Molecular identification and insights into the structural functional characterization of Sortase E transpeptidase of *Corynebacterium glutamicum*

by

Susmitha A
10BB15A39018

A thesis submitted to the
Academy of Scientific & Innovative Research
for the award of the degree of
DOCTOR OF PHILOSOPHY
in
SCIENCE

Under the supervision of
Dr. K. Madhavan Nampoothiri

CSIR-National Institute for Interdisciplinary Science and Technology (NIIST)
Industrial Estate P.O., Thiruvananthapuram-695 019
Kerala, INDIA

Academy of Scientific and Innovative Research
AcSIR Headquarters, CSIR-HRDC campus
Sector 19, Kamla Nehru Nagar,
Ghaziabad, U.P. – 201 002, India

December-2021

राष्ट्रीय अंतर्विषयी विज्ञान तथा प्रौद्योगिकी संस्थान
National Institute for Interdisciplinary Science and Technology
वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद | Council of Scientific and Industrial Research
इंडस्ट्रियल एस्टेट पी. ओ, पापनंकोड, तिरुवनंतपुरम, भारत -695019 | Industrial Estate P.O, Pappanamcode, Thiruvananthapuram, India 695019

CERTIFICATE

This is to certify that the work incorporated in this Ph.D. thesis entitled, **"Molecular identification and insights into the structural functional characterization of Sortase E transpeptidase of *Corynebacterium glutamicum*"**, submitted by **Ms. Susmitha A** to the Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of ***Doctor of Philosophy in Science***, embodies original research work carried out by the student. We, further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research materials obtained from other sources and used in this research work has been duly acknowledged in the thesis. Images, illustrations, figures, tables etc., used in the thesis from other sources, have also been duly cited and acknowledged.

A. Susmitha 9/12/21

Susmitha A
(Candidate)

9.12.2021

Dr. K. Madhavan Nampoothiri
(Supervisor)

Head, Microbial Processes and Technology Division , Tele : +91 471 2515368 (O); +91 9446482557 (M) | Email: rajeevs@niist.res.in, rajeev.csir@gmail.com

STATEMENTS OF ACADEMIC INTEGRITY

I, **Susmitha A,** a Ph.D. student of the Academy of Scientific and Innovative Research (AcSIR) with Registration No. **10BB15A39018** hereby undertake that, the thesis entitled "**Molecular identification and insights into the structural functional characterization of Sortase E transpeptidase of *Corynebacterium glutamicum***" has been prepared by me and that the document reports original work carried out by me and is free of any plagiarism in compliance with the UGC Regulations on *"Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)"* and the CSIR Guidelines for "*Ethics in Research and in Governance (2020)*".

A. Susmitha
9/12/21

Susmitha A
Thiruvananthapuram

It is hereby certified that the work done by the student, under my supervision, is plagiarism-free in accordance with the UGC Regulations on *"Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)"* and the CSIR Guidelines for "*Ethics in Research and in Governance (2020)*".

Poothi
9.12.2021

Supervisor

Dr. K. Madhavan Nampoothiri
Thiruvananthapuram

DECLARATION

I, **Susmitha A** bearing AcSIR Registration No. **10BB15A39018** declare:

(a) that my thesis entitled **"Molecular identification and insights into the structural functional characterization of Sortase E transpeptidase of *Corynebacterium glutamicum*"** is plagiarism free in accordance with the UGC Regulations on *"Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)"* and the CSIR Guidelines for *"Ethics in Research and in Governance (2020)"*.

(b) that I would be solely held responsible if any plagiarized content in my thesis is detected, which is violative of the UGC regulations 2018.

A. Susmitha
9/12/21

Susmitha A
Thiruvananthapuram

Dedicated to my parents

ACKNOWLEDGEMENTS

I wish to offer my thanks first and foremost to my research supervisor, Dr. K. Madhavan Nampoothiri for introducing me to this area of current interest and to many facets of recombinant DNA technology. I will be eternally grateful to him for his unwavering support and compassion over the years, as well as the time, supervision, knowledge, and resources he offered to help me to finish my studies. In person, he was more than a buddy, standing by my side during the most difficult days of my Ph.D. pursuit.

I wish to express my gratitude towards Dr. Rajeev K. Sukumaran (Head, MPTD) for providing timely help, support and advice during my tenure. I would like to extend my profound thanks to Prof. Ashok Pandey (former Head, MPTD) for allowing me to work in the division.

It is my privilege to place on record my gratitude to Dr. A. Ajayaghosh, Director, CSIR-NIIST and Dr. Suresh Das (former Director, CSIR-NIIST) for providing all the necessary facilities in the institute for my research work.

I am grateful to Dr. M. Arumugam, Dr. Karunakaran Venugopal, Dr. Suresh C. H, Dr. Laxmi Varma and Dr. Mangalam S Nair for their warm support as the AcSIR coordinators. I would also like to acknowledge Dr. Rajeev K. Sukumaran, Dr. Raghu K. G and Dr. Binod Parameswaran, members of my Doctoral Advisory Committee (DAC), for their sound advice and support and the constructive suggestions.

I am thankful to the Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India, for allowing my enrolment for the PhD program.

I would like to acknowledge the financial support from CSIR, Govt. of India in the form of a Research Fellowship for my PhD work.

I would like to acknowledge with much appreciation Dr. Harsha Bajaj for her immense support and ideas that helped me to carry out my work forward.

I would like to thank Dr. Kaustabh Kumar Maiti and Arya J. S for helping me to carry out the chemistry part of my work.

I take this opportunity to thank Dr. N. Ramesh Kumar for his valuable suggestions, Er. M. Kiran Kumar, Dr. R. Sindhu, Dr. D. Leena, Dr. Anil K. Mathew, Dr. Gincy Mathew, Dr. B.V.

Thirumalesh, Dr. Rakesh Yasarla, Dr. D. Vijayan and Dr. P.A. Balakumaran for providing timely help, support and encouragement during my tenure at CSIR- NIIST.

I am very thankful to all the members of "Team Biotech" for creating a unique, friendly and healthy working environment. I sincerely thank my seniors Dr. Varsha, Dr. Kiran S. Dhar, Dr. Divya, Dr. Anusree and Dr. Nishant for their advices during my tenure. I am deeply thankful to Mr. Muneeb Hamza, Mrs. Keerthi Sasikumar, Mrs. Varsha, Mrs. Devi, Mrs. Amritha, Mr. Arun, Mrs. Varsha, Mrs. Dhanya M, Mrs. Lekshmi Sundar, Mrs. Soumya, Mr. Anandhu, Mr. Shahabas, Mr. Godan, Mrs. Reeba for their motivation and support throughout my work. I would like to thank Dr. Karthik, Dr. Anju Alphonsa, Dr. Aravind, Dr. Vivek, Dr. Sabeela Beevi, Dr. Vani, Dr. Ayman for their friendly help and moral support. I am deeply thankful to Mr. Rahul, Dr. Remya, Mr. Dileep R, Mr. Anoop, Mr. Adarsh, Ms. Meera, Ms. Athira, Dr. Reshma, Ms. Reshma Mathew, Dr. Sujitha, Dr. Aswathi, Dr. Aswathy Udhayan and Mrs. Meena, Mr. Salman, Mr. Rajesh, Ms. Alphy, Dr. Hazeena, Ms. Sreepriya, Ms. Anju Tomy, Ms. Athulya, Mr. Valan, Ms. Lakshmi M Nair, Ms. Gayathri, Ms. Salini, Mr. Binoop, Mr. Prabhu, Mrs. Kirti, Mrs. Gayathri, Mr. Nitheesh and Mr. Abijith who helped me one way or the other during the entire period of my work.

I am deeply grateful to my mother, father, brother, husband, all my family members and friends for their constant love, emotional support, inspiration and their prayers had been a morale booster through all the difficult times.

Last but not least, I firmly hold that the Almighty have blessed me and enabled me to complete the book without any hindrance.

-Susmitha A

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction and Review of Literature

Abbreviation

CWSS	Cell wall sorting signal
eGFP	Enhanced green fluorescent protein
GRAS	Generally recognized as safe
HPLC	High performance liquid chromatography
LAB	Lactic acid bacteria
MSCRAMMs	Microbial surface components recognizing adhesive matrix molecules
PI	phosphatidyl-myo-inositol
SDP	Sortase-dependent protein

1.1. Introduction

Gram-positive bacteria have a thick cell wall that protects the cell from mechanical stress and also serves as a scaffold for the display of a large number of surface proteins arbitrated by sortase enzymes. The functional roles regulated by sortase while anchoring substrates include pilus polymerization, heme transport, nutrient uptake, sporulation, aerial hyphae formation, and general housekeeping function of the cells (Bradshaw et al., 2015). Based on the amino acid sequences, sortases are classified into six classes (A-F), of which Sortase A of *Staphylococcus aureus* (SaSrtA) is the well-characterized housekeeping enzyme and recognizes LPXTG sorting signal where P (underlined) is the conserved residue in the sorting motif of SrtA (Ton-That et al., 1999). The transpeptidation activity of sortase from SaSrtA and its variants were increasingly being used in a variety of biotechnological applications such as protein ligation, covalently protein binding to the cells, protein labeling, cell-surface modification, protein cyclization, and immobilization of proteins (Dai et al.,

2019). Except for SaSrtA, the sortases from other classes are not being utilized for any application due to low activity and relatively poor enzyme characterization.

The majority of our present understanding of sortases is based on research on Firmicutes (low G + C Gram-positive bacteria where SrtA is the predominant class). Interestingly, in Actinobacteria (high G + C Gram-positive bacteria) Class E sortases are reported but they were not explored much to date and are the focus of the study.

Class E sortases that recognize an unusual LAXTG sorting motif in which the conserved proline residue is replaced with alanine (underlined). However, limited studies were reported on Class E sortases in *Corynebacterium diphtheriae*, *Streptomyces coelicolor, Streptomyces avermitilis, and Streptomyces mobaraensis* (Ton-That and Schneewind, 2003; Kattke et al., 2016; Das et al., 2017; Anderl et al., 2019). There have been few reports in the *Corynebacteriaceae* family, which is a suprageneric actinomycete taxon of Gram-positive bacteria, with the most prominent members being the human diseases *Mycobacterium tuberculosis* and *Mycobacterium leprae* (Bloom and Murray, 1992). Furthermore, the only existing report of sortase was found in human pathogen *C. diphtheriae*, which consists of two classes of sortase which includes class C and class E, of which class C sortase is involved in assembling different types of pilus structures and class E is involved in anchoring the assembled pili. All of these bacteria have a unique cell wall matrix made of mycolic acids, arabinogalactan, and peptidoglycan, which is known as the mycolyl-arabinogalactan-peptidoglycan complex. They also have a comparable set of cell wall-associated glycolipids, such as phosphatidyl-myoinositol (PI) (Dover et al., 2004).

However, the *Corynebacteriaceae* family also includes nonpathogenic bacteria, such as *Corynebacterium glutamicum*, which is a soil-dwelling, nonsporulating bacterium with a G+C content of 53.8 % for the total DNA and employed in the industrial manufacture of

amino acids with the major products being the flavor enhancer L-glutamate (as monosodium salt) and the feed additive L-lysine (Kalinowski et al., 2003; Ikeda and Takeno, 2013). The organism is generally recognized as a safe (GRAS) organism and is also used as a model organism in industrial biotechnology. During in silico analysis, it is observed that the genome of this nonpathogenic wild-type *C. glutamicum* ATCC 13032 contains a single Sortase E (NCgl2838) and was also able to predict a sortase-dependent protein (SDP) Cgl0614 with a LAXTG sorting motif. Since this enzyme has not yet been identified and its functional role is unknown, it was decided to conduct more research on it.

1.2. Objectives

To study in detail the Class E sortase of *C. glutamicum*, the following major objectives were addressed in this thesis.

- In silico structural modeling and characterization of sortase and sortase-dependent proteins (SDPs) in wild-type *C. glutamicum* ATCC 13032.
- Construction of sortase deletion mutant, complemented strain, and overexpressing strain to study the cellular morphology and physiological alteration.
- Cloning, expression, and purification of putative sortase (*NCgl2838*) of *C. glutamicum* in *E. coli* BL21(DE3).
- Standardization and validation of sortase activity and substrate specificity by FRET and HPLC analysis.
- To study substrate recognition and structural function importance of the active site residues by site-directed mutagenesis approach.

- Demonstration of sortase-mediated protein (eGFP and XylB) immobilization on gold nanoparticles.

Chapter 1 gives a brief introduction to the thesis and an extensive review of literature on all recent studies which describes sortases and their various classes in Gram-positive bacteria, biochemistry of sortase transpeptidation, the structure of sortase family, bioengineering application, and the major objectives of the thesis. Chapter 2 is basically about the general materials and methods used in this thesis work. Chapter 3 deals with the insilico characterization of sortase and SDPs of *C. glutamicum*. Chapter 4 describes three distinct variants of *C. glutamicum, srtE* deletion mutant, complement, and overexpressed strain to study the potential influence of *srtE* on cell development, morphology, and physiology. Chapter 5 describes the biochemical characterization, substrate specificity, and site-directed mutagenesis of recombinant Sortase E, Chapter 6 demonstrates the C-terminal mediated immobilization of two recombinant proteins using engineered Sortase E and Chapter 7 summarizes the major highlights of the present work and conclusions from the results obtained and critically analyzed the prospects of the work.

1.3. Review of Literature

1.3.1. Overview of sortase

Sortases are membrane-bound enzymes that cleave the five amino acid long sorting motif at the C-terminus of the secreted protein to form an isopeptide bond between the secreted protein and the peptidoglycan. The sortases are membrane-bound cysteine transpeptidase enzymes, exclusively found in Gram-positive bacteria involved in covalent attachment of the surface proteins to the cell wall of peptidoglycan after secretion via sec-

dependent pathway (Pallen et al., 2001; Comfort, 2004). They belong to a family of membrane-anchored enzymes which covalently mount specific target proteins to the growing cell wall of Gram-positive bacteria by peptide bond formation. Sortase plays a pivotal role in virulence and pathogenesis and is not indispensable for growth and viability in Gram-positive bacteria (Mazmanian et al., 2000; Bierne et al., 2002; Weiss et al., 2004; Cheng et al., 2009). Sortases are generally classified within cysteine transpeptidases which associates with hydrolysis rather than peptide bond formation. All sortase contains His-Cys-Arg as the catalytic triad to catalyze transpeptidation reaction (Ton-That et al., 1999). The cysteine in the active site of the enzyme is involved in bond cleavage and formation of a stable thioacyl intermediate that is relieved by the nucleophilic attack of the amino group (pentaglycine cross-bridge) in peptidoglycan synthesis precursors. In site-directed mutagenesis, the replacement of cysteine (at position 184 in *S. aureus*) with alanine abolishes sortase catalytic activity in vitro and in vivo (Frankel et al., 2007). The His residue is believed to have a dual-acid/base role, donating a proton to the leaving amide nitrogen during the cleavage reaction and accepting a proton from the amino group of the second substrate to allow nucleophilic attack by the unprotonated amine. The Arg side chain is implicated in substrate binding and possibly in the stabilization of a presumed oxyanion intermediate.

A very few studies were reported on sortases of nonpathogenic bacteria with more focus given on sortase in probiotics and their relevance in the interaction with host cells. The presence of sortase in probiotic lactic acid bacteria (LAB) provided a new avenue to look into the role of this enzyme on probiotic attributes, such as adhesion, mucus barrier function, immune signaling, and nutrient uptake (Call and Klaenhammer, 2013). The hypothesis that sortase enzymes may play a crucial role in bacterial physiology as well as mediating bacterial-host interactions, has accelerated the study of this enzyme in different species of LAB.

Sortase-displayed proteins are involved in a wide range of physiological processes that are essential for cell survival, including cell adhesion, nutrition acquisition, immunological evasion, aerial hyphae growth, and sporulation (Weiss et al., 2004; Wang et al., 2009) **(Figure 1.1)**.

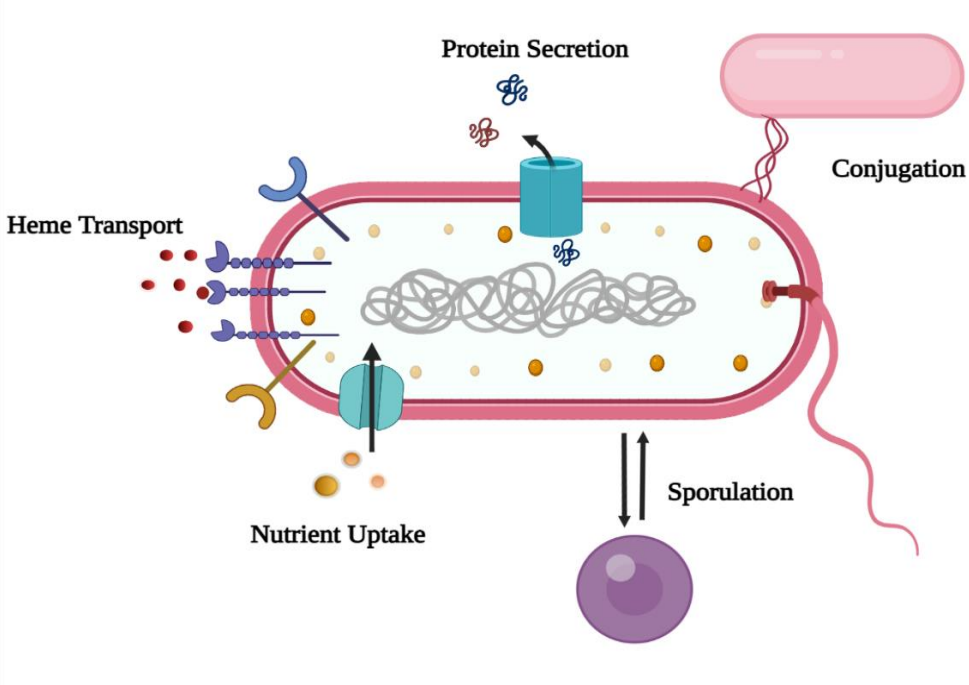


Figure 1.1. Physiological functions of sortases in a Gram-positive bacterial cell wall
Cellular functions of sortases include pili assembly, cell adhesion to the host tissues, binding of surface proteins to the cell wall, spore formation, uptake of nutrients and iron from the surrounding environment.

Sortase-recognized proteins have a C-terminal cell wall sorting signal which consists of a pentapeptide motif, a transmembrane region, and a positively charged lysine or arginine tail (Fischetti et al., 1990).

Sortases are not only found in Gram-positive bacteria but some sortase genes and potential substrates have also been discovered in Gram-negative bacteria that have yet to be studied. Some Gram-negative bacteria such as *Shewanella putrefascienes*, *Shewanella*

oneidensis, *Microbulbifer degradans*, *Colwellia psychrerythraea*, and *Bradyrhizobium japonicum* have a gene encoding a single sortase-like protein and a potential sortase substrate (Comfort and Clubb, 2004).

1.3.2. Sortase-mediated cellular functions

Sortases in bacteria have two distinct functions: they either covalently attach surface proteins directly to the cell envelope or they assemble long proteinaceous fiber-like structures called pili that are involved in bacterial adhesion (Jacobitz et al., 2017).

1.3.2.1. Attachment of surface protein

SaSrtA has served as a model for studying the mechanism of sortase by covalently anchoring surface proteins to the cell wall (Ton-That et al., 1999). The SrtA of *S. aureus* covalently anchors the surface proteins onto the bacterial cell wall via a C-terminal cell wall sorting signal with an LPXTG recognition motif followed by a stretch of hydrophobic amino acids and a positively charged tail (Das et al., 2017; Novick, 2000). The cell wall anchoring proteins are synthesized within the cytoplasm and translocated across the membrane through the Sec machinery. The sortase recognizes the anchoring proteins followed by a nucleophilic attack at the active site of the cysteine and cleaves the C-terminal of the LPXTG motif between threonine and glycine forming a thioester intermediate complex which is then covalently anchored on the pentaglycine cross-bridge of lipid II. The lipid II-protein complex then gets attached to the cell wall via transglycosylation and transpeptidation reactions (Perry et al., 2002) **(Figure 1.2)**.

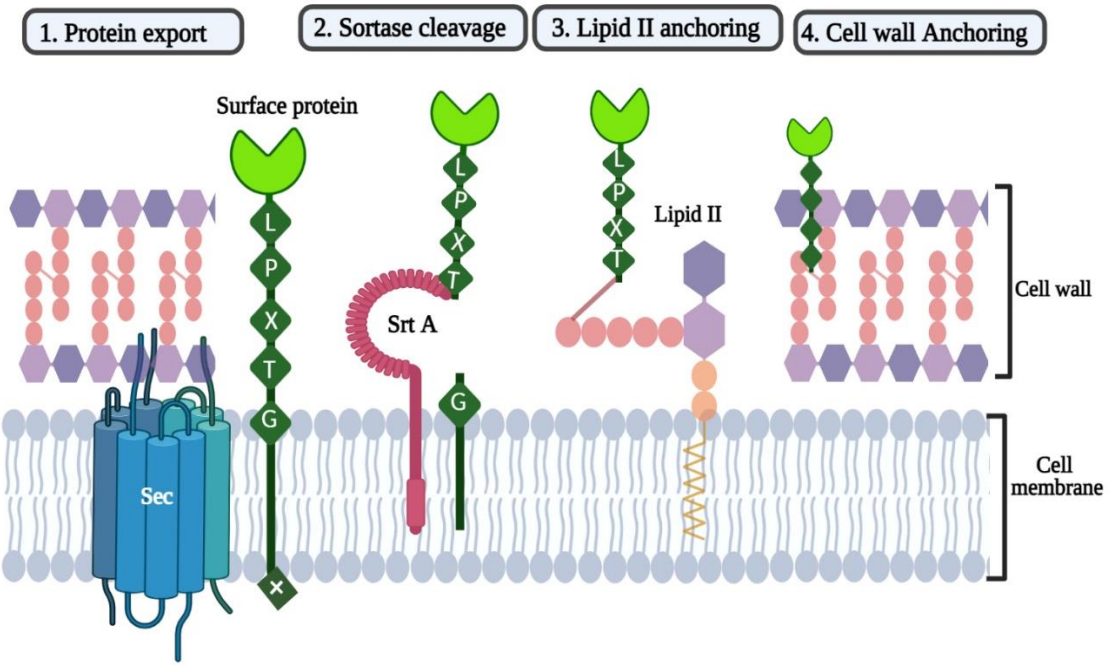


Figure 1.2. Illustration of Sortase A transpeptidation reaction in *S. aureus*

1) Protein synthesized in the cytosol is translocated by the Sec machinery and anchored on the cell membrane. 2) Sortase recognizes the LPXTG sorting motif at C-terminus and cleaves between threonine and glycine. 3) Sortase assembles a protein complex and undergoes a nucleophilic reaction with a lipid II molecule. 4) Through the transpeptidation reaction, the lipid II–protein molecule is further anchored to the cell wall.

1.3.2.2. Pili assembly

The genome of *C. diphtheriae* NCTC 13129 harbors six sortases like genes (named *srtA-F*), five of which presumably assembles three distinct types of pilus structures- SrtA for the SpaA-type pilus, SrtB or SrtC for the SpaD-type pilus, and SrtD or SrtE for the SpaH-type pilus (Spa for sortase-mediated pilus assembly) which are polymerized by specific Class C sortases and SrtF which belongs to class A sortase, catalyzes the anchoring of pilin monomers on the bacterial surface (Ton-That and Schneewind, 2003; Gaspar and Ton-That, 2006; Swaminathan et al., 2007). All three pilus structures share a similar architecture, a major pilin

(designated as SpaA, SpaD, and SpaH) along the pilus shaft joined to the minor pilins (designated as SpaB, SpaC, SpaE, SpaF, SpaI, and SpaG) located at the tip and base of the pilin **(Figure 1.3)**. To analyze the functions of sortase in pili formation, Mandlik and his coworkers (2007) constructed an isogenic mutant strain of NCTC 13129 devoid of all six sortase genes (*srtA–F* mutant) that exhibited a severe defect in adherence to epithelial cells.

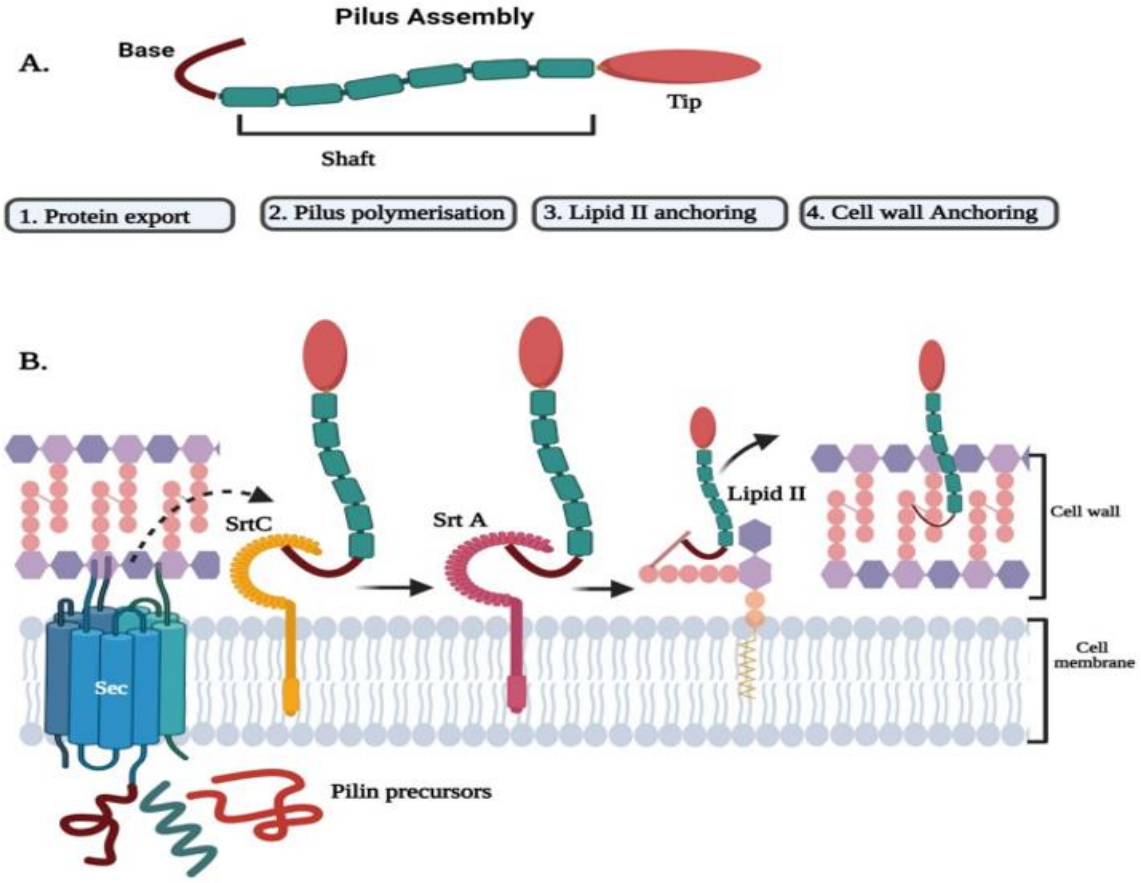


Figure 1.3. Schematic illustration of sortase-mediated pili

(A) Pilus assembly which consists of tip pilin (red), shaft (green), and the base (brown); **(B)** 1) The pilin precursors synthesized from the cytosol enter through the sec machinery. 2) The pilin-specific sortase SrtC (yellow) recognizes the tip pilin and cleaves at the sorting motif and forms an acyl sortase complex. The pilus specific sortase receives a nucleophilic attack from the lysine side chain from the backbone pilin to form a covalent bond between the pilins and undergoes pilin polymerization. 3) The housekeeping sortase undergoes a nucleophilic attack from lipid II molecules. 4) The polymerized pilin is further anchored to the cell wall.

The single pilus-specific SrtA encoded within the *spaA* gene cluster specifically catalyzes the covalent crosslinking of individual pilin monomers and also anchors pili to the cell wall. To analyze the functions of SrtA, immunoelectron microscopy, and biochemical analysis showed that a strain expressing only SrtA secretes significant amounts of polymerized pilins into the culture medium, indicating that one or more sortases might be involved for efficient cell wall anchoring of pili (Ton-That et al., 2004; Mandlik et al., 2007). Indeed, the strain with the deletion of housekeeping gene *srtF* releases SpaA polymers into the culture medium. Thus, two sortases are involved in pilus biogenesis, a pilus-specific sortase for pilin polymerization and the housekeeping sortase for efficient anchoring of pili to the cell wall (Swaminathan et al., 2007; Mandlik et al., 2010). The deletion of *srtA* or *spaA* gene completely abrogates the assembly of SpaA pili and deletion of *spaC* and *spaB* did not abolish SpaA pilus formation. This evidence suggests that SrtA catalyzes the assembly of SpaA pilus and SpaA alone is sufficient to mediate the polymerization of a secreted protein (Ton-That and Schneewind, 2003; Marraffini et al., 2006). Unlike the SpaA-type pili, which are assembled by a single sortase, two sortases; SrtB and SrtC catalyze the assembly of the SpaD-type pili. The deletion of *srtB* alone or both *srtB* and *srtC* abrogated the incorporation of SpaE into SpaDF pili. These results demonstrate that SpaDEF pilus assembly specifically requires SrtB for the incorporation of SpaE into SpaDF pili and whose assembly requires either SrtB or SrtC (Gaspar and Ton-That, 2006). Likewise, the SpaH pilus is independently assembled and different from the other two corynebacterial pili. The SrtD is specifically required for the incorporation of SpaH into SpaIG pili, whose assembly requires either SrtD or SrtE, while other remaining sortases are dispensable (Gaspar and Ton-That, 2006; Swierczynski and Ton-That, 2006). Thus, the housekeeping sortase contributes to efficient cell wall anchoring with other sortases involved in SpaD and SpaH–type pilus (Swaminathan

et al., 2007).

1.3.3. Sortase transpeptidation

The multiple kinds of sortase all use the same 'ping-pong bi-bi' transpeptidation reaction mechanism, in which the sortase initially binds the 5-amino acid recognition motif in the C-terminal region of the substrate protein (Frankel et al., 2005, 2007). Between the catalytic cysteine and the substrate threonine, sortases create thioacyl enzyme intermediates, which are resolved by a nucleophilic attack by bacterial cell wall components. Since the discovery of several types of sortases, *S. aureus* sortase A (SrtA) has served as a model for studying how these enzymes work (Mazmanian et al., 2000). SrtA catalyzes two consecutive processes after recognizing cell wall surface protein: (i) thioesterification and (ii) transpeptidation. The pentaglycine sequence on the surface proteins released via the cytoplasmic membrane is recognized first by the enzyme. The SrtA cleaves the scissile bond between threonine and glycine residues in the second phase, forming an acyl-enzyme intermediate that then transfers the carboxyl of threonine, which is amide-linked, to the pentaglycine cross-bridge of lipid II (Marraffini et al., 2006). Finally, transglycosylation and transpeptidation processes insert the lipid II-surface protein complex into the peptidoglycan (Paterson and Mitchell, 2004; Spirig et al., 2011). Because the composition of peptidoglycan layers in the cell envelope varies from strain to strain, the sortase enzyme accepts nucleophiles that may differ in different Gram-positive bacteria. For example, in *Bacillus anthracis*, diaminopimelic acid, which cross-bridges the peptidoglycan, is assumed to be the point of attachment for the sortase substrate proteins (Severin et al., 2004; Zhang et al., 2009; Spirig et al., 2011). The acyl-enzyme intermediate complex was demonstrated to be

hydrolyzed in the absence of a dedicated nucleophile, resulting in recognition sequence cleavage but no formation of a new peptide bond (Frankel et al., 2005) **(Figure 1.4)**.

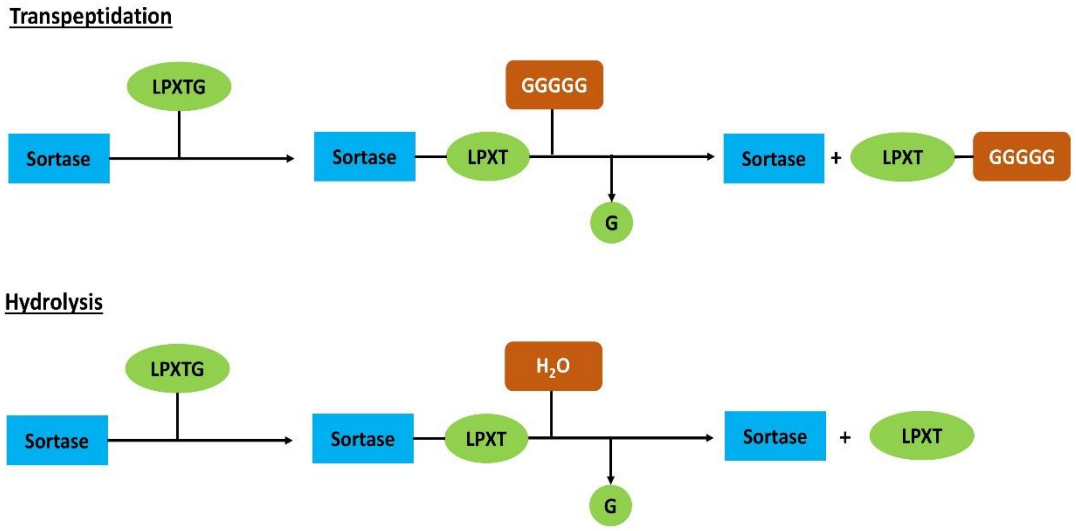


Figure 1.4. Ping-pong transpeptidation by sortase

A glycine residue is released during the cleavage of the LPxTG motif to produce the acyl-enzyme LPET-SrtA. The enzyme undergoes transpeptidation with pentaglycine (GGGGG) to produce an 'LPETGGGGG' peptide or hydrolyzed to release an 'LPET' peptide from this acyl state.

1.3.4. Sortase classification

Many pathogenic and nonpathogenic bacteria have many sortases in their genomes, each of which catalyzes comparable transpeptidation processes to sort different proteins to the cell surface by detecting their class-specific sorting signals. Based on their phylogeny, projected substrate preference, and functionality, sortase transpeptidase can be classified into several classes, A to F (Bradshaw et al., 2015) **(Table 1.1)**. At the active site of sortase superfamilies, there is a conserved signature pattern TLXTC (X can be any amino acid). Many Gram-positive bacteria have class A sortases, which play a housekeeping role in anchoring a vast number of proteins to the cell membrane. By connecting iron acquisition

proteins to the cell membrane, class B sortases have been linked to iron homeostasis (Mazmanian et al., 2002). In bacteria, the class C sortases catalyze the transpeptidation events by catalyzing the polymerization of pilin subunits. The class D sortase anchor proteins are involved in sporulation. Actinobacteria have classes E and F, with class E anchoring proteins to the cell surface that promotes aerial hyphae development (Kattke et al., 2016), whereas the role of class F is still not known.

Table 1.1. Classification of sortases

Sortase Class	Cleavage site	Significant role	Bacterial genus
A	LPXTG	Surface protein anchoring	*Staphylococcus, Listeria, Streptococcus, Bacillus, Clostridium, Enterobacter, Lactobacillus*
B	(N/S/P) PXTG	Heme uptake	*Bacillus, Listeria, Bacillus,*
C	LPXTG	Pili assembly	*Corynebacterium, Streptococcus, Clostridium, Actinomyces, Enterobacter, Lactobacillus*
D	LPXTA	Spore formation	*Bacillus*
E	LAXTG	Aerial hyphae formation, Surface protein anchoring, Pilus attachment	*Corynebacterium, Streptomyces, Actinomyces*
F	LPXTG	Unknown	*Propionibacterium*

1.3.4.1. Class A sortase

Class A sortases (SrtA, EC number: 3.4.22.70) are membrane-bound transpeptidase that catalyzes the transfer and covalent immobilization of surface proteins to peptidoglycan in the cell wall of Gram-positive bacteria (Cascioferro et al., 2014; Bradshaw et al., 2015; Jacobitz et al., 2017). *S. aureus* anchors a vast array of virulence-associated surface proteins

to the cell wall, which is catalyzed by a cysteine transpeptidase enzyme called Sortase A (Mazmanian et al., 2000). *S. aureus* Sortase A (SrtA) has been the prototype for understanding the mechanism of action of these enzymes (Ton-That et al., 1999). *S. aureus* attaches several surface proteins which are characterized by a C-terminal LPXTG motif, including protein A (Spa), two fibronectin-binding proteins (FnbpA and FnbpB), two clumping factors (ClfA and ClfB), a collagen-binding protein (Cna), and three serine-aspartate repeat proteins (SdrC, SdrD, and SdrE). The deletion of *srtA* has led to the failure of surface protein anchoring to the cell wall (Clancy et al., 2010).

The genome of *Listeria monocytogenes* contains the highest number of genes encoding surface proteins in the range of 40–45 with an LPXTG bearing motif at the C-terminal. Inactivation of the *srtA* gene in *L. monocytogenes* altered the expression of specific anchored surface proteins containing the canonical LPXTG motif, ultimately decreasing the ability of the bacterial adhesion, invasion of eukaryotic cells, and affecting host immune responses (Bierne et al., 2002). The RT-PCR and Western-blot data analysis demonstrates that the lack of SrtA alters the expression of LPXTG surface proteins and does not completely abolish the strong attachment of certain surface proteins to cell-wall peptidoglycan (Mariscotti et al., 2012).

The genome analysis of *Streptococcus agalactiae* NEM 316 encompasses one class A and 35 surface proteins containing a cell wall sorting signal motif (26 proteins had an LPXTG motif, 4 had an IPXTG motif, 2 had an LPXTS motif, 2 had an LPXTN motif, and 1 had an FPXTG motif (Glaser et al., 2002). *S. agalactiae* NEM 316 strain lacking *srtA* gene was found defective in anchoring cell surface proteins Alp2 (GBS 0470) and ScpB (GBS1308) bearing LPXTG and LPXTN signature sequence at the C-terminal (Lalioui et al., 2005). To determine the activity of SrtA in fibronectin and fibrinogen binding, a simple binding assay

(ELISA) was performed to compare the binding properties of SrtA$^-$ mutant with those of wild-type and complemented strains. This resulted in reduced binding of fibronectin and fibrinogen in SrtA$^-$ mutant strain. Thus, it is conceivable that the ScpB, a fibronectin protein (Beckmann et al., 2002), and FbsA (GBS1087) a major GBS fibrinogen-binding protein is a SrtA-dependent LPXTG-containing protein (Schubert et al., 2002, 2004). The inactivation of *srtA* in the NEM 316 strain decreased its adherence to human epithelial cell lines (A549, Caco-2, and HeLa) and rat cell lines (L2). Interestingly, the deletion of *srtA* strains did not alter the virulence in the neonatal rat sepsis model as compared to the wild-type parental strain (Lalioui et al., 2005).

Streptococcus sanguinis is a member of the oral mitis group of Streptococci and the initial colonizer for dental plaque formation (Kolenbrander, 2017). Although, during oral injuries, the harmless members of the group invade into the bloodstream causing bacteremia and infective endocarditis (Morita et al., 2014). The deficiency of *srtA* in *S. sanguinis* causes an overall reduction in virulence in association with cell surface proteins and decreased cell surface hydrophobicity. Thus, SrtA of oral streptococci is considered an important molecule for colonization on the smooth surface of the teeth and a drug target to prevent dental biofilm formation (Yamaguchi et al., 2006).

Streptococcus uberis contains only a single copy of sortase A (*srtA*), encoding a transamidase capable of anchoring surface proteins bearing the LPXTG or LPXXXD motifs at the bacterial cell surface (Egan et al., 2010). The *srtA* deficient strain of *S. uberis* was unable to colonize the bovine mammary gland to induce clinical mastitis in dairy cattle indicating that several SrtA-anchored proteins are likely to be involved in the pathogenesis of this bacterium.

B. anthracis variants lacking the *srtA* gene did not anchor the collagen-binding MSCRAMM (microbial surface components recognizing adhesive matrix molecules) BasC protein to the bacterial cell wall as opposed to its parent strain. Recombinant expressed and purified SrtA catalyzed the cleavage reaction with LPETG and LPATG peptides, consistent with the notion that *B. anthracis* SrtA is responsible for the cell wall anchoring of surface proteins with an LPXTG motif (Gaspar et al., 2005). GamR, a *B. anthracis* phage receptor, is known to be anchored by SrtA. The strains containing the three chimeric proteins BasB, BasE, and BasJ, were resistant to lysis in the *srtA* mutant and complementation experiments restored phage lysis indicating they are anchored by Sortase A. Furthermore, BasA was shown to be anchored by SrtA using immunoblot analysis of peptidoglycan fractions (Aucher et al., 2011). Apart from the functional roles, BaSrtA has been characterized structurally. The NMR structure of BaSrtA shows several unique active site features that include the presence of an N-terminal extension that contacts the catalytically essential histidine which might be involved in lipid II recognition. Another feature very unique to this protein is a large structurally disordered active site loop correlated to the attachment of proteins to the m-DAP moiety of lipid II. Based on the NMR structure a lock-and-key mechanism is proposed for recognizing the sorting signal (Weiner et al., 2010).

Dieye et al. (2010) identified and studied a class A sortase in *Lactococcus lactis* IL1403 and showed that it is responsible for the cell wall anchoring of at least five LPXTG-containing proteins. We, therefore, propose that SrtA is the housekeeping sortase in *L. lactis*. Surface proteins are important factors in the interaction of probiotic and pathogenic bacteria with their environment or host. The sortase mutant and one sortase-dependent protein (mucus-binding homolog) mutant showed a significant reduction in adherence to human epithelial cell lines in the case of *Lactobacillus salivarius* UCC118. Van Pijkeren et al. (2006) identified 10

sortase-dependent surface proteins in *L. salivarius* UCC118, by the comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *L. salivarius.*

A sortase gene, *srtA* was identified in *Lactobacillus acidophilus* NCFM (LBA1244) and *Lactobacillus gasseri* ATCC 33323 (LGAS_0825). Additionally, eight and six intact sortase-dependent proteins were predicted in *L. acidophilus* and *L. gasseri*, respectively. Inactivation of sortase did not cause significant alteration in growth or survival in simulated gastrointestinal juices. Meanwhile, both srtA$^-$ mutants showed decreased adhesion to porcine mucin in vitro. Murine dendritic cells exposed to the srtA$^-$ mutant of *L. acidophilus* or *L. gasseri* induced lower levels of pro-inflammatory cytokines TNF-α and IL-12, respectively, compared with the parent strains (Call et al., 2015). This study shows that sortase-dependent proteins contribute to gut retention of probiotic microbes in the gastrointestinal tract.

Among the various classes of identified sortases, most biochemical investigations have been conducted with Sortase A (SrtA). SrtA is usually found in a single copy per genome and plays a critical role in cell adhesion and invasion of host cells (Muñoz-Provencio et al., 2012). The calcium ions stimulate cysteine at the active site of SrtA which attacks the carbonyl carbon of the threonine residue in the LPXTG motif, breaking the threonine and glycine peptide bond, which creates a thioester-linked acyl-enzyme intermediate (Spirig et al., 2011).

1.3.4.2. Class B sortase

The mechanism of Sortase B is similar to that of Sortase A, where *S. aureus* Sortase B enzyme attaches the heme transporter IsdC protein which is a major component of the iron-regulated surface determinant system that scavenges the heme-iron from hemoglobin. The SrtB anchors IsdC to uncross-linked peptidoglycan instead of heavily cross-linked peptidoglycan. The *srtB* and *isdC* genes are located together in the *isd* iron-acquisition

operon. However, in contrast to SrtA, SrtB recognizes NPQTN sorting signals from *S. aureus* (Mazmanian et al., 2002). Gene knockout studies in *S. aureus* revealed that the abolition of the *srtB* gene is responsible for virulence and does not affect cell viability.

Therefore, as in *S. aureus,* the listerial SrtB represents the second class of sortase in *L. monocytogenes,* generally expressed in operons containing genes encoding their substrates with NPKSS/NAKTN recognition motifs. The *srtB* deletion mutants do not have defects in bacterial entry, growth, or motility in tissue-cultured cells and do not show attenuated virulence in mice. SrtB-mediated anchoring could therefore be required to anchor surface proteins involved in the adaptation of this microorganism to different environmental conditions (Garandeau et al., 2004).

B. anthracis srtB, which encodes Sortase B, anchors IsdC to the cell wall envelopes of vegetative bacilli a heme-iron binding surface protein. Sortase B cleaves IsdC between the threonine and the glycine of its NPKTG motif sorting signal. Isogenic variants lacking either *srtB* or *isdC* display significant growth defects due to deficiencies in heme-iron scavenging, suggesting that IsdC binding to heme-iron in the cell wall envelope contributes to bacterial uptake of heme.

Clostridium difficile encodes for a single sortase enzyme Sortase B (CD2718), belonging to the class B sortases, recognizes an entirely different sorting motif (SPxTG or PPxTG). Remarkably, unlike SrtB from other organisms, the function of this enzyme is not associated with heme or heme acquisition proteins instead possibly functions as a housekeeping sortase. The mutation of the catalytic cysteine or the addition of small inhibitors like MTSET abolishes the Sortase B activity. In vitro ligation of a natural cell wall nucleophile, DAP is demonstrated (Donahue et al., 2014; van Leeuwen et al., 2014).

Based on bioinformatics analyses, seven putative sortase substrates have been identified (van Leeuwen et al., 2014). However, two putative substrates, CD0183 and CD2768, containing an SPXTG motif were not cleaved or anchored to the cell wall by sortase in an experimental observation (Peltier et al., 2017). Another two substrates, CD0386 and CD3392, display very high (94 %) sequence similarity, of which CD0386 is attached to the cell wall by the sortase as demonstrated by biochemical analysis of subcellular fractions in a sortase knockout strain (Chambers et al., 2015). Adhesins, CD2831, and CD3246, are anchored proteins on the cell wall by SrtB activity. The functional role, particularly of CD2831 in binding to collagen, has been demonstrated in vitro through the cleaving activity of a Pro-Pro endopeptidase to release CD2831 and CD3246 from the cell surface and is regulated by C-di-GMP (Hensbergen et al., 2015; Peltier et al., 2015). The last putative substrate CD2537 contains only a weak signal peptide (Corver et al., 2017).

However, in contrast to SrtA, SrtB recognizes an NPKTG, NPQTN, and (S/P)PXTG sorting signals from *B. anthracis*, *S. aureus*, and *C. difficile* strain 630 respectively (Mazmanian et al., 2002; Maresso et al., 2006). Unlike Sa-SrtA, Sa-SrtB does not require metal binding for activity (Mazmanian et al., 2002).

1.3.4.3. Class C sortase

Among all sortases, class C sortases represent the largest and most heterogeneous group in a genome (Cozzi et al., 2011) and they are integral membrane cysteine transpeptidase of Gram-positive bacteria (Marraffini et al., 2006). Class C sortase cleaves at LPXTG-like motifs similar to class A sortase, but only sortase C can polymerize the pilus-forming proteins to form high molecular weight structures via transpeptidation mechanism. As pilin subunits emerge, sortases cleave the peptide bond between threonine and glycine of

the LPXTG motif of the pilin proteins and covalently join the C-terminus of one pilin subunit to a Lys side-chain NH_2 group on the next subunit (Ton-That et al., 2004; Marraffini et al., 2006). *C. diphtheriae* type strain NCTC 13129 produces three distinct types of pilus structures, SpaA, SpaD, and SpaH-type pili, which are polymerized by specific class C sortases (Ton-That and Schneewind, 2004) and covalently linked to the cell surface by Sortase F (Dramsi et al., 2005).

Streptococcus agalactiae contains four genes encoding class C sortases (SrtC) were found in NEM316, 2603V/R, and A909 genome sequences which are arranged tandemly in two different loci, *srtC1-srtC2* and *srtC3-srtC4* coding for pilus biogenesis (Dramsi et al., 2006; Khare et al., 2011). Based on the electron microscopy and immunogold labeling, the NEM316 strain assembles pili from *the srtC3-srtC4* locus and encodes three pilin subunits, the major pilin, and two minor pilins. Either SrtC3 or SrtC4 is required for polymerization of pili and housekeeping SrtA anchors the polymerized pili to the cell wall (Dramsi et al., 2006).

The genome of *S. pneumoniae* TIGR4 contains *srtC-1, srtC-2,* and *srtC-3* for pilus assembly. The primary SrtC-1 catalyzes the polymerization of major pilin subunit RrgB. SrtC-2 binds with RrgA and attaches to other pilins. SrtC-3 preferentially binds with the RrgC pilin subunit but does not have a strong affinity as SrtC-1 with RrgB (Lemieux et al., 2008; Shaik et al., 2014; Naziga and Wereszczynski, 2017).

The Class C sortase is involved in pili assembly, especially in the *B. cereus* vegetative cells. The pilus operon consists of three genes (*bcpA-srtD-bcpB*) and deletion of the complete operon/inactivated *srtD* leads to the absence of the pili (Budzik et al., 2007). BcpA is polymerized by Sortase D even in the absence of the BcpB unit. BcpB the minor pilin is cleaved by Sortase D at C-terminal threonine of IPNTG sorting signal and then amide-linked to the YPKN motif of BcpA (Budzik et al., 2009). The cleavage reaction of BcpB is very

specific to Sortase D and is not cleaved by Sortase A. In the case of BcpA, the LPXTG sorting signal is cleaved at the C-terminal threonine and linked to the amino group of lysine in the YPKN motif of another BcpA subunit by Sortase D. BcpA is a substrate for both Sortase D and A, and Sortase A cleaves the BcpA unit at its LPXTG to terminate the pilus assembly and bond to the cell wall cross-bridge (Budzik et al., 2009).

Actinomyces oris consist of class C sortases, SrtC1 and SrtC2, which are involved in the assembly of two distinct forms of pili on the *Actinomyces* cell surface (Mishra et al., 2007). The SrtC1 is involved in the type I fimbriae formation which is further involved in the adhesion of *Actinomyces* to the tooth surface by recognition of proline-rich receptors (Wu et al., 2011). However, SrtC2 results in the assembly of type II fimbriae which are essential for bacterial adhesion to the oral streptococci and host cells by recognizing the polysaccharide receptors (Wu et al., 2012).

Genome analysis of *Enterococcus faecalis* V583 revealed the presence of one class C sortase (SrtC [EF_0194 for biofilm and pilus-associated sortase]) (Paulsen et al., 2003; Dramsi et al., 2005). The *ebp* operon in *E. faecalis* encodes the Ebp pilus structural subunits EbpA, EbpB, and EbpC and the pilus-associated SrtC. The *srtC* is shown to be necessary for the production of the Ebp pili and important for biofilm formation and endocarditis.

1.3.4.4. Class D sortase

Sortase D enzymes can be found in the genome of many bacilli and clostridia. *B. anthracis* Class D SrtC anchors two substrates, BasH and BasI to the cell wall of sporulating *B. anthracis.* LPNTA sorting signal of two substrates is cleaved by Sortase C (SrtC) at the C-terminal threonine (T) of the substrate to the amino group of DAP cross-bridges targeting the polypeptides to the cell wall of sporulating bacilli. Sortase C is also required for the formation

of infectious spores in the tissues of animals. Where Sortase C acts on two different substrates in two different subcellular compartments, the different surface proteins decorate different compartments of the sporulating cell, with BasI present in divisional cells and BasH present in forespores (Marraffini and Schneewind, 2006, 2007). The knockout experiments in *B. anthracis* demonstrated the requirement of sortase activity in sporulation. The spore formation requires BasH and BasI proteins, which contain LPNTA as a potential sorting signal (Marraffini and Schneewind, 2006). Moreover, *Clostridium perfringens* transpeptidase which belongs to the class D family of sortase enhances its catalytic activity with the presence of magnesium ion concentration towards the LPQTGS signal motif (Suryadinata et al., 2015).

1.3.4.5. Class E sortase

Class E sortases are predominant in soil and freshwater-dwelling GC rich genome of Actinobacteria (e.g. *Corynebacterium* and *Streptomyces* genera) (Comfort, 2004).

S. coelicolor is one of the best-studied members of Actinobacteria which uses two class E enzymes, ScSrtE1 and ScSrtE2 to decorate its surface (Duong et al., 2012; Kattke et al., 2016). Its life cycle encompasses three morphologically distinct stages: vegetative hyphae, aerial hyphae, and spores. It is known to encode two sortases of class E (Pallen et al., 2001), SrtE1 and SrtE2 which recognize and cleave LAXTG-containing peptides in vitro. Chaplin proteins (Chp A, B, and C) required for the aerial development of this organism contain a LAXTG-sorting signal rendering them as putative sortase substrates (Elliot et al., 2003). Sortase class E enzymes (SrtE1 and SrtE2) anchor ChpC protein to the cell wall in vivo. Indeed, *srtE1*/*srtE2* double mutants delay the formation of aerial hyphae with hindered sporulation and cease to display short Chaplin proteins (Duong et al., 2012).

S. avermitilis encodes for at least four putative class E sortase enzymes (Duong et al.,

2012) of which only, SrtE3 (SAV4333) has been biochemically characterized. Class E sortase of *S. avermitilis* produces useful transpeptidation yield and can fruitfully complement SaSrtA in peptide ligation and protein engineering endeavors. The Ca^{2+}-independent activity together with the preference for an altered substrate LAXTG instead of LPXTG makes this enzyme an attractive tool for intracellular protein labeling.

Actinomyces oris is an oral bacterium, formerly known as *Actinomyces naeslundii*, which results in the formation of dental plaque (Persson, 2011). It consists of a housekeeping sortases (SrtA), which belongs to class E sortase (Spirig et al., 2011). The housekeeping sortases generally do not show any role in cell viability, however, SrtA of *A. oris* was observed to show some contradictory results. The deletion of the *srtA* gene was found to be lethal for the bacterial cells, which resulted in excessive membrane accumulation of the surface glycoprotein protein perturbing the cell envelope to block the growth and viability of the cells (Li et al., 2014).

Class E enzymes of *C. diphtheriae* appear to perform a housekeeping function similar to class A sortase enzymes (Ton-That and Schneewind, 2003) which contains a conserved tyrosine residue that may enable it to preferentially recognize alanine instead of proline in the cell wall sorting signal.

1.3.4.6. Class F sortase

Sortase F is the only housekeeping sortase characterized in the genome of *Propionibacterium acnes*. Sortase F from *P. acnes* shows a behavior similar to sortases from class A in terms of pH dependence, recognition sequence, and catalytic activity, furthermore its activity is independent of divalent ions, which contrasts to Sortase A from *S. aureus*. The Sortase F can be used as a powerful tool alternative to Sortase A for protein engineering

applications (Girolamo et al., 2019). *S. coelicolor* encodes five class F enzymes, however, their functions are yet unknown (Spirig et al., 2011).

1.3.5. Structural elucidation of sortases

Sortases have offered a tempting target for structural and mechanistic investigation due to their critical involvement in bacterial physiology and lack of sequence similarity to known eukaryotic proteins. The first structure of *S. aureus* SrtA **(Figure 1.5)** was determined by NMR spectroscopy, was the first to be reported (PDB 1IJA, Ilangovan et al. (2001), and was further analyzed by X-ray crystallography (Zong et al., 2004). The N-terminal signal peptide/membrane anchor (N59) was deleted from *S. aureus* SrtA to express the recombinant protein in soluble form along with an N-terminal six-histidyl tag. The purified sortase was used for in vitro activity testing and structural analyses. The structure revealed an eight-stranded beta-barrel flanked by one short alpha-helix and one 3_{10} helices, revealing a new structure. The interaction of two structural motifs, each with four strands, results in the sortase fold. One strand from each motif (β4 and β7) interacts for much of their length, while the remaining strands (β1, β2, β3 and β8, β6, β5) work together to construct the rest of the barrel, which is completed by the interactions of β1 and β5.

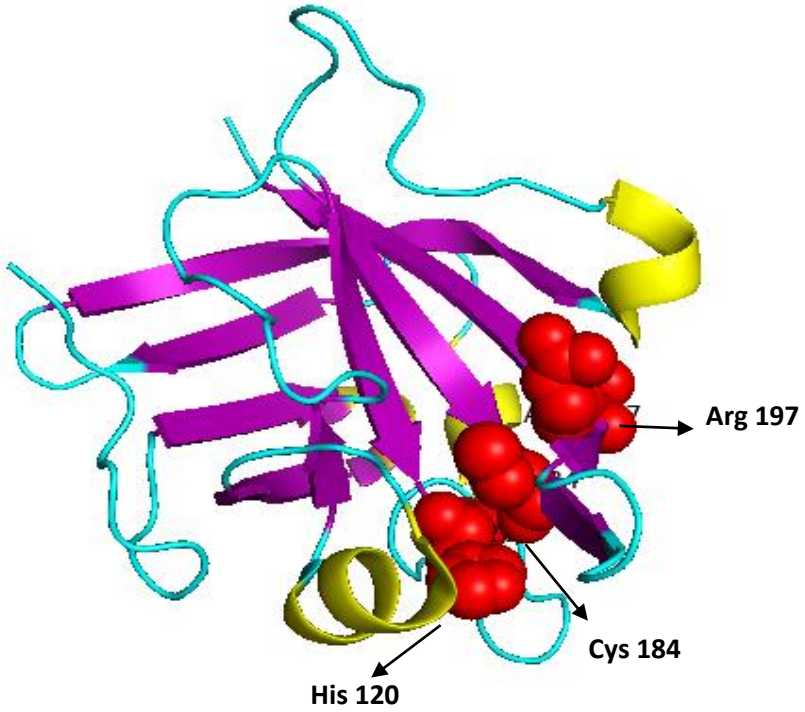


Figure 1.5. **Structure of *S. aureus* SrtA showing active sites**

Cartoon rendering of the NMR structure of *S. aureus* SrtA (PDB 1IJA, Ilangovan et al. (2001) with active sites.

Cys 184 and His 120 are found in or near the active site, and Arg 197 is close to the cysteine in the active site. His and Cys are located on two separate β-sheets, and the thiol group of Cys184 in SrtA N59 is positioned 7 Å away from the imidazole ring of His120 in sortase. Measuring mutant enzyme activity shows that the replacement of Cys 184, His 120, and Arg 197 with Ala completely reduces the enzyme activity. Therefore, three key residues that are responsible for the enzyme activity are Cys 184, His 120, and Arg 197. Between strands β7 and β8, the LPXTG substrate-binding site is placed in a concave plane, and the T and G peptide bond is near to Cys 184 and Arg 197. As a result, the enzyme action is driven by a Cys-Arg catalytic dyad, and the Arg 197 side chain, which can be shifted close to Cys 184 (3.5 Å), acts as an ionizable group to protonate the amide atom of the substrate scissile

bond and facilitate the Cys nucleophile attack. Additionally, Thr 180, Ile 182, and Ala 118 are conserved in most sortases seem to be responsible for binding to leucine and threonine of the LPXTG substrate (Maresso and Schneewind, 2008).

In *S. aureus* SrtA, the side chains of E105, E108, and D112 within the β3–β4 loop, as well as E171 from the β6–β7 loop, appear to create a structurally well-ordered calcium-binding region. In the absence of calcium, the β6/β7 loop is disordered and undergoes some conformational changes. As a result, Ca^{2+} activation activates sortase by a mechanism that affects the shape and dynamics of the active site loop, perhaps facilitating substrate binding (Ilangovan et al., 2001) **(Figure 1.6)**.

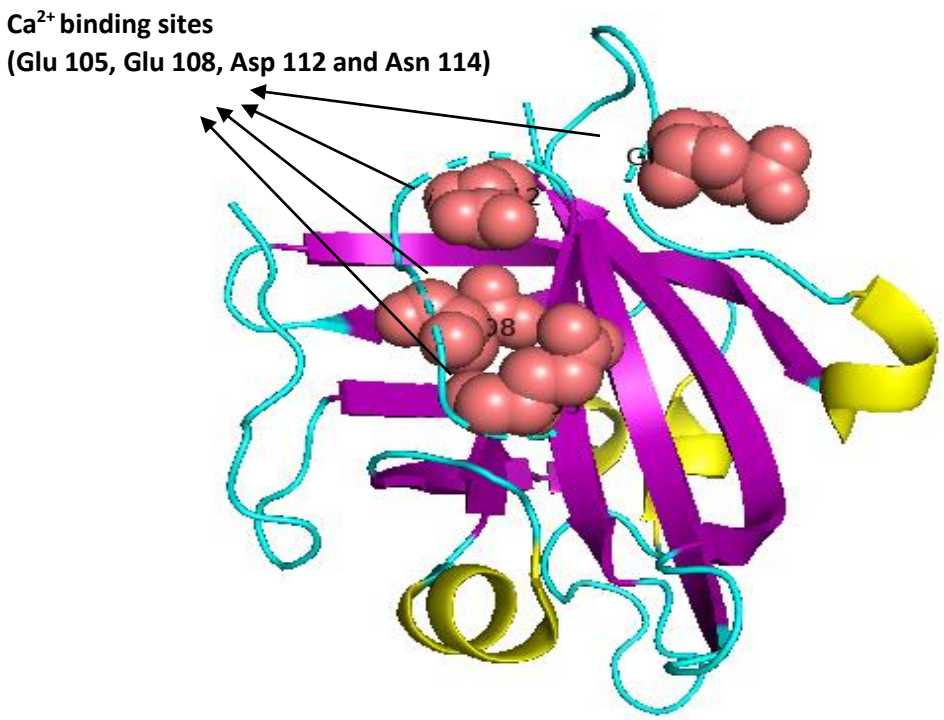


Figure 1.6. **Structure of *S. aureus* SrtA with Ca^{2+} binding sites**

Cartoon rendering of the NMR structure of *S. aureus* SrtA (PDB 1IJA, Ilangovan et al. (2001) with Ca^{2+}- binding sites at Glu 105, Glu 108, Asp 112, and Asn 114.

Unlike Sa-SrtA (*S. aureus* Sortase A), Sp-SrtA (*Streptococcus pyogenes* Sortase A) does not need Ca^{2+} to be activated. Although they have a 24 % sequence identity in the core catalytic domain, Sa-SrtA and Sp-SrtA have substantially similar structures. These two structures differ in some ways, particularly in the connecting loops and the N/C termini. There have been several shifts in the loop connecting the β2/ β3, β3/ β4, β6/ β7, β7/ β8, and β3 loop. The essential catalytic residues are Cys 208, His 142, and Arg 216, and unlike other sortases, His 142 is located close to the active site to play the job of protonating the intermediate (Marraffini et al., 2006).

Structures for other class A sortases, as well as members of classes B, C, D, and E, have been documented so far, in addition to SaSrtA. Sortases from different classes have a similar overall structure, but varied numbers of peripheral α-helices and other differences that presumably affect the function and regulate substrate specificity, according to structural studies. For example, in *S. aureus* Sortase B, an extended β6/β7 loop is involved in recognizing the NPQTN sorting signal substrate, while the two short extra N-terminal helices may play a role in attaching substrates to buried parts of the cell wall (Marraffini et al., 2006; Jacobitz et al., 2014). Nonetheless, in class C sortases, an extended N-terminal region has been demonstrated to function as a lid, occluding the active site and possibly regulating enzyme activity (Mandlik et al., 2008; Manzano et al., 2009).

Class A and B sortases detect their substrates and catalyze the transpeptidation reaction comparably, according to structural and computational analyses of Sortase A from *B. anthracis* and Sortase B from *S. aureus* in complex with respective sorting signals peptides (Jacobitz et al., 2014). These findings show that all sortases may share a common mode of action.

1.3.6. In vitro sortase activity and substrate specificity

Ton-That et al. (2004) demonstrated SaSrtA proteolytic activity against peptides with the LPXTG motif and its transpeptidation process in the presence of the nucleophile NH_2-Gly_3 demonstrated in experiments shortly after its discovery. These studies showed the SaSrtA cleaves between threonine and glycine at the LPXTG motif, however, in the absence of the nucleophile, the thioacyl intermediate is hydrolyzed, resulting in in the cleavage of the LPETG motif without the formation of a new peptide bond. Later, in addition to identifying the 3D structure of SaSrtA, Ilangovan and his colleagues (2001) discovered that in the presence of calcium ions, SaSrtA activity increased by 8-fold. FRET-based experiments, which use sorting signal peptides with donor and quencher fluorophores at either end were initially used to evaluate the kinetic characteristics of SaSrtA. The donor is liberated from the quencher when this peptide is cleaved either by hydrolysis, hydroxylaminolysis, or the native transpeptidation reaction, and this activity can be observed by an increase in fluorescence. While the first study only showed cleavage and hydroxylaminolysis, further investigations showed that SaSrtA can catalyze the native transpeptidation reaction in vitro using just short peptide substrates, one of which has a sorting signal motif and the other a series of 1 to 5 glycine. However, due to constraints associated with the fluorescence inner filter effect quenching at high substrate concentrations, HPLC-based methods were developed, allowing for more precise measurements (Kruger et al., 2004; Frankel et al., 2005). Although, enzyme kinetics of SaSrtA vary widely with other sortases depending on their substrate sequences **(Table 1.2)**.

Table. 1.2. Kinetic constants K_m and K_{cat} of different sortases

Enzyme	Peptide	K_m	K_{cat}	Reference
S. aureus SrtA	LPETG	7.33 ± 1.01 mM	0.086 ± 0.015 s^{-1}	(Frankel et al., 2005)
S. aureus SrtA	LPATG	17.5 μM	-	(Zhang et al., 2014)
S. avermitilis SrtE	LAETG	1.14 ± 0.14 mM	58.18 ± 3.47 s^{-1}	(Das et al., 2017)
C. difficile SrtB	SPKTG	74.7 ± 48.2 μM	$1.1\times10^{-3} \pm 6\times10^{-4}$ min^{-1}	(Peltier et al., 2017)
C. difficile SrtB	PPKTG	53.3 ± 25.6 μM	$8.3\times10^{-4} \pm 3\times10^{-4}$ min^{-1}	(Peltier et al., 2017)
S. suis	LPATG	6.7 μM	-	(Chen et al., 2017)

The in vitro enzymatic activity of additional sortases like *Streptococcus suis* has been explored, which showed a maximal activity at pH 6.0–7.5, 45 °C with abz-LPATG-dnp, and the enzyme was found to be Ca^{2+}-independent (Chen et al., 2017), unlike SaSrtA which is found to be calcium-dependent. The sequence specificities of most sortases have been hardly defined due to the low reaction rate and extended incubation time necessary for reaction product characterization. Therefore, the substrate specificity and recognition sequences were mostly predicted using bioinformatics methods (Kruger et al., 2004).

1.3.7. Sortase as bioengineering tool

The discovery of Sortase A from *S. aureus* could break a peptide bond inside the LPXTG motif in vitro and then reform a new bond with an entering oligo-glycine nucleophile (Mazmanian et al., 1999) paved the path for sortases to become useful tools for protein engineering. Sortases catalyzed transpeptidation reaction known as "sortagging" or sortase-mediated ligation (Popp et al., 2007) has been used to site-specifically bind fluorescent dyes, carbohydrates, and other moieties to protein substrates and the surface of cells (Antos et al., 2016).

SML has been used in several bioconjugate chemistry and protein engineering applications. Protein N/C-terminal labeling, protein sugar modification, protein-lipid modification, protein-protein fusion, peptide/protein cyclization, and cell-surface labeling are only a few examples. Surface functionalization, hydrogel modification, biomolecules PEGylation, virus-like particle decorating, protein immobilization/purification, and in vivo protein labeling have all been added to the SML toolbox in recent years. SaSrtA and its engineered versions were employed in the majority of cases.

Immobilized sortase on the surface was used to modify biomolecules on a large scale (Steinhagen et al., 2013). Sortase-mediated ligation was also used to mark proteins at their N and C termini at particular sites (Williamson et al., 2012). Using an orthogonal sortase approach, the M13 bacteriophage was doubly labeled (Hess et al., 2012). Sortase was used to conjugate unnatural chemicals such fluorescein, biotin, PEG, and tetramethylrhodamine to proteins.

The site-specific sortase transpeptidation reaction can be used to ligate protein/peptides with the short pentapeptide recognition sequence LPETG at the C-terminus of the protein to another protein or peptide with the other sortase substrate GGG at the N-terminus, or to the surface modified with this peptide. Sortase reaction can be used to accomplish site-specific protein oligomerization or circularization if these two tags, LPETG and GGG, are positioned at two termini of a protein.

1.3.8. Sortase as drug targets

The most studied sortase is Sortase A (SrtA) from *S. aureus* because this bacterium has increased pathogenicity, virulence, and high level of drug resistance (Shrestha and Wereszczynski, 2016). Many of the MSCRAMMs are LPXTG-bearing proteins linked to the

cell wall of *S. aureus* by Sortase A. These proteins are virulence factors that play key roles in staphylococcal infections, such as bacterial attachment to host tissues, biofilm development, and immune response evasion (Foster et al., 2014). The role of sortase in the virulence of many Gram-positive pathogens, including staphylococci, streptococci, enterococci, and *L. monocytogenes*, is proved by the fact that mutants lack genes for SrtA display attenuated virulence, without affecting the growth of the bacteria (Marraffini et al., 2006). The development and also the search for new inhibitors of SrtA use the *S. aureus* enzyme as a prototype. The deletion of *srtA* in *S. aureus* reduced the ability to infect the host was demonstrated in several models. SrtA activity has been observed to improve *S. aureus* survival inside macrophages after phagocytosis, as *srtA* deletion mutants are more sensitive to macrophage death (Josefsson et al., 2008). Other Gram-positive bacteria, such as *L. monocytogenes* (Bierne et al., 2002), *S. pneumonia* (Kharat and Tomasz, 2003), *Enterococcus faecalis* (Kemp et al., 2007), and *Streptococcus suis* (Fittipaldi et al., 2012), have shown a relationship between sortase activity and pathogenicity. Overall, these investigations have demonstrated the potential of sortase inhibition for the treatment of Gram-positive infectious illnesses, and the search for compounds that might limit sortase activity has become a focus of research. The identification and characterization of a consistent number of reversible and irreversible SaSrtA inhibitors were made possible by screening natural or synthesized chemical libraries, as well as molecular modeling, pharmacophore hypotheses, 3D-QSAR models, and virtual screening techniques (Cascioferro et al., 2014). The first reported inhibitor of SrtA was MTSET, a thiol-reactive compound that reacts irreversibly with the cysteine thiol of the sortase active site (Zong et al., 2004). However, the short half-life and broad reactivity of this compound render it useful only as an experimental in vitro inhibitor. However, the

majority of these compounds showed activity in the high micromolar range, and none of them have been developed as drugs as a result of their low potency.

1.4. Summary

The first part of this chapter provided a brief introduction of the topic selected and described the rationale for the present study, the gap areas to be addressed, and the defined objectives. The second part contains a relevant updated review of the literature about the area to highlight the major studies and finding so far. It covers cellular functions, sortase classification, structure elucidation, biochemical properties, substrate specificity, and recent engineering applications using sortase.

Chapter 2

Materials and Methods

Abbreviations

%	Percent
µg	Microgram
µM	Micromolar
µL	Microliter
Abz	2-aminobenzoyl
APS	Ammonium persulphate
ATP	Adenosine triphosphate
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
BHI	Brain Heart Infusion
bp	Base pair
BSA	Bovine Serum Albumin
Dap	Diamino propionic acid
DNA	Deoxy Ribonucleic acid
Dnp	2,4- dinitrophenyl
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediamine tetra acetic acid
EtBr	Ethidium bromide
FRET	Fluorescence Resonance Energy Transfer
g	Gram
x g	Times gravity
h	Hour
IPTG	Isopropylβ-D-1-thiogalactopyranoside
LB	Luria Bertani
min	Minutes
mL	Milliliter

mg	Milligram
NAD^+	Nicotinamide Adenine Dinucleotide
NBT	Nitro blue tetrazolium
NEB	New England Biolabs
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
PVDF	Polyvinylidene fluoride
s	Seconds
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SOC	Super optimal broth with catabolites repression
TB	Terrific broth
TBST	Tris buffered saline with tween
TEMED	Tetramethyl ethylenediamine
TFA	Trifluoro acetic acid

2.1. General Materials

2.1.1. Chemicals, culture media and reagent kits

Luria-Bertani (LB) broth or LB agar medium, Brain Heart Infusion (BHI) broth, or BHI agar medium were procured from HiMedia, India. Plasmid isolation kits, PCR clean-up kits, and Gel extraction kits were obtained from Qiagen, Germany. Q5® Site-directed mutagenesis kit, was obtained from New England Biolabs, USA. Ni-NTA spin columns for His-tag assisted Ni-affinity chromatography was purchased from GE Healthcare, USA. Antibodies supplied by Abcam (Cambridge, UK).

Restriction enzymes (BamHI, EcoRI, NdeI, NheI, NotI, SalI), 2X *Taq* Master mix, T_4 DNA ligase were purchased from New England Biolabs, USA. Isopropylβ-D-1-thiogalactopyranoside (IPTG), Lysozyme, Bovine Serum Albumin (BSA), Antibiotics like kanamycin, Nicotinamide Adenine Dinucleotide (NAD^+), Dithiothreitol (DTT), Xylose were

procured from HiMedia, India. Molecular biology grade chemicals like Sorbitol, Polyethylene glycol (PEG) MW 4000, Tris-base, Agarose, Glycine, Ethylenediaminetetraacetic acid (EDTA), Glycerol were obtained from Sigma-Aldrich, India. Analytical grade solvents like Dimethyl sulfoxide (DMSO), Trifluoroacetic acid (TFA), Acetonitrile, and Glacial acetic acid were purchased from Merck, Germany.

2.1.2. Fluorescent peptides

Fluorescently self-quenched peptides were synthesized by Shanghai GL Biochem, China, and solubilized in 50 % DMSO before use. The peptides were predicted based on the reports available on the substrate sequence. The peptide substrates were tagged with a 2-aminobenzoyl (Abz) fluorophore at the N-terminus and 2, 4-dinitrophenyl (Dnp) as the quencher at the C-terminus. All peptides used in the study are listed in **Table 2.1**.

Table 2.1. FRET peptides used in the study

Peptide Sequence*	Description	Mass (Da)
Abz-LAHTG-Dap (Dnp)	substrate of SrtE	859.86
Abz-LAETG-Dap (Dnp)	substrate of SrtE	867.88
Abz-LPETG-Dap (Dnp)	The substrate of SrtA (negative control in this study)	885.90
G-Dap (Dnp)	product after the substrate cleavage	326.27

*Where Abz is 2-aminobenzoyl and Dap (Dnp) is diaminopropionic acid-dinitrophenyl

2.1.3. Bacterial strains and plasmids

Corynebacterium glutamicum ATCC 13032 (Abe et al., 1967) obtained from Bielefeld University, Germany as a part of the Indo-German, bilateral exchange program (DST-DAAD), was used for the study. The bacterial strains and plasmids used in this study are listed in **Table 2.2**.

Table 2.2. Bacterial strains and plasmids

Strains/Plasmids	Description/Genotype	Reference/Source
E. coli		
DH5α	*Fthi-1 endA1 hsdr17(r-, m-) supE44 _lacU169 (f80lacZ_M15) recA1 gyrA96 relA1*	(Hanahan, 1983)
BL21 (DE3)	*F- ompT gal dcm lon hsdSB(rB- mB-) λ(DE3)*	Novagen, USA
C. glutamicum		
ATCC 13032	Wild-type (WT)	(Abe et al., 1967)
Plasmids		
pET28a	T7 expression vector (N-terminal His_6-Tag), Kan^r	Novagen, USA
pVWEx1	*E. coli-C. glutamicum* shuttle vector for regulated gene expression (*Ptac, lacIq, pCGl, oriVCg*)	(Peters-Wendisch et al., 2001)
pK19mobsacB	Km^r, vector for integration of insert into the genome of *C. glutamicum*	(Schagfer et al., 1994)
Recombinant plasmids		
pK19Cgl2838	pK19mobsacB harboring the upstream and downstream flanking regions of *NCgl2838* (*srtE*) gene with EcoR1 and BamHI restriction site	This study
pVWCgl2838	*NCgl2838* cloned at SalI and BamHI site of pVWEx1 along with the insertion of ribosome binding site	This study
pSrtE	*srtE* cloned at Nhe1 and NotI of pET28a with an N-terminal His_6-tag	This study
pTSrtE	*srtE* cloned at NdeI and SalI of pET28a without the signal peptide and transmembrane domain of the gene and contains an N-terminal His_6-tag	This study
pGFP	pRSET vector harboring eGFP	(Madhavan et al., 2017)
pXylB	XylB cloned at PstI and BamHI of pVWEx1	(Sundar et al., 2020)
pGFP-LAHTG	*GFP-LAHTG* cloned at Nde1 and BamHI of pET28a with a C-terminal His_6-tag	This study
pXylB-LAHTG	*XylB-LAHTG* cloned at EcoRI and BamHI of pET28a with a C-terminal His_6-tag	This study

2.1.4. Oligonucleotides

The oligonucleotides used for the present work are listed in **Table 2.3**.

Table 2.3. Oligonucleotides for the study

Primer Name	Sequence (5'-3') *	Characteristics
Sortase wild and truncated		
NCgl2838 F	TTTGCTAGCATGACAGCCACGTTAAGCGCGGAATC	NheI
NCgl2838 R	TTTGCGGCCGCGTTTTCCTCCAAAGCTGCAGGGCGTTC	NotI
CgSrtE F	CCCCATATGGCCTATTGGACCAACGTGGAATC	NdeI
CgSrtE R	CCCGTCGACTTAGTTTTCCTCCAAAGCTGCAGGGCGTTC	SalI
Site-directed mutagenesis		
C240A F	CTTGACCACGGCACACCCGCAGTTC	
C240A R	GTAAGCAGAGCTTCTGATC	
Y118A F	TCCTGGCCGTGCTGTGGATTCC	
Y118A R	CCGGCAAGAAGGTCTTCC	
H135A F	AGTGGCAGGCGCGCGAGTGGGCAAG	
H135A R	GCAAAGTTTCCGGCTTCA	
R249A F	CAACGCTGAGGCCATGATTGTGCAC	
R249A R	GAGAACTGCGGGTGACAC	
Vector confirmation		
T7 Promoter F	TAATACGACTCACTATAGGG	
T7 Terminator R	GCTAGTTATTGCTCAGCGG	
pVW F	GAAACAGAATTAAAAGA	
pVW R	CCAGTGAATTCGAGC	
Sortase application		
eGFP-LAHTG F	GCCCATATGGTGAGCAAGGGCGAGGAGCTGTTCAC	NdeI
eGFP-LAHTG R	CCCGGATCCACCAGTATGAGCTAACTTGTACAGCTCGTCCAG	BamHI
XylB-LAHTG F	GATGATGGATCCATGTCCTCAGCCATCTATCC	BamHI
XylB-LAHTG R	TATGAATTCTCAACGCCAGCCGGCGTCGATCCAG	EcoRI
***C. glutamicum* mutant strains**		
Cgl2838 F	TTTGTCGACGAAAGGAGGCCCTTCAGAATGACAGCCACGTTAAG CG	SalI
Cgl2838 R	TTTGGATCCTTAGTTTTCCTCCAAAGCTGCAGGGCGTTCG	BamHI
2838-1A F	GCGAATTCTGCGGCCCGACTTATACATC	EcoRI
2838-2A R	*CCCATCCACTAAACTTAAACA*CCGCGCTTAACGTGGCTGTC	Linker 21 bp
2838-3A F	*TGTTTAAGTTTAGTGGATGGG*CGAACGCCCTGCAGCTTTGG	Linker 21 bp
2838-4A R	GCGGATCCCGGTGATTGTCCGCGATATG	BamHI
Cgl 38 F	GCTTCTGCAACTCCCGAATG	
Cgl 38 R	TCACGGGTTACGGCACAGAC	

*Restriction sites are underlined and linker sequences are shown in italics.

2.2. General Methods

2.2.1. Bioinformatics and software

The in silico and genome analytic tools employed are listed in **Table 2.4**.

Table 2.4. Online tools and other software used in the study

Softwares	Application	Source
Bioedit	Sequencing editing and analysis	www.mbio.ncsu.edu/BioEdit
BLAST	Sequence homology search	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Clone Manager 9	Primer designing	https://scied.com/dl_cm9.htm
ClustalW	Multiple sequence alignment	https://www.genome.jp/tools-bin/clustalw
ClusPro	Protein-protein and protein-peptide docking	https://cluspro.bu.edu/login.php
Expasy ProtParam tool	Physicochemical properties of protein	https://web.expasy.org/protparam/
ESPript 3	Sequence similarity and secondary structure prediction	https://espript.ibcp.fr/ESPript/ESPript/
Image Lab	Gel documentation (Nucleic acid and protein)	www.biorad.com
I-TASSER	3D structure prediction	https://zhanggroup.org/I-TASSER/
LC solution	HPLC chromatograms	www.shimadzu.com
MEGA X	Phylogenetic analysis	https://www.megasoftware.net/
PDB	Structural analysis	https://www.rcsb.org/
PSIPRED	Secondary structure prediction	http://bioinf.cs.ucl.ac.uk/psipred/
PyMOL	Structure visualization	http://www.pymol.org
SWISS-MODEL	3D structure prediction	https://swissmodel.expasy.org/

2.2.2. Culture growth conditions

C. glutamicum strain ATCC 13032 was generally cultivated in BH1 medium **(Annexure-I)** at 30 °C. All *E. coli* strains were routinely grown in liquid or agar Luria Bertani (LB) medium **(Annexure-I)** at 37 °C. *E. coli* DH5α was used for cloning and maintenance purposes and *E. coli* BL21 (DE3) for expression of recombinant proteins. Antibiotics were used at the following concentration: kanamycin, 25 μgml^{-1} for *C. glutamicum* and 50 μgml^{-1} for *E. coli* wherever necessary. The fully grown cultures were stored at 4 °C, and their glycerol stocks were maintained at -80 °C.

2.3. Molecular Methods

The molecular biology techniques were followed as described by Sambrook et al., (2001) with some necessary modifications.

2.3.1. Isolation of DNA

2.3.1.1. Genomic DNA isolation

The genomic DNA of *C. glutamicum* was isolated from overnight static culture by cell lysis, phenol/chloroform extraction, and ethanol precipitation with some modification (Andreou, 2013). The 10 mL of the overnight culture broth was centrifuged at 6000 x g for 5 min. After discarding the supernatant, the cell pellet was frozen at -20 °C for 30-60 min. Resuspended the cell pellet in 250 µL fresh 10 $mgmL^{-1}$ lysozymes in TE buffer (10 mM Tris; 1 mM EDTA, pH 8.0) and transferred to a microcentrifuge tube. The mixture was incubated at 37 °C for 1-2 h with gentle shaking. 50 µL of 0.5 M EDTA, 50 µL 10 % SDS and 50 µL 5 M NaCl was added to the cell suspension and mixed well, followed by 10 µL RNase (100 $mgmL^{-1}$) and incubated at 37 °C for 30 mins. Around 20 µL of Proteinase K (20 $mgmL^{-1}$) was

added and incubated at 37 °C for 60 mins. The complete lysis of cells can be confirmed based on the lucidity of the suspension. The suspension was mixed thoroughly with an equal volume of Phenol, Chloroform, and Isoamyl alcohol (25:24:1). The mixed solution was centrifuged at 12,000 x g for 15 min at RT. After transferring the aqueous phase to a new microcentrifuge tube, 500 µL of Phenol: Chloroform: Isoamyl alcohol (25:24:1) was added and mixed thoroughly before centrifuging at 12,000 x g for 15 min at room temperature. The aqueous phase was transferred to a new microfuge tube, and an equal volume of 99 % isopropanol was added. The tubes were centrifuged at 12,000 x g for 30 min at 4 °C after gentle mixing by inversion. The supernatant was discarded, and the pellet was thoroughly cleaned in 500 µL of ethanol (70 %, v/v). The pellet was recollected by centrifugation at 12,000 x g for 5 min at 4 °C. The washed pellet was air-dried for 10 min at room temperature (RT). The dried pellet was dissolved in 60 µL nuclease-free water and mixed well by gentle pipetting. The DNA was visualized in 1 % prestained (EtBr) agarose gel (0.8 %) and quantified by UV-Vis spectrophotometer (Nanodrop, USA).

2.3.1.2. Plasmid DNA isolation

Plasmid DNA was isolated from *E. coli* and *C. glutamicum* strains using the QIAprep Spin Miniprep kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Purity and quantification were assessed using a NanoDrop1000 spectrophotometer (NanoDrop Technologies, Inc. USA). Plasmid DNA quality was visualized in 1 % prestained (EtBr) agarose gel.

2.3.2. PCR amplification

Unless otherwise specified, the PCR mixture used to create gene inserts consisted of nuclease-free water, 2X PCR master mix (NEB), 0.5 mM $MgCl_2$ (NEB), 0.24 µM of forward

and reverse primers, and 500 ng template DNA. The PCR was carried out using automated PCR machines from BioRad, USA (MyCycler) and Eppendorf, Germany (Ep gradient). As shown in **Table 2.5**, the reaction was incubated. The finished reaction was kept at -20 °C until it was analyzed on an agarose gel, purified with a gel extraction kit, and used for restriction analysis.

Table 2.5. PCR cycling conditions

Steps	Temperature (°C)	Time	
Initial Denaturation	95	1 min	
Denaturation	95	30 s	**35 cycles**
Annealing	60	1 min	
Extension	72	30 s	
Final extension	72	5 min	
	4	hold	

2.3.3. Quantification of DNA

A Nanodrop ND-1000 spectrophotometer (Thermo fisher scientific, India) was used to calculate the amount of DNA in a sample. The Beer-Lambert law is used to calculate the amount of DNA based on its absorbance at 260 nm: A=ecl, where A represents absorbance at 260 nm, e represents the extinction coefficient, c represents concentration, and l represents path length. The extinction coefficient of double-stranded DNA seems to be 0.02 $ngmL^{-1}cm^{-1}$. The degree of purity of the DNA was determined by the quotient $A_{260/280}$ or $A_{260/230}$, which lies in between 1.8 and 2.0. Lower values of DNA were contaminated by proteins or polysaccharides.

