNA in their corresponding subtype (Brouns *et al.*, 2008, Carte *et al.*, 2008, Haurwitz *et al.*, 2010, Deltcheva *et al.*, 2011, Nam *et al.*, 2012). This typically results in crRNAs that have a repeat derived 5’ handle, followed by the complete spacer sequence and a repeat-derived 3’ handle that forms a hairpin structure. After the initial cleavage, Cas6 proteins remain firmly associated with the hairpin at the 3’ handle of the mature crRNA for cascade formation and subsequent destruction of invading nucleic acid sequences (Jore *et al.*, 2011, Wiedenheft *et al.*, 2012). In type-II, processing of pre-
crRNA relies on a completely different mechanism. In addition to Cas9, crRNA maturation requires a trans-encoded small RNA and the host factor RNase III. Base pairing between the tracrRNA and the repeat region of the pre-crRNA results in a double stranded region that acts as a guide for the pre-crRNA processing by Cas9 and the house keeping ribonuclease RNase III. Cas9 is most probably required for binding and positioning the crRNA-tracrRNA hybrid for cleavage by RNase III (Deltcheva et al., 2011). After initial processing by RNase III, the mature crRNA-tracrRNA hybrid remains firmly associated with Cas9 to bind the target nucleic acid for cleavage (Deltcheva et al., 2011, Gasiunas et al., 2012, Jinek et al., 2012). In type III, expression is similar to that described for type I system. The major difference is that Cas 6 does not bind to a stem-loop but to a non-structured repeat to process the pre-crRNA into crRNA (Carte et al., 2008, Carte et al., 2010). Unlike the type I endoribonuclease, Cas6 then transfer this RNA to a distinct Cas complex termed CMR (type III-B) (Hale et al., 2009, Makarova et al., 2011, Hale et al., 2012) or CSM (type III-A) (Hatoum-Aslan et al., 2011, Hatoum-Aslan et al., 2013) where they are cleaved further at the 3’ end and crRNA maturation (Figure 1-9).

1.6.3. Interference

The third step is interference, during which the foreign DNA or RNA is targeted and cleaved within the proto-spacer sequence (Van Der Oost et al., 2009, Garneau et al., 2010). Finding a proto-spacer sequence that is complementary to the crRNA involves scanning of invader DNA, discriminating self from non-self and base pairing between the 7-8 nucleotide seed region of the spacer and the complementary protospacer, followed by extended base pairing between the spacer and protoscaler, which results in strand cleavage (Sashital et al., 2012). This initial hybridization of crRNA with the complementary target strand results in R-loop formation that function as a signal to recruit nucleases for degradation of double stranded DNA targets. For type I systems scanning and identification of the target seem to be a nonspecific interaction between the cascade complex and the invading DNA. In type I-E systems, the Cse1 recognizes and interacts with the PAM motif to destabilize the DNA duplex and allow the crRNA to access the target DNA for hybridization (Sashital et al., 2012). Base pairing between the crRNA and the
complementary target protospacer is essential for successful interference. In type I-E systems the seed region includes 1-5 and 7-8 at the 5’ end of the spacer (Semenova et al., 2011). Mutations in the seed region abolish CRISPR/Cas mediated immunity. In the case of successful base pairing within this seed region, crRNA strand invasion of the target DNA proceeds and generates an R-loop followed by a conformational change of the cascade complex, bending the target DNA (Westra et al., 2012, Westra et al., 2012). Cas3 nuclease-helicase is recruited to further unwind in an ATP-dependent manner and cleaves using its metal dependent nuclease activity the target DNA. Endonucleolytic cleavage of the displaced strand of the loop is followed by exonucleolytic degradation in a 3’ to 5’ direction (Beloglazova et al., 2011, Mulepati & Bailey, 2011, Sinkunas et al., 2011, defense by CRISPR, 2012, Westra et al., 2012). The other DNA strand undergoes endonucleolytic and exonucleolytic degradation while Cascade dissociates from the target and is recycled (Beloglazova et al., 2011, Mulepati & Bailey, 2011, Sinkunas et al., 2011). In type II interference the signature protein is Cas9 which along with two RNAs tracrRNA:crRNA complex, recognizes and degrades the target (Garneau et al., 2010, Deltcheva et al., 2011, Gasiunas et al., 2012, Jinek et al., 2012). Cas9 binds the complementary strand and proceeds in a step wise manner (Sternberg et al., 2014). Next the Cas9-RNA complex scans the invader DNA to identify a PAM motif that is predominantly similar to that described for the type I. the PAM motif is located on the displaced strand close to the 3’ handle of the crRNA guide. Interaction with the PAM motif is required for DNA binding and subsequent DNA strand displacement and R-loop formation initiate at the PAM (Deveau et al., 2007). Base pairing progress over a 12 nt seed sequence towards the distal end of the target sequence (Jinek et al., 2012, Sternberg et al., 2014). Further, Cas9 uses the HNH like nuclease domain to cleave the DNA strand that base pairs with the crRNA and the RuvC-like nuclease domain cleaves the displaced DNA strand (Jinek et al., 2012, Jinek et al., 2014). The 5’ end of the spacer region has been suggested to contain a seed sequence. In Type III systems, Csm (Csm1-Csm5) complex of type III-A or Cmr (Cmr1-Cmr6) complex of type III-B have been shown to be essential for cleavage. DNA is the target of Csm complex of type III-A system. Type III-A systems discriminate self DNA from non-self DNA in a PAM independent manner with the possible involvement of the Cas10-like
protein Csm1 (Hatoum-Aslan et al., 2014). A seed like sequence at the 5' end of the spacer region of the crRNA has been determined. Recently, Csm6 has been proposed as an additional protein required for interference (Deng et al., 2013). In contrast to type III-A, Cmr complexes of type III-B target RNA rather than DNA. Target RNA cleavage occurs sequentially, at 4-5 cleavage site (Hale et al., 2009, Staals et al., 2013). It is currently not known whether a seed like sequence at the 5' end of the spacer is required for interference. The S. islandicus Cmr complex has been found to target DNA using a mechanism analogue of the CSM complex of III-A (Deng et al., 2013). However, the S. solfataricus Cmr complex targets RNA using a mechanism that differs from those described for the type III-B suggesting functional diversity among type III systems (Zhang et al., 2012) (Figure 1-9).
Figure 1-9: Three stages of CRISPR-Cas action.

A schematic diagram of CRISPR-Cas defense describing the three stages: adaptation, maturation and interference. In type I and type II, but not in type III, the selection of protospacers in invading nucleic acid depends on the PAM sequence.
immediately adjacent to the protospacer. Cas 1 and Cas 2 incorporate the protospacer into the CRISPR locus to form spacers. During the expression stage, CRISPR locus containing the spacers is expressed, producing a long primary transcript (the pre-crRNA). In type I systems, pre-crRNA is cleaved in the presence of Cas5/Cas6-homologues respectively resulting in crRNAs. In Type II systems, trans-encoded small RNA (tracrRNA) binds to the pre-crRNA and the host RNase III and Cas9 are recruited for processing to crRNA. Further maturation might occur by cleavage at a fixed distance within the spacers in a process that require Cas9. In type III systems, Cas6 binds the pre-crRNA, which is then transferred to a distinct Cas complex (Csm in subtype III-A and Cmr in subtype III-B). In subtype III-B systems, the 3′ end of the crRNA is trimmed further. During the interference step, the invading nucleic acid is cleaved. In type I systems, the crRNA guides the Cascade complex to targets that contain the complementary DNA, and the Cas3 subunit is recruited for cleaving the invading DNA. The PAM also plays an important part in target recognition in type I systems. In type II systems, Cas9 loaded with crRNA directly targets invading DNA, in a process that requires the PAM. The two subtypes of CRISPR–Cas type III systems target either DNA (subtype III-A systems) or RNA (subtype III-B). In type III systems, a chromosomal CRISPR locus and an invading DNA fragment are distinguished by either base pairing to the 5′ repeat fragment of the mature crRNA (resulting in no interference) or no base pairing (resulting in interference).

1.6.3.1. Putative role of PAM in spacer interference

The first evidence implicating PAM sequences in the interference stage of the immune response was provided for the type II-A system of *Streptococcus thermophilus* (Deveau *et al.*, 2007, Horvath *et al.*, 2008, Garneau *et al.*, 2010). Mutations in the AGAA motif, located downstream from the protospacer (at the 3′ flank) permitted phages to escape CRISPR defense (Gudbergsdottir *et al.*, 2011). A critical role for PAM sequences in interference was also observed in the crenarchaeal thermoacidophile *Sulfolobus* (Gudbergsdottir *et al.*, 2011). Constructs carrying protospacers and CCN PAM sequence elicited effective interference with a few viable transformants primarily carrying deletions in CRISPR loci that included the
matching spacer. When the PAM sequence was replaced by GGN, GAN or TTN there was no detectable interference while in the presence of TCN there was a significant reduction in transformation efficiency. In the presence of CTN there were also a reduction in transformation efficiency however to a lesser degree than TCN or CCN, suggesting an altered mode of PAM recognition occurring during interference, where possibly a C, at position -2 or -3 was sufficient (Gudbergsdottir et al., 2011). Using a similar plasmid based invader assay to Sulfolobus, Fischer et al. performed a systematic analysis of interference motifs on the archaeal species Haloferax volcanii, which harbor a type I-B CRISPR type. Six different PAM sequences (TTC, ACT, TAA, TAT, TAG and CAC) upstream of the protospacers were found to be active for targeting (Fischer et al., 2012). Since only positions -2 and -3 showed some conservation, it was suggested that the PAM sequence might be TAN (Fischer et al., 2012). For type-B system of another haloarchaeon Haloquadratum walsbyi it was predicted that the TTC motif might be the PAM sequence required during interference (Garcia-Heredia et al., 2012). Based on these results it is possible that for type I systems, PAM recognition during the interference stage is limited to one nucleotide, with some permutations being permitted (Shah et al., 2013). A similar conclusion was provided for a type II-A system of S. agalactiae. In the presence of the NGG PAM sequence, no transformants were formed. When the PAM sequence was converted to AC, CA, AG or GT, there was no detectable interference. However, GA and GG showed an equal reduction in transformation efficiency, consistent with a reduced level of PAM recognition specificity operating at the interference stage (Lopez-Sanchez et al., 2012). Almendros et al., identified the PAM sequences during interference for type I-F system in E. coli. One C located either at PAM position -2 or -3 strongly reduced interference, whereas maintenance of both C produced enhanced interference effects (Almendros et al., 2012). For the type I-E system of E. coli it was shown that T located either at PAM triplet position -2 or A at position -3 appears to be critical for interference (Semenova et al., 2011). Mutations of the A and T resulted in strongly reduced interference. Semenova et al., showed further that in the type I-E system a seven-nucleotide region of a protospacer (“seed sequence”) immediately following the PAM motif is essential for CRISPR/Cas mediated immunity. Mutations in the seed region reduced the binding affinity of the
crRNA-guided Cascade complex to protospacer DNA.

This highly specific interaction could also explain why for type I-E systems, point mutations in the protospacer or PAM sequence is a more common response to interference than in other type-I, type-II and type-III B systems where elimination of matching spacers or loss or mutation of cas genes produce reduced interference effects (Tyson & Banfield, 2008, Gudbergsdottir et al., 2011, Cady et al., 2012, Deng et al., 2012, Fischer et al., 2012, Lopez - Sanchez et al., 2012, Deng et al., 2013). Moreover, structural studies on the type I-E of E. coli demonstrated that CasA and Cas8 form a interference complex that binds to the PAM sequence located on the DNA strand complementary to the crRNA rather than dsDNA or the non-targeted DNA strand (Sashital et al., 2012). This finding led to the proposal that the PAM sequence is implicated in protospacer recognition. In contrast, Jinek et al., have found for the type II system of S. pyogenes that the PAM sequence is recognized on non-complementary DNA strand during interference (Jinek et al., 2012). A summary of experimental data showing the dependence of DNA interference on PAM is presented in Table 1-6 adapted from (Shah et al., 2013).

Table 1-6: PAM-dependent DNA targeting.

<table>
<thead>
<tr>
<th>Bacterial species</th>
<th>CRISPR Subtype</th>
<th>Predicted PAM</th>
<th>Successful Interference</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. solfataricus</td>
<td>I-A1</td>
<td>CCN</td>
<td>CCA; TCA</td>
<td>(Lillestøl et al., 2009, Gudbergsdottir et al., 2011)</td>
</tr>
<tr>
<td>S. solfataricus</td>
<td>I-A2</td>
<td>TCN</td>
<td>TCG</td>
<td>(Manica et al., 2011)</td>
</tr>
<tr>
<td>H. volcanii</td>
<td>I-B</td>
<td>Not determined</td>
<td>TTC; ACT; TAA; TAT; TAG; CAC</td>
<td>(Fischer et al., 2012)</td>
</tr>
<tr>
<td>H. walsbyi</td>
<td>I-B</td>
<td>TTC</td>
<td>Not</td>
<td>(Fischer et al., 2012)</td>
</tr>
</tbody>
</table>
For type III systems there is evidence for PAM-independent interference. The type III-A of *S. epidermis* was shown to confer interference despite mismatched base pairing between the 5’ flank of the crRNA and the PAM region (Marraffini & Sontheimer, 2008). Similar evidence for PAM independent interference was obtained for the type II-B system of *P. furiosus* where target RNA was cleaved despite perfect matching of the 5’ flank of the crRNA to the antisense RNA substrate (Hale et al., 2012). Consistent with these findings, two more studies have demonstrated PAM-independent interference for another RNA-targeting type III-B system of *S. solfataricus* and for a different type DNA-targeting type III-B of *S. islandicus* (Zhang et al., 2012, Deng et al., 2013).

Since spacer acquisition and protospacer interference utilize different molecular mechanisms Shah *et al.*, suggested using the acronym SAM for spacer acquisition motif and TIM for target interference motif (Shah *et al.*, 2013) (Figure 1-10).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong></td>
<td>I-F</td>
<td>CC</td>
<td>GCC; CCC; GCT; CTT; CAA</td>
<td>(Almendros <em>et al.</em>, 2012)</td>
</tr>
<tr>
<td><strong>P. aeruginosa</strong></td>
<td>I-F</td>
<td>CC</td>
<td>AG</td>
<td>(Cady <em>et al.</em>, 2012)</td>
</tr>
<tr>
<td><strong>S. thermophilus</strong></td>
<td>II-A</td>
<td>NNAGAA</td>
<td>AGAA</td>
<td>(Deveau <em>et al.</em>, 2007)</td>
</tr>
<tr>
<td><strong>S. agalactiae</strong></td>
<td>II-A</td>
<td>NGG</td>
<td>NGG; NGA</td>
<td>(Lopez - Sanchez <em>et al.</em>, 2012)</td>
</tr>
<tr>
<td><strong>S. thermophilus</strong></td>
<td>II-A</td>
<td>NGGNG</td>
<td>GGNG</td>
<td>(Sapraauskas <em>et al.</em>, 2011)</td>
</tr>
</tbody>
</table>
Figure 1-10: Putative PAM, SAM and TIM associated with adaptation and interference steps in type I and type II CRISPR systems.

During the adaptation stage by the type I systems the spacer acquisition motif (SAM) is recognized on the invader DNA (upstream of the protospacer) by the spacer acquisition machinery and once the protospacer is selected and excised, it is reinserted within the CRISPR locus. In type II systems, the SAM/PAM motif is recognized on the invader DNA (downstream of the protospacer) by a mechanism similar to the type I systems. During the interference stage by the type I systems, the target interference motif (TIM) is recognized by the Cas protein-crRNA complex on the crRNA-complementary DNA strand. Both TIM recognition and crRNA annealing are required for successful cleavage. In type II systems, TIM is recognized on the invader DNA by the DNA interference complex on the non-complementary DNA strand to the crRNA. Figure adapted from (Shah et al., 2013).
1.7. CRISPR–Cas systems- mechanism(s) of self versus non-self discrimination

Except for type III-B system, which targets and cleaves single-stranded RNA, all other type systems have been shown to target and cleave double-stranded DNA (Hale et al., 2009, Jinek et al., 2012, Westra et al., 2012, Zhang et al., 2012). Hence DNA-targeting CRISPR-Cas systems must efficiently discriminate self versus non-self DNA and avoid aberrant cleavage of genomic DNA. The protospacer itself cannot be used for discrimination. The absence of such discrimination leads to a suicidal autoimmune response (Marraffini & Sontheimer, 2008, Stern et al., 2010, Vercoe et al., 2013). With restriction/modification (R-M) systems this problem is resolved, the host’s genomic DNA is methylated and therefore protected against cleavage, whereas unmodified invader DNA is cleaved. On the other hand, for CRISPR-Cas systems the mechanisms of self/nonself discrimination are poorly understood (Westra et al., 2012). Some recent studies indicate that both type I and II systems have a mechanism to discriminate self (the CRISPR DNA) from nonself (any other DNA) that relies on the PAM conserved sequence motifs (Bolotin et al., 2005, Deveau et al., 2007, Horvath et al., 2008, Mojica et al., 2009). Although Type I systems require PAM sequences (2-3 nt) at the 3’ end of the protospacer on the target strand, Type II systems require the PAM sequence (4-5 nt) at the 5’ end of the protospacer (Bolotin et al., 2005, Deveau et al., 2007, Horvath et al., 2008, Garneau et al., 2010). In contrast, all Type III systems are thought to lack such a PAM requirement to avoid targeting their own CRISPR arrays (Marraffini & Sontheimer, 2010). The absence of a PAM for the RNA targeting type III-B may be explained by the fact that self-targeting of mRNA is less deleterious for the host or serves a regulatory purpose (Hale et al., 2012). Type III-A systems, on the other hand, have been shown to discriminate self versus non-self target through a mechanism that relies on sensing base-pairing between the 5’ handle of the crRNA and the sequence flanking the protospacer (Marraffini & Sontheimer, 2010). Therefore, Type-III-A system targets any DNA sequences with a complimentary protospacer, whereas Type-I and II systems target only those protospacers flanked by the PAM sequence.
1.8. Alternative functions of CRISPR-Cas systems in virulence, gene regulation and physiology

1.8.1. CRISPR-Cas systems, virulence and gene regulation

Recently sequenced bacterial, bacteriophage, and plasmid genomes have given substantial insights into CRISPR-Cas diversity, as well as the foreign elements that are targeted by CRISPR systems. Analyses of genomic sequences of several pathogenic bacteria revealed a direct or indirect link between CRISPR-Cas systems and control of virulence. Mojica et al. (2005) were the first to suggest that pathogenicity of prokaryotic populations is controlled by phages and conjugative plasmids, and that CRISPR spacers targeting these mobile elements might therefore affect bacterial evolution, including pathogenicity and virulence (Mojica et al., 2005). CRISPR sequences were shown to be much less frequent in some bacteria, suggesting that these organisms take the risk of incorporating harmful DNA such as viruses to be able to accumulate virulence genes from other bacteria (Otto, 2013). Comparative analyses of CRISPR spacers revealed sequences with high identity not only to “nonself” DNA of mobile genetic elements but also to “self,” endogenous chromosomal DNA (Mojica et al., 2005). By analyzing CRISPR from 330 organisms it was shown that one in every 250 spacers is self-targeting and that such self-targeting occurs in 18% of all organisms carrying CRISPRs (Stern et al., 2010). The lack of conservation of self targeting spacers among organisms together with abundance of degenerated repeats (not totally conserved) near these spacers, suggest that the acquisition of self spacers is a form of autoimmunity rather than a regulatory mechanism to control gene expression. The authors concluded that accidental incorporation of self-targeting spacers is detrimental to genome integrity and this form of autoimmunity may explain the abundance of degenerated CRISPR systems in prokaryotes with self-targeting spacers (Stern et al., 2010). The existence of prophage-encoded anti-CRISPR genes in Pseudomonas aeruginosa that diminish CRISPR activity could also explain the incidental acquisition of spacers matching endogenous genes thus suggesting a role for self targeting spacers to control the expression of these endogenous genes (Bondy-Denomy & Davidson, 2014). Several
analyses of genomic sequences of pathogenic bacteria suggested a potential role of CRISPR-Cas in processes other than immunity against mobile elements such as regulation of genes involved in virulence (Bourgogne et al., 2008, Shimomura et al., 2011, Kuenne et al., 2013). Bourgogne et al. studied the genome sequences of two Enterococcus faecalis isolates one with and one lacking type II-A CRISPR-Cas systems. CRISPR-Cas harboring strain had enhanced capacity to form biofilms; it promoted host colonization more efficiently than the isolate lacking the system (Bourgogne et al., 2008). Two independent studies looked at the virulence of Listeria monocytogenes during growth in macrophages and both suggested a potential involvement of CRISPR-Cas locus in controlling bacterial pathogenicity in this organism virulence (Kuenne et al., 2013).

