

**SORTASE– MEDIATED TRANSPEPTIDATION: THE APPLICATION OF  
SORTASE A AS A TOOL FOR SITE- SPECIFIC MODIFICATION OF  
RECOMBINANT ANTIBODIES**

**by**

**NUR FAEZEE BINTI ISMAIL**

**Thesis submitted in fulfilment of the requirements  
for the degree of  
Master of Science**

**March 2016**

## ACKNOWLEDGEMENTS

All praises and gratitude is due to Allah, the Most Merciful and Compassionate.

The completion of this research and thesis was not an individual effort as I had support from many amazing individuals. First of all, I would like to thank my project supervisor, Dr Lim Theam Soon for his encouragement, guidance and support in all the time for research and writing the thesis. The experience that I have gained here in INFORMM was truly invaluable. His knowledge, constructive criticism, and excellent advice have been of great help to me. He has always been approachable at all the times which help guided me to complete my research with a truly rewarding experience.

I would like to extend my heartfelt gratitude to the members of the Antibody Engineering lab for their endless guidance and support throughout the entire duration of the project. Thank you for your patience and encouragements to me throughout the research. Thanks again to them for providing me with a wonderful laboratory experience and environment.

Not forgetting, many thanks to my parents, Mr. Ismail bin Ibrahim and Mrs Fatimah binti Ishak, my husband, Zulhilmi bin Husni, my family members, and my closest friends for their endless support and motivation. Thank you so much for being there for me, be it virtually or physically.

The financial support from Malaysian Ministry of Science, Technology and Innovation ScienceFund Grant (Grant No. 305/CIPPM/613229), MyMaster and USM graduate research assistant scheme are gratefully acknowledged.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	ii
LIST OF TABLES .....	vii
LIST OF FIGURES .....	ix
LIST OF SYMBOL AND ABBREVIATIONS.....	xiv
ABSTRAK .....	xvii
ABSTRACT .....	xix
CHAPTER 1: INTRODUCTION .....	1
1.1 General Introduction .....	1
1.2 Statement of problems and rationale of study.....	4
1.3 Research Objectives .....	6
CHAPTER 2: LITERATURE REVIEW .....	7
2.1 Transpepidases .....	7
2.2 Sortase A .....	9
2.3 <i>Escherichia coli</i> as a host cell .....	21
2.4 Expression vector .....	24
2.5 Biotinylated protein purification system.....	27
2.6 Histidine tag purification system .....	29
2.7 Enzyme-linked immunosorbent assay.....	31
CHAPTER 3: MATERIALS AND METHODS .....	44
3.1 Materials.....	44
3.2 Methods.....	47
3.2.1 General methods.....	47
3.2.2 General cloning methods .....	51
3.2.3 General protein expression, extraction and purification protocol.....	54
3.2.4 Buffer exchange and concentration protocol for protein .....	56
3.2.5 Determining protein concentration .....	57
3.2.6 Sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS- PAGE).....	58
3.2.7 Western blot .....	60
PART I: SORTASE A .....	61
3.2.8 SrtA Cloning and Expression studies.....	61
PART II: LPXTG motif .....	65

3.2.9 anti Ubi scFv-LPETG expression, extraction and purification studies.....	66
3.2.10 Cloning and expression of anti Ubi scFv-LPETGG in pRSET-B .....	67
3.2.11 Expression, extraction and purification studies of anti Ubi scFv-LPETGG .....	71
PART III: Oligoglycine motif.....	73
3.2.12 Enhanced green fluorescent protein and invertase.....	73
3.2.13 Cloning of oligoglycine in pRSET-BH6 <i>eGFP</i> and pRSET-BH6 <i>invB</i> .	76
3.2.14 Expression, extraction and purification studies of G <sub>3</sub> -eGFP, G <sub>5</sub> -eGFP, and G <sub>5</sub> -invB.....	79
3.2.15 Preparation of antigens.....	80
3.2.16 Western blot analysis of purified product .....	80
3.3 Functionality test of purified SrtA conjugation components .....	81
3.3.1 anti Ubi-scFv functionality test.....	81
3.3.2 eGFP detection limit .....	81
3.3.3 Detection limit of invB .....	82
3.4 SrtA mediated conjugation and its reaction optimization .....	82
3.4.1 Efficiency of motif and temperature of reaction .....	82
3.4.2 Optimization of pH of reaction .....	83
3.4.3 Optimization of CaCl <sub>2</sub> concentration in reaction .....	83
3.4.4 Optimization of SrtA concentration .....	84
3.4.5 Optimization of reactant concentration.....	85
3.5 Detection of anti Ubi scFv-eGFP and anti Ubi scFv-invB using western blot	86
3.6 Scaling up of the conjugation reaction.....	86
3.8 Detection of purified conjugated product using western blot .....	86
3.9 Direct assay of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP and anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	87
3.9.1 Limit of detection of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP.....	87
3.9.2 Limit of detection of anti Ubi scFv-LPETGG <sub>5</sub> -invB.....	87
3.10 Competitive assay of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP and anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	88
3.10.1 Competitive assay of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP .....	88
3.10.2 Competitive assay of anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	88
CHAPTER 4: RESULTS .....	89
4.1 Preparation of Sortase A conjugation components .....	89
4.1.1 Sortase A .....	89
4.1.2 LPXTG motif .....	94

4.1.3 Oligoglycine motif .....	103
4.1.4 Antigen preparation.....	115
4.1.6 Functionality test of anti Ubi-scFv, eGFP and invB .....	118
4.2 Sortase mediated conjugation optimization .....	124
4.2.1 Selecting efficient motif and optimal temperature of reaction.....	124
4.2.2 pH optimization of Sortase A conjugation of G <sub>5</sub> -eGFP, G <sub>5</sub> -invB to anti Ubi scFv-LPETGG .....	129
4.2.3 CaCl <sub>2</sub> concentration optimization .....	131
4.2.4 Optimization of Sortase A concentration.....	133
4.2.5 Optimization concentration of reactants .....	135
4.2.6 Relationship of temperature, incubation time and Sortase A concentration .....	137
4.2.7 Conjugation of Avi-tagged anti Ubi scFv-LPETGG and pentaglycine motif proteins at optimized condition .....	139
4.3 Western blot analysis of conjugated anti Ubi scFv-LPETGG <sub>5</sub> -eGFP and anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	142
4.3.1 Western blot analysis of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP .....	142
4.3.2 Western blot analysis of anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	143
4.4 Purification of conjugated product.....	145
4.4.1 Purification and western blot studies of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP.....	145
4.4.2 Purification and western blot analysis of anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	148
4.5 Direct assay of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP and anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	151
4.5.1 Limit of detection of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP.....	151
4.5.2 Limit of detection of anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	153
4.6 Competitive assay of conjugated anti Ubi scFv-LPETGG <sub>5</sub> -eGFP and anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	155
4.6.1 Competitive assay of conjugated anti Ubi scFv-LPETGG <sub>5</sub> -eGFP .....	155
4.6.2 Competitive assay of anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	158
CHAPTER 5: DISCUSSION.....	160
5.1 Preparation of Sortase A conjugation components .....	160
5.1.1 Sortase A .....	160
5.1.2 LPXTG motif .....	163
5.1.3 Oligoglycine motif .....	167
5.1.4 Antigen preparation.....	170
5.1.5 Western blot analysis of purified Srt A components .....	170

5.1.6	Functionality test of anti Ubi-scFv, eGFP and invB .....	171
5.2	Sortase mediated conjugation optimization .....	173
5.2.1	Selecting efficient motif and optimal temperature of reaction.....	173
5.2.2	pH of reaction optimization .....	175
5.2.3	CaCl <sub>2</sub> optimization .....	176
5.2.4	Sortase A concentration optimization .....	177
5.2.5	Optimizing concentration of reactants .....	178
5.2.6	Relationship between temperature, incubation time and SrtA concentration .....	179
5.2.7	Conjugation of Avi-tagged anti Ubi scFv-LPETGG and pentaglycine motif proteins at optimized condition .....	180
5.3	Western blot analysis of conjugated anti Ubi scFv-LPETGG <sub>5</sub> -eGFP and anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	181
5.4	Purification of conjugated product.....	182
5.5	Western blot analysis of purified anti Ubi scFv LPETGG <sub>5</sub> -eGFP and anti Ubi scFv LPETGG <sub>5</sub> -invB .....	184
5.6	Overview of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP and anti Ubi scFv-LPETGG <sub>5</sub> - invB assay development.....	186
5.6.1	Direct assay of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP.....	189
5.6.2	Limit of detection of anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	191
5.7	Competitive assay of conjugated anti Ubi scFv-LPETGG <sub>5</sub> -eGFP and anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	192
5.7.1	Competitive assay of conjugated anti Ubi scFv-LPETGG <sub>5</sub> -eGFP .....	192
5.7.2	Competitive assay of anti Ubi scFv-LPETGG <sub>5</sub> -invB .....	193
CHAPTER 6:	CONCLUSION.....	195
6.1	Conclusion .....	195
6.2	Suggestion for future studies.....	199
REFERENCES	.....	201
APPENDICES	.....	218
PUBLICATIONS	.....	230