2.3.4. Agarose electrophoresis of DNA

PCR fragments, plasmid, and genomic DNA were separated on agarose gel at varied concentrations between 0.8 % -2 % with TAE buffer (40 mM Tris-base, 1 mM EDTA, and 20 mM Glacial acetic acid). Samples were mixed with 6X loading dye (Thermo fisher scientific) and separated at a voltage of 70-100 V for 1 h depending on the varied gel size and were visualized in 1 % prestained (Ethidium bromide) gel under far UV (320 nm) illumination using Chemi, Biorad, USA and kept at -20 °C for further cloning studies.

2.3.5. Sequencing of DNA

Colony PCR was performed with heat-denatured bacterial biomass pricked from a single colony containing crude DNA as a template. Sequencing was performed with Sanger's (dideoxy termination) Genetic Analyzer 3500 using SeqScape® Software v2.7 Applied Biosystems. The instrument uses a fluorescence-based DNA analysis system that uses high-resolution capillary electrophoresis technology. The standard operating procedure and consumables used in this method were provided by the instrument manufacturer. When the gene source was genomic DNA, specific primers were used for gene amplification, whereas T7 F/R primers were used when the gene was inserted in the pET28a vector **(Annexure-II)** and pVW F/R primers were used when the gene was inserted in the pVWEx1 vector **(Annexure-II)**. 10 µL PCR reactions were set up with concerned forward and reverse primer as per the protocol mentioned earlier. PCR amplified genes were purified by gel extraction method and analyzed by UV-Spectrophotometer (Nanodrop, USA). 50 ng of high purity DNA (A_{260}/A_{280} ratio between 1.7 and 1.9) was used for sequencing.

2.3.6. Restriction digestion

Amplified PCR products and the vectors were purified, eluted, and digested with the respective restriction enzymes. The 30 µL reaction contains DNA (1 µg), compatible buffer (1X), restriction enzymes (1 U), and nuclease-free water. The components were gently mixed by flicking the tube, and spun down for a few seconds. The contents were incubated at 37 °C for 1 h.

2.3.7. Ligation

Digested PCR products were ligated into respective vectors for expression. A ligation reaction was performed in a 20 µL set-up. A typical reaction contains water, 1X T4 DNA ligase buffer (40 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP), DNA fragments, and finally T4 DNA ligase (1 U). The contents were added and mixed by keeping the tube in ice, and incubated at 16 °C for 12 h or 4 °C for 18 h.

2.3.8. Preparation of competent cells and transformation

2.3.8.1. Preparation of competent E. coli- $CaCl_2$ and TSS method

Chemical methods of competent cell preparation were used for *E. coli* strains. Competent cell preparation by using TSS (Transformation & Storage Solution) was used for *E. coli* DH5 alpha cells (Chung and Miller, 1993). 5 mL of the starter culture was prepared from a single colony from a fresh LB plate and incubated aerobically at 37 °C for overnight. 1 mL starter culture was inoculated to 100 mL LB broth and incubated aerobically at 37 °C for 2-4 h till it reaches an OD_{600} of 0.4. The cells were harvested by centrifugation in a cooling centrifuge (Kubota, Japan) at 2,700 x g for 10 min. After decanting the supernatant, the cells were gently suspended in 10 mL of ice-cold TSS buffer containing 85 % LB medium, 10 %

PEG (w/v, MW 4000), 5 % DMSO (v/v), and 50 mM $MgCl_2$ (pH 6.5), and incubated 15 min on ice. Following incubation, 200 µL of cells were aliquoted into a new sterile tube and stored immediately at -80 °C.

Competent cells of *E. coli* BL21 (DE3) were prepared by using the calcium chloride method. 5 mL of the starter culture was prepared from a single colony from a fresh LB plate and incubated aerobically at 37 °C for overnight. 1 mL starter culture was inoculated into 100 mL LB broth and incubated aerobically at 37 °C for 2-4 h till it reaches an OD_{600} of 0.4. The culture sample was immediately chilled in ice and harvested the cells by centrifugation in a cooling centrifuge (Kubota, Japan) at 3,600 x g for 10 min. The supernatant was decanted and cells were gently suspended in 100 mL ice-cold $MgCl_2$ (100 mM). The cell suspensions were spun down by centrifuging at 3,600 x g for 10 min. After discarding the supernatant, the pellet was re-suspended in 20 mL ice-cold $CaCl_2$ solution (100 mM) and incubated in ice for 20 min. The cells were separated in a cooling centrifuge operated at 3,600 x g for 10 min. After decanting the supernatant, the pellet was re-suspended in 5 mL of pre-chilled $CaCl_2$ (85 mM)-Glycerol (15 %, w/v) solution. The cells were harvested again by centrifugation at 3,600 x g for 15 min, and the supernatant was discarded. Finally, the pellet was resuspended in 2 mL of pre-chilled $CaCl_2$ (85 mM)-Glycerol (15 %) solution. The cells were aliquoted to 1.5 mL pre-chilled microfuge tubes and stored immediately at -80 °C.

2.3.8.2. Transformation into competent E. coli

Competent cell aliquots (50-100 µL) were gently mixed in ligation reaction mixtures (10-20 µL) or purified plasmid DNA (1 µL) after being thawed on ice for 30 min. The reaction mixture was kept cold for 30 min before being heat shocked in a temperature-controlled water bath (42 °C, 45-90 s). The cultures were then placed on ice, and SOC

medium **(Annexure-I)** 400-500 µL was added. After 1 h in a shaker (37 °C, 200 rpm), cultures were spread onto agar plates containing the appropriate antibiotic and incubated overnight (37 °C). Mini and midi-preps were used to process colonies, which were then characterized using analytical digestion. The transformation efficiency is calculated as per the formula given,

$$\textbf{Transformation efficiency} = \frac{\textbf{No. of Transformants (colonies) x Final volume at recovery}}{\textbf{µg of plasmid DNA x Volume plated (mL)}}$$

2.3.8.3. Preparation of electrocompetent C. glutamicum cells

The complex cell wall structure of *C. glutamicum* resists the action of $CaCl_2$ on its cell wall. Thus, the cells were electroporated to achieve the transformation (Van der Rest et al., 1999).

A fresh single colony of *C. glutamicum* was inoculated into a 10 mL LB medium containing 2 % (w/v) filter-sterilized glucose and cultivated overnight (16 h) at 200 rpm at 30 °C. 100 mL Epo medium containing Trypton 1g, Yeast Extract 0.5 g, and NaCl 1 g per 100 mL of LB medium was dissolved in water and autoclaved in a final volume of 80 mL. Shortly before inoculation of the medium, 400 mg isonicotinic acid hydrazide (isoniazid), 2.5 g glycine, and 0.1 mL Tween 80 were solubilized in 20 mL water and added to the medium filter-sterilized. The overnight culture was inoculated into Epo medium to an optical density of 0.3 at OD_{600}. The cells were then allowed to grow for 28 h in a shaker at 120 rpm at 18 °C until the OD_{600} reached approximately 1. The cells were chilled on ice for 10 min before being harvested using a 4,000 x g centrifuge for 10 min. Before being resuspended in 0.5 mL

(v/v) glycerol, the cells were washed four times with 50 mL of ice-cold 10 % (v/v) glycerol. 100 µL of cells were aliquoted and stored at -80 °C in pre-chilled microfuge tubes.

2.3.8.4. Transformation into electrocompetent C. glutamicum cells

The electro-competent cells were thawed in ice for 30 min. 1 mL regeneration medium containing Brain Heart Infusion and 0.5 M Sorbitol was taken in a microfuge tube and kept at 46 °C for pre-warming. 100-300 ng plasmid DNA and for suicidal vectors like pK19mobsacB **(Annexure-II)**, a higher amount of DNA (1-20 µg) (Van der Rest et al., 1999) was taken in a fresh pre-chilled microfuge tube and mixed with 100 µL of the electro-competent cells and kept in ice for 15 min. 100 µL of the content was transferred to pre-chilled fresh and sterile electroporation cuvette (2 mm) and placed in ice for 5 more min. The cells were electroporated at 2500 V, 25 µF, 600 Ω. The contents were transferred to warm regeneration media and incubated at 46 °C for 6 min in a dry incubator. The cells were regenerated at 30 °C for 1 h by vigorous mixing (~200 rpm) in a shaker. 100 µL of the cell suspension was plated in BHI medium supplemented with suitable antibiotics. The transformants appeared after 24-36 h of incubation at 30 °C. The transformation efficiency is calculated as per the formula given,

$$\textbf{Transformation efficiency} = \frac{\textbf{No. of Transformants (colonies) x Final volume at recovery}}{\textbf{µg of plasmid DNA x Volume plated (mL)}}$$

2.4. General Protein Methods

2.4.1. Recombinant protein production and purification

For enzymatic assays, *E. coli* BL21 (DE3) cells harboring recombinant plasmids were cultivated in complex media (TB and LB) supplemented with kanamycin (50 μgmL^{-1}) at an OD_{600} between 0.6-0.8. The cells were harvested by centrifugation at 3500 x g at 4 °C for 30

min. The cells were resuspended in appropriate lysis buffer and lysed by sonication for 10 min (Sonics Vibra cell, U.S.A.) at 35 % amplitude and pulse: 10 s ON and 10 s OFF in an ice bath. Cell debris was removed further by centrifugation at 20,000 x g for 30 min at 4 °C. The lysate was then loaded onto a HisTrap column (GE Healthcare, USA), rinsed with lysis and wash buffers, and the protein was eluted using an imidazole concentration gradient. The excess imidazole was removed with the help of a PD-10 desalting column.

2.4.2. Protein quantification

The purified recombinant protein was pooled and concentrated using 10 kDa Amicon Ultra centrifugal filters (Millipore). Concentrated protein samples were quantified spectrophotometrically using Infinite 200 PRO microplate reader at 595 nm along with Bradford reagent. Bovine serum albumin (BSA) was used as the standard. 10 µL of the sample was mixed with 300 µL of Bradford reagent and incubated for 5 min at RT. A standard curve was plotted against BSA concentration within a range of 0.1-1 $mgmL^{-1}$ **(Figure 2.1)**. All the readings were taken in triplicates in a 96-well microtiter plate.

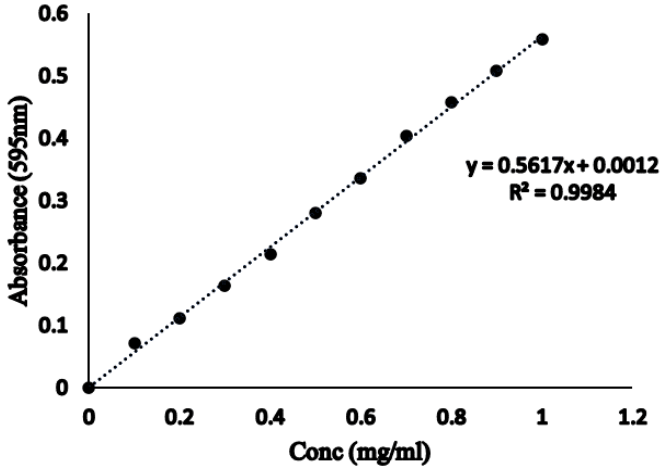


Figure 2.1. Bradford assay standard curve

2.4.3. SDS Polyacrylamide gel electrophoresis

The polyacrylamide gel is cast between 10 cm X 10 cm glass plates provided with Bio-Rad protean mini electrophoresis unit (Bio-Rad, USA). A separating gel solution (8 mL) with 30 % acrylamide-bis acrylamide (30:1) in 1.5 M Tris-HCl (pH 8.8) with 10 % SDS is poured with 10 % freshly prepared ammonium persulphate (APS) and 8 µL of Tetramethyl ethylenediamine (TEMED). A small layer of isopropanol layer is poured over this layer to prevent oxidation of the end groups. This is followed by decanting the isopropanol layer and pouring the stacking gel (5 mL) with 30 % acrylamide-bis acrylamide (30:1) and 0.5 mM Tris-HCl with pH 6.8, 10 % SDS and 10 % APS with 5 µL TEMED. A comb for forming wells is placed before the gel sets. The buffer tank is filled with Tris-glycine buffer (5X) for 100 mL contains 1.5 g Tris-base, 9.4 g Glycine, and 10 % SDS. 25 µL of the protein samples were mixed with 4 µL of 4X sample buffer (50 mM Tris-base (pH 6.8), 2 % SDS, 10 % glycerol, 0.1 % bromophenol blue, 0.9 µL β-mercaptoethanol freshly added), heated to 95 °C for 7 min. The protein expression profiles were analyzed within a range of 12-15 % SDS-PAGE with a voltage of 70-100 V depending on the size of the protein. The SDS gels were either stained with Coomassie Brilliant Blue R-250 for visualization or transferred to nitrocellulose membrane for western blotting (Laemmli, 1970).

2.4.4. Western blotting of His-tagged protein

20 µg and 40 µg of purified protein were mixed with 4X sample buffer and loaded on to 15 % SDS-PAGE with a voltage of 70 V. 10 µL of 10-245 kDa colored protein marker (NEB) loaded on the first lane. The gel was equilibrated in a small container of transfer buffer (10X contains 250 mM Tris-base and 1.92 M glycine) for 15 min. Approximately 4 pieces of blot papers were rinsed with transfer buffer and activated the PVDF membrane by immersing

for 1 min in 100 % methanol. The proteins were transferred into nitrocellulose membrane using transfer buffer at 18 V for 30 min. The membrane was blocked with blocking buffer (5 % non-fat milk (skim) in TBST (10X) containing 200 mM Tris-base, 1.5 M NaCl, 1 % Tween 20) at RT for 30-60 min. The blocked membrane was then labeled with 10 mL of primary antibody (anti-his tag; 1:1000 dilution) at 4 °C O/N and washed 5 times with TBST for 5 min. The membrane was blocked with 10 mL of secondary antibody (goat anti-mouse IgG H/L (alkaline phosphatase); 1:5000) at RT for 1 h. 2-3 mL of alkaline phosphatase chromogen (BCIP/NBT) was added over the membrane and incubated in dark for 5 min. The images were taken using the Chemi, Biorad imaging system.

2.4.5. Site-directed mutagenesis

Q5-site directed mutagenesis kit protocol (NEB, USA) was used for creating specific substitutions nucleotides from the recombinant construct. The PCR reactions (25 µL) were done following the protocol of New England Biolabs (NEB) in terms of contents amounts and concentrations. Contents of the kit contained Q5 hot start high-fidelity 2X master mix, KLD reaction buffer for Kinase, Ligase and DpnI treatment, and KLD enzyme mix. The oligonucleotide primers were designed as per the free online primer designed tool NEBaseChanger. The entire steps for site-directed mutagenesis were shown in **Figure 2.2**.

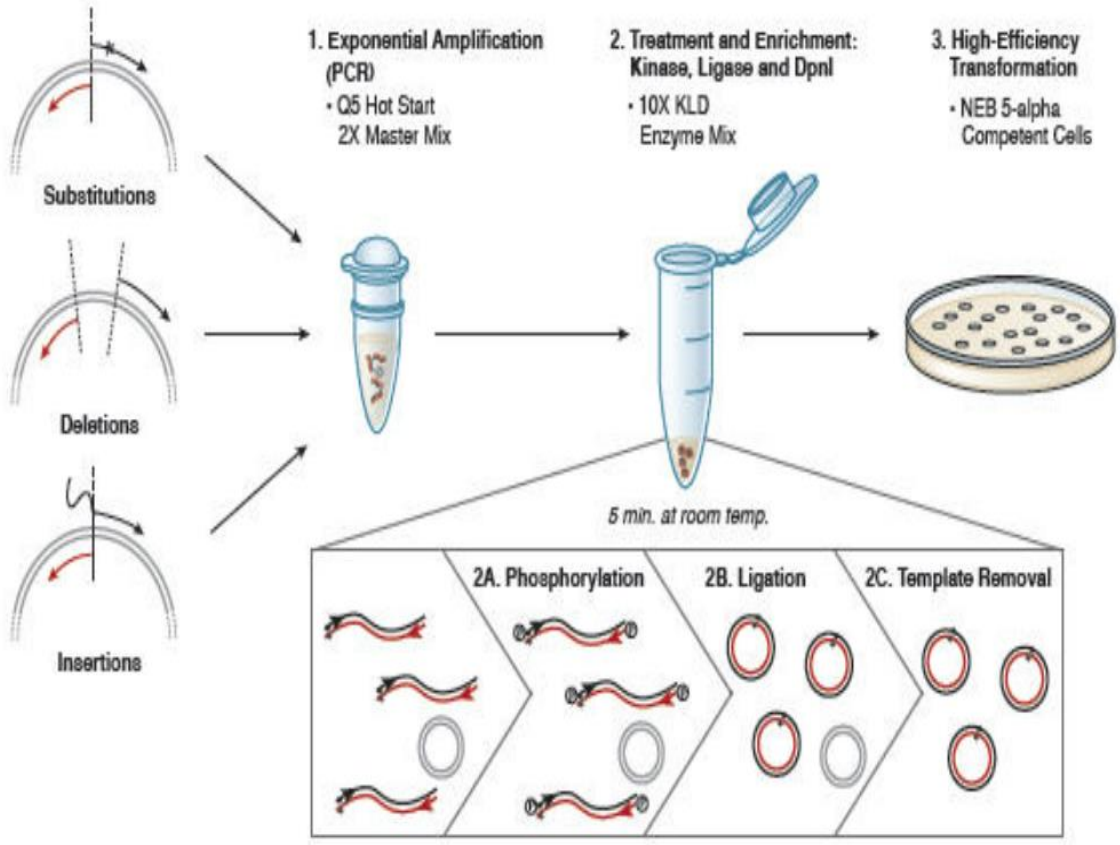


Figure 2.2. Schematic representation of steps involved in introducing site-directed mutations into intact plasmids by PCR based NEB protocol

The composition of the PCR reaction mixture was shown in **Table 2.6** and the programs used for PCR was shown in **Table 2.7**.

Table 2.6. PCR reaction mixture for mutagenesis

Reagents	Quantity (µL)
Q5 hot start high-fidelity 2X master mix	12.5
10 µM Forward primer	1.25
10 µM Reverse primer	1.25
Template DNA (1-25 ngµL^{-1})	1
Nuclease free water	9

Table 2.7. PCR program for mutagenesis

Steps	Temperature (°C)	Time	
Initial Denaturation	98	30 s	
Denaturation	98	10 s	**25 cycles**
Annealing	60	30 s	
Extension	72	30 s/kb plasmid length	
Final extension	72	2 min	
	4	hold	

After the PCR reactions, a total of 10 µL reaction was set up for treatment and enrichment of kinase, ligase, and DpnI, for which 5 µL KLD reaction buffer (10X) and 1 µL KLD enzyme mix was added in a tube containing 1 µL of PCR product. The reaction mix was incubated for 5 min at room temperature and transformed to *E. coli* DH5alpha competent cells **(Section 2.3.8.1)** and confirmed by sequencing. The plasmids bearing the intended mutants were transformed to *E. coli* BL21 (DE3) competent cells **(Section 2.3.8.1)** for expression and purification according to the standard protocol **(Section 2.4.1)**.

2.5. Summary

This chapter deals with general materials and methods such as biological, molecular, and analytical protocols. It followed all standard protocols. However, certain specific protocols will be described in the appropriate working chapters.

Chapter 3

In silico structural modeling and characterization of sortase and sortase-dependent protein (SDP) in *C. glutamicum*

Abbreviation

ATCC	American Type Culture Collection
BLAST	Basic Local Alignment Search Tool
C-score	Confidence score
Da	Dalton
FFT	Fast Fourier Transform
GMQE	Global Model Quality Estimation
GRAVY	Grand Average Hydropathicity
I-TASSER	Iterative Threading Assembly Refinement
kDa	Kilodalton
ORF	Open Reading Frame
PDB	Protein Data Bank
pI	Isoelectric point
QMEAN	Qualitative Model Energy Analysis
RMSD	Root Mean Square Deviation
SOPMA	Self-Optimized Prediction Method with Alignment
TM-score	Template Model score

3.1. Introduction

Sortases are membrane-bound enzymes that catalyze the covalent binding of surface proteins to the cell wall. Each class of sortases described so far performs a specific functional role in attaching a range of surface proteins in Gram-positive bacteria. Sortases not only have virulence and pathogenesis properties in host cells, but also play a significant role in human gut retention and immunomodulation in non-pathogenic bacteria, like probiotics.

Sortase-dependent proteins (SDPs) recognized by sortase share some common characteristics which include a pentapeptide motif at the C-terminus, followed by a transmembrane domain and a positively charged tail (Schneewind et al., 1992). However, based on the pentapeptide recognition motif, predicted a surface protein from the genome of *C. glutamicum* which is considered to be hypothetical. The overall genome annotation has not yet been achieved due to some constraints, including cost and time for creative methodologies (Munir et al., 2016). In silico methodologies for explaining hypothetical proteins are less expensive and take less time to investigate their functionality (Naveed et al., 2017). To establish protein functions, computational approaches incorporating several databases and diverse algorithms are effective alternative tools to laboratory methods (Singh and Chaube, 2014).

Protein structure clarity is critical for understanding biological activities at the molecular level. In *C. glutamicum* ATCC 13032, however, little is known about sortase and its substrate proteins. It is extremely tough and complex to determine the structure of the protein. The protein structure was determined using X-ray crystallography or NMR spectroscopy, although these approaches take time and aren't always successful, especially with membrane proteins. Thus, an alternative promising approach was generated to build the 3D structure of proteins models of unknown function by in silico analysis, and proteins showing more than 35 % similarity served as an objective with greater validation (Fiser, 2010).

The genome sequencing of *C. glutamicum* suggests that the gene *NCgl2838* later designated as *srtE,* encodes a single sortase like transpeptidase. *C. glutamicum* SrtE (CgSrtE) consists of 274 amino acids and is estimated to be 29.86 kDa in size was classified as a class E sortase based on sequence similarities to *C. diphtheriae* Sortase F (CdSrtF). LAXTG was

proposed as the expected recognition sequence for CgSrtE. Based on the recognition sequence LAXTG can predict a hypothetical membrane protein Cgl0614 as the substrate of CgSrtE. The present study was to use the in-silico analysis to model Sortase E and substrate protein Cgl0614 from *C. glutamicum* which includes physicochemical properties of the designed secondary structure, homology modeling, evaluation, and analysis of the modeled structures (CgSrtE and Cgl0614) and their docking analysis using various standard computational tools to understand and detect the possible residues which modulate the substrate specificity.

3.2. Materials and Methods

3.2.1. Sequence retrieval

The *srtE* (*NCgl2838*) gene from *C. glutamicum* ATCC 13032 was annotated in the database records as GenBank: BA000036 and UniProt: Q8NLK3 was retrieved from National Center for Biotechnology Information (NCBI). The protein contains 274 amino acids with an appropriate open reading frame (ORF) and is employed in the current study for further analysis.

3.2.2. Multiple sequence alignment

A multiple sequence alignment of the Sortase E of *C. glutamicum* (CgSrtE) with those of other reported class E sortases of *Streptomyces* sp. was done using multiple sequence alignment tool (Clustal W).

3.2.3. Phylogenetic analysis

A phylogenetic tree was created to examine the evolutionary relationships of *C. glutamicum* sortase with other *Corynebacterium* species. Mega X software (Kumar et al., 2018) was used to perform the phylogenetic analysis, and the Tree View Program was used to

create a phylogenetic tree using the neighbor-joining method and JTT matrix-based model (Jones et al., 1992). A bootstrap study was used to verify the branching pattern repeatability.

3.2.4. Primary structure prediction of CgSrtE

The physicochemical properties of sortase were calculated using the ExPasy ProtParam tool (Colovos and Yeates, 2020). The theoretical isoelectric point (pI) of the protein, molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index, and grand average hydropathicity (GRAVY) were computed using the default parameters.

3.2.5. Secondary structure prediction of CgSrtE

The secondary structure (alpha helices and beta sheets) of CgSrtE protein was predicted using the self-optimized prediction method with alignment (SOPMA) (Geourjon and Deléage, 1995). Alternatively, PSIPRED (PSI-blast-based secondary structure prediction) server (Cuff and Barton, 2000) was also used to validate the results obtained by the SOPMA server.

3.2.6. Template sequence and alignment

The appropriate template for the CgSrtE protein was found using ClustalW (Clustal, 1994) and SWISS-MODEL. The ClustalW (Clustal, 1994) was used to confirm the similarity identity between the amino acid sequences of the homology model of CgSrtE obtained from SWISS-MODEL with the template structure, which was then further processed with the ESPript programs (Robert and Gouet, 2014).

3.2.7. Homology modeling of CgSrtE

The experimental crystal structure of CgSrtE is not available in the Protein Data Bank (PDB) (Burley et al., 2019); thus, its 3D structure was modeled. The protein ID of the target (*C. glutamicum* ATCC 13032 sortase) was obtained from the UniProt Knowledgebase (UniProtKB) with the accession number Q8NLK3 (Bateman, 2019). Following that, the protein ID was submitted to the SWISS-MODEL (Waterhouse et al., 2018) webserver to develop a model with adequate query sequence coverage and sequence identity. Based on the Global Model Quality Estimation (GMQE) (Cardoso et al., 2018) and Qualitative Model Energy Analysis (QMEAN) (Benkert et al., 2009) values, the most reliable 3D structure was chosen. The GMQE values are typically between 0 and 1, with the higher the number, the more reliable the predicted structure, but a value less than 4.0 suggests reliability for QMEAN. The best model structure was then compared with the template protein structure by superimposing both the structures via PyMOL for visualization.

3.2.8. Structure validation of the generated model

The generated SrtE protein structure of *C. glutamicum* by SWISS-MODEL was validated by Ramachandran plot using PROCHECK (Laskowski et al., 1996) and checking the Z-score value acquired by ProSA (Wiederstein and Sippl, 2007).

3.2.9. Identification and analysis of *C. glutamicum* sortase-dependent protein

The genome of *C. glutamicum* ATCC 13032 was searched for proteins containing a LAXTG motif using the Basic Local Alignment Search Tool (BLAST) of the National Centre for Biotechnology Information (Altschul et al., 1990). Thus, we extensively searched for proteins having the signal peptide at the N-terminus and a LAXTG motif, a membrane-

spanning region, and positively charged residues within 50 amino acids of the C-terminus of the protein.

3.2.10. Primary structure prediction of Cgl0614

The physicochemical properties of Cgl0614 were calculated using the ExPasy ProtParam tool (Colovos and Yeates, 2020). The theoretical isoelectric point (pI) of the protein, molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index, and grand average hydropathicity (GRAVY) were computed using the default parameters.

3.2.11. Secondary structure prediction of Cgl0614

The secondary structure (alpha helices and beta sheets) of Cgl0614 protein was predicted using the self-optimized prediction method with alignment (SOPMA) (Geourjon and Deléage, 1995). Alternatively, the PSIPRED server (Cuff and Barton, 2000) was also used to validate the results obtained by the SOPMA server.

3.2.12. Modeling of 3D structure of Cgl0614

The three-dimensional structure of SDP (Cgl0614) was predicted by I-TASSER (Iterative Threading Assembly Refinement) (Roy et al., 2010). The I-TASSER server is a web-based platform for predicting protein structure and function. It enables users to produce high-quality 3D structure and biological function predictions for protein molecules, based on fold recognition and comparing sequence similarity of known proteins (templates) in the Protein Data Bank (PDB). The quality of the anticipated protein model was predicted using the C-score, TM-score, and RMSD. To measure the confidence of each model, the C-score (Confidence score) is calculated based on the importance of threading template alignments

and the convergence factors of structure assembly models. The C-score is normally in the range of [-5, 2], with a higher C-score indicating a more confident model and vice versa. Following the observed correlation between these attributes, TM-score (Template model score) and RMSD (Root Mean Square Deviation of atomic position) are calculated using C-score and protein length (Yang et al., 2014).

3.2.13. Validation of the generated model

The stereochemical quality of the generated Cgl0614 protein structure of *C. glutamicum* was validated using various bioinformatics tools such as PROCHECK and ProSA. The best model structure was then compared with the template protein structure by superimposing both the structures via PyMOL for visualization.

3.2.14. Protein-protein docking

Protein-protein docking was carried out using ClusPro 2.0 server (Kozakov et al., 2017) between the generated structures, CgSrtE (enzyme) and Cgl0614 (substrate protein). The docking results were visualized by using PyMOL to confirm and understand the binding energy of substrate protein Cgl0614 with the CgSrtE enzyme. The ClusPro server performs three steps in the computation process: (i) rigid-body docking utilizing the fast Fourier transform (FFT) correlation technique, (ii) RMSD-based grouping of the structures developed to locate the largest cluster that would represent the most likely models of the complex, and (iii) refining of selected structures.

3.3. Results

3.3.1. Genome search for *srtE* gene in *C. glutamicum* ATCC 13032

The full genome sequence of *C. glutamicum* ATCC 13032 was available in the NCBI database, and it revealed only one copy of the *srtE* (*NCgl2838*) gene, which codes for sortase-like transpeptidase. The translated amino acid sequences were also retrieved from the same website and it was found to be 274 amino acid polypeptides, expected to have an approximate molecular mass of 29.86 kDa **(Table 3.1)**.

Table 3.1. Nucleotide and amino acid sequences of *srtE* gene

***C. glutamicum* ATCC 13032|NCgl2838|*srtE*: 825 bp - Putative sortase-like transpeptidase (Sortase E)**

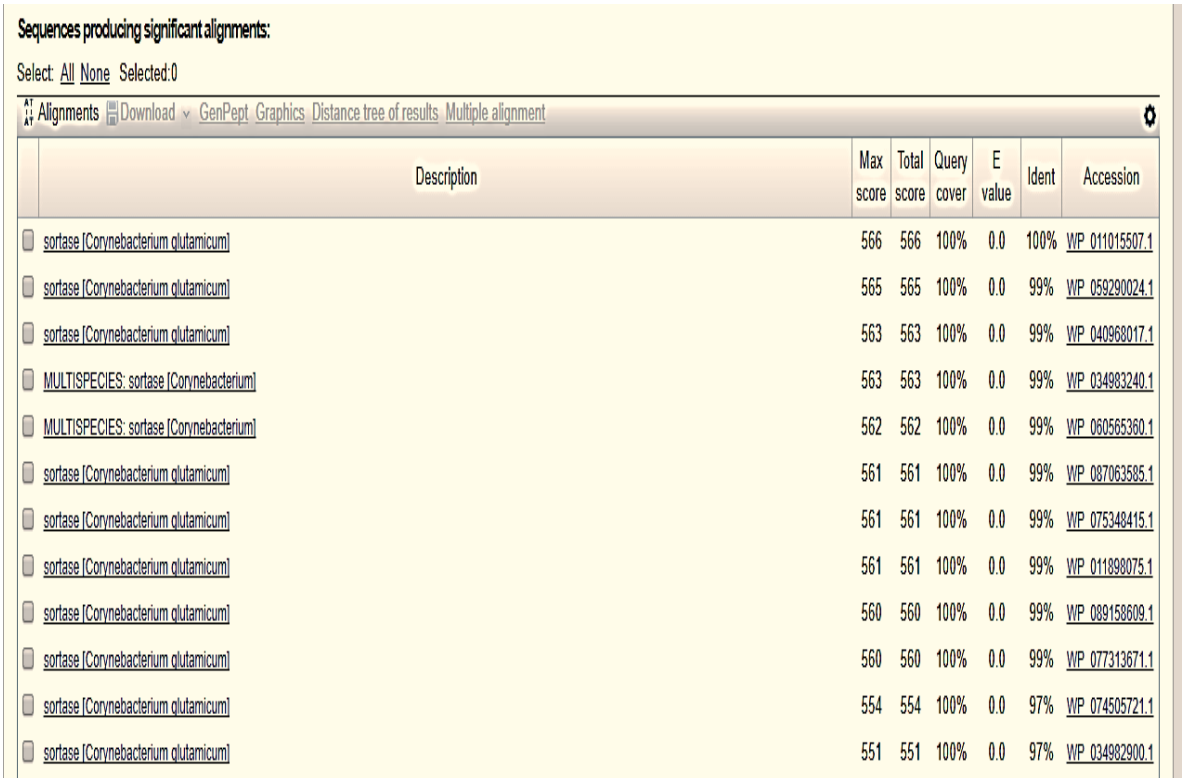
```
atgacagccacgttaagcgcggaatcttctcgcaatggtaaaaagccgcggcctcgagtgagtgtttcccag
gttgttggtgaaatcttgctcaccgtaggcattttggccttgttattcgcatactatgaggcctattggacc
aacgtggaatctgggaaattacaagaatcggctggtcaaaagcttgatgaagactggaatgaagctcgggtg
aatcctcgacaaaagctcaccccggaacttggtgaggcatttgcccggatgtatgttccagctttcggctct
gacttcaacttcgcagtgattgaaggaaccgatgaggaagaccttcttgccggtcctggccgttatgtggat
tcccaaatgcctggtgaagccggaaactttgcagtggcaggccaccgagtgggcaagggtgcgccattcaat
gatctaggaaacctggaagtctgcgatgcgatcgtggtggagacttacaattcctgggatgtgtaccgcgtg
atgccgatgtccaccaacggtgcagatcgtgcagcagaagctgcggattgcttcaacgaaaaccaggtcagc
cgcatggctgaaggtgactatgtgaatgtgtccggacgaagcatcaccactccggatcgcatcgatgccacc
taccccacaccgggcgtcttcgacactgcagtgcgtgaaggatcagaagctctgcttaccttgaccacgtgt
cacccgcagttctccaacgctgagcgcatgattgtgcacgcaatgttggtggaagaaatcgataaatcaagt
ggcgaacgccctgcagctttggaggaaaactaa
```

***C. glutamicum* ATCC 13032|SrtE: 274 aa - Putative sortase-like transpeptidase (Sortase E)**

```
MTATLSAESSRNGKKPRPRVSVSQVVGEILLTVGILALLFAYYEAYWTNVESGKLQESAGQKLDEDWNEARV
NPRQKLTPELGEAFARMYVPAFGSDFNFAVIEGTDEEDLLAGPGRYVDSQMPGEAGNFAVAGHRVGKGAPFN
DLGNLEVCDAIVVETYNSWDVYRVMPMSTNGADRAAEAADCFNENQVSRMAEGDYVNVSGRSITTPDRIDAT
YPTPGVFDTAVREGSEALLTLTTCHPQFSNAERMIVHAMLVEEIDKSSGERPAALEEN
```

3.3.2. Amino acid sequence analysis

The BlastP analysis of the SrtE protein sequence showed a 100 % similarity to class E sortase of *C. glutamicum* **(Figure 3.1)**.



Sequences producing significant alignments:

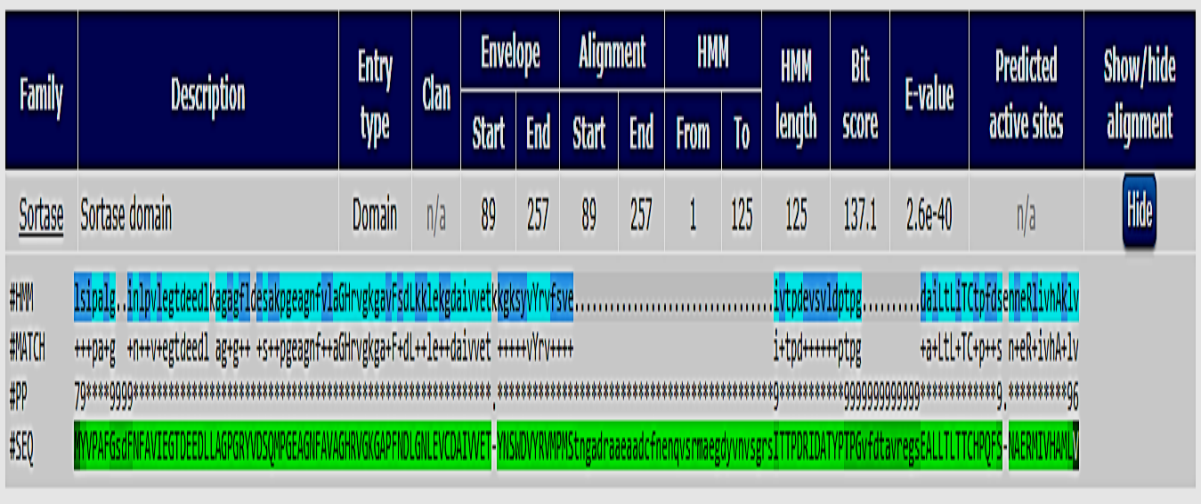
Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession
sortase [Corynebacterium glutamicum]	566	566	100%	0.0	100%	WP_011015507.1
sortase [Corynebacterium glutamicum]	565	565	100%	0.0	99%	WP_059290024.1
sortase [Corynebacterium glutamicum]	563	563	100%	0.0	99%	WP_040968017.1
MULTISPECIES: sortase [Corynebacterium]	563	563	100%	0.0	99%	WP_034983240.1
MULTISPECIES: sortase [Corynebacterium]	562	562	100%	0.0	99%	WP_060565360.1
sortase [Corynebacterium glutamicum]	561	561	100%	0.0	99%	WP_087063585.1
sortase [Corynebacterium glutamicum]	561	561	100%	0.0	99%	WP_075348415.1
sortase [Corynebacterium glutamicum]	561	561	100%	0.0	99%	WP_011898075.1
sortase [Corynebacterium glutamicum]	560	560	100%	0.0	99%	WP_089158609.1
sortase [Corynebacterium glutamicum]	560	560	100%	0.0	99%	WP_077313671.1
sortase [Corynebacterium glutamicum]	554	554	100%	0.0	97%	WP_074505721.1
sortase [Corynebacterium glutamicum]	551	551	100%	0.0	97%	WP_034982900.1

Figure 3.1. BLAST analysis of amino acid sequence

Using the Pfam database the conserved residues in the SrtE protein sequences belong to the sortase domain **(Figure 3.2).**



Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
Sortase	Sortase domain	Domain	n/a	89	257	89	257	1	125	125	137.1	2.6e-40	n/a	Hide

#HMM
#MATCH
#PP
#SEQ

Figure 3.2. The conserved domain of CgSrtE by Pfam database

3.3.3. Multiple sequence alignment of class E sortase

Multiple sequence alignment reveals several motifs which were conserved in the class E sortases. His, Cys, and Arg residues of other reported sortases were aligned well with active sites residues of *C. glutamicum* (*Cgl2838*) respectively, indicating that these three highly conserved residues might constitute the catalytic triads of the enzyme. The Tyr residue is located in a similar place in numerous Class E sortase sequences, suggesting that this is a common feature that explains Sortase E specificity for putative LAXTG-containing substrates in their genomes **(Figure 3.3)**.

Figure 3.3. Multiple sequence alignment of Class E sortase

Tyrosine (Y) residue conserved only in Sortase E enzyme, which is depicted in a pink triangle, TLxTC (marked by a red rectangle) motif which is characteristic of sortase enzymes containing the active site cysteine (C) residue (marked by green triangle), a conserved arginine (R) downstream of the TLxTC motif (marked by green triangle) and the catalytic Histidine (H) residue is marked by a green triangle.

3.3.4. Phylogenetic analysis

The phylogenetic study of *C. glutamicum* SrtE demonstrates that the protein is conserved in the genomes of *Corynebacterium* sp. and acts as a housekeeping gene within the organism **(Figure 3.4)**. The study found 29 *Corynebacterium* sp., of which class E sortase of *Corynebacterium diphtheriae* was reported as a housekeeping gene in the organism.

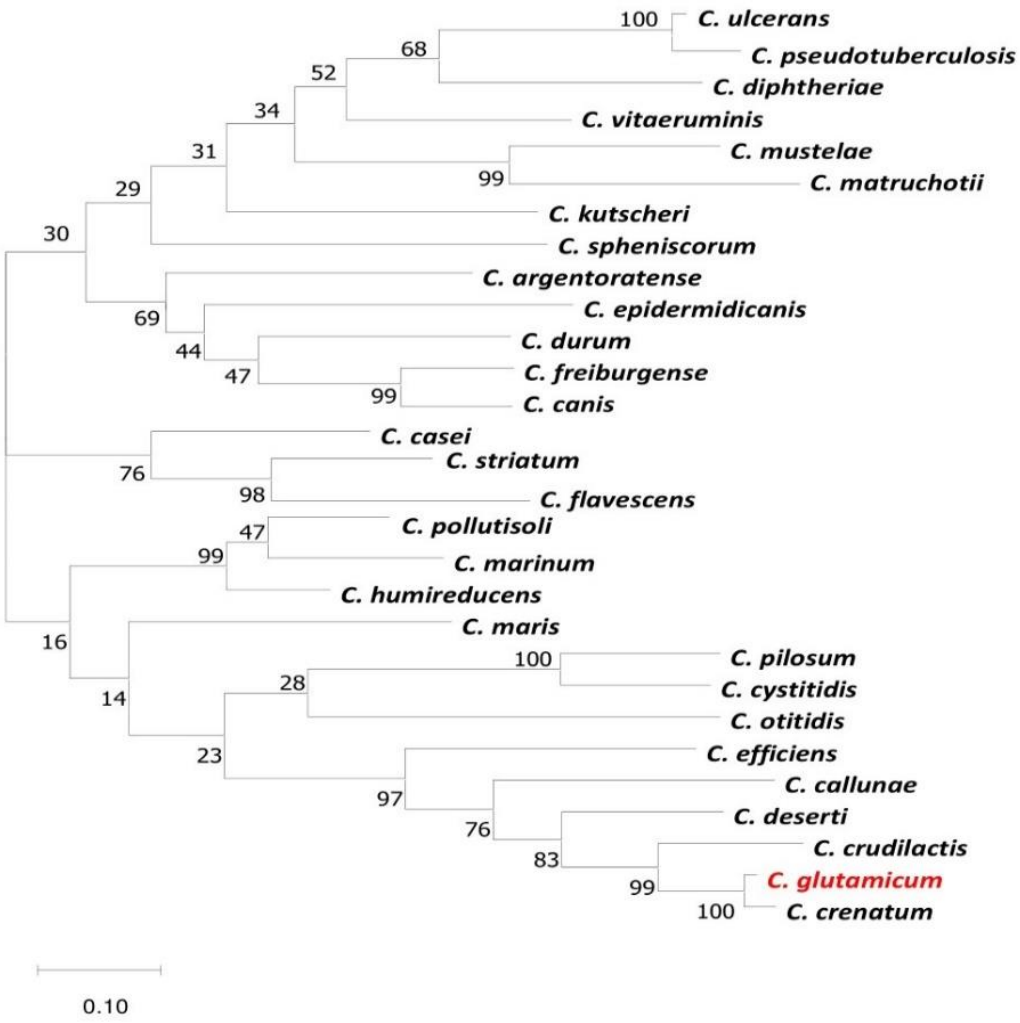


Figure 3.4. Phylogeny of Sortase E homologs of *Corynebacterium* sp.

The phylogenetic tree of the housekeeping Sortase E was built by MEGA X using maximum likelihood strategy and JTT matrix-based model. The proportion of trees with the relevant taxa grouped is presented beside the branches. The scale bar represents 0.1 substitutions per site. The analysis showed 29 genomes of *Corynebacterium* sp.

3.3.5. Primary structure prediction of CgSrtE

The physicochemical properties of the CgSrtE protein sequence were analyzed by using the ExPasy ProtParam tool. The calculated molecular weight of the protein (CgSrtE) containing 274 amino acids was 29857.21 Da, and the theoretical pI was 4.52, indicating that the protein will precipitate in acidic conditions. The number of negatively charged protein residues (Asp and Glu) is 23, whereas the number of positively charged protein residues (Arg and Lys) is 42. The instability index of the protein showed 29.55, indicating that the protein is stable, since the obtained result is less than the cut-off of 40. The aliphatic index and GRAVY values were 74.78 and -0.33 respectively. Overall, the GRAVY score (-0.33) suggested that the protein is hydrophilic and soluble **(Table 3.2)**.

Table 3.2. Physicochemical properties of *C. glutamicum* SrtE

S. No	Physicochemical properties of CgSrtE	Values
1	Amino acid residues	274
2	Molecular weight (Da)	29857.21
3	Theoretical pI	4.52
4	Positively charged residue	42
5	Negatively charged residue	23
6	Total number of atoms	4121
7	Molecular formula	$C_{1302}H_{2024}N_{360}O_{424}S_{11}$
8	Instability index	29.55
9	Aliphatic index	74.78
10	GRAVY	-0.33

3.3.6. Secondary structure prediction of CgSrtE

The secondary structure prediction of CgSrtE was performed by using the tool SOPMA. The random coil was determined to be the most abundant, accounting for 53.65 %,

alpha-helix for 27.37 %, extended strand for 16.42 %, and beta turns for 2.55 %. These results were confirmed by PSIPRED were presented in **Figure 3.5**.

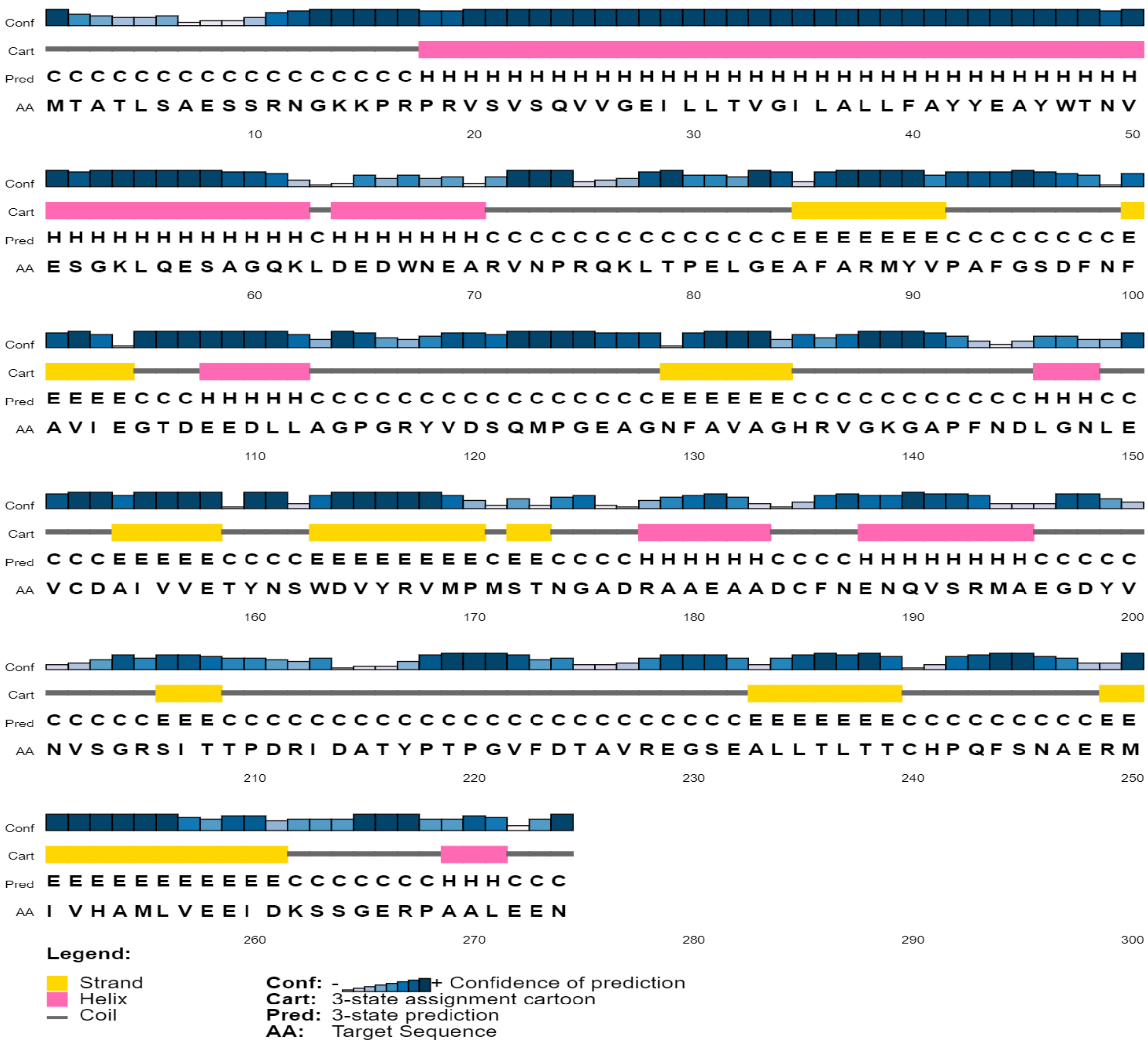


Figure 3.5. Validation of predicted secondary structure of *C. glutamicum* SrtE protein by PSIPRED server

3.3.7. Secondary structure sequence alignment between CgSrtE and CdSrtF

There is no crystal structure for SrtE of *C. glutamicum*, yet its sequence is highly conserved within the genus of *Corynebacterium* sp. Our studies showed that the crystal structure of 5UUS (CdSrtF) from *C. diphtheriae*, has the highest level of sequence homology

in the Protein Data Bank (PDB). Based on this fact, we opted to develop a three-dimensional (3D) structural model of CgSrtE via homology with CdSrtF (PDB ID: 5UUS). The first stage was comparing the expected secondary structure of CgSrtE to the template (CdSrtF) by using ESPript 3.0. The sequence and secondary structure analysis between CgSrtE and CdSrtF were shown in **Figure 3.6.**

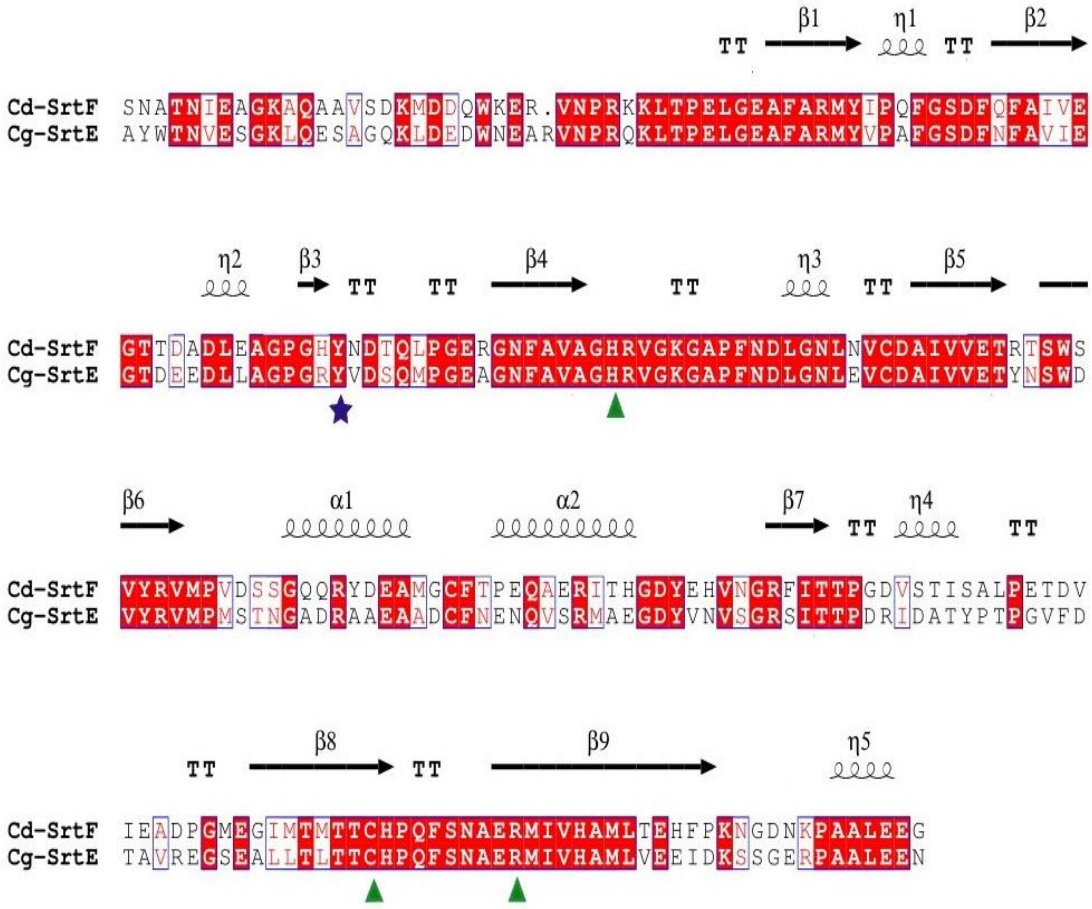


Figure 3.6. Secondary structure comparison between CgSrtE and CdSrtF

ClustalW was used to construct multiple sequence alignments of *C. glutamicum* Sortase E (CgSrtE) with *C. diphtheriae* sortase F (CdSrtF), and ESPript 3 was used to create the figure. Above the alignment are the secondary structural components which contain helices (α), strands (β), 3_{10} helices (η), and turns (T). The green triangle represents the catalytic triads His, Cys, and Arg, whereas the blue star represents the Tyr residue, which is conserved solely in the Sortase E enzyme.

3.3.8. CgSrtE protein homology modeling and structure validation

In silico modeling of SrtE of *C. glutamicum* was generated with the available crystal structure in the PDB database and the best hit was found to be 5UUS (crystal structure of SrtF from *C. diphtheriae*, CdSrtF) with 62.22 % sequence identity **(Figure 3.7)**.

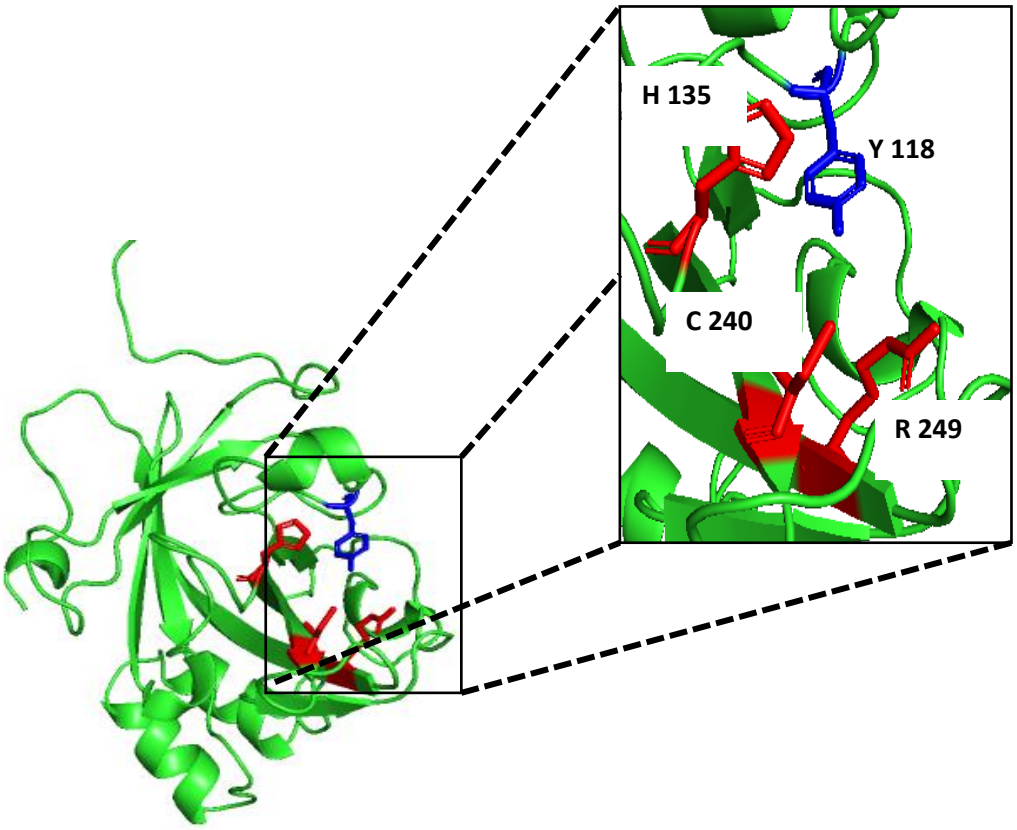


Figure 3.7. Homology modeling of CgSrtE

The modeled CgSrtE is shown as a cartoon representation in green color along with conserved residues at C 240, R 249, H 135, and Y 118.

The modeled protein was indeed obtained to be monomeric by SWISS-MODEL. Inside the SWISS-MODEL, the GMQE score reflects the accuracy of the alignment, and values range from 0-1, with higher values representing better models. The GMQE score for the CgSrtE protein structure is 0.83 and the QMEAN value was found to be -2.33 were shown in **Figure 3.8**.

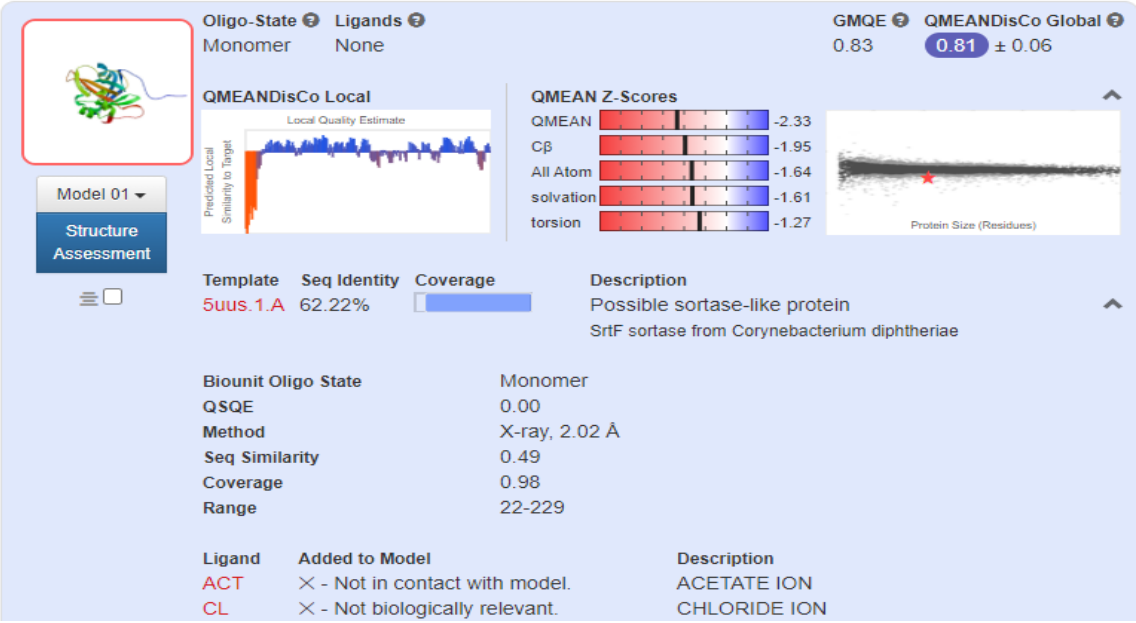


Figure 3.8. Protein structure prediction of CgSrtE

The protein structure was predicted by SWISS-MODEL showing a sequence similarity of 62.22 % with sortase F from *C. diphtheriae.*

Figure 3.9 generated by PyMOL, shows a structural superposition of the modeled CgSrtE on CdSrtF.

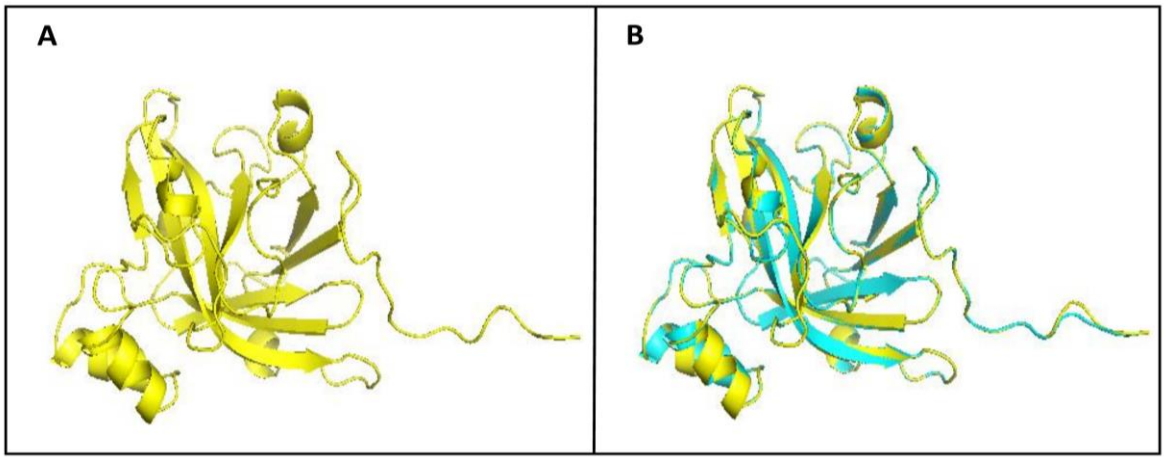
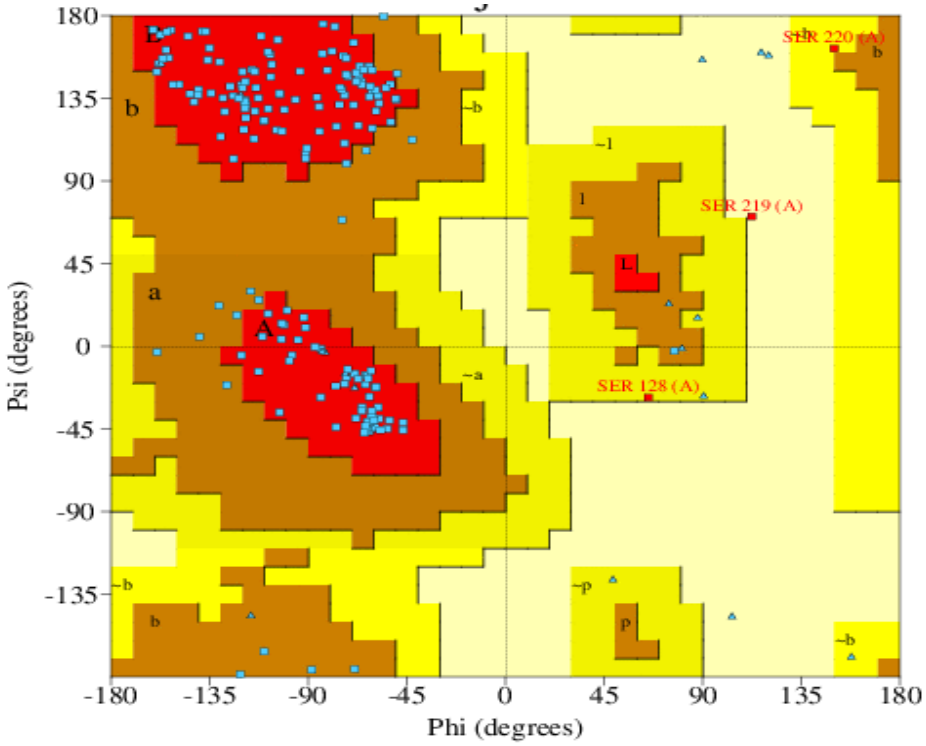


Figure 3.9. Model of CgSrtE on CdSrtF

(A) Homology modeled CgSrtE (yellow); **(B)** The structural superimposition of the homology modeled CgSrtE (yellow) on the CdSrtF template (cyan) (PDB ID: 5UUS). The figure was generated using PyMOL.

The quality of the modeled structure was assessed using PROCHECK and ProSA score analysis. The stereochemical quality and accuracy of the predicted model were evaluated using the Ramachandran Map calculation by the PROCHECK program. The Ramachandran plot showed a tight grouping or clustering of phi~ -50 and psi~ -50 and the backbone conformation of the models. Indeed, the Ramachandran plot grouped the residues based on their quadrangle sections. The red graph areas reflect the most allowed regions, while the yellow regions represent allowed regions. Other residues were depicted by squares, whereas glycine is represented by triangles. The Ramachandran plot shows 157 amino acid residues excluding glycine and proline (88.7 %) in most favorable regions, 17 amino acid residues (9.6 %) fall into additionally allowed regions, 2 amino acid residue (1.1 %) falling into the generously allowed regions and one amino acid residues (0.6 %) in the disallowed region **(Figure 3.10)**. These indicate that the modeled CgSrtE is in good agreement with the template structure (PDB ID: 5UUS).



Plot Statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	157	88.7%*
Additional allowed regions	[a,b,l,p]	17	9.6%
Generously allowed regions	[~a,~b,~l,~p]	2	1.1%
Disallowed regions	[XX]	1	0.6%*
Non-glycine and non-proline residues		177	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		17	
Proline residues		12	
Total number of residues		208	

Based on an analysis of **118** structures of resolution of at least **2.0** Angstroms and *R*-factor no greater than **20.0** a good quality model would be expected to have over **90%** in the most favoured regions [A,B,L].

Figure 3.10. Structure validation of CgSrtE by PROCHECK

The red, bright yellow and light yellow color in the Ramachandran plot represents that 88.7 % residues of CgSrtE protein residues were present in the favorably allowed region (red color), 1.1 % residues were present in the generously allowed region (bright yellow) and 0.6 % residues were present in the disallowed region (light yellow).

In addition, the model quality was assessed using the Z-score derived from the ProSA tool. ProSA compares the Z-Scores of predicted structures to protein structures obtained by NMR and X-ray crystallography of the same size. The projected model had a Z-score of -6.67, which is within the range of Z-scores for protein structures of similar size. The structural validity of the CgSrtE model was validated by these results **(Figure 3.11)**.

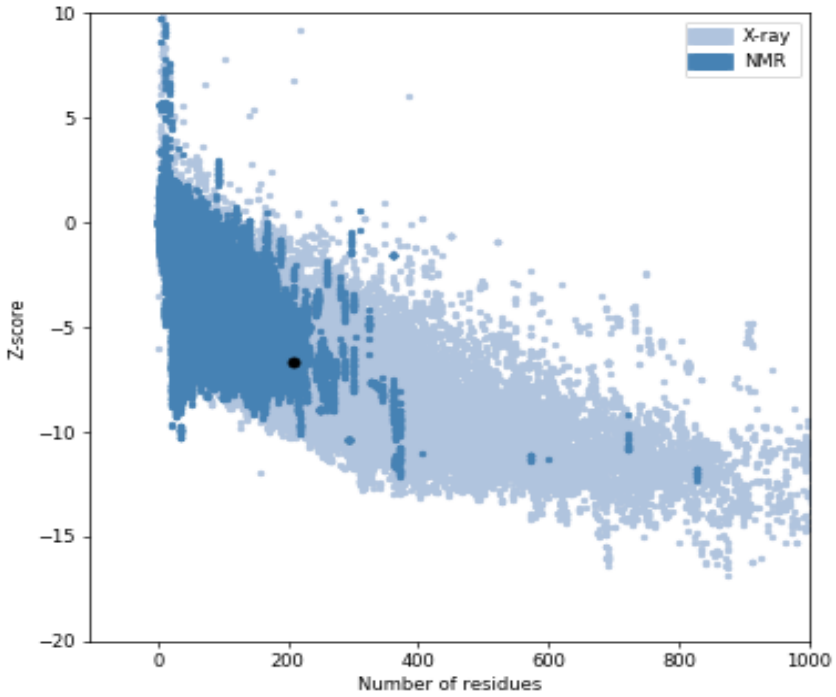


Figure 3.11. Structure validation of CgSrtE by ProSA

ProSA showing Z-score (highlighted as a black dot) of the predicted structure in a plot relative to Z-scores of all experimentally determined protein chains, which were currently accessible in the Protein Data Bank solved using NMR and X-ray. The light blue and dark blue colors were used to differentiate the structures from different sources (X-ray and NMR).

3.3.9. Bioinformatic prediction of Sortase E substrates

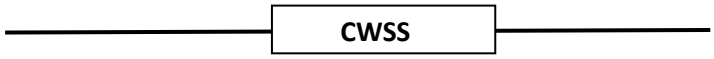
The preliminary identification of sortase substrate proteins in *C. glutamicum* ATCC 13032 was done using a bioinformatics technique. LAXTG has been proposed as the expected

recognition sequence for CgSrtE. The potential candidates were then tested for sortase substrate characteristics which should contain an N-terminal signal peptide sequence and a C-terminal cell wall sorting signal, which includes a hydrophobic transmembrane region followed by a sortase pentapeptide recognition sequence along with two consecutive basic residues (arginine or lysine). Following all these conditions, predicted a substrate protein, Cgl0614 which fulfills all the criteria to be the substrate of Sortase E of *C. glutamicum* **(Table 3.3)**.

Table 3.3. *C. glutamicum* Sortase E substrate protein

***C. glutamicum* ATCC 13032|Cgl0614: 163 aa – Hypothetical membrane protein**

MHSCSMGFGKAILLEIMDPMNSYNTMKSHVTPIPTPVSVRLSSSVLVGAAIAS
LTTDLAPAVKFSVTGVGLALALLIAFAHPYRGEMRMYRFQNNISPVPTIGQV
MPLFFTWLALMLAPIISGAPLWATILVFLAATGWMYLTFPHVDGSRKLAFAE
GPRRNT



	LAXTG	Hydrophobic region	charged tail
Cgl0614 -	LAATG	WMYLTFPHVDGSRKLAFAEGP	RRNT

3.3.10. Primary structure prediction of Cgl0614

The physicochemical properties of the Cgl0614 protein sequence were analyzed by using the ExPasy ProtParam tool. The calculated molecular weight of the protein (Cgl0614) containing 163 amino acids was 17668.98 Da, and the theoretical pI was 9.8, indicating that the protein will precipitate in basic conditions. The number of negatively charged protein

residues (Asp and Glu) is 11, whereas the number of positively charged protein residues (Arg and Lys) is 6. The instability index of the protein showed 37.46, indicating that the protein is stable, since the obtained result is less than the cut-off of 40. The aliphatic index and GRAVY values were 101.78 and 0.548 respectively **(Table 3.4)**.

Table 3.4. Physicochemical properties of *C. glutamicum* Cgl0614

Sr. No	Physicochemical properties of CgSrtE	Values
1	Amino acid residues	163
2	Molecular weight (Da)	17668.98
3	Theoretical pI	9.80
4	Positively charged residue	6
5	Negatively charged residue	11
6	Total number of atoms	2517
7	Molecular formula	$C_{812}H_{1276}N_{206}O_{212}S_{11}$
8	Instability index	37.46
9	Aliphatic index	101.78
10	GRAVY	0.548

3.3.11. Secondary structure validation of Cgl0614

The secondary structure prediction of Cgl0614 was obtained using the SOPMA tool which contains only a combination of alpha-helix (69.3 %) and random coil (30.7 %). These results were further confirmed by PSIPRED were presented in **Figure 3.12**.

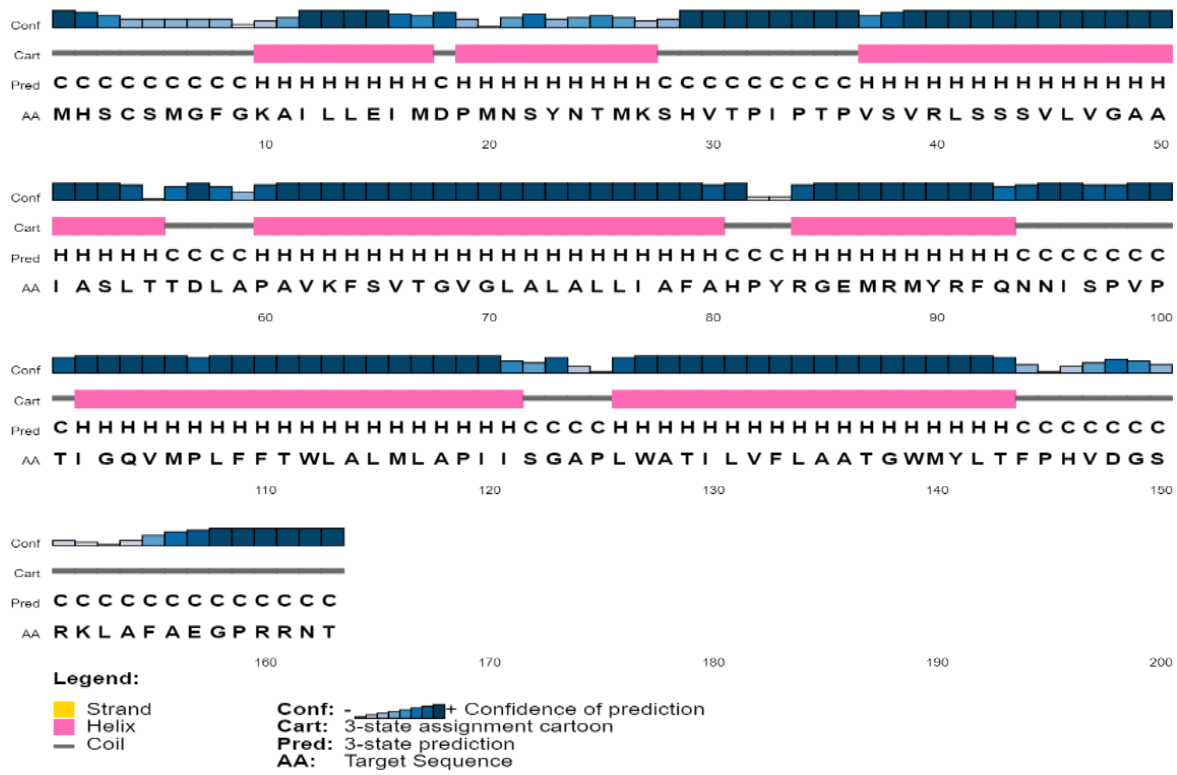


Figure 3.12. Validation of predicted secondary structure of SDP (Cgl0614) of *C. glutamicum* protein by PSIPRED server.

3.3.12. Modeling and validation of Cgl0614

I-TASSER is used for estimating the 3D structure and function of unknown proteins. Using I-TASSER threading approach, the 3D structure of SDP (Cgl0614) of *C. glutamicum* was predicted. For protein structure prediction, I-TASSER (Iterative-Threading/Assembly/ Refinement) was proven to be a better server. I-TASSER structural models are of high quality and resolution. The server generates five models, from which the best one can be chosen using C-score. The correlation quality of the model prediction outcomes was demonstrated by the C-score, which is a measure of the quality of the resultant models. Model 1 was chosen as the best-projected structure model of Cgl0614 based on the C-score. RMSD and TM-score were used to assess the overall model predictions. The Cgl0614 model had a TM score of 0.37

± 0.13 and RMSD of 11.9 ± 4.5Å. The results confirmed the ability of I-TASSER to generate high-quality models for Cgl0614 protein **(Figure 3.13)**.

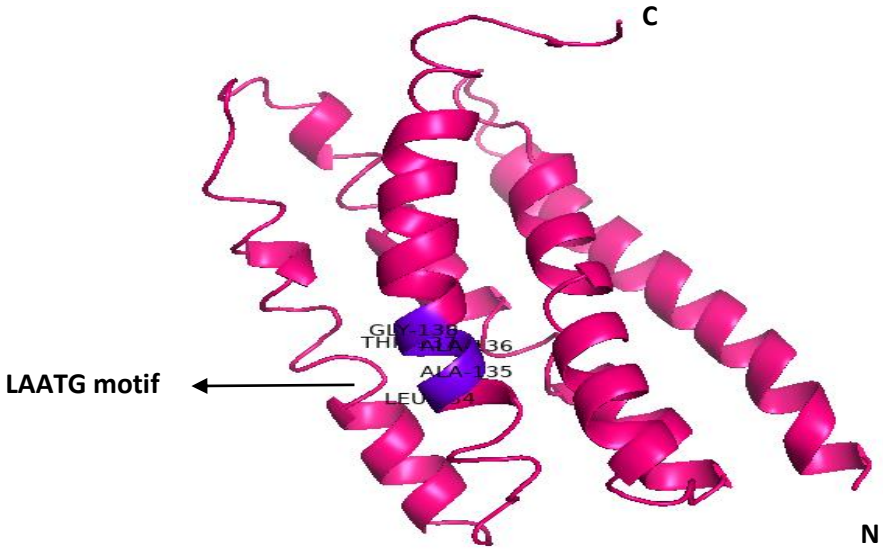


Figure 3.13. Homology modeling of Cgl0614

The modeled Cgl0614 is shown as cartoon representation in magenta color and the sortase substrate motif LAATG in purple color

The Cgl0614 model created by I-TASSER was used to scan the PDB for proteins with structural homologous regions. Ten of the structures were found to be structurally and functionally comparable to Cgl0641 from the PDB database, based on the TM-score **(Table 3.5)**. To analyze the function of the modeled protein, we selected the second-best model, crystal structure of mycolic acid transporter MmpL3 from *Mycobacterium smegmatis* complexed with NITD-349 (PDB ID: 72CM) which function as a transporter protein in *M. smegmatis.* **Figure 3.14** generated by PyMOL, shows a structural superposition of the modeled Cgl0614 on the crystal structure of mycolic acid transporter MmpL3 from *M. smegmatis* complexed with NITD-349.

Table 3.5. Proteins structurally close to the Cgl0614 in the PDB

SL. No	PDB-HIT	TM-score	RMSD	IDENTITY	COVERAGE	FUNCTION
1	7DZQ	0.851	2.22	0.080	0.994	Cryo-EM structure of patched lipid nano disc-the wildtype
2	7C2M	0.780	2.68	0.064	0.951	Crystal structure of mycolic acid transporter MmpL3 from *M. smegmatis* complexed with NITD-349
3	2V50	0.778	2.49	0.093	0.945	The missing part of the bacterial MexAB-OprM system: structural determination of the multidrug exporter MexB
4	7NVH	0.774	2.61	0.090	0.939	Cryo-EM structure of the mycolic acid transporter MmpL3 from *M. tuberculosis*
5	6N3T	0.771	2.71	0.063	0.951	Crystal structure of MmpL3 from *M. smegmatis* complexed with phosphatidylethanolamine
6	3AQP	0.770	2.55	0.064	0.932	Crystal structure of SecDF, a translocon-associated membrane protein, from *Thermus thermophilus*
7	7M4P	0.769	2.56	0.102	0.926	Multidrug efflux pump as AdeJ with eravacycline bound
8	7KGD	0.766	2.70	0.108	0.969	Cryo-EM structures of AdeB from *Acinetobacter baumannii*: AdeB-I
9	6VEJ	0.765	2.30	0.103	0.908	TriABC transporter from *Pseudomonas aeruginosa*
10	5LQ3	0.763	2.90	0.108	0.969	Structures and transport dynamics of the *Campylobacter jejuni* multidrug efflux pump CmeB

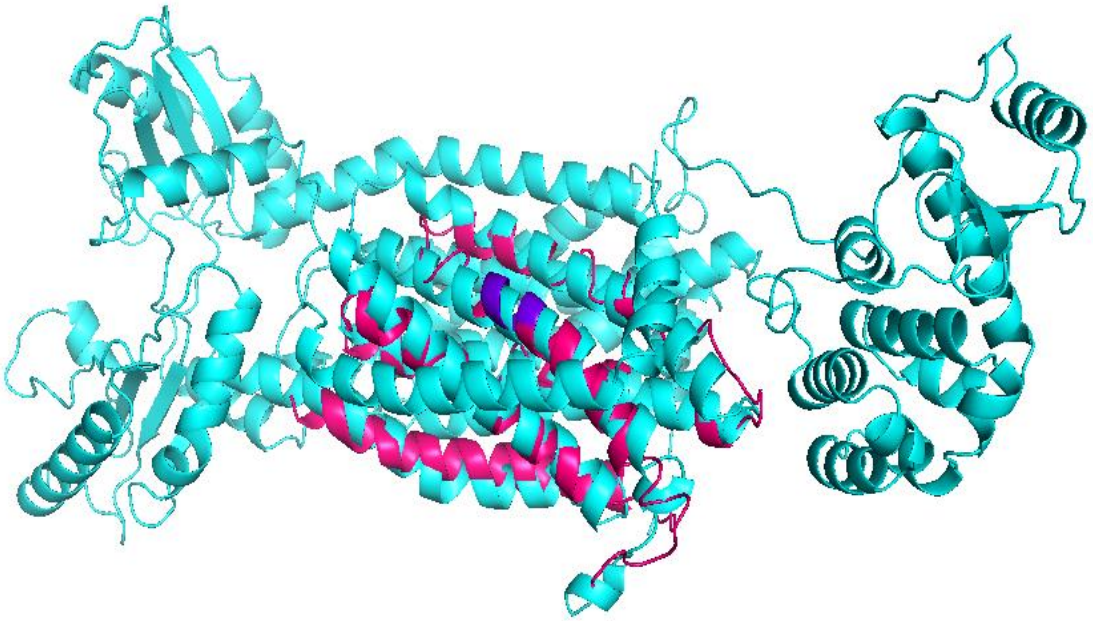
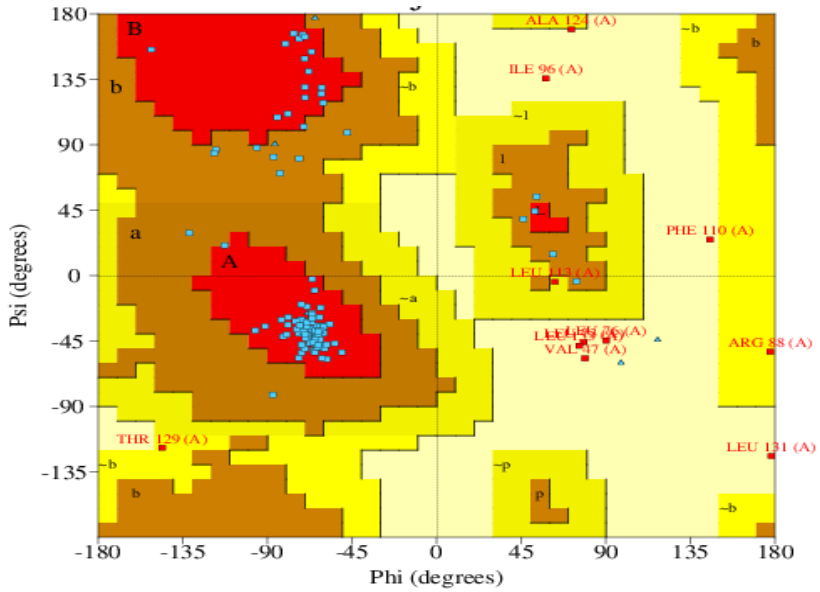


Figure 3.14. Model of Cgl0614 on transporter MmpL3 from *M. smegmatis*

The structural superimposition of the homology modeled Cgl0614 in (magenta) along with the sortase substrate motif LAATG (purple) on the crystal structure of mycolic acid transporter MmpL3 from *M. smegmatis* complexed with NITD-349 (Cyan) (PDB ID: 7C2M) and the figure was generated using PyMOL.

The quality of the modeled structure was assessed by PROCHECK and ProSA score analysis. The stereochemical quality and accuracy of the predicted model were evaluated using the Ramachandran Map calculation using the PROCHECK program. The Ramachandran plot showed a tight grouping or clustering of phi~ -50 and psi~ -50 and the backbone conformation of the models. Indeed, the Ramachandran plot grouped the residues based on their quadrangle sections. The red graph areas reflect the most allowed regions, while the yellow regions represent allowed regions. Other residues were depicted by squares, whereas glycine is represented by triangles. The Ramachandran plot shows 112 amino acid residues excluding glycine and proline (81.8 %) in most favorable regions, 14 amino acid residues (10.2 %) fall into additionally allowed regions, 3 amino acid residue (2.2 %) falling

into the generously allowed regions and 8 amino acid residues (5.8 %) in the disallowed region **(Figure 3.15)**. This indicates that the modeled Cgl0614 is in good agreement with the template structure (PDB ID: 7C2M).



Plot Statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	112	81.8%*
Additional allowed regions	[a,b,l,p]	14	10.2%
Generously allowed regions	[~a,~b,~l,~p]	3	2.2%
Disallowed regions	[XX]	8	5.8%**
Non-glycine and non-proline residues		137	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		11	
Proline residues		13	
Total number of residues		163	

Based on an analysis of **118** structures of resolution of at least **2.0** Angstroms and *R*-factor no greater than **20.0** a good quality model would be expected to have over **90%** in the most favoured regions [A,B,L].

Figure 3.15. Structure validation of Cgl0614 by PROCHECK

The red, bright yellow and light yellow color in the Ramachandran plot represents that 81.8 % residues of Cgl0614 protein residues were present in the favorably allowed region (red color), 2.2 % residues were present in the generously allowed region (bright yellow) and 5.8 % residues were present in the disallowed region (light yellow).

In addition, the model quality was assessed using the Z-score derived from the ProSA tool. ProSA compares the Z-Scores of predicted structures to protein structures obtained by NMR and X-ray crystallography of the same size. The projected model had a Z-score of -2.46, which is within the range of Z-scores for protein structures of similar size. The structural validity of the Cgl0614 model was validated by these results **(Figure 3.16)**.

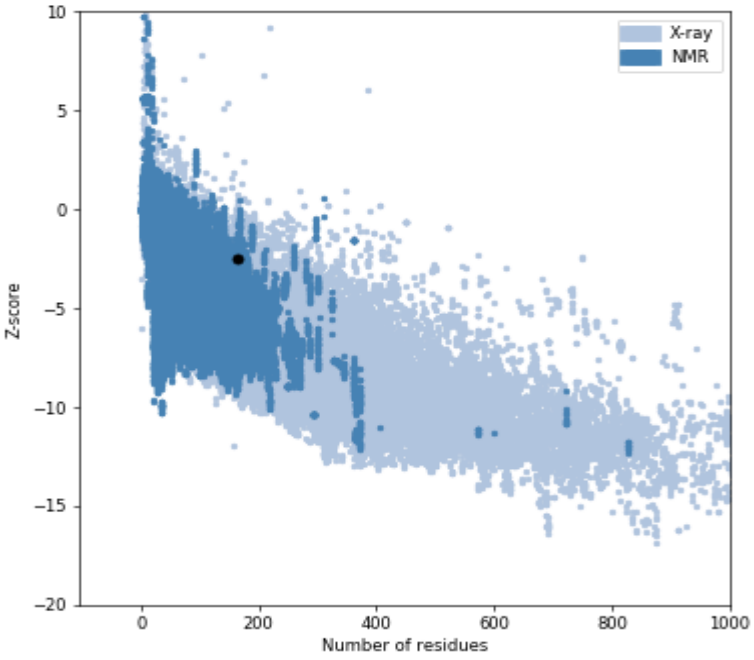


Figure 3.16. Structure validation of Cgl0614 by ProSA

ProSA showing Z-score (highlighted as a black dot) of the predicted structure in a plot relative to Z-scores of all experimentally determined protein chains, which were currently accessible in the Protein Data Bank solved using NMR and X-ray. The light blue and dark blue colors were used to differentiate the structures from different sources (X-ray and NMR).

