Engineering Sortase A for generating site-specific protein 3D assemblies

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Maryam Raeeszadeh-Sarmazdeh

May 2014

Copyright © 2014 by Maryam Raeeszadeh-Sarmazdeh.

All rights reserved.

"When you are with no one but me, you are with everyone…"

-Rumi

To **Reza**, my dad, for always believing in me

To **Mehri**, my mom, for inspiring me to work hard

To **Saeed**, my husband, for his continual and comprehensive support

And to **Sophia**, my little angel for being a huge source of energy

ACKNOWLEDGEMENTS

My Ph.D. journey would not be possible without support and help of many people. First, I want to thank Dr. Eric Boder, my advisor, for his kind support, guidance, and patience through my Ph.D. He always supported my ideas and encouraged my independence. I want to thank Dr. Ranganath Parthasarathy for being such a great help in my PhD research, for his great ideas and his generous donation of information and material. I want to also thank my committee members for their help and support: Dr. Bamin Khomami, Dr. Paul Frymier and Dr. Chris Stephens. I always used their guidance and intellectual views and advices. I was also very welcomed in their labs and offices. I want to thank the undergraduate assistants who helped me with my experiments: Molly Hunt, Carrie Lloyd, Leila Owen, and Nikhil Patel. I want to thank Vince Price, my labmate, who spent a good five years with me in SERF 719 and made my Ph.D. journey more pleasant. I want to thank Alex Meyers, Rosemary Le, Sarah Cruise, Jaana Manik, Sarah Williamson, and Mauricio Valverede my friends and labmates for sharing ideas and materials with me. I also want to thank my parents and my husband for supporting me during this time.

ABSTRACT

There is a great interest for protein immobilization and generating protein three-dimensional assemblies in nano-biotechnology. The main challenge in such protein immobilization and oligomerization is stability and lack of control over the protein ligation site. An enzymatic method overcomes these issues by site-specific protein ligation using conditions compatible with protein structure and stability. Sortase A, a transpeptidase that naturally binds threonine in a LPXTG sequence and glycine in a GGG sequence, provides a covalent site-specific protein immobilization and ligation tool that can be used in generating site-specific protein 3D assemblies. Sortase A accepts various nucleophile substrates. Therefore, engineering Sortase A substrate specificity to create a more powerful tool to site-specifically ligate proteins together and to the surface is advantageous. Toward the goal of generating protein 3D assemblies and engineering Sortase A, we accomplished the following:

First, we site-specifically immobilized recombinant fluorescent proteins on the surface using sortase reaction. A self-assembled layer of GGGC or AAAC peptides, *S. aureus* or *S. pyogenes* Sortase nucleophile substrates, were made on the gold surface. Recombinant fluorescent proteins with LPETG/A tag at the C-terminus were then immobilized on this layer using corresponding sortase A reactions.

We then generated protein wires and assemblies in solution and on the surface in a controlled manner using two strategies. In the first strategy, we used two orthogonal sortases with a slightly different substrate specificity, *S. pyogenes* and *S. aureus* sortases, to control the protein assembly. In the second strategy, we used an enterokinase-cleavable protecting sequence at the N-terminus of a bifunctional protein, such that enterokinase-

mediated removal will activate the N-terminus as a nucleophile substrate after each round of sortase-mediated immobilization and enable sequential deposition of single layers.

As the last goal of this study, we used yeast surface display and directed evolution to engineer Sortase A substrate specificity toward primary amine in the side chain of lysine in a pilin box sequence. A library of sortase mutants and LPETG substrate were displayed on the yeast surface. Sortase library was screened using FACS and mutants with the most activity toward the pilin box sequence were selected and analyzed.

TABLE OF CONTENTS

Chapter 1 Background and significance 1
1-1- Introduction 1
1-2-Background 3
1-2-1-Sortase structure, function, mechanism, and applications 3
1-2-1-1-Sortase function 3
1-2-1-2-Crystal Structure and mechanism 7
1-2-2-Substrate specificity 9
1-2-1-4-Sortase reaction applications 13
1-2-2-protein immobilization and 3D assembly on the surface 15
1-2-2-1-Self assembly on the surface 15
1-2-2-2- Protein immobilization on the surface 16
1-2-2-3- Patterning proteins on the surface 18
1-2-2-4- Protein 3D assemblies and molecular printboards 20
1-2-3- Engineering sortase using yeast surface display and directed evolution 22

Chapter 2 Protein immobilization on the surface 23
2-1- Introduction 23
2-2-Material and methods 25
2-2-1-Protein production and purification 25
2-2-2-Peptide assembly on the gold surface 27
2-2-3-Protein immobilization on the peptide SAM on gold 27
2-2-4-Fluorescence Microscopy 28
2-2-5-Atomic Force Microscopy 28
2-2-6-Western blot 28
2-3-Results and discussion 29
2-3-1- GFP-LPETG immobilization on the gold surface using Sa Srt 29
2-3-2- Protein immobilization on the surface using two different sortases 33
2-3-3- Protein immobilization and oligomerization on gold 36

Chapter 3 Generating protein 3D assemblies using OrthoGonal sortases 40
3-1-Introduction 40
3-2-Material and methods 41
3-2-1-Protein production and purification: 41
3-2-2-Peptide assembly on the gold surface 42
3-2-3-Protein immobilization on the peptide SAM on gold 43
3-2-4-Fluorescence Microscopy 43
3-2-5-Atomic Force Microscopy 44
3-2-6-Western blot 44
3-3- Result and Discussion 45
3-3-1- Generating protein 3D assemblies using orthogonal sortases 45
3-4- Conclusion 53

Chapter 4 Controlling Sortase A-mediated site-specific protein immobilization on the surface using a protease cleavable protecting group 55
4-1-Introduction 55

4-2-Material and Methods **57**
4-2-1-Plasmids and strains 57
4-2-2- Protein production and purification 57
4-2-3- Generating protein wires in solution 58
4-2-4- Protein assembly on the bead surface 59
4-2-5- Western blot 60
4-2-6- Fluorescence Microscopy 60
4-2-7-Peptide assembly on the gold surface 61
4-2-8-Protein immobilization on the peptide SAM on gold 61
4-3- Result and discussion **62**
4-3-1-Generating protein wires in solution using Sortase A 62
4-3-2- Site-specific assembly of protein layers in a controlled manner on the bead surface 63
4-3-3- Site-specific assembly of protein layers in a controlled manner on the gold surface. 67
4-4- Conclusion **69**

Chapter 5 Engineering Sortase A substrate specificity **71**
5-1-Introduction **71**
5-2-Materials and methods **73**
5-2-1-Plasmids and strains 73
5-2-2-Generating sortase library 74
5-2-3-Sortase reaction on the surface of yeast cell 75
5-2-4-Flow cytometry and Immunolabeling 75
5-3- Results and Discussion **77**
5-3-1- Sortase A site-specific peptide ligation on the yeast surface 77
5-3-2- Library screening and sorting 80
5-3-3- Mutations 84
5-4- Conclusion **88**

Chapter 6 Conclusion and future work **93**
6-1- Summary and conclusion **93**
6-2- Future works **95**
6-2-1- Further applications of protein patterning and 3D structures 95
6-2-2- Generating smart protein assemblies 95
6-2-3- Engineering substrate specificity of Sortase 96

LIST OF REFERENCES **99**

APPENDIX **105**

Appendix 1 Atomic force microscopy images **106**

Appendix 2 PLASMID MAPS **107**

Appendix 3 Sequences **109**

Appendix 4 Flow cytometry data **113**

VITA **122**

LIST OF TABLES

Table 1-1- Substrate specificity of different sortases 11

Table 2-1- AFM parameters for protein immobilization on gold..... 39

Table 3-1- analyzing protein dimerization reactions using orthogonal sortases 49

Table 3-2- Mean fluorescent intensity values of mcherry and GFP immobilized on the surface using Sa and Sp Srt reactions …….……51

Table 4-1- Mean fluorescent intensity values of mcherry and GFP immobilized on the PS beads using Sa Srt reaction and EK method 66

Table 4-2- Mean fluorescent intensity values of mcherry and GFP immobilized on the gold surface using Sa Srt reaction and EK method ……70

Table 5-1-Effect of the nucleotide analogue (dPTP and oxo-dGTP) concentration in mutation rate of sortase library 82

Table 5-2- peptides used in sortase reaction on the yeast surface ……...……..84

Table 5-3- Mutations from directed evolution of sortase library for pilin peptide …….....89

LIST OF FIGURES

Figure1-1. Sortase sorts surface proteinin gram positive bacteria. 5
Figure 1-2- *S. aureus* Sortase A active site.. 8
Figure 1-3. Sortase A structure. layover of S. aureus and S. pyogenes sortase structure.....10
Figure1-4. Comparison of different sortases structures. 12
Figure1-4. Sortase A- mediated ligation applications. 14
Figure 2-1. Sortase-mediated immobilization of proteins on gold.....30
Figure 2-2. Fluorescent micrographs of immobilized protein on the gold surface using Sortase . 31
Figure 2-3. The impact of enzyme: substrate ratio on sortase-mediated protein immobilization.. 32
Figure 2-4. Atomic Force Microscopy images of GFP-LPETG immobilized on the gold surface using Sortase A reaction 34
Figure 2-5. Fluorescent protein immobilization on the gold surface 35
Figure 2-6. Sp and Sa Srt reactions in solution 37
Figure 2-7. AFM image of immobilized protein oligomers on the gold surface. 38
Figure 2-8. Protein removal from the gold surface after sortase reaction. 39
Figure 3-1. Sequential site-specific protein immobilization using orthogonal sortases.....47
Figure 3-2. Fluorescent protein oligomerization using orthogonal sortases reactions in solution. 48
Figure 3-3. Fluorescence micrographs of protein layers immobilized on the gold surface using orthogonal sortases 51
Figure 3-4. AFM image of the protein layers immobilized on the gold surface. 52
Figure 4-1. Generating protein wires in solution using Sortase A reaction and controlling it using an N-terminal cleavable sequence. 64
Figure 4-2. Site-specific controlled protein immobilization on the bead surface using SortaseA. 66
Figure 4-3. Anti-DsRed immunoblot to detect protein oligomers immobilized on the bead surface using Sortase A and EK protecting method.....68
Figure 4-4. Protein immobilization in layers on the gold surface using Sortase A and EK protecting method 70
Figure 5-1. Sortase reaction on the yeast surface using purified sortase.....78
Figure 5-2. Srt reaction on the yeast surface 81
Figure 5-3. Sorting library of Srt mutants displayed on the yeast surface against P0 peptide using FACS. 83

Figure 5-4. Mutations found in sortase library after 3 rounds of sorts against P0 peptide 85
Figure 5-5. Sortase structure, space filling model, showed using Pymol.....87
Figure 5-6. NCMFI value for different colonies. The experiments have been done at least two times.....90
Figure 5-7. Sortase library screening against P0 peptide.. 91
Figure 5-8. Sortase library substrate specificity against 3G-Biotin or Biotin-P0 peptide.....92
Figure A1-1- Atomic force microcopy for bare gold, peptide self assembly, and GFP-LPETG layer on the gold surface……..106
Figure A2-1- Map of pMR2 plasmid ……..107
Figure A2-2- pCUAga1Srt construct …….. 108
Figure A4-1. GFP immobilization on polystyrene beads……113
Figure A4-2 . Biotin-LPETG peptide immobilization on 3G polystyrene beads.....114
Figure A4-3. Srt library sorts against P0 peptide comparing to YPA and YPK peptides.....115
Figure A4-4. Sortase library and sorts against 3G-biotin peptide. FL1: AU1 expression, FL4: Streptavidin binding to the 3G-biotin bound to the surface.....116
Figure A4-5. Control-P0 (K to G) peptide reaction tests. Naïve sortase library and sorts were tested against P0 control peptide, that has its lysine replaced by a glycine.....117
Figure A4-6. Tests against Ethylenediamine-biotin conjugated (EDA) peptide for substrate specificity against alkylamines. WT, Inactive, Naïve library and sort 3 (against P0 peptide) were tested against EDA peptide 118
Figure A4-7- Flow cytometry data for individual colonies.....119

List of Abbreviations:

Srt: *S. aureus* Sortase A

Srtm4: Tetramutant *S. aureus* Sortase A

Bifunctional GFP: GGG-GFP-LPETG

EK: Entrokinase

EK-bifunc-GFP: bifunctional GFP with enterokinase recognition sequence; DDDDK-GGG-GFP-LPETG

AFM: Atomic Force Microscopy

FACS: Fluorescence-Activated Cell Sorting

YSD: Yeast Surface Display

CHAPTER 1

BACKGROUND AND SIGNIFICANCE

1-1- Introduction

Immobilization of proteins on the surface has drawn particular attention in the past decades, mainly due to its importance in many nano-biotechnological applications such as proteomics, biosensors, or tissue engineering (1,2). There was a huge progress in genomic analysis recently, and there is a shift to proteome analysis. Therefore, generating protein microarrays and microplates has a great importance (3). Furthermore, immobilizing multilayers of proteins on the surface to generate a three dimensional protein assembly in a site-specific manner is advantageous. Although many different techniques have been used to date to ligate proteins/peptides together and immobilize proteins on the surface, there is still a great need for a robust universal method of protein immobilization that can site-specifically ligate proteins to the surface and other biomolecules. Most of the techniques used for protein ligation and immobilization such as physical, chemical or affinity tags lack stability and control over the ligation site. Therefore, an enzymatic approach has advantages since such protein immobilization and ligation can be achieved covalently and site-specifically in mild conditions compatible with protein structure and function.

Sortase A, a transpeptidase that ligates LPXTG and GGG, has been recently used for sequence-specific immobilization, oligomerization (linear or branched) and circularization of protein. Sortase ligates proteins/peptides using two short peptide sequences. These short tags will not significantly change protein structure and function. They can be used to bind

two recombinant proteins, circularize a bi-functional protein that has both tags at two termini, or ligate a recombinant protein with one tag to the surface that has been functionalized with the other tag. These properties make sortase an ideal tool for generating multidimensional protein assemblies on the surface and molecular printboards.

Sortase A has a broad range of specificity for the nucleophile or acyl acceptor substrate (GGG), and other nucleophiles such as alkylamines, hydroxylamines and primary amine in the side chain of lysine also work as sortase substrate. Furthermore, different sortases have different recognition motifs: for example LPXTG for *S. aureus* sortase A and NPQTN for sortase B. *S. aureus* and *S. pyogenes* sortases also have slightly different substrate specificity. Orthogonal oligomerization of proteins and site-specific attachment of layers of proteins can be achieved using sortases with different substrate specificities. Therefore, engineering sortase substrate specificity is an interesting area of study.

The work in this thesis was focused on generating protein three-dimensional assemblies on the surface in a controlled manner. As the first step, we immobilized protein on the surface site-specifically using Sortase A reaction. We used gold thiolate chemistry to immobilize one of the sortase substrates, (G/A)GG, on the surface. Then we site-specifically immobilized a layer of recombinant fluorescent protein with a LPET(G/A) sequence at the C-terminus on the surface using Sa and Sp Sortase A reactions.

As the next step toward generating protein 3D assemblies, we tried two different approaches to immobilize proteins on the surface layer-by-layer. In the first approach, we used orthogonal sortases to sequentially immobilize layers of protein on the surface. In the second approach, N-terminal protected bifunctional protein was used to generate protein assemblies layer by layer. As the final step, we used yeast surface display and directed

evolution to engineer Sortase A substrate specificity toward ε-amine in side chain of lysine in a pilin box sequence. After screening a library of sortase mutants using FACS, colonies have been selected and analyzed.

Taken together, we immobilized fluorescent proteins site-specifically on the surface using sortase reaction. We developed two different approaches to control layer-by-layer protein immobilization on the surface. We then engineered substrate activity and specificity of sortase toward primary amine in the side chain of lysine in a specific sequence.

1-2-Background

1-2-1-Sortase structure, function, mechanism, and applications

1-2-1-1-Sortase function

Sortases are transpeptidase enzymes found in the plasma membrane of most gram-positive bacteria. Houskeeping sortases attach surface proteins to the cell wall and pilin sortases are responsible for assembly of long protein fibers termed pilus. Pilin sortase recognizes a LPXTG tag (X can be any amino acid) at the C-terminus of the pilin subunit (Fig.1-1). Cysteine at the active site of the enzyme attacks the carbonyl group of threonine and cleaves between T and G to make an acyl enzyme intermediate. Pilus polymerization continues with a nucleophile attack from ε-amine of lysine in a pilin motif, which resolves the intermediate and makes a covalent bond between two subunits. The pilin then will be attached to the cell wall by a house keeping sortase with a similar mechanism as of the pilin sortase. It recognizes the LPXTG sorting signal and cleaves between the Thr and Gly residues to form an acyl intermediate between the Thr residue of the surface protein and a reactive Cys in the TLXTC catalytic pocket of the sortase. Subsequently, this intermediate will be resolved by a

nucleophile attack of a pentaglycine (G_5) in lipid II. In *S. aureus*, lipid II consists of an undecaprenylpyrophosphate anchor linked to N-acetylmuramyl (MurNAc), which is in turn bound to N-acetylglucosamine (GlcNAc) and to the cell wall pentapeptide l-Ala–γ-d-Gln-l-Lys–d-Ala–d-Ala; the l-Lys is bound to a short peptide (a pentaglycine in *S. aureus*) that will normally form the crossbridge between two pentapeptides of different peptidogly- can strands. However, the terminal amino acid of the pentaglycine can also 'attack' the linkage between the sortase and the surface protein, forming an amide bond that tethers the surface protein to lipid II in the bacterial plasma membrane. Subsequent steps then transfer the polysaccharide residues and the attached protein to the peptidoglycan cell wall

(4-6).