## LIST OF TABLES

<b>Table</b>		<b>Page</b>
<b>3.1</b>	Bacteria strains used in this study	45
<b>3.2</b>	Plasmid used in this study	46
<b>3.3</b>	Double digestion of vector and insert	51
<b>3.4</b>	DNA ligation of digested vector and digested insert	52
<b>3.5</b>	PCR mixture of <i>SrtA</i> amplification	62
<b>3.6</b>	Primers used to amplify <i>SrtA</i> gene from pRSET- <i>SrtA</i>	62
<b>3.7</b>	PCR condition of <i>SrtA</i> amplification	63
<b>3.8</b>	PCR mixture of anti Ubi scFv LPETGG amplification	68
<b>3.9</b>	Primers used for anti Ubi scFv-LPETGG amplification	68
<b>3.10</b>	PCR mixture of Avi-tag amplification	70
<b>3.11</b>	Primers used in Avi-tag amplification	70
<b>3.12</b>	PCR condition of Avi-tag amplification	70
<b>3.13</b>	PCR mixture of <i>eGFP</i> and <i>invB</i>	74
<b>3.14</b>	Primers used in the amplification of <i>eGFP</i> and <i>invB</i>	74
<b>3.15</b>	PCR condition of <i>eGFP</i> and <i>invB</i>	75
<b>3.16</b>	Oligo sequences used in generating double stranded DNA of G <sub>3</sub> H6X	76
<b>3.17</b>	Filling-in mixture of triglycine motif	77
<b>3.18</b>	Oligo sequence used in generating double stranded DNA of G <sub>5</sub> H6X	77
<b>3.19</b>	Filling in reaction of G <sub>5</sub> H6X	78
<b>3.20</b>	Filling in condition of dsDNA G <sub>5</sub> H6X generation	78

<b>3.21</b>	Reaction mixture of conjugation reaction	83
<b>3.22</b>	Reaction mixture of different CaCl <sub>2</sub> concentration	84
<b>3.23</b>	Reaction mixture of different SrtA concentration	84
<b>3.24</b>	Reaction mixture of reactant optimization	85



## LIST OF FIGURES

Figure		Page
1.1	Overview of the study carried out	6
2.1	Ribbon structure of Sortase A	11
2.2	Sortase A reaction mechanism	17
2.3	Schematic diagram of detection of protein biomarker and small molecular toxins using antibody-conjugated magnetic beads with invertase by personal glucose meter (PGM)	42
4.1	Agarose gel electrophoresis of amplified <i>SrtA</i> and its cloning	90
4.2	Sequencing result of pRSET-B <i>SrtA</i> and BLAST result.	91
4.3 a	Overexpression study of Sortase A	93
4.3 b	SDS-PAGE analysis of His-tag purification of Sortase A	93
4.4 a	SDS-PAGE analysis of purification of periplasmic extract of BL21(DE3) pQDKG2 anti Ubi scFv-LPETG	95
4.4 b	SDS-PAGE analysis of purification of cytoplasmic extract of BL21 (DE3) pQDKG2 anti Ubi scFv-LPETG	95
4.5	Agarose gel electrophoresis of anti Ubi scFv-LPETGG and its cloning	97
4.6	Sequencing result of pRSET anti Ubi scFv-LPETGG.	98
4.7a	Agarose gel electrophoresis of Avi-tag sequence from pRSET-BH6	99
4.7 b	Agarose gel electrophoresis of digested pRSET Avi-tag anti Ubi scFv LPETGG at <i>NdeI</i> and <i>NcoI</i>	99
4.8	Sequencing result of pRSEt Avi-tag anti Ubi scFv-LPETGG	100

<b>4.9</b>	Overexpression study of anti Ubi scFv-LPETGG	101
<b>4.10 a</b>	SDS-PAGE analysis of His-tag purification of anti Ubi scFv-LPETGG containing His-tag	102
<b>4.10 b</b>	SDS-PAGE analysis of biotinylated protein purification of anti Ubi scFv-LPETGG containing Avitag	103
<b>4.11</b>	Agarose gel electrophoresis of amplification of <i>eGFP</i> and its cloning	104
<b>4.12</b>	Sequencing and BLAST results of pRSET-BH6 <i>eGFP</i>	105
<b>4.13 a</b>	SDS-PAGE analysis of overexpression study of eGFP	106
<b>4.13 b</b>	SDS-PAGE analysis of His-tag purification of eGFP	107
<b>4.14</b>	Agarose gel electrophoresis of <i>invB</i> and its cloning	108
<b>4.15</b>	Sequencing result of <i>invB</i> and its BLAST result	109
<b>4.16 a</b>	SDS-PAGE analysis of overexpression study of invertase (InvB)	110
<b>4.16 b</b>	SDS-PAGE analysis of His-tag purification of invertase ( <i>invB</i> )	111
<b>4.17</b>	Agarose gel electrophoresis of generation of G <sub>3</sub> H6X	112
<b>4.18</b>	Sequencing result of G <sub>3</sub> H6X analysed using ContigExpress	113
<b>4.19</b>	Agarose gel electrophoresis of generation of G <sub>5</sub> H6X	114
<b>4.20 i</b>	Sequencing result of pRSET G <sub>5</sub> - <i>eGFP</i> analysed using ContigExpress	113
<b>4.20 ii</b>	Sequencing result of pRSET G <sub>5</sub> - <i>invB</i> analysed using ContigExpress	115
<b>4.21</b>	SDS-PAGE analysis of His-tagged purified ubiquitin	116
<b>4.22 a</b>	Western blot analysis of purified proteins against anti-His	117

<b>4.22 b</b>	Western blot analysis of anti Ubi scFv.	118
<b>4.23</b>	Graph of OD reading at 405nm versus samples	119
<b>4.24</b>	Determination of eGFP emission peak	121
<b>4.25 i</b>	Fluorescent reading of different eGFP amounts	122
<b>4.25 ii</b>	Analysis of detection limit of eGFP at 510nm eGFP emission wavelength	122
<b>4.26</b>	Analysis of detection limit of invB in three hour incubation	124
<b>4.27 a</b>	Optimization of scFv-eGFP conjugation with different combination of motifs at a variation of temperature	127
<b>4.27 b</b>	Optimization of scFv-invB conjugation with different combination of motifs at a variation of temperature	128
<b>4.28 i</b>	pH optimization of anti Ubi scFv-LPETGG and G <sub>5</sub> H6X-eGFP conjugation via sortase A reaction	130
<b>4.28 ii</b>	pH optimization of anti Ubi scFv-LPETGG and G <sub>5</sub> -invB conjugation via sortase A reaction	130
<b>4.29</b>	SDS-PAGE analysis of optimization of CaCl <sub>2</sub> in conjugation of (i) anti Ubi scFv-LPETGG and G <sub>5</sub> -eGFP (ii) anti Ubi scFv-LPETGG and G <sub>5</sub> -invB	132
<b>4.30 i</b>	SDS-PAGE analysis of optimization of sortase A concentration in anti Ubi scFv-LPETGG and G <sub>5</sub> -eGFP conjugation	134
<b>4.30 ii</b>	SDS-PAGE analysis of optimization of Sortase A concentration in conjugation of anti Ubi scFv-LPETGG and G <sub>5</sub> -invB	134
<b>4.31 a</b>	SDS-PAGE analysis of optimization of reactant	136