3.3.13. Molecular docking between CgSrtE and Cgl0614

The molecular process of protein-protein docking was done on the ClusPro web server. Our results showed the docking of active sites (Cys, Arg, and His) of CgSrtE on the LAATG motif of Cgl0614, the results were displayed in PyMol software in the form of rigid

surface structure, the lowest energy obtained was -936.7 KJ/mol. The energy might be required to form a stable complex between CgSrtE and Cgl0614 **(Figure 3.17)**.

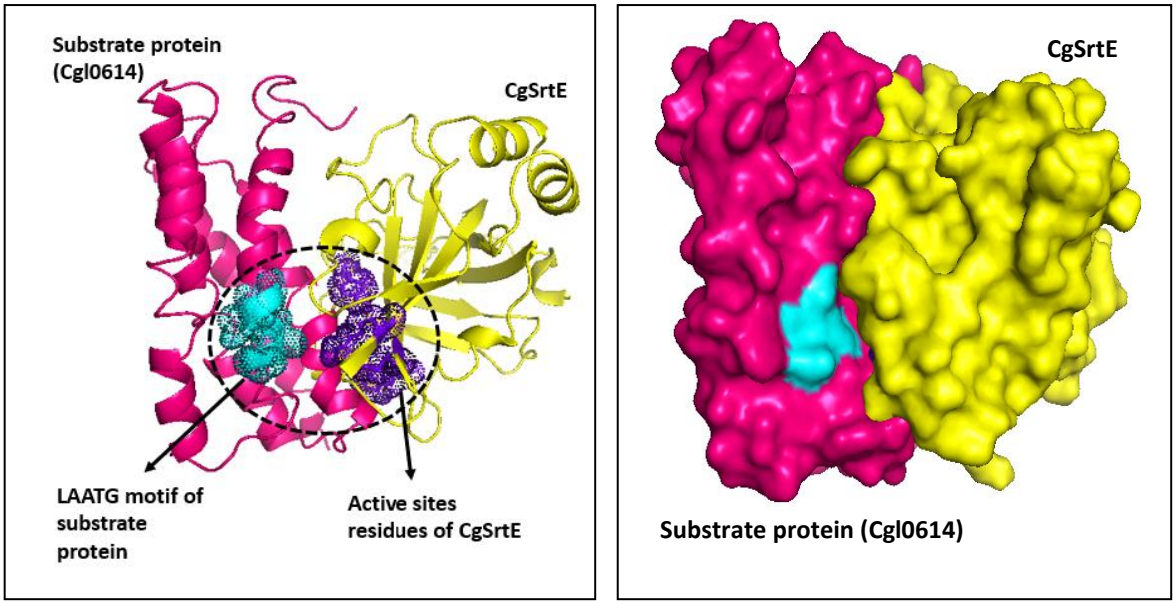


Figure 3.17. CgSrtE-Cgl0614 molecular docking

CgSrtE active sites Cys, Arg, and His binding to the LAATG motif of Cgl0614. The predicted binding site region LAATG showed cyan color on Cgl0614 (magenta color). The CgSrtE (yellow color) active site regions are shown as purple color.

3.4. Discussion

The development of novel sortases with increased catalytic efficiency and specificity is the major focus (Antos et al., 2016). The current section is structured around the idea of determining the protein function from its structure. Bioinformatics analyses were carried out to better understand the distribution of sortase-like proteins in *C. glutamicum*, as well as to uncover putative sortase substrate proteins. The genome analysis of *C. glutamicum* ATCC 13032 predicted a single sortase like transpeptidase (Boekhorst et al., 2005). The SrtE from *C. glutamicum* shows sequence similarity with *C. diphtheriae* and *Streptomyces* species (Das et al., 2017). Based on the expected recognition patterns of CgSrtE, our investigations led to the

revelation that CgSrtE is conserved throughout all the *Corynebacterium* lineages, as well as the identification of a possible sortase substrates protein Cgl0614 with LAATG at the C-terminal in *C. glutamicum* ATCC 13032. The TLxTC sequence towards the C-terminus is the signature motif of sortase enzymes and contains the equivalent of catalytic Cys 239 in *C. diphtheriae* SrtF obtained from the template sequence alignment. The conserved His and Arg of CgSrtE were also aligned well with His 134 and Arg 248 of CdSrtF respectively (Spirig et al., 2011). Despite being missing in other sortases, the tyrosine residue at this site is highly conserved in class E enzymes predicted to recognize LAXTG sorting signals, according to multiple sequence alignment (Comfort and Clubb, 2004; Dramsi et al., 2005). The phylogenetic analysis, multiple alignments, and homology modeling with *C. diphtheriae* revealed that sortase plays a housekeeping role in the organism (Chang et al., 2011).

This study has demonstrated the physicochemical, structural and functional characterization of sortase E and the identification of its substrate protein. The predicted aliphatic index of CgSrtE protein was 74.78 % and Cgl0614 101.78 %, indicating the thermostable nature of the protein. The isoelectric point is at which an amino acid retains the same level of positive and negative charges while having a net charge of zero. The pI value of CgSrtE was 4.52, considering the acidic nature of the enzyme whereas Cgl0614 was found to be basic in nature with a pI value of 9.8. The instability index provides an estimate stability of proteins with value less than 40 (Kim et al., 2011). The instability index of CgSrtE and Cgl0614 were 29.55 and 37.46, suggesting the stability of the proteins.

3.5. Summary

This chapter deals with in silico characterization of sortase and identification sortase-dependent protein anchored by Sortase E in *C. glutamicum*. The bioinformatic analysis

revealed that the *NCgl2838* gene encodes for Sortase E transpeptidase. The phylogenetic analysis of *C. glutamicum* SrtE revealed that the protein is conserved within the genomes of *Corynebacterium* sp. and functions as a housekeeping gene within the organism. The 3D structure of CgSrtE was built by the SWISS-MODEL server and showed 62 % similarity with that of *C. diphtheriae* SrtF in PDB. Using the I-TASSER threading technique, a 3D structural model of Cgl0614 was created. C-scores, TM scores, and RMSD were used to evaluate structural models. The generated substrate protein (Cgl0614) with a LAATG sorting motif was found to be a similar to the crystal structure of mycolic acid transporter MmpL3 from *M. smegmatis* complexed with NITD-349 (PDB ID: 72CM) and predicted to function as a transporter in *C. glutamicum*. The generated modeled protein by SWISS-MODEL and I-TASSER was then validated using multiple quality assessment methods such as PROCHECK and ProSA. The Ramachandran plot using PROCHECK for CgSrtE and predicted substrate protein (Cgl0614) showed more than 80 % amino acid residues in the most favored region. In addition, the model quality was assessed using the Z-score derived from the ProSA tool. According to the X-ray and NMR experiments, the Z-score for the CgSrtE and Cgl0614 model were -6.67 and -2.46 respectively, which were within the acceptable range. The key residues at the binding site of the CgSrtE were determined to be Cys 240, Arg 249, His 135, and Tyr 118. Based on the computational studies we were able to predict a single sortase-dependent protein Cgl0614 which fulfills all the criteria; an N-terminal signal peptide, a C-terminal sorting motif, a transmembrane region, and a positively charged tail in *C. glutamicum*. The lowest docking energy between Sortase E and substrate was -936.7 KJ/mol, which was displayed using PyMOL using the ClusPro web-server.

Chapter 4

Construction of *srtE* deletion and overexpression strains of *C. glutamicum* to study the morphological and physiological alterations

Abbreviation

BHI	Brain Heart Infusion
nt	Nucleotides
PBS	Phosphate Buffered Saline
SEM	Scanning Electron Microscope
v/v	% volume per volume
WT	Wild-type

4.1. Introduction

Most pathogenic and nonpathogenic Gram-positive bacteria contain sortase enzymes that are involved in two functions within the cell wall, one is surface protein anchoring and another one is pilus assembly. In pathogenic bacteria, surface proteins or pili proteins play a significant role in contributing towards cell adhesion and pathogenesis by attaching to specific organ tissues of the host cells during infection or providing a way to escape from the host immune response (Pallen et al., 2001; Cheng et al., 2009). Since, sortase is widely distributed among pathogenic bacteria, they constitute a promising therapeutic target for the development of novel antibiotics. However, in nonpathogenic bacteria especially in probiotics, the sortase plays a pivotal role in eliciting adhesion to epithelial cell lines, modulating the immune response of host cells, and also aiding in bile salt stress resistance (Dieye et al., 2010; Muñoz-Provencio et al., 2012; Kebouchi et al., 2016).

Among the Coryneform bacteria, *C. glutamicum* is a well-known industrial microbe for the production of amino acids and other low-molecular-weight substances (Suzuki et al., 2005; Becker and Wittmann, 2012; Wendisch et al., 2016). Besides this, *C. glutamicum* expression system has been commercialized as CORYNEX®, an alternative potential platform for the secretion of heterologous proteins (Yokoyama et al., 2010; Matsuda et al., 2014). The secretory proteins in *C. glutamicum* are generally translocated via the secretory pathway in an unfolded manner through the SecYEG translocation pore. Upon the arrival of the secretory protein, the signal peptidase cleaves the signal peptide followed by the folding of the protein into its functional conformation (Hemmerich et al., 2019). While the membrane-bound sortase E enzyme recognizes the surface proteins translocated via a secretory pathway with a LAXTG motif. The sorting signal stops translocation across the membrane, allowing sortase to cleave the LAXTG motif and anchor to the surface. Generally, deletion of sortase prevents virulence factors from attaching to the cell wall in Gram-positive bacteria.

In this work, *srtE* deletion was created using the pK19mobsacB suicide vector to better understand the impacts on bacterial physiology and morphology, which have yet to be completely investigated. In *C. glutamicum*, gene insertion, deletion, or replacement is often accomplished using homologous recombination with non-replicative integration vectors and genome alteration can be accomplished through two rounds of positive selection (Schagfer et al., 1994). The non-replicative plasmid pK19mobsacB is made up of a Tn5-derived kanamycin resistance cassette, flanking sequences of about 500 nt homologous to chromosomal sequences upstream and downstream of the DNA sequence to be exchanged, and a *sacB* gene encoding levansucrase, which catalyzes sucrose. When cultivated in the presence of sucrose, this chemical kills the *C. glutamicum* cells. Since, pK19mobsacB plasmid cannot multiply in *C. glutamicum*, only cells with the chromosomally integrated plasmid are chosen following electroporation, as kanamycin resistance may be developed.

After culture, recombinant cells are chosen for vector sequence loss by a second homologous recombination event by spreading cells on sucrose-containing plates. Sortase E was also overexpressed in *C. glutamicum* utilizing the pVWEx1 vector which is a well-established expression vector for *C. glutamicum* that uses the P_{tac} promoter and contains a *lacI* gene for IPTG-inducible gene expression (Peters-Wendisch et al., 2001). Because *C. glutamicum* lacks a homolog, these IPTG-inducible expression methods utilize the Lac repressor from *E. coli* and supply lac operator (Oehler et al., 1994).

In this study, sortase deletion mutants and overexpression strains in *C. glutamicum* were created with the goal of better understanding the morphological and physiological changes in the cells. In the course of the studies, it became evident that *srtE* being a housekeeping gene of *C. glutamicum* did not affect the growth and viability when sortase was deleted. However, when sortase was overexpressed, the results showed retarded growth as compared with the wild-type or the Δ*srtE* deletion mutant cells.

4.2. Material and Methods

4.2.1. Construction of deletion mutant of *C. glutamicum*

The plasmid cloning was done in *E. coli* DH5α bacteria. A non-replicative pK19mobsacB was used to make the in-frame deletion mutant. pK19mobsacB is a suicide plasmid that performs homologous recombination in two steps (Schagfer et al., 1994). A kanamycin resistance cassette and the *sacB* gene-producing levansucrase are both present on this plasmid. The amplification of upstream and downstream flanking regions of ~500 bp of the gene was PCR amplified using the primers srtE-1A F/srtE-2A R and srtE-3A F/srtE-4A R to create a *srtE* deletion plasmid. The amplification conditions were mentioned in **Table 2.4 (Chapter 2)**. After the first round of PCR, the gel extraction was used to purify the PCR products by QIAquick gel extraction kit (Qiagen, Germany). Purified PCR products from

both flanking areas were subjected to crossover PCR by 21 bp overlap sequence and amplified using the primer pairs srtE-1A F/ srtE-4A R. The 1000 bp segments were digested with EcoRI/BamHI and ligated into the pK19mobsacB plasmid. Electroporation **(Section 2.2.4.4 in Chapter 2)** is used to fully incorporate the plasmid into the genome of electrocompetent cells of *C. glutamicum* **(Section 2.2.4.3 in Chapter 2)** after the first homologous recombination, and selection was performed on a BHI agar plate with 25 μgmL^{-1} kanamycin. The kanamycin-resistant clones were chosen and grown for 2 h at 30 °C and 180 rpm on a BHI medium. The cultures were diluted to 1:10, 1:100, and 1:1000 dilutions after 2 h and plated on BHIS agar plates with 10 % sucrose. SacB permits sucrose to be broken down into glucose and fructose. It also transforms fructose into the oligomeric levan, which causes death (Jager et al., 1992). As a result of the second recombination, only kanamycin-sensitive and sucrose-resistant cells were developed, resulting in the deletion of the *sacB* gene **(Figure 4.1)**. The clones were confirmed by colony PCR using the primer pair Cgl F/Cgl R and DNA sequencing carried out to determine the chromosomal deletion of the gene of interest. All the primers synthesized were listed in **Table 2.3 (Chapter 2)**.

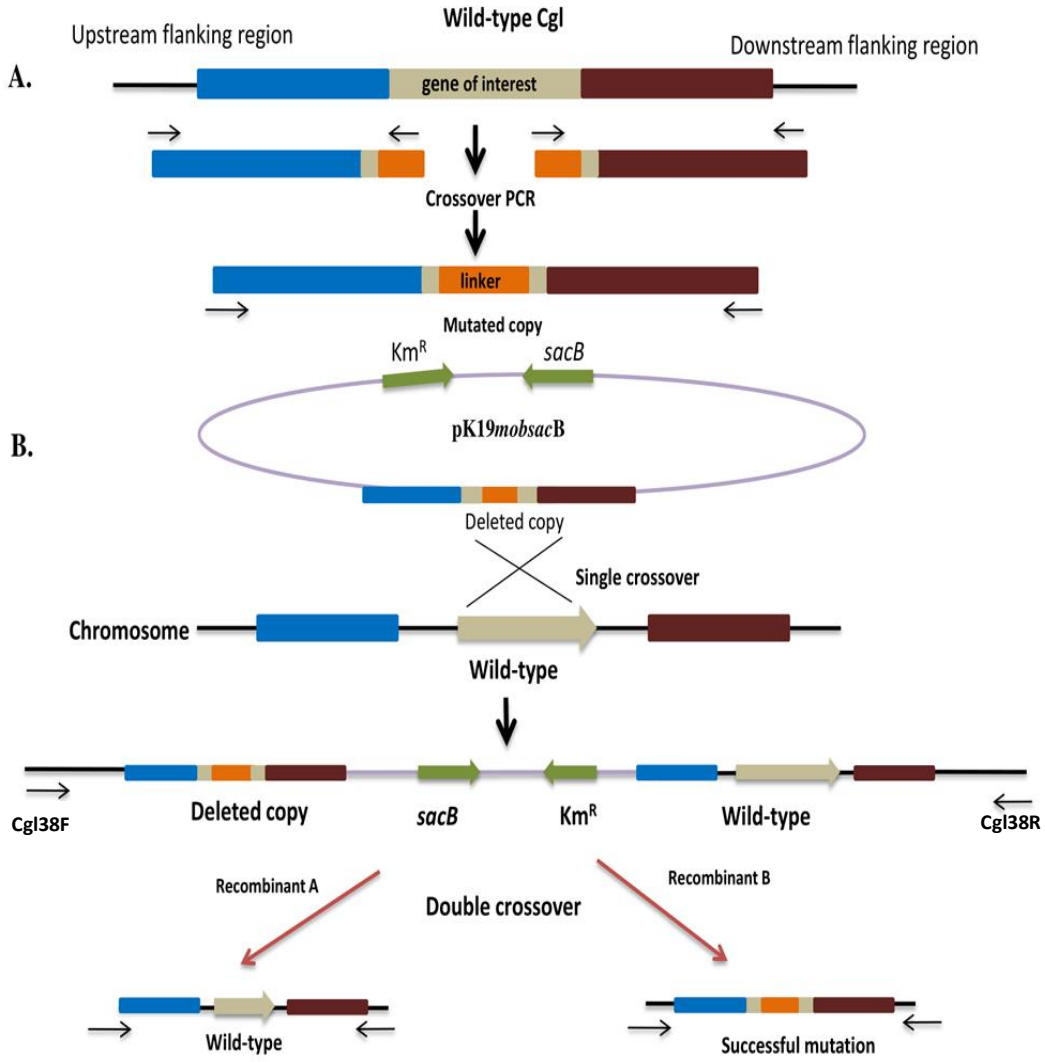


Figure 4.1. Generation of a gene deletion cassette

(A) pK19mobsacB, a non-replicating vector in *C. glutamicum* used to generate gene deletion mutants in this organism. This vector contains a kanamycin resistance gene (Km^R), *sacB*, a multiple cloning site (MCS) within a *lacZα* fragment, and origins of replicon *oriV* and *oriT*. To delete a gene of interest (off-white color), two sets of primers (srtE-1A F/srtE-2A R and srtE-3A F/srtE-4A R) were designed for two PCR reactions (PCR1 and PCR2) that respectively generate ~500 bp fragments. Restriction enzyme sites are incorporated into primers srtE-1A F and srtE-4A R which permits the annealing of third PCR reaction by crossover PCR to generate 1 kb fragment and cloned into pK19mobsacB; **(B)** The recombinant plasmid is electroporated into the genome of *C. glutamicum* after two rounds of homologous recombination.

4.2.2. Construction of plasmids for complementation of *srtE*

To complement the *srtE* deletion in *C. glutamicum*, the *srtE* native expression cassette was amplified using the primer pair NCgl2838 F/R, digested with SalI and BamHI, and ligated into similarly digested pVWEx1 to generate pVWEx1-com*srtE*. The pVWEx1-com*srtE* plasmid was electroporated into *C. glutamicum*-Δ*srtE* to generate the sortase E complemented strain of *C. glutamicum* Δ*srtE*/*srtE*. Colony PCR was used to confirm the transformants, followed by PCR on DNA isolated from pure cultures. PCR and sequencing were used to verify the final complemented strain.

4.2.3. Homologous overexpression of *srtE* from *C. glutamicum*

For overexpression of *srtE*, the following primers NCgl2838 F/ NCgl2838 R **(Table 2.3 in Chapter 2)** were designed for amplification of the gene from genomic DNA of wild-type (WT) *C. glutamicum* **(Table 2.4 in Chapter 2)**. The PCR product was purified and double digested at the SalI/BamHI (NEB, Japan) restriction sites in the pVWEx1 (*Corynebacterium/E. coli)* shuttle vector. The digested PCR product and pVWEx1 vector were ligated together with T4 DNA ligase (NEB, Japan) by an overnight reaction at 4 °C, transformed into *E. coli* DH5 alpha cells and selected on Luria Bertani (LB)-kanamycin (50 μgmL^{-1}) plates. The resulting plasmid p*srtE* was confirmed by sequencing and electroporated into *C. glutamicum* for studying overexpression of the *srtE* gene.

4.2.4. Growth dynamics

For the cultivation of *C. glutamicum* and its derivatives (Δ*srtE*, Δ*srtE*/*srtE,* and p*srtE*), initially, a pre-culture was inoculated with single colonies from agar plates into 10 mL of BHI medium and incubated overnight with shaking at 180 rpm. Secondary cultures were inoculated by a dilution factor of 1/100 with shaking at 180 rpm at 30 °C. A reading of optical

density at 600 nm (OD_{600}) was recorded up to 30 h, of which every 2 h from 0-12 h was recorded and then 30 h.

4.2.5. Cell surface hydrophobicity

Hexadecane was used to test cell surface hydrophobicity as a measure of bacterial adherence (Divya et al., 2012). A freshly produced culture was harvested for 5 min by centrifugation (8000 x g) at 4 °C. The cells were washed twice in PBS (pH 7.2) before being resuspended in the same buffer with the optical density (OD_{600}) set to 0.8–1.0. Then, by vortexing for 2 min, this cell suspension was completely mixed with an equal amount of hexadecane. For 30 min, the two stages were allowed to separate at room temperature. The optical density of the aqueous phase was measured spectrophotometrically at 600 nm before and after mixing with n-hexadecane. The percentage hydrophobicity was calculated as follow:

$$\textbf{Hydrophobicity (\%)} = [(OD_{before} - OD_{after})/OD_{before}] \times 100$$

4.2.6. Scanning electron microscopy

The WT *C. glutamicum* and its derivatives (Δ*srtE*, Δ*srtE*/*srtE*, and p*srtE*) were cultured for 20 h at 30 °C and 180 rpm. The culture was centrifuged for 5 min at 4 °C at 3500 x g. The cells were centrifuged at 3500 x g for 2 min after being rinsed with 0.1 M PBS. The supernatant was discarded, and the cells were fixed in 0.1 M PBS for 4 h at 4 °C with 3 % (v/v) glutaraldehyde. The cells were centrifuged at 3,500 x g for 3 min, washed in 0.1 M PBS for 15 min, and dehydrated twice in an ascending acetone series (30, 50, 70, 90, and 100 percent) for 10 min, with the last step done three times. The stubs were then carefully removed and 10 µL of the samples dehydrated in 100 % acetone were coated on a thin coating of aluminium foil and a gold layer on top for 15–20 min under vacuum and viewed on the scanning electron microscope (JSM - 5600LV, JEOL, Japan).

4.2.7. Spot dilution assay

WT *C. glutamicum* and its derivatives (Δ*srtE*, Δ*srtE*/*srtE* and p*srtE*) were grown overnight in a 5 mL BHI medium at 30 °C. Measured OD_{600} and prepared dilutions of OD_{600} = 1 with 0.9 % saline. Prepared a series of sequential from 10^{-1}-10^{-8}. The final three dilutions (10^{-6}, 10^{-7} and 10^{-8}) of each 10 µL of WT and Δ*srtE* were dropped on BHI plate and Δ*srtE*/*srtE* and p*srtE* were dropped on BHI plate supplemented with 25 µgµL^{-1} of kanamycin. The plates were allowed to dry fully before being incubated for 48 h at different temperatures (30 °C, 37 °C, and 40 °C).

4.3. Results

4.3.1. Construction of sortase-deficient mutant in *C. glutamicum*

The ~500 bp both upstream and downstream flanking regions of *srtE* were amplified from the genomic DNA of *C. glutamicum* using upstream primer pair srtE-1A F/srtE-2A R and downstream primer pair srtE-3A F/srtE-4A R shown in chapter 2 **(Table 2.3)** and *Taq* DNA polymerase at an annealing temperature of 60 °C. The optimized PCR condition of both flanking regions was similar to that of the *srtE* gene as shown in Chapter 2 (**Table 2.4)**. The purified PCR products from both flanking areas were subjected to crossover PCR and amplified using the primer pairs srtE-1A F/ srtE-4A R. The ~1000 bp segments were digested with EcoRI/BamHI and ligated into the pK19mobsacB plasmid and transformed into *E. coli* DH5 alpha cells **(Figure 4.2)**.

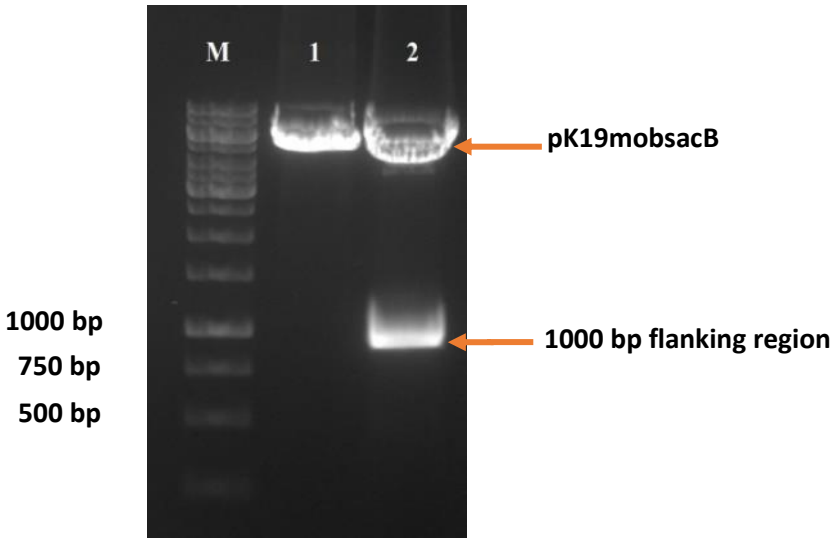


Figure 4.2. Restriction digestion of deletion mutant in pK19mobsacB

Lane M: 1 kb DNA ladder; **Lane 1:** Native plasmid pK19mobsacB with EcoRI/BamHI double digest; **Lane 2:** pK19mobsacB-Δ*srtE* with EcoRI/BamHI double digest showing insert release at 1000 bp flanking region (upstream and downstream flanking region).

A sortase deletion mutant of *C. glutamicum* was generated by two-step homologous recombination. The Δ*srtE* mutant of *C. glutamicum* was confirmed by colony PCR and PCR isolated from DNA of WT and Δ*srtE* mutant of *C. glutamicum* using primer pair Cgl F/Cgl R flanking the *srtE* deletion target **(Figure 4.3)**. The resulting mutants were sequenced by Cgl F/Cgl R primers, sequences were verified and the ORF analysis proved that the amplicon obtained was devoid of *srtE* gene sequence. The following deletion mutant was designated as Δ*srtE*, which comprises around 1000 bp both upstream and downstream flanking regions together.

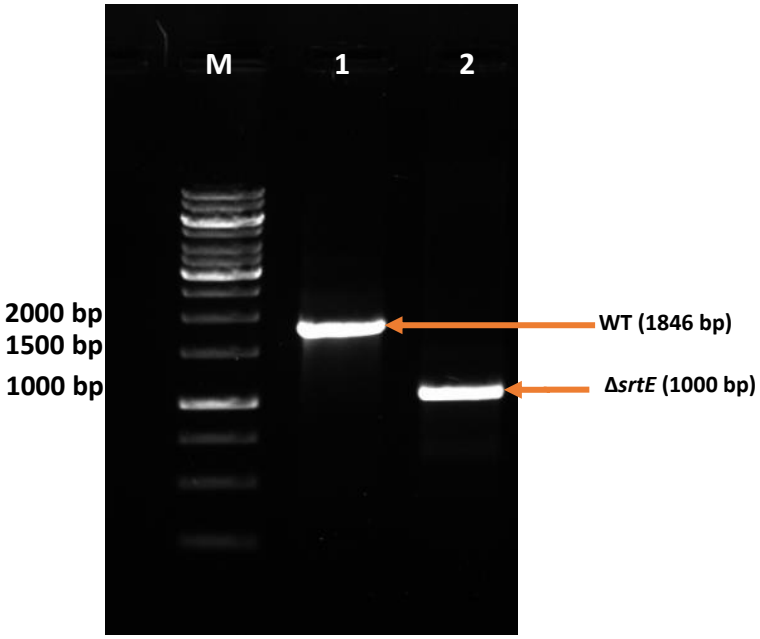


Figure 4.3. Confirmation of *srtE* gene deletion by PCR

Gel picture showing Δ*srtE* gene amplification in *C. glutamicum* 13032, **Lane M:** 1 kb DNA ladder, **Lane1**: WT *srtE* (1846 bp), **Lane 2:** Δ*srtE* (1000 bp).

4.3.2. Complementation of *srtE* gene in the deletion mutant

The identity of the deleted gene was verified by complementing the Δ*srtE* strain with a plasmid. Native *srtE* gene was amplified from the genomic DNA of *C. glutamicum* and *Taq* DNA polymerase at an annealing temperature of 60 °C. The optimized PCR condition was shown in Chapter 2 (**Table 2.4)**. The PCR product was digested with SalI/BamHI and cloned into pVWEx1 digested with SalI and BamHI, creating pVWEx1-com*srtE*. The plasmid was further electroporated into electrocompetent cells of *srtE* mutant. The complemented strain was confirmed by upstream and downstream flanking primers Cgl F/R and by gene-specific primer pair Cgl2838 F/R **(Figure 4.4)**.

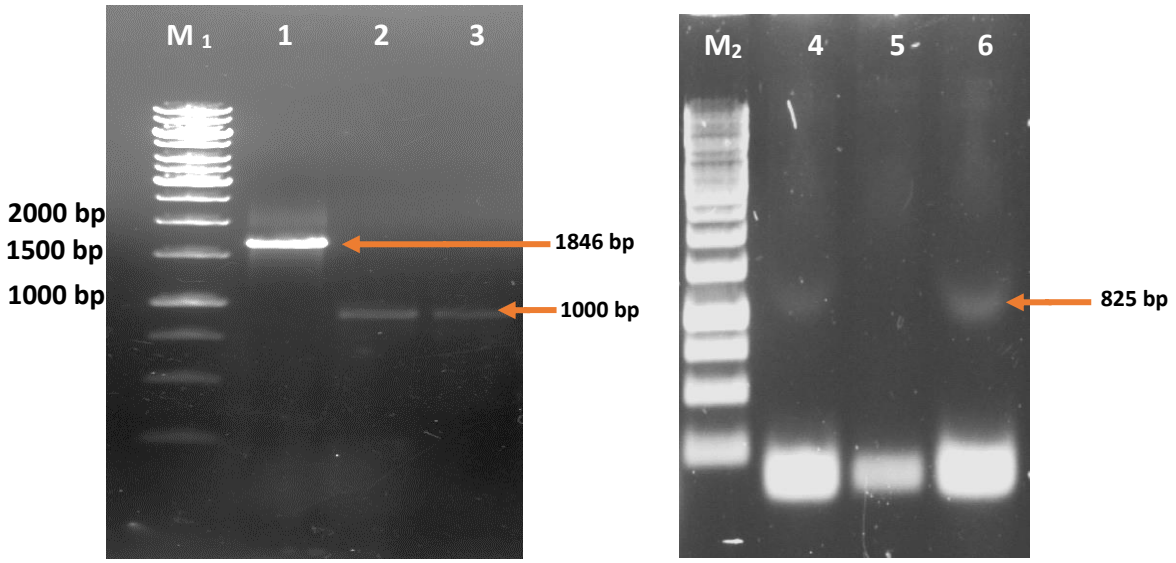


Figure 4.4. Confirmation of complement strain by upstream and downstream primers (Cgl F/R) and *srtE* gene-specific primers (Cgl2838 F/R)

Amplification by Cgl F/R primers; **Lane M_1:** 1kb DNA ladder; **Lane 1**: WT; **Lane 2:** Δ*srtE*; **Lane 3:** Δ*srtE/srtE*; Amplification by Cgl2838 F/R primers; **Lane M_2:** 1kb DNA ladder; **Lane 4**: WT; **Lane 5:** Δ*srtE*; **Lane 6:** Δ*srtE/srtE*

4.3.3. Overexpression of *srtE* in *C. glutamicum*

srtE was overexpressed in *C. glutamicum* 13032 to determine the rate of surface protein anchoring in vivo. *srtE* of 825 bp were amplified from isolated *C. glutamicum* genomic DNA. The genes were cloned into pVWEx1 and electroporated within the host cell. The recombinant plasmids were confirmed by double digestion with SalI/BamHI restriction enzymes **(Figure 4.5)**. The clones were sequenced by pVWEx1 primers, sequences were verified and the ORF analysis proved that the amplicons were completed without any kind of mutation.

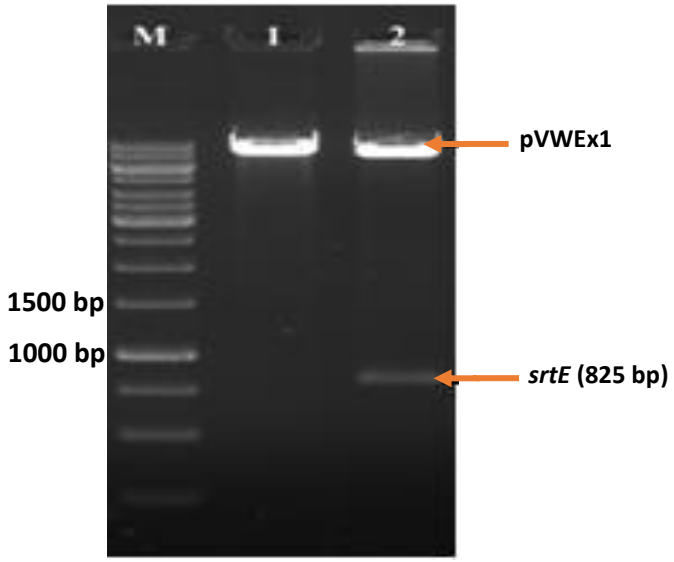


Figure 4.5. Homologous expression of *srtE* in *C. glutamicum*

Lane M: 1 kb DNA ladder; **Lane1:** Native plasmid pVWEx1 with EcoRI/BamHI double digest; **Lane 2:** pVWEx1-*srtE* with EcoRI/BamHI double digest showing insert release at 825 bp.

4.3.4. Effect of *srtE* deletion and overexpression on cell growth

To investigate the role of *srtE*, we created a *C. glutamicum* Δ*srtE* mutant, complement and overexpressing strain, and observed its growth characteristics in BHI medium at an optimum temperature of 30 °C. As shown in **Figure 4.6**, the wild-type and the Δ*srtE* mutant strains showed an almost identical growth rate (0.082 ± 0.007 h^{-1}). The complemented strain showed a slightly reduced rate of growth (0.065 ± 0.001 h^{-1} and this could be due to the presence of the expression vector. However, cells overexpressing the *srtE* gene, grew at a much slower rate (0.042 ± 0.001 h^{-1}). The *srtE*-overexpressing cells initially grew slowly in the BHI medium, but later the turbidity was nearly comparable to that of the original strain.

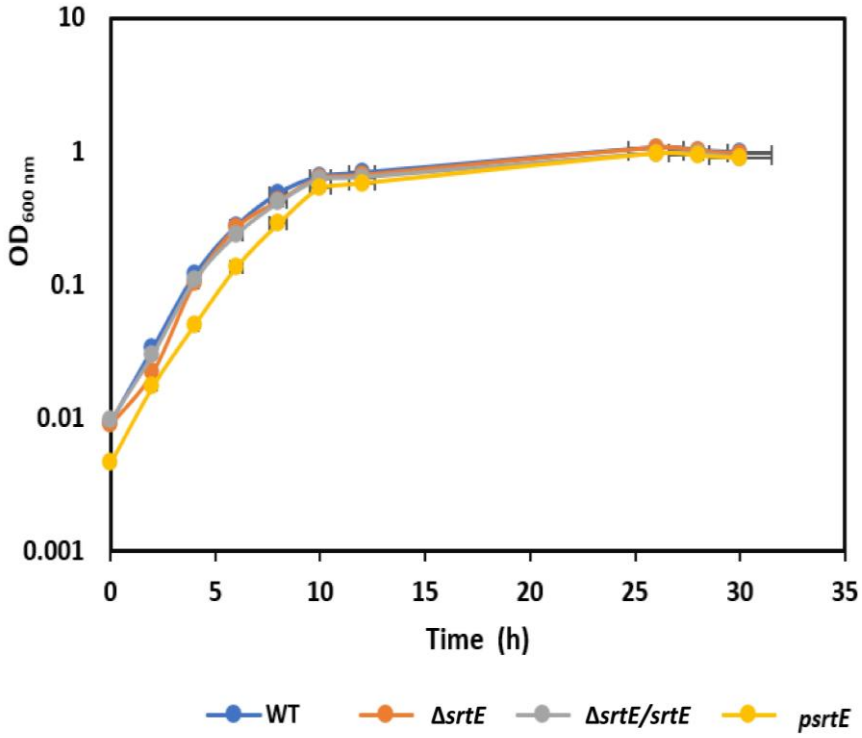


Figure 4.6. Comparison of growth profile of *C. glutamicum* and its derivatives

Comparative growth of wild-type (blue color) and *srtE* deletion (orange color), complemented strains (grey color) and overexpression (yellow color). The cultures were grown for 30 h at 30 °C in BHI medium, and the optical density at 600 nm was measured at the indicated times.

4.3.5. Cell surface hydrophobicity of *Corynebacterium* strain and its derivatives

The bacterial adherence to hydrocarbon by WT *C. glutamicum* and its derivatives were examined by the liquid partitioning method. The *C. glutamicum* wild-type was highly hydrophobic (64 % adhesion to hexadecane), while the adhesion of Δ*srtE* to hexadecane was significantly reduced to 15 %. The complement was able to restore the hydrophobicity similar to that of WT. However, the overexpressed sortase gene showed a reduction in hydrophobicity around 30 %, similar to that of the sortase deletion mutant **(Figure 4.7)**. The reduction in cell surface hydrophobicity in *C. glutamicum* mutant might be influenced by sortase-dependent protein, anchored by SrtE-dependent mechanism.

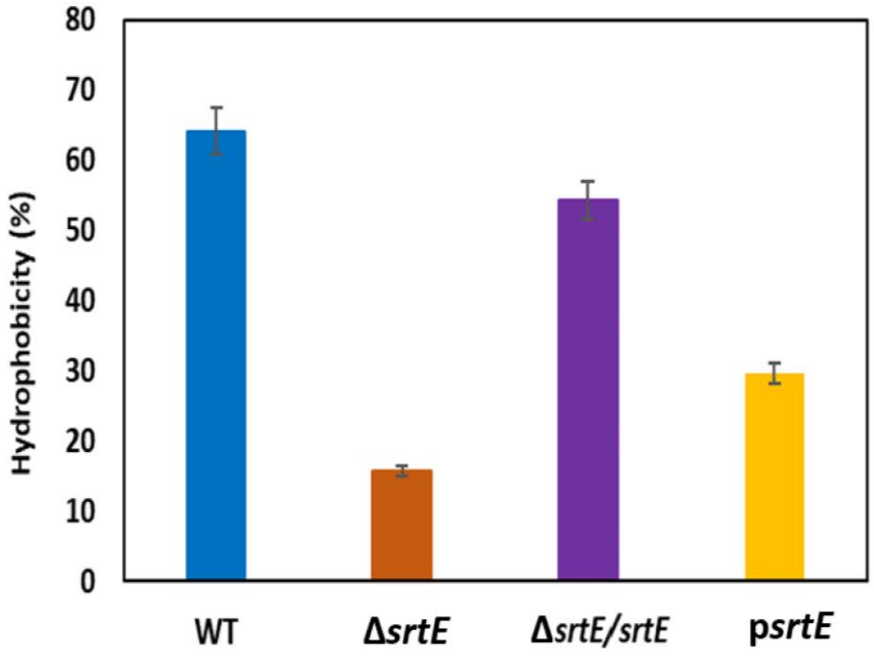


Figure 4.7. Cell surface hydrophobicity of *C. glutamicum* and its derivatives

Results are the means from three experiments, and the bars represent standard deviations.

4.3.6. SEM analysis

To identify the morphological changes in the cells, WT *C. glutamicum* and its derivatives (Δ*srtE*, Δ*srtE*/*srtE*, and p*srtE*) were examined under a scanning electron microscope. The morphology of the Δ*srtE* mutant cells, which had been grown on BHI media was unchanged, according to observations made under a scanning electron microscope. The complement did not show much changes in the morphology and was more or less similar to that of the parental strain. However, the cells of an overexpressed strain containing p*srtE*, showed an elongated cell growth when compared to WT *C. glutamicum* **(Figure 4.8)**.

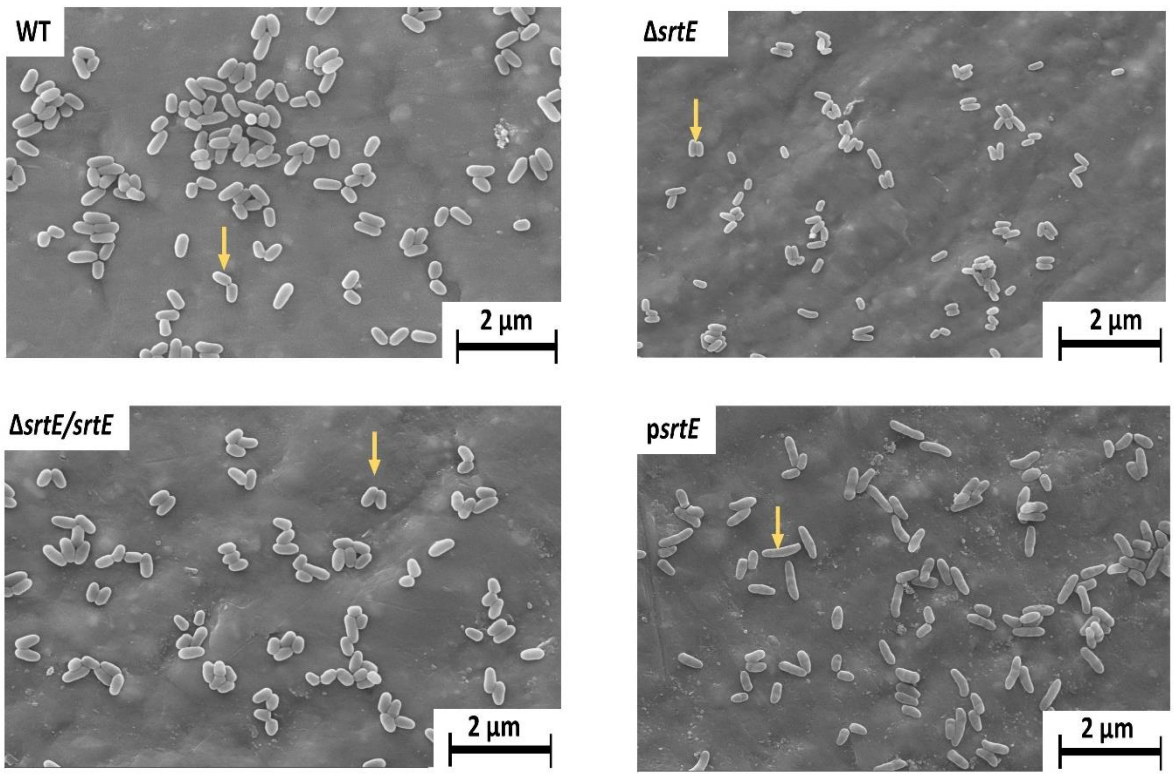


Figure 4.8. Cellular morphology analyzed by SEM

Late-exponential-phase bacterial cultures were treated with acetone gradient were fixed on a specimen holder with a thin coating of aluminium foil and a gold layer on top. The samples (WT, Δ*srtE*, Δ*srtE*/*srtE*, and p*srtE*) were examined by scanning electron microscope.

4.3.7. Involvement of *srtE* in heat stress response

The spot dilution assay at temperature 30 °C and 37 °C were carried out with WT *C. glutamicum* and its derivatives (Δ*srtE*, Δ*srtE*/*srtE*, and p*srtE*) to analyze the viability of the cells after heat stress. The deletion mutant as measured by optical density, noted a significant difference in the numbers of cells. Surprisingly, in the deletion mutant strain, the number of viable cells after the temperature had shifted from 30 °C to 37 °C was significantly reduced by 20 % compared to the parental strain **(Figure 4.9)**. However, the complement was able to restore its viability as that of the parental strain. In the wild-type strain, the number of viable cells did not change appreciably. This indicates that the *srtE* gene is relevant to survival after heat stresses. On the other hand, p*srtE* the overexpressing strain did not show any viable

growth when the temperature was shifted from 30 °C to 37 °C, indicating the loss of membrane integrity which may result in cell death.

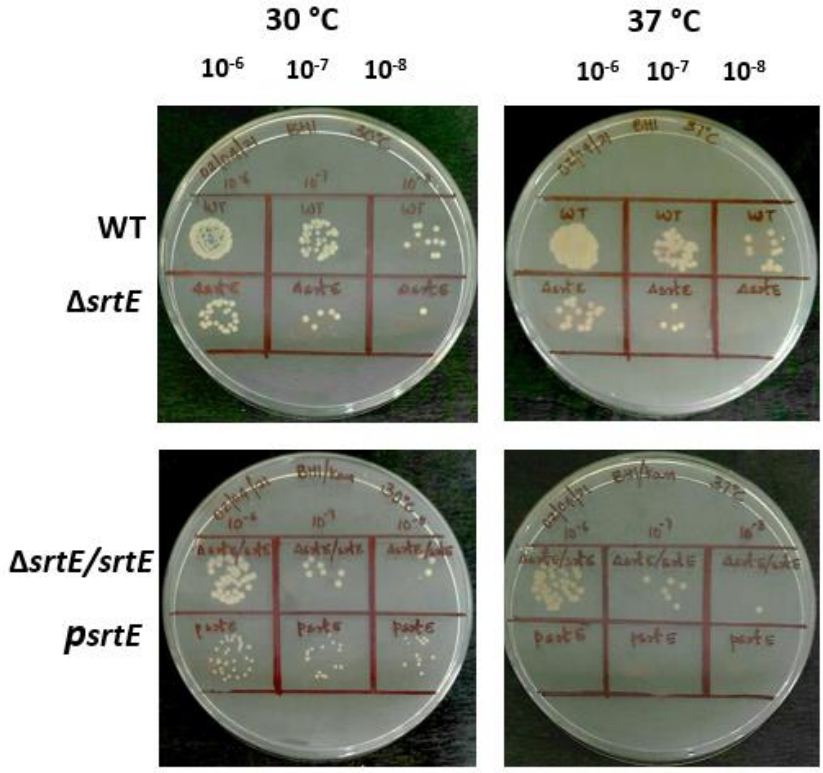


Figure 4.9. Heat stress response of mutant and overexpressing strain of *C. glutamicum*

C. glutamicum and its derivatives were grown at 30 °C and 37 °C in BHI medium and viable cells were counted by plating 10 µL of 10^{-7} diluted cell cultures on BHI medium.

4.4. Discussion

We demonstrated the physiological and morphological changes that occur in *C. glutamicum* when sortase is deleted or overexpressed. The absence of the *srtE* gene in *C. glutamicum* had no effect on cell growth or morphology when compared to wild-type. This is similar to the previous research on *srtA* gene of most pathogenic Gram-positive bacteria, where inactivation of the sortase gene had no effect on microbial cell culture growth or morphology (Mazmanian et al., 1999; Bierne et al., 2002; Lalioui et al., 2005; Wang et al., 2009). The studies also showed that inactivation of sortase does not cause any significant alterations in growth in lactobacilli (Malik et al., 2013). The deletion of *Streptococcus*

pyogenes sortase A, on the other hand, resulted in the accumulation of sorting intermediates, particularly at the septum, altering cellular morphology and physiology and impeding membrane integrity. *Actinomyces oris* is a notable exception, where sortase deletion is lethal due to membrane accumulation of a surface glycoprotein (Wu et al., 2014).

The overexpressed strain of *C. glutamicum* was found to be thermal sensitive with the increase in temperature when compared to mutant cells. The heat-induced stress causes changes in the physical characteristics of live cell membranes. When lipids become more fluid as temperature rises above the organism tolerance limit, this can result in increased membrane permeability, disorder of the lipid/protein membrane order, and finally loss of membrane integrity and cell death (Chapman, 1975; Los and Murata, 2004). The *srtE* gene on overexpression resulted an elongated rod-shaped cell when compared to wild-type and mutant cells. Sortase A overexpression has a dominant negative effect on *L. monocytogenes* invasion, most likely by anchoring an excess of surface proteins that mask other surface factors required for invasion (Garandeau et al., 2004).

The significant reduction of cell surface hydrophobicity in *srtE* mutant cells correlates with the report of *srtA* in *L. gasseri* Kx110A1. The *srtA* inactivation dramatically reduced the cell surface hydrophobicity, indicating the contribution of cell wall associated mechanism of Sortase A dependent proteins (Zuo et al., 2019). The deficiency of *srtA* in *S. sanguinis* causes an overall reduction in virulence in association with cell surface proteins and decreased cell surface hydrophobicity (Yamaguchi et al., 2006).

4.5. Summary

This chapter discusses the variants of *C. glutamicum* mutants, such as the Δ*srtE* deletion mutant created by double-crossover homologous recombination, the sortase complemented strain (Δ*srtE/srtE*) created by inserting a native *srtE* expression cassette into

plasmid pVWEx1 after transformation into the Δ*srtE* mutant, and the overexpression of *srtE* gene using the pVWEx1 shuttle vector. The absence of the s*rtE* gene in *C. glutamicum* did not result in changes in growth rate or morphology, but it did result in a decrease in cell surface hydrophobicity, which might be attributed to the changes in the sortase-dependent protein build-up in the cell membrane. By inserting a *srtE* expression vector, the complement was able to restore its activity to that of WT. Interestingly, p*srtE* overexpressed strains grew slower in the log phase than the wild-type strain, requiring a longer incubation time to reach the stationary phase and resulting in an elongated cell morphology when compared to the parental strain, implying that excessive *srtE* during this phase may interfere with the appropriate expression of other genes. When the incubation temperature was raised from 30 °C to 37 °C, the number of viable cells at 10^{-7} dilution was reduced by 20 % in deletion mutant compared to the parental strain. However, the overexpressed strain did not show any growth when the temperature shifted from 30 °C to 37 °C.

Chapter 5

Molecular cloning, expression and biochemical characterization of Sortase E of *C. glutamicum*

Abbreviation

$CgSrtE_{\Delta N44}$	*C. glutamicum* truncated sortase (N-terminal 44 amino acids removed)
DTT	Dithiothreitol
ESI-MS	Electrospray ionization mass spectrometry
MALDI-TOF/MS	Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometer
ORF	Open Reading Frame

5.1. Introduction

The Class E sortases are widely distributed in GC-rich Actinobacteria, especially in *Corynebacterium* and *Streptomyces* sp. Similar to class A sortase, class E sortase anchors distinct surface proteins on the bacterial cell wall. Sortase E recognizes a novel LAXTG sorting motif in which the conserved proline residue is replaced with alanine (underlined). Class E sortases share just a small amount of basic sequence homology with other sortases and are expected to play a general housekeeping role. In addition, genes producing class A and E sortases are never found together in the same organism, and, similarly to class A sortases, genes encoding class E enzymes are not located near genes encoding possible protein substrates (Comfort and Clubb, 2004). Two class E sortases of *Streptomyces* sp. have been examined in more depth so far, and their structures have recently been discovered. The genome of *S. coelicolor* contains two types of class E sortase, SrtE1 and SrtE2 which anchors Chaplin proteins promoting in aerial hyphae formation. The genome of *S. avermitilis* contains four putative SrtE enzymes of which SrtE3 has been enzymatically and structurally characterized in *E. coli* for in vitro studies and found to be Ca^{2+}- independent unlike SrtA

enzyme. *C. diphtheriae* contains a single class E sortase (named as Cd-SrtF) which helps in anchoring assembled pili to the cell wall peptidoglycan (Chang et al., 2011; Swaminathan et al., 2007). Similarly, based on the in-silico analysis we could able to identify a single class E sortase in *C. glutamicum* that recognizes a LAXTG substrate motif.

However, in the literature reports there are no experimental records that support the sortase activity and substrate specificity of sortase E of *C. glutamicum*, since Sortase E is a membrane-bound enzyme, isolation and characterization of the enzyme in *C. glutamicum* has proven challenging and time-consuming. As a result, planned to express and purify the enzyme in *E. coli*, to investigate the in vitro catalytic activity and substrate selectivity. The detailed characterization of the recombinant enzyme was also carried out to bring out the salient features which enabled the enzyme to be used in various sortagging applications.

5.2. Materials and Methods

5.2.1. Cloning of *srtE* gene into pET28a vector

The genomic DNA of *C. glutamicum* was isolated as previously described **(Section 2.2.3.1)**. The *srtE* from *C. glutamicum* ATCC 13032 is annotated in database records as GenBank: BA000036, Uniprot: Q8NLK3 with an appropriate ORF. A full-length *srtE* gene construct was generated by PCR amplification mentioned in **Table 2.5 (Chapter 2)** from *C. glutamicum* using NCgl2838 F/R primers specified in **Table 2.3 (Chapter 2)**. The PCR product was purified and double digested at the NheI/NotI (NEB, Japan) restriction sites in the pET28a vector (Novagen, U.S.A.). The digested PCR product and pET28a vector were ligated together with T4 DNA ligase (NEB, Japan) by an overnight reaction at 16 °C, transformed into *E. coli* DH5 alpha cells and selected on Luria Bertani (LB)-kanamycin (50 μgmL^{-1}) plates. The positive clones were further confirmed by DNA sequencing and

transformed into *E. coli* BL21 (DE3) expression host strain to generate an N-terminally His_6-tagged recombinant protein designated as pSrtE.

5.2.2. Expression and purification of full-length *C. glutamicum* (CgSrtE)

E. coli BL21 (DE3) harboring pSrtE was allowed to grow in LB medium supplemented with kanamycin at 37 °C with continuous shaking until OD_{600} reaches 0.8. At this point, 1 mM IPTG was used to induce gene expression for 3.5 h at 37 °C. The cells were harvested by centrifugation at 3500 x g for 30 min at 4 °C. Cell pellets were resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 0.1 % Tween 20, 2 mM β-mercaptoethanol, and 20 mM imidazole) and sonicated (Sonics Vibra cell, USA) before centrifugation at 20,000 x g for 30 min at 4 °C. SDS-PAGE was used to examine the isolated protein from the soluble and insoluble fractions.

5.2.3. Synthesis of $CgSrtE_{\Delta N44}$ by DNA manipulation

The $CgSrtE_{\Delta N44}$ residues containing Ala45-Asn274 amino acids were amplified using the primers CgSrtE F/R primers specified in **Table 2.3 (Chapter 2)**, gel-purified and double digested with NdeI/SalI restriction sites. To increase the solubility of the protein, the N-terminal transmembrane regions containing 1-44 amino acids were deleted and the truncated gene was subsequently cloned into the pET28a vector to yield pTSrtE. The construct was verified by DNA sequencing and transformed into *E. coli* BL21 (DE3) cells for protein expression.

5.2.4. Expression and purification of recombinant $CgSrtE_{\Delta N44}$

E. coli BL21 (DE3) harboring the pTSrtE was grown in 1 L of terrific broth (TB) medium **(Annexure-I)** supplemented with kanamycin (50 $\mu g mL^{-1}$) at 37 °C with continuous shaking. When the OD_{600} reached 0.6-0.8, 1 mM IPTG was used to induce $CgSrtE_{\Delta N44}$ expression and incubated further for 3.5 h. The cells were harvested by centrifugation at 3,500

x g for 30 min at 4 °C. Cell pellets were resuspended in lysis buffer containing 50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 0.1 % Tween 20, 2 mM β-mercaptoethanol, and 20 mM imidazole, lysed by sonication (Sonics Vibra cell, USA), and centrifuged at 20,000 x g for 30 min at 4 °C. After loading the lysate onto a HisTrap HP 5-mL column (GE Healthcare, USA), it was rinsed with lysis buffer and wash buffer (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 0.1 % Tween 20, 2 mM β-mercaptoethanol, and 60 mM imidazole), and the protein was eluted with 200 mM imidazole. The excess imidazole was freed out using a PD-10 desalting column. The purified protein was stored in a buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 2 mM β-mercaptoethanol and concentrated using 10 kDa Amicon Ultra centrifugal filters (Millipore). The protein sample was quantified using Bradford reagent with an estimated concentration of ~1.3 $mgmL^{-1}$ and using bovine serum albumin (BSA) as the standard **(Section 2.3.2)**. The purity of the protein was detected using 15 % SDS-PAGE for two independent protein preparation with different concentrations of 20 µg and 40 µg. The His-tagged recombinant protein was confirmed by Western blotting using anti-His antibodies **(Section 2.3.4)**. The molecular weight of the proteins was further analyzed by MALDI-TOF/MS using an Ultraflex TOF/TOF instrument.

5.2.5. Site-directed mutagenesis

Four active site mutations were generated based on multiple sequence alignment and secondary structure analysis of *C. glutamicum* class E sortase. Plasmid pTSrtE was used as the template for the introduction of single amino acid substitution generated by PCR using a Q5 site-directed mutagenesis kit (NEB) **(Section 2.4.5 in Chapter 2)**. The amino acid residues, C240, H135, R249, and Y118 were mutated to encode Ala at these positions using appropriate mutagenic primers shown in **Table 2.3 (Chapter 2)**, designed by using primer designing program from NEBase Changer™ primer design software. After PCR, the mutant plasmids were transformed into *E. coli* DH5α and the transformants were selected on

kanamycin-LB agar plates. All the constructed mutants C240A, Y118A, H135A, and R249A were subsequently detected and confirmed by DNA sequencing and positive mutants were further purified as described in **Section 5.2.4.**

5.2.6. Protein-peptide docking

Protein-peptide docking was carried out using ClusPro 2.0 server (Kozakov et al., 2017) between $CgSrtE_{\Delta N44}$ and LAXTG peptide where X is replaced with three amino acids, H, E, and A. The docking results were visualized by using PyMOL to confirm and understand the binding energy of three peptides with the $CgSrtE_{\Delta N44}$ enzyme. The ClusPro server performs three steps in the computation process: (i) rigid-body docking utilizing the fast Fourier transform (FFT) correlation technique, (ii) RMSD-based grouping of the structures developed to locate the largest cluster that would represent the most likely models of the complex, and (iii) refining of selected structures.

5.2.7. FRET analysis

Peptide substrates such as Abz-LAHTG-Dap (Dnp), Abz-LAETG-Dap (Dnp), and Abz-LPETG-Dap (Dnp) were synthesized by Shanghai GL Biochem, China. The peptide substrates were tagged with a 2-aminobenzoyl (Abz) fluorophore at the N-terminus and 2, 4-dinitrophenyl (Dnp) as the quencher at the C-terminus. The assay was performed in 100 µL reaction volume in a 96-well black microtiter plate (Thermofisher) containing 50 µM of each peptide substrate Abz-LAHTG-Dap (Dnp), Abz-LAETG-Dap (Dnp), and Abz-LPETG-Dap (Dnp) individually in a cleavage buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT and 5 mM $CaCl_2$). The reaction was monitored by the addition of 5 µM of purified $CgSrtE_{\Delta N44.}$ The increase in the fluorescence intensity was measured at 37 °C for 6 h using an Infinite M200 PRO multimode microplate reader (Tecan, Switzerland) at an excitation wavelength of 320 nm and an emission wavelength of 420 nm. Fluorescence was measured by

subtracting the fluorescence of the peptide from the overall reaction fluorescence. Each reaction was performed in triplicate and averaged, and plotted as arbitrary fluorescence units.

5.2.8. Kinetic measurements

To understand the basic enzyme kinetics of $CgSrtE_{\Delta N44}$ with Abz-LAHTG-Dap (Dnp), the sortase assay was set up to monitor the increase in fluorescence with respect to time. The Abz-LAHTG-Dap (Dnp) was dissolved in 50 % dimethyl sulfoxide (DMSO) and incubated at a concentration between 2.5–50 µM with a constant 5 µM enzyme ($CgSrtE_{\Delta N44}$) concentration. The same peptide concentration without adding enzyme for the reaction served as the control. Assay mixture consisted of 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT and 5mM $CaCl_2$ and incubated at 37 °C. The fluorescence intensity was monitored for 6 h at an excitation wavelength of 320 nm and an emission wavelength of 420 nm at regular intervals of 10 min. The fluorescence intensity (V_0) was plotted against substrate concentration [S]. The constants K_m and V_{max} values were determined from the slopes of various concentrations of substrates by applying nonlinear curve fit of Michaelis-Menton equation;

$$V=V_{max}[S]/K_m+[S]$$

The kinetic analysis was performed using GraphPad Prism version 5.0 (GraphPad software).

5.2.9. Biochemical characterization of $CgSrtE_{\Delta N44}$

5.2.9.1. Effect of pH

The pH stability of $CgSrtE_{\Delta N44}$ was carried out in a 100 µL reaction system with appropriate buffers. Na-acetate buffers were used for pH ≤ 5.2, while for pH > 6, Tris-HCI buffers were used with 50 mM concentration along with 5 µM enzyme and reaction mixtures (40 µM Abz-LAHTG-Dap (Dnp), 150 mM NaCl, 1 mM DTT, and 5 mM $CaCl_2$) were incubated at 37 °C for 6 h.

5.2.9.2. Effect of temperature

The optimum temperature was determined by incubating the enzyme reaction mixtures at different temperatures 25, 30, 37, 45, 60, and 70 along with 5 µM $CgSrtE_{\Delta N44}$ enzyme and reaction mixtures (50 mM Tris-HCl (pH 7.5), 40 µM Abz-LAHTG-Dap (Dnp), 150 mM NaCl, 1 mM DTT and 5 mM $CaCl_2$) were incubated for 6 h.

5.2.9.3. Effect of metal ions

The effects of metal ions were determined by incubating 5 µM $CgSrtE_{\Delta N44}$ with 5 mM metal ions (Ca^{2+}, K^{+}, Mg^{2+}, Mn^{2+} and in the absence of Ca^{2+}) in a reaction mixture containing 50 mM Tris-HCl (pH 7.5), 40 µM Abz-LAHTG-Dap (Dnp), 150 mM NaCl and 1 mM DTT incubated at 37 °C for 6 h.

5.2.9.4. Effect of enzyme concentration

The optimum enzyme concentration was checked with a wide range of $CgSrtE_{\Delta N44}$ enzyme concentration (1 µM, 2.5 µM, 5 µM, 7.5 µM, 10 µM, 15 µM, 25 µM, and 35 µM) using 40 µM Abz-LAHTG-Dap (Dnp) substrate in a buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 5 mM $CaCl_2$ at 37 °C for 6 h.

5.2.10. In vitro cleavage assay for sortase activity

For the enzyme activity confirmation, sortase cleavage action was also monitored by HPLC (Kruger et al., 2004). The reaction mixture contains 40 µM Abz-LAHTG-Dap (Dnp), 50 mM Tris-HCl, 150 mM NaCl, 1 mM DTT and 5 mM $CaCl_2$. The reaction was initiated with an addition of 5 µM $CgSrtE_{\Delta N44}$ and incubated at 37 °C for 6 h. The reaction was further quenched by adding 10-fold excess 0.1 % TFA and injected onto a Vydac reversed-phase C18 RP-HPLC column (4.6 X 50 mm, 3 µm). A linear gradient of 10 to 40 % (Acetonitrile/0.1 % TFA) for 10 min at a flow rate of 3 $mLmin^{-1}$ was used for the separation of peptides. The Dnp containing peaks were observed at 355 nm UV absorbance.

5.2.11. Detection of G-Dap (Dnp) product

The sortase cleavage product identity was determined using electrospray ionization mass spectrometry (ESI-MS) on an Agilent 1100 MSD Trap SL mass spectrometer in negative ion mode. The individual product peak G-Dap (Dnp) from HPLC assay was collected, pooled, concentrated using a vacuum concentrator. The dried, concentrated products were dissolved in 500 µL of MillQ/0.1 % TFA and subjected to ESI-MS to confirm the identity.

5.3. Results

5.3.1. PCR amplification and cloning of full-length *srtE* gene

The *srtE* gene was amplified from the genomic DNA of *C. glutamicum* using gene-specific primers **(Table 2.3 in Chapter 2)** and *Taq* DNA polymerase at an annealing temperature of 60 °C. The optimized PCR condition of the *srtE* gene was shown in **Table 2.5 (Chapter 2)**. The amplicon showed an expected size of 825 bp on 1.5 % agarose gel **(Figure 5.1)**.

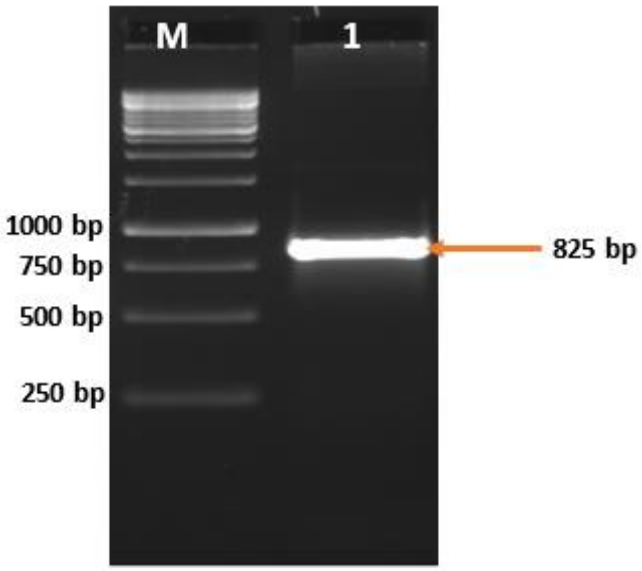


Figure 5.1. PCR amplification of *srtE* gene

Lane: 1 kb DNA ladder; **Lane 2:** *srtE* PCR amplicon

The amplified *srtE* gene was incorporated with Nhe1/NotI restriction sites and ligated into pET28a which was previously digested with the same restriction sites. The recombinant plasmid was confirmed by double digestion with Nhe1/NotI restriction sites **(Figure 5.2).** Clones were sequenced by T7 primers, sequences were verified and the ORF analysis proved that the amplicon was completed without any kind of mutation.

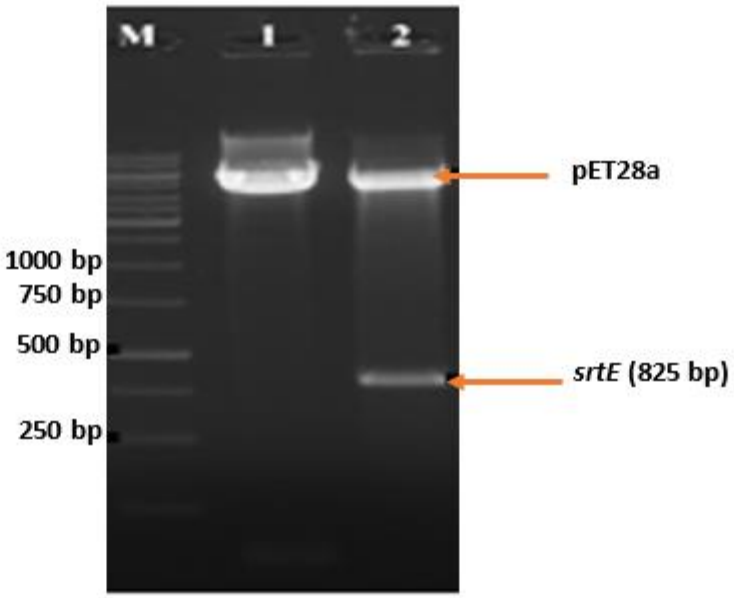


Figure 5.2. Restriction digestion of recombinant plasmid pSrtE

Lane M: 1 kb DNA ladder; **Lane 1:** Native plasmid pET28a Nhe1/NotI double digest; **Lane 2:** pET-28a-*srtE* Nhe1/NotI double digest showing insert release at 825 bp.

5.3.2. Expression of full-length Sortase E (CgSrtE) in *E. coli*

E. coli BL21 (DE3) cells carrying pET28a-*srtE* designated as pSrtE, was chosen for expressing full-length sortase designated as CgSrtE, under the control of IPTG inducible T7 promoter **(Figure 5.3)**.

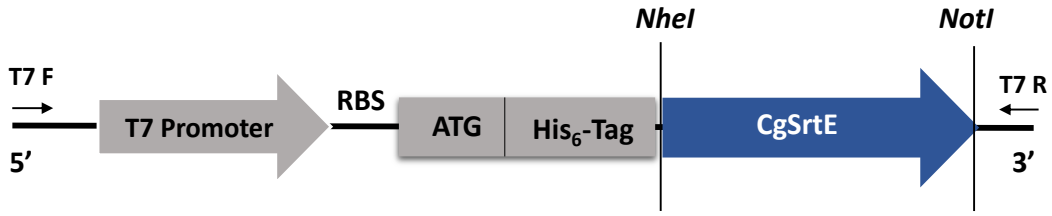


Figure 5.3. Schematic diagram of pSrtE construct

Protein expression from pET28a is under the control of IPTG inducible promoter, and results in an N-terminally His_6-tagged recombinant protein with an expected mass of 32.31 kDa.

The bacterial culture was grown at 37 °C in LB kanamycin medium until the culture density reached an absorbance of 0.8 at OD_{600}. The cultures were induced with 1 mM IPTG for 3.5 h. Then, the cells were harvested by centrifugation and disrupted by sonication as described in **Section 2.3.1 (Chapter 2)**. Both soluble and insoluble fractions were loaded along with crude protein. All the proteins were resolved on 12 % SDS-PAGE, but our protein of interest with a MW of 32.31 kDa (including N-terminal His_6 tag) was able to detect on SDS-PAGE gel in the inclusion bodies **(Figure 5.4)** which was found to be difficult to carry out for the further downstream process.

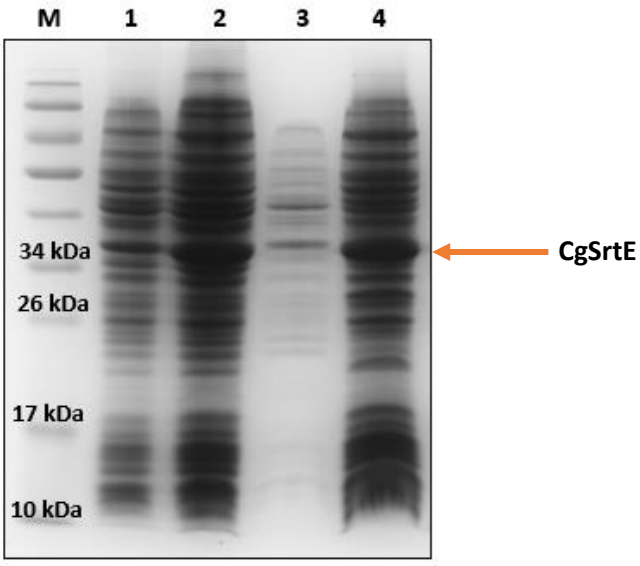


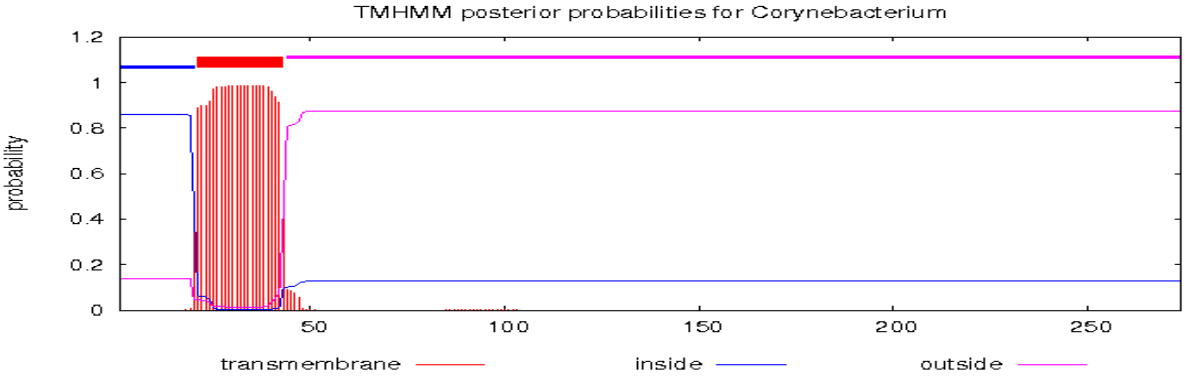
Figure 5.4. SDS-PAGE showing the expression of pSrtE in *E. coli* BL21 (DE3) cells

Lane M: Molecular weight marker; **Lane 1:** Induced pET28a; **Lane 2:** Induced putative sortase; **Lane 3:** Soluble fraction; **Lane 4:** Insoluble fraction

5.3.3. Synthesis of truncated *srtE* gene

The 1-44 amino acids coding the N-terminal transmembrane region were removed from the sortase protein, predicted by the TMHMM server **(Figure 5.5)**. The primers were designed after excluding the signal peptide and transmembrane encoding regions. The *srtE* gene was amplified from the genomic DNA of *C. glutamicum* using gene-specific primers and *Taq* DNA polymerase at an annealing temperature of 60 °C. The optimized PCR conditions of the truncated *srtE* gene was shown in **Table 2.3 (Chapter 2).** The amplicon showed an expected size of 693 bp on 1.5 % agarose gel **(Figure 5.6)**.

```
# Corynebacterium Length: 274
# Corynebacterium Number of predicted TMHs:  1
# Corynebacterium Exp number of AAs in TMHs: 22.47588
# Corynebacterium Exp number, first 60 AAs:  22.4612
# Corynebacterium Total prob of N-in:        0.86054
# Corynebacterium POSSIBLE N-term signal sequence
Corynebacterium TMHMM2.0        inside       1    20
Corynebacterium TMHMM2.0        TMhelix     21    43
Corynebacterium TMHMM2.0        outside     44   274
```




```
        10         20         30         40         50
MTATLSAESS RNGKKPRPRV SVSQVVGEIL LTVGILALLF AYYEAYWTNV
        60         70         80         90        100
ESGKLQESAG QKLDEDWNEA RVNPRQKLTP ELGEAFARMY VPAFGSDFNF
       110        120        130        140        150
AVIEGTDEED LLAGPGRYVD SQMPGEAGNF AVAGHRVGKG APFNDLGNLE
       160        170        180        190        200
VCDAIVVETY NSWDVYRVMP MSTNGADRAA EAADCFNENQ VSRMAEGDYV
       210        220        230        240        250
NVSGRSITTP DRIDATYPTP GVFDTAVREG SEALLTLTTC HPQFSNAERM
       260        270
IVHAMLVEEI DKSSGERPAA LEEN
```

Figure 5.5. Transmembrane region of sortase E in *C. glutamicum*

The transmembrane region was analyzed using the TMHMM server, where 21-43 amino acid sequence contains the transmembrane region (red).

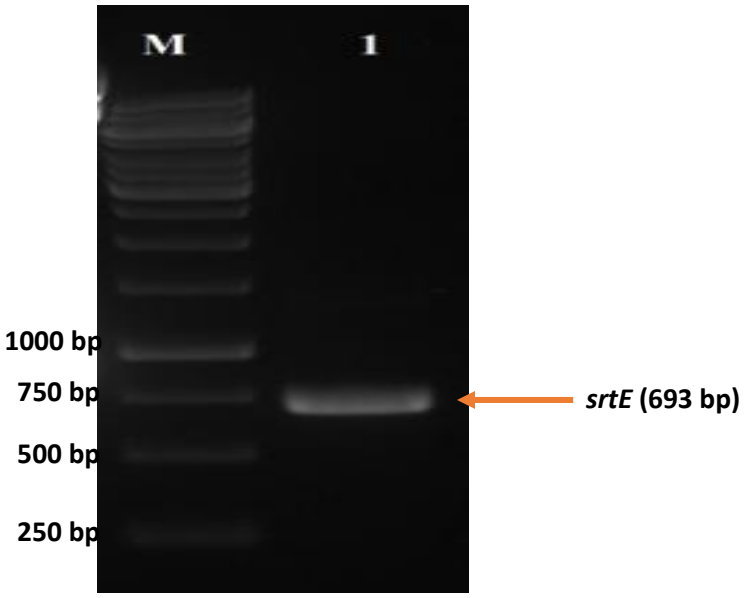


Figure 5.6. PCR amplification of truncated gene

Lane: 1 kb DNA ladder; **Lane 2:** *srtE* PCR amplicon

The gene was incorporated with restriction sites NdeI/SalI and cloned to pET28a vector for expression of protein with N-terminal His-tag and the plasmid was designated as pTSrtE. The pTSrtE plasmid was confirmed by releasing the insert by double digestion with restriction enzymes NdeI/SalI **(Figure 5.7)**. Clones were sequenced by T7 primers, sequences were verified and the ORF analysis proved that the amplicon was completed without any kind of mutation.

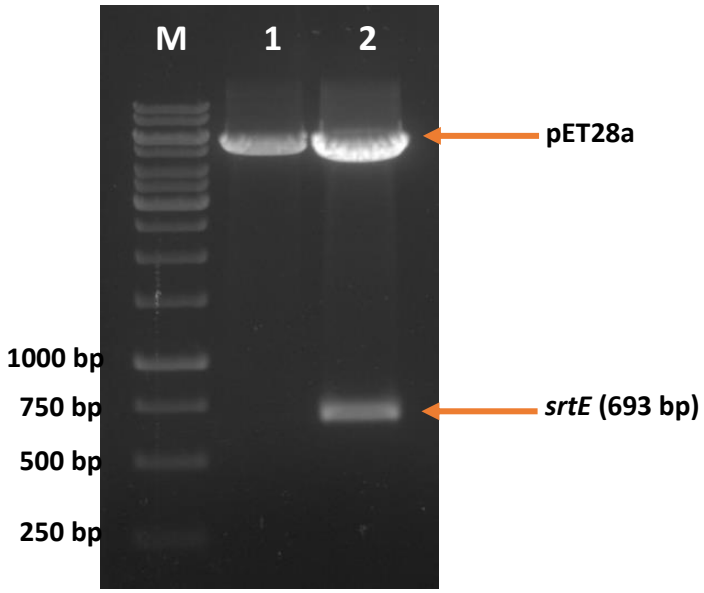


Figure 5.7. Restriction digestion of recombinant plasmid pTSrtE

Lane M: 1 kb DNA ladder; **Lane1:** Native plasmid pET28a NdeI/SalI double digest; **Lane 2:** pET-28a-*srtE* NdeI/SalI double digest showing insert release at 693 bp.

5.3.4. Expression and purification of CgSrtE$_{\Delta N44}$ in *E. coli*

The solubility of the recombinant *C. glutamicum* sortase enzyme was generally improved following the removal of N-terminal transmembrane segment of CgSrtE which includes 2-44 amino acids as transmembrane region, so 45-274 residues were amplified from the full-length gene, expressed in *E. coli* BL21 (DE3) with an N-terminal His$_6$-tag and the expressed protein was designated as CgSrtE$_{\Delta N44}$ **(Figure 5.8)**.

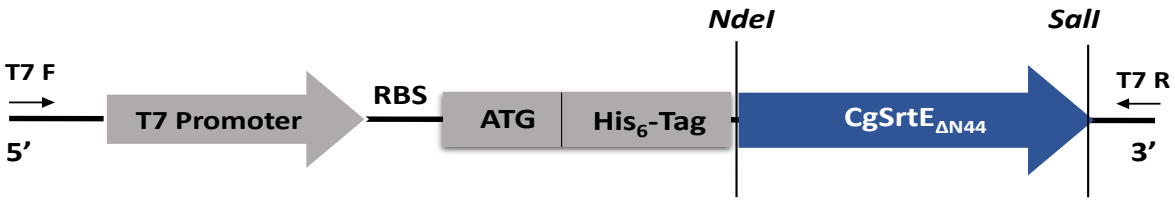


Figure 5.8. Schematic diagram of pTSrtE construct

Protein expression from pET28a is under the control of IPTG inducible promoter, and results in an N-terminally His$_6$-tagged recombinant protein with an expected mass of 27.4 kDa.

The N-terminal His_6-tagged protein was purified using the HisTrap HP 5-mL column. The expression, purity, and homogeneity of $CgSrtE_{\Delta N44}$ were analyzed and confirmed through SDS-PAGE analysis. All the proteins were resolved on 15 % SDS-PAGE, but our protein of interest with a MW of 27.4 kDa (including N-terminal His_6 tag) was able to detect on SDS-PAGE gel in the soluble fraction and able to purify at 200 mM imidazole **(Figure 5.9 A & B)** and its recombinant nature was further confirmed by western blotting with anti-polyhistidine antibody **(Figure 5.9 C)**. MALDI-TOF MS analyses revealed the presence of two peptide fragments, showing identity with the sortase of *C. glutamicum* ATCC 13032 **(Figure 5.10)**.

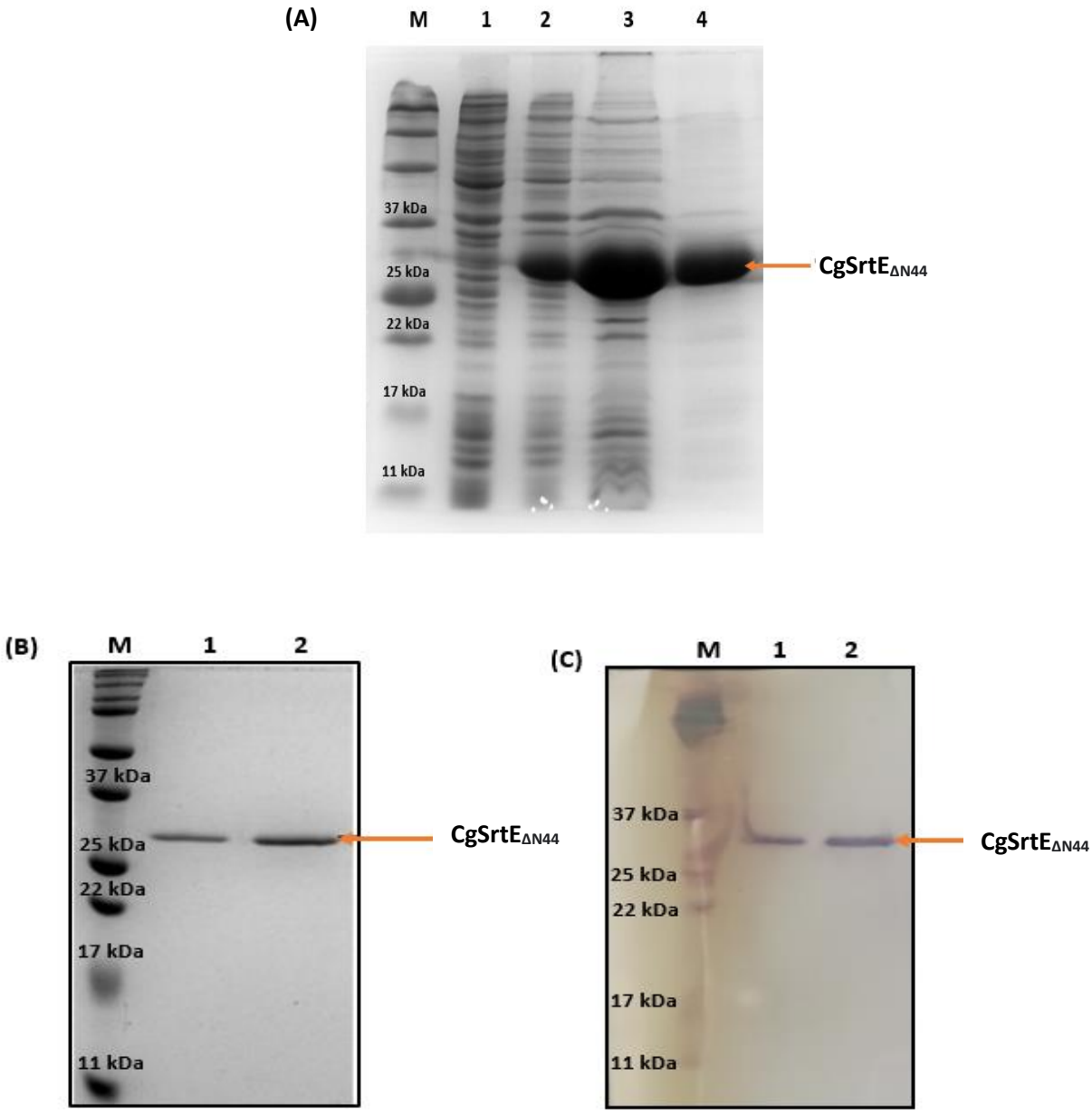


Figure 5.9. Purification and confirmation of CgSrtE$_{\Delta N44}$

(A) Lane M: Molecular protein marker**; Lane 1:** Uninduced crude protein; **Lane 2:** Induced crude protein; **Lane 3:** Soluble fraction; **Lane 4:** Purified protein at 200 mM imidazole gradient; **(B)** Expression and SDS-PAGE analysis of purified recombinant CgSrtE$_{\Delta N44}$ with an N-terminal His tag comprising 45-274 amino acids; **Lane M:** Molecular weight marker; **Lane 1:** 20 µg CgSrtE$_{\Delta N44}$; **Lane 2:** 40 µg CgSrtE$_{\Delta N44}$; **(C)** Western blotting; **Lane 1:** Prestained protein marker; **Lane 1 & 2:** CgSrtE$_{\Delta N44}$ detected on western blot.

Peak Name	Start	End	Observed	Mr (expt)	Mr (calc)	Peptide sequence
P1	55	71	1988.856	1987.948	1987.924	**K. LQESAGQKLDEDWNEAR.V**
P2	213	228	1722.785	1721.777	1721.863	**R. IDATYPTPGVFDTAVR. E**

```
  1   MTATLSAESS RNGKKPRPRV SVSQVFGEIL LTVGILALLF AYYEAYWTNV
 51   ESGKLQESAG QKLDEDWNEA RVNPRQKLTP ELGEAFARMY VPAFGSDFNF
101   AVIEGTAEED LLAGPGRYVD SQMPGEAGNF AVAGHRVGKG APFNDLGNLE
151   VCDAIVVETY NSWDVYRVMP MSTNGADRAA EAADCFNETQ VSRMAEGDYV
201   NVSGRSITTP DRIDATYPTP GVFDTAVREG SEALLTLTTC HPQFSNAERM
251   IVHAMLVEEI DKSSGERPAA LEEN
```

Figure 5.10. Mass spectrometric analysis of purified CgSrtE$_{\Delta N44}$

Matched sortase sequence (red) identified in *C. glutamicum* through MALDI-TOF MS analysis.