Several transcriptome studies found that transcriptional regulators are involved in cas gene expression. For example, OmpR in Yersinia pestis is involved in building resistance against phagocytosis and controls the adaptation to various stressful conditions met in macrophages (Gao et al., 2011). Microarray analysis of the Y. pestis OmpR mutant revealed differential expression of a set of 224 genes including downregulation of cas1 gene expression. The cas1 gene was also repressed during the preadaptation phase, when the strain adapted to the environment in fleas that parasitize rats. However, after passage in rats, cas1 expression was upregulated among other gene transcripts which coincided with escape from the innate immune response and systemic spread leading to fatal sepsis (Sebbane et al., 2006, Gao et al., 2011). Changes in multiple cas gene expression were also observed when transcriptomes obtained from five Salmonella Typhi clinical isolates from the blood circulation system were compared to those strains cultivated in vitro, the authors proposing a potential involvement of Salmonella CRISPR-Cas system in bacterial virulence (Sheikh et al., 2010). In E. faecalis, a bacterial species that can cause opportunistic infectious of the intestine, mutations in stress adaption and virulence genes strongly affect cas gene transcription among a wide variety of genes. Deletion of relA, a GTP pyrophosphokinase involved in (p)ppGpp biosynthesis during amino acid starvation of E. faecalis, treated with antibiotics, led to a strong down regulation of the cas genes (Yan et al., 2009). When the authors used a double mutant RelA and RelB, which is the monofunctional guanosine pentaphosphate synthetase,
expression of the cas genes was induced, suggesting that cas genes function together with the RelAQ system during the stress response and adaptation to potentially unfavorable changes in the environment and contribute to virulence (Yan et al., 2009).

1.8.2. CRISPR-Cas systems and cell physiology

Several studies suggested that the function of CRISPR-Cas goes beyond viral defense, plasmid conjugation and natural transformation. In the soil bacterium, Myxococcus xanthus, three genes within type I-B CRISPR-Cas system (previously named devT, devR and devS), were found to be necessary for triggering the development of fruiting bodies from which bacterial spores are released (Viswanathan et al., 2007). While the exact regulatory mechanism has not been elucidated, the data suggested the CRISPR-CAS system is involved in the regulation of this sporulation event in M. xanthus (Viswanathan et al., 2007, Barrangou, 2013).

In P. aeruginosa, the type I CRISPR/Cas system has been implicated in affecting phage-mediated inhibition of biofilm formation and bacterial swarming (Zegans et al., 2009, Cady & O'Toole, 2011). When phage DMS3 lysogenized its host containing a CRISPR/Cas system with a spacer with complementarity to the phage genome, bacterial population lost its swarming motility and ability to form biofilms. It was suggested that these effects are likely caused by the spacer matching the phage genome (Zegans et al., 2009, Cady & O'Toole, 2011). In E. coli, it was observed an increased expression of type I-E cas operon following stress on the cell envelope (Perez Rodriguez et al., 2011). Apart from H-NS, LeuO, and Rcs/BglJ previously identified as regulators of cas genes (Westra et al., 2010, Arslan et al., 2013), the authors found that BaeSR, a two-component regulatory system, known to respond to envelope stress, upregulated the expression of cas genes (Perez Rodriguez et al., 2011). In the absence of the chaperone DnaK, the expression of ssTorA (Tat-dependent excretion signal peptide of trimethylamine N-oxide reductase) diminished. When the entire cas operon was removed in the DnaK mutant, the expression of ssTorA was restored, as did deletion of the BaeSR system (Perez Rodriguez et al., 2011). These authors showed that upon its expression, the CRISP-Cas system
target ssTor-A encoding sequences affecting protein transport across the membrane (Perez Rodriguez et al., 2011). Although the molecular details await further analysis, this study again suggested that the function of CRISPR-Cas goes beyond viral defense and plasmid conjugation. In a different study, the type I system encoded by E. coli has been shown to play a role in DNA repair (Babu et al., 2010). It was observed that both Cas1 and the crRNA array apart from previously known immunity function, also they have a DNA repair function by interacting with key components of DNA repair systems (Babu et al., 2010). Type II systems have also been identified as having potential roles in bacterial physiology beyond their immunity function. Cas2 within type II CRISPR-Cas system was shown to play an important role in the ability of Legionella pneumophila to replicate within amoebae through a mechanism which is still unclear (Gunderson & Cianciotto, 2013). It was hypothesized that Cas2 acts in combination with a currently unidentified small mRNA that might have an alternative function, either in their processing or affecting the mRNA stability (Gunderson & Cianciotto, 2013). The possibility that CRISPR/Cas systems exert gene regulatory functions or play a role in stress response or virulence is an interesting hypothesis that awaits further mechanistic analysis.

1.9. CRISPR/Cas systems in S. mutans

1.9.1. Role of CRISPR/Cas systems in bacteriophage resistance

Bacterial viruses (bacteriophages or phages) represent important players in shaping the evolution of bacterial communities, including the oral cavity (Labrie et al., 2010, Pride et al., 2011). Even though, relatively little is known about S. mutans phages, they may be required in establishment, maintenance and pathogenicity of this cariogen in the oral cavity (Van der Ploeg, 2009). Phages of S. mutans could also be used in the control and prevention of dental caries (Hitch et al., 2004). In addition, phage-derived peptidoglycan hydrolases (endolysins) may have the potential to treat or prevent dental caries (Delisle et al., 2012). They may remove cariogenic bacteria in vivo, in a targeted manner, leaving beneficial organisms for maintaining good oral health (Delisle et al., 2012). However, recent studies reported that lytic phages
against \textit{S. mutans} strains are difficult to isolate (Van der Ploeg, 2009). To date, there have been only five phages described: M102, M101, M102D, e1 and f2 to have lytic activity against \textit{S. mutans} strains of serotype c, e and f, respectively (Delisle & Rostkowski, 1993, Van der Ploeg, 2007, Van der Ploeg, 2009, Delisle \textit{et al.}, 2012). All \textit{S. mutans} serotype c strains, including UA159, are resistant to phage infection by M102 except strain OMZ381 which shows sensitivity to phage infection resulting in cell lysis (Delisle & Rostkowski, 1993, Van der Ploeg, 2007, Van der Ploeg, 2009). The causes of this limited host range are not known. It is well known that bacteria possess multiple mechanisms to counterattack phage infection. Natural phage resistance mechanisms have been characterized in many bacterial species, which are aimed especially at limiting the invasion of phage DNA, that include: 1) phage adsorption inhibition, 2) DNA injection blocking, restriction/modification of phage DNA, or 4) abortive infection (Abi) (Forde & Fitzgerald, 1999). In the case of \textit{S. mutans}, there is a tremendous lack of known phage resistance mechanisms. During the phage infection cycle, phages need to specifically recognize and bind a cell receptor exposed on the surface of the cell wall and regarding \textit{S. mutans}, phage host receptors seems to be carbohydrate receptors (Delisle \textit{et al.}, 2012). All serotypes of \textit{S. mutans} (serotypes c, e, and f) produce cell wall rhamnose–glucose polysaccharides (RGPs) that has a backbone structure of $\alpha 1,2$- and $\alpha 1,3$-linked rhamnosyl units with glucose side chains linked to alternate molecules on the cell wall (Shibata \textit{et al.}, 2009). Each serotype RGP has unique linkages of its glucose side chains (Pritchard \textit{et al.}, 1986). It was found that the glucose side chain of RGP is required for adsorption and infection of \textit{S. mutans} serotype c phage M102 (Shibata \textit{et al.}, 2009). Since M102 adsorbed also to the strains that do not support propagation, it was suggested the factors different from phage adsorption must contribute to resistance of \textit{S. mutans} serotype c strains to infection by M102 (Shibata \textit{et al.}, 2009, Van der Ploeg, 2009). Recently, numerous studies have shown that bacteria developed a phage resistance mechanism, which relies on CRISPR/Cas systems and shown to protect microbial cells against exogenous DNA of phage or plasmid origin. In the \textit{S. mutans} two distinct CRISPR loci have been identified within the chromosome.
Figure 1-11: Organization of CRISPR1 (A) and CRISPR2 (B) arrays and associated genes in S. mutans UA159.

CRISPR repeats are indicated by white rectangulars and spacers are indicated by ovals.

As outlined in Figure 1-11 (A) CRISPR1 is located between ORFs SMU.1400 and SMU.1398, and contains seven repeats interspersed by six spacers of 30 bp in size. The CRISPR1 array is preceded by four cas genes: SMU.1405c, SMU.1404c, SMU.1403c and SMU.1402c. The CRISPR2 locus of UA159 is located between SMU.1753c and SMU.1752c, and consists of two repeats separated by a spacer (Figure 1-11 B). The genes from SMU.1764c to SMU.1753c are located upstream of the first repeat.

Analysis of CRISPR arrays in S. mutans strains revealed that S. mutans is often attacked by M102-like bacteriophages suggesting frequent occurrence of acquired immunity against infection by M102-like bacteriophages (Van der Ploeg, 2009). Despite this knowledge, the molecular and genetics mechanism responsible for resistance to M102 in serotype c S. mutans strain remains unknown.
1.9.2. CRISPR/Cas link to stress response

Several transcriptome studies in *S. mutans* revealed that deletion in virulence or global regulatory genes affected *cas* gene expression in both type I-C CRISPR/Cas system and type II-A CRISPR/Cas system. The first evidence supporting a role for CRISPR/Cas systems in stress response was presented by Chattoraj *et al.*, who found that the deletion of ClpP affected the expression of more than 100 genes, including the *cas* genes associated with the type I-C CRISPR-Cas system (Chattoraj *et al.*, 2010). In *S. mutans*, the ClpP peptidase is a subunit of Clp proteolytic complex (ClpC, ClpE, or ClpX) (Lemos & Burne, 2002) which play a critical role in cell homeostasis by assisting in refolding or degrading damaged proteins that accumulate in the cytoplasm following exposure to stress-inducing conditions (Gottesman, 2003). How the ClpP subunit controls the regulation of *cas* gene expression is still unknown but it appears that the *cas* genes function together with Clp during the stress conditions and adaptation to changes in the bacterial environment.

Results linking CRISPR/Cas to stress response in *S. mutans* has also been reported by others. Deletion of genes encoding the Spx complex (*spxA, spxB*), revealed increased expression of the *cas* genes associated with the type I-C CRISPR-Cas system (Kajfasz *et al.*, 2010, Kajfasz *et al.*, 2011). The *S. mutans* Spx (suppressor of *clpP* and *clpX*) transcriptional regulator proteins have been shown to modulate stress tolerance, survival and virulence in *S. mutans* (Kajfasz *et al.*, 2010, Kajfasz *et al.*, 2011).

Furthermore, a transcriptome analysis of a mixed culture with *S. mutans*, *S. gordonii* and *Veillonella parvula* revealed differential expression of *cas* genes belonging to the type I-C and type II-A systems (Liu *et al.*, 2011). Finally, a global gene expression profiling study of *S. mutans* cultivated under aerobic or anaerobic conditions revealed that oxygen is a key environmental signal that significantly alters the transcriptome and that these alterations strongly influence bacteriocin production, carbohydrate metabolism, and the expression of known virulence attributes (Ahn *et al.*, 2007). DNA microarray analyses revealed that about 5% of *S. mutans* genome, including some of *cas* genes within CRISPR2/Cas system displayed upregulated
expression in response to growth in the presence of oxygen (Ahn et al., 2007). Whether the CRISPR gene products play a role in modulating the expression of virulence genes and biofilm formation in *S. mutans* in response to changes in oxygen availability remains to be determined.

1.10. Rationale

While CRISPR-Cas systems were originally thought to act as adaptive immunity defense systems, several studies have shown that in addition to protecting the host against invading viruses, plasmid or chromosomal DNA, CRISPR-Cas can participate in the regulation of a range of stress responses in general and bacterial virulence in particular.

The naturally transformable oral bacterium *S. mutans* is well studied at the genetic and physiological level. *S. mutans* strain UA159 harbors two uncharacterized CRISPR/Cas systems, a Type II-A designated CRISPR1 and a type I-C designated CRISPR2. Recent studies indicate that both CRISPR systems within *S. mutans* may play roles in various cellular processes in addition to host defense against phage and plasmid invasion. Despite this indirect evidence that tied CRISPR/Cas systems to cellular processes in *S. mutans*, little is known about the potential role of CRISPR/Cas systems as an adaptive immunity systems nor about their diverse involvement in the stress responses and cell physiology.

The general aim of this dissertation was to examine the role of CRISPR/Cas systems in phage defense, DNA transformation and stress response of *S. mutans*.

**General hypothesis:** *S. mutans* CRISPR/Cas systems target and degrade incoming DNA using CRISPR spacers to guide Cas proteins to cleave invading DNA via phage infection or natural transformation. Function of CRISPR-Cas systems goes beyond viral defense and natural transformation, they play a role in stress response and adaptation to potentially unfavorable changes in the environment to contribute to virulence.
Specific aims:

Specific Aim 1: Investigate whether CRISPR/Cas systems of *S. mutans* constitute a barrier to phage infection.

Specific Aim 2: Examine the role of CRISPR/Cas systems in regulating plasmid transformation in *S. mutans*.

Specific Aim 3: Determine if and how CRISPR/Cas systems participate in the physiology of *S. mutans* with emphasis on the stress response.
1.11. References


pan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. *BMC Genomics* **14**: 47.


Lunsford RD & Roble AG (1997) comYA, a gene similar to comGA of *Bacillus subtilis*, is essential for competence-factor-dependent DNA transformation in *Streptococcus gordonii*. *Journal of Bacteriology* 179: 3122-3126.


pheromone and a two-component regulatory system. *Molecular Microbiology* **21**: 853-862.


Pourcel C, Salvignol G & Vergnaud G (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**: 653-663.


Chapter 2: Role of the *Streptococcus mutans* CRISPR/Cas systems in immunity and cell physiology

2.1. Abstract

CRISPR-Cas systems provide adaptive microbial immunity against invading viruses and plasmids. The cariogenic bacterium *Streptococcus mutans* UA159 has two CRISPR-Cas systems: CRISPR1 (type II-A) and CRISPR2 (type I-C) with several spacers from both CRISPR cassettes matching sequences of phage M102 or genomic sequences of other *S. mutans*. Deletion of *cas* genes of CRISPR1 (∆C1S), CRISPR2 (∆C2E), both CRISPR1+2 (∆C1SC2E) or removal of spacers 2 and 3 (∆CR1SP13E) in *S. mutans* UA159 did not affect phage sensitivity when challenged with virulent phage M102. Using plasmid transformation experiments, we demonstrated that the CRISPR1-Cas system inhibits transformation of *S. mutans* by the plasmids matching the spacers 2 and 3. Functional analysis of the *cas* deletion mutants revealed that in addition to a role in plasmid targeting, both CRISPR systems also contribute to the regulation of bacterial physiology in *S. mutans*. Compared to wild type cells, the ∆C1S strain displayed diminished growth under cell membrane and oxidative stress, enhanced growth under low pH and had reduced survival under heat shock and DNA damaging conditions, whereas the ∆C2E strain exhibited increased sensitivity to heat shock. Transcriptional analysis revealed that the two component signal transduction system VicR/K differentially modulates expression of *cas* genes within CRISPR-Cas systems suggesting that VicR/K might coordinate the expression of two CRISPR-Cas systems. Collectively, we provide *in vivo* evidence that the type II-A CRISPR-Cas system of *S. mutans* may be targeted to manipulate its stress response and to influence the host to control the uptake and dissemination of antibiotic resistance genes.

2.2. Introduction

to 72 bp called spacers derived from exogeneous genetic elements (Bo lotin et al., 2005, Mojica et al., 2005, Pourcel et al., 2005). CRISPRs are often associated with a set of cas genes which encode proteins that mediate the defense process. This CRISPR-mediated defense system targets invading DNA in three steps: 1) adaptation via incorporation of foreign genetic element-derived spacers into the CRISPR array, 2) transcription of CRISPR RNAs containing spacer-repeat units and 3) interference with the invasive nucleic acid leading to their degradation (Makarova et al., 2011). According to the current classification, there are three major types of the CRISPR-Cas systems, type I, type II and type III systems, respectively, that differ by the repertoires of cas genes, the organization of cas operons and the structure of repeats in the CRISPR array (Makarova et al., 2011). In the type I CRISPR-Cas systems the maturation of the precursor CRISPR RNA (pre-crRNA) is mediated by an endonuclease, namely Cse3 (type I-E), Csy4 (type I-F) and Cas5d (type I-C), whereas in the type III systems Cas6 is responsible for crRNA maturation (Carte et al., 2008, Carte et al., 2010, Haurwitz et al., 2010, Gesner et al., 2011, Sashital et al., 2011, Wang et al., 2011, Garside et al., 2012, Nam et al., 2012). In the type II systems trans-activating CRISPR RNA (tracr RNA) binds to the pre-crRNA forming a dual-RNA that is essential for both crRNA maturation by RNAase III and invading DNA cleavage by Cas9 (Garneau et al., 2010, Deltcheva et al., 2011, Jinek et al., 2012, Magadán et al., 2012, Chylinski et al., 2013, Karvelis et al., 2013). Similar to type I, type II-CRISPR-Cas systems require a short protospacer adjacent motif (PAM) that is located immediately adjacent of the protospacer on the foreign DNA element (Horvath et al., 2008, Mojica et al., 2009, Deveau et al., 2010, Bhaya et al., 2011).

Although CRISPR interference was originally defined as a phage resistance mechanism, CRISPR-Cas systems are now known to play a broader role in limiting horizontal gene transfer (Marraffini & Sontheimer, 2008). In Staphylococcus epidermidis and S. pyogenes CRISPR/Cas systems were shown to prevent acquisition of plasmids or prophages by blocking entry in a manner akin to that performed against phage DNA (Marraffini & Sontheimer, 2008, Nozawa et al., 2011). Similar observations have been made in Enterococcus faecalis, Enterococcus
faecium and Campylobacter jejuni (Palmer & Gilmore, 2010, Dugar et al., 2013). However, recently in Streptococcus pneumoniae and Neisseria meningitidis, CRISPR/Cas systems were shown to prevent natural transformation (Bikard et al., 2012, Zhang et al., 2013). Beyond their now canonical function in foreign nucleic acid defense, CRISPR-Cas systems have also been implicated in various aspects of bacterial physiology, virulence and gene regulation (Viswanathan et al., 2007, Hale et al., 2009, Zegans et al., 2009, Babu et al., 2010, Stern et al., 2010, Makarova et al., 2011, Gunderson & Cianciotto, 2013, Sampson et al., 2013).

In Streptococcus mutans, one of the primary pathogens implicated in dental caries, relatively little is known about its virulent phages (Van der Ploeg, 2007, Delisle et al., 2012). Only five phages, designated M101, M102AD, M102, e10 and f1 have been shown to have lytic activity against S. mutans strains of serotype c, e and f, respectively (Van der Ploeg, 2009, Delisle et al., 2012). Except for S. mutans strain OMZ381, all S. mutans serotype c strains, including UA159, are known to be resistant to phage infection by M102; only strain OMZ381 showed sensitivity to phage infection resulting in cell lysis (Van der Ploeg, 2009). Despite this knowledge, the mechanisms responsible for resistance to M102 in S. mutans serotype c remain unknown. S. mutans strain UA159 harbor two distinct CRISPR-Cas systems: a type II-A CRISPR1-Cas system and a type I-C CRISPR2-Cas system (Haft et al., 2005, Horvath et al., 2009, Maruyama et al., 2009, Van der Ploeg, 2009). The analysis of CRISPR cassettes in 29 S. mutans strains revealed that CRISPR spacers had high sequence similarity with M102, a virulent siphophage specific for S. mutans, suggesting that phage-derived spacers present in these strains likely resulted from M102-like phage attacks (Van der Ploeg, 2009). Subsequently, it was shown that M102 adheres to phage sensitive and phage resistant S. mutans serotype c strains, indicating that factors besides phage adsorption determine resistance of S. mutans serotype c strains to infection by M102 phage (Shibata et al., 2009). Despite these studies that explored the role of CRISPR-Cas systems in S. mutans in conferring phage immunity recent transcriptome studies hint towards other functions that CRISPR-Cas systems might have in S. mutans, are however poorly understood.
Here, we investigated the role of CRISPR-Cas systems in phage defense, natural transformation, and stress resistance of S. mutans by utilizing cas gene deletion mutants in S. mutans UA159. We found that S. mutans CRISPR-Cas systems are not essential for phage resistance against M102. However, we demonstrated that the S. mutans type II-A CRISPR1-Cas system inhibits plasmid transformation. Furthermore, we showed that CRISPR-Cas systems are regulated by the VicR/K signaling system to modulate environmental stress tolerance and DNA repair, thereby expanding the role of CRISPR-Cas systems in this pathogen.