	concentration for anti Ubi scFv-LPETGG and G <sub>5</sub> -eGFP	
<b>4.31 b</b>	SDS-PAGE analysis of anti Ubi scFv-LPETGG and G <sub>5</sub> -invB	136
<b>4.32 a</b>	SDS-PAGE analysis of simultaneous effects of temperature, incubation time and Sortase A concentration towards anti Ubi scFv-LPETGG and G <sub>5</sub> -eGFP conjugation.	138
<b>4.32 b i</b>	SDS PAGE analysis of simultaneous effects of temperature, incubation time and sortase A concentration towards anti Ubi scFv-LPETGG and G <sub>5</sub> -invB conjugation at 4°C	140
<b>4.32 b ii</b>	SDS-PAGE analysis of simultaneous effects of temperature, incubation time and Sortase A concentration towards anti Ubi scFv-LPETGG and G <sub>5</sub> -invB at 37°C	140
<b>4.33 i</b>	SDS-PAGE analysis of anti Ubi scFv-LPETGG and G <sub>5</sub> -eGFP	141
<b>4.33 ii</b>	SDS PAGE analysis of anti Ubi scFv-LPETGG and G <sub>5</sub> -invB	141
<b>4.34</b>	Western blot analysis of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP	143
<b>4.35</b>	Western blots analysis of anti Ubi scFv-LPETGG <sub>5</sub> -invB	144
<b>4.36 a</b>	SDS-PAGE analysis of His-tag purification of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP	146
<b>4.36 b</b>	SDS-PAGE analysis of anti Ubi scFv-LPETG <sub>5</sub> -eGFP Avi-tag purification	147
<b>4.36 c</b>	Western blots analysis of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP	147
<b>4.37 a</b>	SDS-PAGE analysis of His-tag purification of anti Ubi scFv-LPETG <sub>5</sub> -invB	150
<b>4.37 b</b>	SDS-PAGE analysis of anti Ubi scFv-LPETG <sub>5</sub> -invB Avi-tag purification	150

<b>4.37 c</b>	Western blot analysis of conjugated anti Ubi scFv-LPETGG <sub>5</sub> -invB.	151
<b>4.38 a</b>	Fluorescent reading of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP direct assay	153
<b>4.38 b</b>	Fluorescent intensity(a.u) at 510nm versus ubiquitin concentration	153
<b>4.39</b>	Reading of glucose concentration versus ubiquitin amount of direct assay of anti Ubi scFv-LPETGG <sub>5</sub> -invB	154
<b>4.40 a</b>	Fluorescent reading of anti Ubi scFv-LPETGG <sub>5</sub> -eGFP competitive assay	157
<b>4.40 b</b>	Fluorescent intensity (a.u) versus anti Ubi scFv-LPETGG amount	157
<b>4.41</b>	Glucose concentrations reading at an hour versus anti Ubi scFv-LPETGG amount	159
<b>5.1</b>	Overview of antibody-eGFP conjugation via SrtA mechanism and its application on the immunoassay platform	187
<b>5.2</b>	Overview of the antibody-invB conjugation via SrtA mechanism and its application of immunoassay platform	189

## LIST OF SYMBOL AND ABBREVIATIONS

	<b>Description</b>	<b>Abbreviations</b>
1	Ampicilin	Amp
2	Base pair	bp
3	Basic Local Alignment Search Tool	BLAST
4	Bovine serum albumin	BSA
5	Coefficient of correlation	$R^2$
6	Degree Celcius	$^{\circ}\text{C}$
7	Deoxyribonucleic acid	DNA
8	Deoxynucleotide phosphate	dNTPs
9	Double strans deoxyribonucleic acid	dsDNA
10	Enhanced green fluorescent protein	eGFP
11	Ethylenediamine tetraacetic acid	EDTA
12	Femto gram	fg
13	Gram	g
14	Hours	h
15	Isopropyl $\beta$ -D-1-thiogalactopyranoside	IPTG
16	Kilobase	kb
17	Liter	L

18	Maximum	max
19	Microgram(s)	μg
20	Microliter(s)	μL
21	Milligram(s)	mg
22	Mililitre(s)	mL
23	Milimolar(s)	mM
24	milimole	mmol
25	Minutes(s)	min
26	Nanogram(s)	ng
27	Nanometre(s)	nm
28	Nanomolar	nM
29	Percentage	%
30	Personal glucose meter	PGM
31	Phosphate-buffered saline	PBS
32	Pico gram	pg
33	Pico mole	pmol
34	Polymerase chain reaction	PCR
35	Revolutions per minute	rpm
36	Relative centrifugal force	xg

37	Room temperature	rpm
38	Second(s)	s
39	Single strand deoxyribonucleic acids	ssDNA
40	Sodium dodecyl sulphate electrophoresis	SDS-PAGE
41	Unit	U
42	Ultra violet	UV
43	Volume	vol
44	Volts	V



# **TRANSPEPTIDASI BERANTARAKAN SORTASE: APLIKASI SORTASE A SEBAGAI ALAT MODIFIKASI SETEMPAT UNTUK ANTIBODI REKOMBINAN**

## **ABSTRAK**

Diagnostik ‘point-of-care’ (POC) untuk kawasan terpencil telah menjadi cabaran besar untuk memerangi penyakit berjangkit di seluruh dunia. Pembangunan kaedah untuk membantu proses pembangunan diagnostik POC baru adalah penting untuk memperbaiki senario penjagaan kesihatan di seluruh dunia. Format utama POC asai adalah pengesanan interaksi antibodi- antigen . Untuk mencapai matlamat ini, asai yang mampu melaporkan interaksi antara antibodi dan antigen adalah diperlukan. Kebiasaannya, asai dibangunkan dengan merangkaikan antibodi-antibodi kepada satu pelapor yang spesifik seperti enzim atau protein. Projek ini menerangkan penggunaan transpeptidasi berantarakkan sortase sebagai satu cara untuk merangkaikan antibodi rekombinan untuk sepasang protein pelapor. Enzim Sortase adalah unik dengan mengenali motif yang mempunyai urutan yang khusus untuk membenarkan tidak balas rangkaian protein-protein secara setempat. Disebabkan sifatnya yang spesifik, ianya mempunyai kelebihan untuk digunakan untuk biokonjugat terutamanya dalam membangunkan platform asai diagnostik. *SrtA*(0.546kb) telah diekstrak daripada *Staphyococcus aureus* dan diklon dalam sistem pRSET dibawah kawalan T7 promoter. Satu set motif telah digabungkan, diklon dan diekspres untuk memilih motif paling efisien pada satu kondisi yang dioptimumkan. Satu eksperimen pembuktian-konsep telah dijalankan menggunakan anti-ubiquitin scFv (0.72kb) sebagai protein motif penderma (LPXTG motif). Disamping itu, protein hijau pendarfluor (eGFP) (0.714kb) dan invertase (1.236kb)

telah dipilih sebagai protein motif penerima (motif glisin) yang berfungsi sebagai sistem pelapor. Strategi penulenan tag berganda telah dibangun untuk meningkatkan hasil penulenan produk konjugat. Pembangunan dua platform yang berbeza menggunakan dua protein pelapor yang berbeza telah digunakan. Kajian awal telah dijalankan dengan menggunakan eGFP untuk aplikasi pada platform immunoassay pendarfluor. Selepas mengoptimumkan keadaan ideal, satu proses konjugasi kedua telah dibangun untuk enzim invertase. Konjugasi untuk invertase akan membolehkan pembangunan asai antibodi-antigen yang pesat menggunakan meter glukosa peribadi (PGM). Invertase konjugat akan berfungsi untuk menghasilkan bacaan di PGM dengan kehadiran glukosa. Proses konjugasi berantaraan sortase di platform PGM membenarkan pembangunan asai POC yang mudah dan cepat. Ini memberi alternatif yang menarik untuk pembangunan antibodi - antigen asai yang menggunakan PGM untuk aplikasi POC.

# **SORTASE-MEDIATED TRANSPEPTIDATION: THE APPLICATION OF SORTASE A AS A TOOL FOR SITE-SPECIFIC MODIFICATION OF RECOMBINANT ANTIBODIES**

## **ABSTRACT**

Point-of-care (POC) diagnostics for resource deprived regions remains a major challenge to combat infectious diseases worldwide. Development of new methods to aid the development process of new POC diagnostics is essential to improve the healthcare scenario worldwide. The major format of POC assay is the detection of antibody-antigen interaction. To achieve this, an assay is required to be capable of reporting the interaction between the antibody and antigen. In most common assays, these detection are achieved by conjugating antibodies to specific reporter molecules such as enzymes or proteins. The work describes the application of sortase mediated transpeptidation as a means to conjugate recombinant antibodies to a pair of reporter proteins. Sortase enzymes are unique in a way that they recognized a specific sequence motif for reaction allowing a site-specific conjugation of proteins. Due to this specificity, they are very beneficial to be used for bioconjugation mainly in developing diagnostic assay platforms. *SrtA* (0.546kb) was isolated from *Staphylococcus aureus* and cloned inside pRSET system under a tight control of T7 promoter. A set of motif were also fused, cloned and expressed to select the most efficient motif at an optimized condition. A proof-of-concept experiment was conducted by using the anti-ubiquitin scFv(0.72kb) as a donor motif (LPXTG motif) protein. Meanwhile, enhanced green fluorescent protein (eGFP)(0.714kb) and invertase(1.236kb) were chosen as the receiver motif (glycine motif) protein functioning as the reporter system. A double tag purification strategy was developed

to improve purification yield of the conjugated product. Development of two different platforms utilizing two different reporter proteins was used. Initial investigation was carried out using eGFP for application on a fluorescent immunoassay platform. After optimizing the ideal conditions, a second conjugation process was developed for invertase enzyme. The conjugation to invertase allowed for rapid development of antibody-antigen assays detectable using a personal glucose meter (PGM). The function of conjugated invertase was to generate a readout on the PGM in the presence of glucose. The sortase mediated conjugation process allows for an easy and rapid development of recombinant antibody derived POC assays on the PGM platform. This would provide an interesting alternative for the development of antibody-antigen assays utilizing a PGM for POC applications.