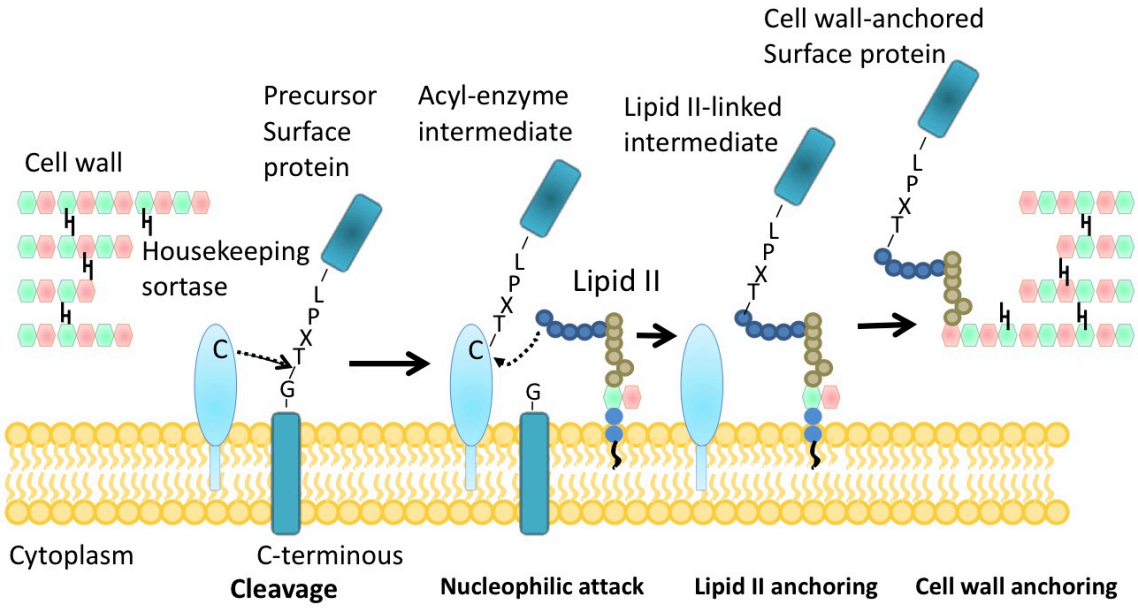
Sortase A fits into the housekeeping sortase group of sortase classification. It recognizes a LPXTG sequence, and needs Gly_n (n>1) to act as a nucleophile substrate or acyl acceptor.

(7-9) (10). *S. aureus* sortase A is a 206 amino acid protein with an N-terminal signal peptide/membrane anchor. This hydrophobic segment that functions as a signal peptide for secretion and as a stop transfer signal for membrane anchoring. The N terminus of the enzyme is located inside the cytoplasm and the C-terminal enzymatic portion located across the plasma membrane (11,12).

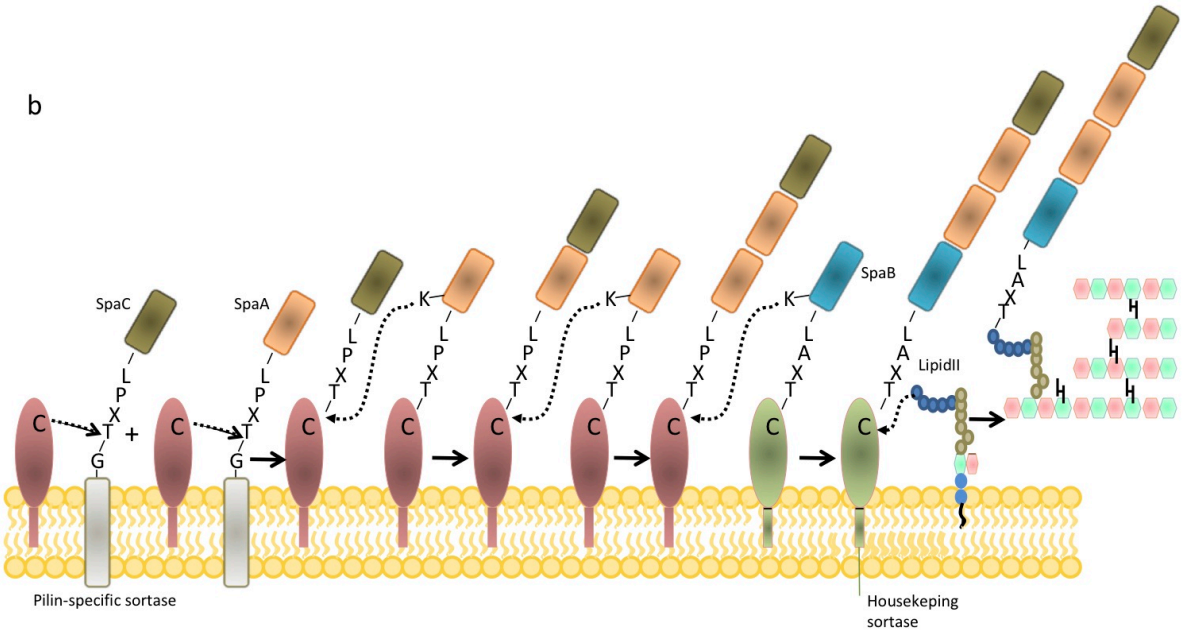
Figure1-1. Sortase sorts surface proteinin gram positive bacteria.

a) Mechanism of sortase-mediated surface protein anchoring to the cell wall. Surface proteins are synthesized in the cytoplasm as precursor proteins with an amino-terminal signal sequence sorting signal. The C-terminal sorting signal consists of a positively charged tail, a hydrophobic region and a LPXTG motif. Following secretion by the Sec secretion system, signal peptidases cleave the signal peptides of surface proteins, thereby producing the precursor surface proteins. The active-site Cys of the sortase cleaves the amide bond between Thr and Gly of the C-terminal pentapeptide LPXTG motif and generates an acyl–enzyme (thioester) intermediate. Nucleophilic attack by the amino group within the pentaglycine crossbridge of lipid II links the C-terminal Thr of the surface protein to lipid II. Penicillin-binding proteins incorporate the precursor into the cell wall as a mature cell wall-anchored surface protein for surface display by catalyzing a transpeptidation reaction. b) Pilin assembly on the cell wall of gram positive bacteria using Sortase. A similar mechanism occurs with SpaCAB pili of Corynebacterium diphtheriae. An internal isopeptide bond is formed between SpaC and SpaA and subsequently between the SpaC–SpaA dimer and additional SpaA subunits to allow pilus polymerization. The pilus is anchored to the cell wall via a SpaB subunit, which is already linked to the housekeeping sortase. Nucleophilic attack by Lys139 of SpaB incorporates this minor subunit into the pilus. Nucleophilic attack of lipid II at the acyl intermediate of the housekeeping sortase with SpaB transfers the pilus to the cell wall envelope and terminates pilus assembly (5).

a


Cell wall
Housekeeping sortase
Precursor Surface protein
Acyl-enzyme intermediate
Lipid II
Lipid II-linked intermediate
Cell wall-anchored Surface protein
Cytoplasm
C-terminous
Cleavage
Nucleophilic attack
Lipid II anchoring
Cell wall anchoring

b


SpaC
SpaA
SpaB
LipidII
Pilin-specific sortase
Housekeping sortase

1-2-1-2-Crystal Structure and mechanism

The soluble sortase A lacks the N-terminal signal peptide/membrane anchor (ΔN59). In order to use soluble sortase for *in vitro* activity assays and structural analysis, this hydrophobic region, was replaced with a six-histidyl tag and recombinant protein was purified. The core of the $SrtA_{N59}$ crystal structure is an 8-strand barrel. One helix and two 3-turn helices connect the strands, forming a novel fold. One side of the barrel, formed by the β4, β7, and β8 strands, is concave in appearance and, along with three of the surrounding loops, forms a tunnel-like hydrophobic pocket in the center of which the catalytic Cys184 is located. Cys 184, His 120, and Asn 98 are located in or around the active site. These three residues are positioned in a configuration similar to catalytic triad of cysteine proteases, although there is a difference between sortase and other cysteine proteases. In papain superfamily, His and Cys are positioned on loops at the end of β strand and α helix and are exposed to the cleft that substrate binds, while 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. Asn 98 is also not in a close proximity of the active site as it is in other cysteine proteases, while Arg 197 is located in the close proximity to the active-site cysteine. Measuring mutant enzyme activity shows that replacement of Cys 184, His 120 with Ala abolishes the enzyme activity, and replacement of Arg 197 greatly reduces the enzyme activity, while mutating Asn 98 with Ala does not affect the activity. Therefore, three key residues that are responsible for the enzyme activity are Cys 184, His 120 and Arg 197 (Fig. 1-2).

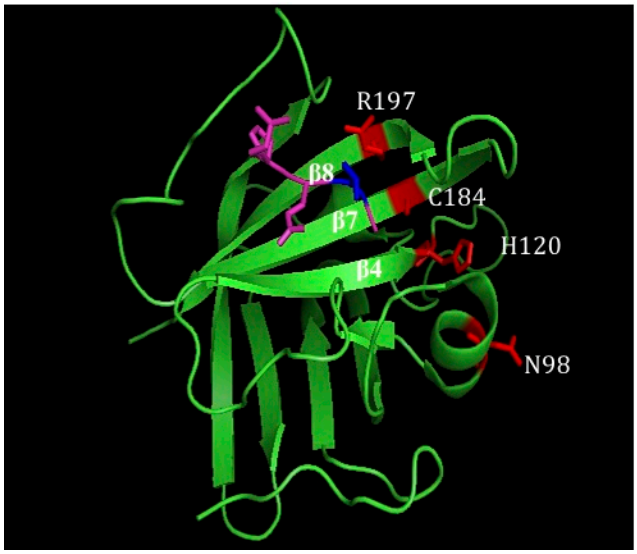


Figure 1-2- *S. aureus* Sortase A active site. The important residues with side chains are shown in red. LPETG substrate is shown in magenta-blue. R197 and Cys184 in the active site are located in a close proximity of the T-G bond in LPETG substrate. T is shown in blue.

The LPXTG substrate-binding site is located in a concave plane between β7 and β8 strands, and T and G peptide bond is positioned in a close proximity of Cys 184 and Arg 197. Therefore, it is possible that the enzyme mechanism leads by a catalytic dyad of Cys-Arg, and the side chain of Arg 197 that can be moved to a close proximity of Cys 184 (3.5 A) works as an ionizable group to protonate the amide atom of the substrate scissile bond and facilitate the Cys nucleophile attack. Thr 180, Ile 182 and Ala 118 that are conserved in most sortases seem to be responsible for binding to leucine and threonine of the LPXTG substrate (13,14).

S. aureus sortase activity will be reduced eight fold in absence of calcium. The side chains of some acidic residues in β3- β4 loop such as Glu105, Glu108 and Asp108 interact

with Ca^{2+}. β6/ β7 loop is disordered in the absence of calcium and Ca^{2+} makes some conformational changes upon Ca^{2+} binding by interacting with Glu 171. Therefore, binding of Ca^{2+} activates sortase by a mechanism that may facilitate substrate binding by effecting structure and dynamics of the active site loop (15,16).

In contrast to Sa-SrtA (*S. aureus* Sortase A), Sp-SrtA (*S. Pyogenes* Sortase A) does not need Ca^{2+} to be activated. The structure of Sa-SrtA and Sp-SrtA are very similar, although they have 24% sequence identity in the core catalytic domain. There are some differences in these two structures especially in the connecting loops, and the N/C termini. Some shifts have been observed in the loop connecting β2/ β3, β3/ β4, β6/ β7, β7/ β8, and β3 loop. Cys-208, His-142 and Arg-216 are the key catalytic residues, and despite other sortases, His-142 is positioned in a close proximity to the active site to play the role of protonating the intermediate. Fig. 2 shows the structure of these two sortases (7).

1-2-2-Substrate specificity

S. aureus sortase A is specific toward the acyl donor substrate LPXTG (X can be any amino acid). An initial rate study using 15 μM enzyme and a peptide library shows that for the amino acid in the third place, M is the best and T is the worst. Sortase does not recognize the inverted substrate (GTEPL). LPETG that mimics the sortase recognition motif of *S. aureus* protein A is a very common sortase substrate (17,18). Sp-Srt recognizes both LPXTA and LPXTG substrates (18,19).

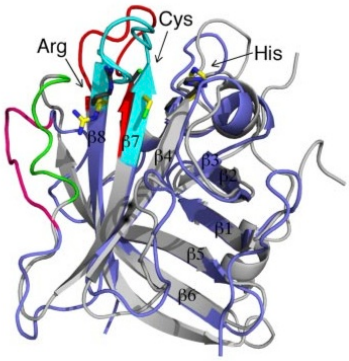


Figure 1-3. Sortase A structure. layover of S. aureus and S. pyogenes sortase structure shown in gray and purple (20). This image has the permission for reuse from the publisher.

For the acyl acceptor substrate, Sp Srt naturally accepts di-alanine as the nucleophile substrate, while Sa Srt prefers pentaglycine of the lipid II. Sortase has a broader range for acyl acceptor substrates that are responsible for the nucleophilic attack the intermediate. A diglycine at the N-terminus of the peptide or protein is sufficient to attack the sortase reaction intermediate, resolve the intermediate and bind to threonine in the LPETG substrate. GlyVal and GlyAla also work as nucleophile substrates although with slower kinetic rates. Alkylamines and their derivatives have been also shown to work as the nucleophile substrates for sortase reaction *in vitro*, although less efficient than the natural substrate (Gly_n). It is likely that aminoethyl works as a relatively good acyl acceptor for the sortase reaction (21). 6-deoxy-6-aminohexose is another sortase substrate that can be used in sugar-peptide conjugations (22). When a nucleophile is not available, water acts as a poor nucleophile and hydrolyzes LPETG and makes LPET-OH (10). Table 1-1 shows substrate specificity of different sortases.

Table 1-1- Substrate specificity of different sortases

	Acyl donor	Acyl acceptor
Sa Srt A	LPXTG	G_5 (lipid II)
Sp Srt A	LPXT(A/G)	AA/GGG
Srt B	NPQTN	G_5 (Peptidoglycan Crossbridge)
Srt C	(I/L)(P/A)XTG	ε-NH_2 of K in pilin

1-2-1-3- Sortase role in pilin assembly in gram-positive bacteria

Gram-negative bacteria, pilus assembly mechanism is not fully understood in gram-positive bacteria. Pili are composed of different subunits, called Spa (sortase-mediated pilus assembly). In *Corynebacterium diphtheria,* Sortase A is necessary for assembly of major subunit SpaA as well SpaB and SpaC. Studies show that assembly of SpaA is affected by deletion of srtA, while deletion of other sortases (B-F) did not interfere with SpaA pilus assembly, although srtF deletion resulted in improper envelope attachment of SpaA. There is a sequence homology between different pilin subunits (Spa A, SpaD, and SpaH), which shows a conserved pilin peptide motif containing a lysine (K) residue (WxxxVxVYPK). Mutating lysine to arginine (R) or alanine (A) results in inhibition of pilus assembly, which suggests that amino group in the side chain of lysine works as a nucleophile group for sortase transpeptidation reaction (23,24). In Group B Streptococcus (GBS), Sortase C is

responsible for pilus assembly. Class C sortase, known as pilin-specific, forms covalent bonds between pilin subunits. Pilin sortases ususally have broader substrate specificity comparing to housekeeping sortases, and multiple pilin proteins with various sequences can act as acyl donors and acceptors, similar to LPXTG and GGG for Sa Srt (table 1-1). Structural studies on Sortase C shows the catalytic triad (Cys, His, Arg) similar to Sortase A. They differ with other sortases in a hydrophobic C-terminal domain (5,25).

The crystal structures of *Streptococcus pneumoniae* pilin-specific sortases (SrtC1, SrtC2 and SrtC3) showed similarities between these structures and those of sortase A and sortase B proteins. Interestingly, the α3 helix forms a flexible lid over the active site; this is a unique feature of pilin-specific sortases and is required for efficient polymerization of pili and for enzyme stability (5).

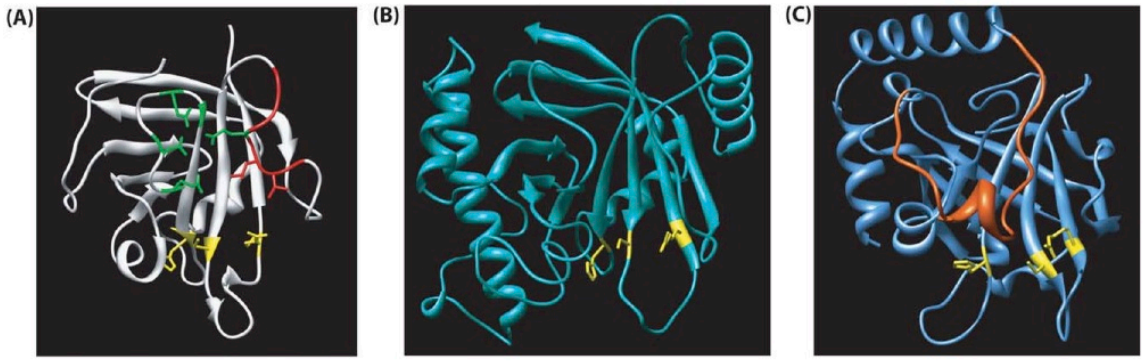


Figure1-4. Comparison of different sortases structures. The active site is shown in yellow. a) S. aureus Sortase A. Ca binding residues are shown in green, and β6/β7 loop is shown in red. b) S. aureus Sortase B has a similar fold with additional helices. c) S. pneumonia SrtC-1 Structure also contains the sortase fold but has an additional N-terminal loop that is located close to substrate binding site (orange) This image was used with permission from the publisher (26).

1-2-1-4-Sortase reaction applications

Sortase-mediated protein ligation has been widely used in protein labeling, oligomerization, circularization and immobilization. Large scale modification of biomolecules was done using immobilized sortase on the surface (27). Site- specific N-terminal and C-terminal labeling of proteins were also done using sortase-mediated ligation (28-31). M13 bacteriophage has been double labeled using an orthogonal sortase method (32). Unnatural compounds such as fluorescein, biotin, PEG, tetramethylrhodamine were conjugated to proteins using sortase A transpeptidase reaction (29,33-35).

Sortase site-specific transpeptidation reaction can be used to ligate protein/peptides with the short Srt recognition sequence at the C-terminus, LPETG, to another protein or peptide with the other Srt substrate , GGG, at the N-terminus or to the surface modified with this peptide. If these two tags, LPETG and GGG, are located at two termini of a protein, site-specific protein ologimerization or circularization can be done using Srt reaction (Fig. 1-4).