2.3. Materials and Methods

**Strains, plasmids, phage, and growth conditions.** Bacterial strains, plasmids and the phage used in this study are listed in Table 2-1. All S. mutans strains were grown in Todd-Hewitt broth supplemented with 0.3% yeast extract (THYE) (Becton Dickinson, Sparks, MD) as static cultures or on THYE medium with 1.5% (wt/vol) agar (Bioshop, Burlington, Ontario, Canada) at 37°C in a 5% (vol/vol) CO\(_2\) atmosphere. Kanamycin (1mg/ml), spectinomycin (1 mg/ml) and/or erythromycin (10 µg/ml) were added as needed. *Escherichia coli* DH5\(\alpha\) was used as the host for propagation of plasmids and was routinely cultured in Luria-Bertani (LB) medium supplemented (when necessary) with spectinomycin (100 µg/ml) at 37°C with aeration. Phage M102 was propagated in S. mutans strain OMZ381 and phage titre was determined using a plaque assay as described previously (Van der Ploeg, 2007). Phage resistance was assayed using both the plaque formation assay as described previously (Van der Ploeg, 2009) and liquid growth assays. For plaque assays, 100 µl of exponentially growing bacterial cultures was mixed with 100 µl of undiluted (~10\(^8\) pfu) and 100-fold serial dilutions of phage M102. After incubation at 37°C for 20 min, 4 ml THYE soft top agar was added and immediately poured on THYE plates and incubated for 48 h at 37°C. Phage sensitivity was assessed based on the number of discrete plaques. For liquid growth assays, 20 µl of mid-log-phase cultures diluted in 350 µl of fresh THYE medium were challenged with 5 µl of
undiluted phage M102 (~10⁸ pfu) and phage sensitivity was monitored using an automated growth reader (Bioscreen C Labsystems, Finland).

Growth kinetic experiments were performed under the following stress conditions: pH 5.5, 0.4 M NaCl, 0.003% H₂O₂, 0.004% SDS, or 25 mM paraquat, as previously described (Senadheera et al., 2007). No antibiotics were used in growth assays in order to avoid additional stress. For heat-shock resistance assays, mid-log cells were incubated at 50°C for 60 min. Samples after heat exposure or incubation at 37°C for 1 h were serially-diluted, plated on THYE plates and colony forming units (CFU) were counted.

Table 2-1: Bacterial strains, plasmids and phages used in this study.

<table>
<thead>
<tr>
<th>Strain, phage or plasmid</th>
<th>Relevant Characteristics</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>UA159</td>
<td>wild-type strain; Erm⁺, Kan⁺, Spec⁺</td>
<td>(Ajdic et al., 2002)</td>
</tr>
<tr>
<td>OMZ381</td>
<td>wild-type strain; Erm⁺, Kan⁺, Spec⁺</td>
<td>(Ajdic et al., 2002, Van der Ploeg, 2009)</td>
</tr>
<tr>
<td>SmuvicK</td>
<td>VicK deficient mutant derived from UA159, Erm'</td>
<td>(Senadheera et al., 2005)</td>
</tr>
<tr>
<td>ΔC1S</td>
<td>lacking cas9 to csn2, Spec'</td>
<td>This study</td>
</tr>
<tr>
<td>ΔC1K</td>
<td>lacking cas9 to csn2, Kan'</td>
<td>This study</td>
</tr>
<tr>
<td>ΔC2E</td>
<td>lacking cas3 to cas2, Erm'</td>
<td>This study</td>
</tr>
<tr>
<td>ΔC1SC2E</td>
<td>lacking all cas genes, Erm', Spec'</td>
<td>This study</td>
</tr>
<tr>
<td>ΔC1KC2E</td>
<td>lacking all cas genes Erm', Kan'</td>
<td>This study</td>
</tr>
<tr>
<td>Δ CR1SP13E</td>
<td>lacking 4 repeats and spacers 1 to 3, Erm'</td>
<td>This study</td>
</tr>
<tr>
<td>Phage M102</td>
<td></td>
<td>(Van der Ploeg, 2007)</td>
</tr>
<tr>
<td>Vector Name</td>
<td>Description</td>
<td>Source</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td>--------</td>
</tr>
<tr>
<td>pCG1</td>
<td>Streptococcus-E. coli shuttle vector, Spec&lt;sup&gt;+&lt;/sup&gt;</td>
<td>(Chen et al., 2011)</td>
</tr>
<tr>
<td>pCG1SP</td>
<td>pCG1 containing irrelevant spacer, Spec&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pCR1SP2 (with flank)</td>
<td>pCG1 containing spacer 2 with potential PAM, Spec&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pCR1SP2 (no flank)</td>
<td>pCG1 containing spacer 2 no potential PAM, Spec&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pCR1SP3 (with flank)</td>
<td>pCG1 containing spacer 3 with potential PAM, Spec&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pCR1SP3 (no flank)</td>
<td>pCG1 containing spacer 3 no potential PAM, Spec&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pCR1SP6 (with flank)</td>
<td>pCG1 containing spacer 6 with potential PAM, Spec&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pCR1SP6 (no flank)</td>
<td>pCG1 containing spacer 6 no potential PAM, Spec&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pCR2SP1 (with flank)</td>
<td>pCG1 containing spacer 1 with potential PAM, Spec&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pCR2SP1 (no flank)</td>
<td>pCG1 containing spacer 1 no potential PAM, Spec&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study</td>
</tr>
</tbody>
</table>

**Construction of mutants in S. mutans.** PCR ligation mutagenesis (Lau et al., 2002) with primers in Table S2-1 was utilized to construct non-polar insertion-deletion mutants in cas genes or CRISPR spacers in S. mutans UA159 wild-type strain: 1) two deletion mutants in the cas9-csn2 operon within CRISPR1-Cas system: (strain ΔC1S and ΔC1K), 2) one deletion mutant in the cas3-cas2 operon within CRISPR2-Cas system (strain ΔC2E), 3) two deletion mutants in all cas genes within CRISPR1-Cas system and CRISPR2-Cas system (strain ΔC1SC2E and strain ΔC1KC2E) and 4) a deletion mutant lacking four repeats and three spacers within
the CRISPR1 array including the spacer identical to M102 (strain ΔCR1SP13E). Successful mutagenesis was validated using nucleotide sequence analysis and quantitative real-time PCR (qRT-PCR).

**Plasmid construction for transformation studies.** Shuttle vector pCG1 that replicates in both *E. coli* and *S. mutans* (Figure 2-2A) was used to clone predicted protospacers for the purposes of plasmid transformation assays (Chen *et al.*, 2011). This plasmid has a beta-galactosidase gene that can be disrupted by inserting the spacer sequence and can be quickly screened (blue/white on X-gal) for successful cloning. Oligonucleotides (50-54 nucleotides) corresponding to protospacer candidates (matching spacer 2,3 and 6 of CRISPR1 locus and 1 of CRISPR2 locus) along with/without 10 nt upstream and downstream sequences were obtained from ACGT Toronto, Ontario (Table S2). Protospacer candidates were selected as containing a sequence with >85% similarity to the *S. mutans* UA159 spacer sequences and originating from virulent phage M102 or genomic DNA from closely related species (*S. mutans* GS-5 and LJ23 genomes). Extra bases corresponding to a *SacI* restriction sites were added onto the synthesized oligonucleotides such that sticky ends were created when annealed oligonucleotides were digested. After digestion these protospacers were ligated to pCG1 digested with *SacI* and dephosphorylated with alkaline phosphatase (New England Biolabs). All constructs were transformed and propagated in *E. coli* DH5α prior to transformation of *S. mutans* UA159. Successful spacer cloning was validated using nucleotide sequence analysis.

**Competence assays.** In this study, natural transformation of planktonic-cell suspensions of UA159 and ΔC1K, ΔC2E, ΔC1KC2E mutants was assessed using streptococcal plasmid pCG1 constructs (Table 2-1, Figure 2-2 and associated text for details) and compared with an empty vector control. Overnight cultures of UA159 and its mutant strains were diluted 20-fold in THYE and incubated at 37°C until an OD600 ~ of 0.1 was reached. Following the incubation period, 1 µg of plasmid DNA and CSP at a final concentration of 1 ug/ml were added. Transformation frequency (TF) assays were conducted as described previously (Senadheera *et al.*, 2007).
**In vivo assay for DNA damage.** Cells in mid-exponential phase were exposed to UV light (intensity of approximately 125 µW/cm²) for 0, 2, 4, 6, 8 and 10 min and then serially-diluted cultures were spotted on THYE agar plates, incubated in the dark at 37°C for 48 h and CFU were counted. For mitomycin C (MMC) sensitivity assays, exponentially growing strains in THYE broth were harvested by centrifugation, resuspended in THYE in the presence of 0.05 µg/ml MMC or absence (control) and incubated at 37°C for 90 min. Sensitivity was quantitatively assessed by plating cells after incubation.

**Gene cloning and protein purification.** Cas5d (SMU.1763c) was cloned using genomic DNA from *S. mutans* UA159 and primers in Table S2-1 into the modified pET15b plasmid as previously described (Beloglazova *et al.*, 2008). For enzymatic assays, Cas5d protein was overexpressed as a fusion with an N-terminal His6-tag in *E. coli* BL21 (DE3) strain (Novagen) which contains no *cas* genes (Brouns *et al.*, 2008). The protein was purified to more than 95% homogeneity using metal-chelate affinity chromatography on a nickel affinity resin and subsequent ion-exchange chromatography on a MonoQ column as previously described (Beloglazova *et al.*, 2008).

**Ribonuclease assays.** RNA1 (39 nt); 5’- AAAUACGUUUUCUCCAUUGUCAUAUUGC GCAUAAGUUGA), and RNA2 (40 nt); 5’UUUCAAUCCUUUUAGGAUAUCUUGAAGAAGAGAU AGAGUUAAA) substrates were obtained from Integrated DNA Technologies. Labeling of RNA substrates was conducted using ATP [γ-32P] and T4 Polynucleotide Kinase per the manufacturer’s instructions (Fermentas). Reaction conditions were as follows; 50 mM Tris pH 7.5, 100 mM KCl, 5 mM MgCl₂, 0.1 nM substrate and 200 ng enzyme for 1 h at 37°C. Samples were run on 15% PAGE gels containing 8 M urea and reaction products were visualized by autoradiography. When RNA from UA159 was used as substrate, total RNA was extracted from cells grown to mid-log phase (OD600 ~ 0.4) in THYE medium. One microgram of total RNA was incubated with 1 µg of Cas5d protein in transcription buffer (50 mM Tris HCl, pH 8.5, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT) for 1 h at 37°C. RNA samples were purified using RNeasy kit
(Qiagen) and then subjected to microarray analysis as described below (Senadheera et al., 2009).

**Cell preparation for gene expression and microarray analysis.** For quantitative real-time PCR (qRT-PCR), cells from S. mutans UA159 and SmuvicK were grown to mid-log phase ($OD_{600}$ of ~ 0.4) under regular or acidic conditions (pH7.5 versus pH5.5). To study cas gene expression under regular conditions, overnight cultures were diluted 1:20 in sterile prewarmed THYE and grown to mid-log phase. Cells were harvested by centrifugation, snap-frozen in liquid nitrogen and stored at - 80°C until required. To measure cas gene expression under acidic condition, overnight cells were diluted 20-fold in sterile prewarmed TYE (tryptone-yeast extract) supplemented with 0.5% glucose (pH 7.5) and grown to mid-log phase ($OD_{600}$, 0.4-0.5). Cultures were then divided into two aliquots, and cells were collected by centrifugation. Pellets were resuspended in 0.5% glucose supplemented TYE adjusted to pH 7.5 or pH 5.5, cultures were incubated for 1h at 37°C with 5% CO$_2$. Cells were then harvested by centrifugation, snap-frozen in liquid nitrogen, and stored at -80°C until RNA isolation. For microarray analysis, cells from S. mutans UA159 at mid-log phase were harvested by centrifugation and utilized for RNA isolation. For qRT-PCR to study non-polar effects on the downstream genes, cells from UA159 and cas deletion mutants were grown in THYE until an $OD_{600}$ of ~0.4 was reached, harvested by centrifugation, snap frozen in liquid nitrogen and stored until used for cDNA synthesis. For Northern blot analysis, cells from S. mutans UA159 were grown to early-log ($OD_{600}$ of ~0.1), mid-log ($OD_{600}$ of ~0.4) and early-stationary phase ($OD_{600}$ of ~1.0) in THYE at 37°C and 5%CO$_2$. For competence stimulating peptide (CSP) treatment, cells from UA159 grown to early-log phase ($OD_{600}$ of ~0.1) were supplemented with 0.2 mM CSP and incubated for 2 hrs at 37°C and 5% CO$_2$ until a final $OD_{600}$ of ~ 0.4 was reached. For heat shock and oxidative stress, cells that reached an $OD_{600}$ of ~0.4 were pelleted, and cell exposed to (0.003 %) H$_2$O$_2$ and 50°C were incubated for 20 min. The cells were then pelleted, snap-frozen in liquid nitrogen and stored at -80°C until used.

**Gene expression analysis using qRT-PCR.** DNase treatment, cDNA synthesis, qRT-PCR and expression analysis were carried out as previously described
Global transcriptome analysis. RNAs from mid-exponential phase cells of *S. mutans* UA159 were incubated in the presence (experimental) or absence (control) of Cas5d (SMU.1763c) protein in transcription buffer (50 mM Tris HCl, pH 8.5, 100 mM KCl, 5 mM MgCl$_2$, 1 mM DTT) for 1 h at 37°C. The RNA samples were transcribed into cDNA using a First-Strand synthesis kit (Invitrogen) as specified in the manufacturer's protocol. Control and experimental cDNAs were used for microarrays (Senadheera et al., 2009). A class comparison analysis was performed to identify statistically significant genes. The statistical algorithm used was the two-sample *t* test (with random variance model) with the parametric *P* value cut-off set to *p* <0.05. Selected genes that showed significant differential expression under experimental conditions were validated utilizing qRT-PCR.

In vitro transcription analysis. RNA transcripts from DNA templates of SMU.995 and SMU.1502c were obtained by *in vitro* transcription using MAXIscript kit (Ambion). Transcription reactions (20 µl) containing 10 mM each of ATP, CTP, GTP, and UTP, 1 µg of PCR product for each transcript, 2 µl of transcription buffer and 2 µl of T7 phage RNA polymerase was incubated for 1 h at 37°C. TURBO DNase (1 µl) was added to the reaction to remove the template DNA. The transcripts were purified using 3M sodium acetate and 100% ethanol and used for ribonuclease activity assay as described above.

Northern Blot Analysis. Total RNA was isolated from UA159 cultures incubated under different growth conditions, as indicated above, using Direct-zol™ RNA MiniPrep (Zymo Research). Five micrograms of total RNA was loaded by lane and resolved on a 8% (wt/vol) polyacrylamide denaturing gel containing 7 M urea. Size-fractionated RNA was transferred to a positively charged nylon membrane (Fermentas) using a Bio-Rad Mini Trans-Blot Cell and subjected to UV crosslinking for 5 minutes. Membranes were pre-hybridized using DIG Easy Hyb (Roche) for 30 min at 42°C and followed by hybridization with DIG High Prime DNA probes (25 ng/
ml) in DIG Easy Hyb hybridization buffer (Roche) at 42°C overnight. The probes for tracrRNA were PCR amplified from UA159 gDNA using tracrRNA-For and tracrRNA-Rev primers (Table S2-1) and labeled using the digoxygenin (DIG) High Prime Labeling Kit (Roche) according to supplier’s instructions. The probed membrane was incubated with CSPD (Roche) and the chemiluminescent signal was visualized using a chemiluminescent detector (Bio-Rad) and photographed. Densitometry was used to determine the transcript expression levels within the detected bands and were quantified using the ImageJ64 program (National Institutes of Health, Bethesda) (Abràmoff et al., 2004). 5S rRNA served as loading control.

2.4. Results

2.4.1. The S. mutans UA159 CRISPR-Cas systems

The S. mutans UA159 genome contains two CRISPR-Cas systems: CRISPR1-Cas system (Figure 2-1A) and CRISPR2-Cas system (Figure 2-1B). The CRISPR1-Cas system of subtype II-A spans approximately 5.8 Kb and has four cas genes organized in an operon: Cas1 (SMU.1404c, 288 aa) and Cas2 (SMU.1403c, 109 aa) encoding core proteins predicted to be implicated together with Csn2 (SMU.1402c, 190 aa) in spacer acquisition, and Cas9 (also known as Csn1) (SMU.1405c, 1345 aa), which encodes the hallmark protein of type II systems associated with the interference step. A CRISPR array (located downstream of SMU.1402) consists of seven repeats (36 bp) interspaced by six spacers (30 bp). Spacer 3 shared 100% nucleotide identity to a 30 bp sequence from S. mutans phage M102, and it was previously hypothesized to confer resistance against M102-like phage infection (Van der Ploeg, 2009). Our sequence similarity searches using the NCBI database for the other spacer sequences within the CRISPR1 locus revealed at least one potential target. For simplicity, we considered only candidate protospacers that matched the CRISPR spacers without or with a up to four mismatches within the CRISPR spacers. Spacer 2 partially matched (26 bp out of 30 bp with 100% sequence identity) to phage M102 and spacer 6 had 100% sequence identity to a segment in the S. mutans GS-5 genome suggesting that CRISPR1-Cas system of S. mutans
UA159 might target not only M102 phages but also incoming DNA from other S. mutans genomes (Table S2-4). This possibility is in agreement with the recent studies that examined in vivo expression of CRISPR locus in S. mutans UA159 suggesting that CRISPR locus is transcriptionally active (Deltcheva et al., 2011, Chylinski et al., 2013). Our deduced protein sequence homology search using BlastP revealed that the cas genes associated with CRISPR encoded hypotetical proteins that are predicted to have nuclease activity involved in defense mechanisms or DNA repair (Table S3) (Haft et al., 2005, Makarova et al., 2006). The CRISPR2-Cas system of subtype I-C spans approximately 8.0 Kb and consists of ten cas genes organized in an operon: three Cas1 subunits: SMU.1757c (94 aa), SMU.1755c (199 aa) and SMU.1754c (131 aa) and Cas2 (SMU.1753c, 97 aa) predicted to act in the adaptation step and genes encoding proteins that are predicted to form Cascade-like complexes involved in the interference stage: Cas3 (SMU.1764c, 131 aa), Cas4 (SMU.1758c, 214 aa), Cas5d (SMU.1763c, 249 aa) and 3 Cas genes representing the CRISPR subtype group Dvulg: SMU.1760c (291 aa), SMU.1761c (469 aa) and SMU.1762c (187 aa). An array of two 32 bp repeats interspaced by one 34 bp spacer was present downstream of its associated cas genes. Spacer 1 within CRISPR2 array matched 100% to a genomic nucleotide sequence of S. mutans LJ23 (Table S4). Using in silico analysis of CRISPR2 array, we found a leader sequence upstream of the first repeat possibly acting as a promoter for the transcription of the array CRISPR (data not shown). However, in accordance with previous findings (Horvath et al., 2009, Van der Ploeg, 2009) we observed that Cas1, required for spacer acquisition, is very unusual. Unlike most other Cas1 that typically are encoded by one gene (forming asymmetrical homodimers) (Wiedenheft et al., 2009, Babu et al., 2010, Kim et al., 2013) Cas1 of type I-C CRISPR2-Cas system has three apparent ORFs in UA159. Since both SMU.1757c (94 aa) and SMU.1754c (133 aa) appear to be too short to encode a functional assembly of Cas1, they might represent truncated regions that prevent the genes from functioning properly. Alternatively, they might result from annotation errors. Since in this locus only one spacer was identified, it is possible that CRISPR2 locus has lost its ability to incorporate novel CRISPR2 spacers as hypothesized previously (Horvath et al., 2009, Van der Ploeg, 2009). However, the in vivo activity of the CRISPR2 array remains to be elucidated.
Similar to cas genes from CRISPR1, our BlastP analysis of Cas proteins revealed high sequence similarity to nucleases, helicases and DNA repair proteins (Table S3) (Haft et al., 2005, Makarova et al., 2006). Promoter analysis of CRISPR1 cas genes and CRISPR2 cas genes revealed classical elements including a putative Pribnow box (-10 box; TATAAT and TAAaAaT, respectively) and the -35 element (TTaAaA and TTGACA, respectively) suggesting that cas genes within the same cluster are likely to be co-transcribed. We detected a putative binding site (TGTWAHNNNNTGTWAH) (Dubrac & Msadek, 2004) for the VicR response regulator protein, which is the regulatory component of VicR/K two component signal transduction system for CRISPR1 cas genes regulation (located at -114 to -136 and +84 to +105 positions from the putative transcriptional start site) and for CRISPR2 cas regulation (located at -44 to -66 positions from the putative transcriptional start site) This finding is consistent with results from our published (Senadheera et al., 2009) and unpublished work. It is likely that VicR may bind these target sequences to activate or repress the expression of cas genes (Dubrac & Msadek, 2004, Senadheera et al., 2005) (Figure 2-1A, Figure 2-1B).
Analysis of the promoter regions of cas genes identified the putative -10 box, -35 box, transcriptional start site (TSS), ribosome binding site (RBS) (all underlined in bold) as well as the putative VicR binding consensus sequence (TGTWAH-6/10 bp-TGTWAH) for cas gene regulation (all in italics).