## CHAPTER 1: INTRODUCTION

### 1.1 General Introduction

The evolution of diagnostics and treatment of diseases has become easier with the expansion of molecular biology approaches. Immunoassays are widely used for preliminary detection of the presence of diseases via antigen-antibody interaction in order to initiate the faster treatments. To bring these complexes towards detectable ranges, commercial antibodies against diseases are commonly fused to a detection module. The signals from these modules are taken to determine the presence of that particular disease biomarker. In a simple explanation, immunoassay is made up of two vital components; the binder and the reporter. The main characteristic of the reporter is their ability to generate signals and how high their sensitivity is so that the complex is well represented. Meanwhile, antibodies in various formats are used to bind the specific antigens with good specificity. As immunoassay is a simple and rapid test, antibody tagging is the most practical method used to develop molecular based diagnostic (H. T. Ta *et al.*, 2011).

Antibody tagging is generally developed by two methods, either by bioconjugation or chemical modification. Until now, chemical modifications are universally used as the tagging reaction. In this format, reactive functional groups are commonly utilized (Yin Zhang *et al.*, 2011). There are three types of conventional chemical modifications that are actively used; NHS ester-maleimide-mediated conjugations with heterobifunctional cross linker containing both amine-reactive NHS ester and sulfhydryl-reactive maleimide (Dai *et al.*, 2013), glutaraldehyde mediated conjugation with stable secondary amine linkages between two proteins (Tresca *et al.*, 1995) , and reductive amination-mediated conjugation where aldehyde groups are

conjugated with amine- or hydrazide-containing molecules by reductive amination (Carmenate *et al.*, 2004). These three methods are the cornerstone of many bioconjugation processes for diagnostic applications. In commercialized diagnostic kits, antibodies used are often conjugated to horseradish peroxidase (HRP) or alkaline phosphatase (AP) which catalyses the substrates such as 3,3',5,5'-tetramethylbenzidine(TMB) and 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) to result in colour formation which in turn is translated as signals readouts (Grandke *et al.*, 2013). The conjugation of the kit components relies on the three methods mentioned earlier. Amine or carboxylate groups are generally involved as the cross linking reagent. Although these methods are widely used, the major downfall is that it occurs in a random orientation and thus increases the possibility of blocking the antigen binding sites on an antibody molecule. Besides, variation from batch to batch production and low recovery yields makes it not cost effective.

As an alternative to the chemical modifications, several approaches have been highlighted to make the production of recombinant antibodies faster and easier (Hu *et al.*, 2013). Due to the increasing interest in recombinant DNA technology, modifications are easily introduced at genetic level thus allowing various genetic fusions to antibodies. Bacterial expression system is commonly favoured since this system is easy to be utilized and at the same time gave high protein yields. However, this method is limited only to protein-protein conjugation and does not allow conjugation to surfaces and particles. As the conjugation of antibody to nanoparticles is greatly improved over the years, this has allowed for antibody conjugated particles to be widely used for therapy and diagnostics (Omidfar *et al.*, 2013).

The other alternative that is constantly being studied is the application of biological enzyme in antibody tagging process. These biological enzymes are usually site-specific and are able to initiate modification on selected functional groups available thus generating new bonds between two molecules. The advantages of this processes includes completely different proteins to be fused together via site directed coupling and polymeric protein assembling without disrupting the function of the fusion components. These type of protein fusion are often affected by various factor for example temperatures, pH, oxidizing condition and enzyme catalysis (Gerrard, 2002). Groups of enzymes catalysing the fusion include transferases, hydrolases and oxidoreductases. In this study, Sortase A from the transpeptidase family will be used to develop an immunoassay platform using this biological enzymatic reaction.

The developed approach was applied to develop a diagnostic test mainly for point-of-care (POC) application. To this effect, the personal glucose meter (PGM) being easy to use and readily available was demonstrated to meet this aim. This type of electrochemical biosensor has been widely studied to detect other molecules besides blood gas in recent years. Looking forward to the future of simplified test assays, PGM offers an easy universal in-home detection platform for the detection of multiple diseases. Therefore, to develop a detection system using PGM, invertase was used as the key reporter enzyme. Invertase is an enzyme which catalyses the conversion of sucrose to glucose. The PGM detection system is a strict process where it only detects the same conformation of glucose so false positive results are reduced. Therefore, conjugation of antibodies to invertase via sortase A mediated conjugation mechanism was investigated to ensure the conjugation process did not alter the functionality of the enzyme as well as the antibody to function on the PGM for detection applications.

## **1.2 Statement of problems and rationale of study**

Since conjugation of biomolecules via chemical methods requires chemical modification to the proteins, inactivation of the conjugates is often the common concern. Being a site specific enzyme, Sortase A is a good candidate for “friendly” conjugation process as it does not modify the structure of the protein involved during the fusion reaction. Besides, it is known that the conjugation aided by a biological process is less harmful compared to chemical conjugation as the process involved is usually a replication of natural mechanism. Unlike chemical methods, the highly selective nature of Sortase A ensures that attachment of proteins is not done randomly but instead targets a specific amino acid motif. Another concern with the use of chemical methods is the requirement of extreme conditions that could affect the folding of the proteins. Sortase A conjugation has been reported to work best at mild conditions whereby disturbances to the integrity of the protein structure can be minimized. The conservation of the protein structure is important to ensure functionality of the protein. This will have a direct effect on the quality of assay developed with the conjugates. Therefore the investigation of alternative processes for bioconjugation of recombinant antibodies could aid to improve the generation of fused antibodies for use on diagnostic platforms. Taking into consideration the nature of Sortase A mediated transpeptidation, it could potentially be an interesting alternative to establish an-in house production of immunoassay components.

For the detection process of immuno-complexes, a sensitive detection module is required in order to allow sensitive detection of the complex. Fluorescence proteins such as the enhanced green fluorescent protein (eGFP) and invertase are suitable for diagnostic applications due to their respective functions. Utilizing eGFP, a fluorescent based detection of antibody-antigen complexes can be accomplished.



This could provide an interesting alternative to the conventional chemically conjugated HRP and AP antibodies. eGFP is often used as biosensor due to its fluorescence properties which does not require activation with a chemical substrate to initiate a readout.

Another biomolecule of interest as a reporter system is invertase. Invertase has gained a lot of interest due to its ability to induce signal production via its glucose conversion activity. The proposed use of it became increasingly attractive as its activity can be detected using a commercialized personal glucose meter (PGM) as reported by several groups. PGM is conventionally used to monitor blood glucose level in diabetic patients. Using this idea, we improvised the use of it by detecting glucose level inside an antibody-antigen reaction, where invertase would function as the key component in producing readout. Compared to HRP and AP, invertase utilizes sucrose as its sole substrates to release glucose and fructose. Utilizing sortase A reaction to tag invertase to antibodies would provide an easier and safer conjugation process as it does not involve modifications to its structure hence preserving the functionality of both the antibody and invertase. In addition to eGFP, invertase is another great tool to be used for antibody-antigen detection. It is hoped that by utilizing the Sortase A mediated transpeptidation to conjugate antibodies to various reporter biomolecules, a new avenue for diagnostic detection system for antibody-antigen reaction can be realised.

A streamlined process and vector series would allow easy and rapid substitution of disease specific recombinant antibodies or antigens for conjugation with invertase. This would reduce the development process of new disease sensors for POC application.