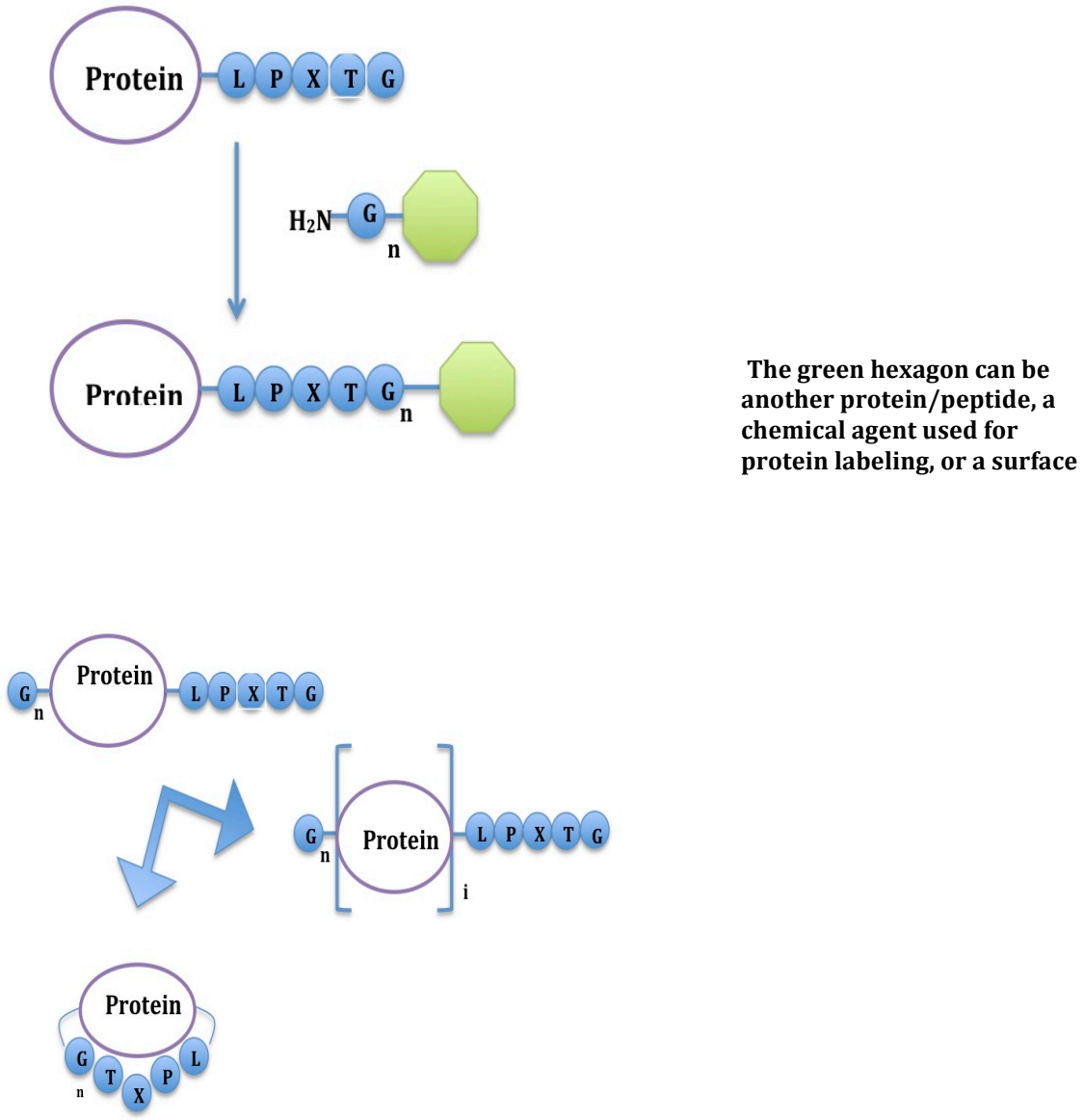


Figure1-4. Sortase A- mediated ligation applications. Sortase-mediated ligation can be used to site-specifically immobilize proteins or peptides on the surface. Using two sortase substrates at two termini sortase-mediated ligation can be used to oligomerize or circularize recombinants proteins.

1-2-2-protein immobilization and 3D assembly on the surface

1-2-2-1-Self assembly on the surface

Surface of gold and silicon can be modified by organic self-assembled monolayers (SAMs) that provide functional groups on the surface. Self-assembly of thiol-containing compounds on gold and silane-containing compounds on glass (or silicon) has a great control over chemical structure of organic compounds on the surface. We will use these two systems to generate a monolayer self-assembly that will be used as a sortase substrate.

The chemistry involved for the chemisorption of thiols on gold is still not fully understood. The formation of SAMs from thiols is not complicated by chemistries that might be required to displace or reduce surface oxides, but the details regarding the nature of the metal-sulfur bond and the spatial arrangement of the sulfur groups on the underlying gold lattice are still controversial. The formation of a thiolate requires the chemical activation of the S-H bond of the thiol (or the S-S bond of the disulfide). Gold-thiolate SAM binding energy studies shows that the adsorption of dimethyl disulfide on Au(111) occurs dissociatively. The reaction is fully reversible, and recombinative desorption of the disulfide is an activated process with a barrier lying near 30 kcal/mol. This energy suggests that a fairly significant degree of charge transfer to sulfur must occur in the thiolates; an inference that has been supported by the results of theoretical calculations (36,37).

We can use the SAM chemistry to immobilize the sortase substrate on the surface. We will use the sulfohydryl group of a cysteine to form gold-thiolate bond on the surface. Proteins with sortase recognition tag (LPXTG) can be covalently immobilized on these functionalized surfaces.

1-2-2-2- Protein immobilization on the surface

Many of the biosensors are based on protein-protein or protein-small molecule interactions. Therefore, protein immobilization on the surface has a great importance in proteomics, biosensors, tissue engineering, and supported catalysts. DNA microarrays had a great progress recently, and the interest has been shifted toward protein microarrays. Generating protein microarrays is more challenging than DNA microarrays, since there are more non-specific interactions between proteins and the surface, and most of the proteins will lose their structure and function when they are oriented in the non-specific direction. Proteins were immobilized on the glass surface using silane-coupling method. A covalent bond was made between the glass surface modified with bifunctional silane and amino groups in protein. The disadvantage of this method is that chemical reaction conditions for binding the aldehyde-containing silane reagent to the amines of the protein may harm protein structure and stability. Protein immobilization on gold was done using Au-S bond between an exposed cysteine residue and the gold surface. The drawback of this method is the need to engineer protein to have Cys at the specific site (38-40). Furthermore, these chemical immobilization methods lacks the control over protein orientation since more than one reactive site may be available in protein structure. Sortase-mediated protein immobilization has the advantage of site-specific ligation under the conditions that will not normally harm protein stability. Furthermore, sortase recognition motif is a short peptide that will not change protein structure and function.

Proteins have been immobilized on the surface using different methods such as physical adsorption, coupling between antibody–antigen, biotin–streptavidin, click chemistries and enzymatic ligation. In the physical adsorption of proteins on the surface, proteins do not

need to be further modified, although this method has the drawbacks of low control on protein orientation, which may result in losing the conformation and function of the protein, and low-protein density on the surface because of the two-dimensional space limitation and protein interactions with the surface. Furthermore, the immobilized protein using physical method may be denatured due to dried state of operation/storage. To overcome this issue, hydrogels are used to provide a semi-wet environment for protein immobilization, although using this technique; non-bound proteins trapped in hydrogels would be difficult to remove. Another method to immobilize proteins is using Ni^{2+} -NTA modified surfaces to bind to the recombinant proteins that have a Hisx6-tag at N- or C- terminus away from the active, functional site. This immobilization method has the disadvantage of being sensitive to pH and some chemicals such as imidazole, and detergents. Furthermore, there are not consistent results showing proteins immobilized on the surface using Hisx6-tagging method maintain their functionality. Zhu *et al.* and Hirabayashi *et al.* showed Hisx6-tag protein immobilization is an efficient method to bind proteins on the Ni modified surfaces while the proteins maintain their functionality comparing to other methods (41,42). Although Vallina-Garcia *et al.* showed no improvement in antigen detection using Ni^{2+} - Hisx6-tag interaction to immobilize Fab fragment of an antibody in comparison with the direct deposition of the protein. There are other methods that use streptavidin-biotin interaction, the drawback is that it is difficult to control the biotinylated site on the protein and therefore site-specific ligation of protein on the surface is still an issue of this method. Antibody-antigen interaction is also used to immobilize proteins in the right orientation. In this approach, the antibody is immobilized on the surface and the protein with antigen tag is then bind to this antibody.

Immobilizing the antibody in the direction that has the antigen recognition site exposed is an issue at first place (43).

In summary, immobilization techniques mentioned above have the drawbacks of having low control on protein orientation that may result in losing the conformation and function of the protein, being sensitive to the pH and harsh chemicals conditions, or having the difficulty to control the biotinylated or antigen site on the protein and lack site-specific protein ligation.

Enzymatic methods have the advantage of making specific bonds between protein and the surface. Transglutaminase (TGase) was used to immobilize recombinant proteins containing TGase substrate to casein surface that provides lysine to react with the TGase substrate (44,45). Proteins were also immobilized on the surface using two aldehyde-containing molecules as substrates for protein farnesyltransferase PFTase and as reactants in both oxime and hydrazone formation (46). Although these enzymatic approaches will provide specific protein ligations, they have a broad substrate specificity range and their substrate specificity is not clearly known. Therefore, sortase-mediated protein immobilization is advantageous because of using more specific substrates (34,47). Sortase A was used to immobilize proteins to polystyrene beads (34), beaded agarose (48), biacore sensor chip (49), and sepharose (50).

1-2-2-3- Patterning proteins on the surface

Various techniques such as lithography and microcontact printing have been used to generate patterns on the surface. Microcontact printing was used to immobilize proteins on surfaces via physical or chemical interactions. Protein micropatterning has been used for controlling cell interactions with biomaterial surfaces. Since proteins are soft materials, they

can be patterned under conditions that are compatible with many biological systems and it is relatively easy to pattern as desired. To date, mostly chemical and physical methods have been used to generate protein patterns on the surface. Protein and antibody microarrays, made using these techniques, were being used in determining antibody specificities, screening vaccines against viral target, analyzing protein-DNA interactions, and cell patterning (51,52).

In tissue engineering, micro-oriented ECM (Extra Cellular Matrix) dictates cell morphology and behavior. Therefore, protein patterning has a great importance in this area. Microcontact printing is used to generate squares and grid lines of fibronectin onto glass, which subsequently direct the self-assembly of fluid lipid bilayers onto the complementary, uncoated regions of the surface. Surfaces with mixed patterns of lipid bilayer and protein surfaces have also been made using microcontact printing. Cell spreading is correlated with both the geometry and dimensions of the fibronectin barriers (53) (54). Immobilizing and patterning antibodies on the surface has application in biosensor technology. Protein patterns were transferred using microcontact printing (55). Spatial organization of signaling complexes impact on cell communication is investigated via patterning anti-CD3 and anti-CD28. In T cell–APC pairs, more IL-2 is produced when CD28 clusters are segregated from central supramolecular activation cluster (cSMAC)-localized CD3 and into the IS periphery (56).

Protein patterning using sortase reaction provides more control over protein orientation and gives less background. Influenza virus A protein with a LPETG tag was immobilized on a G_3 functionalized surface in an array using sortase A-mediated immobilization (57).

1-2-2-4- Protein 3D assemblies and molecular printboards

Proteins are playing a vital role in cellular metabolism and function. Although proteins have different responsibilities according to their function, structure and location in the cell, they have been perfectly synchronized to accomplish the bigger goal of cell proliferation and communication. For instance, antibodies recognize foreign elements; growth factor receptors trigger the cellular response, membrane proteins import/export materials inside and outside the cell, and enzymes catalyzes variety of actions in the cell machinery. This great harmony between different proteins with different functions can be used in generating protein assemblies (58). These protein assemblies can be used as novel biocompatible materials in the development of biosensors, probes, bionanowires, and drug delivery systems (59,60). DNA assemblies were used widely because of the versatile base-pair rule to program DNA assemblies. However, proteins and peptides have properties such as target binding with complex molecular recognition mechanism that make them great building blocks of the nanobiological assemblies. Site-specific ligation of peptides and proteins to generate protein/peptide oligomers and 3D assemblies can affect the stability and shape of the protein assembly.

In nature, some proteins self-assemble into more complicated structures to create a biological function or to provide some structural features. Complex macromolecular assemblies carry out elaborate functions in cellular machinery. Structures of these assemblies have been evolved to result in the desired function such as macromolecular trafficking, motility, adhesion, signaling, sequestration, and defense. Studying the interactions between the subunits of macromolecular assemblies provide insight to design such complicated structures (61).

For instance, basement membrane assembly is composed of various proteins such as laminins, a few characteristic collagen and proteoglycan types and several other glycoproteins. These proteins can participate in other supramolecular assemblies, indicatingtheir high versatility (62). Programming complex nanoscale 3D assemblies composed of nanometric building blocks Nature has many good examples of self assembly of peptide and proteins. S-layer proteins organize their assembled structure on membranes by molecular recognition of proteins and nucleation process. For example, collagens are assembled from triple helix peptides in micron-size with precise recognition between peptides. By mimicking this assembly process, it is possible to build up micron-scale materials that still feature excellent nano-scale ensembles, which essentially bridges the nano-world and the micro-world. The conformation change of peptides and proteins can be applied for smart responsive materials. For example, protein assembly can undergo structural change with variations in pH, ionic strength, temperature, electric/ magnetic fields (58).

To date, a molecular printboard is made of a monolayer of host molecules on a solid substrate on which guest molecules can be attached with control over position, binding strength, and binding dynamics. β -cyclodextrin (β-CD) was used in generating surface patterns and molecular printboards as a general platform for the immobilization of proteins through small multivalent, orthogonal linker molecules through host-guest chemistry. Guest molecules (for example, adamantane and ferrocene derivatives) bind to these host surfaces through supramolecular, hydrophobic inclusion interaction. Multivalent interactions are exploited to tune the binding strength and dynamics of the interaction of guest molecules with the printboard (63-65). As mentioned above, using sortase transpeptidation reaction to

generate these assemblies has advantages over chemical methods. Sortase-mediated protein ligation is a simple, robust method that can site-specifically and covalently binds layers of protein to each other and the functionalized surfaces using mild reaction conditions. Therefore, we can overcome issues of making 3D protein assemblies such as lack of control over homogeneity of layers due to non-specific binding of proteins to each other and the surface.

1-2-3- Engineering sortase using yeast surface display and directed evolution

Directed evolution is a powerful technique to engineer proteins, and yeast surface display (YSD) has been widely used to engineer proteins since its first invention (66,67). Yeast surface display technique use the linkage between Aga2p and Aga1p (a GPI/β-1,6-glucan-anchored protein) to display recombinant proteins fused to Aga2p protein on the cell wall of *Saccharomyces cerevisiae.* A library of protein mutants displayed on the yeast surface can be screened quantitatively using FACS (Fluorescent-Activated Cell Sorting). YSD has many advantages over other molecular displaying techniques such as using eukaryote post-translational modification or the ability to efficiently recombine homologous DNA sequences (68,69).

CHAPTER 2

PROTEIN IMMOBILIZATION ON THE SURFACE

2-1- Introduction

Protein immobilization on the surface has various applications in biosensors, tissue and stem cell engineering and bioenergy (63). The conductivity and surface plasmon resonance properties of gold make it a great platform for biosensors. Many techniques that are detecting biomolecular interactions such as Surface Plasmon Resonance (SPR), Quartz Crystal Microbalance (QCM) and electrochemical biosensors require protein immobilization on the gold surface. Binding proteins in a specific direction to the gold chip has a great impact on the availability of the binding site and therefore on the signal obtained from the protein-protein interactions (70,71). Having control over protein orientation is also important in immobilizing photoconductive proteins such as PhotosystemI (PSI) on the gold surface, since these proteins transfer the electron in one direction, from the P700 reaction center to the *F*A/*F*B sites (72).

Different techniques have been used to immobilize ligands on the gold surface such as chemical or physical methods or engineering the protein for an exposed cysteine (37,43,72-80). Physical method of protein immobilization on the surface has no control over protein orientation and lacks stability. Chemical methods provide more stable protein ligation, although these methods have disadvantages of non-specific binding to the surface caused by availability of more than one reactive site in the protein structure, or reagents and conditions

that are not always compatible with protein stability and functionality. Introducing a cysteine in the amino acid sequence to generate the Au-S bond between engineered proteins and the gold surface, or disulfide bond between protein layers requires mutating other exposed Cys in the amino acid sequence, which might affect protein structure and therefore stability and functionality of the protein. Therefore, an enzymatic method can be used to immobilize and oligomerize proteins covalently with the advantage of using a site-specific ligation in conditions close to neutral pH without exploiting any harsh chemical reagents. Sortase A is a transpeptidase that recognizes a five amino acid sequence near the C-terminus of the protein and ligates that to the N-terminus of a short peptide (two to five amino acids). These short substrate tags can be used in a sortase reaction to immobilize proteins to the surface or ligate them together site-specifically, without altering the protein structure and function. Sortase A-mediated protein immobilization was used to immobilize GFP (Green Fluorescent Protein), and PEG (polyethylene glycol) to agarose and polystyrene beads, an antibody to carboxymethylated biacore sensor chip, glycosyltransferase to the sepharose surface, recombinant thrombomodulin, and Influenza virus on the modified glass slides, recombinant RFP (Red Fluorescent Protein) to the modified polystyrene microparticles. Sortase-mediated immobilization was also used to biotinylate the surface (34,49,57,77,81,82).

Here we used sortase-mediated protein immobilization to immobilize recombinant fluorescent proteins on the gold surface in a layer or multiple layers. Two different Sortase A with slightly different substrate specificity have been used. *S. aureus* Sortase A (Sa Srt) is specific toward LPETG and GGG substrate pair. Sa Srt cleaves between threonine and glycine in the LPXTG sequence and makes an acyl-enzyme intermediate that can be

resolved by a nucleophile attack of the amine in a triglycine sequence (7,83,84). *S. pyogenes* Sortase A (Sp Srt) is more specific toward LPETA and AAA substrate pair, although LPETG and GGG are less efficient substrates for Sp Srt reactions. Substrate specificity of these two sortases was first determined by oligomerization reactions of recombinant bifunctional proteins using Sp and Sa Srt tags at two termini. Sp Srt and Sa Srt were then used to immobilize recombinant GFP and mcherry with LPETG or LPETA sequence at the C-terminus on the gold surface modified with 3G/A peptide, Sa and Sp Srt substrates, respectively. Protein immobilization was investigated using fluorescence and atomic force microscopy.

2-2-Material and methods

2-2-1-Protein production and purification

GFP-LPETG plasmid was used from the cited reference (34). The mcherry variants were made by PCR using "pJK148-Pbip1-signal peptide-linker-mcherry-AHDL" (from Dr. Roger Tsien lab, Caltech) plasmid as a template. *S. aureus* sortase A was cloned out of the pHTT27 plasmid (provided by Dr. Olaf Schneewind, University of Chicago). Both constructs were inserted into a pET-26 vector using NdeI and XhoI restriction sites. A hisx6-tag sequence is located after the multiple cloning sites in the pET-26 vector. In sortase plasmid (pMR5), a hisx6-tag sequence is inserted at the N-terminus and a stop codon is placed before XhoI restriction site to prevent expression of the second hisx6-tag at the C-terminus (Appendix 3). An evolved Sa Srt with higher activity toward LPETG substrate, Srtm4 were made using directed mutagenesis. Four mutations D160N, K137T, D165A,

P94S (clone 8.3 in ref (85)) were made using Quickchange Site-Directed Mutagenesis (pMR5m4). Mutagenesis primers were designed following the primer design direction in Quickchange Site-Directed Mutagenesis kit (Stratagene). Phusion polymerase was used to generate the mutated plasmid by PCR. 1 µl (20 Units) of DpnI (NEB) was added to the PCR product to digest the methylated, non-mutated plasmid and incubated at 37 °C for 1 hour. The plasmids then transformed to XL1-Blue cells (Stratagene) following the company protocol.