2.4.2. Loss of the M102–specific CRISPR spacers or cas genes of both CRISPR-Cas systems have no effect on the phage resistance phenotype of S. mutans UA159

Since spacers 2 and 3 within the CRISPR1 array matched sequences of the M102 genome, we hypothesized that their presence might facilitate phage defense in S. mutans UA159. To test whether the M102-targeting spacers and CRISPR-Cas systems confer immunity against M102 phage infection, S. mutans UA159, OMZ381 (phage sensitive strain) and CRISPR-Cas deficient strains were assessed for phage resistance by challenging with the virulent phage M102 in both liquid growth assays and plaque formation assays. Deletion of cas genes of CRISPR1 and/or CRISPR2 or removal of spacer 2 and 3 within CRISPR1 array in S. mutans UA159 did not
affect phage sensitivity of UA159, since none of these strains were lysed by M102 in plate or liquid-based lytic assays (data not shown). However, in accordance with previous findings by Van der Ploeg (Van der Ploeg, 2009), the control OMZ381 strain displayed sensitivity to phage, as judged by its complete lysis in the presence of M102 (Figure S1 A,B). Since all mutant strains remained resistant to infection by phage M102, we concluded that yet unidentified and CRISPR-independent mechanism(s), are responsible for the M102 resistant phenotype displayed by UA159.

2.4.3. The type II-A CRISPR-Cas system prevents natural transformation by plasmids in S. mutans UA159

In addition to conferring phage immunity, CRISPR-Cas systems were shown to constitute an effective barrier against artificial means of transformation (e.g. electroporation) in several bacteria (Marraffini & Sontheimer, 2008, Deltcheva et al., 2011, Sapranaukas et al., 2011, Semenova et al., 2011). Further, it was shown that the introduction of engineered Streptococcus pyogenes CRISPR/Cas system reduced the transformation efficiency in a heterologous host, S. pneumoniae (Bikard et al., 2012). Zhang et al., (Zhang et al., 2013) found that the native meningoccal CRISPR-Cas system was able to prevent natural transformation of spacer-matching sequences, suggesting that it can limit the horizontal spread of virulence genes. These studies raised the question whether naturally transformable S. mutans employs CRISPR-Cas systems to form an effective barrier to limit foreign DNA acquisition by transformation. The S. mutans spacer sequences have potential matches to either phage M102 or other bacterial species present in the dental plaque (data not shown), however for simplicity we selected only candidate protospacers that fully matched spacer sequences from UA159 or had only a few mismatches, namely CR1SP2, CR1SP3, CR1SP6 within CRISPR1 locus and CR2SP1 within CRISPR2 locus of S. mutans UA159 (Figure 2-2). Previously it was shown that in type II systems, the PAM sequence is located at the protospacer 3’ end, whereas for the type I systems it is located at the 5’ end of the protospacer (Garneau et al., 2010, Sapranaukas et al., 2011, Jinek et al., 2012). To deduce
putative PAM motifs for *S. mutans* UA159, the identified 10 nt sequences located directly downstream and upstream of the protospacer sequences were aligned using WebLogo (data not shown). Sequence logos revealed that all potential natural targets except for the one matching spacer 2, with 3’ PAM (5’-TGGTGAATT-3’) downstream of the protospacer 2, have flanking sequences that deviate significantly from the PAM consensus identified in other *S. mutans* strains (Table S4) where 5’-NGG-3’ is located at 3’ end for type II-A system and 5’TTC-3’ at the 5’ end for type I-C system (Van der Ploeg, 2009, Fonfara et al., 2014). Since we were not able to identify the most common nucleotides that could represent the PAM sequence, we designed our plasmid constructs containing the protospacers matching CR1SP2, CR1SP3, CR1SP6 within CRISPR1 locus and CR2SP1 within CRISPR2 locus of *S. mutans* UA159 and included ten nucleotides on both sides of the protospacer. For comparison, we also cloned protospacers lacking flanking sequences (Figure 2-2B). The resulting pCG1 vectors were used in transformation assays into wild-type UA159 and its cas deletion mutant strains. The transformation frequency of each plasmid carrying a protospacer was compared to that of an empty vector (Figure 2-2C, Figure 2-2D). Empty pCG1 consistently exhibited transformation frequencies (%TF) of 2-4x 10^{-2} consistent with previous work (data not shown). Protospacers with identity to the spacers 2 and 3 within CRISPR1 had dramatically decreased transformation frequencies into wild-type UA159 compared to the native plasmid suggesting that the type II-A CRISPR/Cas system is functional against the plasmids carrying the protospacers matching CR1SP2 and CR1SP3. In contrast, plasmids carrying targets for spacer 6 within CRISPR1 and spacer 1 within CRISPR2 (both matching sequences from *S. mutans* genomes) exhibited transformation frequencies into wild-type strain comparable to those of the empty vector suggesting that they were recognized as self after introduction by transformation and were not targeted by CRISPR machinery (Figure 2-2C). Similar to previous work (Cady et al., 2012), we found that the degree of inhibition of transformation varied widely depending on the protospacer tested. For example, the protospacer from M102 phage with 86% identity to the CR1SP2 spacer caused approximately 1000 fold inhibition of transformation while the protospacer from M102 phage with 100% identity to the CR1SP3 resulted in only about 10 fold inhibition. Intriguingly, despite the previously
demonstrated importance of the PAM sequence in interference in type II-A systems (Deveau et al., 2007, Sapranauskas et al., 2011, Gasiunas et al., 2012, Jinek et al., 2012, Zhang et al., 2013), protospacers matching CRISPR spacers 2 and 3 cloned without any flanking sequences, consistently failed to yield transformants indicating that they likely elicited CRISPR interference (Figure 2-2D). Transformation frequency of ΔC1K and ΔC1KC2E in the presence of plasmids targeting CR1SP2 was restored to that of empty vector as well as in the presence of CR1SP3, however to a lesser extent, indicating that the CRISPR1-Cas function was abolished by deletion of cas genes.
Figure 2-2: The CRISPR1-Cas system of S. mutans UA159 provides immunity against plasmid transformation.
(A) Schematic representation of cloning vector pCG1 used for the construction of plasmids for the transformation interference assay. Plasmids for interference assays were produced by inserting a protospacer and 10 nts on both sides of the protospacers into pCG1 plasmid. (B) pCG1 constructs containing potential targets for different S. mutans UA159 spacers (2, 3, 6 within CRISPR1 array and 1 within CRISPR2 array). (C) and (D) pCG1 derivatives were tested by natural transformation assays using the wild-type S. mutans UA159, ΔC1K, ΔC2E and ΔC1KC2E strains. Transformation frequency was calculated as transformant CFU divided by the total number of viable cells. Results shown are representative of at least two independent experiments. *** indicates constructs showing targeting phenotype approximately over 10 fold change.

2.4.4. Purified recombinant SMU.1763c (Cas5d) has ribonuclease activity

Cas5d protein belongs to the Subtype I-C/Dvulg CRISPR-Cas system and a recent study of B. halodurans provided evidence that pre-crRNA processing, which is the key molecular event that initiates CRISPR interference, is mediated by Cas5d protein which after maturation process assembles with crRNA, Csd1 and Csd2 proteins to form an interference complex (Nam et al., 2012). Further it was also found that Cas5d proteins present in S. pyogenes and X. oryzae possess not only specific endoribonuclease activity for CRISPR RNAs but also nonspecific double-stranded DNA binding affinity suggesting that Cas5d may possess functions other than pre-crRNA processing in bacteria (Koo et al., 2013). Since we observed the production of the soluble S. mutans Cas5d protein (SMU.1763c) during recombinant expression in E. coli, we decided to investigate if SMU.1763c possesses RNA cleavage activity. The purified SMU.1763c was assayed for nuclease activity in a dose dependent manner using the $^{32}$P-labeled single stranded (ss) synthetic oligoribonucleotides as substrates. As shown in Figure 2-3A, purified Cas5d cleaved ssRNA substrates into small fragments. A control protein fraction purified from E. coli cells containing the empty plasmid was used to confirm that the observed RNase activity was associated with Cas5d and not with contaminating E. coli RNases (Figure 2-3B). Although the purified SMU.1763c was over 95% pure (Figure S2), at
this stage we cannot exclude co-purification of contaminant proteins, and we will address this in our future work using site-directed mutagenesis. Similar results were obtained when the DNase treated RNA from UA159 was used as a substrate (data not shown). These results indicated that Cas5d protein is a ribonuclease with activity against ssRNA.

Figure 2-3: Cleavage of the synthetic ssRNA substrates by the Cas5d protein from S. mutans UA159.

5'-[32P] labeled RNA1 or RNA2 (0.05 μm) were incubated in the absence or in the presence of 100 ng (lanes 1, 3) or 200 ng of Cas5d (SMU.1763c) (lanes 2, 4) or in the presence of 100 ng or 200 ng of purification product obtained from E.coli cells transformed with an empty p15TvL vector (lanes 5, 6) at 37°C for 30 min in the presence of 50 mM Tris-HCl (pH 7.0), 5 mm MnCl₂, 100 mM KCl and 1 mM DTT. Reaction products were separated on a 15% PAGE/8M urea gel and visualized by phosphorimaging. 39 nt RNA1 and 40 nt RNA2 were prepared using oligonucleotides 5’- AAAUACGUUUUCUCCAUUGUCAUAUUGCGCAUAAGUUGA and 5’- UUUCAAUUCCUUUJAGGAUAAUCUUGAAGAAGAGUUAA.
2.4.5. Cas5d cleaves cellular RNAs

Since we found that Cas5d protein was not involved in the stage that initiates interference of CRISPR immunity we hypothesized that the Cas5d protein is co-opted in other cellular processes where interaction with RNA substrates for nuclease activity is required. To identify potential RNA substrates targeted and cleaved by Cas5d protein, we used RNAs from mid-exponential phase cells of *S. mutans* UA159 in the presence (experimental) or absence (control) of this protein. DNase treated RNA samples were converted to cDNA and used for global microarrays analysis. Five transcripts were down-regulated over 1.8-fold by the addition of Cas5d protein (*P*<0.05) (Figure S3) and four were confirmed as significantly reduced by qRT-PCR (*p*<0.05) including a putative ABC transporter (SMU.995), a putative cell envelope protein (SMU.246c) and two hypothetical proteins (SMU.1502c and SMU.2075c). To confirm the ability of Cas5d to specifically target these substrates, we performed “*in vitro*” transcription analysis using full-length DNAs of SMU.995, SMU.1502c and SMU.385 (a random substrate from UA159) and T7 phage RNA polymerase. Cas5d cleaved all targets including the control SMU.385, suggesting that Cas5d did not exhibit sequence specificity in its RNA cleavage activity as anticipated from the microarray experiment (Figure 2-4). It is possible that those RNAs that appear as differentially regulated in the presence of Cas5d were likely more accessible to RNase activity as result of their abundance or possibly their location on the genome.
Figure 2-4: Cas5d (SMU.1763c) cleavage of RNA transcripts of SMU.995 and SMU.1502c generated by *in vitro* transcription.

### 2.4.6. Cas proteins are involved in sensitivity to DNA damage

Cas1 and Cas2 proteins have been predicted to be involved in DNA repair (Makarova *et al.*, 2002). Recent work in *E. coli* demonstrated that a mutant deficient in Cas1 had a DNA repair-deficient phenotype (Babu *et al.*, 2010). Since proteins encoded by SMU.1403c, SMU.1404c, SMU.1754c, SMU.1753c, SMU.1755c and SMU.1757c within CRISPR-Cas systems display sequence similarity to the Cas1 and Cas2 family proteins (Table S3), we tested *cas* deletion mutants under DNA damaging conditions to evaluate their putative roles in DNA repair. Hence, we examined the survival of *S. mutans* UA159 and mutant strains under DNA damaging conditions induced by 0.05 µg/ml mitomycin C (MMC) or UV irradiation. Survival of ∆C1S, and ∆C1SC2E cells was drastically altered when exposed to MMC or UV irradiation relative to wild-type strain suggesting a role in DNA repair (P<0.05) (Figure 2-5). Further, to validate that our phenotypic changes were caused only by the lack of *cas* genes, we carried out qRT-PCR reactions on their downstream genes. Deletion of *cas* genes within CRISPR1 (∆C1S strain), CRISPR2 (∆C2E strain) and CRISPR1+2 (∆C1SC2E) had no polar effects on the downstream genes as judged by expression analysis using UA159 and mutant strains (Figure S4). Although, the *S. mutans* CRISPR proteins do not play a prominent role in immunity
against M102, our results suggested a role for these proteins in conferring protection to DNA damaging agents.

Figure 2-5: Effects of 0.05 µg/ml MMC (A) or UV irradiation (B) on viability of *S. mutans* UA159 and mutant strains.

Results shown are representative of at least two independent experiments. Differences are statistically significant (P≤0.05, Student’s t test).

2.4.7. ΔC1S responds to oxidative, SDS, acid and high temperature stressors

Since previous transcriptome studies in *S. mutans* linked CRISPR-Cas systems with environmental stress tolerance (Senadheera *et al.*, 2009, Chatteraj *et al.*, 2010, Kajfasz *et al.*, 2010, Xie *et al.*, 2010, Kajfasz *et al.*, 2011, Liu *et al.*, 2011), we monitored growth kinetics of UA159 and cas deficient mutants under low pH (5.5), H₂O₂ (0.003%), SDS (0.004%), paraquat (25mM), NaCl (0.4M) and EtOH (2%), using an automated growth reader. ΔC1S strain grew faster under low pH and had a
higher yield compared with the wild-type strain, suggesting that cas genes associated with CRISPR1 locus have a negative role in the acid tolerance of S. mutans (Figure 2-6A). \(\Delta\)C1S strain was impaired in its ability to tolerate stresses induced by paraquat, \(\text{H}_2\text{O}_2\) and SDS suggesting the \(\Delta\)C1S played a role in responding to intracellular oxidative stress (paraquat), extracellular oxidative stress (\(\text{H}_2\text{O}_2\)), and cell membrane stress (SDS) (Figure 2-6B and Figure 2-6C). In the presence of other stresses induced by NaCl and EtOH, the \(\Delta\)C1S mutant grew similarly to UA159 (data not shown). While the \(\Delta\)C2E mutant did not reveal drastically altered growth rates compared with wild-type UA159 strain under any of the environmental stressors tested (Figure S5), \(\Delta\)C1SC2E mutant displayed growth phenotypes similar to those of the \(\Delta\)C1S strain (data not shown).
Figure 2-6: Growth kinetics of *S. mutans* UA159 and ΔC1S under varying stressors (A) at pH 7.0 and pH 5.5, (B) 25mM paraquat, 0.003% H$_2$O$_2$, or (C) 0.004% SDS.

Each point is the average of four independent OD values per sample. These results shown are representative of two independent experiments conducted with the mutant and UA159 parent strain.

Under high temperature stress, survival of all mutant strains was impaired ($P < 0.05$) compared with the wild-type, suggesting *S. mutans* cas genes have a role in temperature stress tolerance (Figure 2-7). Since the double mutant displayed sensitivity higher than either of the single mutants it is possible that both CRISPR/Cas systems may work cooperatively or sequentially to combat temperature stress.
2.4.8. Loss of VicK modulates the transcription of *S. mutans* cas genes

Search of the promoter regions located upstream of the CRISPR1 cas and CRISPR2 cas genes revealed the presence of putative binding sites for the VicR response regulator protein on their expression. To test the regulatory role of VicR/K signal transduction system in modulating the activity of cas genes we performed qRT-PCR using cDNAs isolated from a VicK-deficient mutant (SmuvicK) and UA159 strains and examined the expression of two candidate genes from each CRISPR operon. High expression levels of cas genes from both CRISPR operons were observed in UA159 cells suggesting that these genes are being expressed under mid-log growth phase. Loss of VicK caused over 2-fold downregulation of SMU.1753c and SMU.1755c expression from CRISPR2-Cas system suggesting that VicK mediated a positive regulatory role on their expression (Figure 2-8B). Conversely, SMU.1403c and SMU.1404c from the CRISPR1-Cas system were 2-fold upregulated by vicK deletion suggesting that VicK mediated a negative regulatory role on their expression (Figure 2-8A). Hence, it is possible that VicR/K may differentially regulate CRISPR systems to prevent or reduce their simultaneous expression.

Figure 2-7: Survival of *S. mutans* UA159 and mutant strains after exposure to 50°C temperature stress for 1 h.

Results represent mean CFU counts ± standard deviation. Differences are statistically significant (P ≤ 0.05, Student’s t test). These results shown are representative of two independent experiments conducted with the mutants and UA159 parent strain.
Figure 2-8: Expression of cas genes from the CRISPR1 (A) and CRISPR2 (B) operons.

RNA analysis from mid-logarithmic cultures of *S. mutans* UA159 and SmuvicK grown under regular or acidic conditions. Results are the average of triplicate samples from 3 independent experiments ± std error.

### 2.5. Discussion

*S. mutans* is one of several bacterial species known to be competent for HGT via natural transformation. Only a few phages are known to infect *S. mutans* and transformation is the key process used by *S. mutans* to acquire exogenous DNA. Frequent HGT occurs in *S. mutans* to promote homology based DNA repair, genetic diversity or other functions. However, the mechanisms that regulate the transfer, uptake and recombination of incoming DNA in naturally transformable *S. mutans* are
still poorly understood. We demonstrated that in S. mutans UA159, CRISPR/Cas systems do not play a prominent role in acquired resistance to M102 phage infection. Loss of M102-specific CRISPR spacers, or deletion of cas genes of CRISPR1-cas and or CRISPR2-Cas alone did not alter phage resistance of S. mutans UA159. These results agree with earlier observations where inactivation of Cas9 (signature protein for CRISPR-encoded interference in type II-A) in wild-type strain did not lead to loss of resistance to infection by M102 (Van der Ploeg, 2009). Based on the high variability of the CRISPR spacers (including M102 sequences) between S. mutans serotype c (Van der Ploeg, 2009), it is unlikely that CRISPR-Cas systems would be so widespread if they were unable to provide adaptive protection to their hosts. Probably, S. mutans has a variety of natural phage resistance mechanisms including restriction/ modification systems and/or CRISPR-Cas systems to target diverse steps of the phage life cycle to prevent M102 phages from attacking these genomes (Ajdic et al., 2002, Maruyama et al., 2009).

We also revealed that the native type II-A CRISPR1-Cas system of S. mutans UA159 is important for preventing natural transformation via plasmid DNA. Using transformation assays, pCR1SP2 and pCR1SP3 constructs which contained protospacer sequences matching spacers SP2 (86% sequence identity to M102) and SP3 (100% identity to M102) in the CRISPR1 locus, yielded drastically reduced TF compared to that of the empty plasmid. Consistent with previous reports, we found that CRISPR1-Cas machinery can tolerate a few nucleotide mismatches between spacer and protospacer at certain positions (Garneau et al., 2010, Gudbergsdottir et al., 2011, Manica et al., 2011, Sapranauskas et al., 2011, Semenova et al., 2011). As previously observed (Garneau et al., 2010, Cady et al., 2012), the degree of inhibition of transformation varied widely depending on the protospacer tested. The UA159 strain could be transformed with the protospacer with no mismatch (pCR1SP3) at higher frequencies than the protospacer with mismatches (pCR1SP2) suggesting that the CRISPR-Cas machinery is more permissive for pCR1SP3. Although it's still unclear, it might reflect a weak or altered interaction between the CRISPR-Cas system and plasmid DNA however that is yet to be elucidated. Intriguingly, effective interference was not observed with pCR1SP6 and pCR2SP1
constructs (100% identity to a *S. mutans* GS-5 spacer) and respectively (100% identity to a *S. mutans* LJ-23 spacer) possibly suggesting that crRNA transcripts complementary to these targets are only weakly expressed to produce interference. Since UA159 contains only one CRISPR2 spacer, further investigations are warranted to confirm that the observed phenotype is explained by the function of the CRISPR2-Cas machinery and rule out the lack of effective interference activity due to *cas* gene mutations (Horvath *et al.*, 2009, Van der Ploeg, 2009). Surprisingly, the presence or absence of flanking sequences within pCR1SP2 and pCR1SP3 had no effect on the ability to interfere with plasmid transformation. These results contrast to those data obtained using *in vitro* plasmid cleavage assays where dual-tracrRNA:crRNA guided Cas9 from *S. mutans* could efficiently cleave target DNA in the presence of a NGG sequence (Fonfara *et al.*, 2014). Consistent with our findings, in *S. thermophilus* it was also observed that plasmids carrying protospacers associated with consensus or with non-consensus (degenerate) PAMs could not be transformed into the corresponding plasmid-interfering strains whereas phages carrying the degenerative PAMs could infect the matching phage-insensitive mutants (Deveau *et al.*, 2007, Garneau *et al.*, 2010). The authors suggested that the tolerance of PAM degeneracy for CRISPR-Cas function could be due to the lower selective pressure for plasmids as compared to phages. Such an activity could theoretically produce a lower level of TF in the presence or absence of flanking sequences observed in our study. Alternatively, the presence of plasmids inside the cell could increase the expression of Cas proteins reflecting higher interference activity that might not require a PAM site (Makarova *et al.*, 2012).