### 1.3 Research Objectives

This research was carried out to develop an immunoassay via Sortase A mediated transpeptidation. The approaches used consist of:

- 1) Designation and generation of a family of sortase acceptor and donor vector for streamline antibody-protein ligation.
- 2) Cloning, expression and purification of Sortase A and motifs protein
- 3) Development of sortase-mediated antibody-protein ligation
- 4) Evaluation of the functionality of sortase-mediated antibody for immunoassay

The overview of this study was sketched out as shown in **Figure 1.1**

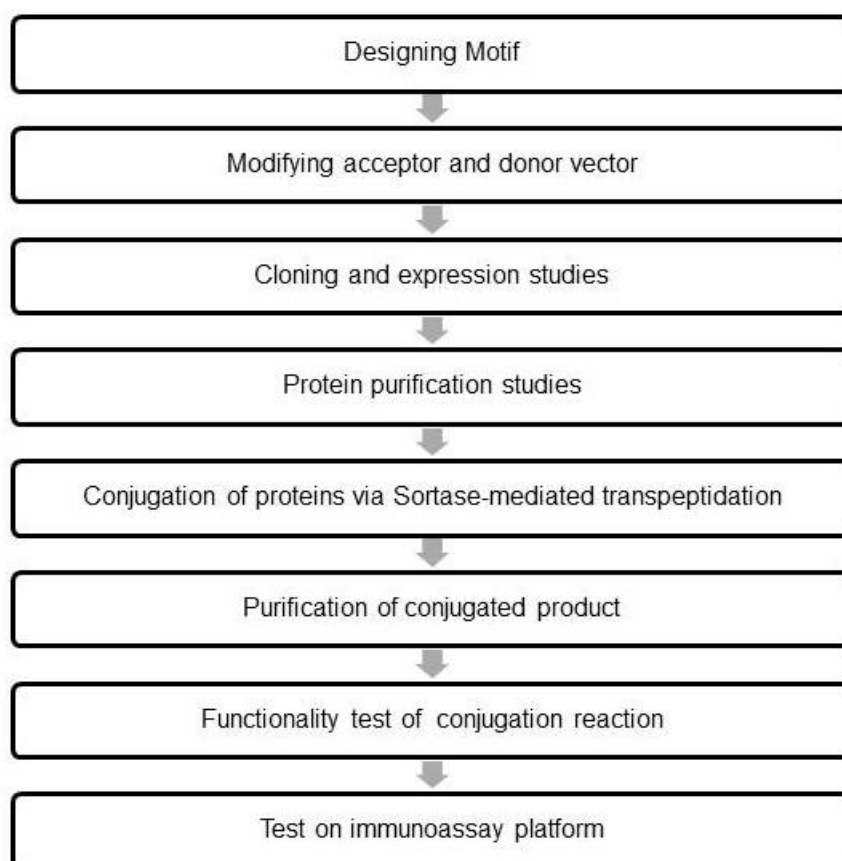


Figure 1.1 : Overview of the study carried out

## CHAPTER 2: LITERATURE REVIEW

### 2.1 Transpeptidases

Transpeptidation is a process of transferring an amino acid from a peptide chain to another via enzyme catalysis or of a peptide chain itself. Sortase family members are well known transpeptidases which actively carry the transpeptidation in Gram positive bacteria like *Staphylococcus aureus*.

#### 2.1.1 *Staphylococcus aureus subsp. aureus* (ATCC®25923™)

*Staphylococcus* sp. is grouped under Micrococcaceae family. *Staphylococcus aureus* is a well-known pathogen causing various infections to human and animal. They usually reside on human skin, nails and nostril areas, infiltrating the inner layer of the host barrier and causes formation of lesions with pus all over the systems (L. A. Marraffini *et al.*, 2006). Classified under Gram positive bacteria, like other members, its pili generation was also made by crosslinking individual units of pilin monomers and finally joining the polymers to the cell wall peptidoglycan via sortase enzymatic reaction (Mandlik *et al.*, 2008). MSCRAMMs (microbial surface component recognizing adhesive matrix molecules) which are involved in the adherence of the bacteria to host cell during infection are also revealed to use sortase transpeptidation system for its binding to the cell wall peptidoglycan (Foster & Höök, 1998). Since sortases was revealed to be crucial for Gram positive bacteria life process, the interest of studies on its system increased.

### 2.1.2 Sortase superfamily

Sortase was discovered in Gram positive bacteria with a variation of motifs. Generally, it is classified into three isoforms, which are SrtA, SrtB and SrtC. All the three isomers of sortases have different recognition motif sequence and functions. SrtA recognizes the LPXTG amino acid motif whereas, SrtB recognize NPQTN and SrtC recognize QVPTG motif (Clancy *et al.*, 2010). SrtA is present in all Gram positive bacteria, meanwhile, SrtB and SrtC are discovered in selected species only. SrtA is mainly responsible for the pathogenesis of the bacteria where it anchors the protein with its signalling motif to lipid II which most of the anchored proteins are the virulent factors of the bacteria itself, meanwhile, SrtB is in charge of the iron deprivation response and SrtC act as the pilin polymerase in the bacteria.

In a study conducted by Weiss *et al.* (2004), SrtA knockout *Staphylococcus aureus* was able to greatly reduce the virulence factor to acute bacterial arthritis in several animal models. The absence of SrtB also gave impact to the adherence of pathogen to the host cell, however, the influence is lower compared to SrtA. SrtB plays a bigger role during later stages of infection where the environmental sources of iron become scarce (Jonsson *et al.*, 2002; Manetti *et al.*, 2007). In a deficient SrtA and SrtC pili-producing species, the setback are impairment of biofilm formation and halting the infection process (Trotonda *et al.*, 2008).

Due to the importance of sortase in Gram positive bacteria infectivity, there has been increasing interest on making sortase an anti-infective agent. This is because in sortase knockout strains, the growth rates of the strains were never impaired by the knockout when compared with the wild-type strain. More importantly, the modification resulted mainly in the loss of virulence only.

## 2.2 Sortase A

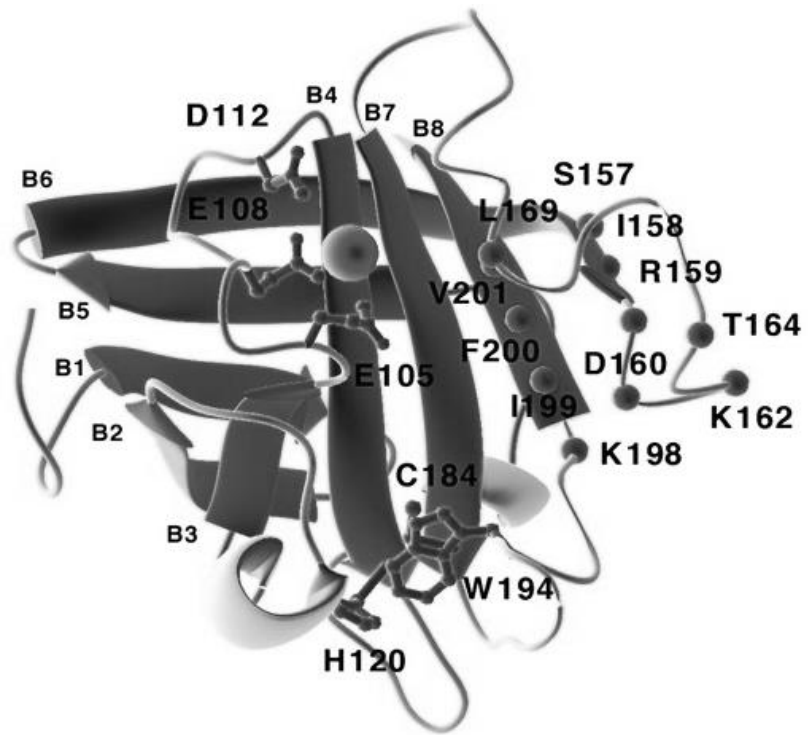
### 2.2.1 Background

Sortase A is known as a housekeeping enzyme to the Gram positive bacteria and exists in almost all species of pathogenic bacteria. As an important enzyme catalysing the attachment of virulent factors to the bacterial peptidoglycan and generation of pili, the deletion of it generally causes the bacteria to lose its pathogenicity (Jonsson *et al.*, 2002; Manetti *et al.*, 2007; Nallapareddy *et al.*, 2011; Sabet *et al.*, 2005). To the realization of this phenomenon, it sparked a wave of study on Sortase A effects in gene therapy (Maresso & Schneewind, 2008).