Proteins were expressed in BL21 (DE3) cells (Invitrogen). An overnight cell culture was diluted (1:50-1:100) in 250ml-1000ml of LB Kan (50 µg/ml) media. The cells were grown to OD600 of 0.5-0.7 at 37 °C shaker and then induced by adding IPTG to the final concentration of 1mM and pelleted after 3-4 hours of induction. Srtm4 protein was induced by adding IPTG to a final concentration of 0.5 mM and incubating for 3 hours in 30 °C shaker. The cells were frozen at -20 °C overnight. Proteins were extracted after thawing the cells on ice by resuspending them in 10 ml B-per reagent (Thermo Scientific) and 10 µl DNaseI (NEB) was added to 250 ml of the culture following the B-per manual. Soluble proteins can be found in the supernatant. Proteins were purified using his-tag purification by Talon resin (Clontech), following batch/gravity user manual. The purified protein was then dialyzed using dialysis cassette with MW cutoff of 10,000 (Thermo Scientific) or buffer exchanged against TBS buffer (pH=7.5) using Amicon filters (Millipore) with MW cutoff of 10,000 to remove imidazole. Protein concentration was measured using A280 measurement with a Bio-Rad spectrophotometer. 10 to 40 mg of the protein can be produced from 1L culture. Proteins were concentrated to 500 to 1200 µM. The concentrated proteins can be

stored in aliquotes for a few months at -20 °C for future use. For short-term storage, proteins were stored at 4 °C.

2-2-2-Peptide assembly on the gold surface

Gold-coated microscope slides with 100 Å gold thickness were purchased from Platypus. Slides were cut into 5x5 or 10x10 mm pieces and washed with RCA solution (ammonium hydroxide: hydrogen peroxide: H_2O; 1:1:5 ratio) at 70 °C for at least 10 minutes, then rinsed with copious amount of dI H_2O and dried under a nitrogen stream. The GGGC, AGGC and AAAC peptides were purchased from Genscript with >%95 purity. Dimethylformamide (DMF) was added drop-wise to the peptide powder until the solution was clear, then oxygen free water (to prevent disulfide formation between Cys) was added to the solution to make 100 μM – 1mM solutions. The clean gold slides were submerged in the peptide solution 1 to 3 nights. After that the slides were washed with TBS buffer (pH=7.5), and dI H_2O and dried under nitrogen.

2-2-3-Protein immobilization on the peptide SAM on gold

In general, 20-40 μM of mcherry/GFP-LPETG was mixed with 15 μM Sa Srt in a reaction buffer (TBS, 10mM $CaCl_2$, pH=7.5). 50-80 μl of the reaction mixture was put on the functionalized gold slides with the peptide and kept in a moist sealed mini petri dish for at least 3 hr at 37 °C incubator. The slides were washed after the reaction with filtered TBS (50 mM Tris, 150 mM NaCl, pH=7.5) and dIH_2O and dried under nitrogen. For immobilizing layers of proteins, after washing the slides with the first protein layer immobilized on the surface several times with TBS buffer and dIH2O and drying the slide under the nitrogen, second round of sortase reaction was done similarly, followed by washes and drying.

2-2-4-Fluorescence Microscopy

Gold slide pieces were mounted on the glass microscope slide. The fluorescence micrographs were obtained using Olympus or Nikon fluorescent microscopes with 40x magnification. The images were analyzed with ImageJ. FITC filter cube (Ex: 494 , and Em: 518) and Texas Red HQ cube (Ex: 577, Em: 620) in Nikon microscope and GFP and RFP filters have been used for detecting GFP or mcherry fluorescence. Mean Fluorescence Intensity (MFI) was determined using Analyze> measure tool in ImageJ software on the whole slides. The background (negative control) value was subtracted from the actual sample before normalization or comparison.

2-2-5-Atomic Force Microscopy

The AFM images were obtained using Asylum Research AFM using tapping mode. A silicon nitride cantilever with 10 nm radius, spring constant of 2 N/m and frequency of 70 kHz was used (AC240TS from Asylum Research). AFM was done on dried samples in a tapping mode. Igor or Gwyddion software was used to analyze the data.

2-2-6-Western blot

20 µl of the denatured proteins plus 20 µl of SDS-PAGE sample buffer were run on a 10% Bis-Tris NuPAGE gel (Invitrogen) using MOPS buffer (Invitrogen). 8 µl of the single color protein marker (LiCor) was added to the first lane. The proteins were transferred into a 0.45 µm nitrocellulose membrane (Bio-rad) using transfer buffer (Invitrogen). The membrane is blocked either for 2 hr at room temperature or overnight at 4°C on orbital shaker in 1:1 volume ratio of blocking buffer (Licor) to PBS buffer. The blocked membrane was then labeled with 30ml of primary antibody (1:1000 dilution); mouse anti-GFP (Iowa lab), mouse

anti-Hisx6 (Covance or Iowa lab) or rabbit anti-DsRed (Clontech) and secondary antibody; anti-mouse IR800 or anti-rabbit IR680 (Odyssey). The images were taken using the Odyssey Infrared Imaging system.

2-3-Results and discussion

2-3-1- GFP-LPETG immobilization on the gold surface using Sa Srt

We used sortase-mediated protein immobilization to site-specifically immobilize proteins on the gold surface using a simple two-step strategy. First, we generated a self-assembly layer of a GGG or AAA peptide with a cysteine at the C-terminus (GGGC or AAAC). The sulfhydryl group in Cys makes Au-S bond with the gold surface (Fig 2-1). Fluorescent proteins were used as model proteins, since the folded protein can be easily detected using fluorescence microscopy.

Fluorescent microscopy images showed that Sa Srt attached mcherry-LPETG and GFP-LPETG to the GGG modified gold surface, but Sa Srt was not able to immobilize mcherry-LPETA on the AAA modified gold surface. Sp Srt immobilized both mcherry-LPETG and mcherry-LPETA on the modified GGG and AAA gold surface respectively, although the mcherry-LPETA immobilization was more efficient. Inactive sortase mutant (C184G) was not able to bind GFP or mcherry-LPETG to the surface (Fig. 2-2).

The immobilized fluorescent protein was stable and could not be removed by several PBS and water washes, and could still be detectable after a few weeks. There was a little fluorescence due to non-specific binding of the fluorescent protein to the bare gold surface, or when the surface was functionalized with AGGC peptide, known to be a poor substrate for Sa Srt (10) (Figure 2-2).

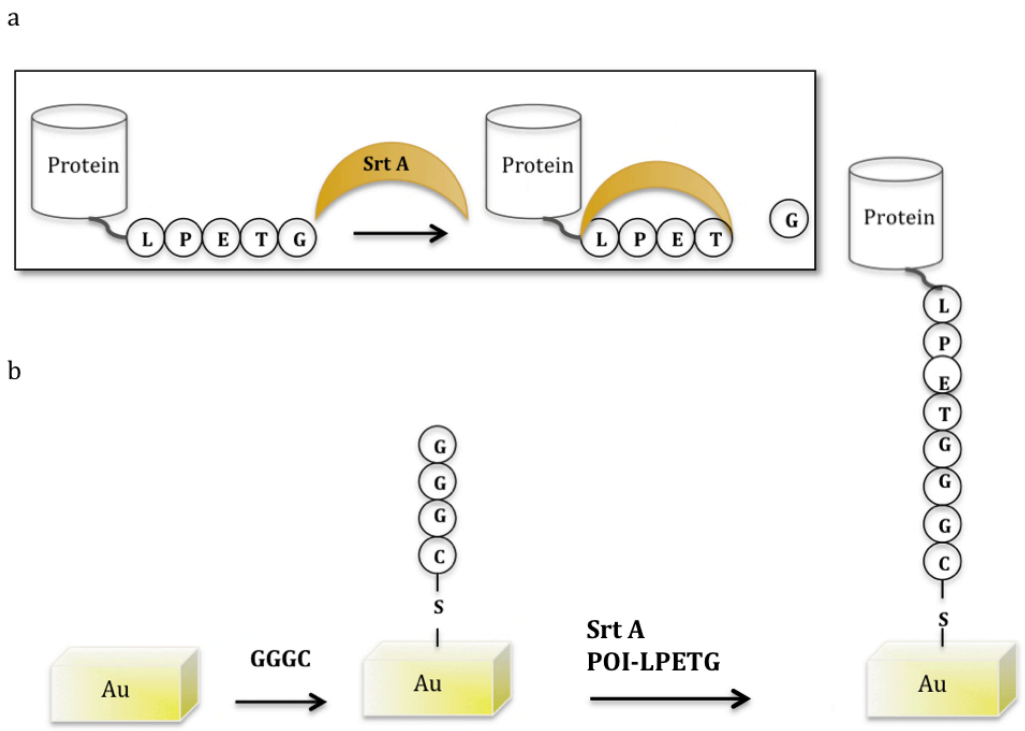


Figure 2-1. Sortase-mediated immobilization of proteins on gold- a) Srt cleaves between T and G in the LPETG sequence at the C-terminus of the recombinant Protein Of Interest (POI) Srt-Protein Of Interest- LPETG intermediate is generated. b) GGGC self-assembly layer on the gold surface is made by formation of Au-S bond between sulfhydryl group of Cys and gold. POI-LPETG binds to the surface site-specifically using *S. aureus* sortase A transpeptidation reaction.

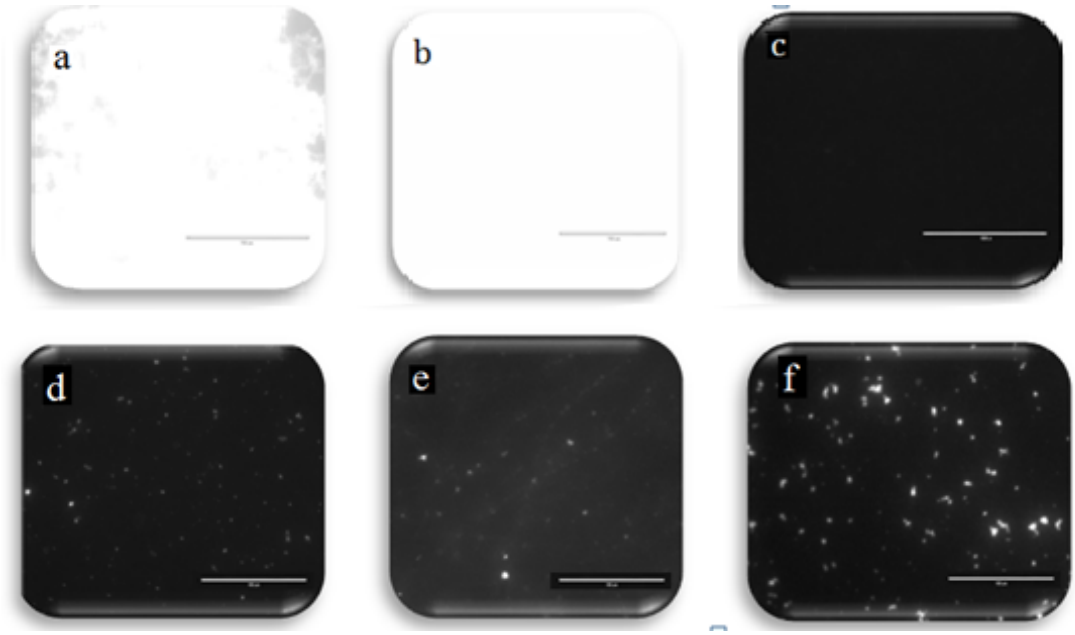


Figure 2-2. Fluorescent micrographs of immobilized protein on the gold surface using Sortase A. a. 20 µM mcherry-LPETG immobilized on GGGC modified gold surface using Sa Srt, b. 25 µM GFP-LPETG immobilized on GGGC modified gold surface using 12 µM *S. aureus* Sortase A, c. GFP-LPETG and inactive sortase, d. GFP-LPETG on GGGC gold slide without Srt, e. GFP-LPETG binding to bare gold, f. GFP-LPETG binding to the AGGC modified gold slides using Sa Srt. The scale bar is 100 µm.

To investigate whether the observed fluorescence is not due to the immobilization of the acyl-enzyme intermediate (GFP-LPET-Sortase) instead of the actual product (GFP_LPETGGG), and to optimize the sortase-mediated ligation reaction conditions, we used different molar concentration ratios of Srt A to GFP-LPETG: 1:10, and 1:1. The fluorescent micrographs did not show an increase in fluorescence upon increasing the sortase concentration in the reaction mixture. The mean fluorescent intensity was increased

by 68% with decreasing the amount of Srt by 10 times molar concentration in the reaction mixture. If the Srt-GFP intermediate was immobilized on the surface, there should have been a decrease in intensity upon decreasing the sortase concentration (Fig 2-3).

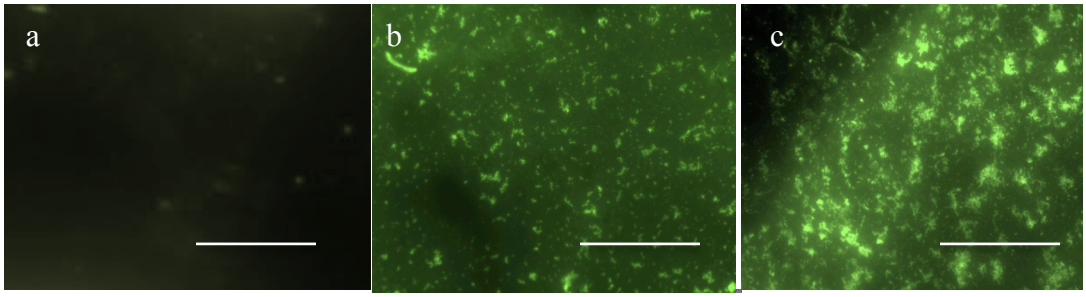


Figure 2-3. The impact of enzyme: substrate ratio on sortase-mediated protein immobilization. GFP-LPETG (30 μM) reaction mixture was added to GGGC modified gold slides; (a) without sortase; MFI: 17.148, (b) with 1:1 concentration ratio of Sa Srt: GFP; MFI: 39.214, (c) with 1:10 ratio Sa Srt: GFP; MFI: 54.436. The scale bar is 100 μm. MFI (Mean Fluorescent Intensity) was calculated using ImageJ.

Protein attachment and surface coverage were investigated using non-contact Atomic Force Microscopy (AFM). As it is shown in Figure 2-4, the gold surface is covered with a layer of GFP. Since the LPETG tag is at the C-terminus of GFP, close to one of the circular sections of the GFP barrel, the height of the immobilized GFPs are expected to be close to the height of the cylinder. GFP has a barrel-like three-dimensional structure with a diameter of about 2.4 nm and a height of 4.2 nm. AFM measures the height of the sample with a high accuracy, while width and length values are larger than actual values (86). The height curve resulted from height values between two points showed that the height of the covered surface is close to a layer of a GGGC peptide (about 0.6 nm) and GFP height (4.2 nm)

(Figure 2-4.c). The bare gold surface and the gold surface after peptide self-assembly were also characterized with AFM (Appendix 1). The RMS values for bare gold, peptide SAM and GFP layer immobilized on the self-assembled peptide layer is shown in table 2-1. Comparing the theoretical value for height GFP and peptide (4.8 nm) and RMS values for GFP layer on the gold surface (1.76 nm), the protein coverage is estimated by dividing the RMS value to the theoretical value as 36.7%.

2-3-2- Protein immobilization on the surface using two different sortases

We used Sp and Sa Srt protein immobilization to site-specifically immobilize fluorescent proteins on the gold surface. First, we generated a self-assembly layer of a GGG or AAA peptide with a cysteine at the C-terminus (GGGC or AAAC). The fluorescent proteins were immobilized on the gold-coated microscope slides that have been functionalized with GGGC or AAAC peptide using Sortase A transpeptidase reaction. Two recombinant Fluorescent Protein, eGFP and mcherry, with a LPETG or LPETA sequence followed by a hisx6 tag at the C-terminus were made and purified using Talon cobalt resins. Either wild-type Sa Srt or the evolved Sa Srt with four mutations (Srtm4) was used in this study.

Fluorescent microscopy images showed that Sa Srt attached mcherry-LPETG and GFP-LPETG to the GGG modified gold surface, but Sa Srt was not able to efficiently immobilize mcherry-LPETA on the AAA modified gold surface. Sp Srt immobilized both mcherry-LPETG and mcherry-LPETA on the modified GGG and AAA gold surface, respectively. Inactive sortase mutant (C184G) was not able to bind GFP or mcherry-LPETG to the surface (Fig. 2-5). The immobilized fluorescent protein was stable and could not be removed by several PBS and water washes, and was still detectable after a few weeks.

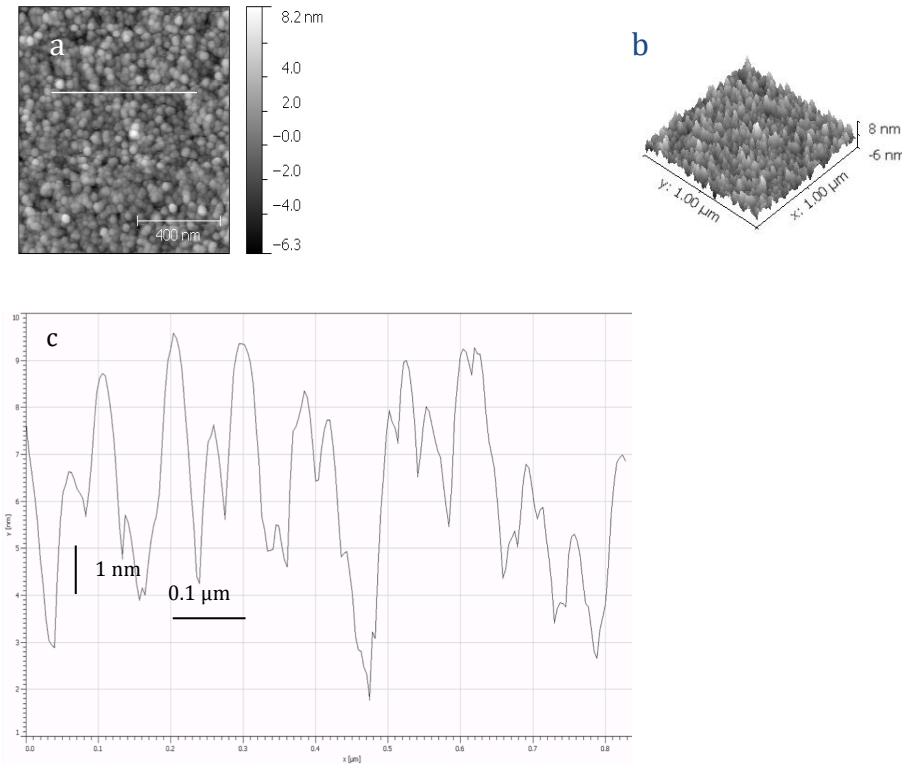


Figure 2-4. Atomic Force Microscopy images of GFP-LPETG immobilized on the gold surface using Sortase A reaction. GFP-LPETG was immobilized on GGGC modified gold slides using Sa Srt reaction. AFM was done in tapping mode using a silicon nitride cantilever (AC240). a) The RMS value is 1.8 nm. The Scale bar is 400 nm. b) 3D image, d) the height curve that shows the height values of the line that was drawn on image b. The peaks heights range between 4-6 nm.