Using purified recombinant protein, we also demonstrated that SMU.1763c possesses RNase activity against synthetic oligoribonucleotides and total RNA extracted from *S. mutans* UA159. Further, based on our *in vitro* transcription and DNA microarray studies, SMU.1763c had no obvious sequence preference in RNA cleavage. Recently, it was also reported that Cas5d ortholog from *B. halodurans* cleaves pre-crRNA by recognizing both the hairpin structure and the 3’ single stranded sequence in the CRISPR repeat region (Nam *et al.*, 2012). Based on these findings and the fact that we were not able to identify sequence specificity in SMU.1763c cleavage, we speculate that RNA secondary structure elements such as
stemloop are required for SMU.1763c to process RNA substrates in a sequence- and site-specific manner. Consistent to previous studies (Babu et al., 2010, Plagens et al., 2012), we found that CRISPR1 cas deficient mutant exhibited enhanced sensitivity to killing by mitomycin C (which inhibits growth by causing DNA cross linkage) (Tomasz, 1995), or UV irradiation (inhibits growth by causing large ds breaks). Additionally, CRISPR2 cas deficient mutant was not sensitive to DNA damage which is in perfect agreement with previous bioinformatic work where Cas1 associated with CRISPR2 appeared to be truncated (Horvath et al., 2009, Van der Ploeg, 2009). S. mutans possesses several DNA repair systems to support functions related to DNA protection or repair: RecA (Quivey et al., 1995), apurinic-apyrimidinic (AP) endonuclease (Hahn et al., 1999), or a UV repair excinuclease (uvrA) (Hanna et al., 2001). The nucleotide excision repair (NER) has been shown to be the major system for repairing damaged DNA caused by UV light and genotoxic agents such as MMC (Hanna et al., 2001). Although, Cas components of the CRISPR1-Cas system possibly act in the NER pathway in response to DNA damage caused by environmental stress, their specific role in this repair pathway remains to be elucidated.

Further, our transcriptional analysis identified that S. mutans cas genes were differentially expressed in the vicK deficient mutant. The VicR/K, one of 14 TCSTSs in S. mutans (Ajdic et al., 2002, Senadheera et al., 2005, Biswas et al., 2008), is comprised of a VicK histidine sensor kinase and an essential VicR response regulator. It was previously shown to be involved in biofilm formation, genetic competence, stress tolerance, bacteriocin production and cell viability (Senadheera et al., 2005, Senadheera et al., 2007, Senadheera et al., 2009, Senadheera et al., 2012). Based on our finding that VicR/K modulates the expression of cas genes, a role of CRISPR-Cas systems in contending with various environmental stressors was not surprising. In S. mutans UA159 it has been proposed that dual crRNA: tracrRNA participates in type II-A CRISPR function (Deltcheva et al., 2011), therefore it raised the question whether the entire CRISPR1-Cas system is necessary to mediate the stress responses observed for the ΔC1S mutant. Using Northern blot analysis, the expression of the tracrRNA under stress conditions was
noted for the wild type (Figure S6). Although our observations may indicate that tracrRNA possibly with crRNA mediate stress response in vivo, our assumption warrants further investigations. We also noted that ΔC2E did not share the same sensitivity to the tested stressors as that seen with ΔC1S suggesting that CRISPR1 cas and CRISPR2 cas genes are differentially regulated to function independently within the environment. In fact, this is the case in S. thermophilus where CRISPR-Cas systems where observed to function independently (Carte et al., 2014). Furthermore, several transcriptome studies revealed that deletion of virulence or global regulatory genes of S. mutans (including genes involved in stress response) differently affected transcription of cas genes within CRISPR-Cas systems suggesting different roles for cas genes within the cell (Nascimento et al., 2008, Chattoraj et al., 2010, Kajfasz et al., 2011). As already proven in other systems (Shinkai et al., 2007, Agari et al., 2010, Pul et al., 2010, Medina-Aparicio et al., 2011), it is possible that different regulatory systems, in addition to the VicR/K interact with S. mutans CRISPR/Cas systems to mediate gene expression in response to cues such as oxidative stress and cell membrane changes or alterations in the internal pH of the cell. The presence of diverse and complex regulatory strategies to modulate the CRISPR/Cas activity might also explain why some phenotypes displayed by ΔC1S and ΔC2E are not compatible with those of a VicK mutant as shown previously (Senadheera et al., 2009). Currently, studies are underway to examine whether VicR exerts a direct regulatory role on the transcription of CRISPRs by binding to their respective promoters. Complementation studies of cas mutations in S. mutans and experiments with the overexpressing VicR strains should be considered in future studies. Examination of other regulatory systems in S. mutans on CRISPRs transcription and function would be also of interest.

In summary, our data provide the first experimental evidence that CRISPR1-Cas system of S. mutans UA159 play novel roles in resistance against incoming plasmids that carry matching proto-spacer sequences and stress response. Given their multiple roles in the cell physiology, the type II-A system may prove to be useful
target for therapeutics to diminish the virulence and also to influence S. mutans species to prevent the uptake and dissemination of antibiotic resistance genes.

2.6. Acknowledgements

We thank Richard Mair for assistance with bioinformatic analyses, Deanna Del Re for constructing the ΔC2E mutant, Greg Brown for cloning cas5d gene. D.G.C. is a recipient of the NIH Grant (R01DE013230-03) and CIHR Grant (MT-15431) and MAS is a recipient of a CIHR Strategic Training Fellowship in Cell Signaling in Mucosal Inflammation and Pain and Harron Scholarship. This work was also partially supported by the Government of Canada through Genome Canada and Ontario Genomics Institute (2009-OGI-ABC-1405) and Ontario Research Fund (ORF-GL2-01-004).
2.7. References


Bolotin A, Quinquis B, Sorokin A & Ehrlich SD (2005) Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151: 2551-2561.


2.8. SUPPLEMENTARY MATERIAL

Table S1: Primers used in this study (nucleotide sequence from 5' to 3')

PCR-ligation mutagenesis

ΔC1S/K-P1  5'ATTATGGAGCCTTTTCGTC3'
ΔC1S/K-P2  5'GG^CGCGCTAACTCAGATGGTTCAGACTC3'
ΔC1S/K-P3  5'GGCCGG^CCGTTTATTTCTTGAAACC3'
ΔC2S/K-P4  5'AACAGAGCCTAACACTGCG3'
ΔC2E-P1    5'TTGTCTTGCAGGTTCAGTC3'
ΔC2E-P2    5'GG^CGCGCATCTTCCCTTGCCTTGTG3'
ΔC2E-P3    5'GGCCGG^CCGGAGATGACGGTTATGACC3'
ΔC2E-P4    5'GCAAGGAACAAATCAGCTC3'
ΔCR1SP13E-P1 5'TGCCAGTGATGAAACGAC3'
ΔCR1SP13E-P2 5'GG^CGCGCCATCTTCCCTTGCCTTGTG3'
ΔCR1SP13E-P3 5'GGCCGG^CCCTCTAAAACCTCGTGGACC3'
ΔCR1SP13E-P4 5'TGGATGGGAAGAAGCTTC3'

Erm-For    5'GG^CGCGCCCGGGGCGCAATAATACGTGGACTG3'
Erm-Rev    5'GGCCGG^CCAAGCTGCGAGCTGAGAAT3'
Spec-For   5'GG^CGCGCCGCGCACTAATAACGAAGTGGACTG3'
Spec-Rev   5'GGCCGG^CCGCTGCGGCAAGTAGTTTCCAGATATGGAG3'
Kan-For    5'GG^CGCGCCGATAAACCAGCGAACCATT3'
Kan-Rev    5'GGCCGG^CCAAGCTTTTATGACATCTA3'

1 Ascl restriction sites are in boldface, and Fsel restriction sites are underlined
Cloning primers

SMU.1763c-For
5’CAAGCTTCGTCATCAGCCTTCCCCTTTCTCCATAG3’

SMU.1763c-Rev
5’TGTATTTCAGGTGATTTCTATAGTAAGAATTTC3’

qRT-PCR primers

SMU.1403c-For 5’AAAGACCAGGCTGGAATC3’
SMU.1403c-Rev 5’CGCTTGAAGGAGGATAATC3’
SMU.1404c-For 5’TACGGTCAAAAGGCTCTAAGTC3’
SMU.1404c-Rev 5’CTGGCAACAATCATCGCTC3’
SMU.1753c-For 5’CGGATGCTGTCTGGATTTG3’
SMU.1753c-Rev 5’CGCAAAACTCTCTGTGATTATGG3’
SMU.1754c-For 5’ATAACACTGCCATTTTC3’
SMU.1754c-Rev 5’CCATACAGATAGACCTGG3’
SMU.1755c-For 5’TAGCCACCATTTCAGAG3’
SMU.1755c-Rev 5’GTATCTATTACGGTTTCG3’
SMU.1757c-For 5’CAAAATCGTCCCTGAGGCG3’
SMU.1757c-Rev 5’TGTTCATCGTCTCCCTATCGG3’
SMU.1758c-For 5’CCAGTCTCAATAGGACACCCAAAC3’
SMU.1758c-Rev 5’TGGAAGGCAAAGGAGAAAGTG3’
SMU.1761c-For 5’TGGATTGGTAGTGGCTTCCATC3’
SMU.1761c-Rev 5’CGCTGTACGGTGGACAGAG3’
SMU.995-For 5’TCGAGAGGAATGCCTATG3’
SMU.995-Rev 5’CCGATGCCATAAGTAG3’
SMU.1502c-For  5'CTAATGCCGTAATCGCTG3'
SMU.1502c-Rev  5'TTATCCAACCGTCAGTGG3'
SMU.2075c-For  5'GGGACAAGAACACAAAAAGGC3'
SMU.2075c-Rev  5'TGGCAAAGAAGGAGTGACG3'
SMU.246-For  5'CGCCATTATCAGACGAACC3'
SMU.246-Rev  5'TGCCATAGACGACTAAACAGC3'
SMU.1802c-For  5'GTCATACTGTCCAGCCATCC3'
SMU.1802c-Rev  5'TGATAGCCTGCGAAAGC3'
SMU.1400-For  5'ACTTTCTAACATCTGACG3'
SMU.1400-Rev  5'GGAAGAGGAAGGTATTG3'
SMU.1752-For  5'CCTGCTTGAAGACCAAAC3'
SMU.1752-Rev  5'CAAGGGAAAGAGGCTCC3'
GtfB-For  5'ACACTTTCCGGGTGCTT3'
GtfB-Rev  5'GCTTAGATGTCATTGCTT3'
16S rRNA-For  5'CTTACCAGGTCATTGAC3'
16S rRNA-Rev  5'ACCCGAGATCTCAGAGAG3'

Primers for in vitro transcription
SMU.995-For  5'TAATACGACTCTATAGGCCATGGAGATCGAGGAGAATGCCTATG 3'
SMU.995-Rev  5'CCGATGCCAATAAGTAGG 3'
SMU.1502-For  5'TAATACGACTCTATAGGCCAGGAGCCGTCCACAGAAATCTC 3'
SMU.1502-Rev  5'TGATATTAGGGGGTAGCG 3'
Primers for Northern blot

<table>
<thead>
<tr>
<th>Oligo name For/Rev</th>
<th>Oligonucleotide Sequence ²  (5’ → 3’)</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR1SP2 (with flank)-For</td>
<td>cgggagctcGGCATGAAGCGATTGAACCAGCTAGCGCAGTTAGTGC TCTTGGTGTAATTgagctccgg</td>
<td>pCR1SP2 with PAM</td>
</tr>
<tr>
<td>CR1SP2 (with flank)-Rev</td>
<td>cgggagctcAATTACACCAAAGAGCACTAAGCTAGCTGGTTCAA TCGCTTCATGCGagctcccgg</td>
<td>pCR1SP2 without PAM</td>
</tr>
<tr>
<td>CR1SP2 (no flank)-For</td>
<td>cgggagctcGGATTGAACCAGCTAGCGCAGTTAGTGCTCTTgagctccgg</td>
<td>pCR1SP2 without PAM</td>
</tr>
<tr>
<td>CR1SP2 (no flank)-Rev</td>
<td>cgggagctcAGAGCAGTAAGCTAGCTGGTTCAATCATGagctcccgg</td>
<td>Irrelevant protospacer</td>
</tr>
<tr>
<td>CR1SP-For</td>
<td>cgggagctcTAGAATTTCTACACTCGAGAGCTAGATGAGAGGAGGCTAGagctcccgg</td>
<td>Irrelevant protospacer</td>
</tr>
<tr>
<td>CR1SP-Rev</td>
<td>cgggagctcATCAGCCTCTTCATCTACAGACACAGTTAGAATGAGTTCTAGAAATTTCTagctcccgg</td>
<td>Irrelevant protospacer</td>
</tr>
<tr>
<td>CR1SP3 (with flank)-For</td>
<td>cgggagctcCGTTGCCCCCTGCCTAAAGCTGGTGTCATCATAGGT AGTTGAATTTTGagctcccgg</td>
<td>pCR1SP3 with PAM</td>
</tr>
<tr>
<td>CR1SP3 (with flank)-Rev</td>
<td>cgggagctcGAATTTGACACTGAGATGACACAACAGTTTAGTACAGGTA GGAAGGCAAGGagctcccgg</td>
<td>pCR1SP3 with PAM</td>
</tr>
<tr>
<td>CR1SP3 (no flank)-For</td>
<td>cgggagctcCGCTAAAGCTGGTGTCATCATAGTATGagctcccgg</td>
<td>pCR1SP3 without PAM</td>
</tr>
</tbody>
</table>

² SacI restriction sites are indicated by lower case.
<table>
<thead>
<tr>
<th>CR1SP3 (no flank)-Rev</th>
<th>cggagctcCTAACTATGATGACACAACAGCTTTTAGCGgagctcccg</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR1SP6 (with flank)-For</td>
<td>cggagctcGTTCAAAAACATGTGAAAAAGAAAGTAACTACTACATTTGAGTTTTagagctcccg</td>
</tr>
<tr>
<td>CR1SP6 (with flank)-Rev</td>
<td>cggagctcGCTCTAAAAACTCAATGTAGTAGTTACTTTTCATACTACATTTTGAGTTTTAgagctcccg</td>
</tr>
<tr>
<td>CR1SP6 (no flank)-For</td>
<td>cggagctcATGTGAAAAAGAAAGTAACTACTACATTTTACGagctcccg</td>
</tr>
<tr>
<td>CR1SP6 (no flank)-Rev</td>
<td>cggagctcTCAATGTAGTAGTTACTTTTCATTTTACATTGagctcccg</td>
</tr>
<tr>
<td>CR2SP1 (with flank)-For</td>
<td>cggagctcGGATTGAAATTGAAATCAATCGGATGATGAAATGAGTTTTCAGAGCTCCGagctcccg</td>
</tr>
<tr>
<td>CR2SP1 (with flank)-Rev</td>
<td>cggagctcAGGGTGCGACAAAACACTATTTTACTATACTATGCGATTGATTCACTTTTACATGGagctcccg</td>
</tr>
<tr>
<td>CR2SP1 (no flank)-For</td>
<td>cggagctcTGAAATCAATCGGATGATGAAATGAGTTTTCAGAGCTCCG</td>
</tr>
<tr>
<td>CR2SP1 (no flank)-Rev</td>
<td>cggagctcAAAAACTCCATTTTCATATGCGATTTAGTTTACATGGagctcccg</td>
</tr>
</tbody>
</table>

pCR1SP6 with PAM
pCR1SP6 without PAM
pCR2SP1 with PAM
pCR2SP1 without PAM
<table>
<thead>
<tr>
<th>Strain</th>
<th>CRISPR1/Cas system</th>
<th>Gene</th>
<th>Protein (putative function)</th>
<th>Homolog (NCBI accession no.; %protein identity)</th>
<th>CRISPR2/Cas system</th>
<th>Gene</th>
<th>Protein (putative function)</th>
<th>Homolog (NCBI accession no.; %protein identity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMU.1405c</td>
<td>Cas9 (nuclease)</td>
<td>S. anginosus (ZP07865023); 73</td>
<td>SMU.1764c</td>
<td>Cas3 (helicase)</td>
<td>S. sanguinis (EGJ38121); 77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMU.1404c</td>
<td>Cas1 (nuclease/DNA repair)</td>
<td>S. maceae (ZP09135329); 97</td>
<td>SMU.1763c</td>
<td>Cas5d (nuclease)</td>
<td>S. sanguinis (EGJ38120); 86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMU.1403c</td>
<td>Cas2 (nuclease/DNA repair)</td>
<td>S. gallecticus (BAK28299); 89</td>
<td>SMU.1762c</td>
<td>Csd1 (nuclease)</td>
<td>S. parasanguinis (YP004621356); 46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMU.1402c</td>
<td>Csn2 (nuclease)</td>
<td>S. maceae (ZP09134698); 88</td>
<td>SMU.1761c</td>
<td>Csd1 (nuclease)</td>
<td>S. costellatus (EID21028); 74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMU.1399-SMU.1398</td>
<td>CRISPR array</td>
<td></td>
<td></td>
<td></td>
<td>SMU.1760c</td>
<td>Csd2 (nuclease)</td>
<td>S. sanguinis (EGJ38118); 71</td>
<td></td>
</tr>
<tr>
<td>UA159</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SMU.1758c</td>
<td>Cas4 (exonuclease)</td>
<td>S. agalactiae (CDD21808); 85</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SMU.1757c</td>
<td>Cas1 (nuclease/DNA repair)</td>
<td>S. sanguinis (EGJ38116); 94</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SMU.1755c</td>
<td>Cas1 (nuclease/DNA repair)</td>
<td>S. pseudoporcinus (ZP07823512); 77</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SMU.1754c</td>
<td>Cas1 (nuclease/DNA repair)</td>
<td>S. costellatus (EID21029); 96</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SMU.1753c</td>
<td>Cas2 (nuclease/DNA repair)</td>
<td>S. costellatus (EID21004); 94</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SMU.1753c - SMU.1752c</td>
<td>CRISPR array</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S4: Alignment of potential natural targets for *S. mutans* UA159 spacers to identify possible PAM motifs.

4 out of 7 spacers matched phage M102 sequences or other *S. mutans* genomes. Protospacers and 10 flanking nts (on both sides) from the representative target genomes were aligned. The PAM regions in the putative targets with consensus nucleotides are shown in bold and with non-consensus nucleotides in italics. Spacers with possible targets in *S. mutans* genomes are shown in red, and spacers with possible phage-like targets are in blue.

<table>
<thead>
<tr>
<th><em>S. mutans</em> CRISPR spacer number</th>
<th>Representative target genome</th>
<th>5' flank</th>
<th>(Protospacer) Potential donor sequence</th>
<th>3' flank</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR1SP2</td>
<td>M102 phage</td>
<td>GGCATGAAGC</td>
<td>GATTGAACCAGCTAGCGCAGTTAGTG CTCT</td>
<td>TGGTGTAATT</td>
</tr>
<tr>
<td>CR1SP3</td>
<td>M102 phage</td>
<td>CTTTAAAGTT</td>
<td>GATTGATACTACTGTGTTGTGCGAAAAAT CGC</td>
<td>TTCCCGTTGC</td>
</tr>
<tr>
<td>CR1SP6</td>
<td><em>S. mutans</em> GS-5</td>
<td>GTTCCAAAAC</td>
<td>ATGTGAAAAAGAAAGTAACTACATT TGA</td>
<td>GTTTAGACG</td>
</tr>
<tr>
<td>CR2SP1</td>
<td><em>S. mutans</em> LJ-23</td>
<td>AGGGTGCGAC</td>
<td>TGAATCAATCGCATAGTATATGAAAT GAGTTTT</td>
<td>GTCGCACCCCT</td>
</tr>
</tbody>
</table>
Figure S1 (A and B): Growth curves of *S. mutans* OMZ381 and UA159 in the presence of M102.

Each point is the average of four independent OD values per sample. Results shown are representative of at least three independent experiments conducted with OMZ381 and UA159 strains.
Figure S2: Coomassie-blue staining of SDS-PAGE gel showing the purified Cas5d (SMU.1763c) protein.
Figure S3: Fold expression of genes likely targeted for cleavage by SMU.1763c Cas5d protein.

Cultures were grown in THYE until mid-log phase (OD_{600nm} ~ 0.4), RNAs were extracted and incubated in the presence or absence of SMU.1763c Cas protein. Resulting samples were used for microarray analysis. Affected transcripts were confirmed using qRT-PCR. Expression data represent normalized values of three independent experiments, each amplified in triplicate. Error bars represent ± std error.
Figure S4: Agarose gel electrophoresis of the RT-PCR amplification products of downstream genes SMU.1752 and SMU.1400 respectively, in S. mutans wild-type strain and mutants defective in the cas genes.

The constitutively expressed 16S rRNA and gtfB genes served as controls. For each strain cDNA samples were derived from three independent experiments.
Figure S5 (A, B and C): Growth kinetics of *S. mutans* UA159 and ΔC2E under varying stressors (A) at pH 7.0 and pH 5.5, (B) 25mM paraquat, 0.003% H₂O₂, or (C) 0.004% SDS.

Each point is the average of four independent OD values per sample. These results shown are representative of two independent experiments conducted with the mutant and UA159 parent strain.
Figure S6 (A and B): Northern blot analysis of tracrRNA expression in *S. mutans* UA159 at different time points of growth and various stressors suggest processing of tracrRNA.