In its native environment, Sortase A catalyses the covalent linkage between MSCRAMMs (microbial surface component recognizing adhesive matrix molecules) to the lipid II on the peptidoglycan layer, in the case of *Staphylococcus aureus*, is the pentapeptide sequence which acts as the cell wall sorting signal(CWSS) (Clancy *et al.*, 2010). Examples of MSCRAMMs are Protein A and fibrinogen-binding protein (ClfA). These MSCRAMMs which are commonly known as surface proteins are important for infectivity and pathogenicity of Gram positive bacteria where they promote the bacterial adhesion to the host and evade the immunity system (L. A. Marraffini *et al.*, 2006). Protein A (spA) takes a role of binding to the Fc region of an immunoglobulin (IgG) and Fab regions of B-cell receptor. These phenomenon causes the blockage of opsonophagocytosis and death of B-cell *in vitro* (Kobayashi & DeLeo, 2013). Meanwhile, fibrinogen-binding MSCRAMMs (ClfA) are used to facilitate the attachment of bacteria to the plasma clots. Mutations of these MSCRAMMs causes the bacteria to reduce the virulence tremendously compared to the wild type strain (Foster & Höök, 1998). Therefore, Sortase A plays an important role in preserving the virulence of these bacteria.

### 2.2.2 Sortase A structure

Clubb and co-worker reported on the development of Sortase A structure in 2001 (Figure 2.1). Sortase A structure from *Staphylococcus aureus* consist of a 206-residue enzyme was divided into three domain with a stretch of non-polar amino acids signal peptide at its N-terminal (1-24 amino acid residues) which functions as the catalyst for its insertion in the cytoplasm (Sarkis K. Mazmanian *et al.*, 2000) followed by a linker of transmembrane domain and catalytic domain (25-59 amino acid residue) and a catalytic domain (60-206 amino acid residue) (J. Zhu, 2010). The deletion of the first 59 amino acid group of the enzyme expressed in *E.coli* expression system resulted in production of a fully functional and soluble polypeptide in the cytoplasm of *E.coli* (Ilangoan *et al.*, 2001; Hung Ton-That *et al.*, 1999). The affinity-purified sortase, labelled as SrtA<sub>ΔN</sub> is responsible for cleaving the LPXTG motif between threonine and glycine, transpeptidation and peptidoglycan substrates. From the sortase A structure elucidated by NMR spectroscopy and x-ray crystallography, its backbone consists of eight stranded beta-barrel with random loops. These random loops termed as  $\beta 2/\beta 3$ ,  $\beta 3/\beta 4$ ,  $\beta 6/\beta 7$  and  $\beta 7/\beta 8$  strands are highly mobile and unstable. A conserved cysteine residue at position 184 resides in the  $\beta 6/\beta 7$  loops (Zong *et al.*, 2004). As a sortase class I member, sortase A meets the standard by containing the signature active site of the family (LXTC) (S. K. Mazmanian *et al.*, 2001). This residue is believed to take part as the active site on stops the activity of sortase A *in vitro* and *in vivo* (Hung Ton-That *et al.*, 2000). Other evidence is that when wild-type sortase was treated with hydroxylamine, the acyl-intermediate was resolved thus releasing the surface protein together with the threonine hydroxamate at the C-terminal of the protein (Hung Ton-That *et al.*, 1999). Besides, when methyl methanethiosulfonates was added, the activity of sortase A is



**Figure 2.1** Ribbon structure of Sortase A(adopted from (Ilangovan *et al.*, 2001))

immediately stopped (Hung Ton-That & Schneewind, 1999). With the addition of DTT, the activity was restored. DTT is a reducing agent which generates the sulfhydryl active site(R-SH) important for reactivation (Hung Ton-That *et al.*, 1999). The substitution of cysteine at position 184 to alanine completely abolishes the activity of Sortase A.

Together with Cys<sup>184</sup>, His<sup>120</sup> and Arg<sup>197</sup> are directly involved in the substrate binding to the enzyme. The first proposal of His<sup>120</sup> function claims that the imidazole ring of His<sup>120</sup> may activate Cys<sup>184</sup> causing the formation of thiolate followed by nucleophilic attack of the carbonyl group of the motif protein. In the absence of substrate, the side chain of Cys<sup>184</sup> which are located at  $\beta$ 6/ $\beta$ 7 loop and His<sup>120</sup> located at  $\beta$ 2/ $\beta$ 3 loop does not interact with each other where the side chain of Cys<sup>184</sup> is projected out from His<sup>120</sup> into the hydrophobic pocket. However, the recent evidences showed that Cys<sup>184</sup> do not need activation from His<sup>120</sup> to form an imidazolium-thiolate bond in order to attack the carbonyl group of threonine in the LPXTG motif. In fact, at physiologic pH, both of the residues are not charged. According to the proposed determination of solvent isotope effect, only a small number of populations are active in their ionization form (Frankel *et al.*, 2005). Imidazolium ion of His<sup>120</sup> is found to take part in protonation of glycine when the covalent bond of threonine and glycine is broken. Besides, it also takes part in the second phase of the mechanism; the nucleophilic attack of pentaglycine of the lipid II by activating the terminal amine on the glycine residue (Frankel *et al.*, 2005; Suree *et al.*, 2009).

Meanwhile the function of Arg<sup>197</sup> is still in debates. Arg<sup>197</sup> which are anchored in  $\beta$ 8 existed near the Cys<sup>184</sup> is also shown to be important structure of sortase A. In a proposal by Luciano A Marraffini *et al.* (2004), they speculate the



major function of Arg<sup>197</sup> is for the ionization of Cys<sup>184</sup> or lipid II substrates (Zong *et al.*, 2004). Alanine substitution at position 197 causes a major decrement in enzyme activity of sortase A. (Hung Ton-That *et al.*, 1999; H. Ton-That *et al.*, 2002; Hung Ton-That *et al.*, 2000). According to the dynamic NMR structures in the recent study, Arg<sup>197</sup> showed potential to be a stabilizer for the oxyanion transition state of the formed intermediates (Suree *et al.*, 2009).

Sortase A from *Staphylococcus aureus* is strongly dependent on the Ca<sup>2+</sup> ion in the reaction. The LPXTG substrates are recognized by sortase A via contact with valine at position 166,168 and leucine at position 169 in the  $\beta$ 6/ $\beta$ 7 loop (Suree *et al.*, 2009). Upon substrate binding, the loop will undergo transition from a non-sequential open loop to a well ordered close state loop suggesting that the binding will cause a conformational change to the residue position. This conformational change is stabilized by the Ca<sup>2+</sup> ion binding to Glu<sup>171</sup> in the  $\beta$ 6/ $\beta$ 7 loop and a residue at  $\beta$ 3/ $\beta$ 4 loop (Naik *et al.*, 2006). Hence, substrate affinity is improved with the binding of Ca<sup>2+</sup> ion. If low concentration of Ca<sup>2+</sup> is used, the catalytic efficiency drops with increasing K<sub>m</sub>. Due to the strong dependency towards Ca<sup>2+</sup> ion, it is difficult to conduct the reaction in Ca<sup>2+</sup> scarce environment and in the presence of the reaction inhibitor, the calcium binding compound such as phosphate, carbonate and EDTA(ethylenediaminetetraacetic acid) (H. Hirakawa *et al.*, 2012).

### **2.2.3 Sortase mediated transpeptidation**

Sortase A conducts transpeptidation in a highly specific manner. There are three important proteins that was involved in the sortase A transpeptidation, the enzyme; Sortase A, the donor motif; LPXTG motif protein and the acceptor motif;

pentaglycine motif. According to Clancy *et al.* (2010), leucine, proline, threonine and glycine were preferred at position 1,2,4 and 5 in the Sortase A signalling motif but it does not have any preference in the amino acid selection at position 3. Preceding the transpeptidation reaction, sortase will cleave the LPXTG motif between threonine and glycine residue and form a thioacyl-intermediate with the captured LPXT-motif protein.

In 1999, a model of the sortase structure predicting its active site, the cysteine 184(Cys<sup>184</sup>) was developed presenting the nucleophilic attack from sortase to the carbonyl group of the peptide bond of threonine and glycine, followed by the thioester complex formation between the interacting active sites of Cys<sup>184</sup> and the carboxyl terminal group of threonine(Hung Ton-That *et al.*, 1999). The cleaved threonine carboxyl end was then attached to the pentaglycine motif of peptidoglycan covalently.