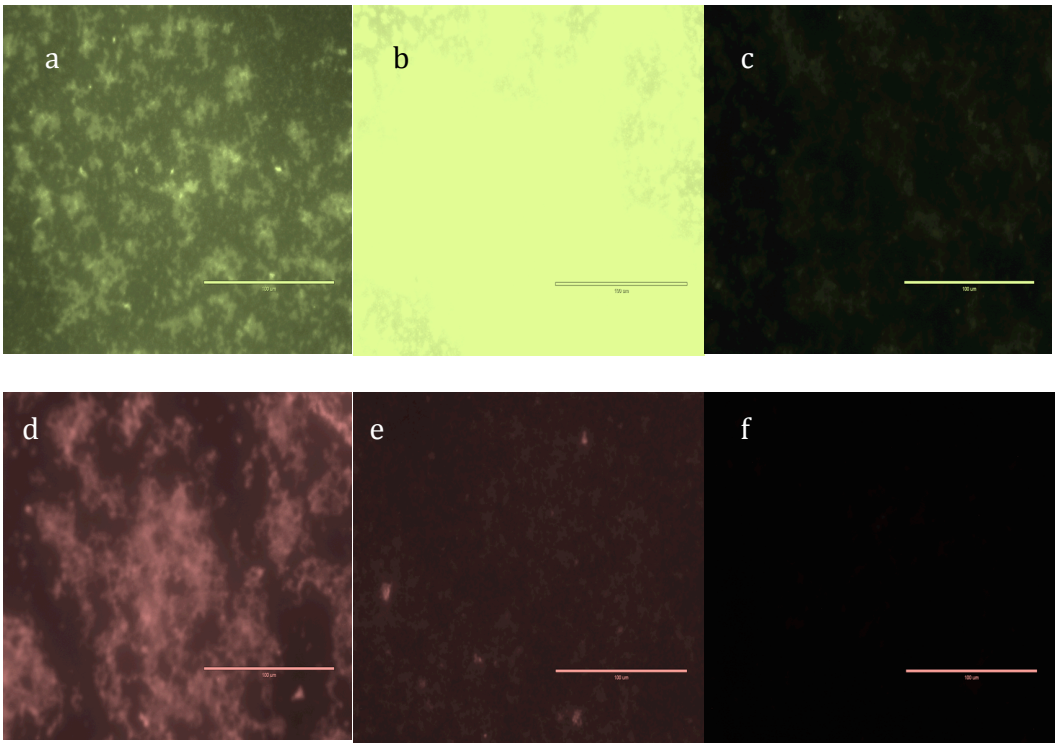


Figure 2-5. Fluorescent protein immobilization on the gold surface. a-c) GFP-LPETG immobilization on the self-assembled layer of GGGC on gold , d-f) AAA- mcherry-LPETA immobilization on self-assembled layer of AAAC on gold. a,d) using 25 µM S. pyogenes Sortase A. b,e) 15 µM S. aureus Sortase A. c,f) using inactive S. aureus Sortase A. Scale bars are 100 µm.

2-3-3- Protein immobilization and oligomerization on gold

We immobilized bifunctional proteins with two specific sortase tags at two ends (AAA-mcherry-LPETA and GGG-GFP-LPETG) using Sp and Sa Srt on the gold surface with AAAC and GGGC peptides assembled on the surface. The recombinant fluorescent proteins ligation in solution using Srtm4 or Sp Srt is shown in Figure 2-6. Sa Srt catalyzed GGG-GFP-LPETG oligomerization reaction up to the formation of tetramers, while it was not able to make AAA-mcherry-LPETA dimer. Sp Srt made GGG-GFP-LPETG or AAA-mcherry-LPETA dimer (Fig. 2-6-a). AFM images showed that Sp Srt immobilized AAA-mcherry-LPETA oligomers on the gold surface (Fig 2-7). Three different regions can be found on the surface with heights corresponding to mcherry monomer, dimer, and trimer on the gold surface.

To further investigate the presence of protein oligomers immobilized on the surface, we removed protein oligomers from the surface using 2-mercaptoethanol (bME), a reducing agent that breaks gold thiolate (Au-S) bond. 50 µl of a mixture of TBS buffer and 14% bME (1:1 volume ratio) was placed on the sample slide for at least 2 hours at room temperature. Then the drop was carefully removed after re-suspending the protein the drop. The denatured protein was run on SDS-PAGE gel and transferred to a nitrocellulose membrane and blotted against mouse anti-GFP (Iowa lab- hybridoma bank) and anti mouse IR-800 (odyssey) antibodies as explained in material and methods. Two bands corresponding monomer and dimer was observed from the proteins removed from the surface (Fig 2-8).

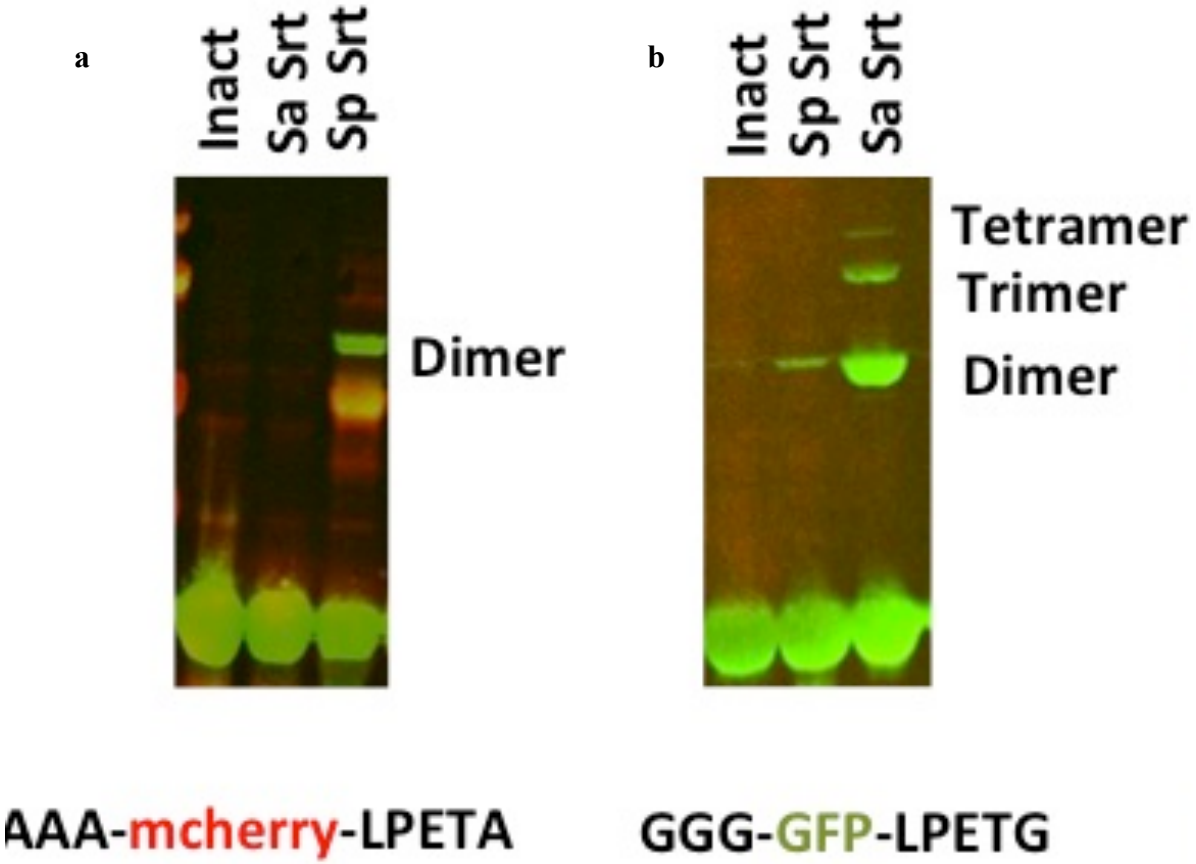
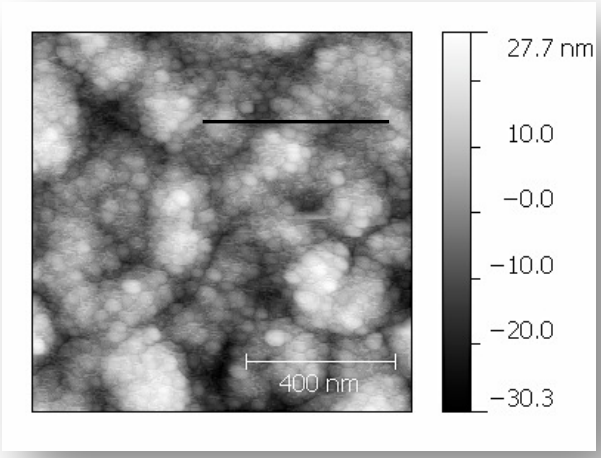


Figure 2-6. Sp and Sa Srt reactions in solution. Western blot against GFP or mcherry. a) AAA-mcherry-LPETA without Srt, with Sa Srtm4, and with Sp Srt. b) GGG-GFP-LPETG without Srt, with Sp Srt and with Sa Srtm4. Oligomers are shown on the images.



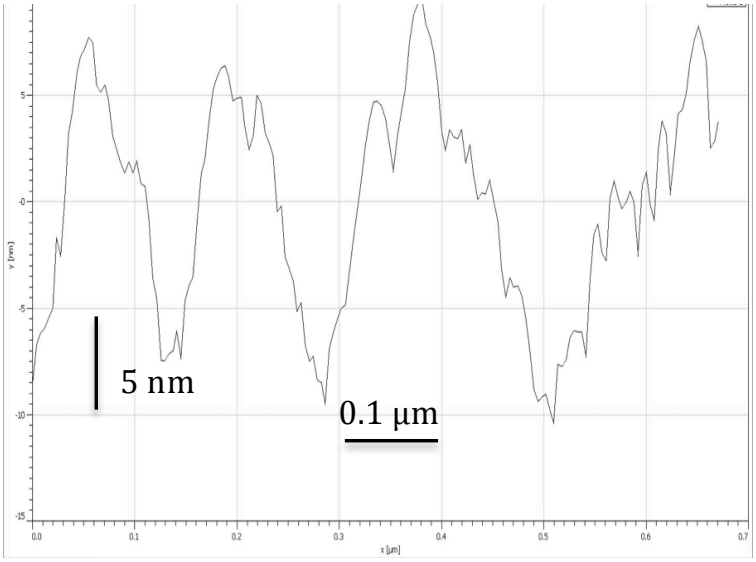


Figure 2-7. AFM image of immobilized protein oligomers on the gold surface. AAA-mcherry-LPETA was immobilized on the surface using *S. pyogenes* Sortase A. The scale bar is 400 nm. The curve is corresponding to the line drawn on the surface was shown in the bottom.

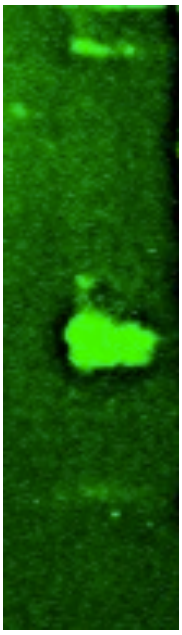


Figure 2-8. Protein removal from the gold surface after sortase reaction. Immobilized protein (GGG-GFP-LPETG) was removed from the gold surface by reducing agent. Western blot was done against GFP using mouse anti-GFP. GFP monomer and dimer is shown in the image.

Table 2-1- AFM parameters for protein immobilization on gold

	RMS (nm)
Bare Gold	0.54
Peptide	1.11
GFP-LPETG	1.76
AAA-mcherry-LPETA	8.9

CHAPTER 3

GENERATING PROTEIN 3D ASSEMBLIES USING ORTHOGONAL SORTASES

3-1-Introduction

Protein assemblies are important elements in biological processes. They attracted major attention as they can be used as novel biocompatible material. Protein assemblies are made of peptides and proteins building blocks, which can be used to generate nanotubes, scaffolds, and nanowires. These protein nanostructures have been used in the development of biosensors, probes, bionanowires, and drug delivery systems (59,60). It is important to have a control over protein assembly. Site-specific protein/peptide ligation can direct the stability and shape of the protein assembly. Site-specific ligation of peptides and proteins to generate protein/peptide oligomers and 3D assemblies has many applications in nano biotechnology. As previously discussed in background and Chapter 2, sortase-mediated ligation can be used as a powerful tool to generate such assemblies. Proteins were conjugated to the self-assembled peptides in hydrogels using sortase-mediated site-specific ligation (87).

We tried two different ways to generate protein assemblies site-specifically and in a controlled manner. In one approach, the N-terminus of triglycine (GGG) of the bifunctional sortase substarte (3G at N-terminus and LPETG at C-terminus) was protected with a protease cleavable site. Therefore, sortase-mediated oligomerization is dependent upon

removal of the protecting group. In the second approach, two sortases with slightly different substrate specificity were used to sequentially ligate proteins with specific tags to each other. As a proof of concept, we first generated protein wires in solution.

3-2-Material and methods

3-2-1-Protein production and purification:

GFP-LPETG plasmid was used from the cited reference (34). The mcherry vaiants were made by PCR using "pJK148-Pbip1-signal peptide-linker-mcherry-AHDL" (from Dr. Tsien lab) plasmid as a template and inserted into pET26 vector backbone using NdeI and XhoI. Sortase A from (*S. aureus*) was cloned out of pHTT27 plasmid (provided by Dr Olaf Schneewind, University of Chicago). Both constructs were inserted into a pET vector by introducing NdeI and XhoI restriction sites. A hisx6-tag was located at the C-terminus of the construct after XhoI site. Sortase A plasmid (pMR5) has a hisx6-tag at the N-terminus. There is stop codon before XhoI site to inhibit the C-terminus hisx6-tag. Srtm4 is an evolved Sa Srt with higher activity toward LPETG substrate. Four mutations D160N, K137T, D165A, P94S (clone 8.3 in ref (85)) were made by site-directed mutagenesis (pMR5m4). Mutagenesis primers were designed following the primer design direction in Quickchange (Strategene). Phusion polymerase was used to generate the mutated plasmid by PCR. 1 µl of DpnI was added to the PCR product on ice to digest the methylated, non-mutated plasmid and kept at 37 °C for 1 hour. The plasmids then transformed to XL1-Blue cells (Stratagene) following the company protocol.

Proteins were expressed in BL21 (DE3) cells (Invitrogen). An overnight cell culture was diluted (1:50-1:100) in 250ml-1000ml of LB Kan media. The cells were grown to OD600 of 0.5-0.7 at 37 °C shaker and then induced by adding IPTG to the final concentration of 1mM and pelleted after 3-4 hours of induction. Srtm4 protein was induced by adding IPTG to a final concentration of 0.5 mM and incubating for 3 hours in 30 °C shaker. The cells were frozen at -20 °C overnight. Proteins were extracted after thawing the cells on ice by resuspending them in 10 ml B-per reagent (Thermo Scientific) and 10 µl DNaseI (NEB) was added to 250 ml of the culture following the B-per manual. Soluble proteins can be found in the supernatant. Proteins were purified using his-tag purification by Talon resin (Clontech), following batch/gravity user manual. The purified protein was then dialyzed using dialysis cassette with MW cutoff of 10,000 (Thermo Scientific) or buffer exchanged against TBS buffer (pH=7.5) using Amicon filters (Millipore) with MW cutoff of 10,000 to remove imidazole. Protein concentration was measured using A280 measurement. 10 to 40 mg of the protein can be produced from 1L culture. Proteins were concentrated to 500 to 1200 µM concentration, and aliquoted and freezed for a few months at -20 °C for future use. For short-term storage, proteins were kept at 4 °C.

3-2-2-Peptide assembly on the gold surface

Gold-coated microscope slides with 100 Å gold thickness were purchased from Platypus. Slides were cut into 5x5 or 10x10 mm pieces and washed with RCA solution (ammonium hydroxide: hydrogen peroxide: H_2O; 1:1:5 ratio) at 70 °C for at least 10 minutes, then rinsed with copious amount of dI H_2O and dried under a nitrogen stream. The GGGC and AAAC peptides were purchased from Genscript with >%95 purity. Dimethylformamide (DMF) was

added drop-wise to the peptide powder until it was dissolved, then oxygen free water (to prevent disulfide formation between Cys) was added to the solution to make 100 μM – 1mM solutions. The clean gold slides were submerged in the peptide solution 1 to 3 nights. After that the slides were washed with TBS buffer (pH=7.5), and dI H_2O and dried under nitrogen.

3-2-3-Protein immobilization on the peptide SAM on gold

Sortase reaction conditions are explained in the results section, but in general the sortase reaction condition is as follows. 20-40 μM of mcherry/GFP-LPETG was mixed with 15 μM Sa Srt in a reaction buffer (TBS, 10mM $CaCl_2$, pH=7.5). 50-80 μl of the reaction mixture was put on the functionalized slides with the peptide and kept in a moist sealed mini petri dish at least 3 hr at 37 °C incubator. The slides were washed the next day with filtered TBS (pH=7.5) and dIH2O and dried under nitrogen. For immobilizing layers of proteins, after washing the slides with the first protein layer immobilized on the surface several times with TBS buffer and dIH2O and drying the slide under the nitrogen, second round of sortase reaction will be done similarly, followed by washes and drying.

3-2-4-Fluorescence Microscopy

Gold slide pieces were mounted on the glass microscope slide using superglue. The fluorescence micrographs were obtained using Olympus or Nikon fluorescent microscopes with 40x magnification. The images were analyzed with ImageJ. FITC filter cube (Ex: 494 , and Em: 518) and Texas Red HQ cube (Ex: 577, Em: 620) in Nikon microscope and GFP and RFP filters have been used for detecting GFP or mcherry fluorescence. Mean Fluorescence Intensity (MFI) was determined using Analyze> measure tool in ImageJ

software on the whole slides. The background (negative control) value was subtracted from the actual sample before normalization or comparison.