A) The expression of the tracrRNA at an OD$_{600}$ of 0.1 and OD$_{600}$ of 0.4 showed three transcripts (~100 bp, ~90 bp and ~80 bp) whereas at an OD$_{600}$ of 1.2 it displayed two products only. Under CSP or oxidative stress three products were visualized. Note as well that under heat stress a fourth faint ~200 bp band was observed. B) The 5S rRNA was used as loading control. Data are representative of three independent experiments.
Chapter 3: Insights into molecular function of the CRISPR-associated Cas5d protein SMU.1763c from *Streptococcus mutans*

N. Beloglazova\(^1\), M. A. Serbanescu\(^2\), R. Flick\(^1\), K. Krastei\(^2\), S. Lemak\(^1\), D.G. Cvitkovitch\(^2\) and A.F. Yakunin\(^*1\)

\(^1\)Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Canada

\(^2\)Dental Research Institute, Faculty of Dentistry, University of Toronto, Toronto, Canada

This manuscript is currently under preparation for submission to the Journal of Biological Chemistry

Running title: *S. mutans* Cas5d ribonuclease

Key words: Cas5d, endoribonuclease activity, type I-C/Dvulg, *S. mutans*, stress response

*To whom correspondence should be addressed

Tel.: 416-978-4013

Fax: 416-978-8528

E-mail: a.iakounine@utoronto.ca.
3.1 Abstract

**Aims:** Structural, biochemical and genetic analyses were used to provide novel functional insights into the SMU.1763c protein (Cas5d), a putative endoribonuclease associated with the CRISPR system I-C (CRISPR2) of *Streptococcus mutans*.

**Methods and Results:** Using stringent reaction conditions we found that the purified protein SMU.1763c is an endoribonuclease that exhibits metal-independent RNase activity against structured single-stranded RNA substrates. The crystal structure of the SMU.1763c protein has been determined in an apo-form and revealed a ferredoxin domain-based architecture and provided insight into the molecular mechanisms of substrate binding and cleavage. Structure-based site-directed mutagenesis of SMU.1763c identified a catalytic triad formed by Lys-120, His-121 and Tyr-50 residues which are important for RNA cleavage. The *S. mutans* SMU.1763c deletion strain was generated, and the effect of gene disruption on the sensitivity to various cell wall inhibitors that perturb cell membrane integrity was tested by microtiter plate-based MIC and MBC assays. The SMU.1763c mutant strain (ΔCas5d) showed sensitivity comparable to the parental strain UA159 when exposed to compounds that compromise cell envelope integrity. Furthermore, the survival of the mutant strain was not severely affected by exposure to high temperature shock (50°C). **Conclusions and Significance of Study:** The SMU.1763c belongs to the Cas5d family of CRISPR RNA endonucleases acting on structured RNA substrates that is likely to be involved in CRISPR RNA processing but not in sensing cell envelope stress or preserving cell integrity.
3.2 Introduction

The clustered regularly interspaced short palindromic repeats (CRISPR) and associated proteins (Cas) constitute microbial defense systems against foreign DNA derived from phages and plasmids (Barrangou et al., 2007, Van Der Oost et al., 2009, Horvath & Barrangou, 2010, Wiedenheft et al., 2012, Zhang et al., 2013). CRISPR consist of short repetitive sequences (repeats) interspaced by sequences derived from invading nucleic acid (spacers) (Bolotin et al., 2005, Mojica et al., 2005, Pourcel et al., 2005). CRISPR associated cas genes are also found adjacent to CRISPR loci and encode Cas proteins with a variety of nucleic acid related functions (Haft et al., 2005, Makarova et al., 2006). CRISPR immune systems function in three stages: adaptation, expression and interference (Haft et al., 2005, Makarova et al., 2006, Barrangou et al., 2007, Brouns et al., 2008, Carte et al., 2008, Hale et al., 2009, Haurwitz et al., 2010, Deltcheva et al., 2011). During the adaptation stage, Cas proteins recognize foreign DNA and incorporate short sequences (25–50 nt) of this DNA (protospacers) into the CRISPR locus of the host genome as new spacers. During the expression stage, the CRISPR locus is transcribed to generate and process the CRISPR RNA (crRNAs). Finally, in the interference stage the invading nucleic acid is destroyed using the processed crRNA in some form of effector complex containing Cas proteins. Based on the set of Cas proteins and their mode of action, the CRISPR-Cas systems have been classified into three main types Type (I-III) (Makarova et al., 2011). The type I system, the most widespread system, is comprised of six subtypes (I-A to I-F) based on the presence of subtype specific genes. Cas3, the signature protein interacts with complexes of subtype-specific Cas proteins that process pre-crRNA and assembles into a cascade-like interference complex to recognize and cleave DNA targets (Brouns et al., 2008, Sorek et al., 2008, Makarova et al., 2011). The CRISPR II systems (II-A to II-B), found in bacteria only, contain Cas9 protein, as a signature protein and is guided by a dual- trans-activating crRNA (tracrRNA):crRNA molecule and endogenous RNAIII for processing pre-crRNA into crRNA and cleaving the target (Deltcheva et al., 2011, Chylinski et al., 2013). The type III systems (subtype III-A and III-B) contain the Cas6-RAMP module (Csm in subtype III-A and Cmr in subtype III-B systems). Cas6 protein is
responsible for the processing step of pre-crRNA into crRNAs whereas the interference complexes (RAMP module) target nucleic acid for cleavage (Carte et al., 2008, Scholz et al., 2013). Consistent with other CRISPR systems, subtype III-A targets DNA invader, whereas the subtype III-B is unique, targets RNA rather than DNA (Marraffini & Sontheimer, 2008, Hale et al., 2009). In addition to their canonical function in defense against foreign nucleic acid, CRISPR-Cas roles in various aspects of bacterial physiology are now being uncovered with CRISPR/Cas systems having been linked to alternative functions including gene regulation, virulence and bacterial stress responses (Louwen et al., 2014).

*Streptococcus mutans*, a normal resident of dental plaque, is considered one of the primary oral opportunistic pathogens that promotes the development of dental caries in humans (Loesche, 1986). Life in the oral cavity is typically characterized by periodic shifts in the availability of nutrients, pH, oxygen tension, the presence of bacteriocins and antimicrobial compounds; all of which strongly influence the survival of *S. mutans* within the dental biofilms (Lemos & Burne, 2008). The *S. mutans* strain UA159 contains two CRISPR loci: a type II-A CRISPR-Cas system and a type I-C CRISPR-Cas system. The type I-C CRISPR locus has 1 spacer and 2 repeats and 10 *cas* genes which encode six core Cas proteins Cas1 (three copies), Cas2, Cas3 and Cas4 and four non-core Cas proteins Cas5d, Csd1 (two copies), Csd2. We and others have previously provided evidence that type I-C CRISPR/Cas system plays a role in environmental stress tolerance (Ahn et al., 2007, Kajfasz et al., 2010, Xie et al., 2010, Kajfasz et al., 2011, Liu et al., 2011, Serbanescu et al., 2015). A mutant strain lacking all of *cas* genes associated with the type I-C CRISPR-Cas system displayed increased sensitivity to heat shock (Serbanescu et al., 2015). Moreover, a recent transcriptome comparison by Chattoraj et al. (Chattoraj et al., 2010) between a ClpP mutant (ClpP protease plays a vital role in adaptation under stress conditions induced by high temperature, low pH, cell membrane antibiotics, DNA damage conditions) and its UA159 progenitor strain identified over 100 ClpP-dependent genes in *S. mutans*, including several stress-related genes, genes encoding bacteriocin-related peptides and many transcription factors. These authors also reported that ClpP deficiency increased the expression of *cas* genes associated with
the type I-C CRISPR/Cas system (including SMU.1763c gene), thus implying a role for the cas genes in the cell envelope of \textit{S. mutans}. Similarly, another microarray analysis of \textit{S. mutans} UA159 wild-type and its VicK-deficient mutant (sensor kinase of the VicRK signaling system (a prominent signaling system involved in stress responses) revealed down regulated expression of all \textit{cas} genes within type I-C system in the mutant when compared with the wild-type strain (unpublished data). Kajfasz \textit{et al.}, also provided evidence that there might be a more general role of type-I-C CRISPR-Cas system in bacterial physiology in \textit{S. mutans} (Kajfasz \textit{et al.}, 2010, Kajfasz \textit{et al.}, 2011). Deletion of \textit{spxA} and \textit{spxB} from the \textit{spx} operon, which is involved in the regulation of the stress response, led to an induction of transcription of the \textit{cas} genes associated with type I-C system (including the SMU.1763c gene) (Kajfasz \textit{et al.}, 2010). Further it was also proposed that all of the \textit{cas} genes within the Type I-C CRISPR-Cas system including \textit{cas5d} were upregulated during biofilm formation in the presence of starch or sucrose (Ahn \textit{et al.}, 2007). Another transcriptome analysis of a mixed culture with \textit{S. mutans}, \textit{S. gordonii} and \textit{Vibrio parvula} revealed differential expression of \textit{cas} genes belonging to the type I-C system (Liu \textit{et al.}, 2011). Although these findings suggested that type I-C CRISPR-Cas components are induced in response to various cellular stimuli, the molecular determinants within this system responsible for the regulation of bacterial physiology in \textit{S. mutans} are still unknown.

Components of the type I-C system have been studied in more detail in the soil bacterium \textit{Bacillus halodurans} (Garside \textit{et al.}, 2012, Nam \textit{et al.}, 2012, Punetha \textit{et al.}, 2013). The \textit{B. halodurans} Cas5d has been shown to mediate the pre-crRNA maturation, a key molecular event that initiates the CRISPR interference. Cas5d catalyzed sequence- and structure-specific recognition and cleavage of long primary CRISPR transcripts (pre-crRNA) into the mature crRNAs (Garside \textit{et al.}, 2012, Nam \textit{et al.}, 2012, Punetha \textit{et al.}, 2013) (Nam \textit{et al.}, 2012). Recent structural studies revealed that Cas5d possesses a single RNA recognition motif (RRM) domain that is able to perform both RNA recognition and cleavage (Garside \textit{et al.}, 2012, Nam \textit{et al.}, 2012, Koo \textit{et al.}, 2013). Like other subtype-specific RNases, Cas5d was found to exhibit RNase activity that was independent of divalent metal ions (Nam \textit{et al.},
It was further found that after pre-crRNA processing, Cas5 assembled with crRNAs, Csd1, and Csd2 proteins to form a multi-subunit interference complex that initiates the process of binding to the complementary dsDNA (Nam et al., 2012). This enables Cas3 to be co-opted for dsDNA degradation (Nam et al., 2012). A recent study has also shown that the B. halodurans Cas5d and Csd1, but not Csd2, in addition to having RNase activity, also exhibited a metal-dependent DNase activity without specificity towards the DNA substrates, suggesting that Cas5d and Csd2 may be coopted in adaptation and interference stages where interaction with DNA is involved (Punetha et al., 2013). Structural and biochemical studies of Cas5d proteins from Thermus thermophilus and Mannheimia succiniciproducens have also revealed the presence of sequence-specific RNase activity that cleaved CRISPR repeats and was thus responsible for processing of pre-crRNA (Garside et al., 2012). The molecular and physiological significance of Cas5d in pathogenic bacteria has also been explored. Crystal structures of Cas5d from the human pathogen S. pyogenes and the plant pathogen Xanthomonas oryzae revealed that the two Cas5d proteins possessed both specific endoribonuclease activity against CRISPR RNAs and nonspecific dsDNA binding activity (Koo et al., 2013).

To gain insight into the possible role of Cas5d protein (SMU.1763c) in S. mutans, we sought to conduct a characterization of SMU.1763c using biochemical, genetic and structural approaches. Previously, we showed that the SMU.1763c protein from S. mutans possesses ribonuclease activity against synthetic RNA substrates and total RNA extracted from S. mutans (Serbanescu et al., 2015). Here we report that under stringent reaction conditions (low concentrations of substrate and protein) purified SMU.1763c cleaves RNA with a preference to structured RNAs. We also present the crystal structure of SMU.1763c protein, which revealed a ferredoxin domain-based architecture. We also identified a catalytic triad formed by Tyr-50, Lys-120 and His-121 residues. Using a SMU.1763c knockout strain (∆Cas5d) we further analyzed if the enzymatic activity of SMU.1763c protein contributes to the tolerance of S. mutans to a variety of cell envelope targeting antibiotics or survival after thermal shock. MIC/MBC assays conducted with lipid II inhibitors (vancomycin and bacitracin) as well as non-lipid inhibitors (penicillin G and oxacillin) revealed no
drastic effect on the sensitivities of cas5d mutant relative to parent strain. In addition, the SMU.1763c deletion mutant exhibited similar survival after heat shock as that seen with UA159 parent strain.

3.3 Material and Methods

Bacterial strains

Bacterial strains used in this study are listed in Table 3-1. The wild type S. mutans UA159 and the SMU.1763c deletion strain (ΔCas5d) were routinely grown in solid/liquid Todd-Hewitt-Yeast Extract (THYE) medium and incubated at 37°C in air with 5% CO₂. When required, erythromycin (10 μg/ml) was added in the medium for the selection and growth of the mutant strain. To delete the SMU.1763c gene in strain UA159 we used a PCR ligation mutagenesis and allelic replacement strategy (a non-polar erythromycin (Erm) resistance marker) as described previously (Lau et al., 2002). The primers used for mutant construction and confirmation are listed in Table 3-1 DNA sequencing was also performed to further confirm correct in-frame insertion of the antibiotic cassette into the target gene.

Table 3-1: Bacterial strains and primers used in this study.

<table>
<thead>
<tr>
<th>Primer</th>
<th>5′-3′ Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMU.1763c-P1</td>
<td>GCAAGATTTGTAAGAGATGCC</td>
<td>Cas5d mutagenesis</td>
</tr>
<tr>
<td>SMU.1763c-P2</td>
<td>ggcgagccGATTTCACTCGCCCATAAGACAC</td>
<td>Cas5d mutagenesis</td>
</tr>
<tr>
<td>SMU.1763c-P3</td>
<td>gcggcgcccAATCCTGTCTTCTTTCTCGC</td>
<td>Cas5d mutagenesis</td>
</tr>
<tr>
<td>SMU.1763c-P4</td>
<td>GAAAGGGGAAGGTATAGGG</td>
<td>Cas5d mutagenesis</td>
</tr>
<tr>
<td>Erm-F</td>
<td>gcggcgcccCAGGGCCCCTTTTGTGAT</td>
<td>Erythromycin cassette</td>
</tr>
<tr>
<td>Erm-R</td>
<td>gcggcgcccAGTCGGCGCAGCTGAT</td>
<td>Erythromycin cassette</td>
</tr>
</tbody>
</table>

3 Restriction enzyme sites are in lower case Ascl – gcggcgcc; FseI – gcggcgcc
<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant characteristics</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>UA159</td>
<td>Wild-type strain; Em&lt;sup&gt;s&lt;/sup&gt;</td>
<td>J. Ferretti, U. of Oklahoma</td>
</tr>
<tr>
<td>ΔCas5d</td>
<td>UA159 Δ (SMU.1763); Em&lt;sup&gt;f&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>ΔC2E</td>
<td>lacking cas3 to cas2, Erm&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Serbanescu et al., 2015</td>
</tr>
</tbody>
</table>

**S. mutans** heat shock resistance assays

Overnight cultures grown in THYE medium were diluted 1:20 using prewarmed THYE. Cultures were then grown until mid-logarithmic phase OD<sub>600</sub> of ~ 0.4 equally divided into two aliquots. One aliquot of each culture was incubated at 50°C for 60 min and the other aliquot was used to assay survival at time 0. Samples were gently sonicated and serially diluted in PBS and sensitivity was quantitatively assessed by plating cells before and after incubation. Each dilution was then spotted in triplicate (20 ul each) onto THYE agar plates and incubated at 37°C for 2 days, and the CFU were counted. Heat shock resistance was calculated by dividing the number of CFU obtained after incubation at 50°C for 60 min by the number of CFU present at time zero and multiplying the result by 100.

**S. mutans** cell envelope antimicrobial susceptibility assays

Disc diffusion, minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) assays of cell membrane inhibitors were examined in the S. mutans wild type UA159 and CRISPR mutant strains. Cell membrane inhibitors tested included the antibiotics vancomycin, bacitracin, and β-lactam antibiotics (penicillin G and oxacillin) (all obtained from Sigma-Aldrich Canada, Ltd., Oakville, Ontario, Canada). The methodology was based on (Suntharalingam et al., 2009). Briefly, 100 µl of mid-log phase bacterial cells adjusted to an OD<sub>600</sub> of ~ 0.01 was added to a 96-well microtiter plate containing THYE medium supplemented with two-
fold serial dilutions of cell envelope inhibitors. Bacterial growth after 24 h was spectrophotometrically measured by using an ELISA microtiter plate reader (model 3550; Bio-Rad Laboratories, Richmond, CA) at an absorbance of 560 nm (OD560). Relative cell density percentages were calculated by using the following equation: (OD560 of culture in the presence of each concentration of antibiotics)/(OD560 of culture in the absence of antibiotics) × 100. MBC testing was carried out using MIC microtiter plates. Briefly, aliquots of 20 µl were taken from each well and spot plated onto THYE agar and incubated for 48h. MBCs were observed after 48 h of incubation at 37°C and 5%CO₂. Disc diffusion assays were performed as described previously (Chattoraj et al., 2010) with the following modifications. Briefly, antibiotic discs (6 mm in diameter; Becton Dickinson Laboratories) were placed on THYE agar plates that were overlaid with 5 ml of soft THYE agar containing 100 µl of overnight culture of S. mutans strain of choice. The plates were incubated overnight at 37°C under microaerophilic conditions, and the zones of inhibition were measured. The antibiotics used for this study were bacitracin (10 U), penicillin G (1 µg/ml), oxacillin (0.5 µg/ml) and vancomycin (5 µg/ml).

**Protein expression, purification and mutagenesis**

Cas5d (SMU.1763c) was cloned using genomic DNA from S. mutans UA159 into the modified pET15b plasmid as previously described using the Primers in Table 3-1 (Beloglazova et al., 2008). For enzymatic assays, Cas5d protein was overexpressed as a fusion with an N-terminal His₆-tag in E. coli BL21 (DE3) strain (Novagen) which contains no cas genes (Brouns et al., 2008). The protein was purified to more than 95% homogeneity using metal-chelate affinity chromatography on nickel affinity resin and subsequent ion-exchange chromatography on MonoQ column as previously described (Beloglazova et al., 2008). Site directed mutagenesis (alanine replacement) of SMU.1763c protein was performed as described previously (Beloglazova et al., 2008) using a protocol based on the QuikChange site-directed mutagenesis kit (Stratagene).

**In vitro Enzymatic Assays**

RNA substrates: RNA1 (39nt); AAAUACGUUUUCUCCAUUGUCAUAUUGCGCAUAAGUUGA; RNA2 (40nt); UUUCAAUUCUUUUUGGAUUAUCUUGAAGAUAGAGU
UAA and RNASMU
(50nt) GAGUUUAGAGCUGUGUUGUUUCGAAUGGUUCCAAA
ACGAUUGAACCAGCand DNA substrate (39nt)
AAATACGTTTTCTCCATTGTCATATT

GCGCATAAGTTGA were obtained from Integrated DNA Technologies. Labeling of oligonucleotides was conducted using ATP [γ-32P] (6,000 Ci/mmol; Amersham Biosciences) and T4 polynucleotide kinase (PNK) per the manufacturer’s instructions (Fermentas). Reaction conditions were as follows; 50 mM Tris pH 7.5, 100 mM KCl, 2 mM MgCl₂, 0.1 nM substrate and 200ng enzyme for 15 min at 37˚C. Samples were run on denaturating PAGE (15% polyacrylamide and 8 M) urea and reaction products were visualized by autoradiography as previously described (Beloglazova et al., 2008).

Protein crystallization and structure determination of SMU.1763c

SMU.1763c crystals were grown using the sitting drop vapor diffusion protocol by mixing 1 μl of the purified selenomethionine incorporated protein of 10 mg/ml with 1 μl of the crystallization buffer containing 25 % (w/v) PEG 2000 MME, 0.2 M NaCl, 100 mM Na-acetate (pH 5.6), and 0.3 M NDSB-201. For diffraction studies, the crystals were stabilized with the crystallization buffer containing 3 M NaCl followed by cryoprotection in 20% ethylene glycol prior to flash-frozen in liquid nitrogen. Data collections were carried out at the beamlines 19-ID and 19-BM of the Structural Biology Center, Advanced Photon Source, Argonne National Laboratory (Rosenbaum et al., 2005). A single-wavelength anomalous diffraction dataset was collected at a wavelength of 0.9794 Å. Reflection data were collected and processed using the program HKL3000 (Rosenbaum et al., 2005) (data not shown). The structure of SMU.1763c was determined by the Se-methionine SAD phasing, density modification, and initial model building as implemented in the Auto-Rickshaw package (Panjikar et al., 2005). The initial model was improved by manual rebuilding using the COOT (Emsley & Cowtan, 2004) and the REFMAC programs (Murshudov et al., 1997). The final structure was validated using MolProbity and COOT validation tools (Murshudov et al., 1997, Lovell et al., 2003). Sequence alignment was
performed using ClustalW and ESPript (Thompson et al., 1994, Gouet et al., 1999).