5.3.5. Docking of CgSrtE$_{\Delta N44}$ and substrate (LAXTG) peptides by ClusPro

The generated structure of CgSrtE by homology modeling **(Section 3.3.8 in Chapter 3)** was found to be similar even after removing 44 amino acids at the N-terminal of the protein. The molecular docking CgSrtE$_{\Delta N44}$ with the peptide substrates (LAHTG, LAATG, and LAETG) were performed by ClusPro to discover and synthesize the best and least affinity substrates for in vitro sortase activity. Our results showed that the docking of the enzyme with LAHTG substrate displayed the lowest binding energy when compared to the other two substrates **(Figure 5.11)**. Thus, the enzyme showed a strong binding affinity with LAHTG and the least affinity with LAETG peptide as shown in **Table 5.1**.

Table 5.1. ClusPro results for Sortase E protein and LAXTG peptide sequence

Peptide	Energy			
	Total-ACE	Interaction-ACE	Total vacuum	Interaction vacuum
LAATG	-2747.86567	-5.97200	-159.90818	-53.82283
LAHTG	-2747.09073	-6.42418	-170.44357	-6828139
LAETG	-2758.47843	-3.37608	-171.40988	-68.74154

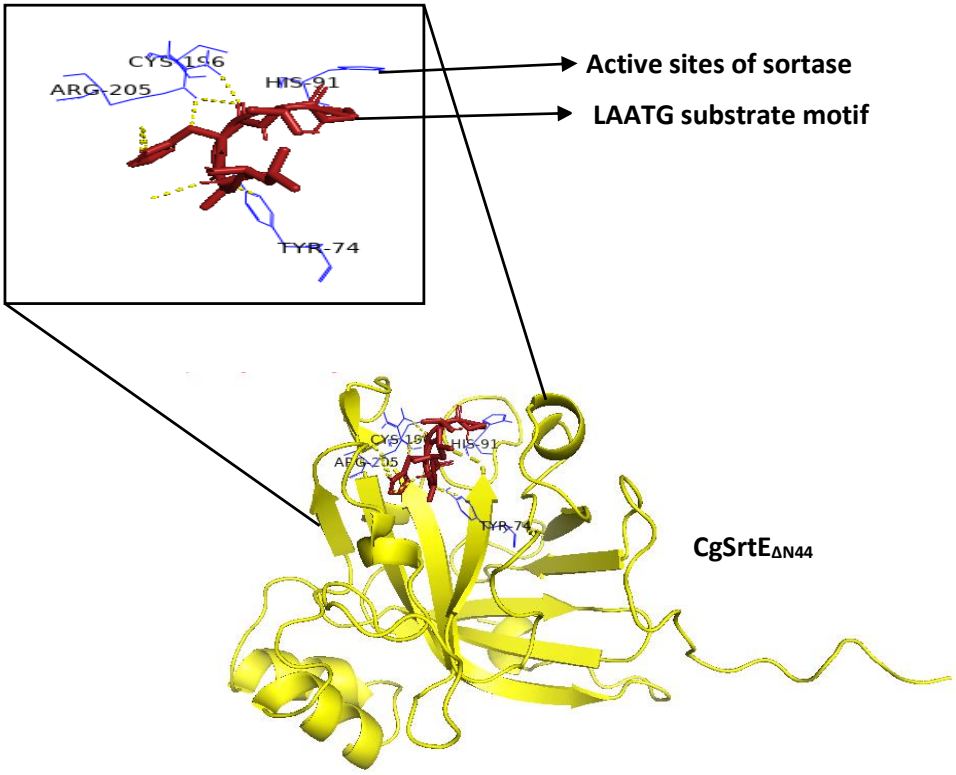


Figure 5.11. CgSrtE$_{\Delta N44}$-LAHTG molecular docking

CgSrtE$_{\Delta N44}$ active sites Cys, Arg, and His binding to the LAHTG peptide. The predicted binding site region LAHTG showed with red color. The CgSrtE$_{\Delta N44}$ (yellow color) active site regions are shown as blue color.

5.3.6. In vitro sortase activity by FRET

To confirm the sortase activity and substrate specificity of CgSrtE$_{\Delta N44}$, three peptides that encompass the sortase sorting motifs were fluorescently labeled with a 2-aminobenzoyl (Abz) fluorophore at the N-terminus and dinitrophenyl (Dnp) quencher group at the C-

terminus were tested: Abz-LAHTG-Dap (Dnp), Abz-LAETG-Dap (Dnp) and Abz-LPETG-Dap (Dnp)**.** The fluorescence signals get reduced when the peptides are nearby due to the quencher in the peptides, while the fluorescence signal gets enhanced when the peptide is cleaved by the enzyme, separating the fluorophore and quencher apart **(Figure 5.12)**.

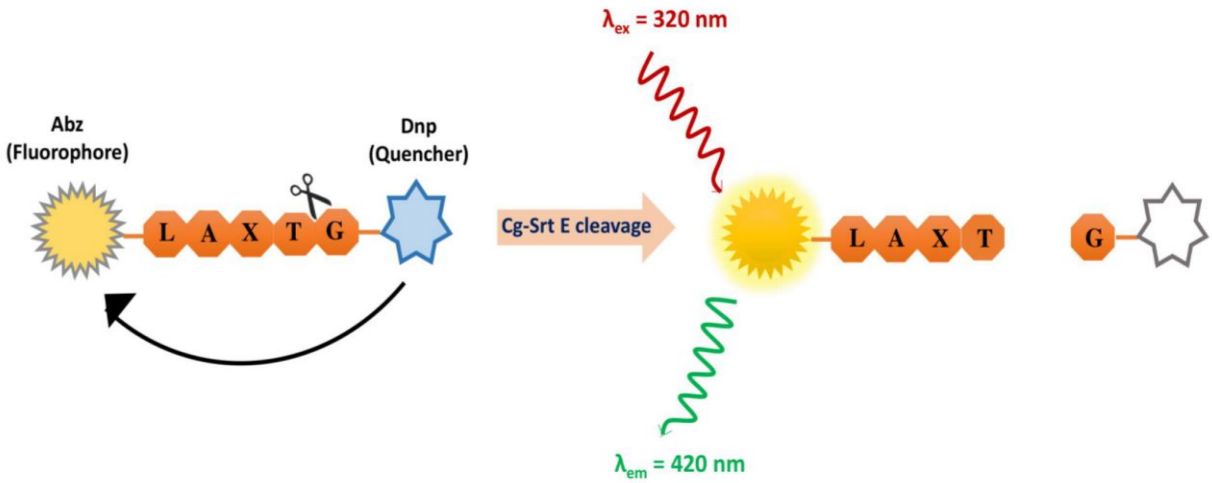


Figure 5.12. Enzyme activity of CgSrtE$_{\Delta N44}$ by using a FRET-based cleavage assay

Schematic representation of Abz-LAXTG-Dap (Dnp) cleavage by CgSrtE$_{\Delta N44}$. The fluorophore (Abz) and quencher (Dnp) are sandwiched between the peptide substrate. The fluorescence is measured when the sortase E-mediated reaction separates the fluorophore from the quencher.

The peptides Abz-LAHTG-Dap (Dnp) and Abz-LAETG-Dap (Dnp) were synthesized based on the reports available on a substrate of class E sortases such as SrtE1 and SrtE2 substrates of *S. coelicolor* (Duong et al., 2012) and by protein-peptide docking analysis. The motif LPXTG is the substrate for class A sortase of *S. aureus* (Mazmanian et al., 1999), hence Abz-LPETG-Dap (Dnp) serves as a negative control for the assay. We observed that the CgSrtE$_{\Delta N44}$ were able to cleave both Abz-LAHTG-Dap (Dnp) and Abz-LAETG-Dap (Dnp) peptides within 6 h of incubation **(Figure 5.13)**. However, as expected, the enzyme failed to cleave the Abz-LPETG-Dap (Dnp), the substrate of SrtA.

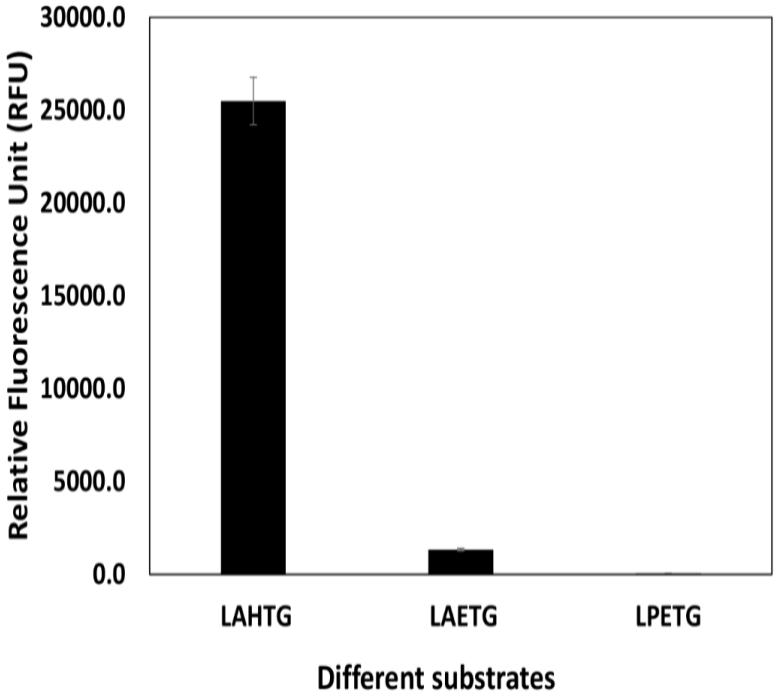


Figure 5.13. Substrate specificity of CgSrtE$_{\Delta N44}$ by FRET-based cleavage assay

Purified recombinant CgSrtE$_{\Delta N44}$ incubated with recognized substrate motifs Abz-LAHTG-Dap (Dnp), Abz-LAETG-Dap (Dnp), and Abz-LPETG-Dap (Dnp) to investigate the substrate specificity. The enzyme failed to cleave the LPETG motif recognized by *S. aureus* sortase A. The results shown here is an average of three independent enzyme assays each done in triplicates.

5.3.7. Kinetic studies by FRET analysis

To determine the kinetic parameters of CgSrtE$_{\Delta N44}$ with Abz-LAHTG-Dap (Dnp) and Abz-LAETG-Dap (Dnp) peptides, kinetic analysis of sortase-catalyzed transpeptidation reaction was performed. Varying concentrations (2.5, 5, 7.5, 10, 15, 20, 25, 30, 35, 40, 45, 50 µM) of each peptide were incubated with 5 µM CgSrtE$_{\Delta N44}$ and the reaction was monitored in every 10 min interval for a period of 6 h. The level of cleavage observed for the Abz-LAETG-Dap (Dnp) peptide was too low to facilitate the kinetic analysis. On the other hand, initial velocities (V_0) obtained from the progress curves were able to plot against the varying concentration of Abz-LAHTG-Dap (Dnp). Thus, with Abz-LAHTG-Dap (Dnp) substrate,

calculated an apparent K_m of 12 ± 1 µM and an apparent V_{max} of 1.3 ± 0.04 RFU/sec for CgSrtE$_{\Delta N44}$ **(Figure 5.14)**.

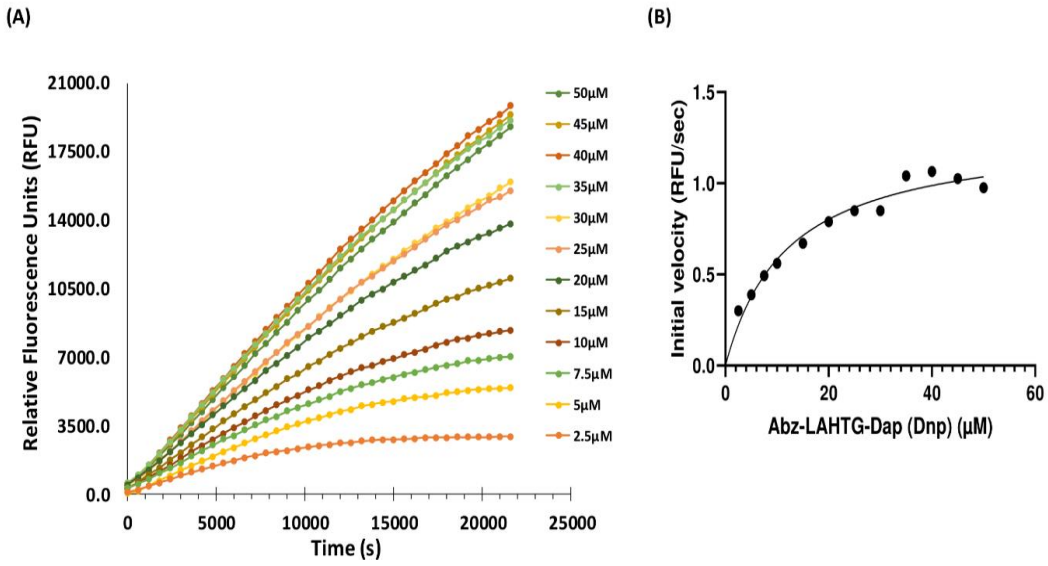


Figure 5.14. FRET assay with Abz-LAHTG-Dap (Dnp)

Progress curves were obtained from the cleavage reaction of Abz-LAHTG-Dap (Dnp) fluorescent peptide catalyzed by recombinant CgSrtE$_{\Delta N44}$. Reactions containing 5 µM of CgSrtE$_{\Delta N44}$ enzyme incubated with 2.5 to 50 µM of fluorescent peptide at 37 °C in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM $CaCl_2$ and 1 mM DTT. **(B)** The kinetic parameters, K_m of 12 ± 1 µM and V_{max} of 1.3 ± 0.04 RFU/sec were determined for CgSrtE$_{\Delta N44}$ with Abz-LAHTG-Dap (Dnp).

5.3.8. Biochemical evaluation of CgSrtE$_{\Delta N44}$

The optimum temperature for CgSrtE$_{\Delta N44}$ activity was determined in a reaction mixture with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT, 5 mM $CaCl_2$ and 40 µM Abz-LAHTG-Dap (Dnp) as the substrate. The increase in fluorescence associated with the catalytic efficiency of the enzyme at different temperatures. Recombinant CgSrtE$_{\Delta N44}$ is less efficient with Abz-LAHTG-Dap (Dnp) substrate when incubated at 25 °C and 30 °C. However, the transpeptidase activity of the enzyme can be improved by incubating the CgSrtE$_{\Delta N44}$ at higher temperatures with the maximal catalytic efficiency being observed at 60 °C **(Figure 5.15 A)**. Among the different concentrations tried, optimum enzyme

concentration was found to be 15 µM **(Figure 5.15 B)**. Similarly, sortase activity was examined over a pH range of 3.6–10.5 of which the enzyme showed significant activity over a wide pH range of 7.5–10.2 with an optimal activity at pH 9.5 **(Figure 5.15 C)**. The activity of $CgSrtE_{\Delta N44}$ was compared with a standard reaction buffer containing Ca^{2+}, with other metal ions (K^{+}, $Mg^{2+,}$ and Mn^{2+}) and also in the absence of Ca^{2+}. None of the metal ions showed any significant effect on the activity of $CgSrtE_{\Delta N44}$ **(Figure 5.15 D)**. A decrease in activity was noted with Mn^{2+}. However, the presence of Ca^{2+} did not have any positive influence on the catalytic efficiency of $CgSrtE_{\Delta N44}$ and it distinguishes this enzyme from the calcium-dependent sortase of *S. aureus*. Besides, the ideal reaction parameters, such as temperature, pH, incubation duration, metal ion, substrate, and enzyme concentrations, were carried out in a single reaction and revealed 7-fold greater sortase E activity with LAHTG substrate within 1 h when compared to the original experiment before optimization **(Figure 5.16)**.

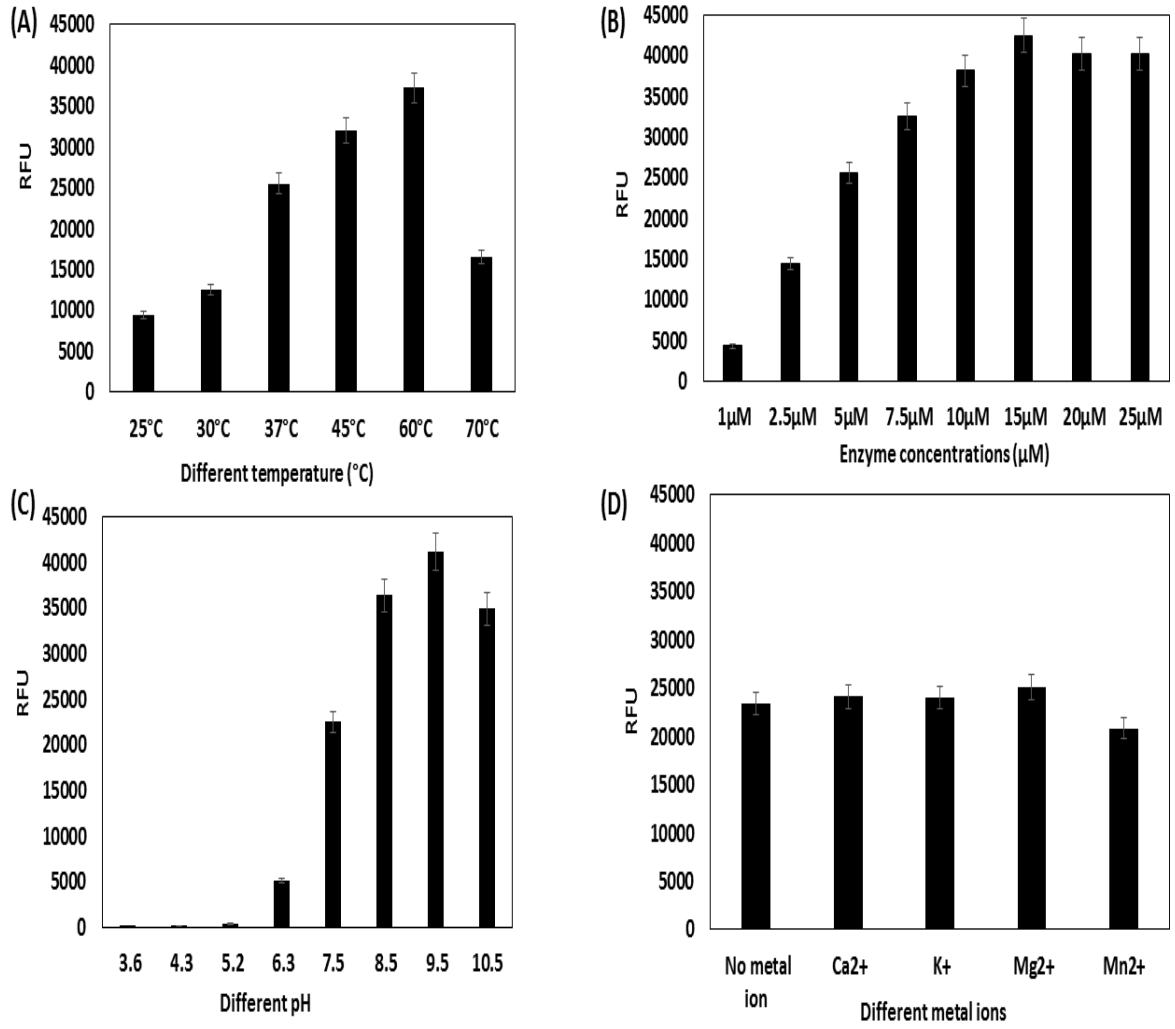


Figure 5.15. Effect of temperature, enzyme concentration, pH, and metal ions on enzyme activity and stability with Abz-LAHTG-Dap (Dnp)

The enzyme assay was performed under standard conditions with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM $CaCl_2$, and 1 mM DTT; **(A)** The effect of temperature was examined. The optimum temperature for the enzymatic activity was 60 °C; **(B)** Maximum sortase E activity was observed at 15 µM enzyme concentration with 40 µM substrate concentration; **(C)** Optimum pH stability was determined at a range of 7.5-10.5; **(D)** The presence or absence of cations Ca^{2+}, Mg^{2+}, $K^{+,}$ and Mn^{2+} showed no significant effect on the activity of $CgSrtE_{\Delta N44}$. The influence of Ca^{2+} did not affect the catalytic efficiency of $CgSrtE_{\Delta N44}$. However, the activity of $CgSrtE_{\Delta N44}$ is observed to be Ca^{2+} independent.

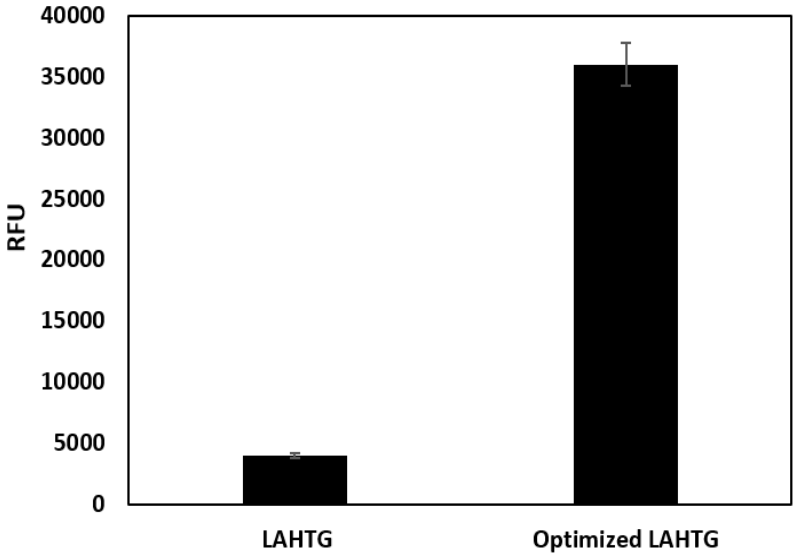


Figure 5.16. CgSrtE$_{\Delta N44}$ activity before and after optimization

All the above-mentioned parameters provided an effective increase with optimized conditions for enzyme assay.

5.3.9. HPLC confirmation of substrate specificity of Sortase E

The hydrolytic activity of CgSrtE$_{\Delta N44}$ was checked by HPLC as well. The CgSrtE$_{\Delta N44}$ (5 µM), was incubated with all the three substrates 40 µM Abz-LAHTG-Dap (Dnp), Abz-LAETG-Dap (Dnp), and Abz-LPETG-Dap (Dnp) in a 500 µL reaction mixture at 37 °C for 6 h. The G-Dap (Dnp) product peak released after the cleavage between threonine and glycine was monitored at a UV absorbance of 355 nm **(Figure 5.17)**. Both Abz-LAHTG-Dap (Dnp) and Abz-LAETG-Dap (Dnp) substrates reacted with an enzyme to form G-Dap (Dnp) products. However, the product formed from LAETG was comparatively less than LAHTG. The LPETG when reacted with CgSrtE$_{\Delta N44}$ did not form any product. The product peak was further collected and analyzed by ESI-MS and it showed a mass of 325.08 Da which is similar to the expected mass of 326.86 Da **(Figure 5.18)**.

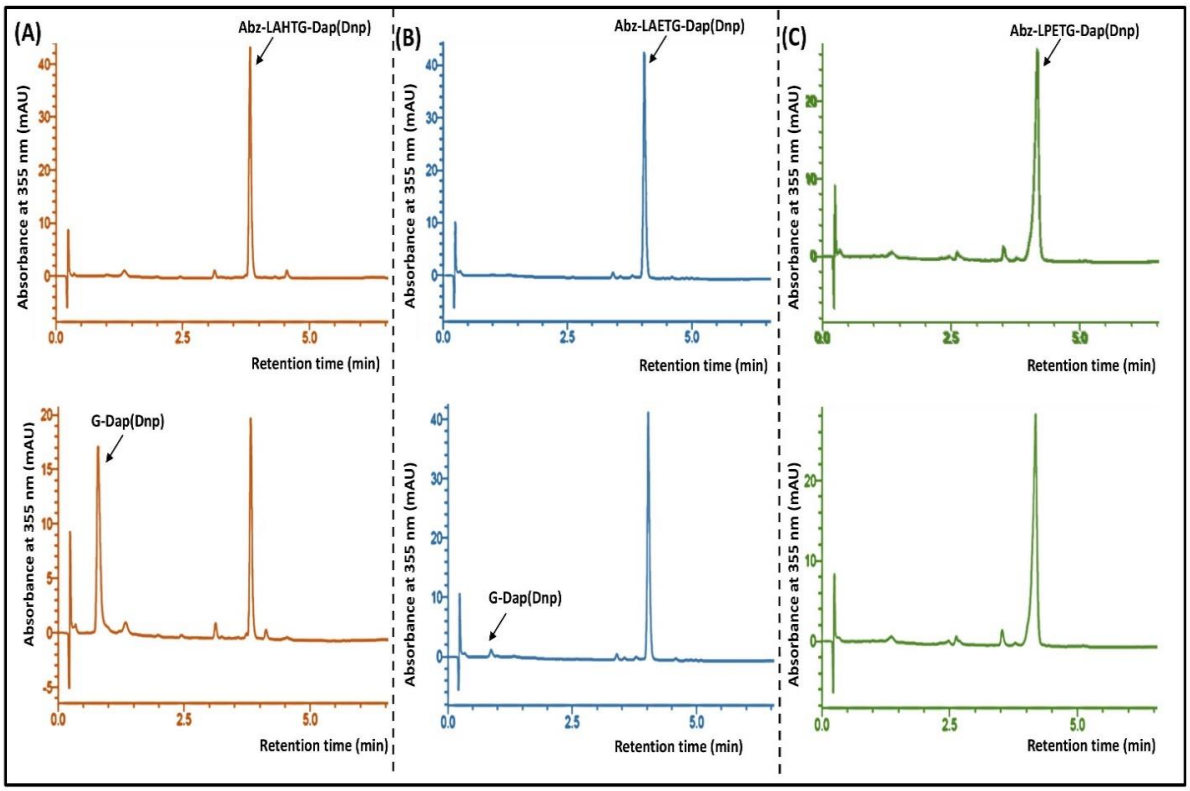


Figure 5.17. Sortase cleavage assay by HPLC with different substrates

(A) Representative HPLC profile of peptide substrate Abz-LAHTG-Dap (Dnp) alone and product G-Dap (Dnp) formed after 6 h reaction with $CgSrtE_{\Delta N44}$ enzyme; **(B)** Representative HPLC profile of peptide substrate Abz-LAETG-Dap (Dnp) alone and product G-Dap (Dnp) formed after 6 h reaction with $CgSrtE_{\Delta N44}$ enzyme; **(C)** Representative HPLC profile of peptide substrate Abz-LPETG-Dap (Dnp) alone and product G-Dap (Dnp) formed after 6 h reaction with $CgSrtE_{\Delta N44}$ enzyme. The reaction products were separated using the C18 RP-HPLC column, and the eluent was monitored at 355 nm UV absorbance.

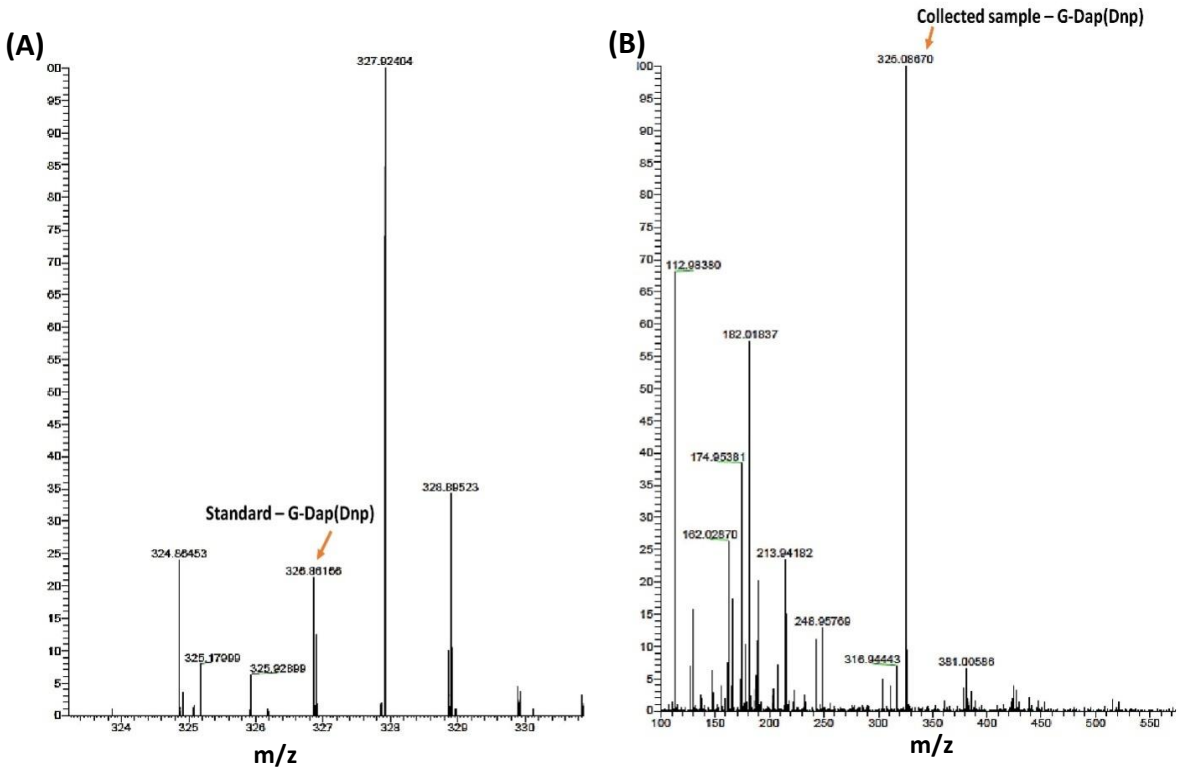


Figure 5.18. ESI-MS analysis of sortase product after cleavage

(A) MS analysis of standard product m/z 326.86; **(B)** MS analysis of purified sample revealed the presence of an ion at m/z 325.08, corresponding to the expected cleavage product G-Dap (Dnp)-NH_2.

5.3.10. Construction of site-directed mutants of $CgSrtE_{\Delta N44}$

The four substitution mutants of $CgSrtE_{\Delta N44}$ (C240A, Y118A, H135A, and R249A) were constructed and the recombinant plasmids bearing these mutations were confirmed by DNA sequencing **(Figure 5.19)**. The mutant plasmids were transformed into *E. coli* BL21 (DE3) cells for overexpression studies.

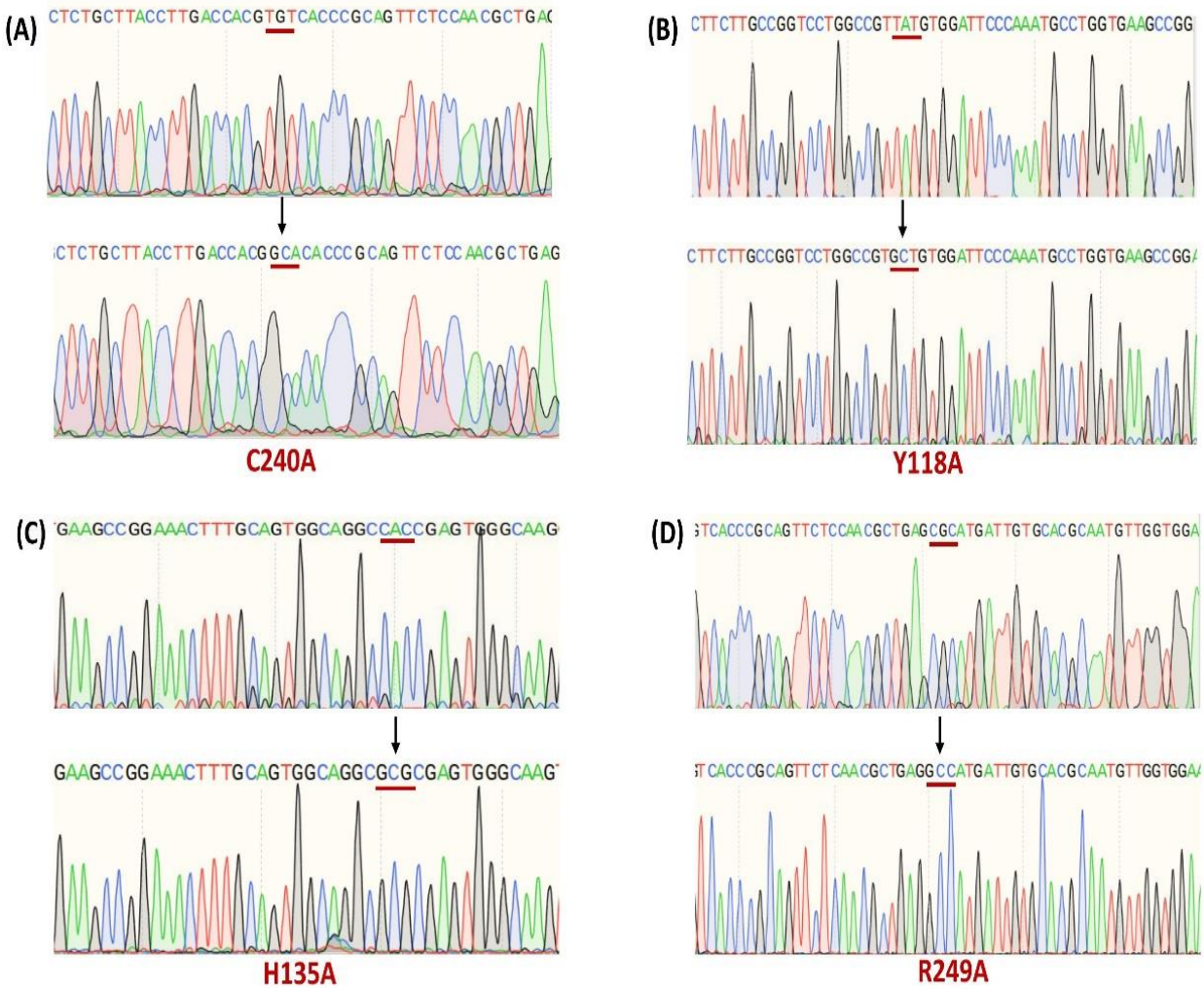


Figure 5.19. Confirmation of site-specific mutations on pTSrtE

(A-D) The chromatogram in each figure shows the wild-type sequence and the lower one shows the mutated sequence. The codons mutated are underlined in red in each case and the corresponding amino acids replaced at those sites are denoted below.

5.3.11. Overexpression and purification of site-directed mutants of CgSrtE$_{\Delta N44}$

Each mutation that occurred on pTSrtE was overexpressed as described in section **(Section 5.2.4)**. The mutants were isolated from the soluble fractions and purified through Ni-NTA columns. The expression, purity, and homogeneity of mutants were analyzed and confirmed through 15 % SDS-PAGE analysis **(Figure 5.20)**. The protein sample was quantified using Bradford reagent and bovine serum albumin (BSA) as a standard in section **(Section 2.3.2 in Chapter 2)**.

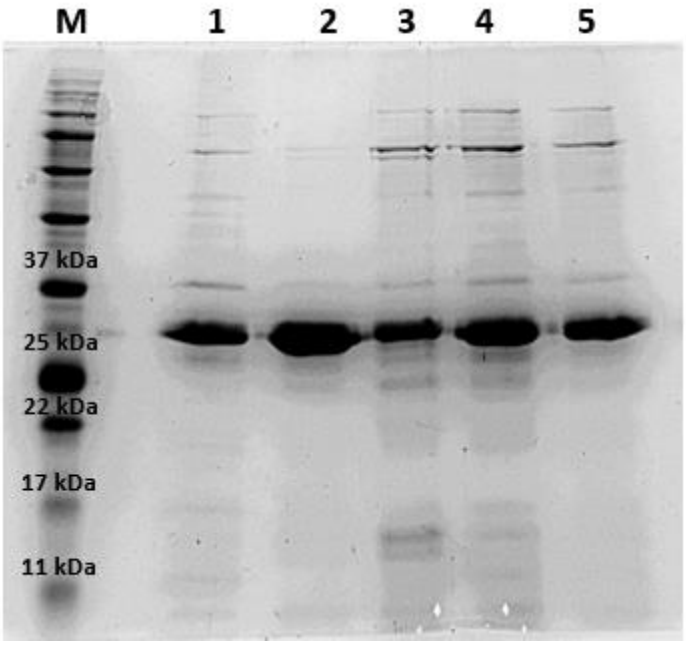


Figure 5.20. SDS-PAGE showing purified site-directed mutant proteins of CgSrtE$_{\Delta N44}$
Lane M: Protein marker 11-245 kDa; **Lane 1:** Wild-type (WT); **Lane 2:** C240A; **Lane 3:** Y118A; **Lane 4:** H135A; **Lane 5:** R249A

5.3.12. Conserved residues in the active site of CgSrtE$_{\Delta N44}$

To functionally confirm the proposed roles of the active site residues H135, C240, R249 based on the previously identified roles in *S. aureus* SrtA and to confirm whether Y118 has a prominent role in class E sortase based on the previous reports on *Streptomyces* sp., site-directed mutagenesis was done. The constructed mutant proteins were incubated with the FRET peptide Abz-LAHTG-Dap (Dnp). All the active site mutants showed more than 50 % loss in catalytic activity **(Figure 5.21)**. Mutation in tyrosine residue also drastically decreased the activity indicating that the -OH group of Y118 could be involved in a hydrogen bond with the backbone nitrogen of the Ala residue for the recognition of the LAXTG sorting signal instead of LPXTG. Thus, it is experimentally validated that the conserved catalytic residues mainly include H135, C240, R249, and Y118.

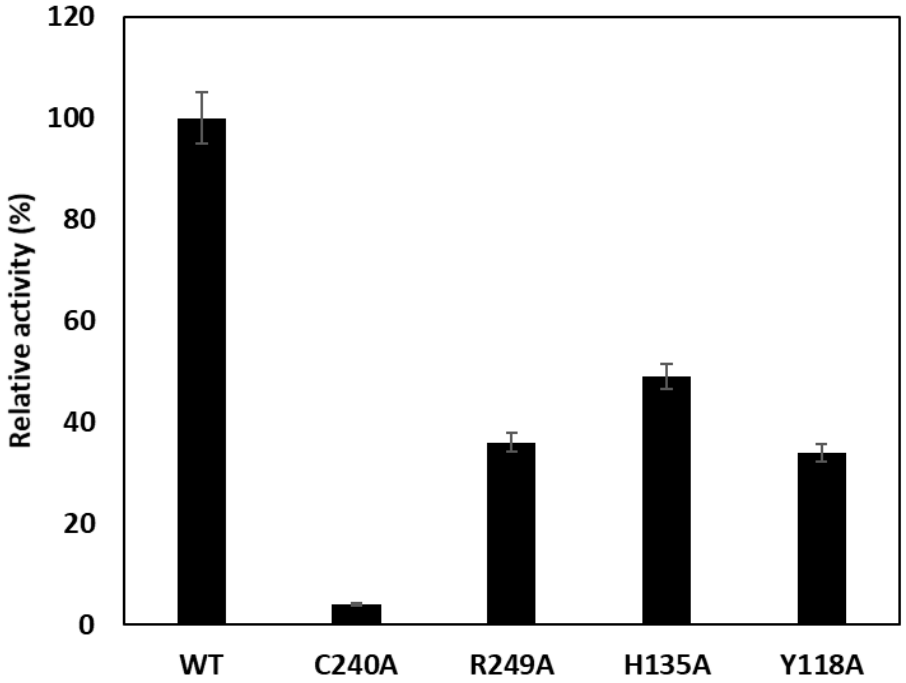


Figure 5.21. Confirmation of conserved residue required for in vitro CgSrtE$_{\Delta N44}$ activity
Single mutations at C240A, H135A, R249A, and Y118A resulted in the loss of enzymatic activities against Abz-LAHTG- Dap (Dnp), indicating that these four residues played critical roles in CgSrtE$_{\Delta N44}$ activity.

5.4. Discussion

Sortase transpeptidases are present in Gram-positive bacteria to covalently attach proteins to their cell wall or to construct pili (Scott and Barnett, 2006). *C. diphtheria* has a single class E sortase (called Cd-SrtF) that aids in the attachment of formed pili to the cell wall peptidoglycan (Swaminathan et al., 2007; Chang et al., 2011) with a sequence resemblance to non-pathogenic *C. glutamicum* with unclear function. There are at least six varieties of sortases (class A-F) based on their amino acid sequences; however, class E enzymes have not been widely explored. Soil and freshwater-dwelling Actinobacteria use class E sortases to display proteins with a non-canonical LAXTG sorting signal, which differs from 90 % of known sorting signals by substituting alanine for proline (Kattke et al., 2016). Among the three peptides, the enzyme CgSrtE$_{\Delta N44}$ demonstrated a higher substrate affinity

towards Abz-LAHTG-Dap (Dnp) with an apparent K_m of 0.012 mM which is higher than the reported SavSrtE K_m of 1.14 mM for the LAXTG motif (Das et al., 2017). The $CgSrtE_{\Delta N44}$ also failed to cleave the Abz-LPETG-Dap (Dnp) substrate of SrtA, hence proving that the $CgSrtE_{\Delta N44}$ belongs to the class E family. The calcium ions enhance the catalytic efficiency of SrtA of *S. aureus* which mainly contains E105, E108, D112, and E171 residues on β3-β4 and β6-β7 loops within the structure (Antos et al., 2009; Das et al., 2017). Since, these calcium-binding sites are not found in Cg-SrtE, which proves that Ca^{2+} is not essential for the enhancement of sortase E activity in *C. glutamicum.* The sortase-catalyzed transpeptidation involves Cys-His-Arg catalytic triads, the Cys residue at the active site of sortase cleaves the Thr-Gly bond and forms a stable thioacyl intermediate complex, His protonates the substrate leaving group and Arg believes to function as an oxyanion hole to stabilize the transition state and further help in proper positioning of substrate in the active site of the enzyme (Frankel et al., 2005, 2007; Bentley et al., 2008). The Tyr residue which is highly conserved in class E sortase recognizes the alanine residue of the LAXTG sorting signal. However, mutating at the conserved residues H135, C240, R249, and Y118 of $CgSrtE_{\Delta N44}$ has led to the conformational destabilization of the protein with a drastic reduction in catalytic efficiency, believing that these residues might play a significant part in the enzyme.

5.5. Summary

Due to the hydrophobic nature of the N-terminus membrane anchor sequence, the full-length CgSrtE protein is not favorable for recombinant expression and purification as it is insoluble and sediments along with the membranous fraction of the cell lysate. As a result, the signal peptide and transmembrane region of *C. glutamicum* sortase fraction which comprises around 1-44 amino acids were removed from the protein to increase the solubility of the protein. $CgSrt_{\Delta N44}$ protein was purified from soluble fraction using Ni-NTA columns in 200

mM imidazole fraction. The recombinant sortase comprising residues 45-274, $CgSrtE_{\Delta N44}$ has been demonstrated to be enzymatically active indicating that removal of nonconserved residues did not affect the catalytic activity. The HPLC and FRET were used to measure sortase E cleavage activity in vitro with synthesized peptides. $CgSrtE_{\Delta N44}$ had a higher substrate affinity for Abz-LAHTG-Dap (Dnp), with an apparent K_m of 0.012 mM, but failed to cleave SrtA substrate Abz-LPETG-Dap (Dnp), indicating that it belongs to class E. The absence of calcium-binding sites in $CgSrtE_{\Delta N44}$ demonstrates that Ca^{2+} is not required for the enhancement of sortase E activity in *C. glutamicum.* This evidence supports the findings of the FRET analysis. Furthermore, the enzyme demonstrated significant activity across a wide pH range of 7.5–10.2, with an optimum pH of around 9.5. When incubated at a higher temperature of around 60 °C, the enzyme catalytic efficiency was improved, and an enzyme concentration of around 15 µM was found to be optimal. Sortase-catalyzed transpeptidation involves Cys-His-Arg catalytic triads, which on mutation showed a 50 % decrease in activity when compared to wild-type.

Chapter 6

Sortase E-mediated site-specific immobilization of green fluorescent protein and xylose dehydrogenase on gold nanoparticles

Abbreviation

AuNPs	Gold nanoparticles
DCM	Dichloromethane
DIC	N, N′-Diisopropylcarbodiimide
DLS	Dynamic Light Scattering
DMF	Dimethylformamide
EDC	N-(3-(dimethylamino)-propyl)-N'-ethyl carbodiimide
eGFP	Enhanced green fluorescent protein
HMPB	4-Hydroxymethyl-3-methoxyphenoxybutyric acid
MBHA	4-Methylbenzhydrylamine
NHS	N-hydroxysuccinimide
NMR	Nuclear magnetic resonance
SERS	Surface-enhanced Raman scattering
TEM	Transmission electron microscopy

6.1. Introduction

Enzymes produced from live cells act as natural biocatalysts, perform biochemical reactions under mild conditions with a high degree of substrate specificity (Girolamo et al., 2019). Compared with chemical catalysts, the ease of production, catalytic efficiency, the chemo-, regio- and stereospecificity greatly promotes the enzyme to be used in pharmaceutical, chemical, and food industry applications (Zhang et al., 2013). However, poor stability of the enzyme, low shelf life, and high cost limits its use in industrial application. One of the best strategies to overcome this problem is to immobilize enzymes using a carrier

such as a hydrogel, biopolymer, synthetic polymers, nanoparticles, etc (Bornscheuer, 2003; Homaei et al., 2013).

The recent reports validate that enzymes immobilized on nanostructured materials have an efficient surface area, improved enzyme loading, and enhanced activity (Ansari and Husain, 2012). The use of nanoparticles has brought significant medical applications in wound healing, cell imaging, tissue engineering, and drug delivery. Among various nanoparticles used, gold nanoparticles (AuNPs) are considered to be an ideal model for studying the effects of immobilization on the activity and stability of different functional proteins (Saware et al., 2015; Ma et al., 2018). The AuNPs have an extremely large surface area which allows a significant amount of proteins to get adsorbed on their surface when introduced into biological entities (Kong et al., 2017). Also, PEGylated AuNPs exhibit promising advantages like enhanced biocompatibility, stability, etc. Interestingly, $AuNP_S$ serves as a promising surface-enhanced Raman scattering (SERS) substrate for getting amplified Raman signals of the molecular fingerprint. SERS provides high sensitivity and molecular specificity along with its capability to resolve complex molecular level biological conformations. The integration of advanced technologies and minimal sample volume improved the utility of SERS in quantitative and qualitative identification of various biomolecules such as nucleic acids and proteins, as they provide characteristic signals (Bantz et al., 2011). Herein, we made use of this technique to track the immobilization of eGFP and XylB on triglycine PEGylated AuNPs via Sortase E-mediated technique.

In recent years, sortase-mediated ligation (sortagging) developed as a robust and powerful tool for peptide or protein conjugation through site-specific modification. The target proteins were altered based on the substrate specificity of the enzyme under mild conditions and without prior chemical modifications (Hata et al., 2015; Zou et al., 2019). Sortases are membrane-bound cysteine transpeptidase that catalyzes a sequence-specific ligation of

proteins to the cell wall of Gram-positive bacteria. Among different classes of sortases, the sortase A from *S. aureus* (SaSrtA) is studied extensively. SrtA recognizes an LPXTG motif at the C-terminal of the protein, cleaves the amide bond between threonine (T) and glycine (G), and covalently links to the N-terminal glycine residues (Mazmanian et al., 1999). Due to its high degree of substrate selectivity, SaSrtA has been widely used in immobilizing proteins or enzymes on solid supports (Chan et al., 2007), nanoparticles (Hata et al., 2015), microgels (Zou et al., 2019), and hydrogels (Cambria et al., 2015). However, *C. glutamicum* ATCC 13032, a non-pathogenic and well-known industrial microbe for amino acid production, contains a Sortase E transpeptidase with high substrate selectivity with LAXTG. The Ca^{2+} independence, high substrate specificity, and ability to accept N-terminal (oligo) residues make *C. glutamicum* Sortase E (CgSrtE) much more efficient than pathogenic SaSrtA in live cells and for industrial sortagging applications.

In this study, we used $CgSrtE_{\Delta N44}$ to immobilize green fluorescent protein (eGFP) and xylose dehydrogenase (XylB) on triglycine functionalized AuNPs. Our initial attempt was to demonstrate successful immobilization of enhanced eGFP, a model protein with a C-terminal LAHTG-His_6 recognition sequence attached to an N-terminal GGG@PEG@AuNP, and once that was accomplished, we subsequently immobilized and analyzed the activity of XylB from *Caulobacter cresentus*, which uses NAD^+ as a cofactor to convert D-xylose to produce xylonic acid, which serves as a platform chemical (Lee et al., 2018).

6.2. Materials and Methods

6.2.1. Instrumentation

The UV-Vis absorption and emission spectra were obtained using Shimadzu UV-2600 UV-Vis spectrophotometer. Absorbance spectra were recorded from 200-800 nm. ^1H NMR spectra were recorded on Bruker Advance 500 NMR spectrometer, and chemical shifts are

expressed in parts per million (ppm). SERS measurements were performed in a WITec Raman microscope (WITec Inc. Germany, alpha 300R) with a laser beam focused on the sample through a 20x objective and a Peltier-cooled CCD detector. Samples were excited using a 633 nm laser, and Raman spectra were accumulated in the Stokes-shifted range of 400−4000 cm^{-1} with 1 cm^{-1} resolution. Before every measurement, calibration using a silicon standard with Raman peak centered at 520 cm^{-1} was carried out. For data analysis, the WITec Project plus (v5.2) software was used. TEM measurements were performed on a JEOL 2010 high-resolution transmission electron microscope with an accelerating voltage of 200 kV. The samples were prepared by pipetting a drop of the aqueous solution of nanoparticles onto a 230-mesh copper grid coated with carbon, and the samples were air-dried before doing the measurement.

6.2.2. AuNP synthesis

AuNPs were synthesized using the standard citrate reduction method (Narayanan et al., 2015). 300 mL of deionized water was heated to 100 °C with continuous stirring (600 rpm). To this, hot deionized water, 300 µL of 250 mM chloroauric acid solution was added and allowed to stir for another 5 min. After this, 750 µL of 100 mM trisodium citrate solution was added and allowed to stir until the color changed to purple. Once the color is changed to wine red, heating was stopped and cooled to room temperature with continuous stirring. The obtained AuNPs were purified by centrifugation followed by resuspension in Milli-Q water and kept in the refrigerator for further use.

6.2.3. Amine PEG encapsulation on AuNPs

To the AuNPs, 3.6 mL of 100 µM HS-PEG-NH_2 was added. This mixture was vortexed and allowed to incubate for 3-4 h to get maximal surface coverage and PEG stabilization. Excess HS-PEG-NH_2 was removed after 3-4 h of thorough mixing and

centrifugation at 4000 x g for 20 min. Repeated centrifugation twice and re-suspended in Milli-Q water (Narayanan et al., 2015; Saranya et al., 2018).

6.2.4. Synthesis of resin-bound GGG peptide sequence

200 mg (0.142 mmol) of HMPB-MBHA resin was swelled using dichloromethane (DCM) for 30 min and then allowed to react with N, N′-Diisopropylcarbodiimide (DIC) activated Fmoc-Gly-OH (337.7 mg, 1.136 mmol) for 24 h. After this, the reaction mixture was washed with dimethylformamide (DMF) (3 x 3 mL) and the Fmoc protection group was cleaved using 20 % piperidine in DMF. The reaction was continued similarly with Fmoc-Gly-OH for another two cycles. The resulting resin-bound $(Gly)_3$ peptide was washed 3 times each with 5 mL DMF and DCM. Then washed with 5 mL n-hexane and dried for 1 h. The resin-bound peptide was kept in the refrigerator (at least for 1 day). Then the desired amount of peptide sequence was cleaved from the resin using 2 % TFA in dichloromethane (5 x 3 mL). Yield 68 % 1H NMR (500 MHz, $CDCl_3$): δ 8.02 (s, 2H), 7.70-7.62 (m, 4H), 7.35-7.29 (m, 4H), 7.26-7.22 (m, 4H), 4.23 (s, 3H), 2.98 (s, 2H) 2.88 (s, 2H). HRMS (ESI): Calcd for $C_{21}H_{21}N_3O_6$: 411.14; Found $[M^+Na^+]$: 434.1318 **(Figure 6.1)** (Saranya et al., 2018; Sujai et al., 2021).

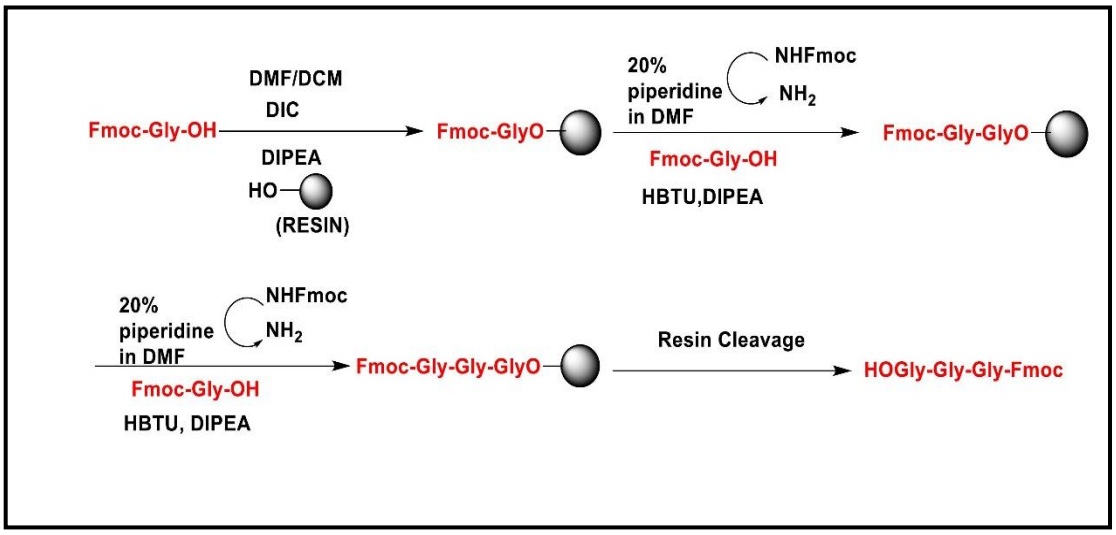


Figure 6.1. The synthetic route adopted in GGG peptide synthesis

6.2.5. Coupling of GGG peptide to PEGylated AuNPs

Here, EDC (N-(3-(dimethylamino)-propyl)-N'-ethyl carbodiimide) (25 mM) and NHS (N-hydroxysuccinimide) (25 mM) solutions were taken together to activate the -COOH groups present on the peptide sequence (2.0 $mgmL^{-1}$). Later, this activated solution (100 µM) was added to the PEGylated AuNPs solution and was allowed to react for 6 h. After this, the solution was centrifuged at 4000 x g for 20 min and the pellet was washed using Milli-Q water. For deprotecting Fmoc, 200 µL of 20 % piperidine in DMF was added to the solution and properly mixed for another 10 min. Further, the solution was centrifuged at 4000 x g for 20 min and the pellet was washed using Milli-Q water. The resulting pellet was redispersed in water and kept at 4 °C for further use (Ramya et al., 2016; Sujai et al., 2021).

6.2.6. Cloning of constructs

LAHTG coding nucleotide residues were added to *eGFP* and *xylB* genes at the 3' end of the sequences. Each gene sequence was amplified using the recombinant plasmids pGFP and pXylB by gene-specific primers eGFP-LAHTG F/R and XylB-LAHTG F/R as shown in **Table 2.3 (Chapter 2).** *eGFP-LAHTG* and *xylB-LAHTG* sequences were flanked by NdeI/BamHI and EcoRI/BamHI restriction sites respectively. The eGFP and XylB fused to C-terminal LAHTG sequences were amplified by PCR and cloned into NdeI/BamHI and EcoRI/BamHI digested pET28a to yield recombinant plasmids pGFP-LAHTG and pXylB-LAHTG, respectively. Similarly, a recombinant $CgSrtE_{\Delta N44}$ enzyme was constructed from *C. glutamicum* ATCC 13032, flanked by NdeI/SalI restriction sites, was prepared by PCR and cloned into pET28a, linked at its N-terminal His_6 sequence to give pTSrtE **(Section 5.2.4 in Chapter 5)**. All the three plasmids were transformed into *E. coli* BL21 (DE3) to generate recombinant eGFP, XylB, and $CgSrtE_{\Delta N44}$ respectively for protein expression and purification.

6.2.7. Protein expression and purification of eGFP-LAHTG and XylB-LAHTG

The calcium-independent $CgSrtE_{\Delta N44}$ was expressed, purified, and showed a higher substrate activity towards LAHTG substrate in **Chapter 5**. The pGFP-LAHTG construct was allowed to grow in LB medium supplemented with 50 μgmL^{-1} kanamycin concentration at 37 °C with continuous shaking until OD_{600} reaches 0.8 and induced with 1 mM IPTG for 20 h at 16 °C. The cells were harvested by centrifugation at 3500 x g at 4 °C for 30 min. Cell pellets were resuspended in lysis buffer containing 100 mM NaCl, 10 mM Tris-HCl, 50 mM NaH_2PO_4, pH 8.0 which was lyzed by sonication (Sonics Vibra cell, U.S.A.), and then centrifuged at 20,000 x g for 30 min at 4 °C. The lysate was purified using HisTrap HP 1-mL column (GE Healthcare, USA), subsequently rinsed with lysis buffer and wash buffer with increasing concentrations (25 mM, 50 mM, 75 mM, and 100 mM) of imidazole, and the protein was eluted with 250 mM imidazole. The excess imidazole was removed using the PD-10 desalting column. The purified eGFP-LAHTG protein was stored in buffer containing 50 mM Tris-HCl, pH 8.0, and 150 mM NaCl and concentrated using a MWCO of 10 kDa (Millipore).

The pXylB-LAHTG was expressed in *E. coli* BL21 (DE3), culturing in LB medium to $OD_{600\ nm}$ = 0.6 at 37 °C, adding 1 mM IPTG, and further culturing for 3 h at 37 °C. Cell pellets were resuspended in 0.5 M phosphate buffer saline (PBS) lysed by sonication and centrifuged at 20000 x g for 30 min at 4 °C. The supernatant was then purified using HisTrap HP 5-mL column, subsequently rinsed with 5 volumes of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0), and wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 100 mM imidazole, pH 8.0), and the protein was eluted directly with 500 mM imidazole. The excess imidazole was removed using the PD-10 desalting column. The XylB was stored in buffer containing 100 mM NaH_2PO_4 and 50 mM NaCl, pH 8.0, and concentrated using 10 kDa Amicon ultra centrifugal filters.

All these protein samples were quantified using Bradford reagent with an estimated concentration of bovine serum albumin (BSA) as a standard **(Figure 2.1 in Chapter 2)** and stored at -80 °C. The purity of the protein was detected using 15 % SDS–PAGE.

6.2.8. Sortase E-mediated protein ligation

Immobilization of protein was performed with a functionalized G-tag modified AuNPs at a total volume of 40 µL and 15 µM of pGFP-LAHTG or pXylB-LAHTG. The reaction was initiated by the addition of 5 µM SrtE along with ligation buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, and 1 mM DTT) for 2 h at 37 °C. The immobilized AuNPs were separated from the reaction mixture by centrifugation at 4000 x g for 20 min at 4 °C. The immobilized AuNPs with GFP-LAHTG or XylB-LAHTG were further stored at 4 °C until subsequent analysis. The protein content after the immobilization of XylB with AuNP was estimated by using the standard Bradford method (Emami Bistgani et al., 2017).

6.2.9. Bioconversion of xylose to xylonic acid by immobilized XylB

The reaction mixture contains 50 mM Na_2HPO_4 (pH 8.0), 2 mM NAD^+ and 200 mM xylose. The reaction was initiated with the addition of 15 µM free or immobilized XylB and incubated at 30 $^{\circ}$C for 72 h. The amount of xylonic acid (xylonate) produced were detected by automated high-performance liquid chromatography (HPLC) system (Prominence UFLC, Shimadzu, Japan) equipped with auto-sampler, Rezex^{TM} ROA-Organic Acid H^+ LC column 300 x 7.8 mm (Phenomenex), 0.01 N H_2SO_4 mobile phase, oven, and PDA detector. The amount of xylonic acid was evaluated by plotting a standard graph **(Figure 6.2)**.

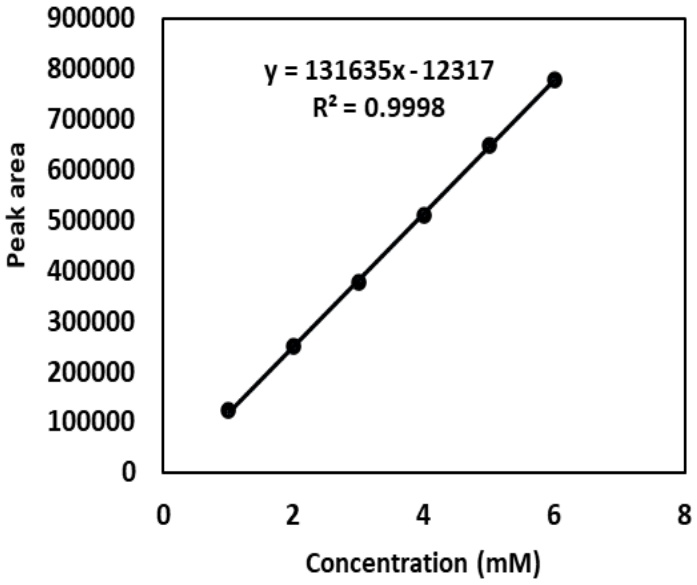


Figure 6.2. Calibration curve for xylonic acid vs peak area

The xylose estimation was carried out spectrophotometrically by Bial's method using the orcinol reagent (Pham et al., 2011). In this test, acid degrades xylose to furfural. Furfural is then combined with $FeCl_3$ and orcinol to create a blue-green precipitate measured at an absorbance of 620 nm wavelength. Bial's reagent was prepared by dissolving 300 mg of orcinol in 5 mL ethanol. Added 3.5 mL of this mixture to 100 mL of 0.1 % solution of $FeCL_3.6H_2O$. The reagent thus formed is to be stored in a dark bottle and used within a couple of hours. A standard curve was plotted against xylose concentration within a range of 1-10 $mgmL^{-1}$ **(Figure 6.3)**. To each 1 mL of the sample, 2.5 mL of the Bial's reagent was added and incubated for 10 min in boiling water. All the readings were taken in triplicates.

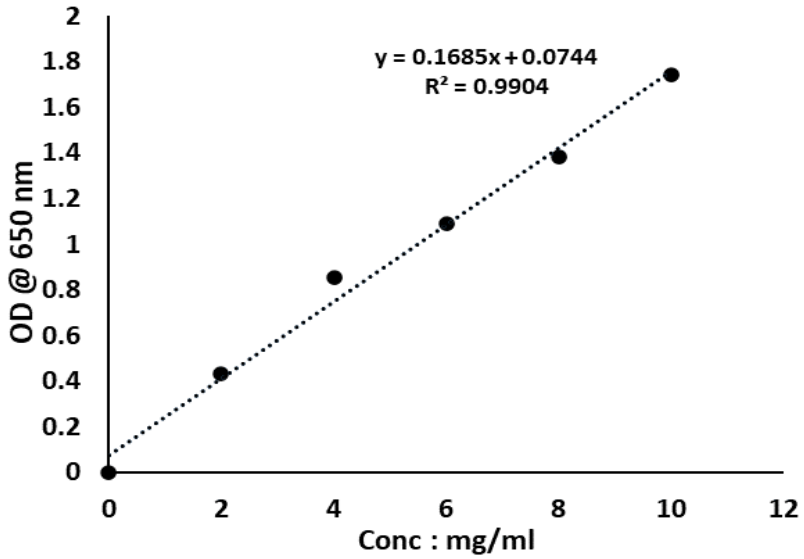


Figure 6.3. Calibration curve for xylose concentration vs absorbance by Orcinol assay

6.2.10. Reusability of immobilized XylB

The reusability of immobilized XylB was evaluated based on the catalytic conversion of xylose to xylonic acid. Each cycle lasted for 72 h and was carried out under optimal conditions using a reaction mixture consisting of 50 mM Na_2HPO_4 (pH 8.0), 2 mM NAD^+ and 200 mM xylose at 30 °C, and the product detection was done by HPLC. After each cycle, the enzyme bounded nanoparticle was recovered by centrifugation at 4000 x g for 20 min and washed twice with MillQ, and resuspended in a new reaction solution. The reaction was repeated for 4 cycles. The xylonic acid produced after the first cycle was defined as 100 %, which was compared with other repeated cycles to calculate the relative activity.

6.3. Results

6.3.1. PCR amplification and cloning of *eGFP-LAHTG* and *xylB-LAHTG* gene

LAHTG sequences were added to *eGFP* and *xylB* genes at the 3' end of the sequences. Each gene sequence was amplified using the recombinant plasmids pGFP and pXylB by gene-specific primers eGFP-LAHTG F/R and XylB-LAHTG F/R as shown in **Table 2.3 (Chapter**

2). The optimized PCR condition of the *eGFP-LAHTG* gene was shown in **Table 6.1** and *xylB-LAHTG* gene conditions were shown in **Table 2.5 (Chapter 2)**. The amplicons *eGFP-LAHTG* and *xylB-LAHTG* gene showed an expected size of 738 bp and 765 bp on 1.5 % agarose gel **(Figure 6.4)**.

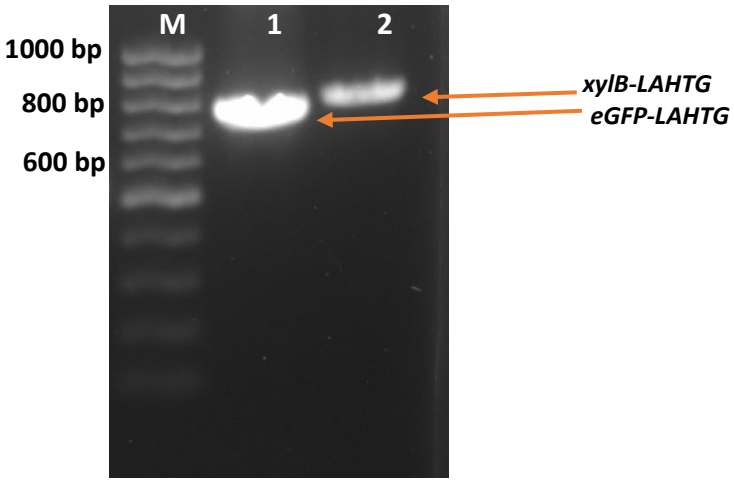


Figure 6.4. PCR amplification of *eGFP* and *xylB* gene

Lane: 1 100 bp DNA ladder; **Lane 2:** *eGFP* (738 bp) PCR amplicon; **Lane 3:** *xylB* (765 bp) PCR amplicon

Table 6.1. PCR cycling conditions of *eGFP* gene

Steps	Temperature (°C)	Time	
Initial Denaturation	94	3 min	
Denaturation	94	30 s	35 cycles
Annealing	57	30 s	
Extension	72	2 min	
Final extension	72	7 min	
	4	hold	

The *eGFP-LAHTG* and *xylB-LAHTG* genes were incorporated with Nde1/BamHI and BamHI/EcoRI restriction sites respectively. The genes were individually ligated into pET28a and the recombinant plasmids pGFP-LAHTG and pXylB-LAHTG were confirmed by double digestion with the respective restriction sites **(Figure 6.5).** Clones were sequenced by T7

primers, sequences were verified and the ORF analysis proved that the amplicons were completed without any kind of mutation.

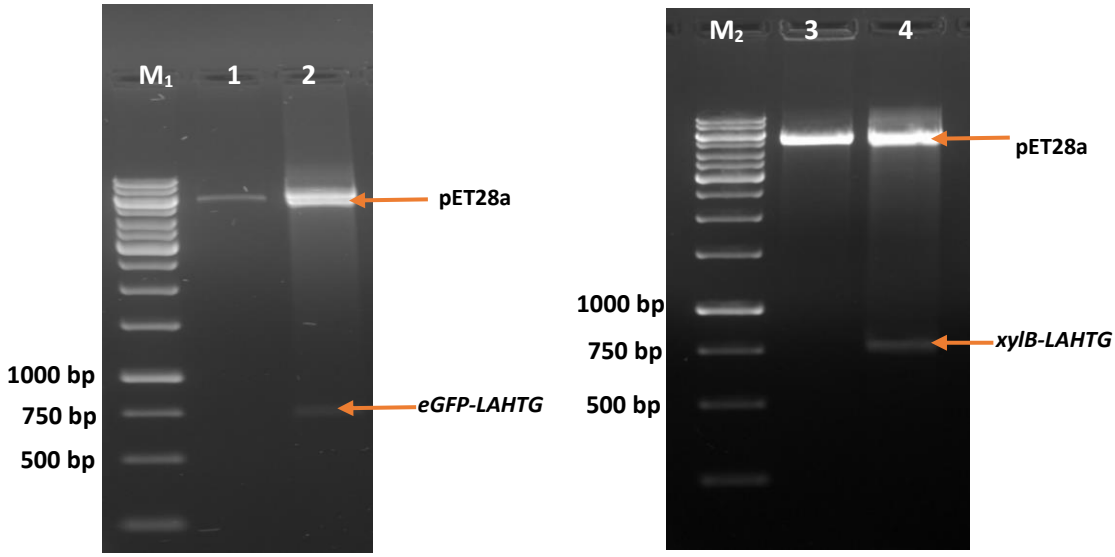


Figure 6.5. Restriction digestion of recombinant plasmids pGFP-LAHTG and pXylB-LAHTG

Lane M_1: 1 kb DNA ladder; **Lane 1:** Native plasmid pET28a NdeI/BamHI double digest; **Lane 2:** pET28a-*eGFP-LAHTG* NdeI/BamHI double digest showing insert release at 738 bp; **Lane M_2:** 1 kb DNA ladder; **Lane 3:** Native plasmid pET28a BamHI/EcoRI double digest; **Lane 4:** pET28a-*xylB-LAHTG* BamHI/EcoRI double digest showing insert release at 765 bp

6.3.2. Expression and purification of eGFP-LAHTG and XylB-LAHTG in *E. coli*

E. coli BL21 (DE3) cells carrying pGFP-LAHTG and pXylB-LAHTG, were chosen for the expression of eGFP-LAHTG and XylB-LAHTG, under the control of IPTG inducible T7 promoter **(Figure 6.6 A and B)**.

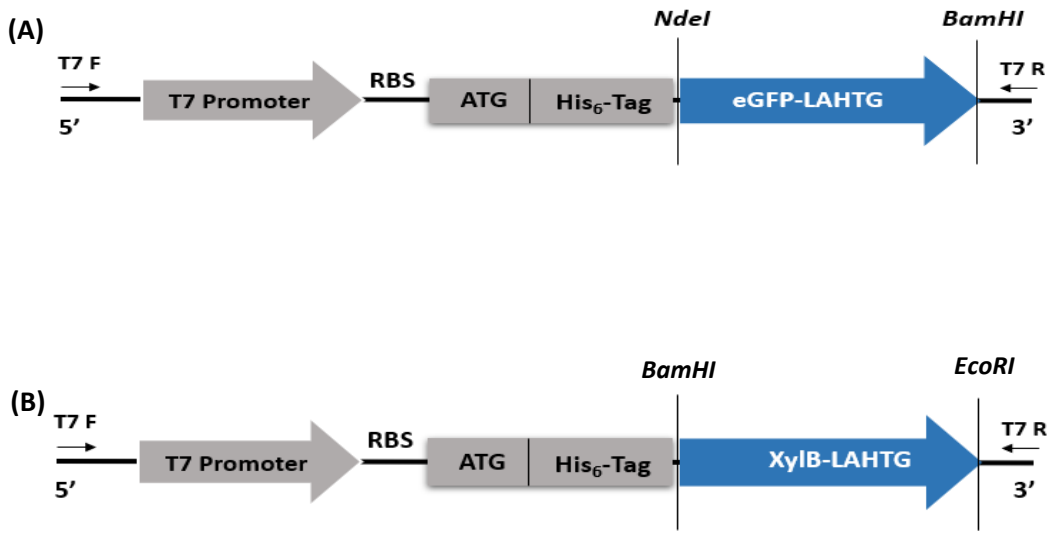


Figure 6.6. Schematic diagram of pGFP-LAHTG and pXylB-LAHTG construct

(A) eGFP-LAHTG expression from pET28a is under the control of IPTG inducible promoter, and results in an N-terminally His_6-tagged recombinant protein with an expected mass of 31.2 kDa; **(B)** XylB-LAHTG expression from pET28a is under the control of IPTG inducible promoter, and results in an N-terminally His_6-tagged recombinant protein with an expected mass of 32.3 kDa

The N-terminal His_6-tagged protein was purified using the HisTrap HP 5-mL column. The expression, purity, and homogeneity of eGFP-LAHTG were analyzed and confirmed through SDS-PAGE analysis **(Figure 6.7).**

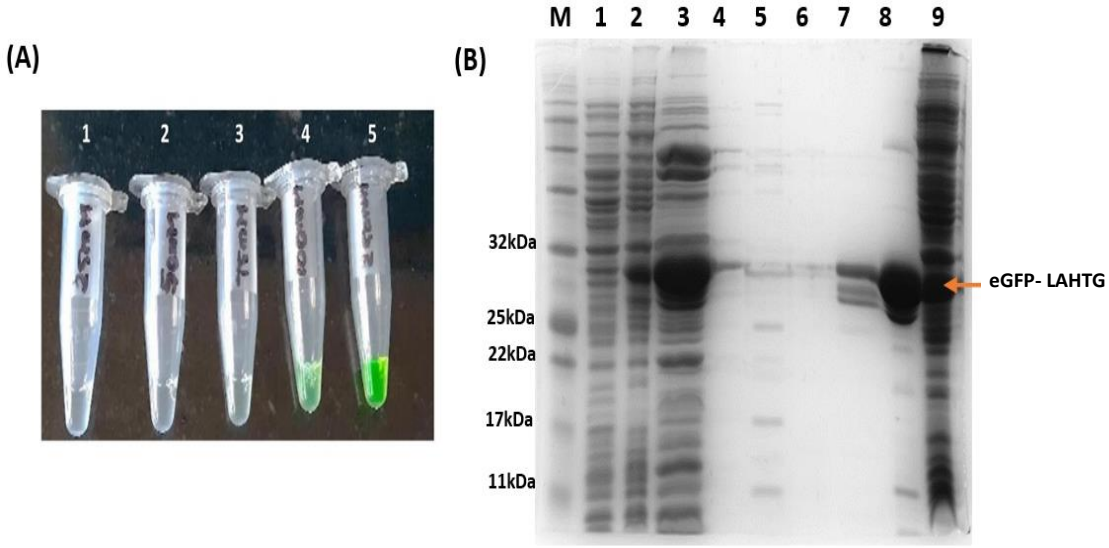


Figure 6.7. Expression and purification of eGFP-LAHTG

(A) Gradient of imidazole fractions [25(1), 50 (2), 75 (3), 100 (4) and 250 mM (5)] collected after passing through Ni-NTA column; **(B) Lane M:** Molecular protein marker**; Lane 1:** Uninduced crude protein; **Lane 2:** Induced crude protein; **Lane 3:** Soluble fraction; **Lane 4:** 25 mM imidazole gradient; **Lane 5:** 50 mM imidazole gradient; **Lane 6:** 75 mM imidazole gradient; **Lane 7:** 100 mM imidazole gradient; **Lane 8:** Partially purified at 250 mM imidazole gradient; **Lane 9:** Expression of eGFP-LAHTG in the insoluble fraction as inclusion body.

The N-terminal His_6-tagged protein was purified using the HisTrap HP 5-mL column. The expression, purity, and homogeneity of XylB-LAHTG were analyzed and confirmed through SDS-PAGE analysis **(Figure 6.8).**

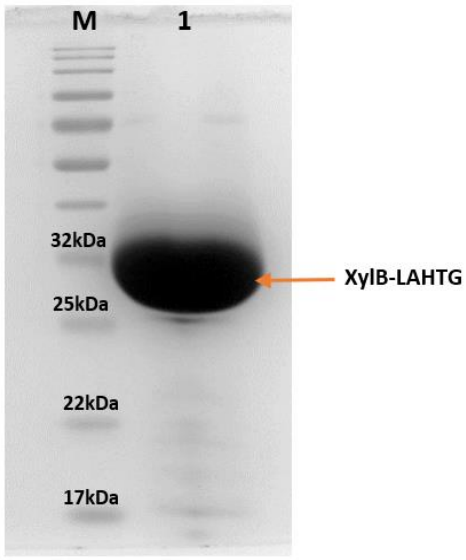


Figure 6.8. Purification of XylB-LAHTG

Expression and SDS-PAGE analysis of purified recombinant protein with C-terminal LAHTG tag; **Lane M:** Molecular protein marker; **Lane 1:** purified XylB-LAHTG.

6.3.3. Fabrication of peptide conjugated AuNPs

The fabrication of NH_2-PEG@AuNP was started with the synthesis of ~ 40 nm AuNP. AuNPs were synthesized through the well-known Turkevich method (Narayanan et al., 2015) and were characterized via UV-Vis spectroscopy, HR-TEM, and DLS. The synthesized AuNPs were further encapsulated with NH_2-PEG-SH for introducing the amine group. Sulfur has a strong binding affinity towards AuNPs; the adsorption of PEG over AuNPs is facilitated by chemisorption. NH_2-PEG coating over AuNPs has been characterized using UV–Vis spectroscopy, which showed a 2-3 nm shift in absorption maxima. The shift in plasmonic absorption peak from 530 nm to 532 nm indicates the PEG coating over AuNPs. This conjugation was further confirmed by HR-TEM and DLS studies. A thin layer of 3-4 nm thickness of PEG was visible over the AuNP surface in the TEM image, and the increase in size after NH_2-PEG encapsulation was also evident from DLS analysis **(Figure 6.9)**. The size observed in TEM and DLS were comparable. For PEG encapsulated AuNP, the size obtained from DLS analysis was slightly higher because of the higher hydrodynamic size possessed by

PEG coating due to the high rate of solvation This hydrodynamic size is always greater than the dry size obtained from HR-TEM. The NH_2-PEG@AuNP was used for further conjugation with tri-peptide.

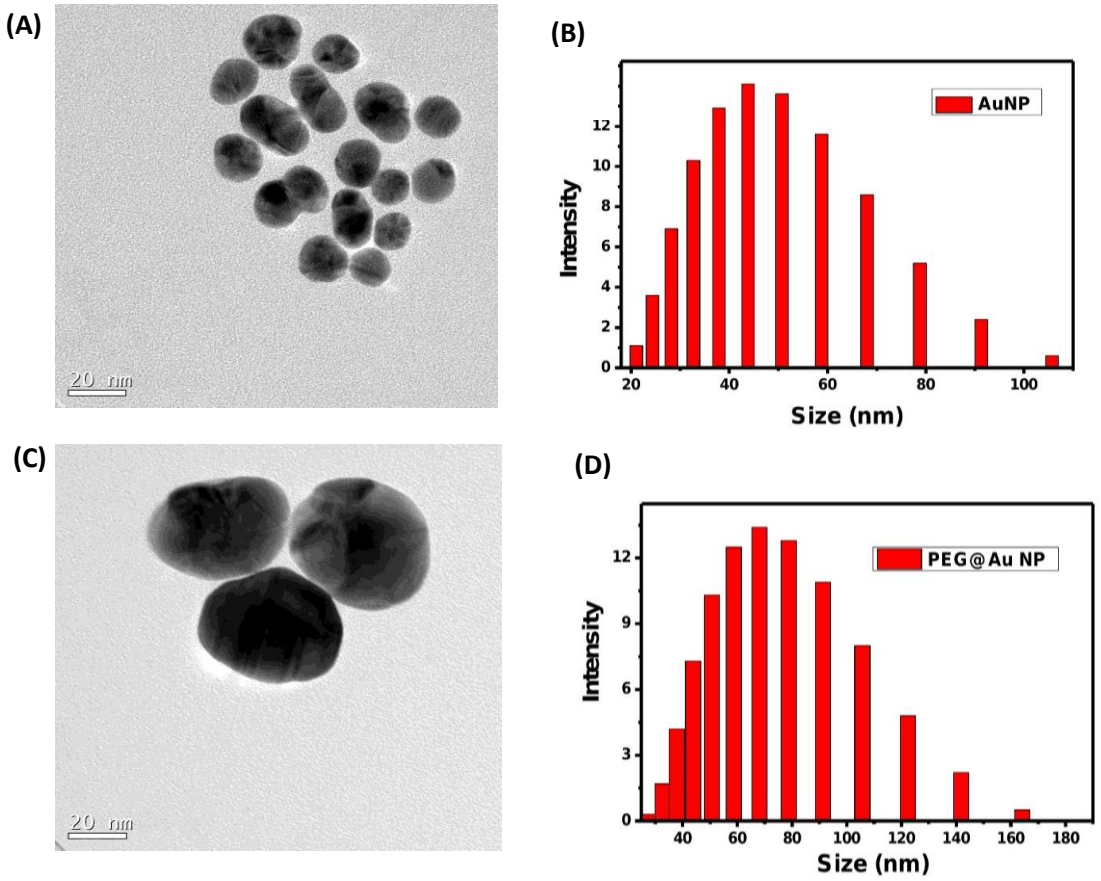


Figure 6.9. TEM and DLS analysis of gold nanoparticles

(A) TEM images of AuNPs with an average diameter of 45 nm; **(B)** DLS analysis of AuNPs; **(C)** TEM images of PEG encapsulated AuNPs with an average diameter of 66 nm; **(D)** DLS analysis of PEG encapsulated AuNPs

Triglycine has been synthesized via solid-phase peptide synthesis, conjugated to the amine terminal of NH_2-PEG@AuNP using carbodiimide chemistry. After the coupling of tripeptide with PEGylated AuNPs, the Fmoc was deprotected using 20 % piperidine in DMF to yield NH_2-terminal peptide conjugated AuNPs (GGG@PEG@AuNP). UV-Vis spectra of GGG@PEG@AuNP showed absorption bands centered at 260 nm and 532 nm **(Figure 6.10)**. This triglycine conjugated AuNPs was further used for enzyme immobilization.

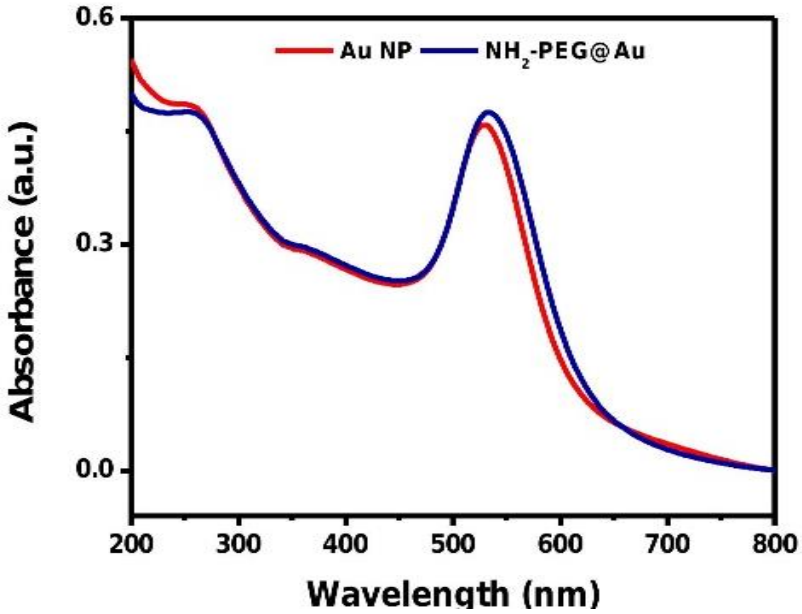


Figure 6.10. UV-Vis spectrum of AuNPs and PEG encapsulated AuNPs

6.3.4. Immobilization of eGFP and XylB on AuNPs using Sortase E

The two recombinant proteins eGFP and XylB, with LAHTG sequence at the C-terminus, were expressed in *E. coli* and purified by Ni-NTA columns. We employed calcium-independent $CgSrtE_{\Delta N44}$, for sortagging a C-terminal oriented immobilization of enhanced green fluorescent protein (eGFP-LAHTG) and xylose dehydrogenase enzyme (XylB-LAHTG) over triglycine functionalized PEGylated gold nanoparticles. $CgSrtE_{\Delta N44}$ cleaves between T and G at the C-terminal of the LAHTG tag of the recombinant proteins to form a new peptide bond between the carboxyl group of T with an amino group of triglycine PEGylated AuNPs **(Figure 6.11)**.