In the absence of nucleophile, sortase starts to hydrolyse the LPETG peptide at the same site in a slow pace with a ping-pong mechanism. A ping-pong mechanism, commonly known as double displacement mechanism is characterized by the formation of a substrate-enzyme intermediate in the first step followed by the disintegration of the complex when another substrate comes into contact to form the final product. At the end of reaction, the enzyme will return to its original state (Cleland, 1973). The limitation of this type of reaction is that the rate of hydrolysis of the acyl-enzyme is dependent on the rate of the same acyl-enzyme formation.

Mimicking the natural phenomenon of sortase conjugation, *in vitro* conjugation of LPXTG-motif protein and oligoglycine motif protein was demonstrated (Popp *et al.*, 2007; Popp *et al.*, 2009; Theile *et al.*, 2013). Substrate containing

LPXTG motif and oligoglycine ( $G_n$ ) was made using the standard molecular cloning protocol. Glutamic acid residue is commonly chosen to be introduced at the position 3 of the recognition motif as it is readily preferable in the native mechanism of sortase A (Boekhorst *et al.*, 2005).

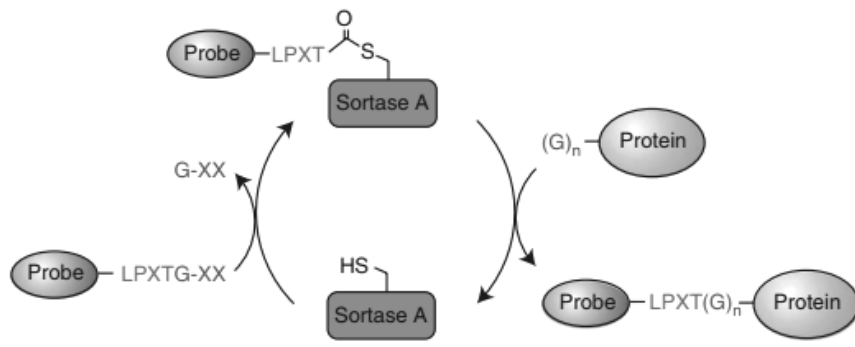
Biotin and fluorescent probes in conjugation with oligoglycine were used as substrates utilizing the sortase mechanism by modifying the LPETG peptide tag and fusing it to the C-terminal of targeted protein (Mao *et al.*, 2004; Ranganath Parthasarathy *et al.*, 2007; Tanaka *et al.*, 2008). As an example, oligoglycine on a folate was used to conjugate to the GFP-containing C-terminal LPETG tag via sortase reaction (Mao *et al.*, 2004). The results shows that one to three glycine residues can serve as substrates for the reaction, however it is rate-limited according to the amount of terminal glycine it possess (Mao *et al.*, 2004). Figure 2.2 shows the overall process of sortase-mediated conjugation and its chemical reaction.

#### **2.2.4 Single-chain fragment variable (scFv) as biosensors**

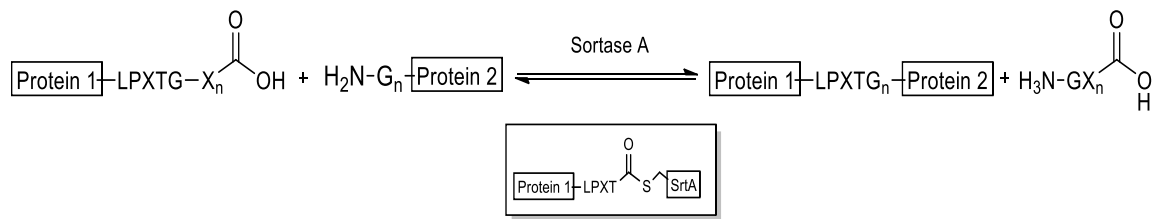
Antibodies are produced as a part of defensive response by vertebrate animals towards the harm caused by toxic materials and foreign organisms. Recombinant antibodies exist in various formats for example scFv, nanobodies, Fabs and minibodies. A commonly used antibody format for *in vitro* detection is the single chain fragment variable(scFv) which contains a variable heavy (VH) and a variable light( VL) chain connected by a sequence of peptide termed as glycine-serine linker without the constant region of the antibodies (Bird *et al.*, 1988; Liu *et al.*, 2009; Zeng *et al.*, 2012). It was first developed in the 1980s as an alternative to the large-sized natural antibody (~150kDa) (Borrebaeck & Wingren, 2011).

scFv is a small form of the antibody idiotype (~27kDa) that retains the antigen specificity as the full antibody (Padlan, 1994). They consist of six complementary regions classified in three groups namely CDR1, CDR2, CDR3 which is present on both VH and VL. Amino acids in the CDRs are complement to the amino acids presented on an antigenic site of an antigen. CDR3 region of VH in general is responsible for antigen-antibody specificity whereas the CDR3 region of VL is important for the affinity of the antibody towards the antigen (Zeng *et al.*, 2012).

Due to its small size, modification at genetic level is easily introduced. Metal-binding tags such as histidine and cysteine are directly introduced in its linker peptide, leading them to self-assemble on sensor surfaces in the correct conformation thus increasing the chances of binding of the antibody to the antigen. Generally, 5-20 amino acids of glycine(G) and serine(S) are being used to bring VH and VL together as a single fragment (Delves *et al.*, 2011). scFv containing less than 15 amino acids or without linker amino acid are often unable to fold properly, however they are able to link or entwin with other scFv to form diabodies, triabodies, tetramers or tandem scFv. These types of metameric scFv together with multivalent antigens are useful in increasing the sensitivity of an assay with respect to the increases in affinity and avidity of the complex itself (Conroy *et al.*, 2009; Holliger & Hudson, 2005). Other types of modification which can be made is the generation of minibodies; the scFv containing constant region of antibodies or conjugation to nanoparticles to generate multifunctional scFv for application in diagnostic or therapeutic field.



(i)



(ii)

**Figure 2.2** Sortase A reaction mechanism (adopted from Theile *et al.* (2013) (i) Structural sketch of reaction mechanism, (ii) Sortase A chemical reaction mechanism

### 2.2.5 Enhanced green fluorescent protein (eGFP)

Wild type green fluorescent protein(GFP) was first isolated by Shimomura *et al.* (1962) from *Aequorea victoria* and the sequence was retrieved later in 1992 by Prasher *et al.* (1992). GFP is a small protein consisting of 238 amino acids sequence with a molecular weight ranging from 27-30kDa. This monomeric protein gained much interest from researchers since its ultimate discovery. The initial work with GFP highlighted a major problem in terms of its low luminescence activity which falls below the standard of conventional reporters (Reichel *et al.*, 1996; G. Zhang *et al.*, 1996).

Due to the highlighted problem, the luminescence and expression of GFP was enhanced by introducing mutation to the chromophore of the GFP including codon optimization. These so-called GFP variants contain amino acid substitutions at position 54 and 66 where Phe<sup>54</sup> was replaced by Leu and Ser<sup>66</sup> was replaced by Thr (Cormack *et al.*, 1996). Aside from that, almost 200 nucleotide bases of GFP underwent silent mutations to suit the preference of the human codon-usage (Haas *et al.*, 1996). These enhanced variants of GFP were named as enhanced green fluorescent protein (eGFP).

In 1994, Chalfie *et al.* (1994) reported the usage of GFP as a reporter system in gene expression. This thought sparked a wave of studies in GFP manipulation as a reporter system in diagnostic and therapeutic application. GFP is commonly used as fusion partners for visualization of whole cells and even as transcriptional probes (March *et al.*, 2003). In building the screening platform, discovery of GFP provided a great impact in providing cost effective and time saving detection system. As a biomarker, GFP was used to select the potent CHO cells which maintain the

recombinant protein expression under repressed conditions (Yuk *et al.*, 2002), loss of fluorescence due to apoptosis in human and mouse cell lines (Strebel *et al.*, 2001), and it also contributed to the discovery of the twin-arginine translocation(TAT) pathway in *E.coli* where proteins are folded in the cytoplasm before being transported out to periplasm layer (DeLisa *et al.*, 2002).

Due to its various applications, GFP variants have been developed in which each and every variant shows different excitation and emission spectra from the wild type GFP. As observed using FL600 microplate fluorescent reader, red-shifted GFP variant has an excitation peak at 488nm and emission peak at 509nm whereas the wild type GFP has the excitation peak at 395nm and emission peak at 509nm and the blue GFP mutant has an excitation peak at almost 380nm and emission at 460nm. The application of GFP can aid to simplify and reduce the time required in a protein immunoassay. The fusion reaction between protein A and GFP for bioassay reported by Aoki and co-workers had successfully probe the SDS-PAGE gel for the presence of recombinant nonspecific enolase (Aoki *et al.*, 1996). After that, the method was improved using protein G originating from *Streptococcus* sp. for targeting a wider array of targets for example goat and sheeps IgGs and mouse IgGs (Arai *et al.*, 1998). The method was then simplified even more by using single molecule detection by fusing anti-hepatitis B scFv to eGFP. The developed probes function as the fluorescent antibody (Casey *et al.*, 2000). These facts indicate how important GFP is to the biotechnological fields because of its aptitude in drug-screening and for protein-protein interaction studies.