3.4 Results

3.4.1 Sequence analysis of the S. mutans SMU.1763c

BLAST analysis revealed that SMU.1763c protein belongs to the Cas5 protein family and specifically with the Cas5d class characteristic of the subtype I-C CRISPR system (Haft et al., 2005). The hypothetical Cas5d protein (249 aa) is encoded by SMU1763c, which is part of an operon that includes core cas genes (cas1, cas2, cas3 and cas4) and a set of cas genes corresponding to the Dvulg CRISPR subtype (csd1, csd2, cas5d). These genes are located upstream of a CRISPR array comprised of two 32 bp partially homologous repeats separated by a 34 bp spacer (Figure 3-1). Based on bioinformatics analysis it was proposed that Cas5d, a subunit of the Cascade complex, belongs to a large class of proteins named RAMP (Repeat Associated Mysterious Protein) (Haft et al., 2005, Makarova et al., 2006). The RAMP superfamily Cas proteins are predicted to have RNA recognition motif (RRM) domains and function as endoribonucleases that process the pre-crRNA to generate short mature crRNAs (Makarova et al., 2011, Makarova et al., 2011). Amino acid sequence alignment of SMU.1763c and Cas5d-like proteins from other streptococci revealed high similarity of SMU.1763c with the Cas5 proteins in several streptococcal microorganisms including S. sanguinis (86% identity; WP_002930620), S. pyogenes (80% identity; WP_010922512), S. dysgalactiae (78% identity; WP_014612478) and S. agalactiae (79% identity; WP_001923097) and CRISPR-associated protein Cas5 in S. sobrinus (80% identity; WP_019772355) suggesting that Cas5d is highly conserved among Gram-positive pathogenic streptococci. S. mutans SMU.1763c shares 46% sequence identity to the Cas5d ortholog from Bacillus halodurans.
Figure 3-1: Organization of the *cas* and CRISPR locus for the Type I-C/Dvulg of *S. mutans* UA159.

It contains an array with two repeats and one spacer and 10 *cas* genes (*cas1*-4, *csd1*-2 and *cas5d*).

### 3.4.2 SMU.1763c protein is an endoribonuclease that recognizes RNA secondary structure elements

Previously, we showed that SMU.1763c (Cas5d) of the CRISPR2 subtype I-C has RNase activity against both synthetic oligoribonucleotide substrates and total RNA extracted from *S. mutans*. Further, based on our *in vitro* transcription and DNA microarray studies, SMU.1763c had no obvious sequence preference in RNA cleavage. However, recently Cas5d from *B. halodurans* was shown to preferentially cleave the structured form of pre-crRNA containing a stem-loop structure (Nam *et al.*, 2012). By systematically varying the sequences of the repeat RNA, Nam *et al.*, found that the *B. halodurans* Cas5d cleaved the hairpin structure at the base of the stem and the 3’ single stranded sequence in the CRISPR repeat region (Nam *et al.*, 2012). Prompted by this, we asked if the RNA recognition by SMU.1763c of *S. mutans* is structure specific too. Using more stringent reaction conditions (lower protein and substrate concentrations), we tested the purified *S. mutans* SMU.1763c protein for RNase activity against several RNA substrates to check if there is secondary structure specificity in its cleavage. Purified SMU.1763c specifically cleaved the 50-nt RNA substrate between the nucleotides A8 and G9 at the 5’-side of the stemloop and between the nucleotides A26 and U27 at the 3’-side of the loop (Figure 3-2) and suggested that RNA secondary structure elements such as stemloops are likely required for SMU.1763c to process RNA substrates in a structure-specific manner. The endoribonucleolytic cleavage activity by SMU.1763c was metal independent, as addition of divalent metal cations (e.g
Mg$^{2+}$, Mn$^{2+}$, Ca$^{2+}$) had no effect on cleavage rate and pattern (data not shown). The structure-specific substrate recognition towards the base of the stemloop may explain why we were not able to identify specificity in SMU.1763c cleavage on the unstructured RNA substrates we tested such as linear RNA substrates.

Figure 3-2: Cleavage of the 50 nt fragment of *Streptococcus mutans* ssRNA by SMU.1763c.

The 5'-$^{32}$P-labelleed ssRNA was incubated at different pH in the presence of 100 mM KCl, 2 mM MgCl$_2$ without protein (lane C) or with 200 ng of SMU.1763c at 37 °C for 15 min and analyzed by denaturing gel electrophoresis. The predicted structure of the substrate used is shown schematically on the top of gel image. The substrate cleavage sites were identified based on the size of cleavage products.
3.4.3 SMU.1763c protein shows no DNase activity

Recently, Nam et al., reported that after initial processing of pre-crRNA, the Cas5 protein from *B. halodurans* remains associated with its crRNA and assembles with other Cas proteins (Csd1 and Csd2) to form a multi-subunit Cascade-like interference complex (Nam et al., 2012). These authors also reported that *B. halodurans* Csd2 exhibits RNase activity but no DNase activity. Further, other reports have described that Cas5d orthologs from *Streptococcus pyogenes* and *Xanthomonas oryzae* possess both specific endoribonuclease activity towards CRISPR RNAs and nonspecific double-stranded DNA binding suggesting that Cas5d may play multiple roles in CRISPR-mediated immunity (Koo et al., 2013) and that Cas5d DNase activity is promoted by the presence of metal ions (Punetha et al., 2013). This prompted us to investigate if the *S. mutans* Cas5d (SMU.1763c) had nuclease activity against DNA substrates and identify if Csd2 (SMU.1760c) exhibited nuclease activity against single stranded RNAs as well as DNAs. As shown in Figure 3-3A, SMU.1763c exhibited no activity against the ³²P-labeled ssDNA as substrate, suggesting that SMU.1763c has a preference for ssRNA but not against ssDNA. We also analyzed the purified Csd2 (SMU.1760c) for the presence of nuclease activity against RNA and DNA substrates. Our enzymatic assays showed that SMU.1760c exhibited nuclease activity against both ssRNA and ssDNA substrates (Figure 3-3A and B). The presence of both RNase and DNase activity hints at the possibility of *S. mutans* SMU.1760c of mediating multiple functions that require RNA and DNA targeting. Therefore, these enzymatic activities of SMU.1760c will be further analyzed using site-directed mutagenesis to determine the catalytically critical residues for the cleavage activity of SMU.1760c and to exclude the possibility of contaminating nuclease activity.
Figure 3-3: Enzymatic assays to probe (A) DNase and (B) RNase activity of SMU.1763c protein and SMU.1760c protein.

5' $^{32}$P labeled DNA and RNA substrates were incubated in the absence of protein (-ve control), or in the presence of YST0580 and YgbT (+ve controls) or 2 µg of SMU.1760c protein or SMU.1763c protein for 1h at 37°C in the presence of 50 mM Tris-HCl (pH 7.0), 5 mm MnCl$_2$, 100 mM KCl and 1 mM DTT. Reaction products showing the hydrolysis of ssRNA or ssDNA substrates were separated on a 15% PAGE/8M urea gel and visualized by phosphorimaging as previously described (Beloglazova et al., 2008).

3.4.4 The SMU.1763c structure exhibits ferredoxin-like architecture

To gain deeper insight into the molecular basis of ssRNase activity of S. mutans Cas5d, we determined the crystal structure of SMU.1763c. The structure revealed a ferredoxin domain based architecture (Figure 3-4 A). The ferredoxin-like domain represents one of the most populated protein folds, with numerous structural and functional derivatives (Beloglazova et al., 2008). This fold is present in numerous RNA-binding proteins including some of the characterized Cas nucleases (Beloglazova et al., 2008). A Dali search (Holm & Rosenström, 2010) for structural homologs of S. mutans Cas5d identified several Cas5d proteins with high structural similarity including the recently characterized Cas5d protein from B. halodurans.
Like SMU.1763c from *S. mutans*, Cas5d *B. halodurans* has a ferredoxin-like fold and a catalytic triad formed by Tyr-46, Lys-116 and His-117 as key residues important for the RNase activity (Nam *et al.*, 2012). However this protein shares relatively low overall sequence similarity (41% sequence identity) to the *S. mutans* Cas5d. The *S. mutans* Cas5d structure also revealed its active site located in the N-terminal ferredoxin-like domain with the conserved residues Tyr-50, Lys-120, and His-121 forming a catalytic triad (Figure 3-4B). SMU.1763c contains a large cluster of conserved residues (His-175, His-51, Arg-135, Arg-134, Tyr-179, Arg-127, Lys-52, Asp-115, Arg-109) positioned around the potential binding and cleavage site.

Figure 3-4: Crystal structure of the *S. mutans* SMU.1763c.

(A), Overall structure showing the presence of the ferredoxin-like fold (two orientations related by 90° rotation). (B), Close-up view of the *S. mutans* SMU.1763c
active site showing the residues potentially involved in RNA binding and cleavage. SMU.1763c catalytic triad is formed by the side chains of Tyr-50, Lys-120, and His-121.

### 3.4.5 Identification of the active site of SMU.1763c

To identify the catalytic residues of SMU.1763c protein, we performed a structure-based sequence alignment of Cas5d proteins whose crystal structures have been solved (sequence identity between *S. mutans* and the other selected organisms varied between 30-50%). Our alignment identified that SMU1763c (Cas5d) from *S. mutans* (SmCas5d) belongs to the Cas5d-A subgroup (contains ~30 additional C-terminal residues) as Cas5d from *S. pyogenes* (SpCas5d) and *B. halodurans* (BhCas5d). We also observed that SmCas5d shares ~57 highly conserved residues with BhCas5d, SpCas5d and XoCas5d (Figure 3-5). We also observed conservation of the Cas5d catalytic residues (Tyr-50, Lys-120, and His-121) suggesting a role of these conserved residues in specific CRISPR RNA processing by Cas5d protein from *S. mutans*. 
Figure 3-5: Structure-based sequence alignment of Cas5d S. mutans (SmCas5d) (PDB code 4R0J) and Cas5d-like proteins from other organisms.

Proteins depicted are from S. pyogenes (SpCas5d)(PDB code 3VZH), B. halodurans (BhCas5d)(PDB code 4F3M) and X. oryzae (XoCas5d)(PDB code 3VZI). Residues conserved in all aligned Cas5d proteins are highlighted in boxes, highly conserved residues in red boxes and similar residues in light blue boxes. The numbers and secondary structure elements (α helix and β sheet) are shown above the alignment and correspond to SmCas5d. The sequences were aligned using structural information provided by the Protein Data Bank, and the figure was generated using ESPrinpt program (Koo et al., 2013).
Among the four Cas5d homologues, structure comparison revealed high similarity between Cas5d from *S. mutans*, Cas5d from *S. pyogenes* and Cas5d from *B. halodurans* (Nam *et al.*, 2012, Koo *et al.*, 2013). To identify the residues of SMU.1763c important for its RNase activity, we performed site-directed mutagenesis and replaced 13 non-core residues to alanine (Tyr-50, His-51, Arg-109, Asp-115, Lys-120, His-121, Arg-127, Arg-134, Arg-135, Lys-52, Ser-128, Tyr-179 and Lys-131). Alanine replacement mutagenesis of three residues of SMU.1763c (Tyr-50, Lys-120 and His-121) resulted in an almost complete loss of RNase activity, confirming that these residues (Tyr-50, Lys-120, and His-121) play critical roles in the RNase activity and form a catalytic triad (Figure 3-6A). Further quantification revealed that alanine replacement of several other conserved or semi-conserved residues of SMU.1763c produced strong negative effects on the nuclease activity of this protein (Lys-52, Ser-128, Lys-131, Arg-134, Arg-135) suggesting that these residues probably contribute to nucleic acid binding. In contrast, alanine substitution of non-conserved residues Tyr-179, His-51, Arg-109, Asp-115, Arg-127 produced mutant proteins with nuclease activity comparable to the wild-type level (Figure 3-6B). Another series of other experiments were carried out to verify metal dependency of RNA cleavage by the SMU.1763c protein. The addition of divalent metal cations did not affect the cleavage of ssRNA confirming no metal-dependent requirement for its cleavage activity (data not shown).
Figure 3-6: Alanine replacement mutagenesis of *S. mutans* SMU.1763c.

(A), A denaturing gel analysis of RNase activity of the purified wild type and mutant SMU.1763c proteins. (B), Gel scanning results showing RNase activity of purified SMU.1763c mutant proteins relative to the wild-type protein (100%). To calculate nuclease activity as % of cleavage activity) the gel images were quantified using imageQuant TL software. Results are the average of two independent experiments conducted with the purified wild-type and mutant proteins.

3.4.6 No effect of SMU.1763c gene deletion on tolerance to heat shock

Since our previous work demonstrated that the ΔC2E strain (deletion mutant in all cas genes within CRISPR2/Cas system) exhibited significant loss of viability under high temperature conditions, it raised the question whether the total operon was necessary for heat stress tolerance or one gene alone mediated this role. To confirm the requirement for Cas5d in heat shock stress, we tested the ability of the Δ SMU.1763c to tolerate elevated temperature (50°C). As shown in Figure 3-7,
survival of ΔSMU.1763c mutant after high temperature shock was similar to wild-type UA159 strain, suggesting no prominent role in temperature stress tolerance.

Figure 3-7: Survival of S. mutans UA159 and ΔCas5d strain after exposure to 50°C temperature stress for 1h.

Results represent mean CFU counts ± standard deviations. Results are representative of at least three independent experiments conducted with the mutant and UA159 parent strain.

3.4.7 No role of SMU.1763c in antibiotic tolerance to cell wall-interfering antimicrobials

The S. mutans, ClpP protease plays a critical role in proper maintenance of the cell-wall structure as well as expression of virulence properties (Lemos & Burne, 2002, Deng et al., 2007, Banerjee & Biswas, 2008, Kajfasz et al., 2009). A recent transcriptome analysis of the ΔclpP strain revealed that the expression of all cas genes belonging to type I-C system, including SMU.1763c, were upregulated in the mutant strain compared to those of wild-type (Chattoraj et al., 2010). Moreover, Sampson et al. also established a link between envelope integrity and the Francisella novicida CRISPR-Cas system (Sampson et al., 2014). A mutant deficient in cas9 was more vulnerable to certain antibiotics such as polymyxin B, streptomycin and kanamycin and this effect was due to increased membrane permeability.
Further, these authors were also able to identify that Cas9 regulates the production of a lipoprotein which can alter membrane permeability (Sampson et al., 2014). Similar findings have also been found in *Campylobacter jejuni*, a bacterium that is a common cause of human gastroenteritis. The Cas9 mutant had enhanced permeability and impaired antibiotic resistance suggesting that Cas proteins of *C. jejuni* may play a regulatory role in membrane permeability in this bacterium (Sampson & Weiss, 2014).

These findings above prompted us to investigate whether the *S. mutans* SMU.1763c Cas5d protein within the CRISPR2-Cas system played a role in proper maintenance of the cell-wall structure through a mechanism mediated by ClpP. The antibiotic tolerance of the wild-type UA159 and mutant strains to cell-wall targeting antibiotics was tested using a panel of four cell envelope inhibitors, each active at different steps in peptidoglycan (PG) biosynthesis. MIC/MBC assays conducted with vancomycin and bacitracin (antibiotics that interfere with lipid II recycling) revealed no fold changes in susceptibility in the ΔCas5d mutant (Table 3-2). Similarly, MIC/MBC assays conducted with penicillin G and oxacillin non-lipid II interacting β-lactams that prevent the later transpeptidation step of PG biosynthesis, revealed no effect on the sensitivities of ΔCas5d mutant relative to parent. ΔC2E mutant (deletion mutant in all cas genes within CRISPR2-Cas system) displayed phenotypes similar to those of the ΔCas5d strain (data not shown).
Table 3-2: Susceptibility profiles of S. mutans strains exposed to cell envelope-interfering antimicrobials.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Lipid II inhibitors</th>
<th>Non-lipid II inhibitors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vancomycin (µg/ml)</td>
<td>Bacitracin (U/ml)</td>
</tr>
<tr>
<td></td>
<td>MIC</td>
<td>MBC</td>
</tr>
<tr>
<td>UA159 (wild-type)</td>
<td>2.5</td>
<td>25</td>
</tr>
<tr>
<td>ΔCas5d</td>
<td>2.5</td>
<td>25</td>
</tr>
</tbody>
</table>

MIC/MBC compared to UA159 parent are indicated.

3.5 Discussion

In type I-C CRISPR systems, the processing of pre-crRNAs to generate mature effector RNAs involves the activity of Cas5d endoribonuclease that specifically recognizes and cleaves pre-CRISPR RNA at the base of the stemloop (Garside et al., 2012, Nam et al., 2012, Nam et al., 2012, Koo et al., 2013, Punetha et al., 2013). The results reported here show that the S. mutans Cas5d exhibits structure-specific endoribonuclease activity. Consistent with previous reports (Garside et al., 2012, Nam et al., 2012, Nam et al., 2012, Koo et al., 2013, Punetha et al., 2013) we found that Cas5d protein from S. mutans recognizes and cleaves the stemloop structure in the CRISPR repeat. The Cas5d from B. halodurans has also been shown to exhibit nuclease activity against DNA substrates in the presence of a divalent metal ion (Punetha et al., 2013). In contrast to Cas5d from B. halodurans that has been suggested to play multiple roles in CRISPR-mediated activity by participating in more than one stage of process, Cas5d activity in S. mutans may be limited by participating in pre-crRNA processing. Unexpectedly, Csd2 in addition to being an endoribonuclease also exhibits endodeoxyribonuclease activity. In B. halodurans and Methanothermobacter thermoautotrophicus Csd1, which assembles into Cascade like complex with Cas5d and Csd2, was also reported to display nuclease activity against DNA and RNA substrates (Guy et al., 2004, Punetha et al., 2013). In conformity with requirements of other type I-C systems, Csd1 in S. mutans may play a similar role, although its nuclease activity needs to be investigated further. Since
DNAase activity of Csd1 was shown to be nonspecific, it raised questions on its possible roles in type I-C. The precise role of this nuclease activity is still unclear, but it was recently speculated that it could be used either during adaptation stage or interference stage (Punetha et al., 2013). Acquisition of new spacers from predator genomes requires DNA fragmentation and subsequent incorporation of spacers into the CRISPR locus. Recent studies suggested the involvement of the Cascade complex along with Cas3, Cas1 and Cas2 in the adaptation stage (Datsenko et al., 2012). Therefore, DNase activity of the Cascade complex would be beneficial to the host by allowing for the acquisition of multiple spacers from multiple regions. Non-specific DNase activity could also be beneficial during the interference stage as well. The CRISPR type I-C system might use multi-protein assemblies to interact with the invading genome and facilitate fragmentation of its nucleic acids. In the case of S. mutans it is also possible that the SMU.1760c (Csd2) nuclease may be employed in clearing the cellular nucleic acid debris generated under stress conditions. This possibility is in agreement with previous observations that the cas genes associated with the type I-C CRISPR/Cas system interact with stress related genes (Chattoraj et al., 2010, Kajfasz et al., 2010, Kajfasz et al., 2011, Serbanescu et al., 2015). Using structural studies, we have identified that S. mutans Cas5d has a ferredoxin-based architecture. Crystal structure showed the presence of an N-terminal ferredoxin-like domain and a C-terminal flexible domain. A high similarity between Cas5d from S. mutans and the previously characterized type I-C Cas5d from B. halodurans emerged from our analysis (Nam et al., 2012). Moreover, the S. mutans Cas5d structure also revealed its active site located in the N-terminal ferredoxin-like domain with the conserved residues Tyr-50, Lys-120, and His-121 forming a catalytic triad. Further, structure-guided mutagenesis confirmed that residues Tyr-50, Lys-120, and His-121 in Cas5d play critical roles in endoribonuclease activity, likely forming a catalytic triad.

Using MIC/MBC assays, we observed that the Cas5d deficient mutant was not sensitive to vancomycin, bacitracin, penicillin G or oxacillin, antibiotics that disrupt the proper maintenance of cell wall structure of S. mutans. These data suggest that in S. mutans, the Cas5d is not part of the protection mechanism to these antibiotics.
Similarly our disk diffusion assays conducted with the mutant and the antibiotics revealed no effect on the tolerance of *S. mutans cas5d* mutant compared to the parent strain (data not shown). However, a transcriptome profiling of *S. mutans* between a *clpP* mutant and its UA159 strain, identified that ClpP (caseinolytic protease) which plays an important role in the regulation of antibiotic tolerance, mediates the expression of CRISPR2 *cas* genes including *cas5d* (2.2 fold change relative to wild-type) (Chattoraj *et al.*, 2010). While these data suggest a role of *cas5d* in ClpP-mediated proteolysis for proper maintenance of cell wall structure its exact function is still not understood and remains to be elucidated.