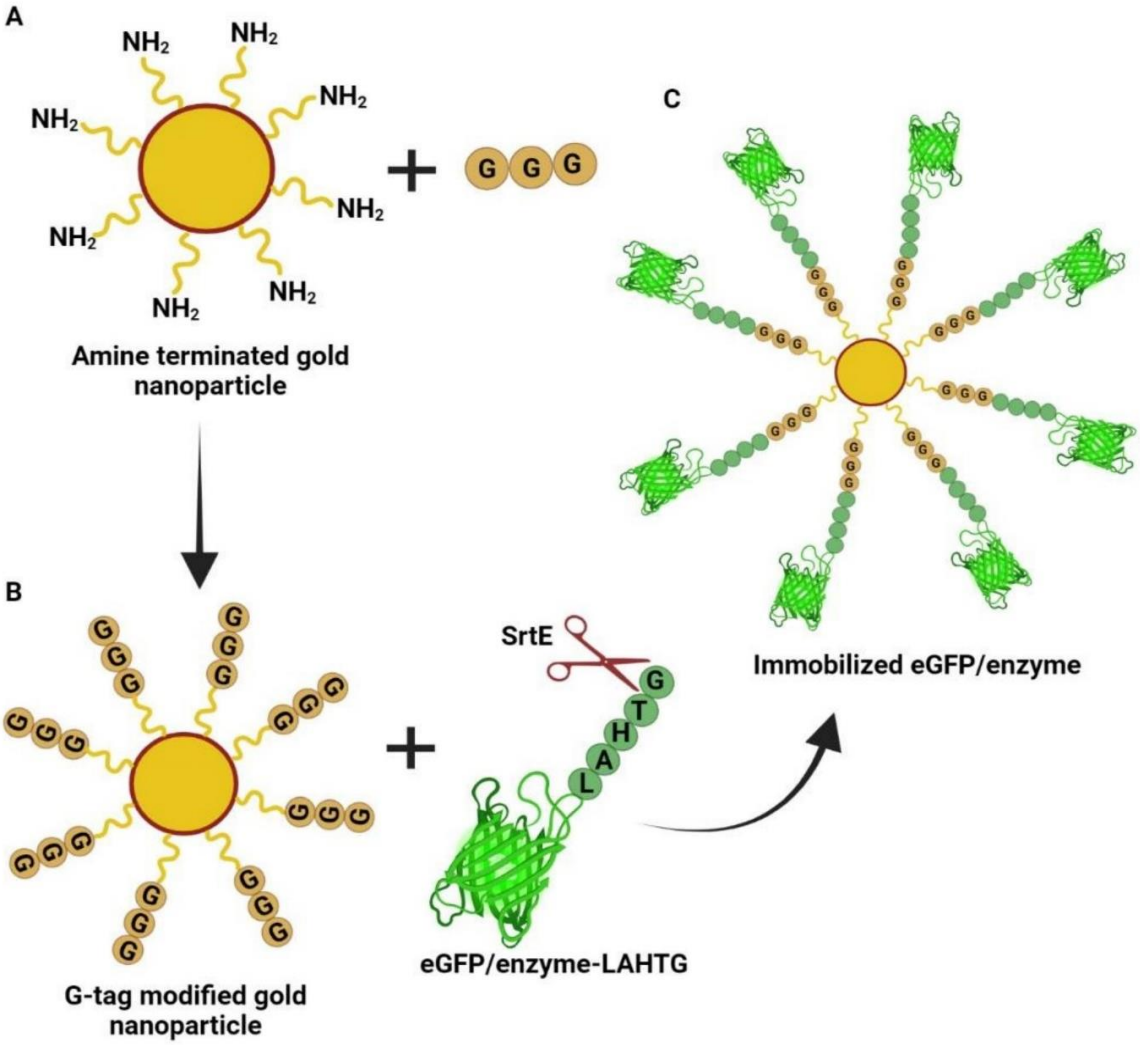


Figure 6.11. Site-specific Sortase E-mediated immobilization of enzymes

(A) Amine terminated PEGylated gold nanoparticle was synthesized; **(B)** The amine terminal of NH_2-PEG@AuNP was coupled with triglycine produced by solid-phase peptide synthesis; **(C)** The purified recombinant proteins with LAHTG tag were immobilized via Sortase E-mediated ligation on functionalized AuNP.

The conjugation of immobilized proteins (eGFP and XylB) was monitored by UV-Vis and Raman spectra. The UV-Vis spectrum of the immobilized protein showed a clear absorption peak of proteins at 265 nm and 532 nm. Thus, immobilization of protein over AuNPs is shown to be evident from the UV-Vis spectrum **(Figure 6.12)**.

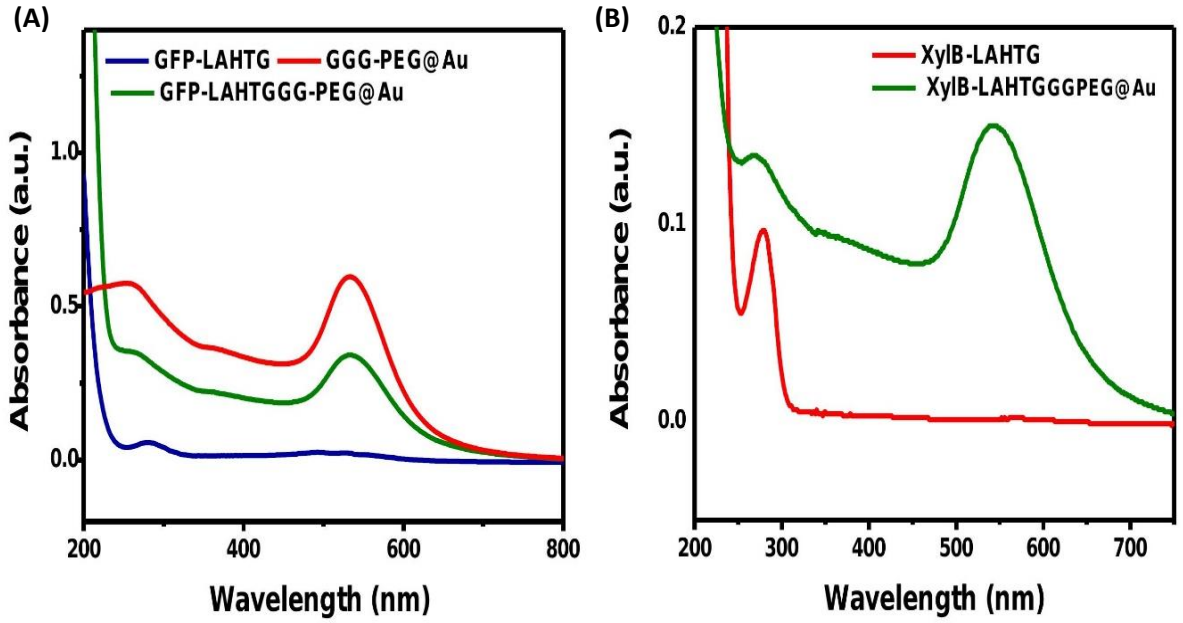


Figure 6.12. UV-Vis spectrum of eGFP/XylB LAHTG and immobilized eGFP/XylB-LAHTG

(A) Absorbance of GFP-LAHTG, GGG-PEG, & GFP-LAHTGGG@Au; **(B)** Absorbance of XylB-LAHTGGG-PEG@Au & XylB-LAHTG

Further, surface-enhanced Raman spectroscopy (SERS) was employed to confirm the enzyme immobilization using the sortase enzyme. The native Raman fingerprints of eGFP-LAHTG were characterized using SERS with prominent peaks at 826 cm^{-1}, 1305 cm^{-1}, 1405 cm^{-1}, and 1460 cm^{-1} **(Figure 6.13)**. These peaks correspond to the significant protein vibration bands, particularly the amide II band, CH_2 wagging, COO stretching, and CH_2/CH_3 stretching **(Table 6.2)**. We utilized these unique Raman patterns to examine the immobilized protein. The distinctive peaks of eGFP-LAHTG were found to be combined with the immobilized protein after purification. This appearance of characteristic peaks of eGFP-LAHTG in the immobilized protein confirms the effective immobilization process.

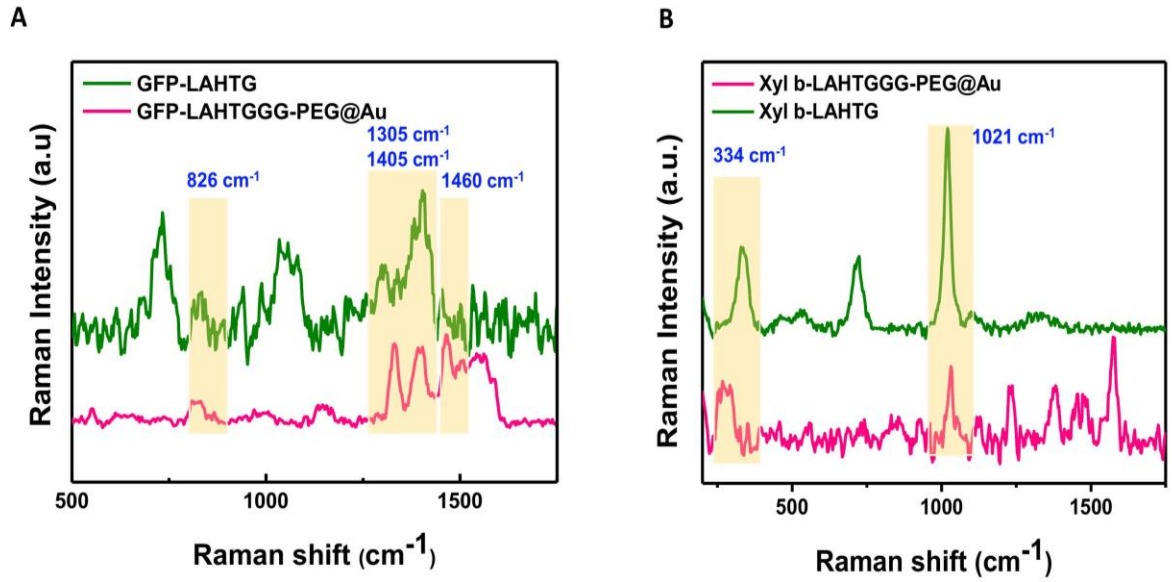


Figure 6.13. Raman spectrum of eGFP/XylB LAHTG and immobilized eGFP/XylB-LAHTG

(A) GFP-LAHTG & GFP-LAHTGGG@Au; **(B)** XylB-LAHTGGGG-PEG@Au & XylB-LAHTG

Table 6.2. The Raman spectra peaks and assignment

Raman spectroscopic peaks	Assignments
334	C-C vibrations in aliphatic chains
826	Protein bands
1021	C-O stretching
1305	d(CH2) twisting, wagging, collagen (protein assignment)
1405	V_s COO
1460	CH_2/CH_3 deformation

6.3.5. Evaluation of eGFP immobilization on AuNP

The relative fluorescence of free eGFP-LAHTG and eGFP-LAHTG immobilized on GGG@PEG@AuNP was evaluated by fluorescence spectroscopy at an excitation wavelength of 488 nm and emission wavelength of 509 nm after 2 h of incubation at 37 °C. After the conjugation of eGFP-LAHTG with AuNPs, the excess protein was removed by centrifugation.

The free GFP-LAHTG showed a high fluorescence intensity when compared to GFP-LAHTG immobilized on AuNPs **(Figure 6.14)**.

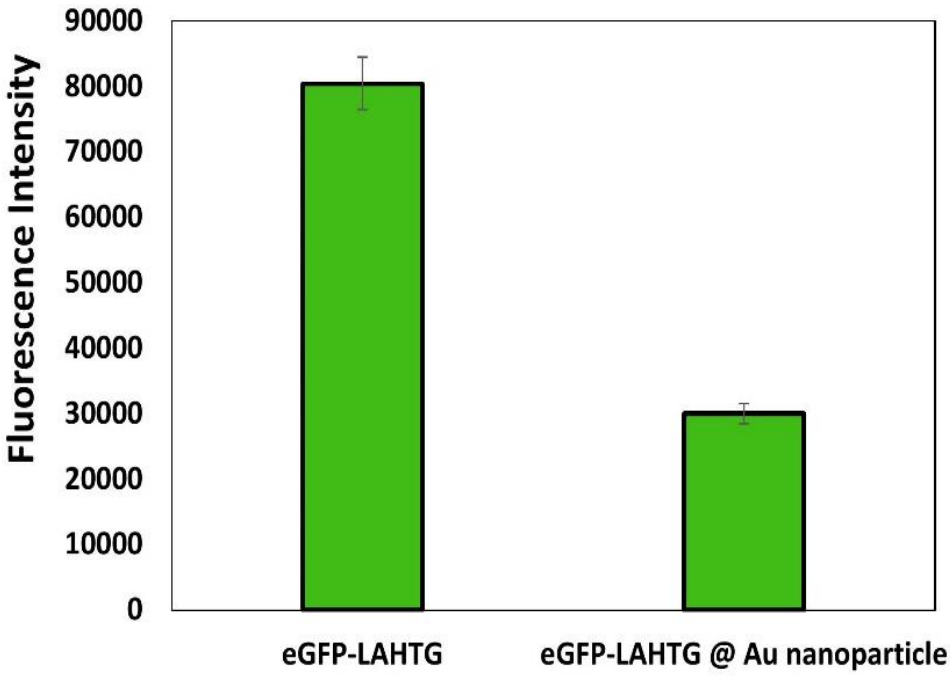


Figure 6.14. Validation of eGFP immobilization on AuNPs

Showing a high fluorescence intensity with free eGFP at excitation of 488 nm and emission of 509 nm when compared to immobilized eGFP-LAHTG on AuNP.

6.3.6. Analysis of xylonic acid production

XylB catalyzes the conversion of xylose to xylonic acid with the concomitant reduction of the NAD^+ cofactor, as presented in **Figure 6.15**.

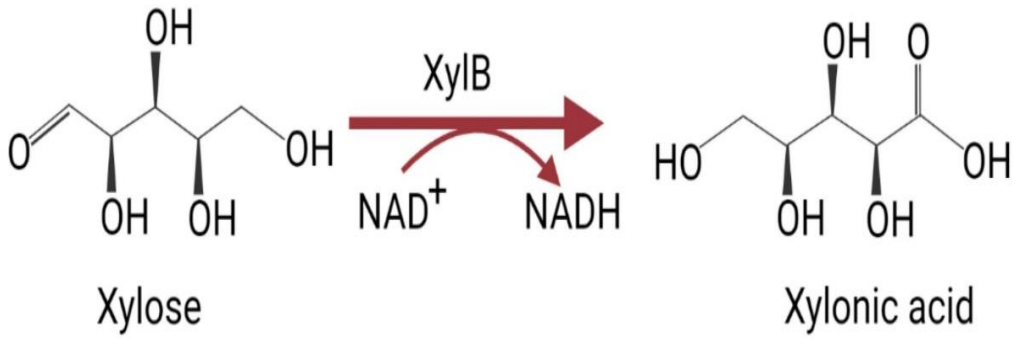


Figure 6.15. Bioconversion of xylose to xylonic acid

XylB catalyzes the conversion of xylose to xylonic acid while simultaneously reducing the NAD^+ cofactor.

HPLC was used to test the detection of xylonic acid generation from free and immobilized XylB with GGG@PEG@AuNP. During a 72-h incubation at 30 °C, the xylose utilization and xylonic acid consumption were evaluated. After 72 h, the free XylB utilizes xylose efficiently around 38.23 mM and produced a maximum of 2.7 mM of xylonic acid. Meanwhile, following immobilization, xylose utilization and xylonic acid generation were 38.96 mM and 1.2 mM, respectively, which were lower than the free enzyme **(Figure 6.16)**. This might be related to changes in the microenvironment of the enzyme after immobilization (Zdarta et al., 2020).

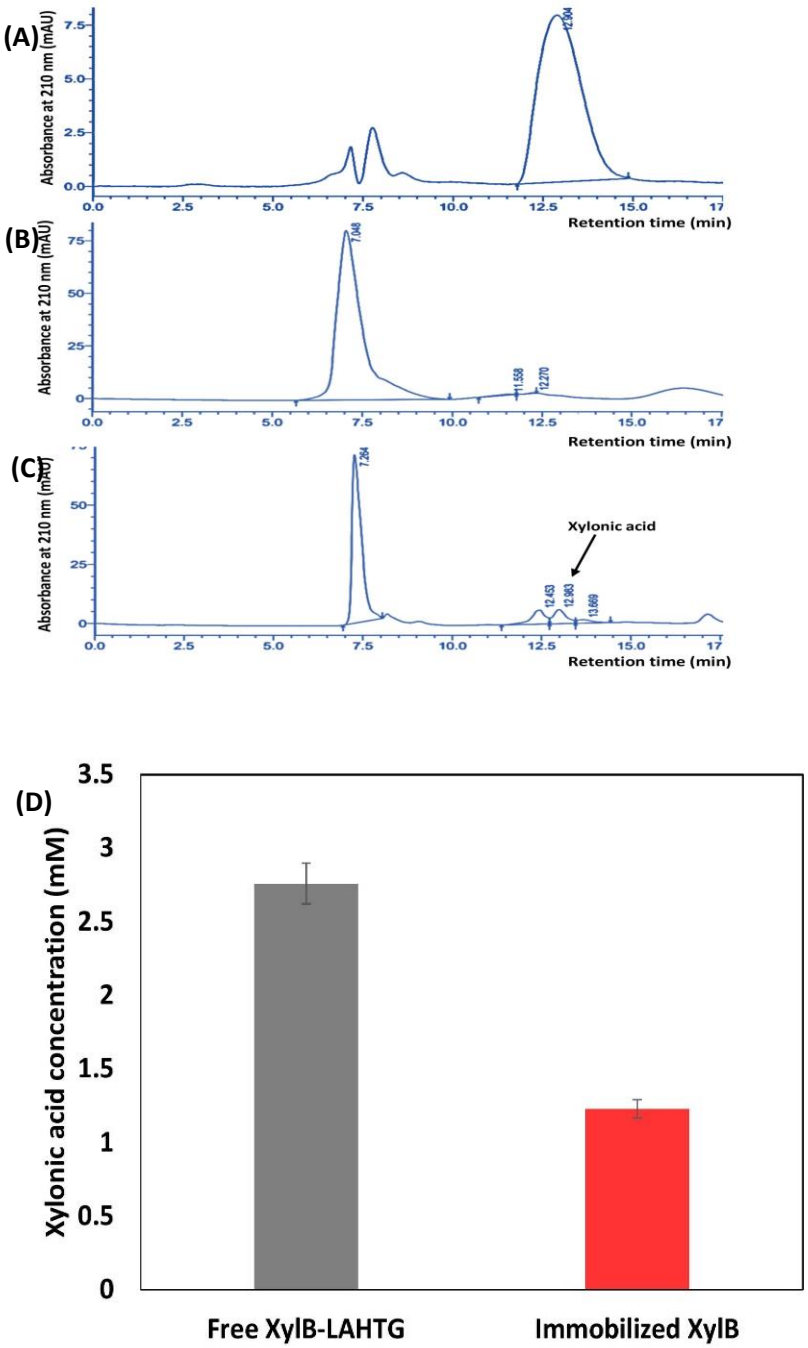


Figure 6.16. Validation of xylonic acid production by HPLC

(A) Xylonic acid standard; **(B)** Substrate control; **(C)** Production of xylonic acid; **(D)** Comparison of Xylonic acid production from free XylB and immobilized XylB, detected by HPLC.

6.3.7. Reusability of the immobilized XylB

The reusability of the immobilized XylB was examined based on the conversion efficiency of xylose to xylonic acid via HPLC. Each cycle was carried out for 3 days at 30 °C,

centrifuged at 4000 x g 20 min at 4 °C, and stored for 1 week to repeat the next cycle. Thus, the enzyme immobilized by site-specific oriented ligation using Sortase E was found to be stable and retain over 80 % reaction yield after four consecutive cycles **(Figure 6.17)**.

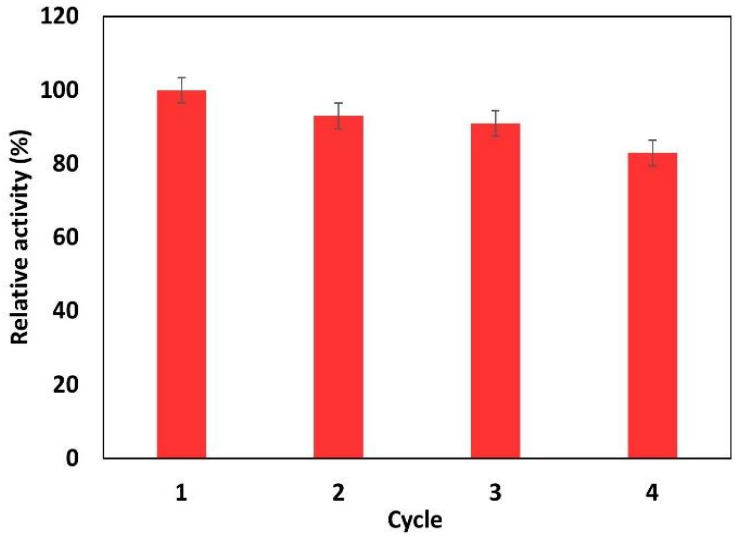


Figure 6.17. The efficiency of the immobilized XylB on repeated cycle use

6.4. Discussion

Enzymatic approaches on site-specific immobilization are gaining considerable interest due to their substrate-binding affinity. Sortase A, a transpeptidase from *S. aureus*, cleaves between the T and G residues in the sequence LPXTG and then forms a native peptide bond between the carboxyl group of the T residue and an amino group of N-terminal glycine oligomers (Ton-That et al., 1999). Although, there are several types of sortases found in Gram-positive bacteria, little is known about their expression and substrate specificity. The Ca^{2+}- independent Sortase E of *C. glutamicum* with a high substrate specificity towards the LAHTG motif was used to immobilize proteins on AuNPs. The Sortase E-mediated surface immobilization on a PEGylated AuNPs was performed under mild reaction conditions in a single step without the supplement of Ca^{2+} ions and without any prior chemical modification. The site-specific covalent attachment enables the protein to be arranged in an ordered fashion

without affecting its biological activity. This was similar to the reports performed by Sortase A on immobilizing α-amylase and β-glucosidase on polystyrene nanoparticles. The catalytic activity of the immobilized α-amylase or β-glucosidase particles was higher than that of the corresponding proteins immobilized by chemical cross-linking (Hata et al., 2015). The immobilized enzyme on AuNPs produced via Sortase E-mediated method exhibited higher reusability (3-4 cycles with more than 80 % activity) and enzyme stability. Since, chemically immobilized enzymes may be more susceptible to denaturation during recycling due to binding of particles at the active site residue and thereby lowering its reusability (Cho et al., 2012).

6.5. Summary

Immobilization on gold nanoparticles is a versatile method to improve the efficacy, recyclability, and stability of the enzymes for industrial application. In this chapter, we used *Corynebacterium glutamicum* Sortase E, which is calcium-independent and has high substrate specificity towards the LAHTG sequence, for sortase-mediated ligation (sortagging). Using an engineered Sortase E-mediated ligation, a pioneering attempt was made on C-terminal oriented immobilization of enhanced green fluorescent protein (eGFP) and xylose dehydrogenase enzyme (XylB) on PEGylated gold nanoparticles (AuNPs). Surface-Enhanced Raman Scattering (SERS) and UV-Vis spectroscopy were used to confirm the successful enzyme/protein conjugation using triglycine functionalized AuNPs. The immobilized XylB was catalytically efficient and able to retain >80 % of its initial activity after four consecutive cycles. Furthermore, the isolation and purification SrtE from a non-pathogenic *C. glutamicum* was easier than isolating and purifying SrtA from organisms that require extensive biosafety procedures. As a result, CgSrtE can be utilized instead of SrtA in bioengineering or biotechnological applications.

Chapter 7

Summary and Conclusion

The transpeptidation reaction catalyzed by sortases, particularly sortase A from *S. aureus* (SaSrtA), has been intensively investigated in recent years, and it is increasingly being used in the synthesis of new protein derivatives and challenging conjugation processes in vitro as well as in cellulo in the presence of Ca^{2+}. Several ways have been developed to improve the sortase performance and broaden the spectrum of substrates that can be used with a sortase-based strategy. However, for the time being, the vast majority of protocols rely on the well-studied *S. aureus* sortase A (SaSrtA) and its derived forms. Although, evolved variants of SaSrtA outperform the wild-type enzyme in terms of reaction speeds and new research suggests that they aren't ideal for all applications. To facilitate intracellular sortagging applications, Ca^{2+}-independent sortase activity is desirable. However, other sortases, except SaSrtA, are not being employed for sortagging applications due to their low activity. However, based on the genome analysis, *C. glutamicum* ATCC 13032 enabled to identify of a single class E sortase and prediction of a single sortase-dependent protein (Cgl0614) with the LAATG motif at the C-terminal of the CWSS in *C. glutamicum*. The biochemical characterization of the enzyme describes the ideal LAXTG substrate, Ca^{2+}-independency, and catalytic efficiency which indicates the possible functional differences between CgSrtE and SaSrtA. The key residues that influence the preference for a non-canonical LAXTG recognition motif over LPXTG have been found by structural modeling, substrate docking, and mutagenesis studies. This is a kind of the first report on the sortase family of transpeptidase enzyme from a nonpathogenic *C. glutamicum* which makes them more friendly for several applications like sortagging.

The sortase E was found to be conserved among the genomes of *Corynebacterium* sp. and function as a housekeeping gene within the organism. The Cys 240, Arg 249, His 135, and Tyr 118 were found to be conserved in the CgSrtE binding site. The SWISS-MODEL server produced the 3D structure of CgSrtE, which revealed a 62 % similarity to CdSrtF in the PDB database. Cgl0614, a surface protein predicted to function as a transporter in *C. glutamicum*, was built using the I-TASSER threading approach and meets all of the criteria for sortase substrate protein: an N-terminal signal peptide, a C-terminal sorting motif LAATG, a transmembrane domain, and a positively charged tail. All the generated models were validated using PROCHECK and ProSA. The in-silico protein-protein docking by ClusPro identified Cgl0614 sorting motif LAATG predicted to interact with catalytic site residues of *C. glutamicum* Sortase E protein.

Three different derivatives such as deletion, complement, and overexpression of *srtE* gene of *C. glutamicum* were constructed to monitor the possible impact of this gene on cell growth, morphology, and physiological properties. The *srtE* deletion mutant was constructed in *C. glutamicum* ATCC 13032 by double-crossover homologous recombination. The identity of *srtE* was verified by constructing a complement, by inserting a native *srtE* expression cassette into the pVWEx1 plasmid. Homologous expression of the *srtE* gene was achieved in *C. glutamicum* with the help of the pVWEx1 plasmid. The gene was cloned under IPTG inducible P_{tac} promoter of the plasmid for controlled expression. The absence of the *srtE* gene did not alter the growth rate or morphology in *C. glutamicum* but resulted in decreased cell surface hydrophobicity. When compared to WT *C. glutamicum*, the overexpressed gene resulted in slowed growth and elongated cell morphology. The growth recovery for Δ*srtE* deletion mutant was 20 % lower than the wild-type *C. glutamicum* but found to be viable when temperature shifted from 30 °C to 37 °C. However, in the overexpressed strain (p*srtE*) the growth was completely retarded with the rise in temperature.

Heterologous expression of sortase gene from *C. glutamicum* was expressed in *E. coli* as a soluble fraction with the removal of 1-44 aa from the N-terminal transmembrane anchor domain of the protein. Expression was carried out after cloning the gene to the pET28a vector and the his-tagged recombinant protein (MW 29.6 kDa) was designated as $CgSrtE_{\Delta N44}$. The cleavage activity of Sortase E was measured in vitro by HPLC and FRET using synthetic peptides. $CgSrtE_{\Delta N44}$ cleaved SrtA substrate Abz-LAHTG-Dap (Dnp) with an apparent K_m of 0.012 mM but failed to cleave Abz-LPETG-Dap (Dnp), indicating that it belongs to class E. The absence of calcium-binding sites in $CgSrtE_{\Delta N44}$ shows that Ca^{2+} is not necessary for sortase E activity enhancement in *C. glutamicum*. Furthermore, the enzyme was active throughout a wide pH range of 7.5–10.2, with an optimal pH of about 9.5. The enzyme catalytic performance was increased when incubated at a higher temperature of roughly 60 °C, and an enzyme concentration of around 15 µM was determined to be ideal. Site-directed mutagenesis of the active site residues of $CgSrtE_{\Delta N44}$ resulted in a 50 % reduction in activity when compared to wild-type, indicating that enzyme is essential for substrate recognition and catalysis.

Demonstrated for the first time the C-terminal mediated sortagging of eGFP and XylB on PEGylated AuNPs by CgSrtE under mild reaction conditions in a single step without the supplement of Ca^{2+} ions or any chemical modification. The resulting recombinant eGFP and XylB from *Caulobacter cresentus* were expressed in *E. coli* with LAHTG-tag at the C-terminal. eGFP-LAHTG and XylB-LAHTG were recombinantly purified by Ni-NTA columns and immobilized separately on a functionalized triglycine gold nanoparticle by SrtE-mediated technique and sortagging was confirmed by SERS and UV-Vis spectral analysis. The effect of XylB oriented sortagging on its activity and reusability were evaluated by comparing free enzymes. The immobilized XylB was found to be catalytically efficient and able to retain >80 % of its initial activity after four consecutive cycles.

To conclude, the present study is about the identification of a novel class E sortase of non-pathogenic *C. glutamicum*, that shows a substrate specificity different from that of class A sortases, and understanding its biological activity at the molecular level requires a clear understanding of protein structure. However, X-ray crystallography and NMR spectroscopy were used to identify the protein structure, albeit both methods take time and aren't always effective, especially with membrane proteins. So, in silico research was used to generate the structure of sortase and its substrate protein and explained by structural analysis and computation. The function of sortase as a housekeeping role without affecting the cell growth and morphology was explained by gene deletion. The Ca^{2+} ion-independent activity and ability to withstand a higher temperature can be utilized as an alternative to SaSrtA. Furthermore, isolating and purifying SrtE from a non-pathogenic *C. glutamicum* was determined to be easier than isolating and purifying SrtA from organisms that need significant biosafety procedures. As a result, CgSrtE can be used as a potential candidate for broadening the capabilities of site-specific modification of protein/enzyme in various bioengineering applications which includes protein dimerization, circularization of proteins to enhance the catalytic activity and stability of the proteins, and site-specifically modifying liposomes with proteins.

Bibliography

Abe, S., Takayama, K. I., and Kinoshita, S. (1967). Taxonomical studies on glutamic acid-producing bacteria. *J. Gen. Appl. Microbiol.* 13, 279–301. doi:10.2323/jgam.13.279.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2.

Anderl, A., Ferlemann, C., Muth, M., Henkel-Gupalo, A., Ebenig, A., Brenner-Weiß, G., et al. (2019). Biochemical study of sortase E2 from *Streptomyces mobaraensis* and determination of transglutaminase cross-linking sites. *FEBS Lett.* 593, 1944–1956. doi:10.1002/1873-3468.13466.

Andreou, L. V. (2013). Preparation of genomic DNA from bacteria. 1st ed. Elsevier Inc. doi:10.1016/B978-0-12-418687-3.00011-2.

Ansari, S. A., and Husain, Q. (2012). Potential applications of enzymes immobilized on/in nano materials: A review. *Biotechnol. Adv.* 30, 512–523. doi:10.1016/j.biotechadv.2011.09.005.

Antos, J. M., Chew, G. L., Guimaraes, C. P., Yoder, N. C., Grotenbreg, G. M., Popp, M. W. L., et al. (2009). Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc.* 131, 10800–10801. doi:10.1021/ja902681k.

Antos, J. M., Truttmann, M. C., and Ploegh, H. L. (2016). Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* 38, 111–118. doi:10.1016/j.sbi.2016.05.021.

Aucher, W., Davison, S., and Fouet, A. (2011). Characterization of the Sortase Repertoire in *Bacillus anthracis*. *PLoS One* 6, e27411. doi:10.1371/journal.pone.0027411.

Bantz, K. C., Meyer, A. F., Wittenberg, N. J., Im, H., Kurtuluş, Ö., Lee, S. H., et al. (2011). Recent progress in SERS biosensing. *Phys. Chem. Chem. Phys.* 13, 11551–11567.

doi:10.1039/c0cp01841d.

Bateman, A. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi:10.1093/nar/gky1049.

Becker, J., and Wittmann, C. (2012). Bio-based production of chemicals, materials and fuels - *Corynebacterium glutamicum* as versatile cell factory. *Curr. Opin. Biotechnol.* 23, 631–640. doi:10.1016/j.copbio.2011.11.012.

Beckmann, C., Waggoner, J. D., Harris, T. O., Tamura, G. S., and Rubens, C. E. (2002). Identification of Novel Adhesins from Group B Streptococci by Use of Phage Display Reveals that C5a Peptidase Mediates Fibronectin Binding. 70, 2869–2876. doi:10.1128/IAI.70.6.2869.

Benkert, P., Künzli, M., and Schwede, T. (2009). QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 37, 510–514. doi:10.1093/nar/gkp322.

Bentley, M. L., Lamb, E. C., and McCafferty, D. G. (2008). Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from *Staphylococcus aureus*. *J. Biol. Chem.* 283, 14762–14771. doi:10.1074/jbc.M800974200.

Bierne, H., Mazmanian, S. K., Trost, M., Pucciarelli, M. G., Liu, G., Dehoux, P., et al. (2002). Inactivation of the srtA gene in Listeria monocytogenes inhibits anchoring of surface proteins and affects virulence. *Mol. Microbiol.* 43, 869–881. doi:10.1046/j.1365-2958.2002.02798.x.

Bloom, B. R., and Murray, C. J. L. (1992). Tuberculosis: Commentary on a reemergent killer. *Science (80-.).* 257, 1055–1064. doi:10.1126/science.257.5073.1055.

Boekhorst, J., De Been, M. W. H. J., Kleerebezem, M., and Siezen, R. J. (2005). Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J. Bacteriol.* 187, 4928–4934. doi:10.1128/JB.187.14.4928-4934.2005.

Bornscheuer, U. T. (2003). Immobilizing enzymes: How to create more suitable biocatalysts.

Angew. Chemie - Int. Ed. 42, 3336–3337. doi:10.1002/anie.200301664.

Bradshaw, W. J., Davies, A. H., Chambers, C. J., Roberts, A. K., Shone, C. C., and Acharya, K. R. (2015). Molecular features of the sortase enzyme family. *FEBS J.* 282, 2097–2114. doi:10.1111/febs.13288.

Budzik, J. M., Marraffini, L. A., and Schneewind, O. (2007). Assembly of pili on the surface of Bacillus cereus vegetative cells. *Mol. Microbiol.* 66, 495–510. doi:10.1111/j.1365-2958.2007.05939.x.

Budzik, J. M., Oh, S.-Y., and Schneewind, O. (2009). Sortase D forms the covalent bond that links BcpB to the tip of *Bacillus cereus* pili. *J. Biol. Chem.* 284, 12989–97. doi:10.1074/jbc.M900927200.

Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., et al. (2019). RCSB Protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* 47, D464–D474. doi:10.1093/nar/gky1004.

Call, E. K., Goh, Y. J., Selle, K., Klaenhammer, T. R., and O'Flaherty, S. (2015). Sortase-deficient lactobacilli: Effect on immunomodulation and gut retention. *Microbiol. (United Kingdom)* 161, 311–321. doi:10.1099/mic.0.000007.

Call, E. K., and Klaenhammer, T. R. (2013). Relevance and application of sortase and sortase-dependent proteins in lactic acid bacteria. *Front. Microbiol.* 4, 1–10. doi:10.3389/fmicb.2013.00073.

Cambria, E., Renggli, K., Ahrens, C. C., Cook, C. D., Kroll, C., Krueger, A. T., et al. (2015). Covalent Modification of Synthetic Hydrogels with Bioactive Proteins via Sortase-Mediated Ligation. *Biomacromolecules* 16, 2316–2326. doi:10.1021/acs.biomac.5b00549.

Cardoso, J. M. S., Fonseca, L., Egas, C., and Abrantes, I. (2018). Cysteine proteases secreted

by the pinewood nematode, *Bursaphelenchus xylophilus*: In silico analysis. *Comput. Biol. Chem.* 77, 291–296. doi:10.1016/j.compbiolchem.2018.10.011.

Cascioferro, S., Totsika, M., and Schillaci, D. (2014). Sortase A: An ideal target for anti-virulence drug development. *Microb. Pathog.* 77, 105–112. doi:10.1016/j.micpath.2014.10.007.

Chambers, C. J., Roberts, A. K., Shone, C. C., and Acharya, K. R. (2015). Structure and function of a *Clostridium difficile* sortase enzyme. *Sci. Rep.* 5, 1–11. doi:10.1038/srep09449.

Chan, L., Cross, H. F., She, J. K., Cavalli, G., Martins, H. F. P., and Neylon, C. (2007). Covalent attachment of proteins to solid supports and surfaces via sortase-mediated ligation. *PLoS One* 2, 1–5. doi:10.1371/journal.pone.0001164.

Chang, C., Mandlik, A., Das, A., and Ton-That, H. (2011). Cell surface display of minor pilin adhesins in the form of a simple heterodimeric assembly in *Corynebacterium diphtheriae*. *Mol. Microbiol.* 79, 1236–1247. doi:10.1111/j.1365-2958.2010.07515.x.

Chapman, D. (1975). Phase transitions and fluidity characteristics of lipids and cell membranes. *Q. Rev. Biophys.* 8, 185–235. doi:10.1017/S0033583500001797.

Chen, F., Xie, F., Yang, B., Wang, C., Liu, S., and Zhang, Y. (2017). *Streptococcus suis* sortase A is Ca 2+ independent and is inhibited by acteoside, isoquercitrin and baicalin. *PLoS One* 12, 1–15. doi:10.1371/journal.pone.0173767.

Cheng, A. G., Kim, H. K., Burts, M. L., Krausz, T., Schneewind, O., and Missiakas, D. M. (2009). Genetic requirements for *Staphylococcus aureus* abscess formation and persistence in host tissues. *FASEB J.* 23, 3393–3404. doi:10.1096/fj.09-135467.

Cho, E. J., Jung, S., Kim, H. J., Lee, Y. G., Nam, K. C., Lee, H. J., et al. (2012). Co-immobilization of three cellulases on Au-doped magnetic silica nanoparticles for the degradation of cellulose. *Chem. Commun.* 48, 886–888. doi:10.1039/c2cc16661e.

Chung, C. T., and Miller, R. H. (1993). Preparation and Storage of Competent *Escherichia coli* Cells. *Methods Enzymol.* 218, 621–627. doi:10.1016/0076-6879(93)18045-E.

Clancy, K. W., Melvin, J. A., and McCafferty, D. G. (2010). Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers* 94, 385–396. doi:10.1002/bip.21472.

Clustal, W. (1994). improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice Thompson, Julie D.; Higgins, Desmond G.; Gibson, Toby J. *Nucleic Acids Res.* 22, 4673–4680. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC308517/.

Colovos, C., and Yeates, T. (2020). Verification of protein structures : Patterns of nonbonded atomic interactions. 1511–1519. doi:10.1002/pro.5560020916.

Comfort, D., and Clubb, R. T (2004). A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. *Infect. Immun.* 72, 2710–2722. doi:10.1128/IAI.72.5.2710.

Corver, J., Cordo', V., van Leeuwen, H. C., Klychnikov, O. I., and Hensbergen, P. J. (2017). Covalent attachment and Pro-Pro endopeptidase (PPEP-1)-mediated release of *Clostridium difficile* cell surface proteins involved in adhesion. *Mol. Microbiol.* 105, 663–673. doi:10.1111/mmi.13736.

Cozzi, R., Malito, E., Nuccitelli, A., D'Onofrio, M., Martinelli, M., Ferlenghi, I., et al. (2011). Structure analysis and site-directed mutagenesis of defined key residues and motives for pilus-related sortase C1 in group B Streptococcus. *FASEB J.* 25, 1874–1886. doi:10.1096/fj.10-174797.

Cuff, J. A., and Barton, G. J. (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins Struct. Funct. Genet.* 40, 502–511. doi:10.1002/1097-0134(20000815)40:3<502::AID-PROT170>3.0.CO;2-Q.

Dai, X., Böker, A., and Glebe, U. (2019). Broadening the scope of sortagging. *RSC Adv.* 9, 4700–4721. doi:10.1039/c8ra06705h.

Das, S., Pawale, V. S., Dadireddy, V., Singh, A. K., Ramakumar, S., and Roy, R. P. (2017). Structure and specificity of a new class of Ca^{2+}-independent housekeeping sortase from *Streptomyces avermitilis* provide insights into its non-canonical substrate preference. *J. Biol. Chem.* 292, 7244–7257. doi:10.1074/jbc.M117.782037.

Dieye, Y., Oxaran, V., Ledue-Clier, F., Alkhalaf, W., Buist, G., Juillard, V., et al. (2010). Functionality of sortase a in *Lactococcus lactis*. *Appl. Environ. Microbiol.* 76, 7332–7337. doi:10.1128/AEM.00928-10.

Divya, J. B., Varsha, K. K., and Nampoothiri, K. M. (2012). Newly isolated lactic acid bacteria with probiotic features for potential application in food industry. *Appl. Biochem. Biotechnol.* 167, 1314–1324. doi:10.1007/s12010-012-9561-7.

Donahue, E. H., Dawson, L. F., Valiente, E., Firth-Clark, S., Major, M. R., Littler, E., et al. (2014). *Clostridium difficile* has a single sortase, SrtB, that can be inhibited by small-molecule inhibitors. *BMC Microbiol.* 14, 1–14. doi:10.1186/s12866-014-0219-1.

Dover, L. G., Cerdeño-Tárraga, A. M., Pallen, M. J., Parkhill, J., and Besra, G. S. (2004). Comparative cell wall core biosynthesis in the mycolated pathogens, *Mycobacterium tuberculosis* and *Corynebacterium diphtheriae*. *FEMS Microbiol. Rev.* 28, 225–250. doi:10.1016/j.femsre.2003.10.001.

Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prévost, M., Kojadinovic, M., et al. (2006). Assembly and role of pili in group B streptococci. doi:10.1111/j.1365-2958.2006.05190.x.

Dramsi, S., Trieu-Cuot, P., and Bierne, H. (2005). Sorting sortases: A nomenclature proposal for the various sortases of Gram-positive bacteria. *Res. Microbiol.* 156, 289–297. doi:10.1016/j.resmic.2004.10.011.

Duong, A., Capstick, D. S., Berardo, C. Di, Findlay, K. C., Hesketh, A., Hong, H., et al. (2012). Aerial development in *Streptomyces coelicolor* requires sortase activity. 83, 992–1005. doi:10.1111/j.1365-2958.2012.07983.x.

Egan, S. A., Kurian, D., Ward, P. N., Hunt, L., and Leigh, J. A. (2010). Identification of Sortase A (SrtA) Substrates in *Streptococcus uberis* : Evidence for an Additional Hexapeptide (LPXXXD) Sorting Motif research articles. 1088–1095.

Elliot, M. A., Karoonuthaisiri, N., Huang, J., Bibb, M. J., Cohen, S. N., Kao, C. M., et al. (2003). The chaplins : a family of hydrophobic cell-surface proteins involved in aerial mycelium formation in *Streptomyces coelicolor*. 1727–1740. doi:10.1101/gad.264403.Filamentous.

Emami Bistgani, Z., Siadat, S. A., Bakhshandeh, A., Ghasemi Pirbalouti, A., and Hashemi, M. (2017). Interactive effects of drought stress and chitosan application on physiological characteristics and essential oil yield of *Thymus daenensis* Celak. *Crop J.* 5, 407–415. doi:10.1016/j.cj.2017.04.003.

Fischetti, V. A., Pancholi, V., and Schneewind, O. (1990). Conservation of a hexapeptide sequence in the anchor region of surface proteins from Gram-positive cocci. *Mol. Microbiol.* 4, 1603–1605. doi:10.1111/j.1365-2958.1990.tb02072.x.

Fiser, A. (2010). Chapter 6. 673. doi:10.1007/978-1-60761-842-3.

Fittipaldi, N., Segura, M., Grenier, D., and Gottschalk, M. (2012). Virulence factors involved in the pathogenesis of the infection caused by the swine pathogen and zoonotic agent *Streptococcus suis*. *Future Microbiol.* 7, 259–279. doi:10.2217/fmb.11.149.

Foster, T. J., Geoghegan, J. A., Ganesh, V. K., and Höök, M. (2014). Adhesion, invasion and evasion: The many functions of the surface proteins of *Staphylococcus aureus*. *Nat. Rev. Microbiol.* 12, 49–62. doi:10.1038/nrmicro3161.

Frankel, B. A., Kruger, R. G., Robinson, D. E., Kelleher, N. L., and McCafferty, D. G.

(2005). *Staphylococcus aureus* sortase transpeptidase SrtA: Insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* 44, 11188–11200. doi:10.1021/bi050141j.

Frankel, B. B., Tong, Y., Bentley, M. M. L., Fitzgerald, M. C., and McCafferty, D. G. (2007). Mutational analysis of active site residues in the *Staphylococcus aureus* transpeptidase SrtA. *Biochemistry* 46, 7269–7278. doi:10.1021/bi700448e.

Garandeau, C., Pucciarelli, M. G., Sabet, C., Newton, S., Portillo, F. G., Cossart, P., et al. (2004). Sortase B , a New Class of Sortase in *Listeria monocytogenes*. 186, 1972–1982. doi:10.1128/JB.186.7.1972.

Gaspar, A. H., Marraffini, L. A., Glass, E. M., DeBord, K. L., Ton-That, H., and Schneewind, O. (2005). *Bacillus anthracis* Sortase A (SrtA) Anchors LPXTG Motif-Containing Surface Proteins to the Cell Wall Envelope. *J. Bacteriol.* 187, 4646–4655. doi:10.1128/JB.187.13.4646-4655.2005.

Gaspar, A. H., and Ton-That, H. (2006). Assembly of Distinct Pilus Structures on the Surface of *Corynebacterium diphtheriae*. 188, 1526–1533. doi:10.1128/JB.188.4.1526.

Geourjon, C., and Deléage, G. (1995). Sopma: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 11, 681–684. doi:10.1093/bioinformatics/11.6.681.

Girolamo, S. Di., Puorger, C., Castiglione, M., Vogel, M., Gébleux, R., Briendl, M., et al. (2019). Characterization of the housekeeping sortase from the human pathogen *Propionibacterium acnes*: First investigation of a class F sortase. *Biochem. J.* 476, 665–682. doi:10.1042/BCJ20180885.

Glaser, P., Rusniok, C., Buchrieser, C., Chevalier, F., Frangeul, L., Msadek, T., et al. (2002). Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. 45, 1499–1513.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557–580. doi:10.1016/S0022-2836(83)80284-8.

Hata, Y., Matsumoto, T., Tanaka, T., and Kondo, A. (2015). C-Terminal-oriented Immobilization of Enzymes Using Sortase A-mediated Technique. *Macromol. Biosci.* 15, 1375–1380. doi:10.1002/mabi.201500113.

Hemmerich, J., Moch, M., Jurischka, S., Wiechert, W., Freudl, R., and Oldiges, M. (2019). Combinatorial impact of Sec signal peptides from Bacillus subtilis and bioprocess conditions on heterologous cutinase secretion by *Corynebacterium glutamicum*. *Biotechnol. Bioeng.* 116, 644–655. doi:10.1002/bit.26873.

Hensbergen, P. J., Klychnikov, O. I., Bakker, D., Dragan, I., Kelly, M. L., Minton, N. P., et al. (2015). *Clostridium difficile* secreted Pro-Pro endopeptidase PPEP-1 (ZMP1/CD2830) modulates adhesion through cleavage of the collagen binding protein CD2831. *FEBS Lett.* 589, 3952–3958. doi:10.1016/j.febslet.2015.10.027.

Hess, G. T., Cragnolini, J. J., Popp, M. W., Allen, M. A., Dougan, S. K., Spooner, E., et al. (2012). M13 Bacteriophage Display Framework That Allows Sortase- Mediated Modi fi cation of Surface-Accessible Phage Proteins.

Homaei, A. A., Sariri, R., Vianello, F., and Stevanato, R. (2013). Enzyme immobilization: An update. *J. Chem. Biol.* 6, 185–205. doi:10.1007/s12154-013-0102-9.

Ikeda, M., and Takeno, S. (2013). Amino Acid Production by Corynebacterium glutamicum. doi:10.1007/978-3-642-29857-8_4.

Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R. T. (2001). Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6056–6061. doi:10.1073/pnas.101064198.

Jacobitz, A. W., Kattke, M. D., Wereszczynski, J., and Clubb, R. T. (2017). Sortase Transpeptidases: Structural Biology and Catalytic Mechanism. 1st ed. Elsevier Inc.

doi:10.1016/bs.apcsb.2017.04.008.

Jacobitz, A. W., Wereszczynski, J., Yi, S. W., Amer, B. R., Huang, G. L., Nguyen, A. V., et al. (2014). Structural and computational studies of the *Staphylococcus aureus* sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. *J. Biol. Chem.* 289, 8891–8902. doi:10.1074/jbc.M113.509273.

Jager, W., Schafer, A., Puhler, A., Labes, G., and Spring, S. (1992). ~ E. 174, 5462–5465.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices. *Comput Appl Biosci.* 8, 275–282. doi:doi.org/10.1093/bioinformatics/8.3.275.

Josefsson, E., Kubica, M., Mydel, P., Potempa, J., and Tarkowski, A. (2008). In vivo sortase A and clumping factor A mRNA expression during *Staphylococcus aureus* infection. *Microb. Pathog.* 44, 103–110. doi:10.1016/j.micpath.2007.08.010.

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., et al. (2003). The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J. Biotechnol.* 104, 5–25. doi:10.1016/S0168-1656(03)00154-8.

Kattke, M. D., Chan, A. H., Duong, A., Sexton, D. L., Sawaya, M. R., Cascio, D., et al. (2016). Crystal structure of the *Streptomyces coelicolor* sortase E1 transpeptidase provides insight into the binding mode of the novel class e sorting signal. *PLoS One* 11, 1–21. doi:10.1371/journal.pone.0167763.

Kebouchi, M., Galia, W., Genay, M., Soligot, C., Lecomte, X., Awussi, A. A., et al. (2016). Implication of sortase-dependent proteins of *Streptococcus thermophilus* in adhesion to human intestinal epithelial cell lines and bile salt tolerance. *Appl. Microbiol. Biotechnol.* 100, 3667–3679. doi:10.1007/s00253-016-7322-1.

Kemp, K. D., Singh, K. V., Nallapareddy, S. R., and Murray, B. E. (2007). Relative

contributions of *Enterococcus faecalis* OG1RF sortase-encoding genes, srtA and bps (srtC), to biofilm formation and a murine model of urinary tract infection. *Infect. Immun.* 75, 5399–5404. doi:10.1128/IAI.00663-07.

Kharat, A. S., and Tomasz, A. (2003). Inactivation of the srtA gene affects localization of surface proteins and decreases adhesion of *Streptococcus pneumoniae* to human pharyngeal cells in vitro. *Infect. Immun.* 71, 2758–2765. doi:10.1128/IAI.71.5.2758-2765.2003.

Khare, B., Krishnan, V., Rajashankar, K. R., Xin, M., and Narayana, S. V (2011). Structural Differences between the *Streptococcus agalactiae* Housekeeping and Pilus-Specific Sortases : SrtA and SrtC1. 6. doi:10.1371/journal.pone.0022995.

Kim, Y. S., Cho, J. H., Park, S., Han, J. Y., Back, K., and Choi, Y. E. (2011). Gene regulation patterns in triterpene biosynthetic pathway driven by overexpression of squalene synthase and methyl jasmonate elicitation in *Bupleurum falcatum*. *Planta* 233, 343–355. doi:10.1007/s00425-010-1292-9.

Kolenbrander, P. (2017). Coaggregation : bacteria specific. 7, 406–413.

Kong, F. Y., Zhang, J. W., Li, R. F., Wang, Z. X., Wang, W. J., and Wang, W. (2017). Unique roles of gold nanoparticles in drug delivery, targeting and imaging applications. *Molecules* 22. doi:10.3390/molecules22091445.

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., et al. (2017). The ClusPro web server for protein-protein docking. *Nat. Protoc.* 12, 255–278. doi:10.1038/nprot.2016.169.

Kruger, R. G., Dostal, P., and McCafferty, D. G. (2004). Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the *Staphylococcus aureus* sortase transpeptidase SrtA. *Anal. Biochem.* 326, 42–48. doi:10.1016/j.ab.2003.10.023.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nat. Publ. Gr.* 228, 726–734. Available at: http://www.mendeley.com/research/discreteness-conductance-chnge-n-bimolecular-lipid-membrane-presence-certin-antibiotics/.

Lalioui, L., Pellegrini, E., Dramsi, S., Baptista, M., Bourgeois, N., Doucet-populaire, F., et al. (2005). The SrtA Sortase of *Streptococcus agalactiae* Is Required for Cell Wall Anchoring of Proteins Containing the LPXTG Motif , for Adhesion to Epithelial Cells , and for Colonization of the Mouse Intestine. 73, 3342–3350. doi:10.1128/IAI.73.6.3342.

Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486. doi:10.1007/BF00228148.

Lee, C. C., Jordan, D. B., Stoller, J. R., Kibblewhite, R. E., and Wagschal, K. (2018). Biochemical characterization of *Caulobacter crescentus* xylose dehydrogenase. *Int. J. Biol. Macromol.* 118, 1362–1367. doi:10.1016/j.ijbiomac.2018.06.124.

Lemieux, J., Woody, S., and Camilli, A. (2008). Roles of the sortases of *Streptococcus pneumoniae* in assembly of the RlrA pilus. *J. Bacteriol.* 190, 6002–6013. doi:10.1128/JB.00379-08.

Li, J. H., Miao, J., Wu, J. L., Chen, S. F., and Zhang, Q. Q. (2014). Preparation and characterization of active gelatin-based films incorporated with natural antioxidants. *Food Hydrocoll.* 37, 166–173. doi:10.1016/j.foodhyd.2013.10.015.

Los, D. A., and Murata, N. (2004). Membrane fluidity and its roles in the perception of environmental signals. *Biochim. Biophys. Acta - Biomembr.* 1666, 142–157.

doi:10.1016/j.bbamem.2004.08.002.

Ma, W., Saccardo, A., Roccatano, D., Aboagye-Mensah, D., Alkaseem, M., Jewkes, M., et al. (2018). Modular assembly of proteins on nanoparticles. *Nat. Commun.* 9, 1–9. doi:10.1038/s41467-018-03931-4.

Madhavan, A., Pandey, A., and Sukumaran, R. K. (2017). Expression system for heterologous protein expression in the filamentous fungus *Aspergillus unguis*. *Bioresour. Technol.* 245, 1334–1342. doi:10.1016/j.biortech.2017.05.140.

Malik, S., Petrova, M. I., Claes, I. J. J., Verhoeven, T. L. A., Busschaert, P., Vaneechoutte, M., et al. (2013). The highly autoaggregative and adhesive phenotype of the vaginal *lactobacillus plantarum* strain cmpg5300 is sortase dependent. *Appl. Environ. Microbiol.* 79, 4576–4585. doi:10.1128/AEM.00926-13.

Mandlik, A., Das, A., and Ton-that, H. (2008). The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. 105, 14147–14152.

Mandlik, A., Swierczynski, A., and Das, A. (2010). Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development Sortase covalently links proteins to peptidoglycan. 16, 33–40. doi:10.1016/j.tim.2007.10.010.

Mandlik, A., Swierczynski, A., Das, A., and Ton-That, H. (2007). *Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells. *Mol. Microbiol.* 64, 111–124. doi:10.1111/j.1365-2958.2007.05630.x.

Manzano, C., Izoré, T., Job, V., Di Guilmi, A. M., and Dessen, A. (2009). Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of Gram-positive pathogens. *Biochemistry* 48, 10549–10557. doi:10.1021/bi901261y.

Maresso, A. W., Chapa, T. J., and Schneewind, O. (2006). Surface protein IsdC and sortase B are required for heme-iron scavenging of *Bacillus anthracis*. *J. Bacteriol.* 188, 8145–8152. doi:10.1128/JB.01011-06.

Maresso, A. W., and Schneewind, O. (2008). Sortase as a Target of Anti-Infective Therapy. *Pharmacol. Rev.* 60, 128–141. doi:10.1124/pr.107.07110.

Mariscotti, J. F., Quereda, J. J., and Pucciarelli, M. G. (2012). Contribution of sortase A to the regulation of *Listeria monocytogenes* LPXTG surface proteins. 43–51. doi:10.2436/20.1501.01.157.

Marraffini, L. A., DeDent, A. C., and Schneewind, O. (2006). Sortases and the Art of Anchoring Proteins to the Envelopes of Gram-Positive Bacteria. *Microbiol. Mol. Biol. Rev.* 70, 192–221. doi:10.1128/MMBR.70.1.192-221.2006.

Marraffini, L. A., and Schneewind, O. (2006). Targeting proteins to the cell wall of sporulating *Bacillus anthracis*. *Mol. Microbiol.* 62, 1402–1417. doi:10.1111/j.1365-2958.2006.05469.x.

Marraffini, L. A., and Schneewind, O. (2007). Sortase C-mediated anchoring of BasI to the cell wall envelope of *Bacillus anthracis*. *J. Bacteriol.* 189, 6425–6436. doi:10.1128/JB.00702-07.

Matsuda, Y., Itaya, H., Kitahara, Y., Theresia, N. M., Kutukova, E. A., Yomantas, Y. A. V., et al. (2014). Double mutation of cell wall proteins CspB and PBP1a increases secretion of the antibody Fab fragment from *Corynebacterium glutamicum*. *Microb. Cell Fact.* 13, 1–10. doi:10.1186/1475-2859-13-56.

Mazmanian, S. K., Liu, G., Jensen, E. R., Lenoy, E., and Schneewind, O. (2000). *Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proc. Natl. Acad. Sci.* 97, 5510–5515. doi:10.1073/pnas.080520697.

Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999). *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science (80-.).* 285, 760–763. doi:10.1126/science.285.5428.760.

Mazmanian, S. K., Ton-That, H., Su, K., and Schneewind, O. (2002). An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. *Proc. Natl. Acad. Sci.* 99, 2293–2298. doi:10.1073/pnas.032523999.

Mishra, A., Das, A., Cisar, J. O., and Ton-That, H. (2007). Sortase-catalyzed assembly of distinct heteromeric fimbriae in *Actinomyces naeslundii*. *J. Bacteriol.* 189, 3156–3165. doi:10.1128/JB.01952-06.

Morita, C., Sumioka, R., Nakata, M., Okahashi, N., and Wada, S. (2014). Cell Wall-Anchored Nuclease of *Streptococcus sanguinis* Contributes to Escape from Neutrophil Extracellular Trap-Mediated Bacteriocidal Activity. 9. doi:10.1371/journal.pone.0103125.

Munir, A., Mehmood, A., and Azam, S. (2016). Structural and function prediction of *Musa acuminata* subsp. malaccensis protein. *Int. J. Bioautomation* 20, 19–30.

Muñoz-Provencio, D., Rodríguez-Díaz, J., Collado, M. C., Langella, P., Bermúdez-Humarán, L. G., and Monedero, V. (2012). Functional analysis of the *Lactobacillus casei* BL23 sortases. *Appl. Environ. Microbiol.* 78, 8684–8693. doi:10.1128/AEM.02287-12.

Narayanan, N., Karunakaran, V., Paul, W., Venugopal, K., Sujathan, K., and Kumar Maiti, K. (2015). Aggregation induced Raman scattering of squaraine dye: Implementation in diagnosis of cervical cancer dysplasia by SERS imaging. *Biosens. Bioelectron.* 70, 145–152. doi:10.1016/j.bios.2015.03.029.

Naveed, M., Tehreem, S., Usman, M., Chaudhry, Z., and Abbas, G. (2017). Structural and functional annotation of hypothetical proteins of human adenovirus: prioritizing the novel drug targets. *BMC Res. Notes* 10, 706. doi:10.1186/s13104-017-2992-z.

Naziga, E. B., and Wereszczynski, J. (2017). Molecular Mechanisms of the Binding and Specificity of *Streptococcus pneumoniae* Sortase C Enzymes for Pilin Subunits. *Sci. Rep.*

7, 1–14. doi:10.1038/s41598-017-13135-3.

Novick, R. P. (2000). Sortase: The surface protein anchoring transpeptidase and the LPXTG motif. *Trends Microbiol.* 8, 148–151. doi:10.1016/S0966-842X(00)01741-8.

Oehler, S., Amouyal, M., Kolkhof, P., Von Wilcken-Bergmann, B., and Müller-Hill, B. (1994). Quality and position of the three lac operators of *E. coli* define efficiency of repression. *EMBO J.* 13, 3348–3355. doi:10.1002/j.1460-2075.1994.tb06637.x.

Pallen, M. J., Lam, A. C., Antonio, M., and Dunbar, K. (2001). An embarrassment of sortases-A richness of substrates? *Trends Microbiol.* 9, 97–101. doi:10.1016/S0966-842X(01)01956-4.

Paterson, G. K., and Mitchell, T. J. (2004). The biology of Gram-positive sortase enzymes. *Trends Microbiol.* 12, 89–95. doi:10.1016/j.tim.2003.12.007.

Paulsen, I. T., Banerjei, L., Hyers, G. S. A., Nelson, K. E., Seshadri, R., Read, T. D., et al. (2003). Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science (80-.).* 299, 2071–2074. doi:10.1126/science.1080613.

Peltier, J., Shaw, H. A., Couchman, E. C., Dawson, L. F., Yu, L., Choudhary, J. S., et al. (2015). Cyclic diGMP Regulates Production of Sortase Substrates of *Clostridium difficile* and Their Surface Exposure through ZmpI Protease-mediated Cleavage. *J. Biol. Chem.* 290, 24453–24469. doi:10.1074/jbc.M115.665091.

Peltier, J., Shaw, H. A., Wren, B. W., and Fairweather, N. F. (2017). Disparate subcellular location of putative sortase substrates in *Clostridium difficile*. *Sci. Rep.* 7, 1–10. doi:10.1038/s41598-017-08322-1.

Perry, A. M., Ton-That, H., Mazmanian, S. K., and Schneewind, O. (2002). Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring. *J. Biol. Chem.* 277, 16241–16248. doi:10.1074/jbc.M109194200.

Persson, K. (2011). Structure of the sortase AcSrtC-1 from *Actinomyces oris*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 212–217. doi:10.1107/S0907444911004215.

Peters-Wendisch, P. G., Schiel, B., Wendisch, V. F., Katsoulidis, E., Möckel, B., Sahm, H., et al. (2001). Pyruvate carboxylase is a major bottleneck for glutamate and lysine production by *Corynebacterium glutamicum*. *J. Mol. Microbiol. Biotechnol.* 3, 295–300.

Pham, P. J., Hernandez, R., French, W. T., Estill, B. G., and Mondala, A. H. (2011). A spectrophotometric method for quantitative determination of xylose in fermentation medium. *Biomass and Bioenergy* 35, 2814–2821. doi:10.1016/j.biombioe.2011.03.006.

Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E., and Ploegh, H. L. (2007). Sortagging: A versatile method for protein labeling. *Nat. Chem. Biol.* 3, 707–708. doi:10.1038/nchembio.2007.31.

Ramya, A. N., Joseph, M. M., Nair, J. B., Karunakaran, V., Narayanan, N., and Maiti, K. K. (2016). New Insight of Tetraphenylethylene-based Raman Signatures for Targeted SERS Nanoprobe Construction Toward Prostate Cancer Cell Detection. *ACS Appl. Mater. Interfaces* 8, 10220–10225. doi:10.1021/acsami.6b01908.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, 320–324. doi:10.1093/nar/gku316.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738. doi:10.1038/nprot.2010.5.

Sambrook, J., Russell, D., (2001) Molecular Cloning: A Laboratory Manual. 2001. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Saranya, G., Joseph, M. M., Karunakaran, V., Nair, J. B., Saritha, V. N., Veena, V. S., et al. (2018). Enzyme-Driven Switchable Fluorescence-SERS Diagnostic Nanococktail for the Multiplex Detection of Lung Cancer Biomarkers. *ACS Appl. Mater. Interfaces* 10,

38807–38818. doi:10.1021/acsami.8b15583.

Saware, K., Aurade, R. M., Kamala Jayanthi, P. D., and Abbaraju, V. (2015). Modulatory Effect of Citrate Reduced Gold and Biosynthesized Silver Nanoparticles on α -Amylase Activity . *J. Nanoparticles* 2015, 1–9. doi:10.1155/2015/829718.

Schagfer, A., Tauch, A., Jsger, W., Kalinowski, J., Thierbachb, G., and Piihler, A. (1994). pK18mobsacB. *Gene* 145, 49–5201.

Schneewind, O., Model, P., and Fischetti, V. A. (1992). Sorting of protein a to the staphylococcal cell wall. *Cell* 70, 267–281. doi:10.1016/0092-8674(92)90101-H.

Schubert, A., Zakikhany, K., Pietrocola, G., Meinke, A., Speziale, P., Eikmanns, B. J., et al. (2004). The Fibrinogen Receptor FbsA Promotes Adherence of *Streptococcus agalactiae* to Human Epithelial Cells. 72, 6197–6205. doi:10.1128/IAI.72.11.6197.

Schubert, A., Zakikhany, K., Schreiner, M., Frank, R., Spellerberg, B., Eikmanns, B. J., et al. (2002). A fibrinogen receptor from group B Streptococcus interacts with fibrinogen by repetitive units with novel ligand binding sites. 557–569.

Scott, J. R., and Barnett, T. C. (2006). Surface proteins of gram-positive bacteria and how they get there. *Annu. Rev. Microbiol.* 60, 397–423. doi:10.1146/annurev.micro.60.080805.142256.

Severin, A., Tabei, K., and Tomasz, A. (2004). cereus RSVF1 , a Strain Closely Related to *Bacillus anthracis*. *Microb. Drug Resist.* 10, 77–82. doi:10.1089/1076629041310082.

Shaik, M. M., Maccagni, A., Tourcier, G., Di Guilmi, A. M., and Dessen, A. (2014). Structural basis of pilus anchoring by the ancillary pilin RrgC of *Streptococcus pneumoniae*. *J. Biol. Chem.* 289, 16988–16997. doi:10.1074/jbc.M114.555854.

Shrestha, P., and Wereszczynski, J. (2016). Discerning the Catalytic Mechanism of *Staphylococcus aureus* Sortase A with QM/MM Free Energy Calculations. *J. Mol. Graph. Model.* 67, 33–43. doi:10.1016/j.jmgm.2016.04.006.

Singh, A., and Chaube, R. (2014). Bioinformatic Analysis, Structure Modeling and Active Site Prediction of Aquaporin Protein from Catfish *Heteropneustes fossilis*. *Int. J. Recent Innov. Trends Comput. Commun.* 2, 3208–3215.

Spirig, T., Weiner, E. M., and Clubb, R. T. (2011). Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* 82, 1044–1059. doi:10.1111/j.1365-2958.2011.07887.x.

Steinhagen, M., Zunker, K., Nordsieck, K., and Beck-sickinger, A. G. (2013). Bioorganic & Medicinal Chemistry Large scale modification of biomolecules using immobilized sortase A from *Staphylococcus aureus*. *Bioorg. Med. Chem.* 21, 3504–3510. doi:10.1016/j.bmc.2013.03.039.

Sujai, P. T., Shamjith, S., Joseph, M. M., and Maiti, K. K. (2021). Elucidating Gold-MnO2Core-Shell Nanoenvelope for Real Time SERS-Guided Photothermal Therapy on Pancreatic Cancer Cells. *ACS Appl. Bio Mater.* 4, 4962–4972. doi:10.1021/acsabm.1c00241.

Sundar, M. S. L., Susmitha, A., Rajan, D., Hannibal, S., Sasikumar, K., Wendisch, V. F., et al. (2020). Heterologous expression of genes for bioconversion of xylose to xylonic acid in *Corynebacterium glutamicum* and optimization of the bioprocess. *AMB Express* 10. doi:10.1186/s13568-020-01003-9.

Suryadinata, R., Seabrook, S. A., Adams, T. E., Nuttall, S. D., and Peat, T. S. (2015). Structural and biochemical analyses of a *Clostridium perfringens* sortase D transpeptidase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 71, 1505–1513. doi:10.1107/S1399004715009219.

Suzuki, N., Okayama, S., Nonaka, H., Tsuge, Y., Inui, M., and Yukawa, H. (2005). Large-Scale Engineering of the. *Society* 71, 3369–3372. doi:10.1128/AEM.71.6.3369.

Swaminathan, A., Mandlik, A., Swierczynski, A., Gaspar, A., Das, A., and Ton-That, H. (2007). Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in

Corynebacterium diphtheriae. *Mol. Microbiol.* 66, 961–974. doi:10.1111/j.1365-2958.2007.05968.x.

Swierczynski, A., and Ton-That, H. (2006). Type III Pilus of Corynebacteria : Pilus Length Is Determined by the Level of Its Major Pilin Subunit. 188, 6318–6325. doi:10.1128/JB.00606-06.

Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F., and Schneewind, O. (1999). Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc. Natl. Acad. Sci.* 96, 12424–12429. doi:10.1073/pnas.96.22.12424.

Ton-That, H., Marraffini, L. A., and Schneewind, O. (2004). Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochim. Biophys. Acta - Mol. Cell Res.* 1694, 269–278. doi:10.1016/j.bbamcr.2004.04.014.

Ton-That, H., and Schneewind, O. (2003). Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol. Microbiol.* 50, 1429–1438. doi:10.1046/j.1365-2958.2003.03782.x.

Ton-That, H., and Schneewind, O. (2004). Assembly of pili in Gram-positive bacteria. *Trends Microbiol.* 12, 228–234. doi:10.1016/j.tim.2004.03.004.

Van der Rest, M. E., Lange, C., and Molenaar, D. (1999). A heat shock following electroporation induces highly efficient transformation of *Corynebacterium glutamicum* with xenogeneic plasmid DNA. *Appl. Microbiol. Biotechnol.* 52, 541–545. doi:10.1007/s002530051557.

Van Leeuwen, H. C., Klychnikov, O. I., Menks, M. A. C., Kuijper, E. J., Drijfhout, J. W., and Hensbergen, P. J. (2014). *Clostridium difficile* sortase recognizes a (S/P)PXTG sequence motif and can accommodate diaminopimelic acid as a substrate for transpeptidation. *FEBS Lett.* 588, 4325–4333. doi:10.1016/J.FEBSLET.2014.09.041.

Van Pijkeren, J. P., Canchaya, C., Ryan, K. A., Li, Y., Claesson, M. J., Sheil, B., et al. (2006). Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *Lactobacillus salivarius* UCC118. *Appl. Environ. Microbiol.* 72, 4143–4153. doi:10.1128/AEM.03023-05.

Wang, C., Li, M., Feng, Y., Zheng, F., Dong, Y., Pan, X., et al. (2009). The involvement of sortase A in high virulence of STSS-causing *Streptococcus suis* serotype 2. 23–33. doi:10.1007/s00203-008-0425-z.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427.

Weiner, E. M., Robson, S., Marohn, M., and Clubb, R. T. (2010). The sortase A enzyme that attaches proteins to the cell wall of *Bacillus anthracis* contains an unusual active site architecture. *J. Biol. Chem.* 285, 23433–23443. doi:10.1074/jbc.M110.135434.

Weiss, W. J., Lenoy, E., Murphy, T., Tardio, L. A., Burgio, P., Projan, S. J., et al. (2004). Effect of *srtA* and *srtB* gene expression on the virulence of *Staphylococcus aureus* in animal models of infection. *J. Antimicrob. Chemother.* 53, 480–486. doi:10.1093/jac/dkh078.

Wendisch, V. F., Jorge, J. M. P., Pérez-García, F., and Sgobba, E. (2016). Updates on industrial production of amino acids using *Corynebacterium glutamicum*. *World J. Microbiol. Biotechnol.* 32. doi:10.1007/s11274-016-2060-1.

Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, 407–410. doi:10.1093/nar/gkm290.

Williamson, D. J., Fascione, M. A., Webb, M. E., and Turnbull, W. B. (2012). Efficient N-Terminal Labeling of Proteins by Use of Sortase. *Angewandte*. 1–5.

doi:10.1002/anie.201204538.

Wu, C., Huang, I. H., Chang, C., Reardon-Robinson, M. E., Das, A., and Ton-That, H. (2014). Lethality of sortase depletion in *Actinomyces oris* caused by excessive membrane accumulation of a surface glycoprotein. *Mol. Microbiol.* 94, 1227–1241. doi:10.1111/mmi.12780.

Wu, C., Mishra, A., Reardon, M. E., Huang, I. H., Counts, S. C., Das, A., et al. (2012). Structural determinants of actinomyces sortase *srtC2* required for membrane localization and assembly of type 2 fimbriae for interbacterial coaggregation and oral biofilm formation. *J. Bacteriol.* 194, 2531–2539. doi:10.1128/JB.00093-12.

Wu, C., Mishra, A., Yang, J., Cisar, J. O., Das, A., and Ton-That, H. (2011). Dual function of a tip fimbrillin of Actinomyces in fimbrial assembly and receptor binding. *J. Bacteriol.* 193, 3197–3206. doi:10.1128/JB.00173-11.

Yamaguchi, M., Terao, Y., Ogawa, T., Takahashi, T., Hamada, S., and Kawabata, S. (2006). Role of *Streptococcus sanguinis* sortase A in bacterial colonization. *Microbes Infect.* 8, 2791–2796. doi:10.1016/j.micinf.2006.08.010.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2014). The I-TASSER suite: Protein structure and function prediction. *Nat. Methods* 12, 7–8. doi:10.1038/nmeth.3213.

Yokoyama, K., Utsumi, H., Nakamura, T., Ogaya, D., Shimba, N., Suzuki, E., et al. (2010). Screening for improved activity of a transglutaminase from *Streptomyces mobaraensis* created by a novel rational mutagenesis and random mutagenesis. *Appl. Microbiol. Biotechnol.* 87, 2087–2096. doi:10.1007/s00253-010-2656-6.

Zdarta, J., Feliczak-Guzik, A., Siwińska-Ciesielczyk, K., Nowak, I., and Jesionowski, T. (2020). Mesostructured cellular foam silica materials for laccase immobilization and tetracycline removal: A comprehensive study. *Microporous Mesoporous Mater.* 291.

doi:10.1016/j.micromeso.2019.109688.

Zhang, D. H., Yuwen, L. X., and Peng, L. J. (2013). Parameters affecting the performance of immobilized enzyme. *J. Chem.* 2013. doi:10.1155/2013/946248.

Zhang, J., Liu, H., Zhu, K., Gong, S., Dramsi, S., Wang, Y. T., et al. (2014). Antiinfective therapy with a small molecule inhibitor of *Staphylococcus aureus* sortase. *Proc. Natl. Acad. Sci. U. S. A.* 111, 13517–13522. doi:10.1073/pnas.1408601111.

Zhang, R., Wu, R., Joachimiak, G., Mazmanian, S. K., Missiakas, D. M., Gornicki, P., et al. (2009). NIH Public Access. 12, 1147–1156. doi:10.1016/j.str.2004.06.001.Structures.

Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O., and Narayana, S. V. L. (2004). Crystal structures of *Staphylococcus aureus* Sortase A and its substrate complex. *J. Biol. Chem.* 279, 31383–31389. doi:10.1074/jbc.M401374200.

Zou, Z., Gau, E., El-Awaad, I., Jakob, F., Pich, A., and Schwaneberg, U. (2019). Selective Functionalization of Microgels with Enzymes by Sortagging. *Bioconjug. Chem.* 30, 2859–2869. doi:10.1021/acs.bioconjchem.9b00568.

Zuo, F., Appaswamy, A., Gebremariam, H. G., and Jonsson, A. B. (2019). Role of Sortase A in *Lactobacillus gasseri* Kx110A1 Adhesion to Gastric Epithelial Cells and Competitive Exclusion of *Helicobacter pylori*. *Front. Microbiol.* 10, 1–11. doi:10.3389/fmicb.2019.02770.

Abstract of the thesis

Name of the Student : **Susmitha A** Faculty of study : **Science**	Registration No : **10BB15A39018** Year of submission : **2021**
AcSIR Academic Centre/CSIR Lab : CSIR-NIIST, Thiruvananthapuram Name of the supervisor (s) : Dr. K. Madhavan Nampoothiri	
Title of the thesis : Molecular identification and insights into the structural-functional characterization of Sortase E transpeptidase of *Corynebacterium glutamicum*	

C. glutamicum ATCC 13032 is an industrial organism recognized for amino acid synthesis. The sequence analysis confirmed that *C. glutamicum* encodes a single sortase, Sortase E (*NCgl2838*), but little is known about its function. In this study, based on the in-silico analysis identified a single sortase-dependent protein with LAXTG sorting motif which is predicted to function as a transporter protein. Three distinct variants of *C. glutamicum*, *srtE* deletion mutant, complement and overexpressed strain were generated to study the potential influence of *srtE* on cell development, morphology, and physiology. A FRET-based assay and HPLC were developed, to confirm that recombinant Sortase E catalyzes the cleavage of fluorescently labelled peptides containing LAXTG motifs in vitro. Mass spectrometry reveals the cleavage site is between the threonine and glycine residues of the LAXTG peptide. The key residues Cys, His, Arg and Tyr that influences the preference for a non-canonical LAXTG recognition motif has been found by structural modeling, substrate docking, and mutagenesis studies. The biochemical characterization of the enzyme shows a K_m of 12 ± 1 µM with LAHTG substrate, Ca^{2+}- independency and catalytic efficiency at higher temperature indicates the possible functional differences between CgSrtE and SaSrtA. Using engineered Sortase E, a pioneer attempt was made to immobilize enhanced green fluorescent protein (eGFP) and xylose dehydrogenase enzyme (XylB) over triglycine functionalized PEGylated gold nanoparticles via sortase-mediated ligation or sortagging. This is a kind of first report on class E sortase family from non-pathogenic *C. glutamicum* which enabled the enzyme to site-specifically modify protein/enzyme for various bioengineering applications.

List of publications

Details of the publications emanating out of the thesis work

1. **Susmitha A**, Nampoothiri K M and Bajaj H; Insights into the biochemical and functional characterization of sortase E transpeptidase of *Corynebacterium glutamicum*, *Biochemical Journal* (2019), 476, 3835–3847.

2. **Susmitha A**, Bajaj H and Madhavan Nampoothiri K; The divergent roles of sortase in the biology of Gram-positive bacteria. *Cell Surface* (2021), 7, 100055.

3. Sundar M S L, **Susmitha A,** Rajan D, Hannibal S, Sasikumar K, Wendisch V F and Nampoothiri K M; Heterologous expression of genes for bioconversion of xylose to xylonic acid in *Corynebacterium glutamicum* and optimization of the bioprocess. *AMB Express* (2020), 10.

4. **Susmitha A,** Jayadev S Arya, Sundar L, Kaustabh Kumar Maiti and Madhavan Nampoothiri K; Sortase E-mediated site-specific immobilization of green fluorescent protein and xylose dehydrogenase on gold nanoparticles. *(Communicated-Biotechnology and Bioengineering Journal)*

5. **Susmitha A,** Amurtha M and Madhavan Nampoothiri K; Structure and functional characterization of Sortase E and sortase-dependent protein from *Corynebacterium glutamicum. (Communicated-Journal of genetic engineering and biotechnology)*

6. **Susmitha A** and Madhavan Nampoothiri K; Implication of *Corynebacterium glutamicum* Sortase E in morphological and physiological changes. *(Communicated in FEMS microbiology letters)*

Details of the publications emanating not related to the thesis work

1. Sundar L, **Susmitha A,** Soumya M, Sasikumar K, and Nampoothiri M. Bioconversion of D-xylose to D-xylonic acid by *Pseudoduganella danionis*. *Indian Journal of Experimental Biology* (2019), 57, 825–838.

2. **Susmitha A,** Sasikumar K, Rajan D, Padmakumar M A, and Nampoothiri K M; Development and characterization of corn starch-gelatin based edible films incorporated with mango and pineapple for active packaging. *Food Bioscience* (2021), 41.

List of Conference presentations

1. Poster presentation on "Substrate recognition and biochemical characterization of sortase enzyme from *C. glutamicum"* in the International Conference on New Horizons in Biotechnology (NHBT-2019) jointly organized by CSIR-NIIST and the Biotech Research Society, India at Thiruvananthapuram during Nov 20-24, 2019. **[Best poster presentation award]**

2. Flash talk and poster presentation on "C-terminal oriented sortagging mediated by Sortase E of *C. glutamicum"* in the International Conference on Biotechnology for Resource Efficiency, Energy, Environment, Chemicals and Health (BREEECH 2021) jointly organized by CSIR-Indian Institute of Petroleum, India at Dehradun during Dec 1-4, 2021.