3-2-5-Atomic Force Microscopy

The AFM images were obtained using Asylum Research AFM using tapping mode. A silicon nitride cantilever with 10 nm radius, spring constant of 2 N/m and frequency of 70 kHz was used (AC240TS from Asylum Research). AFM was done on dried samples in a tapping mode. Igor or Gwyddion software was used to analyze the data.

3-2-6-Western blot

20 µl of the denatured proteins plus 20 µl of SDS-PAGE sample buffer were run on a 10% Bis-Tris NuPAGE gel (Invitrogen) using MOPS buffer (Invitrogen). 8 µl of the single color protein marker (LiCor) was added to the first lane. The proteins were transferred into a 0.45 µm nitrocellulose membrane (Bio-rad) using transfer buffer (Invitrogen) for 1 hr at 33V. The membrane is blocked either for 2 hr at room temperature or overnight at 4 C on a orbital shaker in 1:1 blocking buffer (Licor) and PBS buffer. The blocked membrane was then labeled with primary mouse anti-GFP (Iowa lab), anti-His (Iowa lab) or rabbit anti-DsRed (Clontech) and secondary antibody; anti-mouse IR800 or anti-rabbit IR680 (Licor-Odyssey). The images were taken using the Odyssey Infrared Imaging system.

3-3- Result and Discussion

3-3-1- Generating protein 3D assemblies using orthogonal sortases

As discussed in Chapter 2, protein wires can be immobilized on the surface using sortase reaction. However, there are two issues in immobilizing layers of proteins on the surface in a controlled manner using bifunctional proteins and sortase reaction: 1) The Sortase A reaction product obtained from catalyzing bifunctional proteins with two specific sortase substrate tags at the N- and C-termini, LPETGGG for Sa Srt or LPETAAA for Sp Srt, is also a substrate for the Sa or Sp Srt A which makes the oligomerization reaction less efficient by cleaving the immobilized protein off the surface. 2) There is no control over the homogeneity of each immobilized layer on the surface, which means there might be multilayers in one region, while there is just one layer of protein in another area on the surface. To control these issues, orthogonal sortases can be used to immobilize protein layers on the surface sequentially. Sortase-mediated protein immobilization was used to immobilize recombinant fluorescent proteins on the gold surface in a layer or multiple layers. Two different sortases with slightly different substrate specificities have been used to immobilize proteins in layers in a controlled manner. *S. aureus* Sortase A (Sa Srt) is specific toward LPETG and GGG substrate pair. Sa Srt cleaves between threonine and glycine in the LPXTG sequence and makes an acyl-enzyme intermediate that can be resolved by a nucleophile attack of a triglycine sequence (7,83,84). *S. pyogenes* Sortase A (Sp Srt) is more specific toward LPETA and AA substrate pair. In order to immobilize protein layers on the surface with having control over protein orientation, oligomerization site, and homogeneity of each layer, Sp and Sa Srt A were used sequentially to ligate proteins to the peptides and proteins on the surface site-specifically. Sp Srt reaction was used to immobilize a

bifunctional mcherry (GGG-mcherry-LPETA) on a AAA modified gold surface. Sa Srt was then used to ligate the GFP-LPETG layer to the immobilized mcherry layer. These sequential Sortase A site-specific ligations are using the slightly different substrate specificity of Sa and Sp Srts. Since Sp Srt is more specific toward LPETA and AAA pair, the ligation of the GGG-mcherry-LPETA to the AAA peptide on the surface is favored over the oligomerization of GGG-mcherry-LPETA protein which requires the GGG tag to act as a nucleophile substrate for Sp Srt reaction. On the other hand, the GFP-LPETG ligation reaction using Sa Srt is favored toward catalyzing the amid bond between GFP-LPETG and the the GGG tag, at the N-terminus of the GGG-mcherry-LPETA protein, which generates a second layer of protein rather than immobilizing the GFP on the surface.

Sp Srt is more specific toward the LPETA and AAA sequences at the C- and N- terminus, respectively, although the LPETG or GGG tags work less efficiently. While Sa Srt is more specific toward the LPETG and GGG sequences at the C- or N-termini, and does not cleave the bond between threonine and alanine in LPETA. Therefore, Sp Srt can be used to immobilize the first layer of protein having a LPETA tag at the C-terminus and a GGG tag at the N-terminus. The N-terminus of the GGG sequence of the immobilized bifunctional protein can be used as a nucleophile substrate for immobilizing a layer of GFP-LPETG on top of the immobilized mcherry layer using Sa Srt (Fig. 3-1).

Protein ligation reactions were first done in solution. Two bifunctional fluorescent proteins with two specific Sp and Sa Srt substrate tags at N- and C-terminus, AAA-mcherry-LPETA and GGG-GFP-LPETG, were expressed and purified. Orthogonal Srt reactions on these two substrates were shown in Chapter 2. To investigate Sp and Sa Srt specificity in presence of all four tags, LPET(G/A) and 3(G/A), both substrates were added to the reaction mixture

containing Sa Srt or Sp Srt or both. Reaction products were run on the SDS-PAGE gel and transferred to a nitrocellulose membrane to be analyzed by western blot using antibodies against mcherry and GFP.

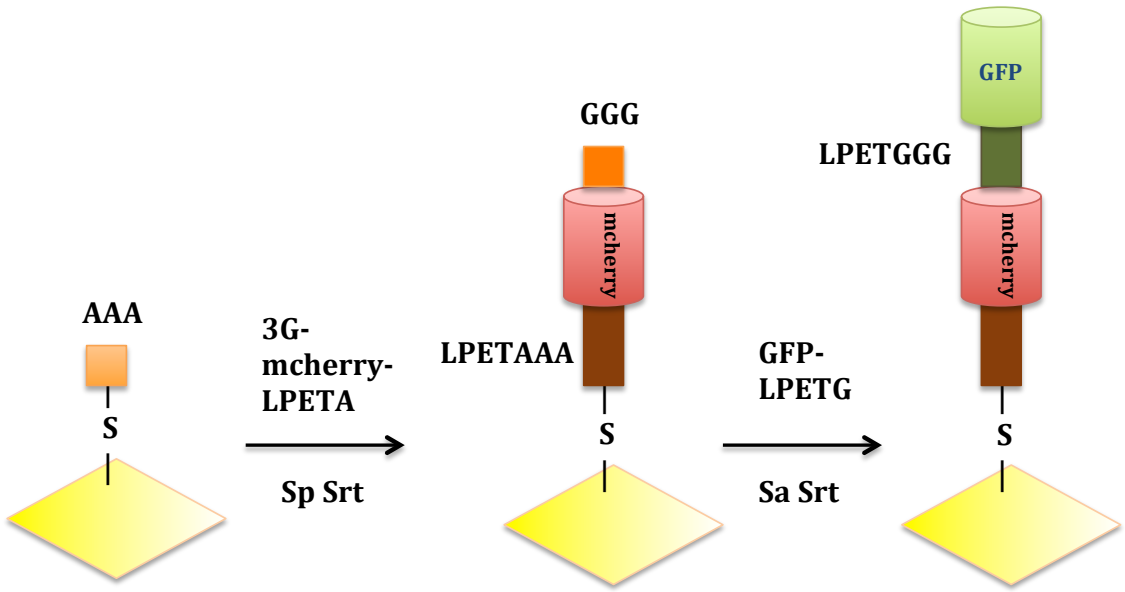


Figure 3-1. Sequential site-specific protein immobilization using orthogonal sortases. GGG-mcherry-LPETA is immobilized on the AAA modified gold surface as the first layer using Sp Srt. Then GFP-LPETG is ligated to the N-terminus of immobilized GGG-mcherry-LPETA using Sa Srt.

The dimerized fusion product was shown in Fig 3-2. Sa Srt made GFP dimers, while Sp Srt made both mcherry and GFP dimers. When both Sa and Sp Srt are present in the reaction mixture the dimer band intensity is more than the sum of the intensities of only one Srt (Sa or Sp) is present (Table 3-1).

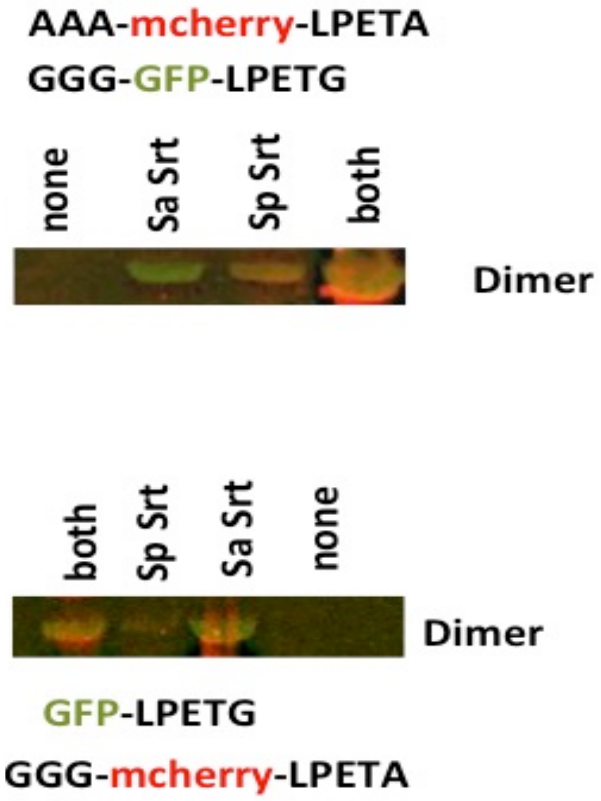


Figure 3-2. Fluorescent protein oligomerization using orthogonal sortases reactions in solution. Western blot against mcherry and GFP. mcherry shows up in 700 (red) channel, and GFP in 800 (green) channel. a) AAA-mcherry-LPETA and GGG-GFP-LPETG were used in the reaction mixture, b) GGG-mcherry-LPETA and GFP-LPETG were used in the reaction mixture. The reaction mixture had 30 μM of each substrate. Reaction buffer is TBS and 10mM $CaCl_2$ (pH=7.5), Lanes are labeled according to the Sp or Sa Srt added to the mixture: **Both**) 20 μM of Sp Srt and 15 μM Sa Srt, **Sp Srt**) 20 μM Sp Srt, **Sa Srt**) 15uM Sa Srt, **None**) without Sortase in mixture.

GGG-mcherry-LPETA and GFP-LPETG substrates were also used in orthogonal sortase reactions in solution. Sp Srt gives a faint dimer band from the little activity it has toward LPETG and GGG tags, while Sa Srt gives a stronger band, 5.88 times more intense than the Sp Srt dimer fusion product. These two experiments provided further evidence that orthogonal sortases can be used in controlling layer-by-layer protein immobilization on the surface. Sp Srt ligates GGG-mcherry-LPETA to AAA peptide on the surface without significant mcherry oligomerization, and Sa Srt ligates GFP-LPETG to the N-terminal GGG of GGG-mcherry-LPETA without cleaving off the first layer.

Table 3-1- analyzing protein dimerization reactions using orthogonal sortases

Sample[i]	**Band Intensity (BI)**[ii]
Sa Srt-a	5.555
Sp Srt-a	9.75
Both-a	33.617
Sa Srt-b	15.019
Sp Srt-b	2.558
Both-b	13.836

i: Bands resulted from protein fusion in Western Blot shown in Fig 3-2 were studied.

ii: The band intensity was corrected by subtracting the background ("none" lane).

We used this approach to ligate a layer of GFP-LPETG to a layer of immobilized GGG-mcherry-LPETA (Fig. 3-3). Fluorescent microscopy was used to detect mcherry and GFP immobilization on the surface. First, GGG-mcherry-LPETA was immobilized on the AAA modified gold slides using Sp Srt reactions. Then Sa Srt was used to immobilize GFP-LPETG on the gold surface. The GFP/mcherry MFI (Mean Fluorescent Intensity) ratio was 12.46 times more than the control slide with inactive Srt (C184G) in the reaction mixture (Table 3-2).

The protein layers immobilized on the surface were also characterized using AFM (Fig 3-4). The average height values, shown in the curves drawn in AFM image, were indicating immobilization of layers of proteins on the surface in most regions with the value of 15-30 nm (black curve at the bottom of Fig 3-4). Although there are regions on the surface with the height much more than what we expected for 2-4 layers of protein. This might be due more non-specific bindings from the non-specific interactions of different proteins with each other and the surface. The effect of scanning such surface with layers of proteins that might be stretched or interact with cantilever should also be considered.

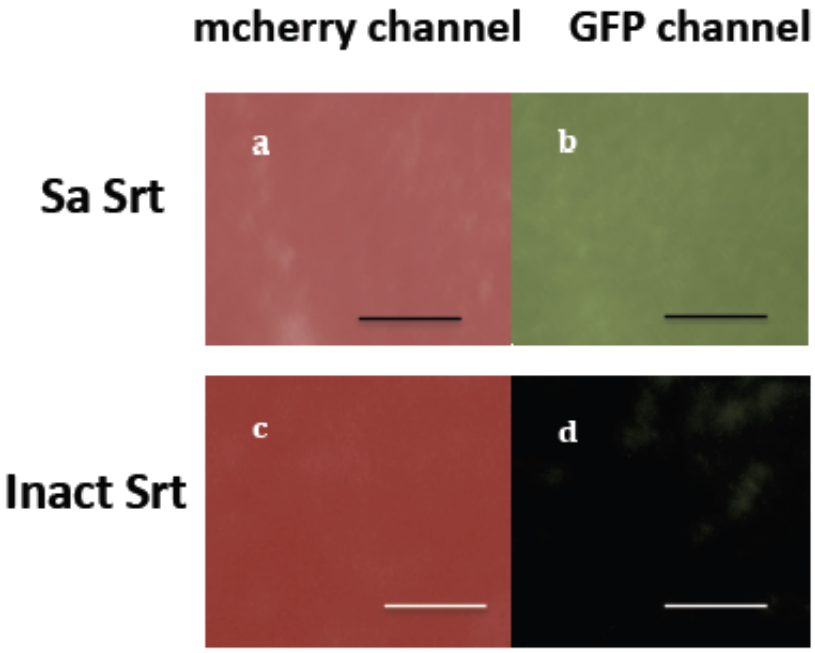


Figure 3-3. Fluorescence micrographs of protein layers immobilized on the gold surface using orthogonal sortases, Sp Srt and Sa Srt reactions. A GFP-LPETG layer immobilized on a GGG-mcherry-LPETA layer on the gold surface modified with AAA peptide using orthogonal sortase A reaction. GGG-mcherry-LPETA was immobilized on the surface using Sp Srt. a,b) Sa Srt was used to immobilize GFP-LPETG layer. c,d) Inactive Srt was used as a control. a,c) mcherry channel, b,d) GFP channel. The scale bar is µm.

Table 3-2- Mean fluorescent intensity values of mcherry and GFP immobilized on the surface using Sa and Sp Srt reactions

Mean Fluorescent Intensity *	Sa Srt (a,b)	Inact Srt (c,d)
mcherry channel	101.094	74.89
GFP channel	91.32	5.43

*: Corresponding to the slides in Fig 3-3

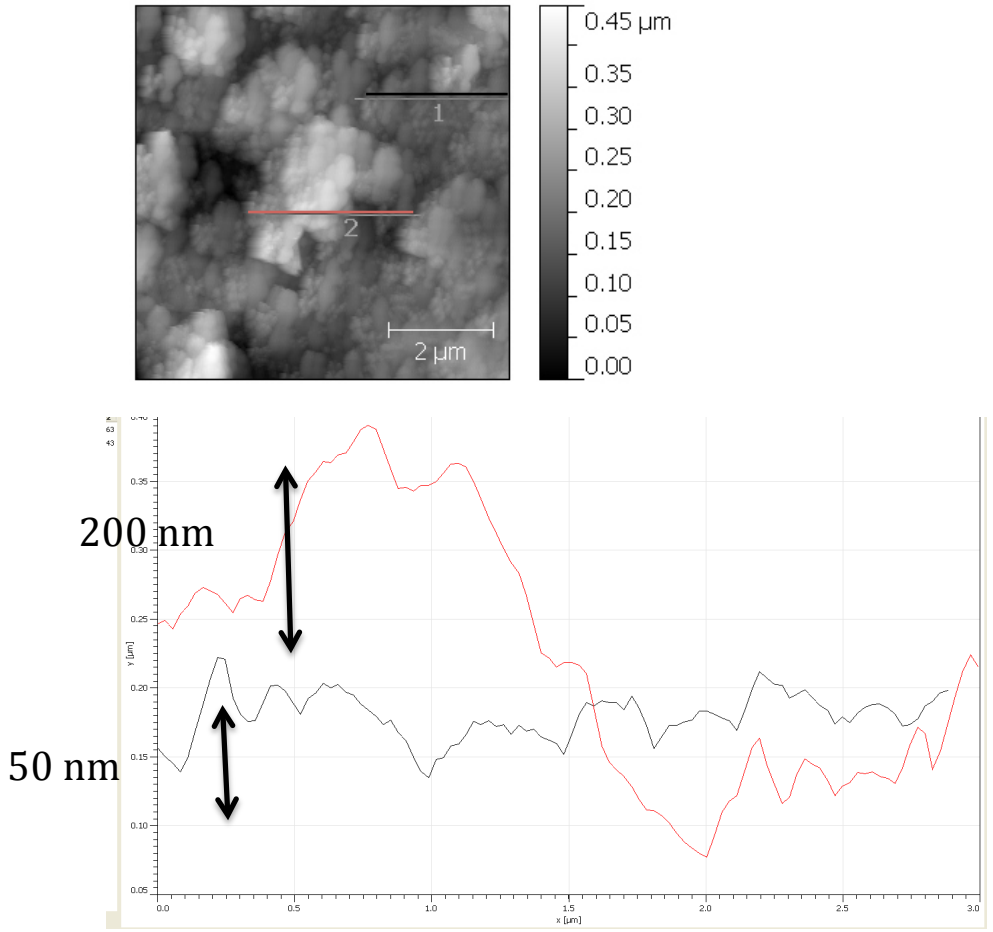


Figure 3-4. AFM image of the protein layers immobilized on the gold surface. GFP-LPETG layer has been immobilized on immobilized on GGG-mcherry-LPETA layer by layer using orthogonal sortases. The scale bar is 2 micron. The RMS value is 76 nm.