Finally, by assessing cell survival after exposure to heat shock at 50°C, we demonstrated that the deletion of *cas5d* gene within type I-C system did not produce significant differences among the survival profile of the mutant strain and the parent strain, suggesting that Cas5d does not play a prominent role in heat shock tolerance. This finding was somewhat surprising, since the mutant lacking all of the ten *cas* genes associated with the type I-C system had an impaired ability to tolerate high temperature stress compared to the parent strain (data not shown). Since mutants did not share the same sensitivity to the stressor, it is likely that Cas5d is not required in heat shock resistance. Hence, more studies are warranted to examine the role of the other *cas* components of type I-C CRISPR in *S. mutans* to support functions related to the heat shock tolerance. It is also highly possible that Cas5d is involved in cellular processes other than tolerance to antibiotics. A recent transcriptome analysis of *S. mutans* cells cultured under aerobic and anaerobic conditions demonstrated that the *cas* genes associated with the type I-C CRISPR/Cas were upregulated between 2.1-2.4-fold in cells exposed to oxygen (Ahn *et al.*, 2007). Although this transcription data imply that Cas proteins may be important factors in initial adherence or biofilm formation by *S. mutans*, further examination would be required to shed light on the potential involvement of SMU.1763c Cas protein in pathways linked to *S. mutans* biofilm formation; It is also of note that in a different study, a mixed-species culture (*V. parvula*, *S. mutans* and *S. gordonii*) had increased expression of multiple *cas* genes including SMU.1763c, suggesting that the *S. mutans* Cas proteins might play an important role in a mixed
biofilm community (Liu et al., 2011). How these genes play any role in the interspecies interactions awaits further clarification. Further examination of Cas5d in *S. mutans* to determine whether type I-C system affects bacterial physiology and fitness would be of interest.

In summary, our biochemical and structural studies demonstrated for the first time that Cas5d from *S. mutans* is an endoribonuclease that recognizes and cleaves the structured RNA substrates. Further, the crystal structure revealed that SMU.1763c Cas protein has a ferredoxin domain-based architecture. Mutagenesis indicated that residues Tyr-50, Lys-120 and His-121 in this protein play a critical role in endoribonuclease activity. Our phenotypic characterization of the ΔCas5d strain revealed that SMU1763c did not play a prominent role in sensing cell envelope stress to preserve envelope integrity or controlling stress response induced by heat shock in *S. mutans*. Future investigations may shed light on the physiological functions of the SMU.1763c Cas5d protein of *S. mutans*.

### 3.6 Acknowledgements

This study was supported by NIH Grant R01DE013230-03 and CIHR Grant MT-15431 to DGC and by the Government of Canada through Genome Canada and the Ontario Genomics Institute 2009-OGI-ABC-1405 to AFY. MAS and NB contributed to experimental design, performed experiments, data analysis and/or to the writing of this manuscript. RF, KK and SL performed experiments. DGC and AFY contributed to data analysis and writing of this manuscript.
3.7 References


Chapter 4: Summary and Conclusions
4.1. Summary of Dissertation

This dissertation examines the ability of two CRISPR/Cas systems (type II-A and type I-C) to provide immunity against invasive genetic elements (viruses and plasmids) and investigates whether these CRISPR systems are involved in other cellular functions in the biofilm-forming organism Streptococcus mutans. Using a plasmid-based interference assay that mimics infection (plasmid transformation) with invading protospacer-containing DNA elements, we demonstrated DNA interference activity in vivo in S. mutans UA159 by its type II-A CRISPR/Cas system. Endogenous CRISPR spacers 2 and 3 from type II-A, both matching sequences from phage M102, were found to be essential for CRISPR interference. In contrast, spacer 6 within type II-A and spacer 1 within type I-C both matching sequences from other S. mutans genomes, had no effect on interference. It is possible that the crRNA complementary to these spacers do not accumulate to levels high enough to be equally effective as suggested in P. aeruginosa (Cady et al., 2012) or lack of interference activity due to loss of function of cas1 (stop codon, frame shift mutation) and csd1 (frame shift mutation) within CRISPR2/Cas system as suggested based on bioinformatics analysis (Van der Ploeg, 2009). Alternatively, they may lack activity to allow for genetic exchange with closely related bacteria to enable the fitness of the host under unfavorable conditions. In addition to anti-viral defense, CRISPR1/Cas system was found to be involved in bacterial physiology in S. mutans. We found that under low pH, oxidative, cell membrane and heat stresses cas deletion strains showed reduced survival compared with UA159. We also provided evidence that CRISPR1/Cas system was involved in the DNA repair pathway, which we suggest may contribute to exchange of fitness enhancing genes under stress and contribute to the evolutionary fitness of the organism. Previous transcriptome data showed that CRISPR2-Cas components of S. mutans have potential roles in various aspects of bacterial physiology including stress response (Chattoraj et al., 2010, Kajfasz et al., 2011), biofilm formation (Ahn et al., 2007) or communication between species (Liu et al., 2011). Utilizing a mutant deficient in CRISPR2 cas genes we revealed a possible role of type-I-C system in high temperature stress response. Further, using structural and biochemical studies we found that the role of the Cas5d protein SMU.1763c, a putative endoribonuclease associated with the CRISPR
system I-C, participates in processing of CRISPR RNAs in *S. mutans*. Lastly, we revealed that VicK, the histidine kinase of VicRK signaling system governs expression of *cas* genes within both CRISPR-Cas systems.

4.2. General Discussion

4.2.1. CRISPR/Cas mediated immunity

CRISPR interference confers adaptive, sequence-based immunity against viruses, plasmids and genomic DNA (Barrangou *et al.*, 2007, Brouns *et al.*, 2008, Sorek *et al.*, 2008, Van Der Oost *et al.*, 2009, Waters & Storz, 2009, Karginov & Hannon, 2010, Marraffini & Sontheimer, 2010, Wiedenheft *et al.*, 2012). *S. mutans* strain UA159 harbors two distinct CRISPR/Cas systems: CRISPR1 and CRISPR2 (Haft *et al.*, 2005, Horvath *et al.*, 2009, Van der Ploeg, 2009). Previous reports in *S. mutans* suggested that phage-derived spacers present in these strains likely resulted from M102-like phage attacks (Van der Ploeg, 2009). Subsequently, it was shown that M102 attaches to phage sensitive and phage resistant *S. mutans* serotype c strains, indicating that factors other than phage adsorption contribute to the resistance of *S. mutans* strains serotype c to infection by M102 phage (Shibata *et al.*, 2009). Based on these findings it was suggested that CRISPR/Cas systems of *S. mutans* possibly play a role in acquired resistance to M102 infection. We set out to determine whether the CRISPR/Cas systems in *S. mutans* prevent phage infection and natural transformation by spacer-matching sequences. In fact, we suggest that *S. mutans* employs CRISPR-independent defense mechanisms against phage M102 to target diverse steps of the phage life infection.

Our study showed that the native type II-A CRISPR1 system of *S. mutans* UA159 is capable of preventing natural transformation via plasmid DNA. We demonstrated that protospacers matching the CR1SP2 and CR1SP3 spacers are targeted by the CRISPR1 system. Similar to previous work, the degree of inhibition of transformation varied widely depending on the protospacer tested (Cady *et al.*, 2012). Our results showed that plasmids carrying intact targets for CR1SP6 (identity to *S. mutans* GS-5) and CR2SP1 (identity to *S. mutans* LJ-23) which were predicted to be targeted for
cleavage by CRISPR/Cas systems, do not provide immunity. Although it is unclear why they do not provide immunity it is possible that crRNA molecules complementary to these spacers do not accumulate to levels high enough to be equally effective. It is possible that secondary structures of inactive crRNAs form stem-loops or other structures that could block binding to Cas9 or tracrRNA or the folding of crRNA. Alternatively, they lack activity for other reasons that are yet to be elucidated.

Unexpectedly, we found that the S. mutans CRISPR/Cas system II-A targets invading plasmid DNA through a PAM-independent mechanism. These results contrast to those data obtained using in vitro plasmid cleavage assays, where dual-tracrRNA/crRNA-guided Cas9 from S. mutans could efficiently cleave target DNA in the presence of an NGG sequence (Fonfara et al., 2014). A PAM-independent CRISPR interference was observed in CRISPR type III system (e.g. type III-A Csm system from Streptococcus thermophilus, which targets RNAs) (Tamulaitis et al., 2014). In type II-A system S. thermophilus it was also observed that plasmids carrying protospacers with intact or degenerate PAMs could not be transformed into the corresponding plasmid-interfering strains, whereas phages carrying the degenerative PAMs could infect the matching phage-insensitive mutants (Deveau et al., 2007, Garneau et al., 2010). It was proposed that the tolerance of PAM degeneracy for CRISPR-Cas function could be due to the lower selective pressure for plasmids compared to phages. Such an activity could theoretically produce a lower level of transformation frequency (TF) in the presence or absence of flanking sequences observed in our study. Alternatively, the presence of plasmids inside the cell could increase the expression of Cas proteins, reflecting higher interference activity that might not require a PAM site (Makarova et al., 2012). Although the magnitude of the reduction in transformation efficiency caused by the action of the type II-A CRISPR/Cas system varied depending on the protospacer being tested, our study clearly demonstrated DNA interference activity in vivo in S. mutans by the system.

4.2.2. Involvement of CRISPR-Cas in bacterial stress responses

Despite our work and other studies (Maruyama et al., 2009, Van der Ploeg, 2009)
that explored the role of CRISPR/Cas systems in \textit{S. mutans} in conferring phage immunity, recent transcriptome studies hint towards other functions that CRISPR/Cas systems might have in bacterial physiology in \textit{S. mutans}, however these mechanisms are poorly understood (Senadheera \textit{et al.}, 2009, Chattoraj \textit{et al.}, 2010, Kajfasz \textit{et al.}, 2010, Xie \textit{et al.}, 2010, Kajfasz \textit{et al.}, 2011, Liu \textit{et al.}, 2011). We found that the VicRK signaling system, previously shown to be involved in tolerance to stress response, differentially regulates the expression of CRISPR/Cas systems. VicR/K may coordinate their expression to prevent the simultaneous expression of both CRISPR-Cas systems. This is in agreement with a recent report where it was suggested that CRISPR-Cas systems of \textit{S. thermophilus} can function independently (Carte \textit{et al.}, 2014). As expected, we found that CRISPR/Cas systems are important for tolerance of \textit{S. mutans} to different environment stresses. Deletion of \textit{cas} genes within the type I-C system resulted in increased sensitivity to heat shock, whereas inactivation of \textit{cas} genes within type II-A showed increased tolerance to low pH and displayed increased sensitivity to oxidative and cell membrane stress. Since \textit{ΔC1S} did not share the same level of sensitivity to the tested stressors as that seen with \textit{ΔC2E}, it is likely that CRISPR1 \textit{cas} and CRISPR2 \textit{cas} genes are differentially regulated to function independently within the environment. It is also possible that different signaling systems and/or global regulators in \textit{S. mutans} form cross-regulatory networks with VicRK components to modulate stress response pathways. In fact, previous studies have speculated the possibility of cross-talk between the LiaS sensor of the \textit{S. mutans} LiaSR system (involved in cell envelope stress tolerance) with one or more RRrs by comparing phenotypic variations (e.g., acidic pH resistance and bacteriocin mutacin IV production) between the liaS and liaR mutants (Li \textit{et al.}, 2002, Chong \textit{et al.}, 2008). Moreover, (Downey \textit{et al.}, 2014) provided the first direct evidence of cross-talk between sensor VicK HK and noncognate GcrR and SloR response regulator components to modulate stress response pathways. It is possible that some of stress response regulatory systems in \textit{S. mutans} interact with VicRK to mediate the expression of CRISPR/Cas system genes in response to cues such as oxidative stress, cell membrane changes or alterations in the internal pH of the cell. Another important finding is that \textit{cas} genes of CRISPR1/Cas system contribute to the protection of \textit{S. mutans} cells against DNA-damaging agents. It is
well documented that one of the major challenges that oral biofilm bacteria have, including \textit{S. mutans} is to cope with environmental fluctuations (in particular low pH and oxidative stress) and many of the stresses induce DNA damage in \textit{S. mutans} (Lemos & Burne, 2008, Derr \textit{et al.}, 2012). To date several DNA repair proteins have been shown to repair DNA damage induced by stress including: RecA recombinase protein that repairs and restarts stalled DNA replication forks (Quivey \textit{et al.}, 1995), an apurinic-apyrimidinic AP endonuclease involved in the DNA base excision reair (BER) (Hahn \textit{et al.}, 1999), a UV repair excinuclease (uvrA) involved in the nucleotide excision repair pathway (NER) for excising larger DNA lesions caused by acid and other DNA-damaging agents (Hanna \textit{et al.}, 2001) and the most recent discovered Smx and Smn exonucleases involved the base excision repair (BER) pathway (Faustoferri \textit{et al.}, 2005). Reduced survival of \textit{S. mutans} observed with DNA damaging agents suggests that Cas components of the CRISPR1-Cas system possibly act in one of the DNA repair pathway in response to DNA damage caused by environmental stress to promote DNA repair leading to survival and proliferation. However, the specific role of CRISPR1-Cas elements in the repair pathway remains to be elucidated.

4.2.3. \textbf{Molecular function of the CRISPR-associated protein SMU.1763c (Cas5d) from \textit{S. mutans}}

Previous reports in \textit{B. halodurans} suggested that Cas5d is a sequence-specific RNA endonuclease that cleaves CRISPR repeats and is responsible for processing of pre-crRNA (Nam \textit{et al.}, 2012). Furthermore Cas5d, together with Csd1 and Csd2, were found to complement Cas-deficient (deletion of each component of Cascade-like assembly) \textit{E. coli} to silence plasmid DNA \textit{in vivo} (Nam \textit{et al.}, 2012). Using biochemical, genetic and structural approaches, we found that the SMU.1763c protein from \textit{S. mutans} possesses ribonuclease activity with a preference to structured RNAs with stem-loops suggesting that this activity might contribute to pre-crRNA processing in \textit{S. mutans}. This finding prompted us to investigate the biological function of Cas5d cleavage activity. Previously we showed that disruption of type I-C CRISPR \textit{cas} genes (including cas5d) affected the ability of \textit{S. mutans} to withstand stress encounter by high temperature. However, the mutant deficient in
Cas5d did not reveal a noticeable defect in its ability to tolerate high temperature stress. We concluded that the enzymatic activity of SMU.1763c protein has no prominent role in stress tolerance of *S. mutans*. Although our investigation revealed that *S. mutans* Cas5d protein is not able to provide the cell with capability to resist and cope with the stress induced by high temperature or cell envelope antibiotics to promote cell survival, we can not exclude that it may be implicated in other aspects of bacterial physiology that remain to be elucidated. In fact it has previously suggested that type I-C CRISPR *cas* genes interact with either virulence or global regulatory genes involved in stress response (Chattoraj *et al.*, 2010, Kajfasz *et al.*, 2010, Kajfasz *et al.*, 2011). Notably, all of the *cas* genes of type I-C in *S. mutans* were induced during biofilm formation in the presence of starch or sucrose (Ahn *et al.*, 2007). In addition, RelA a regulatory protein involved in stringent response in *S. mutans*, has been found to interact with some of the *cas* genes belonging to both type I-C and type II-A CRISPR-Cas systems, including *cas5d* (Nascimento *et al.*, 2008). Together, these findings described above revealed that *cas* genes associated with type I-C system are activated in response to changes in the environment suggesting that there might be a more general involvement of CRISPR-Cas in virulence than previously appreciated. However, the evidence is circumstantial and descriptive and further investigations are required to provide some clues towards a molecular mechanism by which CRISPR/Cas systems function in *S. mutans*.

In summary, we have investigated the potential role of two CRISPR/Cas systems as interference systems in the defense against foreign genetic elements in *S. mutans*. We have identified that *S. mutans* has evolved CRISPR-Cas systems to protect itself from the attack by invading plasmids, We have also shown that CRISPR-Cas systems are important components of the bacterial stress response and DNA repair in *S. mutans* to mount appropriate responses as a result of a wide variety of adverse and fluctuating conditions within the host. We have also identified a regulatory role of the VicKR signaling system in CRISPR-Cas expression. Cas5d protein has been proposed to play a key role in the crRNA processing for CRISPR/Cas system subtype I-C. We showed that the Cas5d protein SMU.1763c from *S. mutans* cleaves RNA substrates with a preference to structured RNAs. The crystal structure of
SMU1763c revealed the active site and together with site-directed mutagenesis identified a catalytic triad.

4.3. Future Directions

While significant progress has been made towards understanding the role of CRISPR/Cas systems in protection against foreign DNA and cell physiology, several important questions remain unanswered. Although our data suggest that *S. mutans* prevents plasmid transformation by type II-A CRISPR-Cas system, the acquired resistance to M102 infection cannot be attributed to the function of this defense system alone. It remains to be elucidated what are the other mechanisms of resistance used by *S. mutans* UA159 that may function cooperatively or independently to confer resistance to M102-like phage attacks. It would also be interesting to address whether UA159 is resistant to the M102 phage as a consequence of adaptation to laboratory growth.

We showed that type II-A CRISPR-Cas system is functionally active in DNA interference via plasmid transformation in UA159 strain. If this system would be moved into OMZ381 it may be effective against phage predation in the susceptible strain.

We demonstrated that the type II-A CRISPR-Cas system targets spacers matching M102 sequences in a PAM-independent manner, however it is not clear how *S. mutans* can distinguish self versus non-self in a population; that requires further investigations. Finally, although our investigations using cas defective mutant strains further corroborate past evidence that the CRISPR-Cas systems may also respond to stress response, their exact role in *S. mutans* stress tolerance, in addition to DNA repair remains to be elucidated.

4.4. Significance

Our study reports an important advance in understanding CRISPR-Cas systems in the human pathogen *S. mutans*. Firstly, we elucidated the role of CRISPR-type-II-A system in protecting the host against incoming DNA, by targeting and excluding
unwanted DNA taken up by natural transformation, a key process used by *S. mutans* to acquire exogenous DNA. These results are important since recent reports have shown that the natural type II-A CRISPR-Cas system of *S. pyogenes* (known as CRISPR-Cas9 system or SpCas9) can be used as a genome editing tool to inactivate or modify any gene of interest in a variety of species and cell types including bacteria, yeast, fruit fly, common crops and especially in mammalian cells (see (Sander & Joung, 2014) for a detailed list). Cleavage by the SpCas9 can achieve precise site-specific mutagenesis if it is targeted either with a pair of crRNA and tracrRNA or with a chimeric sgRNA containing a short guide sequence that directly matches the target site (Cong *et al.*, 2013, DiCarlo *et al.*, 2013, Jiang *et al.*, 2013, Maeder *et al.*, 2013, Wang *et al.*, 2013). One hallmark of the CRISPR-Cas9 system is its ability to modify the machinery to express several crRNAs to target multiple sequences in parallel (Barrangou *et al.*, 2007, Garneau *et al.*, 2010, Deltcheva *et al.*, 2011). Inactivation of RuvC and HNH nuclease domains of Cas9 protein (each of which nicks a DNA strand to generate blunt-end DSBs) may generate nickase mutants that cause single-strand DNA breaks. Catalytically dead Cas9 (dCas9) coupled to functional effectors (transcriptional activators or fluorescent reporters) can facilitate targeted regulation or life imaging (Hsu *et al.*, 2014). Furthermore, inactivated Cas9/crRNAs systems have the ability to both upregulate and repress transcription factors, a technique key in investigating and controlling cellular behavior (Gilbert *et al.*, 2013, Mali *et al.*, 2013, Mali *et al.*, 2013). The native CRISPR-Cas II-A system from *S. mutans* share the conserved features observed in *S. pyogenes* and *S. thermophilus* (Chylinski *et al.*, 2013, Fonfara *et al.*, 2014). Accordingly, the native type II-A system of *S. mutans* could be valuable in applying the CRISPR machinery as a versatile, adaptable tool for multiplex genome editing and gene regulation.

Secondly, our results provide evidence that type II-A system is important in the prevention of natural transformation (key process used by *S. mutans* to acquire virulence genes and antimicrobial resistance genes) via plasmid DNA. Therefore CRISPR-Cas loci can impact the spread of virulence factors. This is actually the case in pathogenic staphylocococci which have acquired resistance primarily through the acquisition of conjugative plasmids carrying resistance genes (Weigel *et
Staph. epidermidis RP62a, contains a CRISPR-Cas system with a spacer matching all staphylococcus conjugative plasmids sequences to date. The spacer provides immunity against the conjugative transfer of these plasmids, thereby preventing the acquisition of the antibiotic resistance that they carry (Marraffini & Sontheimer, 2008). This seems to be the case for the human pathogen N. meningitidis as well. It was revealed that CRISPR spacers matching Neisseria genomic sequences are able to limit natural transformation, which is the primary source of genetic variation that contributes to immune evasion, antibiotic resistance, and virulence in this human pathogen (Zhang et al., 2013). Therefore CRISPR loci could control the dissemination of antibiotic resistance genes in bacterial pathogens.

Lastly, in addition to adaptive immunity to prevent the entrance of unwanted DNA, we showed that CRISPR-Cas systems of S. mutans also play other roles in the cell’s physiology by modulating stress responses and DNA repair. The ability of S. mutans to tolerate the fluctuating environmental conditions and to thrive within those environments is vital to its virulence. S. mutans itself is of importance as one of the primary causative agents of the most prevalent human infection, dental caries. However, it is also a member of the medically relevant Streptococcus and a biofilm-forming organism. A better understanding of the CRISPR-Cas pathways involved in stress response in S. mutans may be critical to generate new findings for possible translation into novel strategies to help engineer bacterial species to allow us to better combat infections beyond the oral cavity.
4.5. References


Derr AM, Faustoferrri RC, Betzenhauser MJ, Gonzalez K, Marquis RE & Quivey RG (2012) Mutation of the NADH oxidase gene (nox) reveals an overlap of the oxygen-