ANNEXURE I- Media Composition

1. Luria Bertani Medium (LB) gL^{-1}	
Tryptone	10
NaCl	10.0
Yeast Extract	5.0
Agar (for solid medium)	15.0
Adjust the pH to 7.0 by 1 N HCl/ NaOH, and sterilized by autoclaving	

2. Brain Heart Infusion (BHI) gL^{-1}	
Beef heart infusion	2.5
Calf brain infusion	6.25
Na_2HPO_4	1.75
Glucose	1.0
Peptone	5.0
NaCl	2.5
Agar (for solid medium)	15.0
Adjust the pH to 7.0 by 1 N HCl/ NaOH, and sterilized by autoclaving	

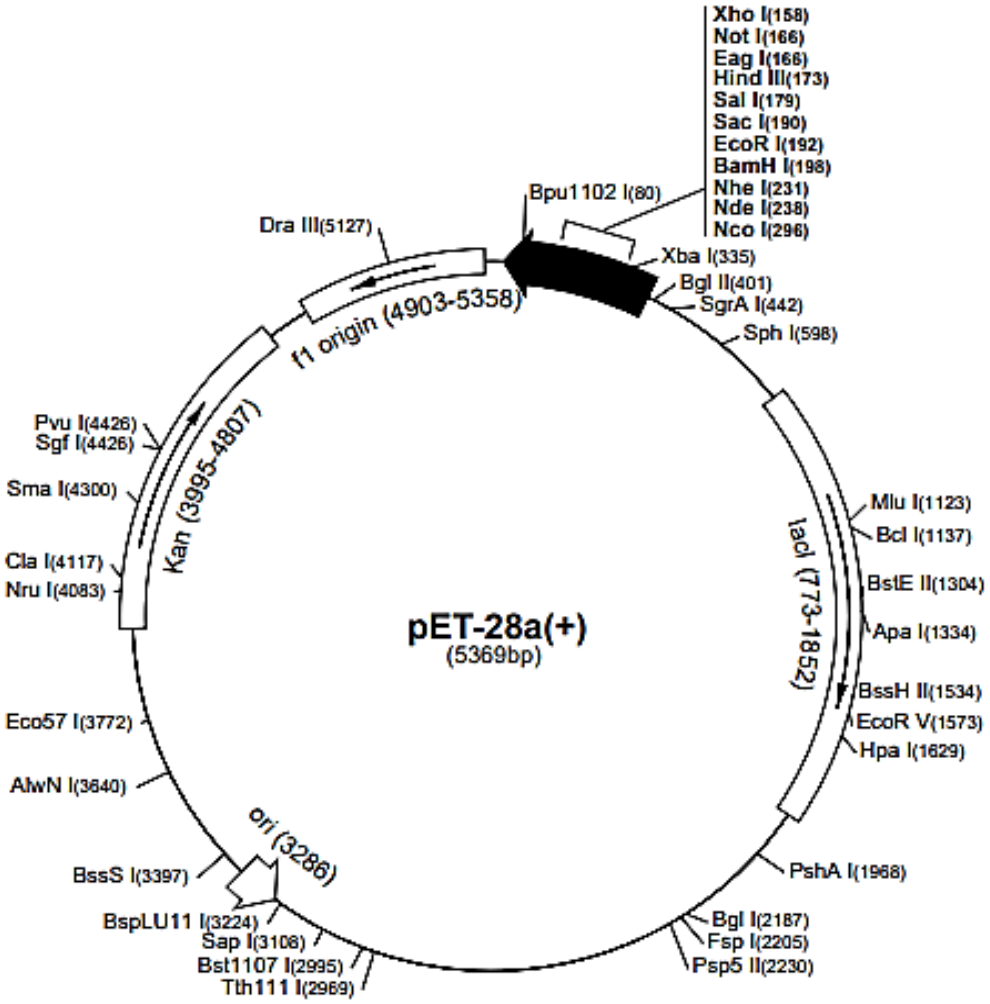
4. SOC Medium (gL^{-1})	
Glucose	20.0 mM
Bacto-tryptone	20 g/L
Yeast extract	5 g/L
$MgCl_2$	10 mM
KCl	0.186 g/L
NaCl	0.5 g/L

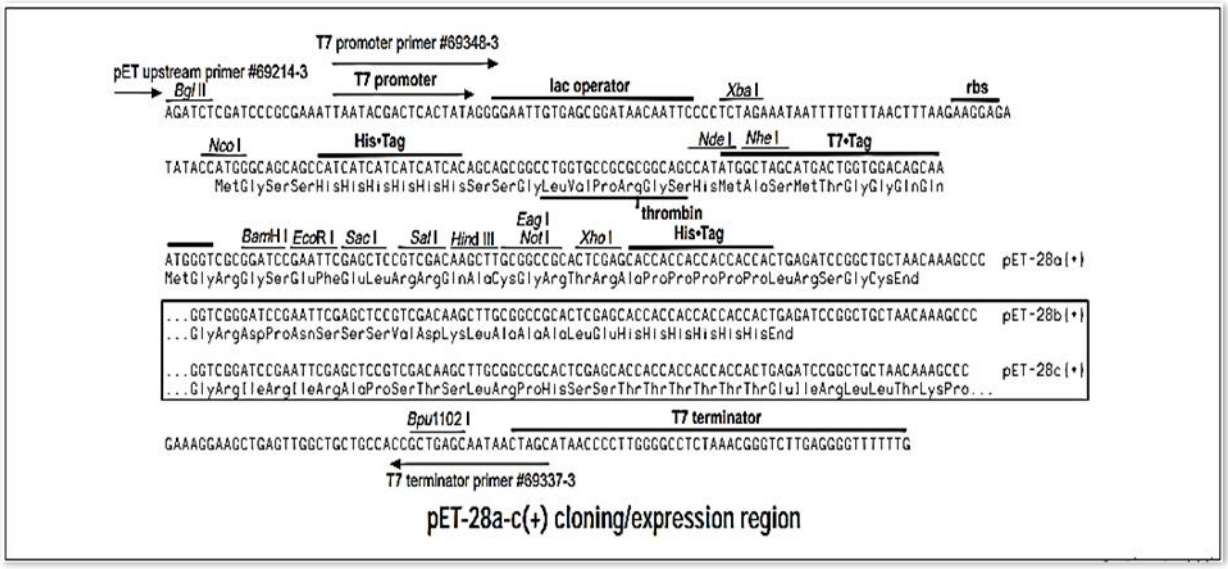
6. Epo Medium (gL^{-1})	
Tryptone	10 gL^{-1}
Yeast extract	5 gL^{-1}
NaCl	10 gL^{-1}
LB broth	1L

8. Terrific broth (TB) gL^{-1}	
TB-A (900 mL)	
Yeast extract	24 g
Trypton	12 g
Glycerol	4 g
Sterilize TB-A by autoclaving and allow media to cool and add 100 mL of TB-B	
TB-B (100 mL)	
KH_2PO_4	2.3 g
K_2HPO_4	12.5 g

ANNEXURE II-Vector map and sequences

pET28a (5569 bp) (Novagen, USA)



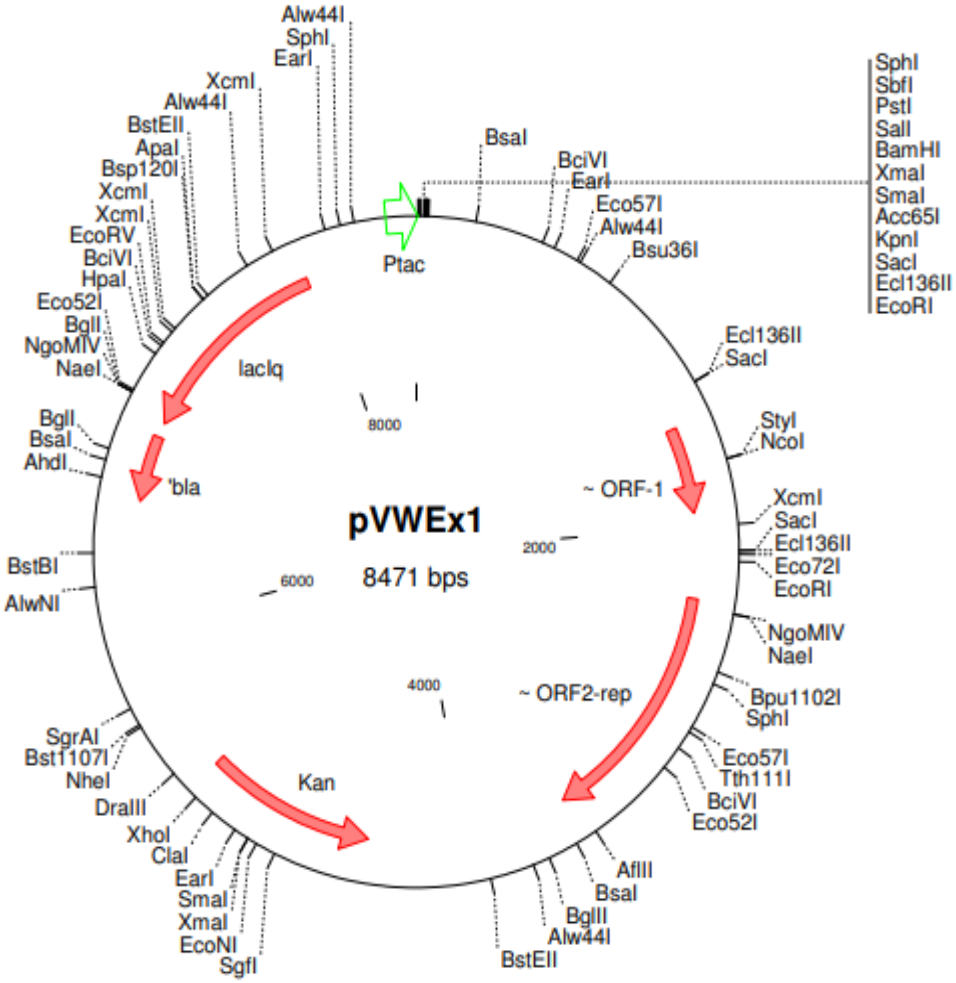


TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA
GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTT
CTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCG
ATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTG
GGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGT
GGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATA
AGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACG
CGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGC
GGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATT
CTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATA
CCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAG
GATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTA
ATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCC
GGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACG
CTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGA
GACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGC
AGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTG
GAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAA
AATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCT
GTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTT
CCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACC
CATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGA
ATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGA
CCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAA
GGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACC
GCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCAC
TTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGC
TGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTAC
ACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAA
GGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAG

GGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGA
TTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTT
ACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATT
CTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACC
GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTAC
GCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGC
CGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCG
ACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACA
GACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAA
CGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGC
CTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAA
AGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGA
TTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTAC
TGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGC
GGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGT
GTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGC
TGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTC
AGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTC
TGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCAT
GCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGG
CGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGG
CCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGG
CACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGC
CCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCC
GGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTC
GGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGC
GTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTT
CACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAA
AATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTAT
CCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCC
CAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTT
GCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGA
ATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACT
TAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCA
GTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCA
AGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAG
CGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTAC
AGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCG
CGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAAC
GCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCA
GCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTC
ACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTAC
TGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAA
AGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCAT
TAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCAT
GCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAA
CAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATA
GGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGA
TCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACA
ATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCA
TCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATGACTG

GTGGACAGCAAATGGGTCGCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACT
CGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGT
TGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTG
AGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

pVWEx1 (8471 bp) (Peters-Wendisch et al., 2001)



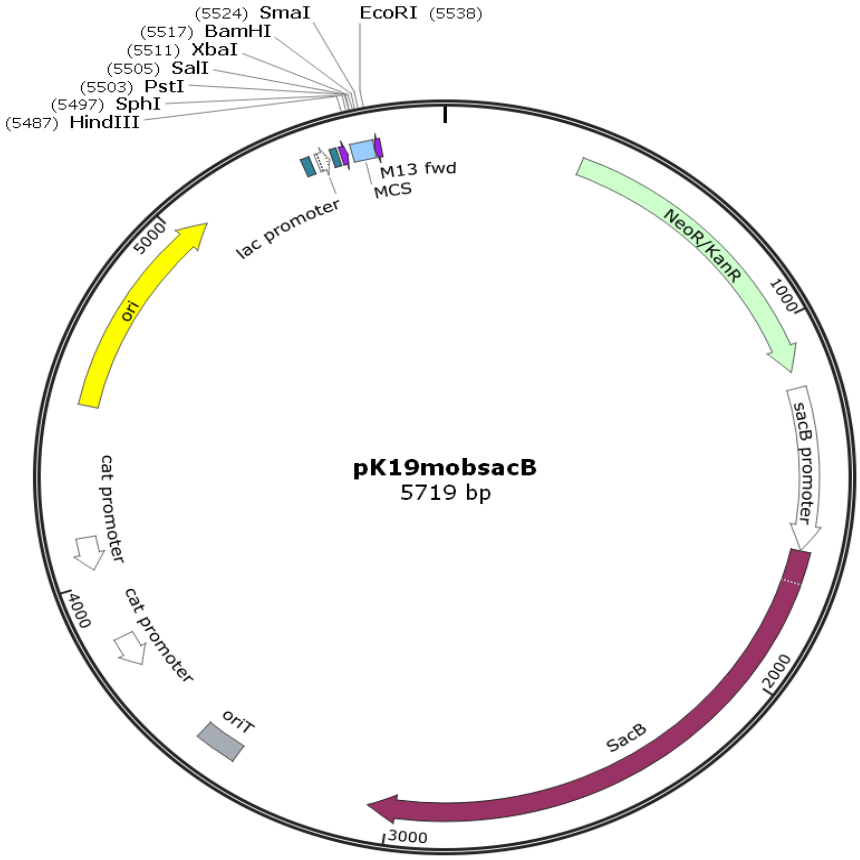
AAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCACTGG
CCGTCGTTTTACAGCCAAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATAC
AGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCG
GTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGT
GGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCG
AAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAA
TCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCC
CGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCG
TTTCTACAAACTCTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGAC

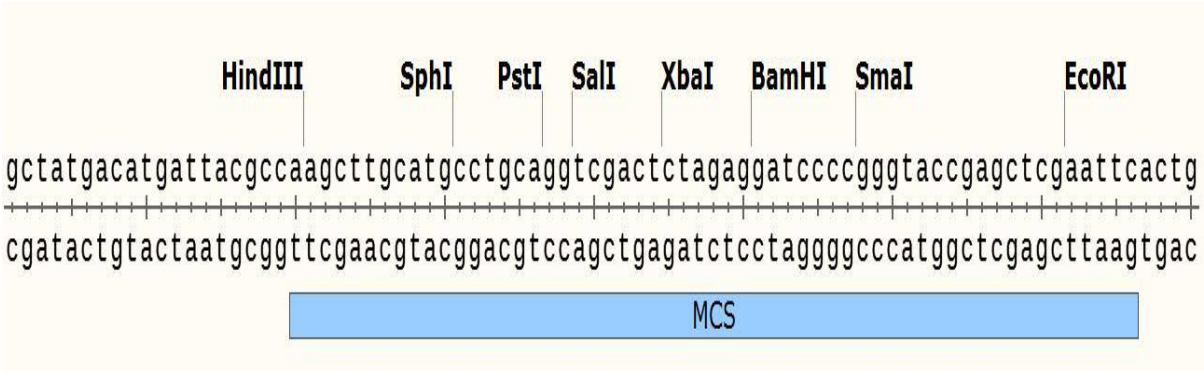
AATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTC
CGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAAC
GCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGG
ATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGC
ACTTTTGATCCCCCTGCGGCGTCGCTGATCGCCCTCGCGACGTTGTGCGGGTGGCTTGTCCC
TGAGGGCGCTGCGACAGATAGCTAAAAATCTGCGTCAGGATCGCCGTAGAGCGCGCGTCGCG
TCGATTGGAGGCTTCCCCTTTGGTTGACGGTCTTCAATCGCTCTACGGCGATCCTGACGCTT
TTTTGTTGCGTACCGTCGATCGTTTTATTTCTGTCGATCCCGAAAAAGTTTTTGCCTTTTGT
AAAAAACTTCTCGGTCGCCCCGCAAATTTTCGATTCCAGATTTTTTAAAAACCAAGCCAGAA
ATACGACACACCGTTTGCAGATAATCTGTCTTTCGAAAAATCAAGTGCGATACAAAATTTTT
AGCACCCCTGAGCTGCGCAAAGTCCCGCTTCGTGAAAATTTTCGTGCCGCGTGATTTTCCGC
CAAAAACTTTAACGAACGTTCGTTATAATGGTGTCATGACCTTCACGACGAAGTACCAAAAT
TGGCCCGAATCATCAGCTATGGATCTCTCTGATGTCGCGCTGGAGTCCGACGCGCTCGATGC
TGCCGTCGATTTAAAAACGGTGATCGGATTTTTCCGAGCTCTCGATACGACGGACGCGCCAG
CATCACGAGACTGGGCCAGTGCCGCGAGCGACCTAGAAACTCTCGTGGCGGATCTTGAGGAG
CTGGCTGACGAGCTGCGTGCTCGGCAGCGCCAGGAGGACGCACAGTAGTGGAGGATCGAATC
AGTTGCGCCTACTGCGGTGGCCTGATTCCTCCCCGGCCTGACCCGCGAGGACGGCGCGCAAA
ATATTGCTCAGATGCGTGTCGTGCCGCAGCCAGCCGCGAGCGCGCCAACAAACGCCACGCCG
AGGAGCTGGAGGCGGCTAGGTCGCAAATGGCGCTGGAAGTGCGTCCCCCGAGCGAAATTTTG
GCCATGGTCGTCACAGAGCTGGAAGCGGCAGCGAGAATTATCCGCGATCGTGGCGCGGTGCC
CGCAGGCATGACAAACATCGTAAATGCCGCGTTTCGTGTGGCCGTGGCCGCCCAGGACGTGT
CAGCGCCGCCACCACCTGCACCGAATCGGCAGCAGCGTCGCGCGTCGAAAAAGCGCACAGGC
GGCAAGAAGCGATAAGCTGCACGAATACCTGAAAAATGTTGAACGCCCCGTGAGCGGTAACT
CACAGGGCGTCGGCTAACCCCCAGTCCAAACCTGGGAGAAAGCGCTCAAAAATGACTCTAGC
GGATTCACGAGACATTGACACACCGGCCTGGAAATTTTCCGCTGATCTGTTCGACACCCATC
CCGAGCTCGCGCTGCGATCACGTGGCTGGACGAGCGAAGACCGCCGCGAATTCCTCGCTCAC
CTGGGCAGAGAAAATTTCCAGGGCAGCAAGACCCGCGACTTCGCCAGCGCTTGGATCAAAGA
CCCGGACACGGGAGAAACACAGCCGAAGTTATACCGAGTTGGTTCAAAATCGCTTGCCCGGT
GCCAGTATGTTGCTCTGACGCACGCGCAGCACGCAGCCGTGCTTGTCCTGGACATTGATGTG
CCGAGCCACCAGGCCGGCGGGAAAATCGAGCACGTAAACCCCGAGGTCTACGCGATTTTGGA
GCGCTGGGCACGCCTGGAAAAAGCGCCAGCTTGGATCGGCGTGAATCCACTGAGCGGGAAAT
GCCAGCTCATCTGGCTCATTGATCCGGTGTATGCCGCAGCAGGCATGAGCAGCCCGAATATG
CGCCTGCTGGCTGCAACGACCGAGGAAATGACCCGCGTTTTCGGCGCTGACCAGGCTTTTTC
ACATAGGCTGAGCCGGTGGCCACTGCACGTCTCCGACGATCCCACCGCGTACCGCTGGCATG
CCCAGCACAATCGCGTGGATCGCCTAGCTGATCTTATGGAGGTTGCTCGCATGATCTCAGGC
ACAGAAAAACCTAAAAAACGCTATGAGCAGGAGTTTTCTAGCGGACGGGCACGTATCGAAGC
GGCAAGAAAAGCCACTGCGGAAGCAAAAGCACTTGCCACGCTTGAAGCAAGCCTGCCGAGCG
CCGCTGAAGCGTCTGGAGAGCTGATCGACGGCGTCCGTGTCCTCTGGACTGCTCCAGGGCGT
GCCGCCCGTGATGAGACGGCTTTTCGCCACGCTTTGACTGTGGGATACCAGTTAAAAGCGGC
TGGTGAGCGCCTAAAAGACACCAAGATCATCGACGCCTACGAGCGTGCCTACACCGTCGCTC
AGGCGGTCGGAGCAGACGGCCGTGAGCCTGATCTGCCGCCGATGCGTGACCGCCAGACGATG
GCGCGACGTGTGCGCGGCTACGTCGCTAAAGGCCAGCCAGTCGTCCCTGCTCGTCAGACAGA
GACGCAGAGCAGCCGAGGGCGAAAAGCTCTGGCCACTATGGGAAGACGTGGCGGTAAAAAGG
CCGCAGAACGCTGGAAAGACCCAAACAGTGAGTACGCCCGAGCACAGCGAGAAAAACTAGCT
AAGTCCAGTCAACGACAAGCTAGGAAAGCTAAAGGAAATCGCTTGACCATTGCAGGTTGGTT
TATGACTGTTGAGGGAGAGACTGGCTCGTGGCCGACAATCAATGAAGCTATGTCTGAATTTA
GCGTGTCACGTCAGACCGTGAATAGAGCACTTAAGTCTGCGGGCATTGAACTTCCACGAGGA

CGCCGTAAAGCTTCCCAGTAAATGTGCCATCTCGTAGGCAGAAAACGGTTCCCCCCGTAGGG
GTCTCTCTCTTGGCCTCCTTTCTAGGTCGGGCTGATTGCTCTTGAAGCTCTCTAGGGGGGCT
CACACCATAGGCAGATAACGGTTCCCCACCGGCTCACCTCGTAAGCGCACAAGGACTGCTCC
CAAAGATCTTCAAAGCCACTGCCGCGACTCCGCTTCGCGAAGCCTTGCCCCGCGGAAATTTC
CTCCACCGAGTTCGTGCACACCCCTATGCCAAGCTTCTTTCACCCTAAATTCGAGAGATTGG
ATTCTTACCGTGGAAATTCTTCGCAAAAATCGTCCCCTGATCGCCCTTGCGACGTTGCTCGC
GGCGGTGCCGCTGGTTGCGCTTGGCTTGACCGACTTGATCCTCCGGCGTTCAGCCTGTGCCA
CAGCCGACAGGATGGTGACCACCATTTGCCCCATATCACCGTCGGTACTGATCCCGTCGTCA
ATAAACCGAACCGCTACACCCTGAGCATCAAACTCTTTTATCAGTTGGATCATGTCGGCGGT
GTCGCGGCCAAGACGGTCGAGCTTCTTCACCAGAATGACATCACCTTCCTCCACCTTCATCC
TCAGCAAATCCAGCCCTTCCCGATCTGTTGAACTGCCGGATGCCTTGTCGGTAAAGATGCGG
TTAGCTTTTACCCCTGCATCTTTGAGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGAC
TCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAG
AGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCACGGAACGGTCTG
CGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAAAGTTCGATTTATTCAACAA
AGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTCT
GATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAAT
ACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATA
GGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATT
AATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATC
CGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTAC
GCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCG
AGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCG
CAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCT
GGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATA
AAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATC
TGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCT
TCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATAC
CCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTG
AATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATG
ATGATATATTTTTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTGT
TGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGATCACGCATCTTCCCGACAACGCAGA
CCGTTCCGTGGCAAAGCAAAAGTTCAAAATCACCAACTGGTCCACCTACAACAAAGCTCTCA
TCAACCGTGGCTCCCTCACTTTCTGGCTGGATGATGGGGCGATTCAGGCCTGGTATGAGTCA
GCAACACCTTCTTCACGAGGCAGACCTCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTT
GGCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTG
CGTCAGCAGAATATGTGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTC
GGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATG
CCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCC
GCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTT
TCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTC
AGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGA
CCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCAC
CACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTA
AGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCA
AAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCA

GAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAAGGGGTCTGA
CGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCT
TCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAA
ACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATT
TCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTAC
CATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCA
GCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTC
CATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC
GCAACGTTGTTGCCATTGCCGATGATAAGCTGTCAAACATGGCCTGTCGCTTGCGGTATTCG
GAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAG
AAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCGGGGAGAGGCGGTTTGCGTATTGGG
CGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCT
GGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGT
TTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTAC
CGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCA
TCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTT
TGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATT
GCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGC
CCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTA
CCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAA
CGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGT
TAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCG
ACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTT
AATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCA
GCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCC
ATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCG
GGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCA
CATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTG
CACCATTCGATGGTGTCAACGTAAATGCATGCCGCTTCGCCTTCGCGCGCGAATTGCAAGCT
GATCCGGGCTTATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGC
TGTGGTATGGCTGTGCAGGTCGTAAATCACTGCATAATTCGTGTCGCTCAAGGCGCACTCCC
GTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGC
TGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACA
CAGGAAACAGAATTAAAAGATATGACCATGATTACGCC

pK19mobsacB (5719 bp) (Schagfer et al., 1994)





TGCCGCAAGCACTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACACGTAGAAAGCCAGTCC
GCAGAAACGGTGCTGACCCCGGATGAATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACG
CAAGCGCAAAGAGAAAGCAGGTAGCTTGCAGTGGGCTTACATGGCGATAGCTAGACTGGGCG
GTTTTATGGACAGCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAA
GCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAA

GATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCA
GGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGG
CTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGA
CCGACCTGTCCGGTGCCCTGAATGAACTCCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCC
ACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCT
GCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAG
TATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTC
GACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGA
TCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCA
AGGCGCGGATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAAT
ATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGA
CCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGG
CTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTAT
CGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGCTAGAGGATCGATCCTTTTT
AACCCATCACATATACCTGCCGTTCACTATTATTTAGTGAAATGAGATATTATGATATTTTC
TGAATTGTGATTAAAAAGGCAACTTTATGCCCATGCAACAGAAACTATAAAAAATACAGAGA
ATGAAAAGAAACAGATAGATTTTTTAGTTCTTTAGGCCCGTAGTCTGCAAATCCTTTTATGA
TTTTCTATCAAACAAAAGAGGAAAATAGACCAGTTGCAATCCAAACGAGAGTCTAATAGAAT
GAGGTCGAAAAGTAAATCGCGCGGGTTTGTTACTGATAAAGCAGGCAAGACCTAAAATGTGT
AAAGGGCAAAGTGTATACTTTGGCGTCACCCCTTACATATTTTAGGTCTTTTTTTATTGTGC
GTAACTAACTTGCCATCTTCAAACAGGAGGGCTGGAAGAAGCAGACCGCTAACACAGTACAT
AAAAAAGGAGACATGAACGATGAACATCAAAAAGTTTGCAAAACAAGCAACAGTATTAACCT
TTACTACCGCACTGCTGGCAGGAGGCGCAACTCAAGCGTTTGCGAAAGAAACGAACCAAAAG
CCATATAAGGAAACATACGGCATTTCCCATATTACACGCCATGATATGCTGCAAATCCCTGA
ACAGCAAAAAAATGAAAAATATCAAGTTTCTGAATTTGATTCGTCCACAATTAAAAATATCT
CTTCTGCAAAAGGCCTGGACGTTTGGGACAGCTGGCCATTACAAAACGCTGACGGCACTGTC
GCAAACTATCACGGCTACCACATCGTCTTTGCATTAGCCGGAGATCCTAAAAATGCGGATGA
CACATCGATTTACATGTTCTATCAAAAAGTCGGCGAAACTTCTATTGACAGCTGGAAAAACG
CTGGCCGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCTATCCTAAAAGACCAA
ACACAAGAATGGTCAGGTTCAGCCACATTTACATCTGACGGAAAAATCCGTTTATTCTACAC
TGATTTCTCCGGTAAACATTACGGCAAACAAACACTGACAACTGCACAAGTTAACGTATCAG
CATCAGACAGCTCTTTGAACATCAACGGTGTAGAGGATTATAAATCAATCTTTGACGGTGAC
GGAAAAACGTATCAAAATGTACAGCAGTTCATCGATGAAGGCAACTACAGCTCAGGCGACAA
CCATACGCTGAGAGATCCTCACTACGTAGAAGATAAAGGCCACAAATACTTAGTATTTGAAG
CAAACACTGGAACTGAAGATGGCTACCAAGGCGAAGAATCTTTATTTAACAAAGCATACTAT
GGCAAAAGCACATCATTCTTCCGTCAAGAAAGTCAAAAACTTCTGCAAAGCGATAAAAAACG
CACGGCTGAGTTAGCAAACGGCGCTCTCGGTATGATTGAGCTAAACGATGATTACACACTGA
AAAAAGTGATGAAACCGCTGATTGCATCTAACACAGTAACAGATGAAATTGAACGCGCGAAC
GTCTTTAAAATGAACGGCAAATGGTACCTGTTCACTGACTCCCGCGGATCAAAAATGACGAT
TGACGGCATTACGTCTAACGATATTTACATGCTTGGTTATGTTTCTAATTCTTTAACTGGCC
CATACAAGCCGCTGAACAAAACTGGCCTTGTGTTAAAAATGGATCTTGATCCTAACGATGTA
ACCTTTACTTACTCACACTTCGCTGTACCTCAAGCGAAAGGAAACAATGTCGTGATTACAAG
CTATATGACAAACAGAGGATTCTACGCAGACAAACAATCAACGTTTGCGCCGAGCTTCCTGC
TGAACATCAAAGGCAAGAAAACATCTGTTGTCAAAGACAGCATCCTTGAACAAGGACAATTA
ACAGTTAACAAATAAAAACGCAAAAGAAAATGCCGATGGGTACCGAGCGAAATGACCGACCA
AGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGG
GCTTCGGAATCGTTTTCCGGGACGCCCTCGCGGACGTGCTCATAGTCCACGACGCCCGTGAT
TTTGTAGCCCTGGCCGACGGCCAGCAGGTAGGCCGACAGGCTCATGCCGGCCGCCGCCGCCT
TTTCCTCAATCGCTCTTCGTTCGTCTGGAAGGCAGTACACCTTGATAGGTGGGCTGCCCTTC
CTGGTTGGCTTGGTTTCATCAGCCATCCGCTTGCCCTCATCTGTTACGCCGGCGGTAGCCGG
CCAGCCTCGCAGAGCAGGATTCCCGTTGAGCACCGCCAGGTGCGAATAAGGGACAGTGAAGA

AGGAACACCCGCTCGCGGGTGGGCCTACTTCACCTATCCTGCCCGGCTGACGCCGTTGGATA
CACCAAGGAAAGTCTACACGAACCCTTTGGCAAAATCCTGTATATCGTGCGAAAAAGGATGG
ATATACCGAAAAAATCGCTATAATGACCCCGAAGCAGGGTTATGCAGCGGAAAAGCGCTGCT
TCCCTGCTGTTTTGTGGAATATCTACCGACTGGAAACAGGCAAATGCAGGAAATTACTGAAC
TGAGGGGACAGGCGAGAGACGATGCCAAAGAGCTCCTGAAAATCTCGATAACTCAAAAAATA
CGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACG
TCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTAT
TTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATG
CTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTC
TATCAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTA
TGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCA
GGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCA
CAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACTTACTGATTTAGTGTATG
ATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGGGCTGGATGATCCTCCAGCG
CGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAAAAGGATCTAGGTGAAGATCCTTTTTG
ATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTA
GAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAAC
AAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC
CGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAG
TTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTT
ACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGT
TACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAG
CGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCC
CGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGA
GGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGA
CTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAA
CGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGT
TATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAA
ACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACT
GGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAG
GCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCA
CACAGGAAACAGCTATGACATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGG
ATCCCCGGGTACCGAGCTCGAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC
CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAG
CGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGAT
AAGCTAGCTTCACGC

ANNEXURE III-List of instruments

Instruments	Model and Country
Autoclave	Labline, India
Balance	Mettler Toledo, Mumbai, India
Centrifuge	Kubota 7780, Japan; Eppendorf, Germany; MICRO CL 17, Thermo Fisher Scientific, India
Cold room	Rinac Pvt. Ltd, India
Deep freezer	Elanpro, India; Haier, China
DNA sequencer	3500 Genetic Analyzer, Applied Biosystems, Hitachi, Japan
Electrophoresis unit	Bio-Rad, USA
Electroporator	Eppendorf, Germany
Fluorescence spectrophotometer	Infinite M200 PRO multimode microplate reader, Tecan, Switzerland
Gel documentation	ChemiDoc, Biorad, USA
Heating water bath	B20G, Lab companion, South Korea
Hot air Oven	Kemi Instruments, India
HPLC	Shimadzu, Japan
Incubator	Infors Ht, Switzerland
Laminar air flow chamber	Labline, India
Microplate reader	Infinite 200, Tecan, Switzerland
Nanodrop spectrophotometer	ND1000, Thermo Fisher Scientific, India
pH meter	Eutech, Thermo Fisher Scientific, India
PCR machine	MyCycler, Bio-Rad, USA; Eppendorf, Germany
Raman microscope	Alpha 300R, WITec Inc. Germany
Scanning electron microscope	JSM - 5600LV, JEOL, Japan
Sonicator	Vibra cell, Sonics and materials Inc., USA
Transmission electron microscope	JEM2010, JEOL, Japan
Thermostat	Eppendorf, USA
UV-Vis Spectrophotometer	UV-160A, Shimadzu, Japan,

ANNEXURE IV-AcSIR course work

SI No.	Level 100	Course No. and Title	Status
1.	BIO-101	Biostatistics	Completed
2.	BIO-102	Bioinformatics	Completed
3.	BIO-103	Basic Chemistry	Completed
4.	BIO-104	Research Methodology, communication/ ethics/ safety	Completed
	Level 200		
1.	BIO-NIIST-201	Biotechnology and Instrumentation	Completed
2.	BIO-NIIST-206	Protein Sciences and Proteomics	Completed
3.	BIO-NIIST-239	Basic Molecular Biology	Completed
	Level 300		
1.	BIO-NIIST-301	Seminar Course	Completed
2.	BIO-NIIST-337	Bioprocess Technology	Completed
3	BIO-NIIST-369	Enzymology and Enzyme Technology	Completed
	Level 400		
1.	BIO-NIIST-4-0001	Project Proposal	Completed
2.	BIO-NIIST-4-0002	Review	Completed
	Level 800		
1.	BIO-NIIST-4-0003	Project work	Completed

Biochemical Journal (2019) **476** 3835–3847
https://doi.org/10.1042/BCJ20190812

PORTLAND PRESS

Research Article

Insights into the biochemical and functional characterization of sortase E transpeptidase of *Corynebacterium glutamicum*

Aliyath Susmitha[1,2,3], **Kesavan Madhavan Nampoothiri**[1,3] **and Harsha Bajaj**[1]

[1]Microbial Processes and Technology Division, CSIR- National Institute for Interdisciplinary Science and Technology (NIIST), Trivandrum 695019, Kerala, India; [2]Academy of Scientific and Innovative Research (AcSIR), CSIR-NIIST, Trivandrum 695019, Kerala, India; [3]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

Correspondence: Kesavan Madhavan Nampoothiri (madhavan@niist.res.in)

Most Gram-positive bacteria contain a membrane-bound transpeptidase known as sortase which covalently incorporates the surface proteins on to the cell wall. The sortase-displayed protein structures are involved in cell attachment, nutrient uptake and aerial hyphae formation. Among the six classes of sortase (A–F), sortase A of *S. aureus* is the well-characterized housekeeping enzyme considered as an ideal drug target and a valuable biochemical reagent for protein engineering. Similar to SrtA, class E sortase in GC rich bacteria plays a housekeeping role which is not studied extensively. However, *C. glutamicum* ATCC 13032, an industrially important organism known for amino acid production, carries a single putative sortase (*NCgl2838*) gene but neither *in vitro* peptide cleavage activity nor biochemical characterizations have been investigated. Here, we identified that the gene is having a sortase activity and analyzed its structural similarity with Cd-SrtF. The purified enzyme showed a greater affinity toward LAXTG substrate with a calculated K_M of 12 ± 1 μM, one of the highest affinities reported for this class of enzyme. Moreover, site-directed mutation studies were carried to ascertain the structure functional relationship of Cg-SrtE and all these are new findings which will enable us to perceive exciting protein engineering applications with this class of enzyme from a non-pathogenic microbe.

Introduction

The surface proteins were displayed on to the cell wall of Gram-positive bacteria by a cysteine transpeptidase enzyme referred as sortase which was first reported and extensively studied in *Staphylococcus aureus* [1]. The SrtA of *S. aureus* covalently anchors the surface proteins onto the bacterial cell wall via a C-terminal cell wall sorting signal with an LPXTG recognition motif followed by a stretch of hydrophobic amino acids and a positively charged tail [2,3]. The nucleophilic attack was carried out by cysteine at the active site of sortase and cleaves the C-terminal of LPXTG motif between threonine and glycine forming a thioester intermediate complex which is then covalently anchored on the pentaglycine crossbridge of the lipid II moiety. The lipid II protein complex gets attached to the cell wall via transglycosylation and transpeptidation reaction. All sortase contains His–Cys–Arg as the catalytic triad to catalyze transpeptidation reaction [4–6].

Sortase-displayed surface structures plays a pivotal role in displaying virulence and pathogenesis without affecting the growth and viability of the cells [7–10]. The inhibition of these surface protein anchoring has led sortase, an ideal drug target for antimicrobial treatment [11]. Moreover, *S. aureus* SrtA and its variants has a high catalytic activity and substrate specificity when compared with other classes of sortase, which has developed sortase as a valuable biotechnology tool for biomedical

Received: 5 November 2019
Revised: 6 December 2019
Accepted: 9 December 2019

Accepted Manuscript online: 9 December 2019
Version of Record published: 23 December 2019

Biochemical Journal (2019) 476 3835–3847
https://doi.org/10.1042/BCJ20190812

engineering [12,13]. In recent years, sortagging has been used for a variety of applications such as protein ligation, covalently attaching proteins to the cells, protein labeling, cell-surface modification, protein cyclization and immobilization of proteins [14–21].

According to Spirig et al. sortases have being classified into six distinct types based on amino acid sequences, the classes A, B, C, D, E and F evolved a high substrate specificity towards LPXTG, NPXTN, LPXTG, LPXTA, LAXTG and LPXTG recognition motifs, respectively [11,22,23]. The prototype and housekeeping class A sortase was well characterized in Gram-positive bacteria and helps in anchoring a large number of proteins to the cell envelope. The class B sortase have been implicated in iron homeostasis by attaching the iron acquisition proteins to the cell envelope [24]. The class C sortase are responsible for the transpeptidation reactions by catalyzing the polymerization of pilin subunits in bacteria [25]. The sortase of class D anchor proteins that are involved in sporulation [26]. In contrast, there are only limited reports on E and F classes of sortases. The first enzymatic characterization on class F sortase has been reported recently in a human pathogen *Propionibacterium acnes* which contains an LPXTG sorting motif similar to class A sortase [11].

Class E sortase are widely distributed in GC rich Actinobacteria especially in *Corynebacterium* and *Streptomyces* sp. Similar to class A sortase, class E sortase anchors distinct surface proteins on the bacterial cell wall and function as a housekeeping sortase in the organism. The sortase E enzyme recognizes an unusual LAXTG sorting motif in which the conserved proline residue is replaced with alanine (underlined) [27]. Class E sortase are reported in two *Streptomyces* sp., one in *S. coeliocolor* and other one in *S. avermitilis*, which are structurally characterized and studied in detail [3,11,23,28]. The genome of *S. coeliocolor* contains two types of class E sortase, SrtE1 and SrtE2 which anchors chaplin proteins promoting in aerial hyphae formation. The genome of *S. avermitilis* contains four putative SrtE enzymes of which SrtE3 has been enzymatically and structurally characterized in *E. coli* for *in vitro*studies and found to be Ca^{2+} independent unlike SrtA enzyme. *C. diphtheria* contains a single class E sortase (named as Cd-SrtF) which help in anchoring assembled pili to the cell wall peptidoglycan [29,30].

Several studies were documented on the role of sortases in the context of virulence and colonization factors on bacterial cell surfaces. A very few studies were reported on sortases of non-pathogenic bacteria with more focus given on sortase in probiotics and their relevance in the interaction with host cells [31]. Similarly, *C. glutamicum* ATCC 13032 is a soil dwelling, non-pathogenic, industrially important microbe metabolically engineered for producing various amino acids and value-added chemicals [32]. The genome sequencing reveals that *C. glutamicum* contains a single sortase gene *NCgl2838* which encodes a putative sortase (Cg-SrtE) [33] and the sequence shows a similarity to Cd-SrtF of *C. diphtheriae* showing a high substrate specificity towards LAXTG recognition motif of class E sortase. However, in the literature reports there is no experimental records that support the sortase activity and substrate specificity of sortase E of *C. glutamicum*.

Therefore, the aim of the study was to produce a recombinant purified Cg-SrtE to identify the *in vitro*catalytic activity, biochemical properties and structure-functional relationship of the enzyme by site directed mutation.

Experimental

Bacterial strains and growth conditions

Bacterial strains and plasmids used are listed in (Table 1). *C. glutamicum* strain ATCC 13032 was cultivated in LB medium at 30°C. *E. coli* DH5α was used for cloning purposes and *E. coli* BL21 (DE3) (Novagen) for expression of recombinant proteins. The *E. coli* cells were routinely grown aerobically in Luria–Bertani (LB) broth at 37°C and 50 μg/ml kanamycin supplementation was done wherever necessary.

Bioinformatic analysis

The signal peptide of sortase protein was predicted by SignalP 4.1 [34]. Transmembrane helices and membrane topology of sortase protein sequences were predicted using TMHMM [35] and multiple protein sequence alignments were performed using ClustalW [36] and further processed with the ESPript programs [37].

DNA manipulation

The genomic DNA of *C. glutamicum* was isolated by previously established method [38]. The *srtE* from *C. glutamicum* ATCC 13032 is annotated in database records as GenBank: BA000036, Uniprot: Q8NLK3 with an appropriate ORF. The Cg-SrtE residues containing Ala45–Asn274 amino acids was amplified using the primers CglSrtE FP 5′-CCCCATATGGCCTATTGGACCAACGTGGAATC-3′ and CglSrtE RP 5′-CCCGTCGACTTAGT TTTCCTCCAAAGCTGCAGGGCGTTCGAT-3′, gel-purified and double digested with appropriate restriction

Biochemical Journal (2019) 476 3835–3847
https://doi.org/10.1042/BCJ20190812

Table 1 Bacterial strains and plasmids used in this study

Strains/plasmids	Description/genotype	Reference/source
C. glutamicum		
ATCC13032	Wild type (WT)	[39]
E. coli		
DH5α	*supE44 ΔlacU169 hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	[40]
BL21 (DE3)	F^- *ompT* $hsdS_B$ (r_B^--m_B^-) *gal dcm* (DE3)	Novagen, U.S.A.
Plasmids		
pET28 (a)	T7 expression vector (N-terminal His_6-Tag), Kan^r	Novagen, U.S.A.

sites (NdeI and SalI sites, underlined). To increase the yield of the protein, N-terminal transmembrane region containing 2–44 amino acids were deleted and the truncated gene was subsequently cloned into pET28a vector, yielding p28Cg-SrtE. The construct was verified by DNA sequencing and transformed into *E. coli* BL21 (DE3) cells for protein expression.

Expression and purification of Cg-SrtE

E. coli BL21 (DE3) harboring p28Cg-SrtE were grown in one liter of terrific broth (TB) medium supplemented with kanamycin (50 μg/ml) at 37°C with continuous shaking. When OD_{600} reached between 0.6 and 0.8, Cg-SrtE expression was induced by 1 mM IPTG for 3.5 h and cells were harvested by centrifugation at 3500×***g*** at 4°C for 30 min. Cell pellets were resuspended in lysis buffer containing 50 mM Tris–HCl, pH 7.5, 300 mM NaCl, 0.1% Tween 20, 2 mM β-mercaptoethanol and 20 mM imidazole which was lysed by sonication (Sonics Vibra cell, U.S.A.), and then centrifuged at 20 000×***g*** for 30 min at 4°C. The lysate was then loaded onto a HisTrap HP 5-ml column (GE Healthcare), subsequently rinsed with lysis buffer and wash buffer (50 mM Tris–HCl, pH 7.5, 300 mM NaCl, 0.1% Tween 20, 2 mM β-mercaptoethanol and 60 mM imidazole), and the protein was eluted with 200 mM imidazole. The excess imidazole was freed out using PD-10 desalting column. The purified protein was stored in buffer containing 50 mM Tris–HCl, pH 7.5, 150 mM NaCl and 2 mM β-mercaptoethanol and concentrated using 10 kDa Amicon ultracentrifugal filters (Millipore). Protein samples were quantified using Bradford reagent with an estimated concentration of ∼1.3 mg/ml and bovine serum albumin (BSA) as a standard. The purity of the protein was detected using 15% SDS–PAGE for two independent protein preparation with different concentration of 20 μg and 40 μg. The molecular mass of the proteins was analyzed by MALDI-TOF/MS using an Ultraflex TOF/TOF instrument.

Site-directed mutagenesis

Plasmid p28Cg-SrtE was used as the template for the introduction of single amino acid substitution generated by PCR. The amino acid residues, C240, H135, R249 and Y118 were mutated to encode Ala at these positions using appropriate primers listed in (Table 2) using Q5 site-directed mutagenesis kit (NEB). After PCR, the mutant plasmids were transformed into *E. coli* DH5α and the transformants were selected on kanamycin-LB

Table 2 Primers designed for site-directed mutagenesis

Mutants	Oligonucleotide sequences (5′–3′) orientation
C240A-F	5′-CTTGACCACGGCACACCCGCAGTTC-3′
C240A-R	5′-GTAAGCAGAGCTTCTGATC-3′
Y118A-F	5′-TCCTGGCCGTGCTGTGGATTCC-3′
Y118A-R	5′-CCGGCAAGAAGGTCTTCC-3′
H135A-F	5′-AGTGGCAGGCGCGCGAGTGGGCAAG-3′
H135A-R	5′-GCAAAGTTTCCGGCTTCA-3′
R249A-F	5′-CAACGCTGAGGCCATGATTGTGCAC-3′
R249A-R	5′-GAGAACTGCGGGTGACAC-3′

Biochemical Journal (2019) 476 3835–3847
https://doi.org/10.1042/BCJ20190812

agar plates. All the constructed mutants C240A, Y118A, H135A and R249A were subsequently detected and confirmed by DNA sequencing and positive mutants were further purified as described for wild type protein.

FRET analysis

Peptide substrates such as Abz-LAHTG-Dap(Dnp), Abz-LAETG-Dap(Dnp) and Abz-LPETG-Dap(Dnp) were synthesized by Shanghai GL Biochem. The peptide substrates were tagged with a 2-aminobenzoyl (Abz) fluorophore at the N-terminus and 2, 4-dinitrophenyl (Dnp) as the quencher at the C-terminus. The assay was performed in 100 μl reaction volume in a 96-well black microtiter plates (Thermofisher) containing 50 μM of each peptide substrate Abz-LAHTG-Dap(Dnp), Abz-LAETG-Dap(Dnp) and Abz-LPETG-Dap(Dnp) individually in a cleavage buffer (50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 1 mM DTT and 5 mM $CaCl_2$). The reaction was monitored by the addition of 5 μM of purified Cg-SrtE. The increase in the fluorescence intensity was measured at 37°C for 6 h using an Infinite M200 PRO multimode microplate reader (Tecan) at an excitation wavelength of 320 nm and emission wavelength of 420 nm. Fluorescence was measured by subtracting the peptide alone fluorescence from the overall reaction fluorescence. Each reactions were performed in triplicate and averaged, and plotted as arbitrary fluorescence units.

Kinetic measurements

Abz-LAHTG-(Dap)Dnp of Cg-SrtE were dissolved in 50% dimethyl sulfoxide and incubated at a concentration between 2.5–50 μM with a constant 5 μM enzyme (Cg-SrtE) concentration. The same peptide concentration without adding enzyme for the reaction served as control. Assay mixture consisted of 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 1 mM DTT and 5 mM $CaCl_2$ and incubated at 37°C. The fluorescence intensity was monitored for 6 h at regular 10 min intervals. The fluorescence intensity (V_0) was plotted against substrate concentration. The constants K_M and V_{max} was calculated by fitting the Michaelis–Menton equation; $V = V_{max}[S]/K_M + [S]$, using GraphPad Prism software.

Effect of pH, temperature, metal ions and enzyme concentration

The pH stability of Cg-SrtE was carried out in 100 μl reaction system with appropriate buffers. Na-acetate buffer was used for pH ≤ 5.2, while for pH > 6, Tris/HCI buffers was used along with 5 μM enzyme and reaction mixtures (40 μM Abz-LAHTG-(Dap)Dnp, 150 mM NaCl, 1 mM DTT and 5 mM $CaCl_2$) incubated at 37°C for 6 h. The optimum temperature was determined by incubating the enzyme reaction mixtures at different temperatures 25°C, 30°C, 37°C, 45°C, 60°C and 70°C. The effects of ions were determined by incubating Cg-SrtE with 5 mM metal ions (Ca^{2+}, K^+, Mg^{2+}, Mn^{2+} and in the absence of Ca^{2+}) at standard conditions. The optimum enzyme concentration was checked with a wide range of enzyme concentration (1, 2.5, 5, 7.5, 10, 15, 25 and 35 μM) using 40 μM Abz-LAHTG-(Dap)Dnp substrate in a buffer containing 50 mM Tris–HCl, 150 mM NaCl, pH 7.5, 5 mM $CaCl_2$ at 37°C for 6 h as mentioned in the above protocol.

In vitro cleavage assay for sortase activity

For the activity confirmation, sortase cleavage assay was also monitored by HPLC [41]. The reaction mixture contains 40 μM Abz-LAHTG-(Dap)Dnp, 50 mM Tris–HCl, 150 mM NaCl, 1 mM DTT and 5 mM $CaCl_2$. The reaction was initiated with an addition of 5 μM Cg-SrtE and incubated for 37°C for 6 h. The reaction was further quenched by adding 10-fold excess 0.1% TFA and injected onto a Vydac reversed-phase C18 RP-HPLC column (4.6 × 50 mm, 3 μm). A linear gradient of 10% to 40% B (Acetonitile/0.1% TFA) for 10 min at a flow rate of 3 ml/min was used for the separation of peptides. The Dnp containing peaks were observed at 355 nm UV absorbance. The G-(Dap)Dnp peaks were collected and approximate mass was confirmed by LC–MS.

Results

In silico analysis of putative sortase of *C. glutamicum*

The complete genome sequence of *C. glutamicum* ATCC 13032 was available in NCBI database and it predicted only a single copy of sortase-like transpeptidase [33]. The sortase protein was identified and confirmed on the basis of TLXTC signature motif. The respective gene *NCgl2838* consisted of 274 amino acids protein (accession- NP_602126), which is further annotated in the context as SrtE. The protein sequence alignment by BLASTP, exhibited a 62% sequence similarity with that of Cd-SrtF of *C. diphtheria*. The multiple alignments revealed that H134, C239 and R248 of Cd-SrtF were aligned well with active sites residues of *C. glutamicum*

Biochemical Journal (2019) **476** 3835–3847
https://doi.org/10.1042/BCJ20190812

SrtE of H135, C240 and R249, respectively, signifying that these three highly conserved residues might be the key catalytic triads of the enzyme (Figure 1). Whereas, the Y118 is highly conserved in class E sortase, which helps in recognition of LAXTG sorting signals. Moreover, the structure prediction analysis of SrtE was pre-formed using Phyre2 Protein Fold Recognition Server and also showed a high level of similarity to the known crystal structure of Cd-SrtF deposited in the Protein Data Bank (PDB code 5UUS).

Expression and purification of Cg-SrtE

To increase the solubility of the recombinant *C. glutamicum* sortase enzyme, 2–44 amino acids at the N-terminal transmembrane region of the protein was deleted and expressed in *E. coli* BL21 (DE3) which was designated as Cg-SrtE. The N-terminal His-tagged protein was purified using HisTrap HP 5-ml column. The expression, purity and homogeneity of Cg-SrtE was analyzed and confirmed through SDS–PAGE analysis (Figure 2A). The theoretical molecular mass of the full protein was found to be 29 857.21 Da which was nearly identical with mass of 30 001 Da obtained from MALDI-TOF/MS analysis (Figure 2B).

In vitro sortase activity and kinetic studies by FRET

To confirm the sortase activity and substrate specificity of Cg-SrtE, three peptides that encompass the sortase sorting motifs and were fluorescently labeled with a 2-aminobenzyl (Abz) fluorophore at the N-terminus and dinitrophenyl (Dnp) quencher group at the C-terminus were tested: Abz-LAHTG-Dap(Dnp), Abz-LAETG-Dap (Dnp) and Abz-LPET-GDap(Dnp). The fluorescence signals gets reduced when the peptides are in close proximity due to the quencher in the peptides, while the fluorescence signal gets enhanced when the peptide is cleaved by the enzyme, separating the fluorophore and quencher apart (Figure 3A). The peptides Abz-LAHTG-Dap(Dnp) and Abz-LAETG-Dap(Dnp) is synthesized based on the reports available on the substrate of E class sortase such as SrtE1 and SrtE2 substrates of *S. coeliocolor* [28]. The motif LPXTG is the

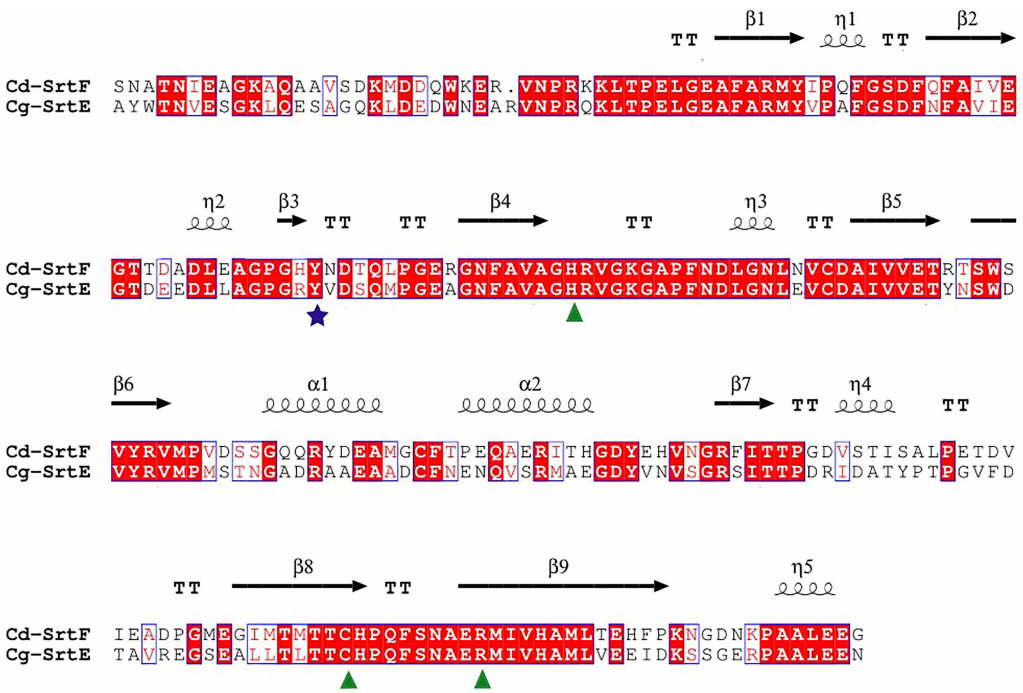


Figure 1. *In silico* analysis of *C. glutamicum* putative sortase.

The multiple sequence alignments of *C. glutamicum* sortase E with *C. diphtheriae* sortase F was performed with ClustalW, and the figure was produced using ESPript 3. The secondary structure elements helices (α), strands (β), 3_{10} helices (η) and turns (T) are shown above the alignment. The catalytic triads His, Cys and Arg are marked by green triangle and Tyr residue conserved only in sortase E enzyme, which is depicted in blue star.

Biochemical Journal (2019) **476** 3835–3847
https://doi.org/10.1042/BCJ20190812

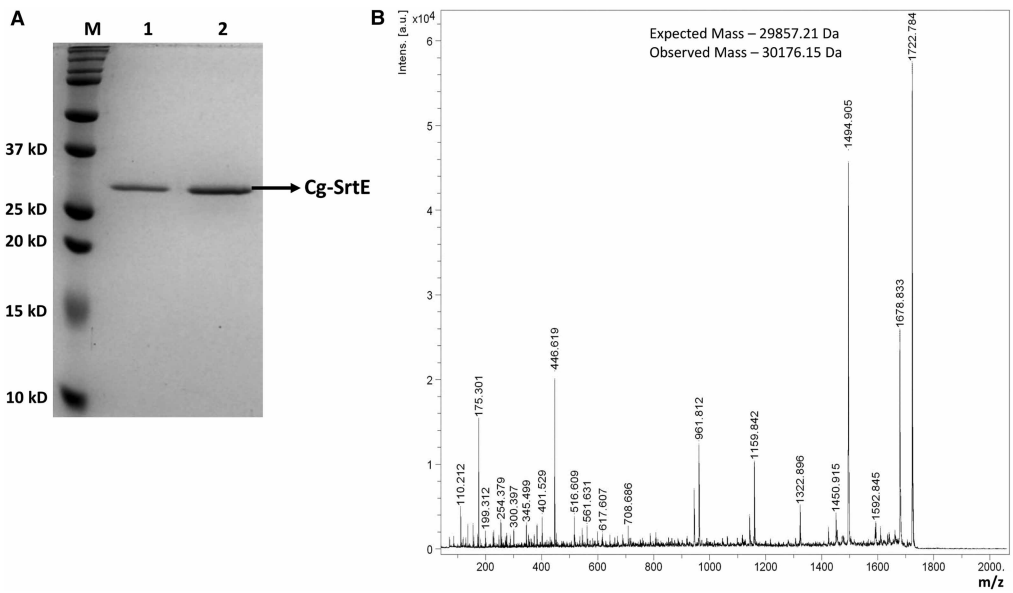


Figure 2. Characterization of recombinant Cg-SrtE.

(**a**) Expression and SDS–PAGE analysis of purified recombinant Cg-SrtE with an N-terminal His tag comprising 45–274 amino acids. Lane M, Molecular mass marker; Lane 1, 20 μg Cg-SrtE; Lane 2, 40 μg Cg-SrtE. (**b**) MALDI-TOF mass spectra of the protein.

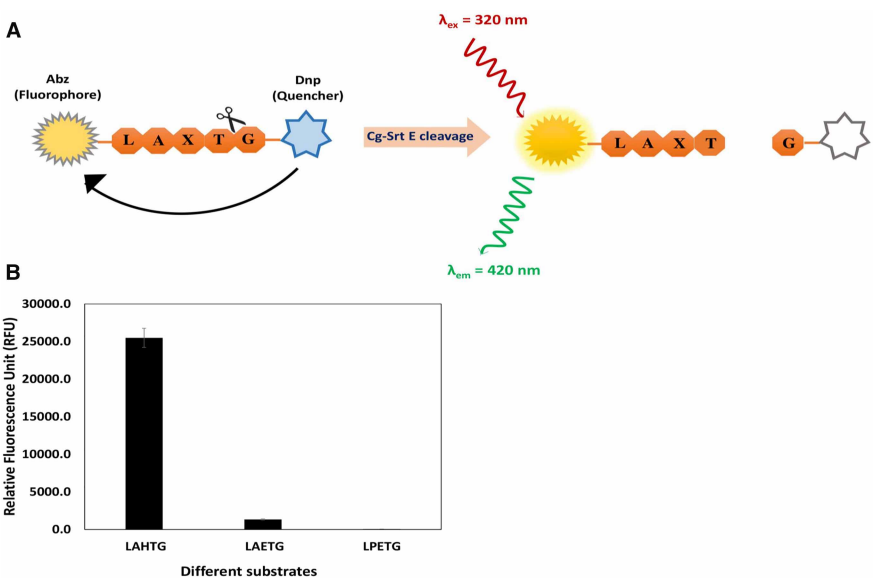


Figure 3. Enzyme activity and substrate specificity of Cg-SrtE by using a FRET-based cleavage assay.

(**a**) Schematic representation of Abz-LAXTG-Dap(Dnp) cleavage by Cg-SrtE. The fluorophore (Abz) and quencher (Dnp) is sandwiched between the peptide substrate. The fluorescence is measured when the sortase E-mediated reaction separates the fluorophore from the quencher. (**b**) Purified recombinant Cg-SrtE incubated with recognized substrate motifs Abz-LAHTG-Dap(Dnp), Abz-LAETG-Dap(Dnp) and Abz-LPETG-Dap(Dnp) to investigate the substrate specificity. The enzyme failed to cleave the LPETG motif recognized by *S. aureus* sortase A. Results shown here is an average of three independent enzyme assay each done in triplicates.

substrate for class A SrtA of *S. aureus* [1], hence Abz-LPETG-Dap(Dnp) serves as a negative control for the assay. We observed that the Cg-SrtE were able to cleave both Abz-LAHTG-Dap(Dnp) and Abz-LAETG-Dap (Dnp) peptide within 6 hours of incubation (Figure 3B). However, as expected the enzyme failed to cleave the Abz-LPETG-Dap(Dnp), the peptide of SrtA.

To determine the kinetic parameters of Cg-SrtE with Abz-LAHTG-Dap(Dnp) and Abz-LAETG-Dap(Dnp) peptides, kinetic analysis of sortase-catalyzed transpeptidation reaction was performed. Varying concentrations (2.5, 5, 7.5, 10, 15, 20, 25, 30, 35, 40, 45, 50 μM) of each peptides were incubated with 5 μM Cg-SrtE and the reaction was monitored in every 10 min interval for a period of 6 h. The level of cleavage observed for Abz-LAETG-Dap(Dnp) peptide was too low to facilitate the kinetic analysis (data not shown). On the other hand, initial velocities (V) obtained from the progress curves was able to plot against the varying concentration of Abz-LAHTG-Dap(Dnp) (Figure 4A). Thus, with Abz-LAHTG-Dap(Dnp) substrate, calculated an apparent K_M of 12 ± 1 μM and an apparent V_{max} of 1.3 ± 0.04 RFU/sec for Cg-SrtE (Figure 4B).

Effect of temperature, enzyme concentration, pH and metal ions on Cg-SrtE catalysis

The optimum temperature for Cg-SrtE activity was determined in a reaction mixture with 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 1 mM DTT, 5 mM $CaCl_2$ and 40 μM Abz-LAHTG-Dap(Dnp) as the substrate. The increase in fluorescence associates with the catalytic efficiency of the enzyme at different temperatures. Recombinant Cg-SrtE is less efficient with Abz-LAHTG-Dap(Dnp) substrate when incubated at 25°C and 30°C. However, the transpeptidase activity of the enzyme can be improved by incubating the Cg-SrtE at higher temperatures with the maximal catalytic efficiency being observed at 60°C (Figure 5A). Among the different concentrations tried, optimum enzyme concentration was found to be 15 μM (Figure 5B). Similarly, sortase activity was examined over a pH range of 3.6–10.5 of which the enzyme showed significant activity over a wide pH range of 7.5–10.2 with an optimal activity at pH 9.5 (Figure 5C). The activity of Cg-SrtE was compared with standard reaction buffer containing Ca^{2+}, with other metal ions (K^+, Mg^{2+} and Mn^{2+}) and also in the absence of Ca^{2+}. None of the metals ions showed any significant effect on the activity of Cg-SrtE (Figure 5D). A decrease in activity noted with Mn^{2+}. However, the presence of Ca^{2+} did not have any positive influence in the catalytic efficiency of Cg-SrtE and it distinguishes this enzyme from the calcium-dependent sortase of *S. aureus*.

Conserved residues in the active site of Cg-SrtE

To functionally confirm the proposed roles of the active site residues H135, C240, R249 based on the previously identified roles in *S. aureus* SrtA and to confirm whether Y118 has a prominent role in class E sortase based

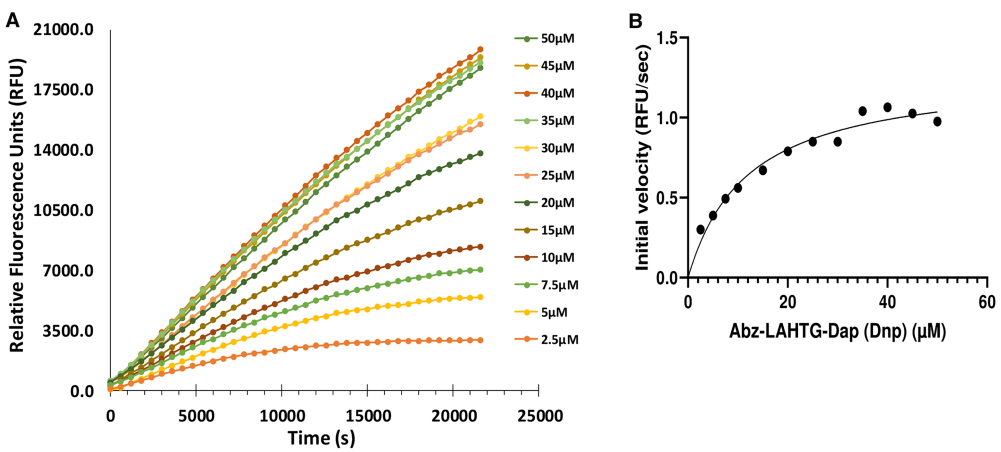


Figure 4. FRET assay with Abz-LAHTG-Dap(Dnp).

(**A**) Progress curves obtained from the cleavage reaction of Abz-LAHTG-Dap(Dnp) fluorescent peptide catalyzed by recombinant Cg-SrtE. Reactions containing 5 μM of Cg-SrtE enzyme incubated with 2.5 to 50 μM of fluorescent peptide at 37° C in 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 5 mM $CaCl_2$ and 1 mM DTT. (**B**) The kinetic parameters, K_M of 12 ± 1 μM and V_{max} of 1.3 ± 0.04 RFU/sec was determined for Cg-SrtE with Abz-LAHTG-Dap(Dnp).

Biochemical Journal (2019) **476** 3835–3847
https://doi.org/10.1042/BCJ20190812

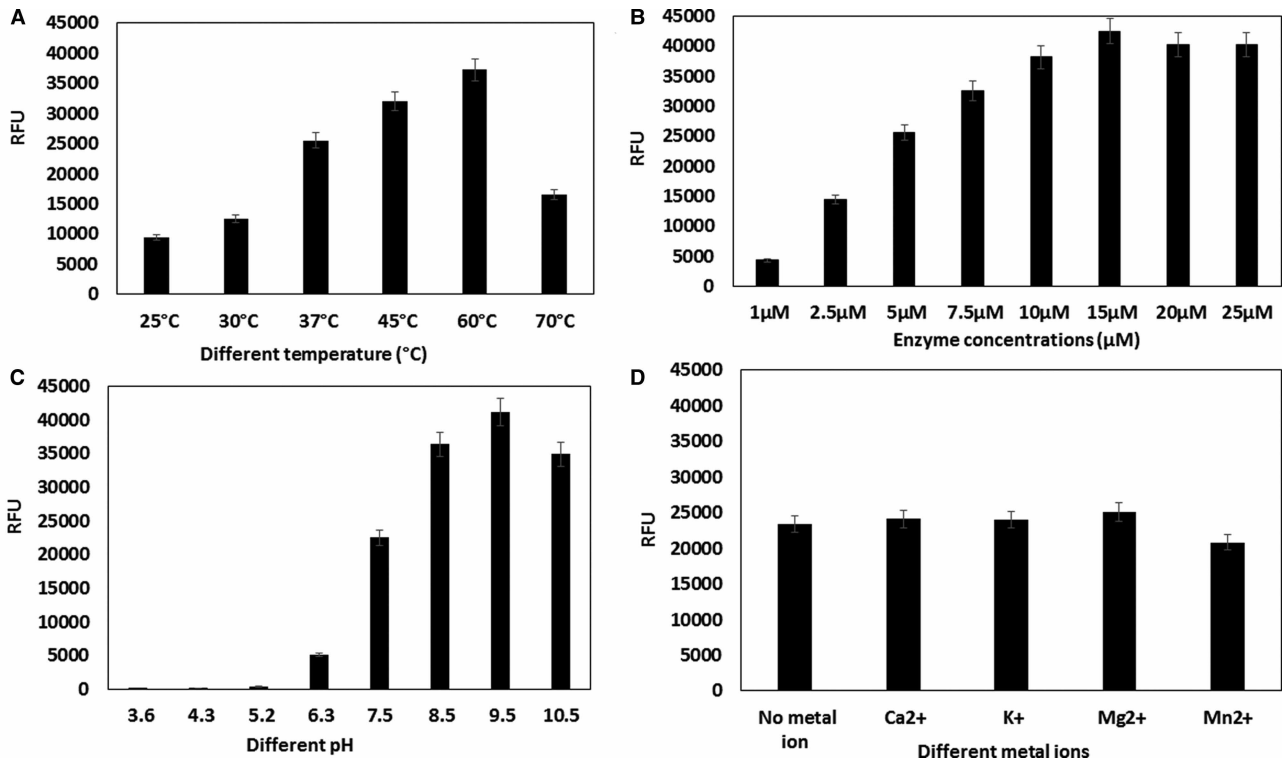


Figure 5. Effect of temperature, enzyme concentration, pH and metal ions on enzyme activity and stability with Abz-LAHTG-Dap(Dnp).
The enzyme assay was performed under standard conditions with 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 5 mM $CaCl_2$ and 1 mM DTT. (**A**) The effect of temperature was examined. The optimum temperature for the enzymatic activity was 60°C.(**B**) Maximum sortase E activity was observed at 15 μM enzyme concentration with 40 μM substrate concentration. (**C**) Optimum pH stability was determined at a range of 7.5–10.5. (**D**) The presence or absence of cations Ca^{2+}, Mg^{2+}, K^{+} and Mn^{2+} showed no significant effect on the activity of Cg-SrtE. The influence of Ca^{2+} did not affect catalytic efficiency of Cg-SrtE. However, the activity of Cg-SrtE is observed to be Ca^{2+} independent.

on the previous reports on *Streptomyces* sp., site-directed mutagenesis were done. The constructed mutant proteins were incubated with the fluorescence resonance energy transfer (FRET) peptide Abz-LAHTG-Dap(Dnp). As expected, all the active site mutants showed more than 50% loss in catalytic activity (Figure 6). Mutation in tyrosine residue also drastically decreased the activity indicating that the -OH group of Y118 could be involved in a hydrogen bond with the backbone nitrogen of the Ala residue for the recognition of LAXTG sorting signal instead of LPXTG. Thus, it is experimentally validated that the conserved catalytic residues mainly include H135, C240, R249 and Y118.

HPLC confirmation of substrate specificity of SrtE to LAXTG

Hydrolytic activity of Cg-SrtE was checked by HPLC as well. The Cg-SrtE (5 μM), was incubated with 40 μM Abz-LAHTG-Dap(Dnp) in a 500 μl reaction mixture at 37°C for 6 h. The G-Dap(Dnp) product peak released after the cleavage between threonine and glycine was monitored at an absorbance of 355 nm (Figure 7A). The product peak was collected and analyzed by ESI–MS and it showed a mass of 325.08 Da (Figure 7B) which is similar to the expected mass of 326.86 Da.

Discussion

Most pathogenic and non-pathogenic Gram-positive bacteria contains sortase enzyme which are involved in two functions within the cell wall, one is surface protein anchoring and other one is pilus assembly. In pathogenic bacteria, surface proteins or pili proteins play a significant role in contributing towards cell adhesion and pathogenesis by attaching to specific organ tissues of the host cells during infection or providing a way to escape from host immune response [8,42]. Since sortase are widely distributed among pathogenic bacteria, they constitute a promising therapeutic target for the development of novel antibiotics. However, in non-pathogenic

Biochemical Journal (2019) **476** 3835–3847
https://doi.org/10.1042/BCJ20190812

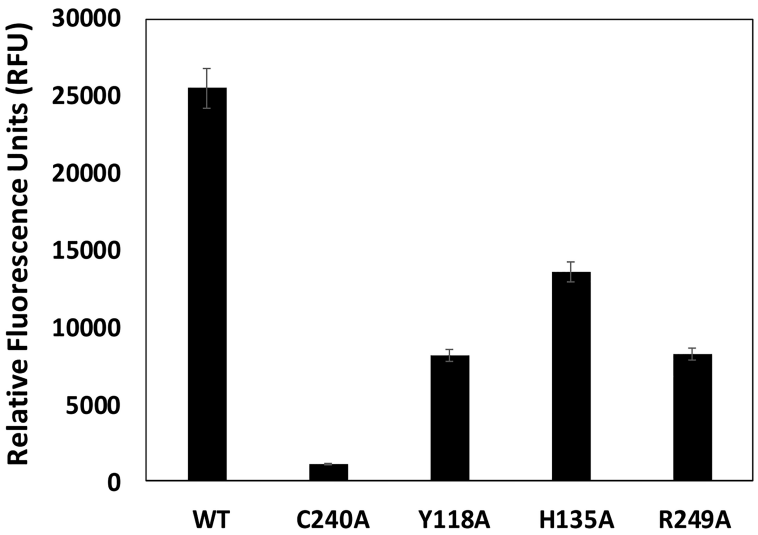


Figure 6. Confirmation of conserved residue required for *in vitro* Cg-SrtE activity.

Single mutations at C240A, H135A, R249A and Y118A resulted in the loss of enzymatic activities against Abz-LAHTG- Dap (Dnp), indicating that these four residues played critical roles in Cg-SrtE activity.

bacteria especially in probiotics, the sortase plays a pivotal role in eliciting health benefits to the host. Sortase and sortase-dependent proteins in lactic acid bacteria promote adhesion to epithelial cell lines, modulating immune response of host cells and also aid in bile salt stress resistance [43–45].

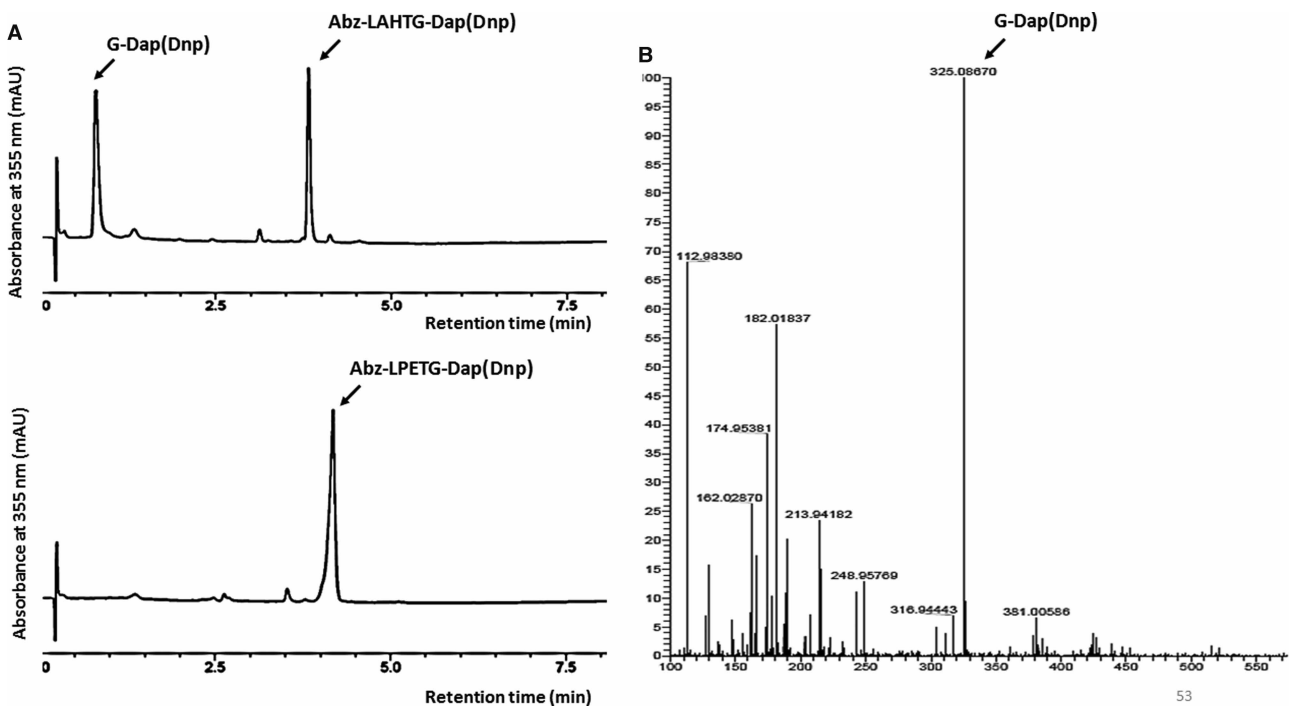


Figure 7. Sortase cleavage assay by HPLC with Abz-LAHTG-Dap(Dnp) substrate.

(**A**) Representative HPLC profile of peptide substrate Abz-LAHTG-Dap(Dnp), non-substrate Abz-LPETG-Dap(Dnp) (as control) and product G-Dap (Dnp) formed by the action of Cg-SrtE. The reaction products were separated using C18 RP-HPLC column; and the eluent was monitored at 355 nm UV absorbance. (**B**) MS analysis of purified sample revealed the presence of an ion at m/z 325.08, corresponding to the expected cleavage product G-Dap(Dnp)-NH_2.

Biochemical Journal (2019) **476** 3835–3847
https://doi.org/10.1042/BCJ20190812

Among the Coryneform bacteria, *C. glutamicum* is a well-known industrial microbe for the production of amino acids and other low-molecular mass substances [46–48]. Besides this, *C. glutamicum* expression system has been commercialized as CORYNEX®, an alternative potential platform for the secretion of heterologous proteins [49,50]. The secretory proteins in *C. glutamicum* are generally translocated via secretory (sec) pathway in an unfolded manner through the SecYEG translocation pore. Upon the arrival of the secretory protein, the signal peptidase cleaves the signal peptide followed by folding of protein into its functional conformation [51]. While the membrane-bound sortase E enzyme recognizes the surface proteins translocated via secretory (sec) pathway with a LAXTG sorting motif at the C-terminal of the protein and covalently anchors the surface proteins on the cell wall. *C. glutamicum* being a non-pathogenic bacterium also possess two *secA* homologous genes (*secA1* and *secA2*) which help in the translocation of proteins and deletion of either of the genes will affect growth and viability of the cells [52]. The phylogenetic analysis of SrtE of *C. glutamicum* reveals that the protein was found to be conserved within the genomes of *Corynebacterium* sp. and act as a housekeeping gene within the organism (Figure 8). The analysis showed 29 genomes of *Corynebacterium* sp. *C. diphtheriae* encodes six sortases of which class E sortase plays a housekeeping role in the organism. However, the multiple alignment and structure homology modeling of *C. glutamicum* SrtE exhibited 62% similarity with the available crystal structure of *C. diphtheriae* of SrtF from the protein data bank.

In this study, *in vitro*sortase E cleavage activity was performed by HPLC and FRET with the synthesized peptides. Among the three peptides, the enzyme Cg-SrtE demonstrated a higher substrate affinity towards Abz-LAHTG-Dap(Dnp) with an apparent K_M of 0.012 mM which is higher than the reported SavSrtE K_M of

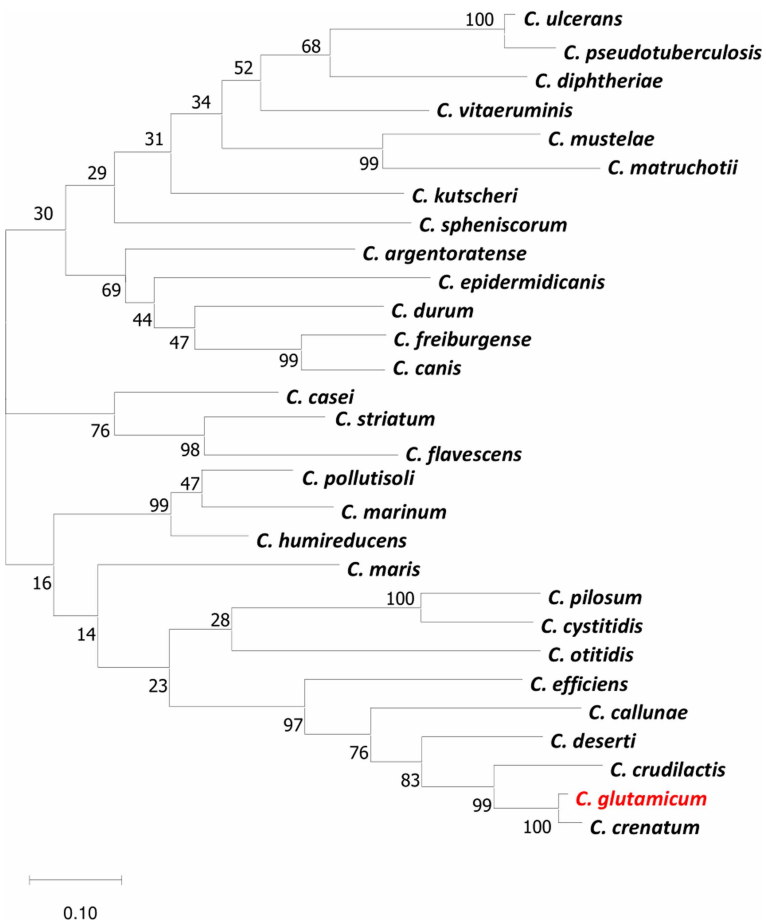


Figure 8. Phylogeny of Sortase E homologs of Corynebacterium.

The phylogenetic tree of the housekeeping sortase E was constructed by MEGA X [53] using maximum likelihood method and JTT matrix-based model [54]. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The scale bar represents 0.1 substitutions per site. The analysis showed 29 genomes of *Corynebacterium* sp.

1.14 mM for the LAXTG motif [3]. The Cg-SrtE also failed to cleave Abz-LPETG-Dap(Dnp) substrate of SrtA, hence proving that the Cg-SrtE belongs to class E family. The calcium ions enhances the catalytic efficiency of SrtA of *S. aureus* which mainly contains E105, E108, D112 and E171 residues on β3–β4 and β6–β7 loops within the structure [3,55]. Since, these calcium binding sites are not found in Cg-SrtE, which proves that Ca^{2+} is not essential for the enhancement of sortase E activity in *C. glutamicum*. These evidences matches with the report obtained from FRET analysis. Moreover, the enzyme showed significant activity over a wide pH range of 7.5–10.2 with an optimum pH ∼9.5 which is similar to the optimum pH of SrtA [11]. The enzyme catalytic efficiency was improved when incubated at higher temperature ∼60°C and enzyme concentration of ∼15 μM was found to be optimum. The sortase-catalyzed transpeptidation involves Cys–His–Arg catalytic triads, the Cys residue at the active site of sortase cleaves the Thr-Gly bond and forms a stable thioacyl intermediate complex, His protonates the substrate leaving group and Arg believes to function as an oxyanion hole to stabilize the transition state and further help in proper positioning of substrate in the active site of the enzyme [5,6,56]. The Tyr residue which is highly conserved in class E sortase recognizes the alanine residue of the LAXTG sorting signal. However, mutating at the conserved residues H135, C240, R249 and Y118 of Cg-SrtE has led to the conformational destabilization of the protein, with a drastic reduction in catalytic efficiency, believing that these residues might play a significant part in the enzyme.

The high catalytic efficiency with LAXTG substrate and the Ca^{2+} independency, allows non-pathogenic Cg-SrtE to be used subsequently in sortagging applications such as protein immobilization, sortase assay kit and self-cleaving tag for protein purification etc., which makes the enzyme more efficient than the pathogenic *S. aureus* SrtA variants.

Abbreviations

C. glutamicum, *Corneybacterium glutamicum*; Cd-SrtF, *Corynebacterium diphtheriae* sortase F; Cg-SrtE, *Corynebacterium glutamicum* sortase E; DTT, dithiothreitol; *E.* coli, *Escherichia coli*; *ESI*/MS, electrospray ionization mass spectrometry; FRET, fluorescence resonance energy transfer; IPTG, isopropyl β-D-thiogalactopyranoside; K_M, Michaelis constant; LC–MS, liquid chromatography–mass spectrometry; MALDI-TOF/MS, matrix assisted laser desorption ionization-time of flight mass spectrometry; RFU, relative fluorescence unit; RP-HPLC, reversed-phase high-performance liquid chromatography; *S. aureus*, *Staphylococcus aureus*; *S. avermitilis*, *Streptomyces avermitilis*; *S. coeliocolor*, *Streptomyces coeliocolor*; SavSrtE, *Streptomyces avermitilis* sortase E; SDS–PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; SrtA, sortase A; TFA, trifluoroacetic acid.

Author Contribution

K.M.N. and A.S. designed the work plan and wrote the manuscript. A.S. performed the research work. H.B. did data analysis and provided critical comments.