3-4- Conclusion

Sortase A-mediated protein ligation on the gold surface is a simple robust method that can be used to immobilize recombinant proteins with the Sortase A recognition sequence (LPETG/A) in the desired orientation on the self-assembled (GGGC/AAAC) peptide on a thin gold layers of biosensor chips or electrodes. This can be used in biosensors for molecular detection techniques that involve protein immobilization to improve their sensitivity. In this study, we used two different Sortase A with two different sets of specific substrates to immobilize proteins on the surface. The most efficient pair of substrates for *S. aureus* Sortase A are LPETG and GGG, and for *S. pyogens* Sortase A, this set is LPETA and AA (19,20). We were able to immobilize recombinant protein with LPETG or LPETA sequence at the C-terminus on the surface that was modified with the nucleophile sortase substrate (GGG or AAA peptide) using sortase mediated protein ligation. We also used bifunctional proteins with two sortase substrates at the C- and N-termini in Sortase A protein/peptide ligation reaction to immobilize protein oligomers on the surface. These surfaces were characterized using fluorescent and atomic force microscopy. The protein removed from the surface was assayed with Western blot to determine protein oligomerization. This technique can be used to generate protein layers on the surface, although it lacks control over uniformity of the layer formation. To overcome this challenge, the orthogonal sortases with slightly different substrate specificities can be used to control protein immobilization layer-by-layer.

The remaining challenge in sortase-mediated protein ligation and immobilization is the yield of Sortase A reaction. Sortase A transpeptidation reaction has a low efficiency in vitro. The product of the sortase reaction (LPETGGG) is a substrate for the enzyme; therefore some of the ligated product can be cleaved off the surface. To overcome this challenge, Yamamura et al. introduced a hairpin structure that is formed upon product formation and inhibits the enzyme access to the product (21).

Moreover, The amine group in the released Gly can act as a nucleophile substrate, which makes the transpeptidation reaction reversible. Liu and coworkers introduces a sequence that starts a chemical reaction after product formation and makes the N-terminus of the released glycine. This makes the last step of sortase transpeptidation reaction irreversible (88). Levary and coworkers controlled these unwanted reactions by optimizing the amount of enzyme/substrate and the incubation time used in the immobilization reaction (89). Although this specific property of sortase-mediated protein ligation can be beneficial if removing of the recombinant protein without changing the surface properties is desirable.

Site-specific immobilization of protein layers on the surface to generate 3D protein assemblies has a great potential as well. These well-oriented protein assemblies can be used as protein scaffolds that mimic Extra Cellular Matric (ECM), bioactive hydrogels, biocompatible drug carries, and layers of photoconductive proteins that has been ligated to each other and surface layer by layer.

CHAPTER 4

CONTROLLING SORTASE A-MEDIATED SITE-SPECIFIC PROTEIN IMMOBILIZATION ON THE SURFACE USING A PROTEASE CLEAVABLE PROTECTING GROUP

4-1-Introduction

As discussed in previous chapters, protein patterns and 3D assemblies control cell morphology and signaling in tissue and stem cell research (90). Among different methods used for protein ligation and immobilization, such as physical or chemical methods, site-specific covalent protein ligation using an enzymatic approach has many advantages such as stability and control over ligation site. In gram-positive bacteria, Sortase A attached surface proteins with a LPXTG (X can be any amino acid) sequence at the C-terminus to the N-terminus of a pentaglycine in the peptidoglycan cell wall (33,84). Sortase breaks the amide bond between T and G by a nucleophile attack of the Cys in the active site and makes an acyl-enzyme intermediate. This enzyme-substrate intermediate will be resolved by the nucleophile substrate (Gly_5), and results in formation of an amide bond between T in LPET and G in pentaglycine.

If LPETG and GGG tags were located at two termini of a protein, cyclized or oligomer proteins were made using sortase reaction.

Sortase-mediate protein immobilization and oligomerization can be used to generate site-specific protein wires or 3D assemblies, although there is no control over homogeneity of

each protein layer. In order to develop a method to immobilize layers of proteins on the surface in a controlled manner, we used an enterokinase cleavable sequence at the N-terminus of a bifunctional GFP, GFP with a GGG and a LPETG at the N- and C- termini respectively, to control site-specific protein ligation and immobilization on the surface using Sortase A reaction. The enterokinase cleavable sequence, DDDDK, was inserted before the GGG sequence at the N-terminus of a bifunctional GFP to control protein ligation in one direction. The first layer of protein is immobilized on the surface by site-specific ligation of the C-terminus EK-bifunc-GFP (DDDDK-GGG-GFP-LPETG) to the N-terminus of GGG modified surface. Then the protecting group will be removed and available GGG at the N-terminus can participate in the second round of protein immobilization with LPETG sequence of EK-bifunc-GFP.

We showed that generating wires in the solution and protein 3D assembly on the surface required the enterokinase recognition sequence to be removed, before adding sortase to the reaction mixture. GFP oligomers, up to heptamers, were made using sortase reaction after removing the protecting group by enterokinase. Recombinant mcherry (GGG-mcherry-LPETG) was immobilized on the layer of GFP that was immobilized on the GGG modified surfaces, after removing the N-terminal protecting group using enterokinase. We immobilized fluorescent protein layers on the GGG modified polystyrene beads and the gold surface.

4-2-Material and Methods

4-2-1-Plasmids and strains

The EK-bifunc-GFP (DDDDKGGG-GFP-LPETG) and GGG-mcherry-LPETG construct were made using PCR and inserted between NdeI and XhoI restriction sites in a pET vector with a hisx6-tag at the C-terminus after XhoI site. *S. aureus* Sortase A with a hisx6-tag at the N-terminus was cloned out of pHTT27 plasmid and inserted into the similar pET vector by introducing NdeI and XhoI restriction sites. Srtm4 is an evolved Sa Srt with higher activity toward LPETG substrate (85). Four mutations D101N, K137T, D106A, P94S were made by site-directed mutagenesis (pMR5m4). Mutagenesis primers were designed following the primer design direction in Quickchange (Strategene). Phusion polymerase (NEB) was used to generate the mutated plasmid using 16 cycles of PCR. Then 1 µl (20 Units) of DpnI (NEB) was added to the PCR product to digest the methylated, non-mutated plasmid and the reaction mixture was incubated at 37°C for 1 hour. 1 µl of the plasmids were then transformed to XL1-Blue cells (Stratagene) following the company protocol.

4-2-2- Protein production and purification

Proteins were expressed in BL21(DE3) cells (Invitrogen). An overnight cell culture was diluted (1:50-1:100) in 250ml of LB Kan media. The cells were grown to OD600 of 0.5-0.7 at 37°C shaker and then induced by adding IPTG to the final concentration of 1mM and pelleted after 3-4 hours of induction. Srtm4 protein was induced by adding IPTG to a final concentration of 0.5mM and incubating for 3 hours in 30 °C shaker. The cells were frozen at -20 °C overnight. Proteins were extracted after thawing the cells on ice by resuspending them in 10 ml B-per reagent (Thermo scientific) and 10 µl DNaseI (NEB) following the B-

per manual. All proteins were soluble and can be found in the supernatant. Proteins were purified using his-tag purification by Talon resin (Clontech), following batch/gravity user manual. The purified proteins were then dialyzed using thermo scientific dialysis cassette with MW cutoff of 10,000 or buffer exchanged against TBS (50 mM Tris, 150 mM NaCl) buffer (pH=7.5) using Amicon filters (Millipore) with MW cutoff of 10,000 to remove imidazole. Protein concentration was measured using A280 measurement. 10 to 40 mg of the protein can be produced from 1L culture. Proteins were concentrated, aliquoted and freezed for a few months at -20 °C if needed for the future use. For short-term storage, proteins were kept at 4 °C.

4-2-3- Generating protein wires in solution

Entrokinase (EK) recognition sequence (DDDDK) was removed by adding 1µl enterokinase (2 µg/ml, NEB) to 50 µl volume of 30 µM EK-bifunctional-GFP in the TBS buffer (pH=8.0) overnight at 23 °C shaker. Since all of the proteins used in the reaction were His- tagged except enterokinase, proteins were separated from EK using Talon resins (Clontech). 50 µl of enterokinase reaction product was added to 100 µl of washed resins in equilibrium /wash buffer (pH= 7.0). The mixture was shaken gently on orbital shaker at room temperature for 30 minutes. Then resins were centrifuged at 700 g for 5 minutes and washed with 900 µl of equilibrium /wash buffer (pH= 7.0). His tagged proteins were then eluted by adding 150 µl of elution buffer (imidazole in equilibrium /wash buffer). The resins were centrifuged at 700 xg for 5 minutes and the supernatant was kept as the purified protein. To remove imidazole from the buffer, the purified proteins were buffer exchanged and concentrated 10 times with TBS (pH=8.0) using amicon filters (MW cutoff = 10 kDa). The reaction mixture

consists of 40 µl of bifunctional GFP after the EK tag and enterokinase have been removed (final concentration of 150 µM), 5 µl 10x reaction buffer (TBS, 10mM $CaCl_2$, pH=8.0), and 10 µM Srtm4, at 37 °C shaker, rpm 290, for 3 hours. Then 1 µl EDTA (0.5 mM) was added and they have been kept on ice.

4-2-4- Protein assembly on the bead surface

To conjugate Gly_3 to Amine-terminated beads (Polysciences), 100 µL of the beads were pelleted and washed twice in 0.1 M 2-morpholinoethane sulfonic acid (MES), pH 4.8. Fmoc-protected Gly_3 (Sigma) was dissolved in dimethylformamide and added at a final concentration of 1 mM to 200 µL of 0.1 M MES (pH 4.8) containing the washed beads. 1-ethyl-3-[3-dimethylaminopropyl]- carbodiimide hydrochloride (EDC, Pierce) was then added in aqueous solution to the bead slurry at a final concentration of 10 mM, and the mixture was shaken for 2 h at room temperature. After shaking, the beads were pelleted, washed once with 0.1 M MES, and twice in deionized water. They were suspended in 20% piperidine in water and shaken for 20 min, after which the beads were pelleted, washed twice in water and then twice in 50 mM Tris, 150 mM NaCl, 0.1% Tween-20, and stored in the same buffer.and protein immobilization on the beads was done as described. 20 µM EK-bifunc-GFP was immobilized on the Gly_3 bead surface using 12 µM Srtm4 by incubating the beads in the reaction mixture at 37 °C shaker for 3 hr. The beads then washed several times with TBS and PBSA (PBS +0.1% BSA). The protecting group (DDDDK) was removed by incubating the beads by adding 1µl of Enterokinase (NEB) to 50 µl of the beads resuspended in TBS buffer, and incubating at 23 °C shaker overnight. The beads were washed the following day several times with TBS and PBSA. 3G-mcherry-LPETG reaction

mixture (TBS, $CaCl_2$ 10 mM, pH=7.5) containing 6 μM Srtm4 was then added to the beads slurry and incubated for 1.5 hr at 37 °C shaker. The beads were washed several times with TBS and PBSA before further analysis. Sortase A reaction conditions has been optimized as explained (89).

4-2-5- Western blot

20 μl of the denatured proteins plus 20 μl of SDS-PAGE sample buffer (Tris, SDS, glycerol, 2- mercaptoethanol, orange dye) were run on a 10% Bis-Tris NuPAGE gel (Invitrogen) using MOPS buffer (Invitrogen). 8 μl of the single color protein marker (LiCor) was added to the first lane. The proteins were transferred for 1 hr at 33 V into a 0.45 μm nitrocellulose membrane (Bio-rad) using transfer buffer (Invitrogen). The membrane was blocked either for 2 hr at room temperature or overnight at 4 °C on a orbital shaker in 1:1 blocking buffer (Licor) and PBS buffer (pH=7.4). The blocked membrane was then labeled with primary antibody (1:1000 dilution); mouse anti-His (Covance) or rabbit anti-DsRed (Clontech) and secondary antibody (1: 10,000 dilution); anti-mouse IR800 or anti-rabbit IR680 (Odyssey). The images were taken using the Odyssey Infrared Imaging system.

4-2-6- Fluorescence Microscopy

The fluorescence micrographs were obtained using Olympus or Nikon fluorescent microscopes with 40x magnification. The images were analyzed with ImageJ. FITC filter cube (Ex: 494 , and Em: 518) and Texas Red HQ cube (Ex: 577, Em: 620) in Nikon microscope and GFP and RFP filters have been used for detecting GFP or mcherry fluorescence. Mean Fluorescence Intensity (MFI) was determined using Analyze> measure

tool in ImageJ software on the whole slides. The background (negative control) value was subtracted from the actual sample before normalization or comparison.

4-2-7-Peptide assembly on the gold surface

Gold-coated microscope slides with 100 Å gold thickness were purchased from Platypus. Slides were cut into 5x5 or 10x10 mm pieces and washed with RCA solution (ammonium hydroxide: hydrogen peroxide: H_2O; 1:1:5 ratio) at 70 °C for at least 10 minutes, then rinsed with copious amount of dI H_2O and dried under a nitrogen stream. The GGGC peptide was purchased from Genscript with >%95 purity. Dimethylformamide (DMF) was added drop-wise to the peptide powder until it was dissolved, then oxygen free water (to prevent disulfide formation between Cys) was added to the solution to make 100 μM–1 mM solutions. The clean gold slides were submerged in the peptide solution 1 to 3 nights. After that the slides were washed with TBS buffer (pH=7.5), and dI H_2O and dried under nitrogen.

4-2-8-Protein immobilization on the peptide SAM on gold

20-40 μM of EK-bifunc-GFP was mixed with 15 μM Srt, or 10 μM Srtm4 in a reaction buffer (TBS, 10mM $CaCl_2$, pH=7.5). 50-80 μl of the reaction mixture was put on the functionalized slides with the peptide and kept in a moist sealed mini petri dish at least 3 hr at 37 °C incubator. The slides were washed the next day with filtered TBS (pH=7.5) and dIH2O and dried under nitrogen. For immobilizing layers of proteins, the slides with the first protein layer immobilized on the surface were washed several times with TBS buffer and dIH2O and dried under nitrogen. Then second round of sortase reaction was done similarly, followed by washes and drying.

4-3- Result and discussion

4-3-1-Generating protein wires in solution using Sortase A

As discussed in Chapter 1, site-specific protein oligomers can be generated using sortase reaction. To investigate whether sortase-mediated site-specific protein oligomerization can be controlled by a removable sequence at the N-terminus of a bifunctional protein Site-specific protein wires, an enterokinase cleavable sequence was inserted after GGG at the N-terminus of a bifunctional protein to limit protein oligomerization in one direction. Enterokinase recognizes a DDDDK sequence and cleaves after the lysine (K). The protecting group at the N-terminus of EK-bifunc-GFP, DDDDK, was removed using entrokinase. Bifunctional GFP with hisx6-tag at the C-terminus was purified using Talon resins to remove enterokinase from the reaction mixture. Sortase A was then used to generate site-specific GFP wires, up to seven units, in solution (Fig. 4-1). It is likely that entrokinase has some inhibitory effect on the sortase reaction, since the oligomerization was not as effective when enterokinase was present at the time of adding Srt to the reaction mixture. When sortase was added to EK-bifunc-GFP, a dimer (running a little above the normal GFP dimer) was made. Since Srt lacks tight specificity for the nucleophile substrate, it is likely that the N-terminual amine of D or the primary amine in the side chain K in the DDDDK sequence at the N-terminus of the EK-bifunc-GFP resolved the enzyme-substrate intermediate in the absence of the original nucleophile substrate (3G). GFP oligomers were not observed when sortase and enterokinase were not added to the mixture. Bifunctional GFP (GGG-GFP-LPETG) was used as a positive control. Sortase-mediated GFP oligomerization is much more effective after removal of the protecting group, upto

heptamers, comparing to the original bifunctional GFP, upto trimmers, under similar conditions. One possible explanation is that the Met at the N-terminus of the bifunctional GFP has not been fully removed after posttranslational modifications and cleaving the protecting group using enterokinase makes the N-terminal GGG tag available for sortase reaction.

4-3-2- Site-specific assembly of protein layers in a controlled manner on the bead surface

Amine terminated polystyrene beads were functionalized with Fmoc-triglycine using EDC chemistry. The Fmoc protection was then removed using piperidine as described in material and method (34). GFP-LPETG or biotin-LPETG peptide was immobilized on Gly_3 modified bead surface using Srtm4. The fluorescent signal corresponding to the protein/peptide immobilization on the bead surface was measured using flow cytometry and shown in Fig A4-1,2. GFP fluorescent intensity was shown in FL1 channel and biotin-LPETG peptide labeled with streptavidin-AlexaFluor 647 showed up in FL4 channel. The mean fluorescent intensity was increased when Srt was added to the bead slurry reaction mixture by 20 times when GFP was attached to Gly_3 beads using sortase reaction. In case of amine terminated beads, the fluorescent intensity was increased by 3 tims (Fig A4-1). There was a 4-fold increase in fluorescent intensity related to LPETG peptide binding to PS beads using sortase reaction (Fig A4-2). 20 μM EK-bifunc-GFP was ligated to the Gly_3 beads using 12 μM Srtm4. To remove the protecting group at the N-terminus of EK-bifunctional-GFP, the beads with immobilized EK-bifunc-GFP were incubated at 23°C shaker overnight after adding enterokinase to the bead slurry. The beads were washed several times with TBS buffer (pH=7.5). The second round of protein immobilization was then done on the beads.

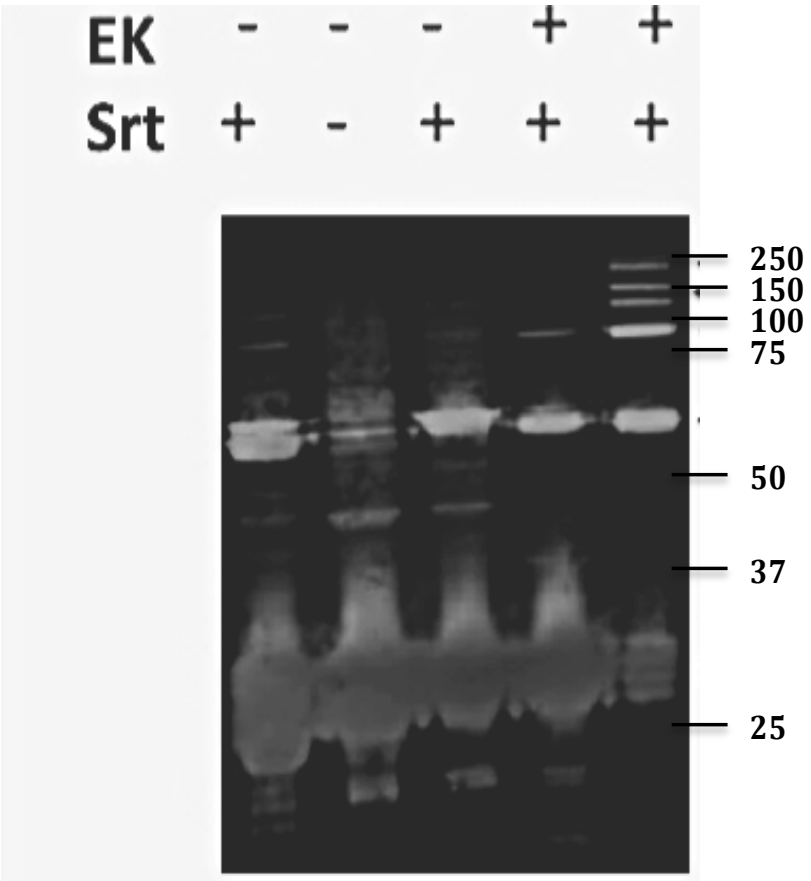


Figure 4-1. Generating protein wires in solution using Sortase A reaction and controlling it using an N-terminal cleavable sequence. Anti –Hisx6 blot shows GFP oligomerization in solution using Sortase A that is controlled by an enterokinase cleavable sequence at the N-terminus of a bifunctional GFP as a protecting group. Left to right: The left lane is bifunctional GFP as a positive control, the rest are EK-bifunc-GFP. Presence of EK (enterokinase) and Srt (Sortase) were stated on the top. In lane 4, enterokinase was not removed before adding Srt, while in lane 5 it has been moved.