### 2.2.6 Invertase

Invertase or its other name beta-fructofuranosidase is an enzyme that converts sucrose to glucose via hydrolysis process. It is commonly expressed in various organisms, including bacteria, fungi and plants. One good example of sucrose-hydrolysing bacteria is *Zymomonas mobilis*. It is chosen as the best candidate in fermenting ethanol, defeating yeast. In *Z.mobilis*, there are three important sucrose-hydrolyzing enzymes, known as invA, invB and levansucrase (Buchholz & Eveleigh, 1990; Ohta *et al.*, 1991). invA was only found intracellularly, while, invB and levansucrase was found in the culture medium of *Z. mobilis*.

More than 60% of sucrose conversion via the bacterial enzyme was catalysed by the extracellular invertase (invB) (O'Mullan *et al.*, 1992). Research found that the catalytic activity of invB is the highest among the prokaryotic invertases (Bugbee, 1984; Vega *et al.*, 1991). Besides, the conversion reaction by invertase is highly specific as it catalyses its specific substrate only; the  $\beta$ -fructosyl end of a glycosidic bond which is commonly seen in sucrose. The product of this enzymatic reaction appeared to be only glucose and fructose at equimolar concentration. Although both levansucrase and invB are classified under extracellular sucrases, they take a different role. InvB do not catalyse the polymerisation of levan and not even take part in degradation of polyfructan as seen in *Bacillus subtilis* levansucrase (Martin *et al.*, 1987).

InvB is a 47kDa extracellular invertase which was released to culture broth and intracellular for its catalytic activity. It is located at the cell surface of the bacterial cells (Vásquez-Bahena *et al.*, 2006). The isoelectric point was calculated to be at 4.3 and the sucrose activity is at its optimum at pH 5.0-5.5 (O'Mullan *et al.*,



1992; Vallejo-Becerra *et al.*, 2008). InvB is the only invertase grouped in 68 families of glucosyl hydrolases which most of the family members come from levansucrases (Kondo *et al.*, 1994; Kyono *et al.*, 1995). Even though the expression level of *invB* in *Zymomonas mobilis* was very low, the saccharolytic activity that it has is high level compared to any other sucrose. Due to its high saccharolytic activities and specificities, *invB* is chosen as a biomarker in the assay development in recent studies (Borman, 2011; Xiang & Lu, 2011, 2013; X. Zhu *et al.*, 2014).

### **2.3 *Escherichia coli* as a host cell**

A full understanding on the properties of *Escherichia coli* strains is crucial in order to perform a successful technique in molecular genetics. *E.coli* has various strains which are genetically engineered to be used in the propagation and manipulation of recombinant DNA. *E.coli* exists as a short rod shaped bacteria from Gram negative family which contains a circular genome of 4.6 mega base pair. It is favoured as a model system due to its short doubling time and ability to grow on various media (Blattner *et al.*, 1997). *E.coli* has a doubling time of 20-30min in a rich media; therefore, it would be enough to generate a population of 1-2billion cells per ml of liquid media. Besides, the high transformability and the genetic manipulation secure its position as a suitable host for propagation, manipulation and also the characterization of recombinant DNA.

The commonly used *E.coli* in recombinant DNA technology is mostly derived from *E.coli* K12 strain. This strain was isolated from the faeces of diphtheria patient in 1922 (Tatum & Lederberg, 1947).

### 2.3.1 DH10B *Escherichia coli* strain

DH10B is a widely used *Escherichia coli* strain for propagation of large insert DNA library clones. Due to its high efficiency of DNA transformation and well preservation of large plasmid, it is fully utilized for cloning. Generating the strains takes up a series of extensive genetic recombination steps with 226 genes being manipulated. Originated from K-12 strain MG1655, mutation rate of this strain was observed to be 13.5 fold higher than its original strains suggesting the effect of sudden increase in insertion sequence (IS) transposition. Due to the transposition activity of IS elements, the genome architecture was remodelled and causing tandem duplication and inversion. Due to *leuLABCD* gene deletion, the strain needs addition of leucine to grow on minimal medium. Deletion of both *relA1* and *spoT1* alleles cause the strains to be sensitive to the nutritional deficiencies and grow slightly slower compared to the wild types strain (Durfee *et al.*, 2008).

MAX Efficiency DH10B<sup>TM</sup> developed by Life Technologies is a *mcrA*, *mcrB*, *mcrC* and *mrr* strain containing the mutation of *mcrA* and knockout of the whole *mrr-hsdRMS-mcrBC* region. This modification allows the production of higher amount of colonies per microgram of the methylated DNA compared to other strains. With the removal of the restriction systems, the chances of getting more clone containing methylated cytosine residue is enhanced. Reports also proved that strains with *mcrA* and *mcrBC* shows increment in efficiency of transformation by more than 1000-folds compared to DH5 $\alpha$  (Lorow & Jessee, 1990).

To efficiently clone methylated DNA, a few genetic markers was introduced. *recA1* was introduced to stabilize the foreign insert, meanwhile *endA1* takes the role of improving the DNA quality from plasmid isolation. *hsdRMS* is also mutated to

enable the cloning of DNA conducted in a restriction endonuclease-free environment. Besides,  $\phi 80lacZ\Delta M15$  marker also was introduced to allow alpha-complementation with the  $\beta$ -galactosidase gene from vectors harbouring it for blue-white screening using X-gal plate (Lorow & Jessee, 1990). Hence, DH10B is a suitable host for cloning purpose.

### **2.3.2 BL21(DE3)**

BL21 host strain is widely used for expression of recombinant protein in bacterial expression system (Sorensen & Mortensen, 2005). BL21 (DE3) is a type of expression B strain of *E.coli*. It contains the chromosomal  $\lambda$  DE3 encoding expression of T7 RNA polymerase (RNAP) under strict control of lacUV5 promoter (Samuelson, 2011) allowing expression under tight control of T7 promoter system. This B21 derivative has mutation of *lon* which encodes for lon protease and *ompT* encodes for outer membrane proteases to increase the expression yield and help in stabilization for the recombinant protein (Casali, 2003; Samuelson, 2011).

### **2.3.3 SHuffle T7 K12**

Originates from K-12 strain, SHuffle T7 contains modified genome with *trxB* and *gor* gene knockout. These genes are the reductases responsible in maintaining reduced condition of cytoplasm of *E.coli* where *gor* encodes for glutathione reductase and *trxB* encodes for thioredoxin reductase. Mutation of both genes allows disulphide bond formation in cytoplasmic region of *E.coli* (Casali, 2003).

This mutant strain was further modified with constitutive expression of disulphide isomerase (DsbC) encoded by *dsbC*. DsbC is a part of reducing system in periplasm of *E.coli* (McCarthy *et al.*, 2000). It takes a role of correcting the misfolded protein by transferring its disulphide bond to the protein by oxidizing the protein while it was reduced (Zapun *et al.*, 1995). The insertion of this feature allow correct folded protein formation within cytoplasm while ensuring the fidelity of disulphide bond as well as improving the yield of recombinant proteins.

Commonly, scFvs are expressed in cytoplasm and form the disulphide bond in periplasm layer where it was trans located with the aids of leader peptide following Sec(PelB) (H. Thie *et al.*, 2008) or Tat pathways (TorA) (DeLisa *et al.*, 2002). Using this systems, mature form of scFv without leader peptide are achieved after the leader peptide was cleaved off upon translocation (Dev & Ray, 1990). Even though the system allows correct protein folding in periplasmic layer, it suffers from low protein yield due to inefficient translocation (Lobstein *et al.*, 2012; Zarschler *et al.*, 2013). Disulphide bond in scFv is essential for the stability and functionality of the antibodies binding to its antigen (Ahmad *et al.*, 2012; Smallshaw *et al.*, 1999). Hence, a cytoplasmic-disulphide bond formation is an ideal to overcome such problem under tight control of T7 promoter.

## **2.4 Expression vector**

### **2.4.1 pRSET-B**

pRSET-B utilizes the T7 based expression system originally constructed by F.William Studier and his group in 1990 (Studier *et al.*, 1990). It is a widely used system for high level recombinant protein expressions adapted from T7 RNA