Acknowledgements

The first author Susmitha acknowledges the Junior and Senior Research Fellowship from CSIR, New Delhi. The corresponding author acknowledge the colleagues of CSIR NIIST namely Dr Ramesh Kumar N for providing sequencing help and to Dr Kaustabh Maiti for his useful suggestions and assistance in HPLC.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

References

1 Mazmanian, S.K., Liu, G., Ton-That, H. and Schneewind, O. (1999) *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760–763 https://doi.org/10.1126/science.285.5428.760

2 Novick, R.P. (2000) Sortase: the surface protein anchoring transpeptidase and the LPXTG motif. *Trends Microbiol.* **8**, 148–151 https://doi.org/10.1016/S0966-842X(00)01741-8

3 Das, S., Pawale, V.S., Dadireddy, V., Singh, A.K., Ramakumar, S. and Roy, R.P. (2017) Structure and specificity of a new class of Ca^{2+}-independent housekeeping sortase from *Streptomyces avermitilis* provide insights into its non-canonical substrate preference. *J. Biol. Chem.* **292**, 7244–7257 https://doi.org/10.1074/jbc.M117.782037

4 Perry, A.M., Ton-That, H., Mazmanian, S.K. and Schneewind, O. (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring. *J. Biol. Chem.* **277**, 16241–8 https://doi.org/10.1074/jbc.M109194200

Biochemical Journal (2019) **476** 3835–3847
https://doi.org/10.1042/BCJ20190812

5 Frankel, B.A., Kruger, R.G., Robinson, D.E., Kelleher, N.L. and McCafferty, D.G. (2005) *Staphylococcus aureus* sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* **44**, 11188–11200 https://doi.org/10.1021/bi050141j
6 Frankel, B.A., Tong, Y., Bentley, M.L., Fitzgerald, M.C. and McCafferty, D.G. (2007) Mutational analysis of active site residues in the *Staphylococcus aureus* transpeptidase SrtA. *Biochemistry* **46**, 7269–7278 https://doi.org/10.1021/bi700448e
7 Bierne, H., Mazmanian, S.K., Trost, M., Pucciarelli, M.G., Liu, G., Dehoux, P. et al. (2002) Inactivation of the srtA gene in *Listeria monocytogenes* inhibits anchoring of surface proteins and affects virulence. *Mol. Microbiol.* **43**, 869–881 https://doi.org/10.1046/j.1365-2958.2002.02798.x
8 Cheng, A.G., Kim, H.K., Burts, M.L., Krausz, T., Schneewind, O. and Missiakas, D.M. (2009) Genetic requirements for *Staphylococcus aureus* abscess formation and persistence in host tissues. *FASEB J.* **23**, 3393–3404 https://doi.org/10.1096/fj.09-135467
9 Mazmanian, S.K., Liu, G., Jensen, E.R., Lenoy, E. and Schneewind, O. (2000) *Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proc. Natl Acad. Sci.* **97**, 5510–5515 https://doi.org/10.1073/pnas.080520697
10 Weiss, W.J., Lenoy, E., Murphy, T., Tardio, L.A., Burgio, P., Projan, S.J. et al. (2004) Effect of srtA and srtB gene expression on the virulence of *Staphylococcus aureus* in animal models of infection. *J. Antimicrob. Chemother.* **53**, 480–486 https://doi.org/10.1093/jac/dkh078
11 Di Girolamo, S., Puorger, C., Castiglione, M., Vogel, M., Gébleux, R., Briendl, M. et al. (2019) Characterization of the housekeeping sortase from the human pathogen *Propionibacterium acnes*: first investigation of a class F sortase. *Biochem. J.* **476**, 665–682 https://doi.org/10.1042/BCJ20180885
12 Mao, H., Hart, S.A., Schink, A. and Pollok, B.A. (2004) Sortase-mediated protein ligation: a new method for protein engineering. *J. Am. Chem. Soc.* **126**, 2670–2671 https://doi.org/10.1021/ja039915e
13 Puorger, C., Di Girolamo, S. and Lipps, G. (2017) Elucidation of the recognition sequence of sortase B from *Bacillus anthracis* by using a newly developed liquid chromatography-mass spectrometry-based method. *Biochemistry* **56**, 2641–2650 https://doi.org/10.1021/acs.biochem.7b00108
14 Bachran, C., Schröder, M., Conrad, L., Cragnolini, J.J., Tafesse, F.G., Helming, L. et al. (2017) The activity of myeloid cell-specific VHH immunotoxins is target-, epitope-, subset-and organ dependent. *Sci. Rep.* **7**, 2–11 https://doi.org/10.1038/s41598-017-17948-0
15 Botulinum, R.I. and Study, R. (2014) HHS Public Access. 4(1):139–48
16 Jia, X., Kwon, S., Wang, C.I.A., Huang, Y.H., Chan, L.Y., Tan, C.C. et al. (2014) Semienzymatic cyclization of disulfide-rich peptides using sortase A. *J. Biol. Chem.* **289**, 6627–6638 https://doi.org/10.1074/jbc.M113.539262
17 Policarpo, R.L., Kang, H., Liao, X., Rabideau, A.E., Simon, M.D. and Pentelute, B.L. (2014) Flow-based enzymatic ligation by sortase A. *Angew. Chem. Int. Ed. Engl.* **53**, 9203–9208 https://doi.org/10.1002/anie.201403582
18 Popp, M.W.-L. and Ploegh, H.L. (2011) Making and breaking peptide bonds: protein engineering using sortase. *Angew. Chem. Int. Ed. Engl.* **50**, 5024–5032 https://doi.org/10.1002/anie.201008267
19 Ritzefeld, M. (2014) Sortagging: a robust and efficient chemoenzymatic ligation strategy. *Chem. A Eur. J.* **20**, 8516–8529 https://doi.org/10.1002/chem.201402072
20 Schmidt, M., Toplak, A., Quaedflieg, P.J. and Nuijens, T. (2017) Enzyme-mediated ligation technologies for peptides and proteins. *Curr. Opin. Chem. Biol.* **38**, 1–7 https://doi.org/10.1016/j.cbpa.2017.01.017
21 States, U. (2017) HHS Public Access. 106(3):201–7
22 Spirig, T., Weiner, E.M. and Clubb, R.T. (2011) Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* **82**, 1044–1059 https://doi.org/10.1111/j.1365-2958.2011.07887.x
23 Kattke, M.D., Chan, A.H., Duong, A., Sexton, D.L., Sawaya, M.R., Cascio, D. et al. (2016) Crystal structure of the *Streptomyces coelicolor* sortase E1 transpeptidase provides insight into the binding mode of the novel class e sorting signal. *PLoS ONE* **11**, 1–21 https://doi.org/10.1371/journal.pone.0167763
24 Mazmanian, S.K., Skaar, E.P., Gaspar, A.H., Humayun, M., Gornicki, P., Jelenska, J. et al. (2003) Passage of ehme-iron across the envelope of *Staphylococcus aureus*. *Science* **299**, 906–909 https://doi.org/10.1126/science.1081147
25 Khare, B. and Narayana S, V.L. (2017) Pilus biogenesis of Gram-positive bacteria: roles of sortases and implications for assembly. *Protein Sci.* **26**, 1458–1473 https://doi.org/10.1002/pro.3191
26 Bradshaw, W.J., Davies, A.H., Chambers, C.J., Roberts, A.K., Shone, C.C. and Acharya, K.R. (2015) Molecular features of the sortase enzyme family. *FEBS J.* **282**, 2097–2114 https://doi.org/10.1111/febs.13288
27 Comfort, D. and Clubb, R.T. (2004) A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. *Infect. Immun.* **72**, 2710–2722 https://doi.org/10.1128/IAI.72.5.2710-2722.2004
28 Duong, A., Capstick, D.S., Di Berardo, C., Findlay, K.C., Hesketh, A., Hong, H.J. et al. (2012) Aerial development in *Streptomyces coelicolor* requires sortase activity. *Mol. Microbiol.* **83**, 992–1005 https://doi.org/10.1111/j.1365-2958.2012.07983.x
29 Swaminathan, A., Mandlik, A., Swierczynski, A., Gaspar, A., Das, A. and Ton-That, H. (2007) Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in *Corynebacterium diphtheriae*. *Mol. Microbiol.* **66**, 961–974 https://doi.org/10.1111/j.1365-2958.2007.05968.x
30 Chang, C., Mandlik, A., Das, A. and Ton-That, H. (2011) Cell surface display of minor pilin adhesins in the form of a simple heterodimeric assembly in *Corynebacterium diphtheriae*. *Mol. Microbiol.* **79**, 1236–1247 https://doi.org/10.1111/j.1365-2958.2010.07515.x
31 Call, E.K. and Klaenhammer, T.R. (2013) Relevance and application of sortase and sortase-dependent proteins in lactic acid bacteria. *Front. Microbiol.* **4**, 73. https://doi.org/10.3389/fmicb.2013.00073
32 Gopinath, V., Murali, A., Dhar, K.S. and Nampoothiri, K.M. (2012) Erratum: *Corynebacterium glutamicum* as a potent biocatalyst for the bioconversion of pentose sugars to value-added products (Applied Microbiology and Biotechnology DOI:10.1007/s00253-011-3686-4). *Appl. Microbiol. Biotechnol.* **93**, 451–453 https://doi.org/10.1007/s00253-011-3789-y
33 Boekhorst, J., De Been, M.W.H.J., Kleerebezem, M. and Siezen, R.J. (2005) Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J. Bacteriol.* **187**, 4928–4934 https://doi.org/10.1128/JB.187.14.4928-4934.2005
34 Petersen, T.N., Brunak, S., Von Heijne, G. and Nielsen, H. (2011) Signalp 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 https://doi.org/10.1038/nmeth.1701
35 Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E.L.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 https://doi.org/10.1006/jmbi.2000.4315
36 Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 https://doi.org/10.1093/nar/22.22.4673

37 Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, 320–324 https://doi.org/10.1093/nar/gku316
38 Lessard, P.A., Brien, X.M.O., Currie, D.H. and Sinskey, A.J. (2004) Pb264, a small, mobilizable, temperature sensitive plasmid from *Rhodococcus*. *BMC Microbiol.* **4**, 15 https://doi.org/10.1186/1471-2180-4-15
39 Abe, S., Takayama, K.I. and Kinoshita, S. (1967) Taxonomical studies on glutamic acid-producing bacteria. *J. Gen. Appl. Microbiol.* **13**, 279–301 https://doi.org/10.2323/jgam.13.279
40 Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580 https://doi.org/10.1016/S0022-2836(83)80284-8
41 Kruger, R.G., Dostal, P. and McCafferty, D.G. (2004) Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the *Staphylococcus aureus* sortase transpeptidase SrtA. *Anal. Biochem.* **326**, 42–48 https://doi.org/10.1016/j.ab.2003.10.023
42 Pallen, M.J., Lam, A.C., Antonio, M. and Dunbar, K. (2001) An embarrassment of sortases- A richness of substrates? *Trends Microbiol.* **9**, 97–101 https://doi.org/10.1016/S0966-842X(01)01956-4
43 Dieye, Y., Oxaran, V., Ledue-Clier, F., Alkhalaf, W., Buist, G., Juillard, V. et al. (2010) Functionality of sortase a in *Lactococcus lactis*. *Appl. Environ. Microbiol.* **76**, 7332–7337 https://doi.org/10.1128/AEM.00928-10
44 Muñoz-Provencio, D., Rodríguez-Díaz, J., Collado, M.C., Langella, P., Bermúdez-Humarán, L.G. and Monedero, V. (2012) Functional analysis of the *Lactobacillus casei* BL23 sortases. *Appl. Environ. Microbiol.* **78**, 8684–8693 https://doi.org/10.1128/AEM.02287-12
45 Kebouchi, M., Galia, W., Genay, M., Soligot, C., Lecomte, X., Awussi, A.A. et al. (2016) Implication of sortase-dependent proteins of *Streptococcus thermophilus* in adhesion to human intestinal epithelial cell lines and bile salt tolerance. *Appl. Microbiol. Biotechnol.* **100**, 3667–3679 https://doi.org/10.1007/s00253-016-7322-1
46 Suzuki, N., Okayama, S., Nonaka, H., Tsuge, Y., Inui, M. and Yukawa, H. (2005) Large-scale engineering of the *Corynebacterium glutamicum* genome. *Appl. Environ. Microbiol.* **71**, 3369–3372 https://doi.org/10.1128/AEM.71.6.3369-3372.2005
47 Becker, J. and Wittmann, C. (2012) Bio-based production of chemicals, materials and fuels: *Corynebacterium glutamicum* as versatile cell factory. *Curr. Opin. Biotechnol.* **23**, 631–640 https://doi.org/10.1016/j.copbio.2011.11.012
48 Wendisch, V.F., Jorge, J.M.P., Pérez-García, F. and Sgobba, E. (2016) Updates on industrial production of amino acids using *Corynebacterium glutamicum*. *World J. Microbiol. Biotechnol.* **32**, 105 https://doi.org/10.1007/s11274-016-2060-1
49 Yokoyama, K., Utsumi, H., Nakamura, T., Ogaya, D., Shimba, N., Suzuki, E. et al. (2010) Screening for improved activity of a transglutaminase from *Streptomyces mobaraensis* created by a novel rational mutagenesis and random mutagenesis. *Appl. Microbiol. Biotechnol.* **87**, 2087–2096 https://doi.org/10.1007/s00253-010-2656-6
50 Matsuda, Y., Itaya, H., Kitahara, Y., Theresia, N.M., Kutukova, E.A., Yomantas YA, V. et al. (2014) Double mutation of cell wall proteins CspB and PBP1a increases secretion of the antibody Fab fragment from *Corynebacterium glutamicum*. *Microb. Cell Fact.* **13**, 1–10 https://doi.org/10.1186/1475-2859-13-56
51 Hemmerich, J., Moch, M., Jurischka, S., Wiechert, W., Freudl, R. and Oldiges, M. (2019) Combinatorial impact of Sec signal peptides from *Bacillus subtilis* and bioprocess conditions on heterologous cutinase secretion by *Corynebacterium glutamicum*. *Biotechnol. Bioeng.* **116**, 644–655 https://doi.org/10.1002/bit.26873
52 Caspers, M. and Freudl, R. (2008) *Corynebacterium glutamicum* possesses two secA homologous genes that are essential for viability. *Arch. Microbiol.* **189**, 605–610 https://doi.org/10.1007/s00203-008-0351-0
53 Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 https://doi.org/10.1093/molbev/msy096
54 Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) The rapid generation of mutation data matrices. *Comput. Appl. Biosci.* **8**, 275–282. https://doi.org/10.1093/bioinformatics/8.3.275
55 Antos, J.M., Chew, G.L., Guimaraes, C.P., Yoder, N.C., Grotenbreg, G.M., Popp, M.W.L. et al. (2009) Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc.* **131**, 10800–1 https://doi.org/10.1021/ja902681k
56 Bentley, M.L., Lamb, E.C. and McCafferty, D.G. (2008) Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from *Staphylococcus aureus*. *J. Biol. Chem.* **283**, 14762–14771 https://doi.org/10.1074/jbc.M800974200

The Cell Surface 7 (2021) 100055

Contents lists available at ScienceDirect

The Cell Surface

journal homepage: www.sciencedirect.com/journal/the-cell-surface

The divergent roles of sortase in the biology of Gram-positive bacteria

Aliyath Susmitha [a,b], Harsha Bajaj [a], Kesavan Madhavan Nampoothiri [a,b,*]

[a] Microbial Processes and Technology Division, CSIR – National Institute for Interdisciplinary Science and Technology (NIIST), Trivandrum 695019, Kerala, India
[b] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

ARTICLE INFO

Keywords:
Sortase
Cell wall
Gram-positive
Pathogenic
Non-pathogenic

ABSTRACT

The bacterial cell wall contains numerous surface-exposed proteins, which are covalently anchored and assembled by a sortase family of transpeptidase enzymes. The sortase are cysteine transpeptidases that catalyzes the covalent attachment of surface protein to the cell wall peptidoglycan. Among the reported six classes of sortases, each distinct class of sortase plays a unique biological role in anchoring a variety of surface proteins to the peptidoglycan of both pathogenic and non-pathogenic Gram-positive bacteria. Sortases not only exhibit virulence and pathogenesis properties to host cells, but also possess a significant role in gut retention and immunomodulation in probiotic microbes. The two main distinct functions are to attach proteins directly to the cell wall or assemble pili on the microbial surface. This review provides a compendium of the distribution of different classes of sortases present in both pathogenic and non-pathogenic Gram-positive bacteria and also the noteworthy role played by them in bacterial cell wall assembly which enables each microbe to effectively interact with its environment.

1. Introduction

The cell wall of Gram-positive bacteria contains a multi-layered cellular component that acts as a cytoskeletal element for maintaining physical integrity and also acts as a scaffold for displaying a large number of surface proteins mediated by sortase enzymes. Sortases are membrane-bound transpeptidases that cleave the sorting signal of the secreted protein to form an isopeptide bond between the secreted protein and peptidoglycan. They are either responsible for covalently anchoring specific surface proteins or polymerizing pilin sub-units to form a proteinaceous structure termed pili (Hendrickx et al., 2011). Sortase-displayed surface structures play a pivotal role in displaying virulence and pathogenesis properties without affecting the growth and viability of cells. They are responsible for cell attachment, heme transport, nutrient uptake, sporulation and aerial hyphae formation (Cheng et al., 2009; Weiss et al., 2004) (Fig. 1). The surface proteins recognized by the sortase enzyme contain a C-terminal pentaglycine recognition motif followed by a stretch of hydrophobic amino acids and a positively charged tail (Schneewind et al., 1992).

Sortases are classified into six different classes (A-F), based on their primary sequences (Spirig et al., 2011) (Table 1). Class A sortase is well characterized and found mostly in low GC content Gram-positive bacteria. They play a housekeeping role in anchoring a variety of functionally distinct surface proteins with an LPXTG recognition sequence. Class B sortase actively participates in iron acquisition by recognizing iron transporter proteins with a NPQTN motif and covalently anchors them to the cell wall. Class C sortases are responsible for constructing complex pili polymers by recognizing the LPXTG motif. Class D sortases display proteins containing a LPXTA recognition motif on the cell wall that enables spore formation. GC-rich actinobacteria, in particular *Corynebacterium* and *Streptomyces* spp. contain Class E sortase which performs a similar function to Class A sortases. Class E sortases are involved in anchoring surface proteins and aerial hyphae proteins by recognizing a LAXTG sorting motif. Class F sortases were initially reported in *Propionibacterium acnes* which contains an LPXTG sorting signal similar to sortase A (Girolamo et al., 2019). All sortases contain a His-Cys-Arg as the catalytic triad, which catalyzes the transpeptidation reaction (Frankel et al., 2005, 2007; Perry et al., 2002). The cysteine in the active site of the enzyme is involved in bond cleavage and formation of a stable thioacyl intermediate that is relieved by the nucleophilic attack of the amino group (pentaglycine crossbridge) in peptidoglycan synthesis precursors (Mazmanian et al., 1999). In site-directed mutagenesis, the replacement of cysteine (at position 184 in S. aureus) with an alanine abolishes sortase catalytic activity *in vitro* and *in vivo*. The

* Corresponding author at: Microbial Processes and Technology Division, CSIR – National Institute for Interdisciplinary Science and Technology (NIIST), Trivandrum 695019, Kerala, India.
E-mail address: madhavan@niist.res.in (K. Madhavan Nampoothiri).

https://doi.org/10.1016/j.tcsw.2021.100055
Received 16 March 2021; Received in revised form 11 June 2021; Accepted 11 June 2021
Available online 13 June 2021

dual acid/base role was carried out by His residue which donates a proton to the leaving amide nitrogen during the cleavage reaction and the second substrate accepts the proton from the amino group to allow the nucleophilic attack by the unprotonated amine. The Arg side chain is implicated in substrate binding and possibly in stabilisation of a presumed oxyanion intermediate (Frankel et al., 2007).

Among different classes of sortases, the SrtA of *Staphylococcus aureus* is the pioneer for understanding the mechanism of these enzymes. After recognition of cell wall surface protein, SrtA catalyzes two sequential reactions: (i) thioesterification and (ii) transpeptidation. The enzyme first recognizes the pentaglycine sequence on the surface proteins, which are being secreted through the cytoplasmic membrane. The pentaglycine sequence contains an LPXTG motif at the C-terminus of the protein. In the second step, the SrtA cleaves the scissile bond between threonine and glycine residues to form an acyl-enzyme intermediate which subsequently transfers the carboxyl of threonine which is amide-linked to the pentaglycine cross-bridge of lipid II (Marraffini et al., 2006). Finally, the lipid II surface protein complex gets incorporated into the peptidoglycan by means of transglycosylation and transpeptidation reactions (Paterson and Mitchell, 2004; Spirig et al., 2011). The sortase enzyme accepts the nucleophiles which might vary in different Gram-positive bacteria, as the composition of peptidoglycan layers in the cell envelope vary from strain to strain..

Sortases are not only restricted to Gram-positive bacteria, however some sortase genes and its potential substrates are also found in Gram-negative bacteria which are not well characterized. Some Gram-negative bacteria such as *Shewanella putrefascienes*, *Shewanella oneidensis*, *Microbulbifer degradans*, *Colwellia psychrerythraea* and *Bradyrhizobium japonicum* consist of a gene encoding a single sortase-like protein and a potential sortase substrate (Comfort and Clubb, 2004).

Although there are many studies on sortases, referencing virulence and colonization factors, there are only a few reports on sortases that display proteins in non-pathogenic bacteria which includes food grade microbes of the lactic acid bacteria (LAB) and in *Corynebacterium glutamicum,* an industrial important microbe for the production of amino acids. In non-pathogenic bacteria, especially in probiotics, the sortase plays a pivotal role in eliciting health benefits to the host. Distinctive sortases in pathogenic and non-pathogenic bacteria are shown in Table 2 and this review provides a description about the virulence and functional aspects of sortases reported in each Gram-positive species that has been published.

Table 1
Sortase classification.

Sortase Class	Cleavage site	Main Function	Bacterial genus
A	LPXTG	Surface protein anchoring	*Staphylococcus, Listeria, Streptococcus, Bacillus, Clostridium, Entrobacter, Lactobacillus*
B	(N/S/P) PXTG	Heme uptake	*Bacillus, Listeria, Bacillus,*
C	LPXTG	Pili assembly	*Corynebacterium, Streptococcus, Clostridium, Actinomyces, Entrobacter, Lactobacillus*
D	LPXTA	Spore formation	*Bacillus*
E	LAXTG	Aerial hyphae formation, Surface protein anchoring, Pilus attachment	*Corynebacterium, Streptomyces, Actinomyces*
F	LPXTG	Unknown	*Propionibacterium*

2. Pathogenic bacteria

2.1. *Staphylococcus aureus*

S. aureus is one of the most predominant pathogens responsible for causing mastitis, an inflammation of breast tissue (Duarte et al., 2015), skin infection, pneumonia, sepsis, and endocarditis in humans (Zhang et al., 2015). However, most of these infections are caused by the anchoring of a vast array of virulence-associated surface proteins to the cell wall, which are catalyzed by a cysteine transpeptidase enzyme called Sortase A (Mazmanian et al., 2000). *S. aureus* sortase A (SrtA) has

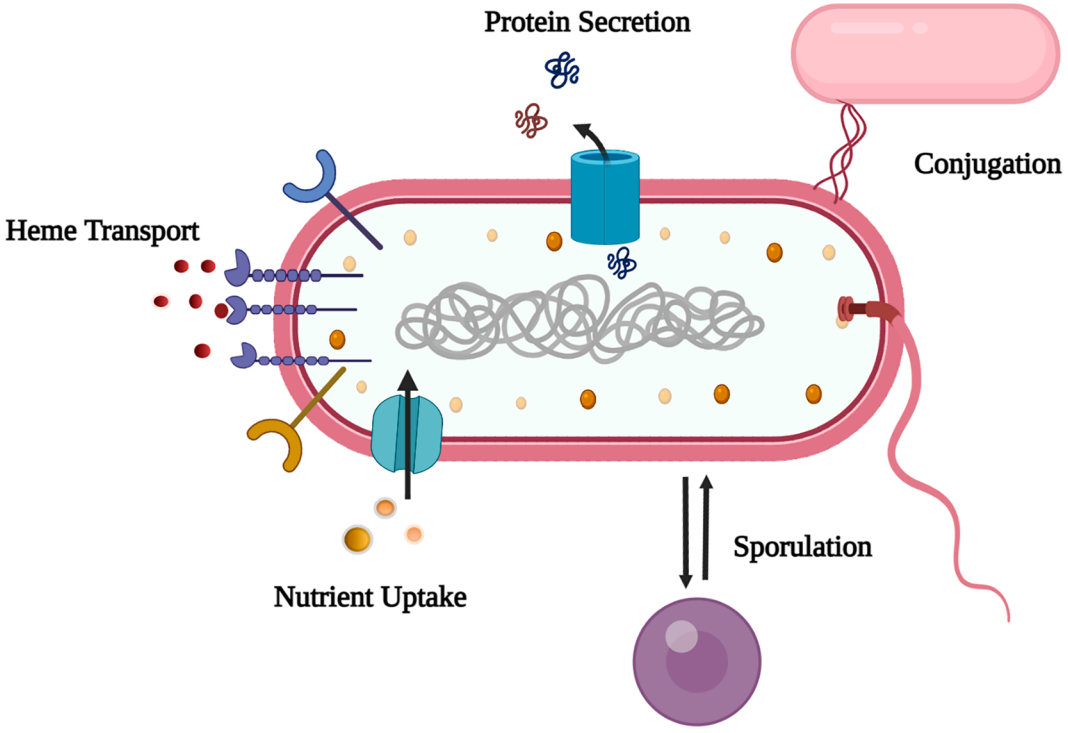


Fig. 1. General functions of sortases in a bacterial cell wall. Sortases are involved in pili formation, cell attachment to the host tissues, anchoring surface proteins to the cell wall, spore formation, uptake of nutrients and iron from the surrounding environment.

Table 2
Distinctive sortases in pathogenic and non-pathogenic bacteria.

Bacterial Species	Sortase protein	Putative CWSS proteins	References
Actinomyces oris	SrtA, SrtC (SrtC1 and SrtC2)	14	(Wu et al., 2011, 2012)
Bacillus anthracis	SrtA, SrtB, SrtC	10	(Gaspar et al., 2005; Maresso et al., 2006; Marraffini et al., 2006)
Bacillus cereus	SrtA, SrtB, SrtC (SrtC1 and SrtC2), SrtD	2	(Budzik et al., 2007)
Bacillus subtilis	YhcS	2	(Duc Nguyen et al., 2011)
Bifidobacterium bifidum	SrtA	14	(Westermann et al., 2012; Wei et al., 2016)
Clostridium perfringens	SrtB	13	(Boekhorst et al., 2005; van Leeuwen et al., 2014)
Corynebacterium diphtheriae	SrtA, SrtB, SrtC, SrtD, SrtE, SrtF	17	(Mandlik et al., 2007)
Corynebacterium glutamicum	SrtE	1	(Susmitha et al., 2019)
Enterobacter faecalis	SrtA, SrtC	41	(Kemp et al., 2007)
Lactobacillus plantarum	SrtA	32	(Pretzer et al., 2005)
Lactobacillus rhamnosus	SrtA, SrtC (SrtC1 and SrtC2)	6	(Douillard et al., 2014)
Lactococcus lactis	SrtA, SrtC	14	(Dieye et al., 2010)
Lactobacillus salivarius	SrtA	10	(Van Pijkeren et al., 2006)
Lactobacillus casei	SrtA (SrtA1 and SrtA2), SrtC (SrtC1 and SrtC2)	23	(Muñoz-Provencio et al., 2012)
Lactobacillus acidophilus	SrtA	12	(Call et al., 2015)
Lactobacillus gasseri	SrtA	12	(Call et al., 2015)
Lactobacillus delbrueckii	SrtA	2	(Van Pijkeren et al., 2006)
Listeria monocytogenes	SrtA, SrtB	43	(Bierne et al., 2002; Garandeau et al., 2004)
Propionibacterium acnes	SrtF	4	(Girolamo et al., 2019; Lodes et al., 2006)
Staphylococcus aureus	SrtA, SrtB	22	(Mazmanian et al., 2000, 2002)
Streptococcus suis	SrtA, SrtB, SrtC, SrtD, SrtE	3	(Osaki et al., 2002; Lu et al., 2011)
Streptococcus pyogenes	SrtA, SrtB	15	(Barnett and Scott, 2002)
Streptococcus agalactiae	SrtA, SrtC (SrtC1, SrtC2, SrtC3 and SrtC4)	35	(Lalioui et al., 2005; Dramsi et al., 2006)
Streptococcus pneumoniae	SrtA, SrtC (SrtC1, SrtC2 and SrtC3)	16	(Kharat and Tomasz, 2003; Lemieux et al., 2008)
Streptomyces avermitilis	SrtE (SrtE3)	16	(Das et al., 2017)
Streptococcus mutans	SrtA	6	(Lee and Boran, 2003)
Streptomyces coelicolor	SrtE (SrtE1 and SrtE2)	17	(Duong et al., 2012)
Streptococcus gordonii	SrtA	7	(Davies et al., 2009; Nobbs et al., 2007)
Streptococcus sanguinis	SrtA	32	(Yamaguchi et al., 2006; Turner et al., 2009)
Streptococcus uberis	SrtA	10	(Egan et al., 2010; Eigh et al., 2010)
Streptococcus thermophilus	SrtA	2	(Kebouchi et al., 2016)

been the prototype for understanding the mechanism of action of these enzymes (Mazmanian et al., 1999). The SrtA of *S. aureus* covalently anchors the surface proteins onto the bacterial cell wall via a C-terminal cell wall sorting signal with an LPXTG recognition motif followed by a stretch of hydrophobic amino acids and a positively charged tail (Das et al., 2017; Novick, 2000). The cell wall anchoring proteins are synthesized within the cytoplasm and translocated across the membrane through the Sec machinery. The sortase recognizes the anchoring proteins followed by a nucleophilic attack at the active site of the cysteine and cleaves the C-terminal of LPXTG motif between threonine and glycine forming a thioester intermediate complex which is then covalently anchored on the pentaglycine cross-bridge of lipid II. The lipid II protein complex then gets attached to the cell wall via transglycosylation and transpeptidation reactions (Perry et al., 2002) (Fig. 2). *S. aureus* attaches several surface proteins which are characterized by a C-terminal LPXTG motif, including protein A (Spa), two fibronectin-binding proteins (FnbpA and FnbpB), two clumping factors (ClfA and ClfB), a collagen-binding protein (Cna), and three serine-aspartate repeat proteins (SdrC, SdrD, and SdrE). The deletion of SrtA has led to the failure of surface protein anchoring to the cell wall (Clancy et al., 2010). The mechanism of sortase B is similar to that of sortase A, where *S. aureus* sortase B enzyme attaches the heme transporter IsdC protein which is a major component of the iron-regulated surface determinant system that scavenges the heme–iron from hemoglobin. The SrtB anchors IsdC to uncross-linked peptidoglycan instead of heavily cross-linked peptidoglycan. The *srtB* and *isdC* genes are located together in the *isd* iron-acquisition operon. However, in contrast to SrtA, SrtB recognizes NPQTN sorting signals from *S. aureus* (Mazmanian et al., 2002). Gene knockout studies in *S. aureus* revealed that the abolition of *srtB* gene is responsible for virulence and does not affect cell viability.

2.2. *Corynebacterium spp.*

C. diphtheriae is the etiological agent of pharyngeal diphtheria in humans (Hadfield et al., 2000). The genome of *C. diphtheriae* NCTC13129 harbors six sortases like genes (named *srtA-F*), five of which presumably assembles three distinct types of pilus structures- SrtA for the SpaA-type pilus, SrtB or SrtC for the SpaD-type pilus, and SrtD or SrtE for the SpaH-type pilus (Spa for sortase-mediated pilus assembly) which are polymerized by specific Class C sortases and SrtF which belongs to class A sortase, catalyzes the anchoring of pilin monomers on the bacterial surface (Gaspar and Ton-That, 2006; Swaminathan et al., 2007; Ton-That and Schneewind, 2003). All three pilus structures share a similar architecture, a major pilin (designated as SpaA, SpaD, and SpaH) along the pilus shaft joined to the minor pilins (designated as SpaB, SpaC, SpaE, SpaF, SpaI and SpaG) located at the tip and base of the pilin (Fig. 3). To analyze the functions of sortase in pili formation, Mandlik and his coworkers constructed an isogenic mutant strain of NCTC13129 devoid of all six sortase genes (*srtA–F* mutant) that exhibited a severe defect in adherence to epithelial cells.

The single pilus-specific SrtA encoded within the *spaA* gene cluster specifically catalyzes the covalent crosslinking of individual pilin monomers and also anchors pili to the cell wall. To analyze the functions of SrtA, immunoelectron microscopy, and biochemical analysis showed that a strain expressing only SrtA secretes significant amounts of polymerized pilins into the culture medium, indicating that one or more sortases might be involved for efficient cell wall anchoring of pili (Mandlik et al., 2007; Ton-That et al., 2004). Indeed, the strain with the deletion of housekeeping gene *srtF* releases SpaA polymers into the culture medium. Thus, two sortases are involved in pilus biogenesis, a pilus-specific sortase for pilin polymerization and the housekeeping sortase for efficient anchoring of pili to the cell wall (Mandlik et al., 2010; Swaminathan et al., 2007). The deletion of *srtA* or *spaA* gene completely abrogates the assembly of SpaA pili and deletion of *spaC* and *spaB* did not abolish SpaA pilus formation. This evidence suggests that SrtA catalyzes the assembly of SpaA pilus and SpaA alone is sufficient to

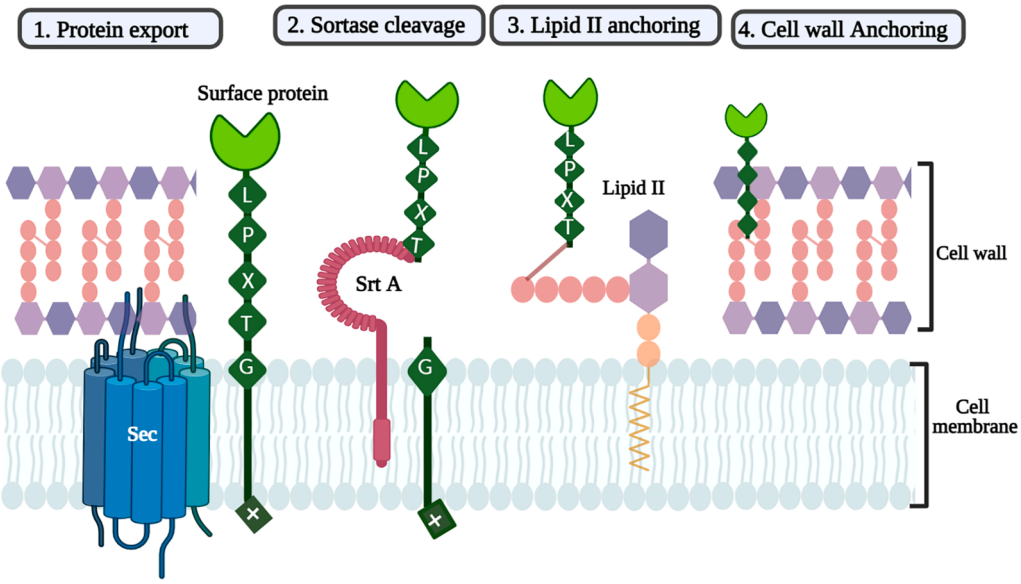


Fig. 2. Illustration of sortase A transpeptidation reaction in *S. aureus*. 1) Protein synthesized from the cytosol gets translocated through the Sec machinery and gets anchored on the cell membrane. 2) Sortase recognizes the C-terminal of the LPXTG sorting motif and cleaves between threonine and glycine. 3) Sortase forms a protein complex and undergoes a nucleophilic reaction from lipid II molecule. 4) Lipid II– protein molecule gets further anchored to the cell wall *via* transpeptidation reaction.

mediate the polymerization of a secreted protein (Marraffini et al., 2006; Ton-That and Schneewind, 2003).

Unlike the SpaA-type pili, which are assembled by a single sortase, two sortases; SrtB and SrtC catalyze the assembly of the SpaD-type pili. The deletion of *srtB* alone or both *srtB* and *srtC* abrogated the incorporation of SpaE into SpaDF pili. These results demonstrate that SpaDEF pilus assembly specifically requires SrtB for the incorporation of SpaE into SpaDF pili and whose assembly requires either SrtB or SrtC (Gaspar and Ton-That, 2006).

Likewise, the SpaH pilus is independently assembled and different from the other two corynebacterial pili. The SrtD specifically required for the incorporation of SpaH into SpaIG pili, whose assembly requires either SrtD or SrtE, while other remaining sortases are dispensable (Swierczynski and Ton-That, 2006). Thus, the housekeeping sortase contributes to efficient cell wall anchoring with other sortases involved in SpaD and SpaH–type pilus (Swaminathan et al., 2007). *Corynebacterium* spp also contains an industrially important non-pathogenic microbe, *C. glutamicum* which encodes a single sortase enzyme. The sortase enzyme shows high substrate specificity towards the LAXTG sorting sequence of class E sortase. The two-dimensional structure of Cg-SrtE was found to be similar to *C. diphtheriae* sortase. The Cg-SrtE was biochemically characterized and shows a Ca^{2+} independent catalytic mechanism (Susmitha et al., 2019). The high catalytic efficiency with LAXTG substrate and the Ca^{2+} independency, allows such non-pathogenic Cg-SrtE to be used in sortagging applications such as protein immobilization, for sortase assay kit and as a self-cleaving tag for protein purification etc., which makes the enzyme more robust than the pathogenic *S. aureus* SrtA variants.

2.3. *Listeria monocytogenes*

L. monocytogenes is a facultative intracellular food-borne Gram-positive bacterium, responsible for life-threatening infections in humans and animals. It is a causative agent of listeriosis, which is characterized by gastroenteritis, meningitis, encephalitis, bacteremia, soft tissue, and parenchymal infections, and mother-to-fetus transmission (Dussurget et al., 2004; Posfay-barbe and Wald, 2009). Among the Gram-positive bacteria, the genus *Listeria* contains the highest number of genes encoding surface proteins in the range of 40–45 with an LPXTG bearing motif at the C-terminal, 2 surface proteins containing an NPKSS/NAKTN motif (Lmo2185 and Lmo2186), and two sortases (SrtA and SrtB) (Boekhorst et al., 2005; Cabanes et al., 2002; Cossart, 2007; Garandeau et al., 2004). Inactivation of the *srtA* gene in *L. monocytogenes* altered the expression of specific anchored surface proteins containing the canonical LPXTG motif, ultimately decreasing the ability of the bacterial adhesion, invasion of eukaryotic cells, and affects host immune responses (Bierne et al., 2002). The RT-PCR and Western-blot data analysis demonstrates that the lack of SrtA alters the expression of LPXTG surface proteins and does not completely abolish the strong attachment of certain surface proteins to cell-wall peptidoglycan (Mariscotti et al., 2012). Therefore, as in *S. aureus,* the listerial SrtB represents the second class of sortase in *L. monocytogenes,* generally expressed in operons containing genes encoding their substrates with NPKSS/NAKTN recognition motifs. The *srtB* deletion mutants do not have defects in bacterial entry, growth, or motility in tissue-cultured cells and do not show attenuated virulence in mice. SrtB-mediated anchoring could therefore be required to anchor surface proteins involved in the adaptation of this microorganism to different environmental conditions (Garandeau et al., 2004).

2.4. *Streptococcus spp.*

Streptococcus pyogenes is a human pathogen which causes life-threatening diseases such as necrotizing facititis, septicemia and toxic shock syndrome which results in 500,000 death per year (Dekker and Boersma, 2018). *S. pyogenes* contains SrtA which localizes LPXTG cell surface virulent factors; M protein, GRAB, protein F, and ScpA. The other class of enzyme SrtB anchors T6 protein and its role on pathogenesis is still unclear (Barnett and Scott, 2002). The deletion of sortase A mutant leads to accumulation of surface proteins onto the cell wall, aberrant morphology, reduced growth, and increased membrane permeability (Raz et al., 2015).

Streptococcus agalactiae is known for life-threatening neonatal infections, such as pneumonia, sepsis, and meningitis (Lalioui et al., 2005). It is a commensal bacterium predominantly found in colonization with gastrointestinal and genitourinary tracts. It is a serious cause for mortality or morbidity in pregnant and non-pregnant adults suffering from significant underlying diseases (Lalioui et al., 2005; Sendi and Johansson, 2008). The genome analysis of *S. agalactiae* NEM316 encompasses one class A, four class C sortases (Dramsi et al., 2005), and 35 surface proteins containing a cell wall sorting signal motif (26 proteins had an LPXTG motif, 4 had an IPXTG motif, 2 had an LPXTS motif, 2 had an LPXTN motif, and 1 had an FPXTG motif (Glaser et al., 2002). *S. agalactiae* NEM316 strain lacking *srtA* gene was found defective in

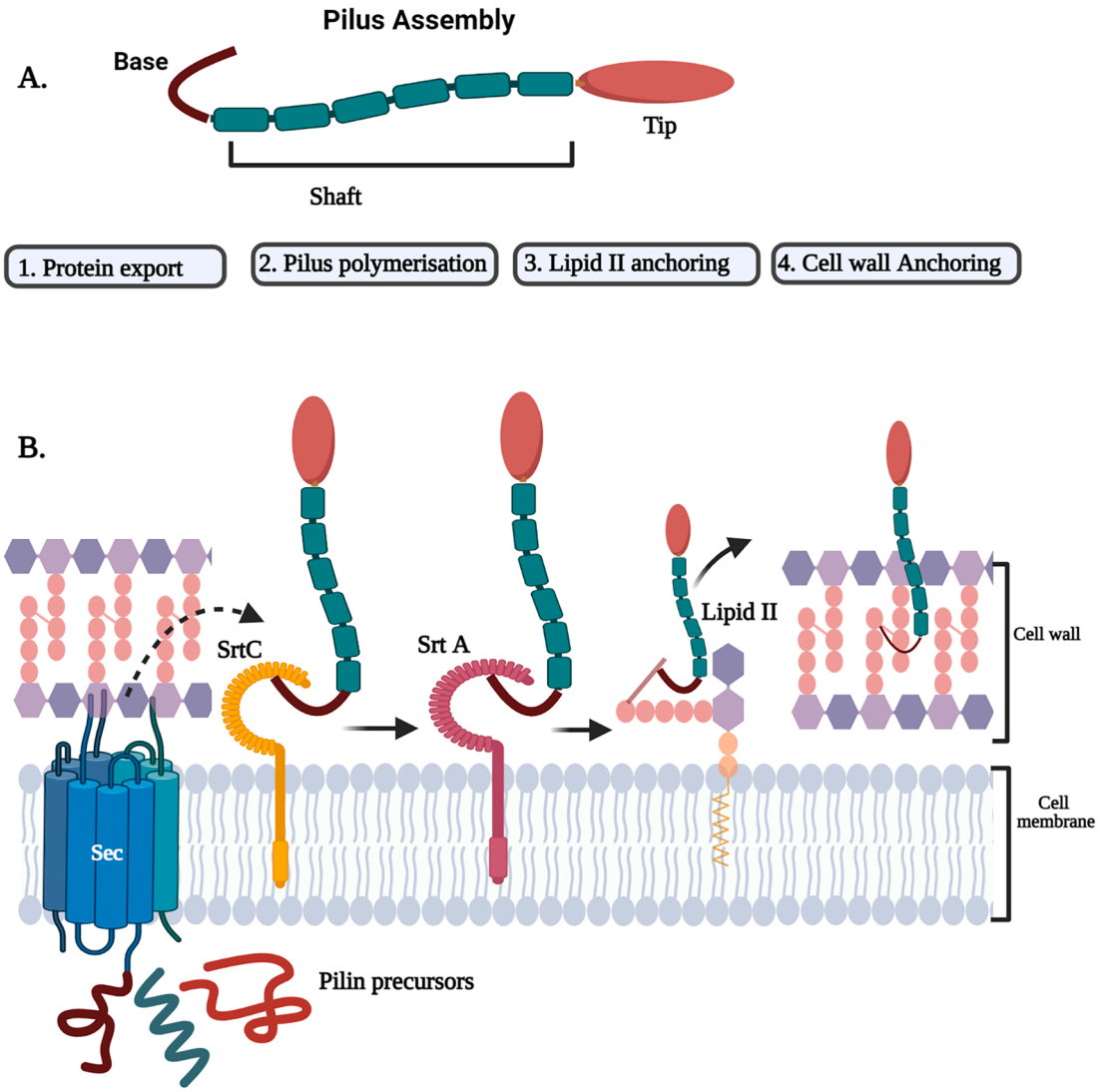


Fig. 3. Schematic illustration of Sortase-mediated pili. A) Pilus assembly which consist of tip pilin (red), Shaft (green) and the base (brown). B. 1) The pilin precursors synthesized from the cytosol enters through the sec machinery. 2) The pilin- specific sortase SrtC (yellow) recognizes the tip pilin and cleaves at the sorting motif and forms an acyl sortase complex. The pilus specific sortase receives a nucleophilic attack from the lysine side chain from the backbone pilin to form a covalent bond between the pilins and undergoes pilin polymerization. C) The housekeeping sortase undergoes a nucleophilic attack from lipid II molecule. 4) The polymerized pilin is further anchored to the cell wall. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

anchoring cell surface proteins Alp2 (GBS 0470) and ScpB (GBS1308) bearing LPXTG and LPXTN signature sequence at the C-terminal (Lalioui et al., 2005). To determine the activity of SrtA in fibronectin and fibrinogen binding, a simple binding assay (ELISA) was performed to compare the binding properties of SrtA$^-$ mutant with those of wild-type and complemented strains. This resulted in reduced binding of fibronectin and fibrinogen in SrtA$^-$ mutant strain. Thus, it is conceivable that the ScpB a fibronectin protein (Beckmann et al., 2002), and FbsA (GBS1087) a major GBS fibrinogen-binding protein is a SrtA-dependent LPXTG-containing protein (Schubert et al., 2004, 2002). The inactivation of *srtA* in NEM316 strain decreased its adherence to human epithelial cell lines (A549, Caco-2, and HeLa) and rat cell lines (L2). Interestingly, the deletion of *srtA* strains did not alter the virulence in the neonatal rat sepsis model as compared to the wild-type parental strain (Lalioui et al., 2005).

Besides SrtA, four genes encoding class C sortases (SrtC) were found in NEM316, 2603 V/R, and A909 genome sequences which are arranged tandemly in two different loci, *srtC1-srtC2* and *srtC3-srtC4* coding for pilus biogenesis (Dramsi et al., 2006; Khare et al., 2011). Based on the electron microscopy and immunogold labeling, the NEM316 strain assembles pili from *the srtC3-srtC4* locus and encodes three pilins subunits, the major pilin, and two minor pilins. Either SrtC3 or SrtC4 is required for polymerization of pili and housekeeping SrtA anchors the polymerized pili to the cell wall (Dramsi et al., 2006).

Streptococcus pneumoniae is a human pathogen responsible for multiple infections, including otitis media, meningitis, pneumonia, and septicemia (Lemieux et al., 2008). The genome of *S. pneumoniae* TIGR4 and R6 contains *srtA* gene and in addition to that TIGR4 also contains *srtC-1, srtC-2,* and *srtC-3* for pilus assembly. The inactivation of *srtA* from *S. pneumoniae* was shown to affect the localization of β-galactosidase and neuraminidase (NanA) surface proteins and decreased bacterial adherence and invasion to human pharyngeal cells *in vitro*. On the other hand, *srtA* inactivation did not affect the virulence of capsular type III strain of *S. pneumoniae* in the mouse intraperitoneal model (Kharat and Tomasz, 2003; Paterson and Mitchell, 2006). The *rlrA* pathogenicity islet of *S. pneumoniae* encodes three SrtC isoforms and three structural PI-1 subunit proteins; RrgA, RrgB, and RrgC. The RrgB pilus subunit forming the major backbone of pilus which comprises a pilin motif, E box, and C-terminal cell wall sorting signal (CWSS). RrgA found at the tip of the pilus and involved in pilus adhesion. RrgC anchors assembled pilus

to the cell wall in association with SrtA. The primary SrtC-1 catalyzes the polymerization of major pilin subunit RrgB. SrtC-2 binds with RrgA and attaches to other pilins. SrtC-3 preferentially binds with RrgC pilin subunit but does not have a strong affinity as SrtC-1 with RrgB (Lemieux et al., 2008; Naziga and Wereszczynski, 2017; Shaik et al., 2014).

Streptococcus mutans plays a significant role in the development of human dental cavities (Hamada and Slade, 1980). The importance of SrtA in modulating the cell-surface-related properties by surface anchoring proteins; WapA, Pac, GbpC, Dex, and FruA in *S. mutans* was confirmed by constructing a *srtA* mutant which was found to be non-adherent, non-aggregating, and less hydrophobic than the *srtA* complemented strain. The pathogenesis of SrtA in Streptococci was demonstrated in a rat model of infection, where the inactivation of *srtA* was incapable of colonizing the oral mucosa in the absence of sucrose and in the presence of sucrose colonization was found to be less effective (Lee and Boran, 2003). These phenomena provides evidence that SrtA could be a novel and attractive drug target for the prevention of cariogenicity (Lee and Boran, 2003; Murai et al., 2005).

Streptococcus gordonii is a non-cariogenic pioneer colonizer in the development of dental plaque (Kuboniwa et al., 2006). They are found at multiple sites within the oral cavity and adhere to teeth, as well as mucosa, through the interaction between macromolecules on the bacterial cell wall and proteins or glycoproteins on the oral surfaces (Aas et al., 2005; Nobbs et al., 2007). *S. gordonii* consists of SrtA which anchors surface adhesins (SspA/B) with an LPXTG containing recognition motifs, hydrophobic spanning regions and positively charged tail at the C-terminal. The disruption of *srtA* gene changes the localization and function of SspA and SspB adhesins, reduces the biofilm formation and binding to specific salivary agglutinin receptor *in vitro* (Davies et al., 2009).

Streptococcus sanguinis is a member of the oral mitis group of Streptococci and the initial colonizer for dental plaque formation (Kolenbrander et al., 1993). Although, during oral injuries, the harmless members of the group invade into the bloodstream causing bacteremia and infective endocarditis (Morita et al., 2014). The deficiency of *srtA* in *S. sanguinis* causes an overall reduction in virulence in association with cell surface proteins and decreased cell surface hydrophobicity. Thus, SrtA of oral streptococci is considered an important molecule for colonization on the smooth surface of the teeth and a drug target to prevent dental biofilm formation (Yamaguchi et al., 2006).

Streptococcus uberis is one of the most common pathogens associated with the lactating bovine mammary gland and impacts on animal health, welfare, and economics of milk production (Bradley et al., 2007; Egan et al., 2010; Eigh et al., 2010). *S. uberis* contains only a single copy of sortase A (*srtA*), encoding a transamidase capable of anchoring surface proteins bearing the LPXTG or LPXXXD motifs at the bacterial cell surface (Egan et al., 2010). The *srtA* deficient strain of *S. uberis* was unable to colonize the bovine mammary gland to induce clinical mastitis in dairy cattle indicating that a number of SrtA-anchored proteins are likely to be involved in the pathogenesis of this bacterium (Egan et al., 2010).

Streptococcus suis is a chain-forming Gram-positive bacteria that causes meningitis in pigs and responsible for the economic losses to the swine industry. The zoonotic pathogen also emerges to be a major risk to humans working in the pig industry (Fittipaldi et al., 2012). *S. suis* possess six genes (*srtA, srtB, srtC, srtD, srtE and srtF*) which encode proteins similar to sortase and sortase-like proteins of other streptococci. The *srtA* gene is linked adjacent to *gyrA*, the other three genes *srtBCD* are found in clusters sequentially within the genome and *srtE* and *srtF* located in a separate chromosome (Osaki et al., 2002). When compared to the wild-type, the deletion of *srtA* gene abolished the anchoring of two virulence-related proteins with LPXTG motifs, MRP (Muramidase-released protein) and Sao on the cell surface and drastically reduced its adherence to human epithelial cells (Wang et al., 2009). The enzymes of class C sortases are involved in pili assembly, of which *Streptococcus suis* sortase has an open-lid confirmation when the substrate binds to the enzyme (Lu et al., 2011).

2.5. *Bacillus spp.*

B. anthracis is a spore-forming, Gram-positive, soil-borne organism that causes lethal anthrax disease in humans (Mock and Fouet, 2001). *B. anthracis* encodes three sortase enzymes: Ba-SrtA, Ba-SrtB, and Ba-SrtC from class A, class B, and class D, respectively. Bioinformatics analysis has identified 9 to 11 putative CWSS proteins, depending on the analysed strain. *B. anthracis* variants lacking the *srtA* gene did not anchor the collagen-binding MSCRAMM (microbial surface components recognizing adhesive matrix molecules) BasC protein to the bacterial cell wall as opposed to its parent strain. Recombinant expressed and purified SrtA catalyzed the cleavage reaction with LPETG and LPATG peptides, consistent with the notion that *B. anthracis* SrtA is responsible for the cell wall anchoring of surface proteins with an LPXTG motif (Gaspar et al., 2005). GamR, a *B. anthracis* phage receptor, is known to be anchored by SrtA. A *srtA* mutant strain displays reduced susceptibility to this phage, whereas the double *srtA-srtB* or *srtA-srtC* and the triple mutant strains displayed similar susceptibility to the *srtA* mutant, indicating that GamR is anchored by SrtA (Davison et al., 2005). Using the GamR function of causing phage lysis, a genetic screen was performed by fusing GamR to proteins containing CWSS and analysed for *in vivo* anchoring to the cell wall. The strains containing the three chimeric proteins BasB, BasE, and BasJ, were resistant to lysis in the *srtA* mutant and complementation experiments restored phage lysis indicating they are anchored by Sortase A. Furthermore, BasA was shown to be anchored by SrtA using immunoblot analysis of peptidoglycan fractions (Aucher et al., 2011). Apart from the functional roles, Ba-srtA has been characterized structurally. The NMR-structure of Ba-SrtA shows several unique active site features that include the presence of an N-terminal extension that contacts the catalytically essential histidine which might be involved in lipid II recognition. Another feature very unique to this protein is a large structurally disordered active site loop correlated to the attachment of proteins to the *m*-DAP moiety of lipid II. Based on the NMR structure a lock-and-key mechanism is proposed for recognizing the sorting signal (Weiner et al., 2010)

B. anthracis srtB, which encodes sortase B, anchors IsdC to the cell wall envelopes of vegetative bacilli a heme–iron binding surface protein. Sortase B cleaves IsdC between the threonine and the glycine of its NPKTG motif sorting signal. Isogenic variants lacking either *srtB* or *isdC* display significant growth defects due to deficiencies in heme–iron scavenging, suggesting that IsdC binding to heme–iron in the cell wall envelope contributes to bacterial uptake of heme (Maresso et al., 2006). The crystal structure of *B. anthracis* sortase B *shows* β7/ β8 loop is structurally disordered similar to sortase A from *B. anthracis* (Zhang et al., 2004).

B. anthracis Class D SrtC anchors two substrates, BasH and BasI, to the cell wall of sporulating *B. anthracis*. LPNTA sorting signal of two substrates is cleaved by Sortase C (SrtC) at the C-terminal threonine (T) of the substrate to the amino group of DAP cross-bridges targeting the polypeptides to the cell wall of sporulating bacilli. Sortase C is also required for the formation of infectious spores in the tissues of animals. Where sortase C acts on two different substrates in two different subcellular compartments, the different surface proteins decorate different compartments of the sporulating cell, with BasI present in divisional cells and BasH present in forespores (Marraffini and Schneewind, 2007, 2006). The NMR solution structure of Sortase C shows structurally disordered surface loops (β2 − β3 and β4 − H1 loops) that surround the active site histidine, suggesting that they may play a key role in associating Ba-SrtC with its lipid II substrate similar to SrtA (Robson et al., 2012). Contradictory reports are found regarding the role of sortases in the virulence of the organism. Interestingly, the *srtA*, *srtB*, *srtC* mutant strains, deleted for sortase genes, are not affected in their virulence as compared to its wild-type parent strain (Fouet, 2009). It has been shown that disruption of either the *srtA* or *srtB* gene results in an

inability of the bacteria to grow in J774A.1 cells and no significant difference in the growth rate was observed in BHI broth between the parent strain and mutants. However, the molecular determinants causing this effect are not known (Zink and Burns, 2005).

B. cereus is a Gram-positive spore-forming bacteria mainly associated with food poisoning and also causes opportunistic skin infections (Shinagawa, 1990). From a bioinformatics analysis, it is known that *B. cereus* encodes Class A sortase (SrtA), Class B sortase (SrtB), Class C sortase (SrtD), and two class D sortases (SrtC1 and SrtC2). However, the function and substrates are only characterized for the Class A and Class C (Dramsi et al., 2005). The Class C sortase is involved in pili assembly especially in the *B. cereus* vegetative cells. The pilus operon consists of three genes (*bcpA-srtD-bcpB*) and deletion of the complete operon/inactivated *srtD* leads to the absence of the pili (Budzik et al., 2007). BcpA is polymerized by Sortase D even in the absence of the BcpB unit. BcpB the minor pilin is cleaved by Sortase D at C-terminal threonine of IPNTG sorting signal and then amide-linked to the YPKN motif of BcpA (Budzik et al., 2009). The cleavage reaction of BcpB is very specific to Sortase D and is not cleaved by sortase A. In the case of BcpA, the LPXTG sorting signal is cleaved at the C-terminal threonine and linked to the amino group of lysine in the YPKN motif of another BcpA sub-unit by sortase D. BcpA is a substrate for both sortase D and A, and sortase A cleaves the BcpA unit at its LPXTG to terminate the pilus assembly and bond to the cell wall cross-bridge (Budzik et al., 2009). Sortase B or Sortase C are not involved in the pili assembly or attachment to the *B. cereus* surface. Also, the operon *bcpA-srtD-bcpB* is not involved in the pili formation of the spore-forming cell surfaces (Budzik et al., 2007).

B. subtilis is generally recognized as a safe (GRAS) organism. *B. subtilis* codes for two putative sortases (probably belonging to class D), YhcS and YwpE, and two surface proteins, YhcR and YfkN, harboring sorting motifs supposed to be recognized by the putative sortase. YfkN contains the potential sorting signal LPDTA and in YhcR the motif is LPDTS (Duc Nguyen et al., 2011). *yhcS* gene is expressed preferentially at the late stationary phase and anchors YhcR on the cell wall of *B. subtilis* cell (Duc Nguyen et al., 2011; Fasehee et al., 2011). However the role or substrate of YwpE, the second putative sortase, is yet to be determined, the genomic analysis has shown that the open reading frame for *ywpE* encodes a truncated sortase-like protein (Liew et al., 2012).

2.6. Propionibacterium acnes

Sortase F is the only housekeeping sortase characterized in the genome of *P. acnes*. Sortase F from *P. acnes* shows a behavior similar to sortases from class A in terms of pH dependence, recognition sequence, and catalytic activity, furthermore its activity is independent of divalent ions, which contrasts to sortase A from *S. aureus*. The sortase F can be used as a powerful tool alternative to sortase A for protein engineering applications (Girolamo et al., 2019).

2.7. Clostridium spp.

C. perfringens, a pathogenic bacterium causes gas gangrene and food poisoning (Adak et al., 2002). Sortase B, C, and D are identified in *C. perfringens* and crystal structures reported (Suryadinata et al., 2015; Tamai et al., 2019, 2017). Pilin formation and its crystal structure are reported in *C. perfringens*. Molecular and biochemical analysis of sortase C mediated catalysis and polymerization of pilin is proposed (Tamai et al., 2019). SrtD in *C. perfringens* is required for effective conjugative transfer of plasmid like pCW3 carrying antibiotic-resistant genes (Revitt-Mills et al., 2021). The role of sortase is established in the pathogenesis of *C. perfringens* strains causing gas gangrene and necrotic enteritis (Choo et al., 2016; Lepp et al., 2021). Sortase-dependent pilus produced by the bacterium binds to collagen and causes necrotic enteritis in poultry. However, the sortase gene presenting the pilus on the cell-surface is yet to be known (Lepp et al., 2021).

C. difficile is a spore-forming and toxin-producing intestinal pathogen associated with high morbidity rates. It causes antibiotic-associated diarrhoea and severe inflammation in the intestines (Rupnik et al., 2009). *C. difficile* encodes for a single sortase enzyme sortase B (CD2718), belonging to the class B sortases, recognizes an entirely different sorting motif (SPxTG or PPxTG). Remarkably, unlike SrtB from other organisms, the function of this enzyme is not associated with heme or heme acquisition proteins instead possibly functions as a housekeeping sortase. The mutation of the catalytic cysteine or the addition of small inhibitors like MTSET abolishes the sortase B activity. *In vitro* ligation of a natural cell wall nucleophile, DAP is demonstrated (Donahue et al., 2014; van Leeuwen et al., 2014).

Based on bioinformatics analyses, seven putative sortase substrates have been identified (Van Leeuwen et al., 2014). However, two putative substrates, CD0183 and CD2768, containing an SPXTG motif were not cleaved or anchored to the cell wall by sortase in an experimental observation (Peltier et al., 2017). Another two substrates, CD0386 and CD3392, displaying very high (94%) sequence similarity, of which CD0386 is attached to the cell wall by the sortase as demonstrated by biochemical analysis of sub-cellular fractions in a sortase knockout strain (Chambers et al., 2015). Adhesins, CD2831 and CD3246, are anchored proteins on the cell wall by SrtB activity. The functional role, particularly of CD2831 in binding to collagen, has been demonstrated *in vitro* through the cleaving activity of a Pro-Pro endopeptidase to release CD2831 and CD3246 from the cell surface and is regulated by C-di-GMP (Hensbergen et al., 2015; Peltier et al., 2015). The last putative substrate CD2537 contains only a weak signal peptide (Corver et al., 2017).

2.8. Actinomyces spp

S. avermitilis encodes for at least four putative class E sortase enzymes (Duong et al., 2012). Only, SrtE3 (SAV4333) has been biochemically characterized, where it recognizes a LAXTG or LPXTG motif in a calcium-independent manner (Das et al., 2017).

S. coelicolor is a soil-dwelling multi-cellular bacterium that has three differentiated states in its life cycles: vegetative hyphae, aerial hyphae, and spores. It is known to encode for seven sortase enzymes two from sortase class E and five class F enzymes (Pallen et al., 2001). Sortase class E enzymes SrtE1 and SrtE2 recognize and cleave LAXTG-containing peptides *in vitro*. Chaplin proteins (Chp A, B, and C) required for the aerial development of this organism contain a LAXTG-sorting signal rendering them as putative sortase substrates (Elliot et al., 2003). Sortase class E enzymes (SrtE1 and SrtE2) anchor ChpC protein to cell wall *in vivo*. Indeed, *srtE1/srtE2* double mutants delay the formation of aerial hyphae with hindered sporulation and cease to display short Chaplin proteins (Duong et al., 2012). The crystal structure of SrtE1 has been determined at a1.93 Å resolution (Kattke et al., 2016).

A. oris is an oral bacterium, formerly known as *Actinomyces naeslundii,* which results in the formation of dental plaque (Persson, 2011). It consists of three cysteine transpeptidases, the housekeeping sortases (SrtA), which belongs to class E sortase (Spirig et al., 2011) and the other belongs to class C sortases, SrtC1 and SrtC2, which is involved in the assembly of two distinct forms of pili on the *Actinomyces* cell surface (Mishra et al., 2007). The SrtC1 is involved in the type I fimbriae formation which is further involved in the adhesion of *Actinomyces* to the tooth surface by recognition of proline-rich receptors (Wu et al., 2011). However, SrtC2 results in the assembly of type II fimbriae which are essential for the bacterial adhesion to the oral streptococci and host cells by recognizing the polysaccharide receptors (Wu et al., 2012).

The crystal structure of *A. oris* SrtA contains an open accessible active site for the attachment of surface proteins covalently bound to the cell envelope and fimbriae/pili-specific sortases (SrtC1 and SrtC2) contains a flexible lid to cover the active site which gets exposed during pili formation (Manzano et al., 2009; Persson, 2011).

The housekeeping sortases generally do not show any role in cell viability, however, SrtA of *A. oris* was observed to show some

contradictory results. The deletion of the *srtA* gene was found to be lethal for the bacterial cells, which resulted in excessive membrane accumulation of the surface glycoprotein protein perturbing the cell envelope to block the growth and viability of the cells (Li et al., 2014).

3. Lactic acid bacteria (LAB)

Even though sortases and sortase dependent proteins (SDPs) were investigated extensively related to pathogenesis, the presence of such proteins in probiotic lactic acid bacteria provided a new avenue to look in to the role of this enzyme on probiotic attributes, such as adhesion, mucus barrier function, immune signalling and nutrient uptake. The hypothesis that sortase enzymes may play crucial roles in bacterial physiology, (as in the case of PrtP in *L. lactis ssp*. *cremoris* MG1363) as well as mediating bacterial-host interactions has accelerated the study of this enzyme in different species of LAB. Call and Klaenhammer (2013) reviewed in detail the reports on such proteins in selected species of health-promoting LAB. Recently the application of the LPXTG motif as a bio-therapeutic in Lactic acid bacteria has been reported. To date, LAB and sortase-mediated cell wall anchoring have been explored in the display of potential vaccine antigens including the tetanus toxin fragment C (Norton et al., 1996), human papillomavirus (HPV) type 16 E7 antigen (Bermúdez-Humarán et al., 2003; Cortes-Perez et al., 2005, 2003), the oncofetal antigen (Fredriksen et al., 2010), and *Salmonella enterica* serovar *typhimurium* flagellin (Kajikawa et al., 2011). All these studies indicate that vaccine delivery in LAB using LPXTG or LPXTG-like cell wall anchors has great potential.

Some of the major studies summarizing the sortases in different LAB species and their roles were briefly covered in this review;

Lactobacillus rhamnosus strain GG, a well-known probiotic bacterium, also displays on its cell surface mucus binding pilus structures, along with other LPXTG surface proteins, which are processed by sortases upon specific recognition of a highly conserved LPXTG motif. Demonstration of the expression and presence of mucus binding pilin-like structures on the surface of *L. rhamnosus* GG has been determined (Kankainen et al., 2009) and interestingly, *L. rhamnosus* GG shows exemplary ability to adhere to Caco-2 cells as compared to other probiotic strains (Jacobsen et al., 1999). The genome sequence of *L. rhamnosus* GG revealed two potential clusters of pilus-encoding genes in tandem with a *srtC* gene. The first cluster identified contained genes for *spaA* (LGG_00442), *spaB* (LGG_00443), and *spaC* (LGG_00444) clustered with *srtC*1(LGG_00441), while the second cluster contained genes for *spaD* (LGG_02370), *spaE* (LGG_02371), and *spaF* (LGG_02372) clustered with *srtC2* (LGG_02369; Bioinformatic analysis of all predicted LPXTG proteins encoded by the *L. rhamnosus* GG genome revealed remarkable conservation of glycine residues juxtaposed to the canonical LPXTG motif. Douillard et al (2014) investigated and defined the role of the triple glycine (TG) motif in determining sortase specificity during pilus assembly and anchoring. Mutagenesis of the TG motif resulted in a lack of an alteration of the *L. rhamnosus* GG pilus structures, indicating that the TG motif is critical in pilus assembly and that they govern the pilin-specific and housekeeping sortase specificity. Chaurasia et al (2016) provided new insights about pilus formation in gut-adapted *L. rhamnosus* GG from the crystal structure of the SpaA backbone pilin subunit. According to the paper, SpaA consists of two tandem CnaB-type domains, each with an isopeptide bond and E-box motif. Von Ossowski (2017) in his review on novel molecular insights about Lactobacillar sortase-dependent piliation described three types of lactobacillar piliation (i.e., SpaCBA, SpaFED and LrpCBA), each has been described as having the basic characteristics common to all sortase-dependent pili, but as well, certain unique properties and associated actions that are inherent to them. The authors investigated two contrasting gut-adapted species from the *Lactobacillus* genus, allochthonous *L. rhamnosus*, and autochthonous *Lactobacillus ruminis*.

Dieye et al (2010) identified and studied a class A sortase in *Lactococcus lactis* IL1403 and showed that it is responsible for the cell wall anchoring of at least five LPXTG-containing proteins. We, therefore, propose that SrtA is the housekeeping sortase in *L. lactis*. Surface proteins are important factors in the interaction of probiotic and pathogenic bacteria with their environment or host. The sortase mutant and one sortase-dependent protein (mucus-binding homolog) mutant showed a significant reduction in adherence to human epithelial cell lines in the case of *Lactobacillus salivarius* UCC118. Van Pijkeren et al (2006) identified 10 sortase-dependent surface proteins in *L. salivarius* UCC118, by the comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *L. salivarius*.

As it has been emphasized, sortase, an enzyme that covalently couples a subset of extracellular proteins containing an LPXTG motif to the cell surface, is of particular interest in characterizing bacterial adherence and communication with the mucosal immune system. *Lactobacillus casei* BL23 harbors four sortase genes, two belonging to class A (*srtA1* and *srtA2*) and two belonging to class C (*srtC1* and *srtC2*). Class C sortases were clustered with genes encoding their putative substrates that were homologous to the SpaEFG and SpaCBA proteins that encode mucus adhesive pili in *L. rhamnosus* GG. Twenty-three genes encoding putative sortase substrates were identified in the *L. casei* BL23 genome with unknown (35%), enzymatic (30%), or adhesion-related (35%) functions (Muñoz-Provencio et al., 2012). In summary, in *L. casei* BL23, around 20 proteins are likely anchored to the cell surface by sortases. Although the specific function of most of them is unknown, most of them would account for an adaptation to persist in the gastrointestinal niche. SrtA1 is the housekeeping sortase in this strain, while SrtA2 can compensate for its absence to a certain extent.

A sortase gene, *srtA*, was identified in *Lactobacillus acidophilus* NCFM (LBA1244) and *Lactobacillus gasseri* ATCC 33,323 (LGAS_0825). Additionally, eight and six intact sortase-dependant proteins were predicted in *L. acidophilus* and *L. gasseri*, respectively. Inactivation of sortase did not cause significant alteration in growth or survival in simulated gastrointestinal juices. Meanwhile, both DsrtA mutants showed decreased adhesion to porcine mucin *in vitro*. Murine dendritic cells exposed to the DsrtA mutant of *L. acidophilus* or *L. gasseri* induced lower levels of pro-inflammatory cytokines TNF-α and IL-12, respectively, compared with the parent strains (Call et al., 2015). This study shows that sortase-dependent proteins contribute to gut retention of probiotic microbes in the gastrointestinal tract.

Streptococcus thermophilus (ST) belongs to the LAB group and is recognized as safe since it has obtained the Generally Recognized as Safe (GRAS) status. Kebouchi et al (2016) investigated, *in vitro*, the implication of sortase A (SrtA) and sortase-dependent proteins (SDPs) in the adhesion of ST LMD-9 strain to intestinal epithelial cells (IECs) and resistance to bile salt mixture (BSM; taurocholate, deoxycholate, and cholate). The mutation in genes *srtA* and *mucBP* leads to a significant decrease in LMD-9 adhesion capacity to Caco-2 TC7, HT29-CL.16E (*mucBP* gene mutation) cells. However, no difference was observed using HT29-MTX cells. The study revealed that SDPs could be involved in the LMD-9 adhesion depending on the cell lines indicating the importance of eukaryotic-cell surface components in adherence and also SDPs could contribute to resistance to bile salts probably by maintaining the cell membrane integrity.

The *Lactobacillus* genomes encode a single copy of the sortase (SrtA) and a variable number of LPXTG-motif-containing proteins, ranging from two proteins in *Lactobacillus delbrueckii bulgaricus* ATCC-BAA-365 and ATCC 11,842 and to 27 functional proteins in *Lactobacillus plantarum* WCFS1.

Four *Lactobacillus* proteins belonging to the sortase-dependent protein family have been functionally characterized. Three of the lactobacilli sortase-dependent proteins correspond to the mucus adhesins of *L. reuteri* 1063 Mub (Roos and Jonsson, 2002), *L. plantarum* WCFS1 Msa (Pretzer et al., 2005) and *L. acidophilus* NCFM Mub (Buck et al., 2005). The fourth characterized sortase-dependent protein is LspA of *Lactobacillus salivarius* UCC118 which has been reported to mediate the adhesion of this strain to human epithelial cells and mucus (Claesson et al.,

2006; Van Pijkeren et al., 2006).

Similarly, bifidobacteria represent one of the dominant groups of microorganisms colonizing the human infant intestine (Turroni et al., 2014). Whole-genome transcription profiling of *Bifidobacterium bifidum* PRL2010, a strain isolated from an infant stool, revealed a small number of commonly expressed extracellular proteins, among which were genes that specify sortase-dependent pili modulating bacterium–host interactions. The genome of *B. bifidum* PRL2010 encompasses three different loci encoding predicted sortase-dependent pili, of which only pil2 and pil3 appear to be functional (Foroni et al., 2011). Similarly, the genome of *B. bifidum* S17 was shown to contain a large number of genes that might be involved in host colonization including Tad and sortase-dependent pili, lipoproteins, and several other genes encoding for surface proteins with domains known to mediate interaction with host structures (Westermann et al., 2012).

Adhesion of bifidobacterial cells to the mucosa of the large intestine is considered a hallmark for the persistence and colonization of these bacteria in the human gut. In this context, Milani et al (2017) analyzed the genetic diversity of the predicted arsenal of sortase-dependent pili of known and sequenced members of the *Bifidobacterium* genus and constructed a bifidobacterial sortase-dependent fimbriome database. Their analyses revealed considerable genetic variability of the sortase-dependent fimbriome among bifidobacterial (sub) species and they concluded that it may be due to horizontal gene transfer events. While it is SrtA reported in most bifidobacteria while searching uniprot we could see there were some genes annotated as ESN35_09070, of *Bifidobacterium gallinarum*, or the one in *Bifidobacterium platyrrhinorum* were classified as class C. However, not many studies were reported on them. Krishnan et al (2016) summarized the latest awareness about pili in probiotics with emphasis on members of lactobacilli and bifidobacteria.

Enterococci are currently leading causes of hospital-acquired infections, such as bloodstream, wound, and catheter-associated urinary tract infections (Murray, 1990). Adhesion to and biofilm formation on damaged tissue and abiotic surfaces, such as central venous and urinary catheters, are critical components of enterococcal pathogenesis that complicate successful treatments. Cell surface proteins have been shown to play significant roles in *E. faecalis* virulence and among these, the sortase-assembled endocarditis and biofilm-associated pilus (Ebp pilus) is important for *in vitro* biofilm formation and virulence in *E. faecalis* (Kline et al., 2009; Singh et al., 2007) and *E. faecium* (Sillanpää et al., 2010) and for infective endocarditis in *E. faecalis* (Nallapareddy et al., 2006).

Genome analysis of *E. faecalis* V583 revealed the presence of two class A sortases (EF_2524 and SrtA [EF_3056]), one class C sortase (SrtC [EF_0194 for *b*iofilm and pilus-associated *s*ortase]) (Dramsi et al., 2005; Paulsen et al., 2003), and 41 surface proteins bearing a cell wall sorting signal motif (Sillanpää et al., 2004). It is believed that atleast some of these surface proteins are microbial surface components recognizing adhesive matrix molecules that play a role in the attachment of *E. faecalis* to extracellular matrix proteins and thus are likely to be important for virulence. Also, two other sortases, named *srt-1* and *srt-2*, were reported in *E. faecalis* strain E99 containing a *bee* (*b*iofilm *e*nhancer in *e*nterococcus) locus; however, their occurrence was rare and they were found only in a few isolates examined from a selection of 40 *E. faecalis* (Telford et al., 2006). The *ebp* operon in *E. faecalis* encodes the Ebp pilus structural subunits EbpA, EbpB, and EbpC and the pilus-associated sortase SrtC. The housekeeping sortase SrtA is encoded elsewhere in the genome (Kemp et al., 2007). The ubiquitous presence of both sortase genes in *E. faecalis* isolates increases the likelihood that sortases in general, and *srtC* in particular, could be a target for disease prevention. *srtC* shown to be necessary for the production of the Ebp pili and important for biofilm formation and endocarditis. Nielsen et al (2013) reported that a *srtA* deletion mutant showed a small (5%) reduction in biofilm formation while a *srtA* - *srtC* double mutant showed a much greater reduction (74%) in comparison to a smaller reduction (44%) with a SrtC mutant. In summary, from this study, it appears likely that SrtC, presumably via SrtC-anchored surface proteins, plays an important role in both *in vitro* biofilm formation and *in vivo* murine kidney infections under the experimental conditions examined.

In general, Class A sortases, which appear to be ubiquitous in many Gram-positive bacteria, anchor a large number and broad range of surface proteins (Marraffini et al., 2006). The sortase C class of enzymes is predicted to anchor a much smaller set of substrates, and the genes coding for these are typically clustered with the substrate genes and involved in pilus biogenesis in addition to surface anchoring (Scott and Zähner, 2006; Telford et al., 2006). Pansegrau and Bagnoli (2015) reviewed the pilus assembly in Gram-positive bacteria. They illustrated the operon structure of selected pilus islands.

4. Concluding remarks

To conclude, sortases are either involved in anchoring proteins to the cell wall (the so-called housekeeping sortases) or in polymerizing pilin proteins (pilin-specific sortases). A significant proportion of the work carried out on sortases so far has focused on SrtA from *S. aureus*, which is of great industrial benefit and representing an important therapeutic target. SrtA of *S. aureus* has been significantly exploited for a variety of industrial purposes which includes, enzyme immobilization, cell surface labeling, antibody-drug conjugates, dimerization and cyclization of proteins. It is also likely that sortases may soon become drug targets for the treatment of a wide range of conditions but, some of which are potentially yet to be conceived. Remarkable progress has been made in the last decade in obtaining the information about the classes of sortases A, B, C, and D from Gram-positive bacteria. As a result, novel information about the class E sortase of Gram-positive bacteria has been determined, which includes its function, the substrate motif, and structurally three-dimensional folds. The class F sortase was recently reported in *P. acnes* with a substrate-specificity similar to class A sortase. Sortases are not essential for bacterial cell survival but do significantly impact the binding to host tissues, signaling to the host, or escaping the host immune response, and thus they are equally crucial for nonpathogenic gastrointestinal bacteria as well. Besides, sortase expression signals in lactobacilli have been exploited as a means to develop oral vaccines targeted to the gastro-intestinal tract. The future is not certainly to look in to the primary sequence homology and classification but rather to focus more on structure–function studies that determine the substrate specificity and also the interaction of sortases with other membrane bound enzymes to study better cell wall assembly, such as pili biogenesis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank Council of Scientific and Industrial Research (CSIR) New Delhi for the fellowship (JRF and SRF) to AS and acknowledges the supported by the ‘Innovative Young Biotechnologist Award’ by the Department of Biotechnology, Government of India (BT/11/IYBA/2018/09) to HB. The figures created with Biorender.com.

References

Aas, J.A., Paster, B.J., Stokes, L.N., Olsen, I., Dewhirst, F.E., 2005. Defining the Normal Bacterial Flora of the Oral Cavity. J. Clin. Microbiol. 43, 5721–5732. https://doi.org/10.1128/JCM.43.11.5721.

Adak, G.K., Long, S.M., O'Brien, S.J., 2002. Trends in indigenous foodborne disease and deaths, England and Wales: 1992 to 2000. Gut 51, 832–841. https://doi.org/10.1136/gut.51.6.832.

Aucher, W., Davison, S., Fouet, A., 2011. Characterization of the Sortase Repertoire in *Bacillus anthracis*. PLoS One. 6, e27411 https://doi.org/10.1371/journal.pone.0027411.

Barnett, T.C., Scott, J.R., 2002. Differential Recognition of Surface Proteins in *Streptococcus pyogenes* by Two Sortase Gene Homologs. J. Bacteriol. 184, 2181–2191. https://doi.org/10.1128/JB.184.8.2181.

Beckmann, C., Waggoner, J.D., Harris, T.O., Tamura, G.S., Rubens, C.E., 2002. Identification of Novel Adhesins from Group B Streptococci by Use of Phage Display Reveals that C5a Peptidase Mediates Fibronectin Binding. Infect. Immun. 70, 2869–2876. https://doi.org/10.1128/IAI.70.6.2869.

Bermúdez-Humarán, L.G., Cortes-Perez, N.G., Le Loir, Y., Gruss, A., Rodriguez-Padilla, C., Saucedo-Cardenas, O., Langella, P., De Oca-Luna, R.M., 2003. Fusion to a carrier protein and a synthetic propeptide enhances E7 HPV-16 production and secretion in *Lactococcus lactis*. Biotechnol. Prog. 19, 1101–1104. https://doi.org/10.1021/bp0340077.

Bierne, H., Mazmanian, S.K., Trost, M., Pucciarelli, M.G., Liu, G., Dehoux, P., Jänsch, L., Garcia-del Portillo, F., Schneewind, O., Cossart, P., 2002. Inactivation of the *srtA* gene in *Listeria monocytogenes* inhibits anchoring of surface proteins and affects virulence. Mol. Microbiol. 43, 869–881. https://doi.org/10.1046/j.1365-2958.2002.02798.x.

Boekhorst, J., Been, M.W.H.J.D., Kleerebezem, M., Siezen, R.J., 2005. Genome-Wide Detection and Analysis of Cell Wall-Bound Proteins with LPxTG-Like Sorting Motifs. J. Bacteriol. 187, 4928–4934. https://doi.org/10.1128/JB.187.14.4928.

Bradley, A.J., Leach, K.A., Breen, J.E., Green, L.E., Green, M.J., 2007. Survey of the incidence and aetiology of mastitis on dairy farms in England and Wales. Vet. Rec. 160, 253–258. https://doi.org/10.1136/vr.160.8.253.

Buck, L.B., Altermann, E., Svingerud, T., Klaenhammer, T.R., 2005. Functional analysis of adhesion factors and signaling mechanisms in *Lactobacillus acidophilus* NCFM. Appl. Environ. Microbiol. 71, 8344–8351. https://doi.org/10.1128/AEM.71.12.8344.

Budzik, J.M., Marraffini, L.A., Schneewind, O., 2007. Assembly of pili on the surface of *Bacillus cereus* vegetative cells. Mol. Microbiol. 66, 495–510. https://doi.org/10.1111/j.1365-2958.2007.05939.x.

Budzik, J.M., Oh, S.-Y., Schneewind, O., 2009. Sortase D forms the covalent bond that links BcpB to the tip of Bacillus cereus pili. J. Biol. Chem. 284, 12989–97. 10.1074/jbc.M900927200.

Cabanes, D., Dehoux, P., Dussurget, O., Frangeul, L., Cossart, P., 2002. Surface proteins and the pathogenic potential of *Listeria monocytogenes*. Trends Microbiol. 10, 238–245. https://doi.org/10.1016/S0966-842X(02)02342-9.

Call, E.K., Goh, Y.J., Selle, K., Klaenhammer, T.R., O'Flaherty, S., 2015. Sortase-deficient lactobacilli: Effect on immunomodulation and gut retention. Microbiol. 161, 311–321. https://doi.org/10.1099/mic.0.000007.

Call, E.K., Klaenhammer, T.R., 2013. Relevance and application of sortase and sortase-dependent proteins in lactic acid bacteria. Front. Microbiol. 4, 1–10. https://doi.org/10.3389/fmicb.2013.00073.

Chambers, C.J., Roberts, A.K., Shone, C.C., Acharya, K.R., 2015. Structure and function of a *Clostridium difficile* sortase enzyme. Sci. Rep. 5, 1–11. https://doi.org/10.1038/srep09449.

Chaurasia, P., Pratap, S., Von Ossowski, I., Palva, A., Krishnan, V., 2016. New insights about pilus formation in gut-adapted *Lactobacillus rhamnosus* GG from the crystal structure of the SpaA backbone-pilin subunit. Sci. Rep. 6, 1–17. https://doi.org/10.1038/srep28664.

Cheng, A.G., Kim, H.K., Burts, M.L., Krausz, T., Schneewind, O., Missiakas, D.M., 2009. Genetic requirements for *Staphylococcus aureus* abscess formation and persistence in host tissues. FASEB J. 23, 3393–3404. https://doi.org/10.1096/fj.09-135467.

Choo, J.M., Cheung, J.K., Wisniewski, J.A., Steer, D.L., Bulach, D.M., Hiscox, T.J., Chakravorty, A., Smith, A.I., Gell, D.A., Rood, J.I., Awad, M.M., 2016. The NEAT Domain-Containing Proteins of *Clostridium perfringens* Bind Heme. PLoS One. 11, e0162981 https://doi.org/10.1371/journal.pone.0162981.

Claesson, M.J., Li, Y., Leahy, S., Canchaya, C., Van Pijkeren, J.P., Cerdeño-Tárraga, A.M., Parkhill, J., Flynn, S., O'Sullivan, G.C., Kevin Collins, J., Higgins, D., Shanahan, F., Fitzgerald, G.F., Van Sinderen, D., O'Toole, P.W., 2006. Multireplicon genome architecture of *Lactobacillus salivarius*. Proc. Natl. Acad. Sci. U. S. A. 103, 6718–6723. https://doi.org/10.1073/pnas.0511060103.

Clancy, K.W., Melvin, J.A., McCafferty, D.G., 2010. Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. Biopolymers 94, 385–396. https://doi.org/10.1002/bip.21472.

Comfort, D., Clubb, R.T., 2004. A Comparative Genome Analysis Identifies Distinct Sorting Pathways in Gram-Positive Bacteria. Infect. Immun. 72, 2710–2722. https://doi.org/10.1128/IAI.72.5.2710.

Cortes-Perez, N.G., Azevedo, V., Alcocer-González, J.M., Rodriguez-Padilla, C., Tamez-Guerra, R.S., Corthier, G., Gruss, A., Langella, P., Bermúdez-Humarán, L.G., 2005. Cell-surface display of E7 antigen from human papillomavirus type-16 in *Lactococcus lactis* and in *Lactobacillus plantarum* using a new cell-wall anchor from lactobacilli. J. Drug Target. 13, 89–98. https://doi.org/10.1080/10611860400024219.

Cortes-Perez, N.G., Bermúdez-Humarán, L.G., Le Loir, Y., Rodriguez-Padilla, C., Gruss, A., Saucedo-Cárdenas, O., Langella, P., Montes-De-Oca-Luna, R., 2003. Mice immunization with live lactococci displaying a surface anchored HPV-16 E7 oncoprotein. FEMS Microbiol. Lett. 229, 37–42. https://doi.org/10.1016/S0378-1097(03)00778-X.

Corver, J., Cordo', V., van Leeuwen, H.C., Klychnikov, O.I., Hensbergen, P.J., 2017. Covalent attachment and Pro-Pro endopeptidase (PPEP-1)-mediated release of Clostridium difficile cell surface proteins involved in adhesion. Mol. Microbiol. 105, 663–673. 10.1111/mmi.13736.

Cossart, P., 2007. Listeria monocytogenes Surface Proteins : from Genome Predictions to Function. Microbiol. Mol. Biol. Rev. 71, 377–397. https://doi.org/10.1128/MMBR.00039-06.

Das, S., Pawale, V.S., Dadireddy, V., Singh, A.K., Ramakumar, S., Roy, R.P., 2017. Structure and specificity of a new class of Ca^{2+}-independent housekeeping sortase from *Streptomyces avermitilis* provide insights into its non-canonical substrate preference. J. Biol. Chem. 292, 7244–7257. https://doi.org/10.1074/jbc.M117.782037.

Davies, J.R., Svensa, G., Herzberg, M.C., 2009. Identification of novel LPXTG-linked surface proteins from *Streptococcus gordonii*. Microbiology. 155, 1977–1988. https://doi.org/10.1099/mic.0.027854-0.

Davison, S., Couture-Tosi, E., Candela, T., Mock, M., Fouet, A., 2005. Identification of the Bacillus anthracis (gamma) phage receptor. J. Bacteriol. 187, 6742–9. 10.1128/JB.187.19.6742-6749.2005.

Dekker, F.J., Boersma, Y.L., 2018. Identification of potential antivirulence agents by substitution-oriented screening for inhibitors of Streptococcus pyogenes sortase A. Eur. J. Med. Chem. 10.1016/j.ejmech.2018.10.027.

Dieye, Y., Oxaran, V., Ledue-Clier, F., Alkhalaf, W., Buist, G., Juillard, V., Lee, C.W., Piard, J.C., 2010. Functionality of sortase a in *Lactococcus lactis*. Appl. Environ. Microbiol. 76, 7332–7337. https://doi.org/10.1128/AEM.00928-10.

Donahue, E.H., Dawson, L.F., Valiente, E., Firth-Clark, S., Major, M.R., Littler, E., Perrior, T.R., Wren, B.W., 2014. *Clostridium difficile* has a single sortase, SrtB, that can be inhibited by small-molecule inhibitors. BMC Microbiol. 14, 1–14. https://doi.org/10.1186/s12866-014-0219-1.

Douillard, F.P., Rasinkangas, P., Von Ossowski, I., Reunanen, J., Palva, A., De Vos, W.M., 2014. Functional identification of conserved residues involved in *Lactobacillus rhamnosus* strain GG sortase specificity and pilus biogenesis. J. Biol. Chem. 289, 15764–15775. https://doi.org/10.1074/jbc.M113.542332.

Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prévost, M., Kojadinovic, M., Lalioui, L., Poyart, C., Trieu-cuot, P., 2006. Assembly and role of pili in group B streptococci. Mol. Microbiol. 60, 1401–1413. https://doi.org/10.1111/j.1365-2958.2006.05190.x.

Dramsi, S., Trieu-Cuot, P., Bierne, H., 2005. Sorting sortases: A nomenclature proposal for the various sortases of Gram-positive bacteria. Res. Microbiol. 156, 289–297. https://doi.org/10.1016/j.resmic.2004.10.011.

Duarte, C.M., Freitas, P.P., Bexiga, R., 2015. Technological advances in bovine mastitis diagnosis: an overview. J. Vet. Diagnostic Investig. 27, 665–672. https://doi.org/10.1177/1040638715603087.

Duc Nguyen, H., Thi, T., Phan, P., Schumann, W., 2011. Analysis and application of *Bacillus subtilis* sortases to anchor recombinant proteins on the cell wall. AMB Express. 1, 22. https://doi.org/10.1186/2191-0855-1-22.

Duong, A., Capstick, D.S., Berardo, C. Di, Findlay, K.C., Hesketh, A., Hong, H., Elliot, M. A., 2012. Aerial development in *Streptomyces coelicolor* requires sortase activity. Mol. Microbiol. 83, 992–1005. https://doi.org/10.1111/j.1365-2958.2012.07983.x.

Dussurget, O., Pizarro-cerda, J., Cossart, P., 2004. Molecular determinants of *Listeria monocytogenes* virulence. Annu. Rev. Microbiol. 58, 587–610. https://doi.org/10.1146/annurev.micro.57.030502.090934.

Egan, S.A., Kurian, D., Ward, P.N., Hunt, L., Leigh, J.A., 2010. Identification of Sortase A (SrtA) Substrates in *Streptococcus uberis* : Evidence for an Additional Hexapeptide (LPXXXD) Sorting Motif. J. Proteome. Res. 9, 1088–1095. https://doi.org/10.1021/pr901025w.

Eigh, J.A.L., Gan, S.A.E., Ard, P.N.W., Ield, T.R.F., Offey, T.J.C., 2010. Original article Sortase anchored proteins of *Streptococcus uberis* play major roles in the pathogenesis of bovine mastitis in dairy cattle. Vet. Res. 41, 63. https://doi.org/10.1051/vetres/2010036.

Fasehee, H., Westers, H., Bolhuis, A., Antelmann, H., Hecker, M., Quax, W.J., Mirlohi, A. F., Van Dijl, J.M., Ahmadian, G., 2011. Functional analysis of the sortase YhcS in *Bacillus subtilis*. Proteomics 11, 3905–3913. https://doi.org/10.1002/pmic.201100174.

Fittipaldi, N., Segura, M., Grenier, D., Gottschalk, M., 2012. Virulence factors involved in the pathogenesis of the infection caused by the swine pathogen and zoonotic agent *Streptococcus suis*. Future Microbiol. 7, 259–279. https://doi.org/10.2217/fmb.11.149.

Foroni, E., Serafini, F., Amidani, D., Turroni, F., He, F., Bottacini, F., O'Connell Motherway, M., Viappiani, A., Zhang, Z., Rivetti, C., Van Sinderen, D., Ventura, M., 2011. Genetic analysis and morphological identification of pilus-like structures in members of the genus Bifidobacterium. Microb. Cell Fact. 10, 1–13. https://doi.org/10.1186/1475-2859-10-S1-S16.

Fouet, A., 2009. The surface of *Bacillus anthracis*. Mol. Aspects Med. 30, 374–385. https://doi.org/10.1016/J.MAM.2009.07.001.

Frankel, B.A., Kruger, R.G., Robinson, D.E., Kelleher, N.L., McCafferty, D.G., 2005. *Staphylococcus aureus* sortase transpeptidase SrtA: Insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. Biochemistry 44, 11188–11200. https://doi.org/10.1021/bi050141j.

Frankel, B.B. a, Tong, Y., Bentley, M.M.L., Fitzgerald, M.C., McCafferty, D.G., 2007. Mutational analysis of active site residues in the Staphylococcus aureus transpeptidase SrtA. Biochemistry 46, 7269–7278. 10.1021/bi700448e.

Fredriksen, L., Mathiesen, G., Sioud, M., Eijsink, V.G.H., 2010. Cell wall anchoring of the 37-kilodalton oncofetal antigen by *Lactobacillus plantarum* for mucosal cancer vaccine delivery. Appl. Environ. Microbiol. 76, 7359–7362. https://doi.org/10.1128/AEM.01031-10.

Garandeau, C., Pucciarelli, M.G., Sabet, C., Newton, S., Portillo, F.G., Cossart, P., Charbit, A., 2004. Sortase B, a New Class of Sortase in *Listeria monocytogenes*. J. Bacteriol. 186, 1972–1982. https://doi.org/10.1128/JB.186.7.1972.

Gaspar, A.H., Marraffini, L.A., Glass, E.M., DeBord, K.L., Ton-That, H., Schneewind, O., 2005. *Bacillus anthracis* Sortase A (SrtA) Anchors LPXTG Motif-Containing Surface

Proteins to the Cell Wall Envelope. J. Bacteriol. 187, 4646–4655. https://doi.org/10.1128/JB.187.13.4646-4655.2005.
Gaspar, A.H., Ton-That, H., 2006. Assembly of Distinct Pilus Structures on the Surface of *Corynebacterium diphtheriae*. J. Bacteriol. 188, 1526–1533. https://doi.org/10.1128/JB.188.4.1526.
Girolamo, S. Di, Puorger, C., Castiglione, M., Vogel, M., Gébleux, R., Briendl, M., Hell, T., Beerli, R.R., Grawunder, U., Lipps, G., 2019. Characterization of the housekeeping sortase from the human pathogen *Propionibacterium acnes*: First investigation of a class F sortase. Biochem. J. 476, 665–682. https://doi.org/10.1042/BCJ20180885.
Glaser, P., Rusniok, C., Buchrieser, C., Chevalier, F., Frangeul, L., Msadek, T., Zouine, M., Couvé, E., Lalioui, L., Poyart, C., Trieu-cuot, P., Kunst, F., 2002. Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. Mol. Microbiol. 45, 1499–1513. https://doi.org/10.1046/j.1365-2958.2002.03126.x.
Hadfield, T.L., McEvoy, P., Polotsky, Y., Tzinserling, V.A., Yakovlev, A.A., 2000. The pathology of diphtheria. J. Infect. Dis. 181, 3–7. https://doi.org/10.1086/315551.
Hamada, S., Slade, H.D., 1980. Biology, immunology, and cariogenicity of *Streptococcus mutans*. Microbiol. Rev. 44, 331–384. https://doi.org/10.1128/mr.44.2.331-384.1980.
Hendrickx, A.P.A., Budzik, J.M., Oh, S.Y., Schneewind, O., 2011. Architects at the bacterial surface-sortases and the assembly of pili with isopeptide bonds. Nat. Rev. Microbiol. 9, 166–176. https://doi.org/10.1038/nrmicro2520.
Hensbergen, P.J., Klychnikov, O.I., Bakker, D., Dragan, I., Kelly, M.L., Minton, N.P., Corver, J., Kuijper, E.J., Drijfhout, J.W., Van Leeuwen, H.C., 2015. *Clostridium difficile* secreted Pro-Pro endopeptidase PPEP-1 (ZMP1/CD2830) modulates adhesion through cleavage of the collagen binding protein CD2831. FEBS Lett. 589, 3952–3958. https://doi.org/10.1016/j.febslet.2015.10.027.
Jacobsen, C.N., Nielsen, V.R., Hayford, A.E., Møller, P.L., Michaelsen, K.F., Pærregaard, A., Sandström, B., Tvede, M., Jakobsen, M., 1999. Screening of probiotic activities of forty-seven strains of Lactobacillus spp. by in vitro techniques and evaluation of the colonization ability of five selected strains in humans. Appl. Environ. Microbiol. 65, 4949–4956. https://doi.org/10.1128/aem.65.11.4949-4956.1999.
Kajikawa, A., Nordone, S.K., Zhang, L., Stoeker, L.L., LaVoy, A.S., Klaenhammer, T.R., Dean, G.A., 2011. Dissimilar properties of two recombinant *Lactobacillus acidophilus* strains displaying Salmonella FliC with different anchoring motifs. Appl. Environ. Microbiol. 77, 6587–6596. https://doi.org/10.1128/AEM.05153-11.
Kankainen, M., Paulin, L., Tynkkynen, S., Von Ossowski, I., Reunanen, J., Partanen, P., Satokari, R., Vesterlund, S., Hendrickx, A.P.A., Lebeer, S., De Keersmaecker, S.C.J., Vanderleyden, J., Hämäläinen, T., Laukkanen, S., Salovuori, N., Ritari, J., Alatalo, E., Korpela, R., Mattila-Sandholm, T., Lassig, A., Hatakka, K., Kinnunen, K.T., Karjalainen, H., Saxelin, M., Laakso, K., Surakka, A., Palva, A., Salusjärvi, T., Auvinen, P., De Vos, W.M., 2009. Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human-mucus binding protein. Proc. Natl. Acad. Sci. U. S. A. 106, 17193–17198. https://doi.org/10.1073/pnas.0908876106.
Kebouchi, M., Galia, W., Genay, M., Soligot, C., Lecomte, X., Awussi, A.A., Perrin, C., Roux, E., Dary-Mourot, A., Le Roux, Y., 2016. Implication of sortase-dependent proteins of *Streptococcus thermophilus* in adhesion to human intestinal epithelial cell lines and bile salt tolerance. Appl. Microbiol. Biotechnol. 100, 3667–3679. https://doi.org/10.1007/s00253-016-7322-1.
Kemp, K.D., Singh, K.V., Nallapareddy, S.R., Murray, B.E., 2007. Relative contributions of *Enterococcus faecalis* OG1RF sortase-encoding genes, *srtA* and *bps* (SrtC), to biofilm formation and a murine model of urinary tract infection. Infect. Immun. 75, 5399–5404. https://doi.org/10.1128/IAI.00663-07.
Kharat, A.S., Tomasz, A., 2003. Inactivation of the *srtA* gene affects localization of surface proteins and decreases adhesion of *Streptococcus pneumoniae* to human pharyngeal cells in vitro. Infect. Immun. 71, 2758–2765. https://doi.org/10.1128/IAI.71.5.2758-2765.2003.
Khare, B., Krishnan, V., Rajashankar, K.R., Xin, M., Narayana, S.V., 2011. Structural Differences between the *Streptococcus agalactiae* Housekeeping and Pilus-Specific Sortases : SrtA and SrtC1. PLoS One. 6, e22995 https://doi.org/10.1371/journal.pone.0022995.
Kline, K.A., Kau, A.L., Chen, S.L., Lim, A., Pinkner, J.S., Rosch, J., Nallapareddy, S.R., Murray, B.E., Henriques-Normark, B., Beatty, W., Caparon, M.G., Hultgren, S.J., 2009. Mechanism for sortase localization and the role of sortase localization in efficient pilus assembly in *Enterococcus faecalis*. J. Bacteriol. 191, 3237–3247. https://doi.org/10.1128/JB.01837-08.
Kolenbrander, P.E., Ganeshkumar, N., Cassels, F.J., Hughes, C.V., 1993. Coaggregation : specific adherence among human oral plaque bacteria. FASEB J. 7, 406–413.
Krishnan, V., Chaurasia, P., Kant, A., 2016. Pili in Probiotic Bacteria. Probiotics Prebiotics Hum. Nutr. Heal. 10.5772/6308787.
Kuboniwa, M., Tribble, G.D., James, C.E., Kilic, A.O., Tao, L., Herzberg, M.C., Shizukuishi, S., Lamont, R.J., 2006. *Streptococcus gordonii* utilizes several distinct gene functions to recruit *Porphyromonas gingivalis* into a mixed community. Mol. Microbiol. 60, 121–139. https://doi.org/10.1111/j.1365-2958.2006.05099.x.
Lalioui, L., Pellegrini, E., Dramsi, S., Baptista, M., Bourgeois, N., Doucet-populaire, F., Rusniok, C., Zouine, M., Glaser, P., Kunst, F., Poyart, C., Trieu-cuot, P., 2005. The SrtA Sortase of Streptococcus agalactiae Is Required for Cell Wall Anchoring of Proteins Containing the LPXTG Motif, for Adhesion to Epithelial Cells, and for Colonization of the Mouse Intestine. Infect. Immun. 73, 3342–3350. https://doi.org/10.1128/IAI.73.6.3342.
Lee, S.F., Boran, T.L., 2003. Roles of sortase in surface expression of the major protein adhesin P1, saliva-induced aggregation and adherence, and cariogenicity of *Streptococcus mutans*. Infect. Immun. 71, 676–681. https://doi.org/10.1128/IAI.71.2.676-681.2003.
Lemieux, J., Woody, S., Camilli, A., 2008. Roles of the sortases of *Streptococcus pneumoniae* in assembly of the RlrA pilus. J. Bacteriol. 190, 6002–6013. https://doi.org/10.1128/JB.00379-08.
Lepp, D., Zhou, Y., Ojha, S., Mehdizadeh Gohari, I., Carere, J., Yang, C., Prescott, J.F., Gong, J., 2021. *Clostridium perfringens* produces an adhesive pilus required for the pathogenesis of necrotic enteritis in poultry. J. Bacteriol. 203, e005578–20. https://doi.org/10.1128/jb.00578-20.
Li, J.H., Miao, J., Wu, J.L., Chen, S.F., Zhang, Q.Q., 2014. Preparation and characterization of active gelatin-based films incorporated with natural antioxidants. Food Hydrocoll. 37, 166–173. https://doi.org/10.1016/j.foodhyd.2013.10.015.
Liew, P.X., Wang, C.L.C., Wong, S.-L., 2012. Functional Characterization and Localization of a *Bacillus subtilis* Sortase and Its Substrate and Use of This Sortase System To Covalently Anchor a Heterologous Protein to the *B. subtilis* Cell Wall for Surface Display. J. Bacteriol. 194, 161–175. https://doi.org/10.1128/JB.05711-11.
Lodes, M.J., Secrist, H., Benson, D.R., Jen, S., Shanebeck, K.D., Guderian, J., Maisonneuve, J.F., Bhatia, A., Persing, D., Patrick, S., Skeiky, Y.A.W., 2006. Variable expression of immunoreactive surface proteins of *Propionibacterium acnes*. Microbiology 152, 3667–3681. https://doi.org/10.1099/mic.0.29219-0.
Lu, G., Qi, J., Gao, F., Yan, J., Tang, J., Gao, G.F., 2011. A novel "'open-form"' structure of sortaseC from *Streptococcus suis*. Proteins. 79, 2764–2769. https://doi.org/10.1002/prot.23093.
Mandlik, A., Swierczynski, A., Das, A., 2010. Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. Trends. Microbiol. 16, 33–40. https://doi.org/10.1016/j.tim.2007.10.010.
Mandlik, A., Swierczynski, A., Das, A., Ton-That, H., 2007. *Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells. Mol. Microbiol. 64, 111–124. https://doi.org/10.1111/j.1365-2958.2007.05630.x.
Manzano, C., Izoré, T., Job, V., Di Guilmi, A.M., Dessen, A., 2009. Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of Gram-positive pathogens. Biochemistry 48, 10549–10557. https://doi.org/10.1021/bi901261y.
Maresso, A.W., Chapa, T.J., Schneewind, O., 2006. Surface protein IsdC and sortase B are required for heme-iron scavenging of *Bacillus anthracis*. J. Bacteriol. 188, 8145–8152. https://doi.org/10.1128/JB.01011-06.
Mariscotti, J.F., Quereda, J.J., Pucciarelli, M.G., 2012. Contribution of sortase A to the regulation of *Listeria monocytogenes* LPXTG surface proteins. Int. Microbiol. 15, 43–51. https://doi.org/10.2436/20.1501.01.157.
Marraffini, L.A., DeDent, A.C., Schneewind, O., 2006. Sortases and the Art of Anchoring Proteins to the Envelopes of Gram-Positive Bacteria. Microbiol. Mol. Biol. Rev. 70, 192–221. https://doi.org/10.1128/mmbr.70.1.192-221.2006.
Marraffini, L.A., Schneewind, O., 2007. Sortase C-mediated anchoring of BasI to the cell wall envelope of *Bacillus anthracis*. J. Bacteriol. 189, 6425–6436. https://doi.org/10.1128/JB.00702-07.
Marraffini, L.A., Schneewind, O., 2006. Targeting proteins to the cell wall of sporulating *Bacillus anthracis*. Mol. Microbiol. 62, 1402–1417. https://doi.org/10.1111/j.1365-2958.2006.05469.x.
Mazmanian, S.K., Liu, G., Jensen, E.R., Lenoy, E., Schneewind, O., 2000. *Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. Proc. Natl. Acad. Sci. 97, 5510–5515. https://doi.org/10.1073/pnas.080520697.
Mazmanian, S.K., Liu, G., Ton-That, H., Schneewind, O., 1999. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. Science. 285, 760–763. https://doi.org/10.1126/science.285.5428.760.
Mazmanian, S.K., Ton-That, H., Su, K., Schneewind, O., 2002. An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. Proc. Natl. Acad. Sci. 99, 2293–2298. https://doi.org/10.1073/pnas.032523999.
Milani, C., Mangifesta, M., Mancabelli, L., Lugli, G.A., Mancino, W., Viappiani, A., Faccini, A., 2017. The Sortase-Dependent Fimbriome of the Genus Bifidobacterium : Extracellular Structures with Potential To Modulate Microbe-Host Dialogue. 83, e01295-17. https://dx.doi.org/10.1128%2FAEM.01295-17.
Mishra, A., Das, A., Cisar, J.O., Ton-That, H., 2007. Sortase-catalyzed assembly of distinct heteromeric fimbriae in *Actinomyces naeslundii*. J. Bacteriol. 189, 3156–3165. https://doi.org/10.1128/JB.01952-06.
Mock, M., Fouet, A., 2001. Anthrax. Annu. Rev. Microbiol. 55, 647–671. https://doi.org/10.1146/annurev.micro.55.1.647.
Morita, C., Sumioka, R., Nakata, M., Okahashi, N., Wada, S., 2014. Cell Wall-Anchored Nuclease of *Streptococcus sanguinis* Contributes to Escape from Neutrophil Extracellular Trap-Mediated Bacteriocidal Activity. PLoS One. 9, e103125 https://doi.org/10.1371/journal.pone.0103125.
Muñoz-Provencio, D., Rodríguez-Díaz, J., Collado, M.C., Langella, P., Bermúdez-Humarán, L.G., Monedero, V., 2012. Functional analysis of the *Lactobacillus casei* BL23 sortases. Appl. Environ. Microbiol. 78, 8684–8693. https://doi.org/10.1128/AEM.02287-12.
Murai, C., Igarashi, T., Inoue, M., Sasa, R., 2005. *Streptococcus mutans* sortase catalyzes cell wall anchoring of WapA and FruA. Pediatr. Dent. J. 15, 127–133. https://doi.org/10.1016/S0917-2394(05)70041-0.
Murray, B., 1990. The Life and Times of Enterococcus. Clin. Microbiol. Rev. 3, 46–65. https://doi.org/10.1128/CMR.3.1.46.
Nallapareddy, S.R., Singh, K.V., Sillanpää, J., Garsin, D.A., Höök, M., Erlandsen, S.L., Murray, B.E., 2006. Endocarditis and biofilm-associated pili of *Enterococcus faecalis*. J. Clin. Invest. 116, 2799–2807. https://doi.org/10.1172/JCI29021.
Naziga, E.B., Wereszczynski, J., 2017. Molecular Mechanisms of the Binding and Specificity of *Streptococcus pneumoniae* Sortase C Enzymes for Pilin Subunits. Sci. Rep. 7, 1–14. https://doi.org/10.1038/s41598-017-13135-3.
Nielsen, H.V., Flores-Mireles, A.L., Kau, A.L., Kline, K.A., Pinkner, J.S., Neiers, F., Normark, S., Henriques-Normark, B., Caparon, M.G., Hultgren, S.J., 2013. Pilin and sortase residues critical for endocarditis- and biofilm-associated pilus biogenesis in

Enterococcus faecalis. J. Bacteriol. 195, 4484–4495. https://doi.org/10.1128/JB.00451-13.
Nobbs, A.H., Vajna, R.M., Johnson, J.R., Zhang, Y., Erlandsen, S.L., Oli, M.W., Kreth, J., Brady, L.J., Herzberg, M.C., 2007. Consequences of a sortase A mutation in Streptococcus gordonii 4088–4097. 10.1099/mic.0.2007/007252-0.
Norton, P.M., Brown, H.W.G., Wells, J.M., Macpherson, A.M., Wilson, P.W., Le Page, R. W.F., 1996. Factors affecting the immunogenicity of tetanus toxin fragment C expressed in *Lactococcus lactis*. FEMS Immunol. Med. Microbiol. 14, 167–177. https://doi.org/10.1016/0928-8244(96)00028-4.
Novick, R.P., 2000. Sortase: The surface protein anchoring transpeptidase and the LPXTG motif. Trends Microbiol. 8, 148–151. https://doi.org/10.1016/S0966-842X(00)01741-8.
Osaki, M., Takamatsu, D., Shimoji, Y., Sekizaki, T., 2002. Characterization of *Streptococcus suis* genes encoding proteins homologous to sortase of gram-positive bacteria. J. Bacteriol. 184, 971–982. https://doi.org/10.1128/jb.184.4.971-982.2002.
Pansegrau, W., Bagnoli, F., 2015. Pilus Assembly in Gram-Positive Bacteria. Curr. Top Microbiol. Immunol. 404, 203–233. https://doi.org/10.1007/82_2015_5016.
Paterson, G.K., Mitchell, T.J., 2006. The role of *Streptococcus pneumoniae* sortase A in colonisation and pathogenesis. Microbes Infect. 8, 145–153. https://doi.org/10.1016/j.micinf.2005.06.009.
Paterson, G.K., Mitchell, T.J., 2004. The biology of Gram-positive sortase enzymes. Trends Microbiol. 12, 89–95. https://doi.org/10.1016/j.tim.2003.12.007.
Paulsen, I.T., Banerjei, L., Hyers, G.S.A., Nelson, K.E., Seshadri, R., Read, T.D., Fouts, D. E., Eisen, J.A., Gill, S.R., Heidelberg, J.F., Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, L., Beanan, M., Daugherty, S., DeBoy, R.T., Durkin, S., Kolonay, J., Madupu, R., Nelson, W., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J., Khouri, H., Utterback, T., Radune, D., Ketchum, K.A., Dougherty, B.A., Fraser, C.M., 2003. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. Science (80-.). 299, 2071–2074. https://doi.org/10.1126/science.1080613.
Peltier, J., Shaw, H.A., Couchman, E.C., Dawson, L.F., Yu, L., Choudhary, J.S., Kaever, V., Wren, B.W., Fairweather, N.F., 2015. Cyclic diGMP Regulates Production of Sortase Substrates of *Clostridium difficile* and Their Surface Exposure through ZmpI Protease-mediated Cleavage. J. Biol. Chem. 290, 24453–24469. https://doi.org/10.1074/jbc.M115.665091.
Peltier, J., Shaw, H.A., Wren, B.W., Fairweather, N.F., 2017. Disparate subcellular location of putative sortase substrates in *Clostridium difficile*. Sci. Rep. 7, 1–10. https://doi.org/10.1038/s41598-017-08322-1.
Perry, A.M., Ton-That, H., Mazmanian, S.K., Schneewind, O., 2002. Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring. J. Biol. Chem. 277, 16241–16248. https://doi.org/10.1074/jbc.M109194200.
Persson, K., 2011. Structure of the sortase AcSrtC-1 from *Actinomyces oris*. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 212–217. https://doi.org/10.1107/S0907444911004215.
Posfay-barbe, K.M., Wald, E.R., 2009. Listeriosis. Semin. Fetal Neonatal Med. 14, 228–233. https://doi.org/10.1016/j.siny.2009.01.006.
Pretzer, G., Snel, J., Molenaar, D., Wiersma, A., Bron, P.A., Lambert, J., De Vos, W.M., Van Der Meer, R., Smits, M.A., Kleerebezem, M., 2005. Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum*. J. Bacteriol. 187, 6128–6136. https://doi.org/10.1128/JB.187.17.6128-6136.2005.
Raz, A., Tanasescu, A., Zhao, A.M., Serrano, A., Alston, T., 2015. *Streptococcus pyogenes* Sortase Mutants Are Highly Susceptible to Killing by Host Factors Due to Aberrant Envelope Physiology. PLoS One. 10, e0140784 https://doi.org/10.1371/journal.pone.0140784.
Revitt-Mills, S.A., Watts, T.D., Lyras, D., Adams, V., Rood, J.I., 2021. The ever-expanding tcp conjugation locus of pCW3 from *Clostridium perfringens*. Plasmid. 113, 102516 https://doi.org/10.1016/j.plasmid.2020.102516.
Robson, S.A., Jacobitz, A.W., Phillips, M.L., Clubb, R.T., 2012. Solution Structure of the Sortase Required for Efficient Production of Infectious *Bacillus anthracis* Spores. Biochemistry 51, 7953–7963. https://doi.org/10.1021/bi300867t.
Roos, S., Jonsson, H., 2002. A high-molecular-mass cell-surface protein from *Lactobacillus reuteri* 1063 adheres to mucus components. Microbiology 148, 433–442. https://doi.org/10.1099/00221287-148-2-433.
Rupnik, M., Wilcox, M.H., Gerding, D.N., 2009. *Clostridium difficile* infection: New developments in epidemiology and pathogenesis. Nat. Rev. Microbiol. 7, 526–536. https://doi.org/10.1038/nrmicro2164.
Schneewind, O., Model, P., Fischetti, V.A., 1992. Sorting of protein a to the staphylococcal cell wall. Cell. 70, 267–281. https://doi.org/10.1016/0092-8674(92)90101-H.
Schubert, A., Zakikhany, K., Pietrocola, G., Meinke, A., Speziale, P., Eikmanns, B.J., Reinscheid, D.J., 2004. The Fibrinogen Receptor FbsA Promotes Adherence of Streptococcus agalactiae to Human Epithelial Cells. Infect. Immun. 72, 6197–6205. 10.1128/IAI.72.11.6197.
Schubert, A., Zakikhany, K., Schreiner, M., Frank, R., Spellerberg, B., Eikmanns, B.J., Reinscheid, D.J., 2002. A fibrinogen receptor from group B Streptococcus interacts with fibrinogen by repetitive units with novel ligand binding sites. Mol. Microbiol. 46, 557–569. doi.org/10.1046/j.1365-2958.2002.03177.x.
Scott, J.R., Zähner, D., 2006. Pili with strong attachments: Gram-positive bacteria do it differently. Mol. Microbiol. 62, 320–330. https://doi.org/10.1111/j.1365-2958.2006.05279.x.
Sendi, P., Johansson, L., 2008. Invasive Group B Streptococcal Disease in Non-pregnant Adults. Infection. 36, 100–111. https://doi.org/10.1007/s15010-007-7251-0.
Shaik, M.M., Maccagni, A., Tourcier, G., Di Guilmi, A.M., Dessen, A., 2014. Structural basis of pilus anchoring by the ancillary pilin RrgC of *Streptococcus pneumoniae*. J. Biol. Chem. 289, 16988–16997. https://doi.org/10.1074/jbc.M114.555854.
Shinagawa, K., 1990. Analytical methods for *Bacillus cereus* and other Bacillus species. Int. J. Food Microbiol. 10, 125–141. https://doi.org/10.1016/0168-1605(90)90061-9.
Sillanpää, J., Nallapareddy, S.R., Singh, K.V., Prakash, V.P., Fothergill, T., Ton-That, H., Murray, B.E., 2010. Characterization of the ebpfm pilus-encoding operon of *Enterococcus faecium* and its role in biofilm formation and virulence in a murine model of urinary tract infection. Virulence. 1, 236–246. https://doi.org/10.4161/viru.1.4.11966.
Sillanpää, J., Xu, Y., Nallapareddy, S.R., Murray, B.E., Höök, M., 2004. A family of putative MSCRAMMs from *Enterococcus faecalis*. Microbiology. 150, 2069–2078. https://doi.org/10.1099/mic.0.27074-0.
Singh, K.V., Nallapareddy, S.R., Murray, B.E., 2007. Importance of the ebp (endocarditis- and biofilm-associated pilus) locus in the pathogenesis of *Enterococcus faecalis* ascending urinary tract infection. J. Infect. Dis. 195, 1671–1677. https://doi.org/10.1086/517524.
Spirig, T., Weiner, E.M., Clubb, R.T., 2011. Sortase enzymes in Gram-positive bacteria. Mol. Microbiol. 82, 1044–1059. https://doi.org/10.1111/j.1365-2958.2011.07887.x.
Suryadinata, R., Seabrook, S.A., Adams, T.E., Nuttall, S.D., Peat, T.S., 2015. Structural and biochemical analyses of a *Clostridium perfringens* sortase D transpeptidase. Acta Crystallogr. Sect. D Biol. Crystallogr. 71, 1505–1513. https://doi.org/10.1107/S1399004715009219.
Susmitha, A., Nampoothiri, K.M., Bajaj, H., 2019. Insights into the biochemical and functional characterization of sortase E transpeptidase of *Corynebacterium glutamicum*. Biochem. J. 476, 3835–3847. https://doi.org/10.1042/BCJ20190812.
Swaminathan, A., Mandlik, A., Swierczynski, A., Gaspar, A., Das, A., Ton-That, H., 2007. Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in *Corynebacterium diphtheriae*. Mol. Microbiol. 66, 961–974. https://doi.org/10.1111/j.1365-2958.2007.05968.x.
Swierczynski, A., Ton-That, H., 2006. Type III Pilus of Corynebacteria : Pilus Length Is Determined by the Level of Its Major Pilin Subunit. J. Bacteriol. 188, 6318–6325. https://doi.org/10.1128/JB.00606-06.
Tamai, E., Katayama, S., Sekiya, H., Nariya, H., Kamitori, S., 2019. Structures of major pilins in *Clostridium perfringens* demonstrate dynamic conformational change. Acta Crystallogr. Sect. D Struct. Biol. 75, 718–732. https://doi.org/10.1107/S2059798319009689.
Tamai, E., Sekiya, H., Maki, J., Nariya, H., Yoshida, H., Kamitori, S., 2017. X-ray structure of *Clostridium perfringens* sortase B cysteine transpeptidase. Biochem. Biophys. Res. Commun. 493, 1267–1272.
Telford, J.L., Barocchi, M.A., Margarit, I., Rappuoli, R., Grandi, G., 2006. Pili in Gram-positive pathogens. Nat. Rev. Microbiol. 4, 509–519. https://doi.org/10.1038/nrmicro1443.
Ton-That, H., Marraffini, L.A., Schneewind, O., 2004. Sortases and pilin elements involved in pilus assemly of *Corynebacterium diphtheriae*. Mol. Microbiol. 53, 251–261. https://doi.org/10.1111/j.1365-2958.2004.04117.x.
Ton-That, H., Schneewind, O., 2003. Assembly of pili on the surface of *Corynebacterium diphtheriae*. Mol. Microbiol. 50, 1429–1438. https://doi.org/10.1046/j.1365-2958.2003.03782.x.
Turner, L.S., Kanamoto, T., Unoki, T., Munro, C.L., Wu, H., Kitten, T., 2009. Comprehensive Evaluation of *Streptococcus sanguinis* Cell Wall-Anchored Proteins in Early Infective Endocarditis. Infect. Immun. 77, 4966–4975. https://doi.org/10.1128/IAI.00760-09.
Turroni, F., Duranti, S., Bottacini, F., Guglielmetti, S., Sinderen, D. Van, Ventura, M., 2014. *Bifidobacterium bifidum* as an example of a specialized human gut commensal. Front. Microbiol. 5, 1–9. https://doi.org/10.3389/fmicb.2014.00437.
Van Leeuwen, H.C., Klychnikov, O.I., Menks, M.A.C., Kuijper, E.J., Drijfhout, J.W., Hensbergen, P.J., 2014. *Clostridium difficile* sortase recognizes a (S/P)PXTG sequence motif and can accommodate diaminopimelic acid as a substrate for transpeptidation. FEBS Lett. 588, 4325–4333. https://doi.org/10.1016/J.FEBSLET.2014.09.041.
Van Pijkeren, J.P., Canchaya, C., Ryan, K.A., Li, Y., Claesson, M.J., Sheil, B., Steidler, L., O'Mahony, L., Fitzgerald, G.F., Van Sinderen, D., O'Toole, P.W., 2006. Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *Lactobacillus salivarius* UCC118. Appl. Environ. Microbiol. 72, 4143–4153. https://doi.org/10.1128/AEM.03023-05.
Von Ossowski, I., 2017. Novel molecular insights about lactobacillar sortase-dependent piliation. Int. J. Mol. Sci. 18, 1551. https://doi.org/10.3390/ijms18071551.
Wang, C., Li, M., Feng, Y., Zheng, F., Dong, Y., Pan, X., Cheng, G., Dong, R., Hu, D., Feng, X., Ge, J., Liu, D., Wang, J., Cao, M., Hu, F., Tang, J., 2009. The involvement of sortase A in high virulence of STSS-causing *Streptococcus suis* serotype 2. Arch. Microbiol. 191, 23–33. https://doi.org/10.1007/s00203-008-0425-z.
Wei, X., Wang, S., Zhao, X., Wang, X., Li, H., Lin, W., Lu, J., Zhurina, D., Li, B., Riedel, C. U., Sun, Y., Yuan, J., 2016. Proteomic profiling of *Bifidobacterium bifidum* S17 cultivated under in vitro conditions. Front. Microbiol. 7, 1–10. https://doi.org/10.3389/fmicb.2016.00097.
Weiner, E.M., Robson, S., Marohn, M., Clubb, R.T., 2010. The sortase A enzyme that attaches proteins to the cell wall of *Bacillus anthracis* contains an unusual active site architecture. J. Biol. Chem. 285, 23433–23443. https://doi.org/10.1074/jbc.M110.135434.
Weiss, W.J., Lenoy, E., Murphy, T., Tardio, L.A., Burgio, P., Projan, S.J., Schneewind, O., Alksne, L., 2004. Effect of *srtA* and *srtB* gene expression on the virulence of *Staphylococcus aureus* in animal models of infection. J. Antimicrob. Chemother. 53, 480–486. https://doi.org/10.1093/jac/dkh078.

Westermann, C., Zhurina, D.S., Baur, A., Shang, W., Yuan, J., Riedel, C.U., 2012. Exploring the genome sequence of *Bifidobacterium bifidum* S17 for potential players in host-microbe interactions. Symbiosis. 58, 191–200. https://doi.org/10.1007/s13199-012-0205-z.

Wu, C., Mishra, A., Reardon, M.E., Huang, I.H., Counts, S.C., Das, A., Ton-That, H., 2012. Structural determinants of actinomyces sortase SrtC2 required for membrane localization and assembly of type 2 fimbriae for interbacterial coaggregation and oral biofilm formation. J. Bacteriol. 194, 2531–2539. https://doi.org/10.1128/JB.00093-12.

Wu, C., Mishra, A., Yang, J., Cisar, J.O., Das, A., Ton-That, H., 2011. Dual function of a tip fimbrillin of Actinomyces in fimbrial assembly and receptor binding. J. Bacteriol. 193, 3197–3206. https://doi.org/10.1128/JB.00173-11.

Yamaguchi, M., Terao, Y., Ogawa, T., Takahashi, T., Hamada, S., Kawabata, S., 2006. Role of *Streptococcus sanguinis* sortase A in bacterial colonization. Microbes Infect. 8, 2791–2796. https://doi.org/10.1016/j.micinf.2006.08.010.

Zhang, J., Yang, F., Zhang, X., Jing, H., Ren, C., Cai, C., Dong, Y., Zhang, Y., Zou, Q., Zeng, H., 2015. Protective Efficacy and Mechanism of Passive Immunization with Polyclonal Antibodies in a Sepsis Model of *Staphylococcus aureus* Infection. Sci. Rep. 5, 1–12. https://doi.org/10.1038/srep15553.

Zhang, R., Wu, R., Joachimiak, G., Mazmanian, S.K., Missiakas, D.M., Gornicki, P., Schneewind, O., Joachimiak, A., 2004. Structures of Sortase B from *Staphylococcus aureus* and *Bacillus anthracis* Reveal Catalytic Amino Acid Triad in the Active Site. Structure. 12, 1147–1156. https://doi.org/10.1016/J.STR.2004.06.001.

Zink, S.D., Burns, D.L., 2005. Importance of *srtA* and *srtB* for growth of *Bacillus anthracis* in macrophages. Infect. Immun. 73, 5222–5228. https://doi.org/10.1128/IAI.73.8.5222-5228.2005.

Sundar *et al. AMB Expr* (2020) 10:68
https://doi.org/10.1186/s13568-020-01003-9

AMB Express

ORIGINAL ARTICLE

Open Access

Heterologous expression of genes for bioconversion of xylose to xylonic acid in *Corynebacterium glutamicum* and optimization of the bioprocess

M. S. Lekshmi Sundar[1,2], Aliyath Susmitha[1,2], Devi Rajan[1], Silvin Hannibal[3], Keerthi Sasikumar[1,2], Volker F. Wendisch[3] and K. Madhavan Nampoothiri[1,2*]

Abstract

In bacterial system, direct conversion of xylose to xylonic acid is mediated through NAD-dependent xylose dehydrogenase (*xylB*) and xylonolactonase (*xylC*) genes. Heterologous expression of these genes from *Caulobacter crescentus* into recombinant *Corynebacterium glutamicum* ATCC 13032 and *C. glutamicum* ATCC 31831 (with an innate pentose transporter, *araE)* resulted in an efficient bioconversion process to produce xylonic acid from xylose. Process parameters including the design of production medium was optimized using a statistical tool, Response Surface Methodology (RSM). Maximum xylonic acid of 56.32 g/L from 60 g/L xylose, i.e. about 76.67% of the maximum theoretical yield was obtained after 120 h fermentation from pure xylose with recombinant *C. glutamicum* ATCC 31831 containing the plasmid pVWEx1 *xylB*. Under the same condition, the production with recombinant *C. glutamicum* ATCC 13032 (with pVWEx1 *xylB*) was 50.66 g/L, i.e. 69% of the theoretical yield. There was no significant improvement in production with the simultaneous expression of *xylB* and *xylC* genes together indicating xylose dehydrogenase activity as one of the rate limiting factor in the bioconversion. Finally, proof of concept experiment in utilizing biomass derived pentose sugar, xylose, for xylonic acid production was also carried out and obtained 42.94 g/L xylonic acid from 60 g/L xylose. These results promise a significant value addition for the future bio refinery programs.

Keywords: *Corynebacterium glutamicum*, Biomass, Heterologous expression, Response surface methodology (RSM), Xylose, Xylonic acid, Xylose dehydrogenase

Key points

- Made *C. glutamicum* recombinants with genes for xylose to xylonic acid conversion.
- Bioprocess development using *C. glutamicum* for xylonic acid.
- Conversion of biomass derived xylose to xylonic acid.

*Correspondence: madhavan@niist.res.in; madhavan85@hotmail.com
[1] Microbial Processes and Technology Division, CSIR–National Institute for Interdisciplinary Science and Technology (NIIST), Thiruvananthapuram 695019, Kerala, India
Full list of author information is available at the end of the article

Introduction

D-xylonic acid, an oxidation product of xylose, is a versatile platform chemical with multifaceted applications in the fields of food, pharmaceuticals, and agriculture. It is considered by the U.S. Department of Energy to be one of the 30 chemicals of highest value because it can be used in a variety of applications, including as a dispersant, pH regulator, chelator, antibiotic clarifying agent and health enhancer (Byong-Wa et al. 2006; Toivari et al. 2012). Xylonic acid may also be used as a precursor for bio-plastic, polymer synthesis and other chemicals such as 1,2,4-butanetriol (Niu Wei et al. 2003). Although xylonic acid production is feasible via chemical oxidation

using platinum or gold catalysts, selectivity is relatively poor (Yim et al. 2017). As the pentose sugar catabolism is restricted to the majority of the industrial microbes (Wisselink et al. 2009), microbial conversion of xylose to xylonic acid gained interest. As of now, biogenic production of xylonic acid has been accomplished in various microorganisms, including *Escherichia coli*, *Saccharomyces cerevisiae* and *Kluyveromyces lactis* by introducing *xylB (*encoding *xylose dehydrogenase)* and *xylC* (encoding *xylonolactonase)* genes from *Caulobacter crescentus or Trichoderma reesei* (Nygård et al. 2011; Toivari et al. 2012; Cao et al. 2013).

As xylose is the monomeric sugar required for xylonic acid production, a lot of interest has been paid on utilizing xylose generated from lignocellulosic biomass (Lin et al. 2012). Bio-transformation of lignocellulosic biomass into platform chemicals is possible only through its conversion to monomeric sugars, mostly by pretreatment, i.e. pre-hydrolysis by alkali or acid at higher temperature or via enzymatic hydrolysis. Monomeric hexose and pentose sugars are generated from lignocellulosic biomass along with inhibitory by-products like furfural, 5-hydroxymethylfurfural, 4-hydroxybenzaldehyde that affect the performance of microbial production hosts (Matano et al. 2014). The concept of biomass refinery is getting more and more attraction for the cost effectiveness of the 2G ethanol program. Microbial production of value-added products such as biopolymers, bioethanol, butanol, organic acids and xylitol were reported utilizing the C5 stream generated by the pretreatment of biomass by different microbes like *Pichia stipitis, Clostridium acetobutylicum, Candida guilliermondii, Bacillus coagulans* (Mussatto and Teixeira 2010; Ou et al. 2011; de Arruda et al. 2011; Lin et al. 2012; Raganati et al. 2015).

Although some of the industrial strains are capable of pentose fermentation, most of them are sensitive to inhibitors of lignocellulosic biomass pretreatment. However, *Corynebacterium glutamicum* showed remarkable resistance towards these inhibitory by-products under growth-arrested conditions (Sakai et al. 2007). *C. glutamicum* is a Gram-positive, aerobic, rod-shaped, non-spore forming soil actinomycete which exhibits numerous ideal intrinsic attributes as a microbial factory to produce amino acids and high-value chemicals (Heider and Wendisch 2015; Hirasawa and Shimizu 2016; Yim et al. 2017). This bacterium has been successfully engineered towards producing a broad range of products, including diamines, amino-carboxylic acids, diacids, recombinant proteins and even industrial enzymes (Becker et al. 2018; Baritugo et al. 2018). A lot of metabolic resurrections were reported in *C. glutamicum* for the production of chemicals like amino acids, sugar acid, xylitol and biopolymers from hemicellulosic biomasses such as wheat bran, rice straw and sorghum stover (Gopinath et al. 2011; Wendisch et al. 2016; Dhar et al. 2016).

Since *C. glutamicum* lacks the genes for the metabolic conversion of xylose to xylonic acid, the heterologous expression of xylose dehydrogenase (*xylB*) and xylonolactonase (*xylC*) genes from *Caulobacter crescentus* was attempted. In addition to ATCC 13032 wild type, we also explored the *C.glutamicum* ATCC 31831 culture which contains a pentose transporter gene (*araE*) which enables the uptake of pentose sugar (Kawaguchi et al. 2009; Choi et al. 2019). Both *xylB* and *xylC* genes individually, as well as together as *xylBC*, were amplified from xylose operon of *C. crescentus* and the plasmids were transformed to both *C. glutamicum* strains and checked the xylonic acid production.

Materials and methods

Microbial strains and culture conditions

Microbial strains and plasmids used in this study are listed in Table 1. For genetic manipulations, *E. coli* strains were grown at 37 °C in Luria–Bertani (LB) medium. *C. glutamicum* strains were grown at 30 °C in Brain Heart Infusion (BHI) medium. Where appropriate, media were supplemented with antibiotics. The final antibiotic concentrations for *E. coli* and *C. glutamicum* were 25 µg/ml of kanamycin. Culture growth was measured spectrophotometrically at 600 nm using a UV–VIS spectrophotometer (UVA-6150, Shimadzu, Japan).

Molecular techniques and strain construction

Standard molecular techniques were done according to the protocol described by (Sambrook et al. 2006). Genomic DNA isolation was done with Gen Elute genomic DNA isolation kit (Sigma, India). Plasmid isolation was done using Qiagen plasmid midi kit (Qiagen, Germany). Polymerase chain reaction (PCR) was performed using automated PCR System (My Cycler, Eppendorff, Germany) in a total volume of 50 µl with 50 ng of DNA, 0.2 mM dNTP in PrimeSTAR™ buffer (Takara), and 1.25 U of PrimeSTAR™ HS DNA polymerase (Takara) and the PCR product was purified by QIA quick PCR purification kit (Qiagen, Germany) as per the instructions provided by the manufacturers. Competent *E. coli* DH5α cells were prepared by Transformation and Storage Solution (TSS) method and transformed by heat shock (Chung and Miller 1993). The *C. glutamicum* competent cells were electroporated to achieve the transformation (van der Rest et al. 1999).

Xylose dehydrogenase (*xylB*) and xylonolactonase (*xylC*) and *xylBC* genes together of *Caulobacter crescentus* were amplified from the xylose-inducible *xylXABCD* operon (CC0823–CC0819) (Stephens et al. 2007) by polymerase chain reaction (PCR) with appropriate primers as

Table 1 Microbial strains, plasmids and primers used in the study

Strains and vectors	Descriptions	References
Microbial strains		
Corynebacterium glutamicum	ATCC13032, wild type (WT)	Abe et al. (1967)
Corynebacterium glutamicum	ATCC 31831	Kinoshita et al. (2004)
Escherichia coli DH5α	*Fthi-1 endA1 hsdr17(r-, m-) supE44 _lacU169 f80lacZ_M15) recA1 gyrA96 relA1*	Hanahan and Harbor (1983)
Plasmid vectors		
pVWEx1	Kan^r; *E. coli-C. glutamicum* shuttle vector	Peters-Wendisch et al. (2001)
pEKEx3 xylXABCD	Spec^r; pEKEx3 derivative for the regulated expression of *xylXABCD*$_{Cc}$ of *C. crescentus*	This study
Primers (sequences 5′–3′)		
xylB-pVW-fw	CGCCAAGCTTGCATGC**CTGCAG**TAAAGGAGATATACATATGTCCTCAGCCATCTATCC	This study
xylB-pVW-rw	CGAGCTCGGTACCCGG**GGATCC**CTTCACGCTGGGCCGGGATG	This study
xylC-pVW-fw	CGCCAAGCTTGCATGC**CTGCAG**TAAAGGAGATATACATATGACCGCTCAAGTCACTTG	This study
xylC-pVW-rw	CGAGCTCGGTACCCGG**GGATCC**GGGCGTGCGGTTAGACAAGG	This study
xylBC-pVW-fw	TGTTTAAGTTTAGTGGATGGGATGACCGCTCAAGTCACTTGCGTATGGG	This study
xylBC-pVW-fw	CCCATCCACTAAACTTAAACATCAACGCCAGCCGGCGTCGATCC	This study

shown in Table 1 and the purified PCR products (747 bp *xylB*, 870 bp *xylC* and 1811 bp *xylBC*) were verified by sequencing and cloned into the restriction digestion site (*Bam* HI/*Pst* I) of pVWEx1 shuttle vector. The engineered plasmids so-called pVWEx1*xylB*, pVWEx1*xylC* and pVWEx1*xylBC* were transformed into *E. coli* DH5α and the transformants bearing pVWEx1 derivative were screened in LB medium supplemented with kanamycin (25 μg mL^{-1}). Competent cells of *C. glutamicum* ATCC 13032 and ATCC 31831 were prepared and the plasmids were electroporated into both the *C. glutamicum* strains with parameters set at 25 μF, 600 Ω and 2.5 kV, yielding a pulse duration of 10 ms and the positive clones were selected in LBHIS kanamycin (25 μg mL^{-1}) plates (van der Rest et al. 1999).

Fermentative production of xylonic acid by *C. glutamicum* transformants

For xylonic acid production, *C. glutamicum* was inoculated in 10 ml of liquid medium (BHI broth) in a test tube and grown overnight at 30 °C under aerobic condition with shaking at 200 rpm. An aliquot of the 10 ml culture was used to inoculate 100 ml CGXII production medium (Keilhauer et al. 1993) containing 35 g/L xylose and 5 g/L glucose as carbon sources, kanamycin (25 μg mL^{-1}). IPTG (1 mM) induction was done along with the inoculation. Fermentation was carried out in 250 mL Erlenmeyer flasks containing 100 mL production medium and incubated as described above. Samples were withdrawn at regular intervals to determine sugar consumption and xylonic acid production. Since *xylB* transformant was found to be the best producer, a comparison of it with *C. glutamicum* ATCC 13032 having *xylB* gene was also carried out to see whether the inbuilt *araE* pentose transporter in ATCC 31831 has any advantage over wild type ATCC 13032.

Media engineering by response surface methodology (RSM)

Response surface methodology was applied to identify the operating variables that have a significant effect on xylonic acid production. A Box Behnken experimental design (BBD) (Box and Behnken 1960) with four independent variables (selected based on single parameter study, data not shown) that may affect xylonic acid production, including $(NH_4)_2SO_4$ (2.5–12.5 g/L), urea (4.5–18.5 g/L), xylose (30–90 g/L) and inoculum (7.5–1.125%) were studied at three levels −1, 0 and +1 which correspond to low, medium and high values respectively. Responses were measured as titer (g/L) of xylonic acid. The statistical as well as numerical analysis of the model was evaluated by analysis of variance (ANOVA) which included p-value, regression coefficient, effect values and F value using Minitab 17 software. Studies were performed using *C. glutamicum* ATCC 31831 harboring pVWEx1-*xylB*.

Dilute acid pretreatment of the biomass

The rice straw was crushed into fine particle (size of 10 mm) and pre-soaked in dilute acid (H_2SO_4) for 30 min, pretreated with 15% (w/w) biomass loading and 1% (w/w) acid concentration at 121 °C for 1 h. After cooling, the mixture was neutralized to pH 6–7 using 10 N NaOH. The liquid portion, i.e. acid pretreated liquor (APL) rich in pentose sugar (xylose) was separated from the pretreated slurry and lyophilized to

concentrate to get desired xylose level which was estimated prior to the shake flask fermentation studies.

Quantification of sugars and xylonic acid in fermentation broth

The qualitative and quantitative analysis of sugars and sugar acid (xylonic acid) was performed using an automated high-performance liquid chromatography (HPLC) system (Prominence UFLC, Shimadzu, Japan) equipped with auto-sampler, column oven and RI Detector. The monomeric sugars (xylose and glucose) were resolved with Phenomenex Rezex RPM Pb^{+} cation exchange monosaccharide column (300 × 7.5 mm) operated at 80 °C. MilliQ water (Millipore) with a flow rate of 0.6 mL/min was used as the mobile phase. For xylonic acid detection, Phenomenex organic acid column (250 mm × 4.6 mm × 5 µm) operated at 55 °C was used with a mobile phase of 0.01 N H_2SO_4 at a flow rate of 0.6 mL/min. The samples were centrifuged (13,000 rpm for 10 min at 4 °C) and filtered using 0.2 µm filters (Pall Corporation, Port Washington, New York) for analysis.

Results

Xylose utilization and xylonic acid production by *C. glutamicum* transformants

Corynebacterium glutamicum recombinants expressing *xylB*, *xylC* and *xylBC* were constructed. The xylose dehydrogenase and xylonolactonase genes were cloned into IPTG-inducible expression vector pVWEx1 and transformed into *C. glutamicum* ATCC 31831. To check xylonic acid production from xylose, the *C. glutamicum* ATCC 31831 transformants harboring pVWEx1-*xylB,* pVWEx1-*xylC* and pVWEx1-*xylBC* were cultivated in CGXII medium containing 5 g/L of glucose as the carbon source for initial cell growth and 35 g/L of xylose as the substrate for xylonic acid production. Cell growth, xylose consumption and xylonic acid production were analyzed during the incubation for a desired period of interval. From analysis, it is clear that compared to the control strain with empty vector (Fig. 1a), the transformants harboring pVWEx1-*xylB* picked up growth very fast compared to the other transformants and utilized xylose effectively (77.2% utilization after 120 h) and resulted in maximum production of 32.5 g/L xylonic acid (Fig. 1b). The pVWEx1-*xylBC* harboring strain produced 26 g/L xylonic acid (Fig. 1d), whereas pVWEx1-*xylC* showed

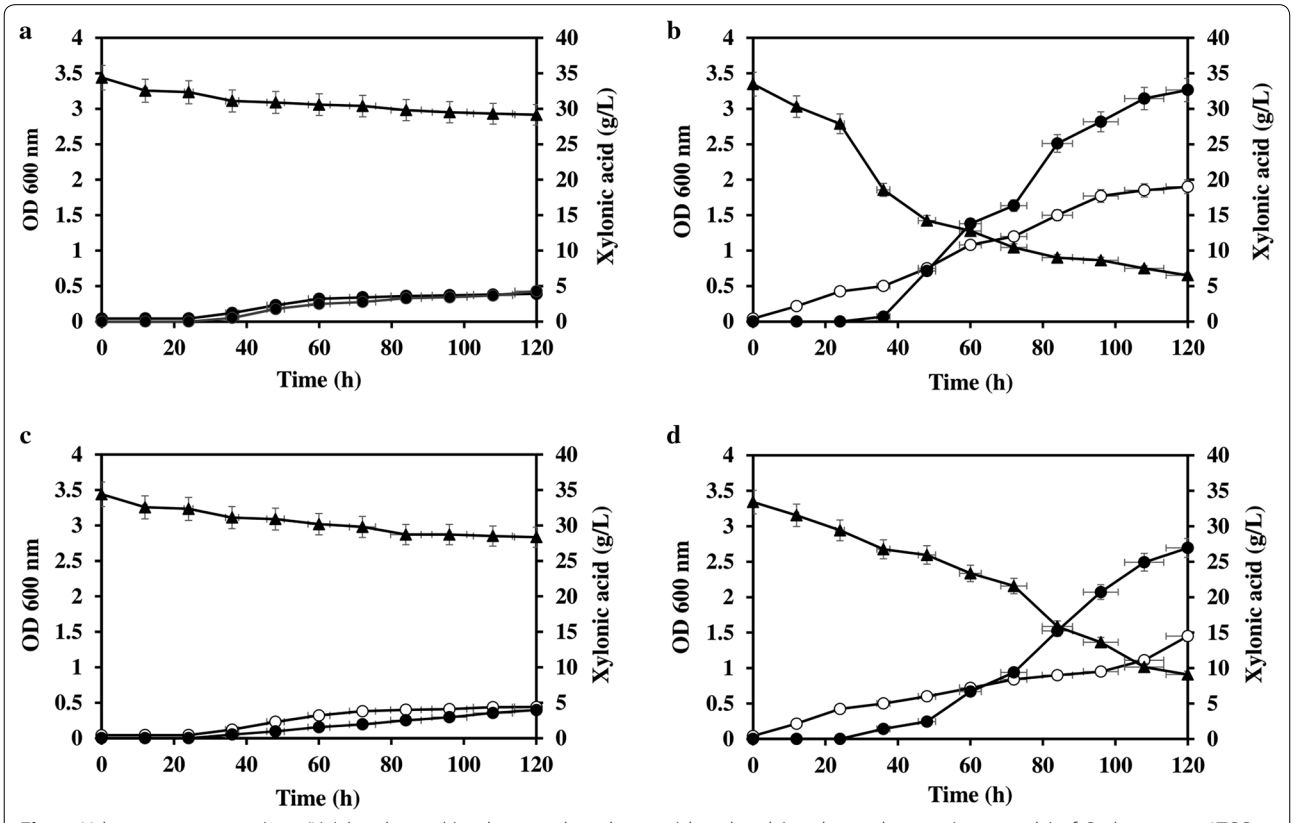


Fig. 1 Xylose consumption (35 g/L) (*closed triangle*), xylonic acid production (closed circle) and growth curve (open circle) of *C. glutamicum* ATCC 31831 (**a**) pVWEx1 (**b**) pVWEx1-*xyl B* (**c**) pVWEx1-*xylC* (**d**) pVWEx1-*xyl BC* respectively

neither any significant xylose uptake nor xylonic acid production (Fig. 1c).

Box–Behnken experimental design (BBD) and operational parameter optimization

The objective of the experimental design was medium engineering for maximum xylonic acid production. There were a total of 15 runs for optimizing the four individual parameters in the current BBD. Experimental design and xylonic acid yield are presented in Table 2. The polynomial equation obtained for the model was as below:

$$\begin{aligned}\text{Xylonic acid}\left(\text{g/L}\right) = & -48.7-0.45\,X_1+3.5\,X_2+0.220\,X_3+2.058\,X_4\\ & -0.019\,X_1^2-0.2139\,X_2^2-0.0423\,X_3^2-0.01943\,X_4^2\\ & -0.075\,X_1X_2+0.0416\,X_1X_3-0.0119\,X_1X_4\\ & +0.526\,X_2X_3+0.0482\,X_2X_4-0.00128\,X_3X_4\end{aligned}$$

where X_1, X_2, X_3 and X_4 are xylose, $(NH4)_2SO_4$, urea and inoculum concentration respectively. Maximum production efficiency (0.47 g^{-1} L^{-1} h^{-1}) was observed with Run No.13 where the concentration of parameters was urea 11.5 g/L, xylose 60 g/L, $(NH_4)_2SO_4$ 7.5 g/L and inoculum 1.125% and xylonic acid titer was 56.32 g/L. It indicates that $(NH_4)_2SO_4$, inoculum concentration and xylose have a significant positive effect than urea on xylonic acid yield.

Response surface curves were plotted to find out the interaction of variables and to determine the optimum level of each variable for maximum response. The contour plot showing the interaction between a pair of factors on xylonic acid yield is given in Fig. 2a–f. Major interactions studied are of inoculum and xylose concentration (a), xylose and urea concentration (b), $(NH_4)_2SO_4$ and urea concentration (c), effect of inoculum and $(NH_4)_2SO_4$ concentration (d), effect of $(NH_4)_2SO_4$ and xylose concentration (e) and the interaction of inoculum and urea concentration (f).

The ANOVA of response for xylonic acid is shown in Table 3. The R^2 value explains the variability in the xylonic acid yield associated with the experimental factors to the extent of 97.48%.

Role of *araE* pentose transporter for enhanced uptake of xylose and xylonic acid production

Using the designed medium standardized for *C. glutamicum* ATCC 31831, which possesses an arabinose and xylose transporter encoded by *araE*, a comparative production study was carried out with recombinant *C. glutamicum* ATCC 13032. Both the strains grew well in the CGXII production medium and metabolized xylose to xylonic acid. After 120 h fermentation, the recombinant strain, ATCC 13032 produced 50.66 g/L of xylonic acid whereas ATCC 31831 produced 56.32 g/L (Fig. 3). It was observed that better uptake of the pentose sugar was also exhibited by *C. glutamicum* ATCC 31831, i.e.,

Table 2 Box–Behnken experimental design matrix with experimental values of xylonic acid production by *Corynebacterium glutamicum* ATCC 31831

Run order	Urea (g/L)	Xylose (g/L)	$(NH_4)_2SO_4$ (g/L)	Inoculum (% v/v)	Xylonic acid (g/L)
1	11.5	60	7.5	11.25	56.119
2	11.5	90	2.5	11.25	59.792
3	11.5	30	12.5	7.5	25.061
4	4.5	30	7.5	15	21.359
5	18.5	60	2.5	15	52.481
6	11.5	30	2.5	7.5	25.061
7	11.5	90	12.5	15	58.418
8	4.5	60	12.5	11.25	30.341
9	18.5	90	7.5	15	58.795
10	4.5	90	7.5	11.25	45.749
11	18.5	60	12.5	15	48.982
12	11.5	60	7.5	15	56.018
13	11.5	60	7.5	15	*56.318*
14	18.5	30	7.5	11.25	28.349
15	4.5	60	2.5	7.5	28.816

Maximum conversion of xylose to xylonic acid indicated in italic

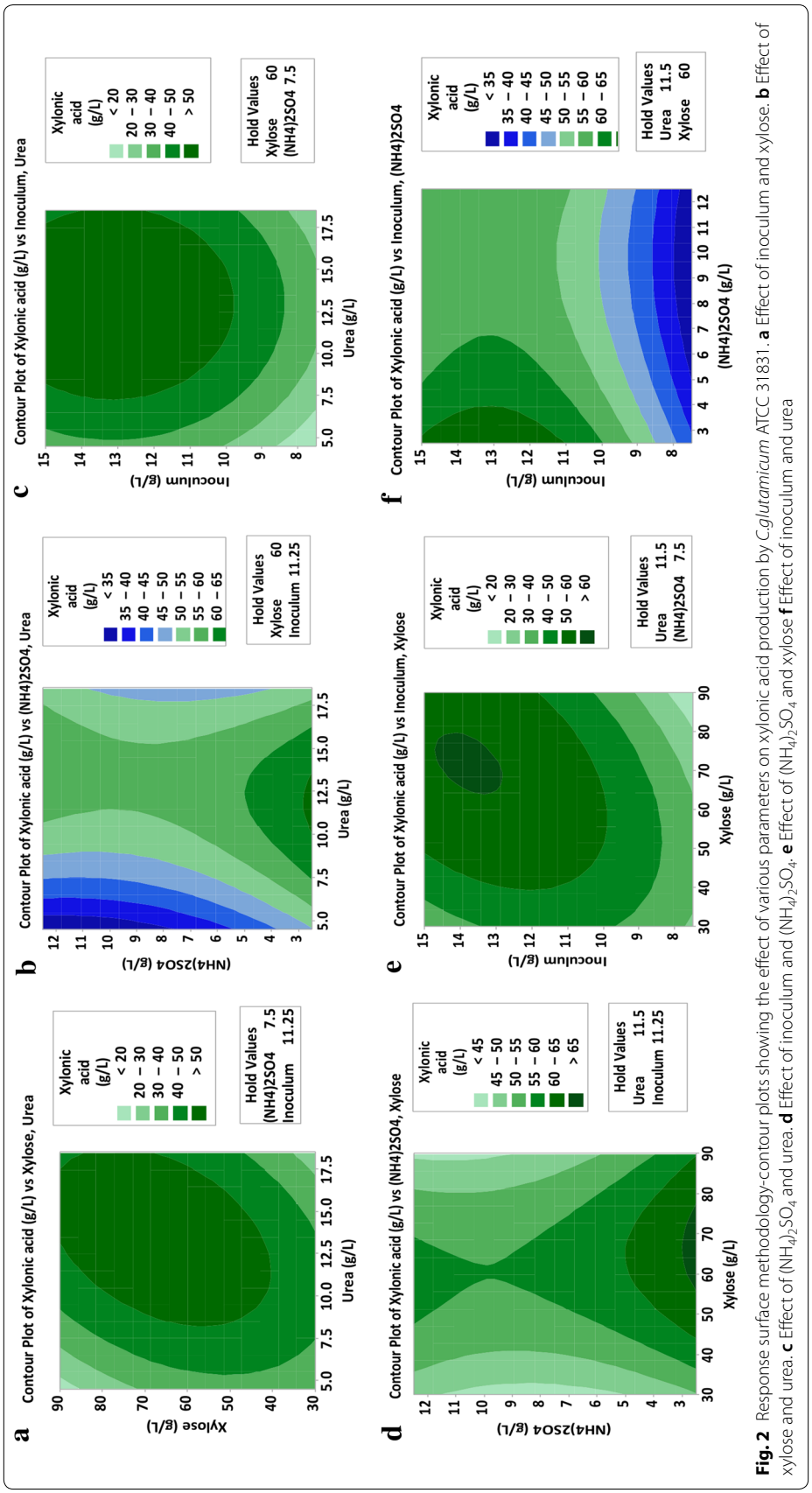


Fig. 2 Response surface methodology-contour plots showing the effect of various parameters on xylonic acid production by *C.glutamicum* ATCC 31831. **a** Effect of inoculum and xylose. **b** Effect of xylose and urea. **c** Effect of $(NH_4)_2SO_4$ and urea. **d** Effect of inoculum and $(NH_4)_2SO_4$. **e** Effect of $(NH_4)_2SO_4$ and xylose **f** Effect of inoculum and urea

Table 3 Analysis of variance for xylonic acid production using *C. glutamicum* ATCC 31831

Source	DF	Adj SS	Adj MS	F	P
Regression	12	3583.09	298.591	6.45	0.142
Linear	4	1688.34	422.234	9.11	0.101
Square	4	1249.59	312.398	6.74	0.133
Interaction	4	284.83	71.208	1.54	0.431
Residual error	2	92.66	46.328		
Lack-of-fit	1	0.00	92.657		
Pure error	1		0.000		
Total	14	3675.75			

S = 6.80649, R-Sq = 97.48%, R-Sq (pred) = 0.00% and R-Sq (adj) = 82.35%

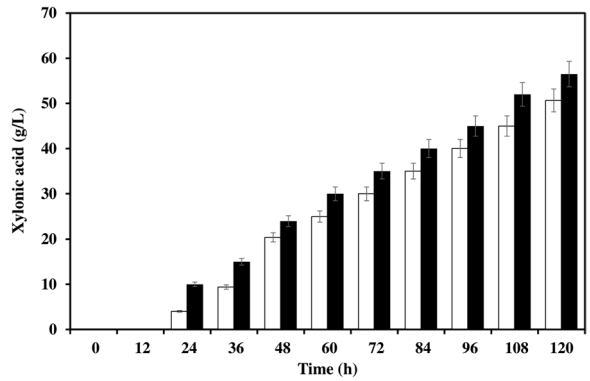


Fig. 3 Xylonic acid production by *C. glutamicum* ATCC 13032 (open bar) and *C. glutamicum* ATCC 31831 (closed bar) harbouring plasmid pVWEx1-*xylB*

75% consumption compared to 60% by ATCC 13032 after 120 h fermentation and same the case with culture growth where ATCC 31831 showed better growth (10× dilution of culture broth for spectrophotometric reading (Additional file 1: Figure S1).

Xylonic acid from rice straw hydrolysate

Fermentation was carried out in rice straw hydrolysate using *C. glutamicum* ATCC 31831 (pVWEx1-*xylB*). The strain could grow in different xylose concentrations (of 20, 40, and 60 g/L) in rice straw hydrolysate, and after 120 h fermentation, maximum titer obtained was 42.94 g/L xylonic acid from 60 g/L xylose (Fig. 4). A production yield of 58.48% xylonic acid in hydrolysate is remarkable for sugar acid production with engineered strain of *C. glutamicum* which is quite tolerant to the inhibitors present in the hydrolysate.

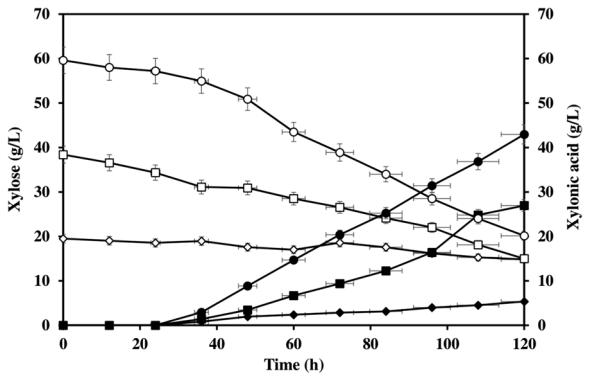


Fig. 4 Xylose utilization (open symbols) and xylonic acid production (closed symbols) by *C. glutamicum* ATCC 31831 (pVWEx1-*xylB*) in rice straw hydrolysate containing different concentrations of xylose 20 g/L (open diamond), 40 g/L (open square) and 60 g/L (open circle). Xylonic acid production from 20 g/L xylose (closed diamond), 40 g/L xylose (closed square) and 60 g/L xylose (closed circle)

Discussion

Heterologous expression of genes for the production of varied value-added chemicals were successfully carried out in *C. glutamicum*, for example, the production of amino acids, sugar alcohol, organic acid, diamines, glycolate and 1,5-diaminopentane (Buschke et al. 2013; Meiswinkel et al. 2013; Zahoor et al. 2014; Pérez-García et al. 2016; Dhar et al. 2016). *C. glutamicum* being a versatile industrial microbe and the availability of genetic engineering tools makes it a rapid and rational manipulation host for diverse platform chemicals. Most corynebacteria are known not to utilize xylose as carbon source. The absence of xylose metabolizing genes restricts the growth of *Corynebacterium* in pentose rich medium. To develop an efficient bioconversion system for xylonic acid synthesis, the genes of *Caulobacter crescentus* were expressed in *C. glutamicum.* The resulting transformants *C. glu*-pVWEx1-*xylB* and *C.glu*-pVWEx1-*xylBC* were able to grow in mineral medium containing xylose and converted it into corresponding pentonic acid.

Xylose can be metabolized in four different routes (I) The oxido-reductase pathway, (II) The isomerase pathway, (III) The Weimberg pathway, an oxidative pathway and (IV) The Dahms pathway (Cabulong et al. 2018). Xylose once inside the cell gets converted to xylonolactone and then into xylonic acid on the expression of two genes namely, *xylB* (xylose dehydrogenase) and *xylC* (xylonolactonase). These two enzymes are involved in both the Weimberg and Dahms pathway where xylose is metabolized to xylonic acid (Brüsseler et al. 2019). In the present study, it is observed that only the xylose dehydrogenase enzyme activity is good enough for xylonic acid production. Without the dehydrogenase

Table 4 Comparison of xylonic acid production and productivity by the best xylonic acid producers

Microorganism	D-xylose (g/l)	D-xylonate (g/l)	Yields (g/g)	Volumetric productivity (g/l/h)	Specific productivity [g(g/biomass)/h]	PH	Biomass (g/l)	Process	References
Gluconobacter oxydans (ATCC 621)	100	109	1.1	2.5	~1.5	5.5	1.7	Batch	Buchert et al. (1988)
Gluconobacter oxydans (ATCC 621)	100	107	1.1	2.2	~1.5	4.5	1.3	Batch	Buchert et al. (1988)
Pseudomonas fragi (ATCC 4973)	150	162	1.1	1.4	0.2	6.5	6.9	Batch	Buchert et al. (1988)
Pseudomonas putida	~0.4	~0.4	~1	~1.9	~0.7	6.8	2.9	Continuous	Meijnen et al. (2009)
Enterobacter cloacea	200	190	~1	~1.6	–	6.5	nd	Batch	Ishizaki et al. (1973)
Escherichia coli	40	39	1.0	1.1	0.14	7.0	~8	Batch	Liu et al. (2012)
Saccharomyces cerevisiae Xyd 1	20	4	0.4	0.03	0.007	5.5	4.6	Batch	Toivari et al. (2010)
Saccharomyces cerevisiae SUS2DD	23	3	0.4	0.02	0.006	5.5	5.3	Batch	Toivari et al. (2012)
Saccharomyces cerevisiae xylB	23	17	0.8	0.23	0.06	5.5	5	Batch	Toivari et al. (2012)
Kluyveromyces lactis Xyd 1	23	8	0.4	0.13	0.01	5.5	9	Batch	Nygård et al. (2011)
Corynebacterium glutamicum (ATCC 13032)	20	6.23	1.04	1.02	–	–	–	Batch	Yim et al. (2017)
Corynebacterium glutamicum (ATCC 31831)	60	56.32	~1	0.93	–	5.5	1.4	Batch	This study

activity, the lactonase activity alone cannot do the conversion of xylose to xylonic acid. Further, the xylonolactonase expression along with xylose dehydrogenase resulted in xylonic acid production but not that efficient as dehydrogenase alone with the case of *C. glutamicum*. It is reported that, xylonolactone once formed can be converted to xylonic acid either by the spontaneous oxidation of lactone or through the enzymatic hydrolysis of xylonolactonase enzyme (Buchert and Viikari 1988). *Corynebacterium glutamicum* being an aerobic organism, direct oxidation of xylonolactone to xylonic acid is more favorable inside the cell. Previous studies have also shown that xylose dehydrogenase (*xylB*) activity alone can result in the production of xylonic acid (Yim et al. 2017).

Corynebacterium glutamicum ATCC 31831 grew on pentose as the sole carbon source. The gene cluster responsible for pentose utilization comprised a six-cistron transcriptional unit with a total length of 7.8 kb. The sequence of the *C. glutamicum* ATCC 31831 *ara* gene cluster containing gene *araE*, encodes pentose transporter, facilitates the efficient uptake of pentose sugar (Kawaguchi et al. 2009). Previous studies have also reported the role of *araE* pentose transporter in *Corynebacterium glutamicum* ATCC 31831 and its exploitation for the production of commodity chemicals like 3HP and ethanol (Becker et al. 2018). In the present study, *Corynebacterium glutamicum* ATCC 31831 with an inbuilt *araE* pentose transporter exhibited effectual consumption of xylose as well as its conversion to xylonic acid. Further studies have to be done to explore the role of the same *araE* pentose transporter as an exporter for xylonic acid.

Micrococcus spp., *Pseudomonas*, *Kluveromyces lactis*, *Caulobacter*, *Enterobacter*, *Gluconobacter*, *Klebsiella* and *Pseudoduganella danionis* (ISHIZAKI et al. 1973; Buchert et al. 1988; Buchert and Viikari 1988; Toivari et al. 2011; Wiebe et al. 2015; Wang et al. 2016; Sundar Lekshmi et al. 2019) are the non-recombinant strains reported for xylonic acid production. Among which *Gluconobacter oxydans* is the prominent wild-type strain exhibits higher titers of xylonic acid up to 100 g L^{-1} (Toivari et al. 2012).

Although these strains are capable of producing xylonic acid from pure sugar, they fail to perform as an industrial strain since some are opportunistic pathogen grade and they are not tested in hydrolysate medium may be due to their lower tolerance towards lignocellulosic inhibitors. There was an earlier report on recombinant *C. glutamicum* ATCC 13032 produced 6.23 g L^{-1} of xylonic acid from 20 g L^{-1} of xylan (Yim et al. 2017). In this study they have employed multiple modules, (i) xylan degradation module (ii) conversion module from xylose to xylonic acid by expression of *xdh* gene and (iii) xylose transport module by expression of *xylE* gene, and optimized gene expression introducing promoters (Yim et al. 2017). The product titers with *C. glutamicum* ATCC 31831 presented in this study are comparable with other wild type and recombinant strains (Table 4) and the volumetric productivity in the feed phase can outperform the titers published employing the recombinant *C. glutamicum* ATCC 13032.

Media engineering was carried out with the statistical tool response surface methodology (RSM) for the enhanced production of xylonic acid. The Box–Behnken model with experimental values containing 15 runs was designed for the optimization study. RSM aided to narrow down the most influencing parameters and its optimization on xylonic acid production. The engineered strain produced up to 56.3 g/L of xylonic acid and is characterized by high volumetric productivity and maximum product yield of 76.67% under optimized conditions applying defined xylose/glucose mixtures in synthetic medium. One of the major challenges is the range of acidic and furan aldehyde compounds released from lignocellulosic pre-treatment. Here, the recombinant *C. glutamicum* ATCC 31831 could resist the inhibitors present in rice straw hydrolysate and produced xylonic acid nearly to 58.5% of the maximum possible yield.

The challenges involve getting sufficient xylose after pretreatment and also the separation of xylonic acid from the fermented broth. For the industrial application, downstream processing of xylonic acid is very important. Ethanol precipitation and product recovery by extraction are the two interesting options described for the purification of xylonic acid from the fermentation broth (Liu et al. 2012). With this industrially streamlined recombinant strain a highly profitable bioprocess to produce xylonic acid from lignocellulosic biomass as a cost-efficient second-generation substrate is well within the reach. The one-step conversion of xylose to xylonic acid and the bioprocess developed in the present study favors pentose sugar utilization in rice straw in a straight forward and cost-effective method. The proof of concept showed the simultaneous utilization of biomass-derived sugars (C5 and C6) and it has to be investigated in detail.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13568-020-01003-9.

Additional file 1: Figure S1. Growth (circles) and xylose consumption (triangles) by *C. glutamicum* ATCC 13032 (pVWEx1-*xylB*) (open symbols) and *C. glutamicum* ATCC 31831 (pVWEx1-*xylB*) (closed symbols) in CGXII medium containing 60 g/L xylose.

Acknowledgements
The first author LS acknowledges the Senior Research Fellowship (SRF) by Council of Scientific and Innovative Research (CSIR), New Delhi. KMN and VFW acknowledge the financial assistance from DBT, New Delhi BMBF, and Germany to work on *Corynebacterium glutamicum*.

Authors' contributions
LS, the first author executed majority of the work and wrote the article.SA, SH and KS contributed in the molecular biology aspects of the work while DR involved in the RSM studies. VFW helped in critical reading of manuscript. KMN, the corresponding author who conceived and designed the research and helped to prepare the manuscript. All authors read and approved the manuscript.

Funding
The study is funded by DBT, New Delhi and BMBF, Germany under Indo German collaboration.

Availability of data and materials
All data generated or analysed during this study are included in this published article and its additional files.

Ethics approval and consent to participate
The authors declare that they have no conflict of interest regarding this manuscript. This article doesn't contain any studies performed with animals or humans by any of the authors.

Consent for publication
Not applicable.

Competing interests
The authors declare(s) that they have no competing interests.

Author details
[1] Microbial Processes and Technology Division, CSIR–National Institute for Interdisciplinary Science and Technology (NIIST), Thiruvananthapuram 695019, Kerala, India. [2] Academy of Scientific and Innovative Research (AcSIR), CSIR-National Institute for Interdisciplinary Science and Technology (CSIR-NIIST), Thiruvananthapuram 695019, Kerala, India. [3] Genetics of Prokaryotes, Faculty of Biology & CeBiTec, Bielefeld University, Bielefeld, Germany.

Received: 26 March 2020 Accepted: 4 April 2020
Published online: 15 April 2020

References
Abe S, Takayama K-I, Kinoshita S (1967) Taxonomical studies on glutamic acid-producing bacteria. J Gen Appl Microbiol 13:279–301. https://doi.org/10.2323/jgam.13.279

Baritugo K-A, Kim HT, David Y, Choi J, Hong SH, Jeong KJ, Choi JH, Joo JC, Park SJ (2018) Metabolic engineering of *Corynebacterium glutamicum* for fermentative production of chemicals in biorefinery. Appl Microbiol Biotechnol 102:3915–3937. https://doi.org/10.1007/s00253-018-8896-6

Becker J, Rohles CM, Wittmann C (2018) Metabolically engineered *Corynebacterium glutamicum* for bio-based production of chemicals, fuels, materials, and healthcare products. Metab Eng. https://doi.org/10.1016/j.ymben.2018.07.008

Box GEP, Behnken DW (1960) Some new three level designs for the study of quantitative variables. Technometrics 2:455–475. https://doi.org/10.1080/00401706.1960.10489912

Brüsseler C, Späth A, Sokolowsky S, Marienhagen J (2019) Alone at last!—heterologous expression of a single gene is sufficient for establishing the five-step Weimberg pathway in *Corynebacterium glutamicum*. Metab Eng Commun 9:e00090. https://doi.org/10.1016/j.mec.2019.e00090

Buchert J, Viikari L (1988) The role of xylonolactone in xylonic acid production by Pseudomonas fragi. Appl Microbiol Biotechnol 27:333–336. https://doi.org/10.1007/BF00251763

Buchert J, Puls J, Poutanen K (1988) Comparison of *Pseudomonas fragi* and *Gluconobacter oxydans* for production of xylonic acid from hemicellulose hydrolyzates. Appl Microbiol Biotechnol 28:367–372. https://doi.org/10.1007/BF00268197

Buschke N, Becker J, Schäfer R, Kiefer P, Biedendieck R, Wittmann C (2013) Systems metabolic engineering of xylose-utilizing *Corynebacterium glutamicum* for production of 1,5-diaminopentane. Biotechnol J 8:557–570. https://doi.org/10.1002/biot.201200367

Byong-Wa C, Benita D, Macuch PJ, Debbie W, Charlotte P, Ara J (2006) The development of cement and concrete additive. Appl Biochem Biotechnol 131:645–658. https://doi.org/10.1385/abab:131:1:645

Cabulong RB, Lee W-K, Bañares AB, Ramos KRM, Nisola GM, Valdehuesa KNG, Chung W-J (2018) Engineering *Escherichia coli* for glycolic acid production from D-xylose through the Dahms pathway and glyoxylate bypass. Appl Microbiol Biotechnol 102:2179–2189. https://doi.org/10.1007/s00253-018-8744-8

Cao Y, Xian M, Zou H, Zhang H (2013) Metabolic engineering of *Escherichia coli* for the production of xylonate. PLoS ONE 8:e67305. https://doi.org/10.1371/journal.pone.0067305

Choi JW, Jeon EJ, Jeong KJ (2019) Recent advances in engineering *Corynebacterium glutamicum* for utilization of hemicellulosic biomass. Curr Opin Biotechnol 57:17–24. https://doi.org/10.1016/j.copbio.2018.11.004

Chung CT, Miller RH (1993) Preparation and storage of competent *Escherichia coli* cells. Methods Enzymol 218:621–627. https://doi.org/10.1016/0076-6879(93)18045-E

de Arruda PV, de Cássia Lacerda Brambilla Rodrigu R, da Silva DDV, de Almeida Felipe M (2011) Evaluation of hexose and pentose in pre-cultivation of *Candida guilliermondii* on the key enzymes for xylitol production in sugarcane hemicellulosic hydrolysate. Biodegradation 22:815–822. https://doi.org/10.1007/s10532-010-9397-1

Dhar KS, Wendisch VF, Nampoothiri KM (2016) Engineering of *Corynebacterium glutamicum* for xylitol production from lignocellulosic pentose sugars. J Biotechnol 230:63–71. https://doi.org/10.1016/j.jbiotec.2016.05.011

Gopinath V, Meiswinkel TM, Wendisch VF, Nampoothiri KM (2011) Amino acid production from rice straw and wheat bran hydrolysates by recombinant pentose-utilizing *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 92:985–996. https://doi.org/10.1007/s00253-011-3478-x

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557–580. https://doi.org/10.1016/S0022-2836(83)80284-8

Heider SAE, Wendisch VF (2015) Engineering microbial cell factories: metabolic engineering of *Corynebacterium glutamicum* with a focus on non-natural products. Biotechnol J 10:1170–1184. https://doi.org/10.1002/biot.201400590

Hirasawa T, Shimizu H (2016) Recent advances in amino acid production by microbial cells. Curr Opin Biotechnol 42:133–146. https://doi.org/10.1016/j.copbio.2016.04.017

Ishizaki H, Ihara T, Yoshitake J, Shimamura M, Imai T (1973) D-Xylonic acid production by *Enterobacter cloacae*. J Agric Chem Soc Jpn 47:755–761. https://doi.org/10.1271/nogeikagaku1924.47.755

Johanna B, Liisa V (1988) Oxidative D-xylose metabolism of *Gluconobacter oxydans*. Appl Microbiol Biotechnol 29:375–379

Kawaguchi H, Sasaki M, Vertes AA, Inui M, Yukawa H (2009) Identification and functional analysis of the gene cluster for L-arabinose utilization in *Corynebacterium glutamicum*. Appl Environ Microbiol 75:3419–3429. https://doi.org/10.1128/AEM.02912-08

Keilhauer C, Eggeling L, Sahm H (1993) Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the ilvB-ilvN-ilvC operon. J Bacteriol 175:5595–5603. https://doi.org/10.1128/JB.175.17.5595-5603.1993

Kinoshita S, Udaka S, Shimono M (2004) Studies on the amino acid fermentation Part 1 Production of L-glutamic acid by various microorganisms. J Gen Appl Microbiol 50(6):331–343

Lin T-H, Huang C-F, Guo G-L, Hwang W-S, Huang S-L (2012) Pilot-scale ethanol production from rice straw hydrolysates using xylose-fermenting *Pichia stipitis*. Bioresour Technol 116:314–319. https://doi.org/10.1016/j.biortech.2012.03.089

Liu H, Valdehuesa KNG, Nisola GM, Ramos KRM, Chung W-J (2012) High yield production of D-xylonic acid from D-xylose using engineered *Escherichia coli*. Bioresour Technol 115:244–248. https://doi.org/10.1016/j.biortech.2011.08.065

Matano C, Meiswinkel TM, Wendisch VF (2014) Amino acid production from rice straw hydrolyzates. In: Wheat and rice in disease prevention and health. pp 493–505

Meijnen JP, De Winde JH, Ruijssenaars HJ (2009) Establishment of oxidative D-xylose metabolism in *Pseudomonas putida* S12. Appl Environ Microbiol 75:2784–2791. https://doi.org/10.1128/AEM.02713-08

Meiswinkel TM, Gopinath V, Lindner SN, Nampoothiri KM, Wendisch VF (2013) Accelerated pentose utilization by *Corynebacterium glutamicum* for accelerated production of lysine, glutamate, ornithine and putrescine. Microb Biotechnol 6:131–140. https://doi.org/10.1111/1751-7915.12001

Mussatto SI, Teixeira JA (2010) Lignocellulose as raw material in fermentation processes. Microbial Biotechnol 2:897–907

Nygård Y, Toivari MH, Penttilä M, Ruohonen L, Wiebe MG (2011) Bioconversion of D-xylose to D-xylonate with *Kluyveromyces lactis*. Metab Eng 13:383–391. https://doi.org/10.1016/j.ymben.2011.04.001

Ou MS, Ingram LO, Shanmugam KT (2011) L(+)-Lactic acid production from non-food carbohydrates by thermotolerant Bacillus coagulans. J Ind Microbiol Biotechnol 38:599–605. https://doi.org/10.1007/s10295-010-0796-4

Pérez-García F, Peters-Wendisch P, Wendisch VF (2016) Engineering Corynebacterium glutamicum for fast production of L-lysine and l-pipecolic acid. Appl Microbiol Biotechnol 100:8075–8090. https://doi.org/10.1007/s00253-016-7682-6

Peters-Wendisch PG, Schiel B, Wendisch VF, Katsoulidis E, Möckel B, Sahm H, Eikmanns BJ (2001) Pyruvate carboxylase is a major bottleneck for glutamate and lysine production by *Corynebacterium glutamicum*. J Mol Biol Biotechnol 3(2):295–300

Raganati F, Olivieri G, Götz P, Marzocchella A, Salatino P (2015) Butanol production from hexoses and pentoses by fermentation of *Clostridium acetobutylicum*. Anaerobe 34:146–155. https://doi.org/10.1016/j.anaerobe.2015.05.008

Sakai S, Tsuchida Y, Okino S, Ichihashi O, Kawaguchi H, Watanabe T, Inui M, Yukawa H (2007) Effect of lignocellulose-derived inhibitors on growth of and ethanol production by growth-arrested *Corynebacterium glutamicum* R. Appl Environ Microbiol 73:2349–2353. https://doi.org/10.1128/AEM.02880-06

Sambrook BJ, Maccallum P, Russell D (2006) Molecular cloning: a laboratory manual to order or request additional information. Mol Clon 1:1–3

Stephens C, Christen B, Fuchs T, Sundaram V, Watanabe K, Jenal U (2007) Genetic analysis of a novel pathway for D-xylose metabolism in *Caulobacter crescentus*. J Bacteriol 189:2181–2185. https://doi.org/10.1128/JB.01438-06

Sundar Lekshmi MS, Susmitha A, Soumya MP, Keerthi Sasikumar, Nampoothiri Madhavan K (2019) Bioconversion of D-xylose to D-xylonic acid by *Pseudoduganella danionis*. Indian J Exp Biol 57:821–824

Toivari MH, Ruohonen L, Richard P, Penttilä M, Wiebe MG (2010) *Saccharomyces cerevisiae* engineered to produce *D*-xylonate. Appl Microbiol Biotechnol 88:751–760. https://doi.org/10.1007/s00253-010-2787-9

Toivari MH, Penttil M, Ruohonen L, Wiebe MG, Nyg Y (2011) Bioconversion of D-xylose to D-xylonate with *Kluyveromyces lactis*. Metab Eng 13:383–391. https://doi.org/10.1016/j.ymben.2011.04.001

Toivari M, Nygård Y, Kumpula EP, Vehkomäki ML, Benčina M, Valkonen M, Maaheimo H, Andberg M, Koivula A, Ruohonen L, Penttilä M, Wiebe MG (2012) Metabolic engineering of *Saccharomyces cerevisiae* for bioconversion of D-xylose to D-xylonate. Metab Eng 14:427–436. https://doi.org/10.1016/j.ymben.2012.03.002

van der Rest ME, Lange C, Molenaar D (1999) A heat shock following electroporation induces highly efficient transformation of Corynebacterium glutamicum with xenogeneic plasmid DNA. Appl Microbiol Biotechnol 52:541–545. https://doi.org/10.1007/s002530051557

Wang C, Wei D, Zhang Z, Wang D, Shi J, Kim CH, Jiang B, Han Z, Hao J (2016) Production of xylonic acid by *Klebsiella pneumoniae*. Appl Microbiol Biotechnol 100:10055–10063. https://doi.org/10.1007/s00253-016-7825-9

Wei Niu, Mapitso Molefe N, Frost JW (2003) Microbial synthesis of the energetic material precursor 1,2,4-butanetriol. J Am Chem Soc 125:12998–12999

Wendisch VF, Brito LF, Gil Lopez M, Hennig G, Pfeifenschneider J, Sgobba E, Veldmann KH (2016) The flexible feedstock concept in industrial biotechnology: metabolic engineering of *Escherichia coli*, *Corynebacterium glutamicum*, Pseudomonas, Bacillus and yeast strains for access to alternative carbon sources. J Biotechnol 234:139–157. https://doi.org/10.1016/j.jbiotec.2016.07.022

Wiebe MG, Nygård Y, Oja M, Andberg M, Ruohonen L, Koivula A, Penttilä M, Toivari M (2015) A novel aldose-aldose oxidoreductase for co-production of D-xylonate and xylitol from D-xylose with *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 99:9439–9447. https://doi.org/10.1007/s00253-015-6878-5

Wisselink HW, Toirkens MJ, Wu Q, Pronk JT, van Maris AJA (2009) Novel evolutionary engineering approach for accelerated utilization of glucose, xylose, and arabinose mixtures by engineered *Saccharomyces cerevisiae* strains. Appl Environ Microbiol 75:907–914. https://doi.org/10.1128/AEM.02268-08

Yim SS, Choi JW, Lee SH, Jeon EJ, Chung WJ, Jeong KJ (2017) Engineering of *Corynebacterium glutamicum* for consolidated conversion of hemicellulosic biomass into xylonic acid. Biotechnol J 12:1–9. https://doi.org/10.1002/biot.201700040

Zahoor A, Otten A, Wendisch VF (2014) Metabolic engineering of *Corynebacterium glutamicum* for glycolate production. J Biotechnol 192:366–375. https://doi.org/10.1016/j.jbiotec.2013.12.020

Publisher's Note

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com