30 μM 3G-mcherry-LPETG was ligated to de-protected bifunctional GFP using 8 μM Srtm4 (Fig 4-2.a). Less Srt was used for second round of reaction to decrease re-cleavage of the attached protein to the surface. Fluorescence microscopy images were taken after several washes with TBS buffer. Fluorescence micrographs showed that mcherry was attached to the beads using Srt reaction only when the protection group has been removed using EK and the N-terminus GGG was available to participate in sortase reaction with the LPETG sequence at the C-terminus of the recombinant mcherry (Fig 4-2. b,c). Analyzing fluorescent micrograph of beads showed that removing the protecting group before adding GGG-mcherry-LPETG and Sa Srt has resulted in 3 times more fluorescent in mcherry channel comparing to the beads with N-terminal protected GFP immobilized on them (Table 4-1).

The beads with immobilized layers of protein were boiled for 30 min at 100°C to melt polystyrene. The mixture was then run on a SDS-PAGE gel and transferred to a Nitrocellulose membrane. The membrane was labeled with antibody against mcherry to detect the mcherry immobilization and oligomerization on the bead surface. GGG-mcherry-LPETG can oligomerize upon binding to the surface using sortase reaction.

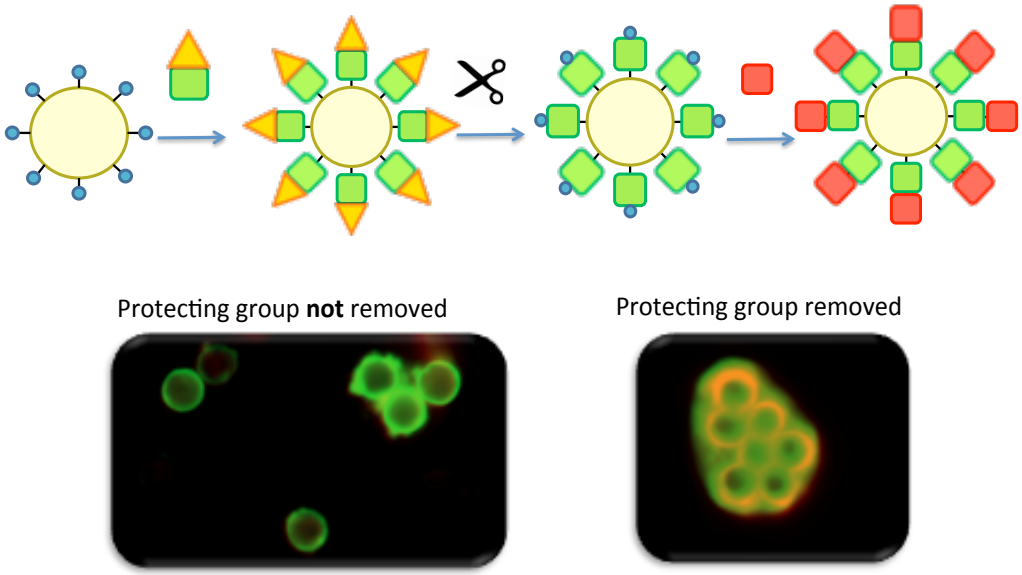


Figure 4-2. Site-specific controlled protein immobilization on the bead surface using SortaseA. a) Polystyrene beads were functionalized with triglycine. EK-bifunc-GFP was site-specifically immobilized on the surface using Srt reaction. The EK protecting group was cleaved off using enterokinase, and therefore triglycine located at the N-terminus of bifunctional GFP was available for another round of protein immobilization. Then, mcherry-LPETG was immobilized on the surface using sortase reaction. b,c) Fluorescent microscopy images of overlay of GFP and RFP channels corresponding to GFP and mcherry fluorescent on the beads. Both of them had an immobilized layer of EK-bifunc-GFP on the Gly_3 surface using Srt. The protecting group was cleaved off by incubating the beads with enterokinase at 23°C overnight in image c. Then mcherry-LPETG was added to in a reaction mixture using Srtm4 (TBS, 10mM $CaCl_2$, pH=7.5).

Table 4-1- Mean fluorescent intensity values of mcherry and GFP immobilized on the PS beads using Sa Srt reaction and EK method

Channel	EK tag not removed	EK tag removed
Mcherry MFI	15.893	45.54
GFP MFI	46.868	43.363

mcherry oligomers were observed in presence of Srt when the protecting group was

removed using enterokinase (Fig. 4-3). When no EK or Srt were added, a weak band corresponding to mcherry monomer were observed, that might be because of Srt reaction between mcherry-LPETG and GGG on the surface from the Srt failed to be removed from the previous step or non specific binding of mcherry to the surface.

4-3-3- Site-specific assembly of protein layers in a controlled manner on the gold surface

We used sortase-mediated protein immobilization to site-specifically immobilize EK-bifunc-GFP on the gold surface using a simple two-step strategy. Protein was immobilized on the gold surface using sortase similar to previous chapters. First, we generated a self-assembly layer of a GGG peptide with a cysteine at the C-terminus (GGGC). The sulfhydryl group in Cys makes Au-S bond with the gold surface. Then we used Srtm4 to site-specifically immobilize DDDDK-GGG-GFP-LPETG on the surface as the first protein layer on the gold surface that has been modified by a GGGC peptide. The protecting group was removed using enterokinase by adding 1 µl of enterokinase to 50 µl to the EK-bifunc-GFP immobilized slide with and incubating at room temperature overnight. The slide was then washed with copious amount of TBS buffer. Second layer of site-specific protein immobilization was done similar to the first layer immobilization. 3G-mcherry-LPETG was immobilized on the GFP layer immobilized on the gold surface using Srt. The reaction mixture added to the slides were incubated at 37°C incubator at a moist sealed plate for 1.5 hr. Fluorescent microscopy showed that the recombinant mcherry was immobilized on the surface when the enterokinase recognition sequence at the N-terminus of bifunctional GFP was removed and the amine group of Gly at the GGG sequence at the N-terminus was

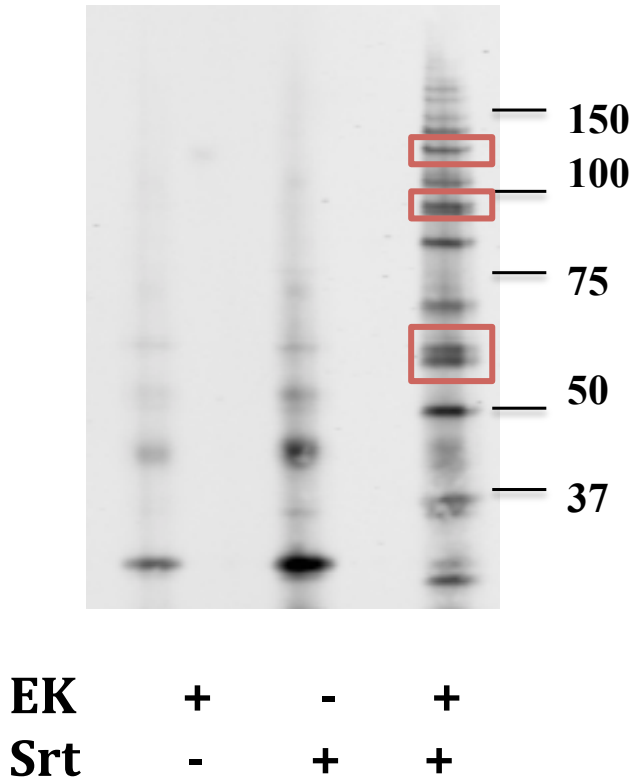


Figure 4-3. Anti-DsRed immunoblot to detect protein oligomers immobilized on the bead surface using Sortase A and EK protecting method. Gly_3 functionalized polystyrene beads with layers of immobilized GFP and mcherry proteins were boiled at 100 °C for 30 min with SDS-PAGE sample buffer. Then western blot was done using rabbit anti-Ds-Red that binds to mcherry and anti rabbit IR-680. EK-bifunc-GFP was immobilized on all of the samples using Srt reaction. From left to right: 1) beads were treated with enterokinase to remove the protecting group, but Srt was not added along with the GGG-mcherry-LPETG reaction mixture, 2) beads were not treated with enterokinase, but Srt was added to the the GGG-mcherry-LPETG reaction mixture, 3) beads were treated with enterokinase to remove the protecting group, and Srt was added along with the GGG-mcherry-LPETG reaction mixture. GFP/mcherry dimer, trimer and tetramer are marked with the red rectangle. GFP is about 30 kDa and mcherry is about 28 kDa.

available for Srt reaction, suggesting that 3G-mcherry-LPETG immobilization on the surface is due to ligation to the N-terminus of the bifunctional GFP immobilized on the triglycine-gold surface (Figure 4-4). Table 4-1 shows the MFI values for each slide. The ratio of mcherry to GFP was increased 2.56 times when the enterokinase tag was removed.

4-4- Conclusion

We used sortase-mediated site-specific ligation to immobilize and oligomerize proteins on the surface. First, we generated GFP wires in solution using sortase-mediated ligation. In order to control protein oligomerization, we introduced an enterokinase cleavable sequence at the N-terminus of a bifunctional GFP. Although the enterokinase recognition sequence (DDDDK) located at the N-terminus of the bifunctional GFP was generating a fusion product close to the GFP dimer in the presence of sortase, the DDDDK protection group at the N-terminus inhibited the formation of higher order GFP oligomers. GFP oligomers were made upto heptamers using sortase after removing enterokinase recognition sequence.

The same approach was used to immobilize layers of protein on the surface in a controlled manner using sortase reaction. GGG-mcherry-LPETG was attached to the recombinant GFP immobilized on the gly_3 functionalized gold or polystyrene bead surface. An efficient second layer (mcherry) attachment required removing of the protecting group by incubating the immobilized proteins with enterokinase. Mean fluorescent intensity of mcherry was increased by 2.5-3 times when the N-terminal protecting group of the EK-bifunc GFP was removed before sortase reaction.

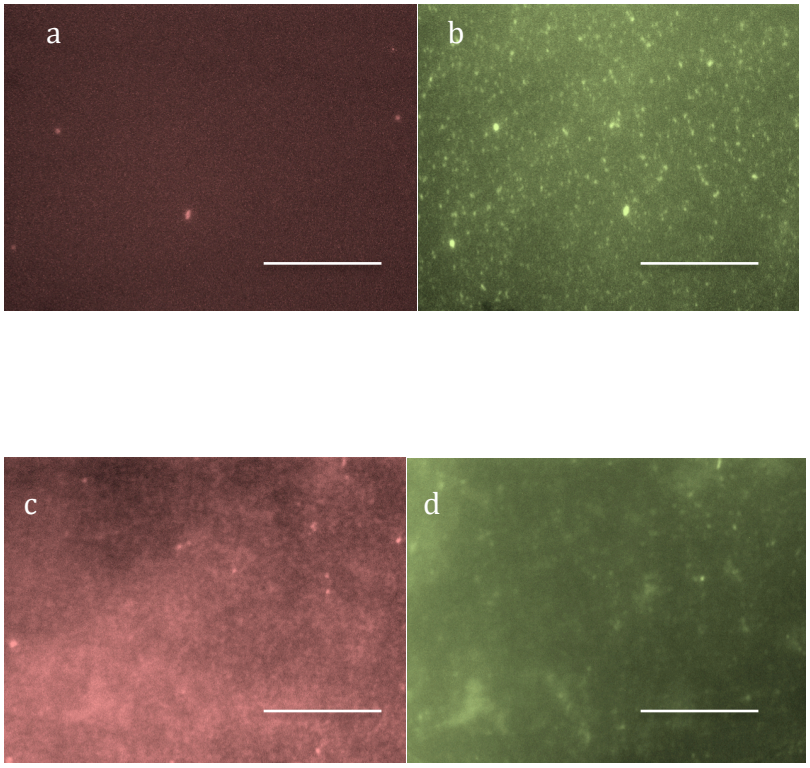


Figure 4-4. Protein immobilization in layers on the gold surface using Sortase A and EK protecting method. EK-bifunc-GFP was immobilized on the surface of Gly_3 gold surface using Srtm4. In samples a,b) The DDDDK protecting group was not removed; while in c,d) The protecting group was removed by incubating the slides with enterokinase at room temperature.

Then 3G-mcherry-LPETG was immobilized on the slides using Srtm4. Fluorescence microscopy was done after washing the samples after 1.5 hr incubation at 37°C incubator. a,c show the mcherry channel, and b,d represent GFP filter. The scale bar is 100 µm.

Table 4-2- Mean fluorescent intensity values of mcherry and GFP immobilized on the gold surface using Sa Srt reaction and EK method

Channel *	EK tag not removed	EK tag removed
Mcherry MFI	47.299	90.774
GFP MFI	88.357	66.126

*: Corresponding to the slides in Fig 4-4

CHAPTER 5

ENGINEERING SORTASE A SUBSTRATE SPECIFICITY

5-1-Introduction

Enzymes are proteins that accelerate rate of biochemical reactions in nature. In order to use enzymes outside the cellular machinery, activation, stability or specificity of the enzymes needs to be improved (91). Designing protein structure to have the desired property requires a vast knowledge of protein structure and relationship of the amino acids in stability and functionality of the protein. Despite all the progress in computational modeling of protein structure, inability to accurately predict the effect of mutations on the complex structure/function of enzymes has limited the general application of rational design for enzyme engineering. More recently, directed evolution has proven to be a powerful tool for the modification of enzyme activities and has become the most widely used approach (92). Different strategies of directed evolution such as random, targeted mutagenesis or recombination are used to correlate enzyme structure and function by screening pool of mutants. Although directed evolution is a powerful approach to engineer proteins, knowledge-based library design can decrease the size of protein library and screening. Therefore, it is essential to follow the strategy that fits the evolution need (93).

Many areas of protein engineering such as antibody engineering are concentrated on manipulating protein binding affinity or specificity against another protein, peptide or reagent. Therefore, the goal is improving association/dissociation parameters. In case of engineering enzymes such as sortase, there are more factors that need to be optimized: 1)

Binding to the acyl donor substrate (LPXTG) to generate the intermediate, 2) Allowing acyl acceptor substrate (natural GGG, or P0 peptide) to access the active site and do the nucleophile attack, 3) releasing the product.

As previously stated, Sortase A has broad substrate specificity toward the nucleophile substrate. Sortase A substrate specificity and activity toward a new nucleophile substrate can be used to make a more powerful tool in generating protein 3D assemblies. Although Srt C is responsible for the pilus assembly in gram-positive bacteria, Srt A has previously shown some activity toward the ε- amine in the side chain of K (lysine) in a pilin box sequence (94). The engineered sortase toward ε- amine in the side chain of K (lysine) can be used in generating branched protein ligation in protein 3D assemblies. Engineering Srt A to be more active and specific toward ε- amine in the side chain of K (lysine) in a specific sequence also provides a powerful enzyme to be used in site-specific bioconjugation of peptides and proteins to biomolecular structures or assemblies that harbor this specific sequence, without perturbing the structure and function of that assembly.

Sortase A activity has been previously evolved using yeast surface display and directed evolution. Chen and coworkers fused sortase to Aga2p and used a chemical conjugation to bind the sortase substrate on Aga1p displayed on the yeast surface. They isolated Sortase A mutants with up to a 140-fold increase in activity compared to the wild-type enzyme (85). Sortase substrate specificity was evolved toward broader acyl donor (LPXTG) substrate range using phage display. The sortase mutant library was made by random mutagenesis of substrate recognition loop (95).

In order to engineer sortase substrate specificity and activity toward the side chain of lysine in a pilin box sequence, we used directed evolution and yeast surface display. P0 peptide,

pilin box motif of a SpaA subunit of *C. diphteriae*, was used as a substrate to engineer Sortase A substrate specificity. First, we used YSD to display sortase and LPETG substrate on the surface of the yeast. Sortase was inserted to the N-terminus Aga1p and the LPETG peptide was fused to the C-terminus Aga2p. Then we used FACS to screen Sortase A library against P0 substrate. Sortase colonies after three rounds of sort have been screened and analyzed.

5-2-Materials and methods

5-2-1-Plasmids and strains

The yeast *Saccharomyces cerevisiae* strains, BJ5465 (MATa ura3-52 trp1 leu2-delta1 his3-delta200 pep4::HIS3 prb1-delta1.6R can1 GAL), and its derivative EBY100 (MATa AGA1::GAL1-AGA1::URA3 ura3-52 trp1 leu2-delta200 his3-delta200 pep4::HIS3 prb11.6R can1 GAL) were used in the experiments. Aga1-sortase plasmids and libraries have URA3 marker, and Aga2-LPETG-myc plasmid has TRP1 marker. EBY100 was used to display the LPETG substrate on the surface of the yeast for the experiments done with purified sortase. Wild-type Srt and library of sortase mutants, along with the sortase substrate were transformed to BJ5465 strain.

Sortase A substrate to be displayed on the yeast surface, Aga2-LPETG-cmyc construct, was made by PCR using Aga2-LPETG-1 and Aga2-LPETG-2 primers, and inserted between EcoRI and XhoI sites in pCT302 plasmid, pMR1 plasmid. The NheI site has been replaced with EagI restriction site (Fig A2-2).

Sortase A construct, with an AU1 epitope tag at the N-terminus, was cloned at the N-terminus of Aga1. Sortase A construct to be displayed on the yeast surface, AU1-Sortase A-

TEV–linker-Aga1, was made using PCR and cloned into the modified pY3 plasmid before aga1 gene.. To insert Sortase at the N-terminus Aga1 with EagI and NheI sites into pY3 (pIU211) vector, two PCRs were done using two sets of primers: 1) Aga1-Srt-Rev1 and Aga1-Srt-RE-insert-For, and 2) Aga1-Srt-For1 and Aga1-Srt-RE-insert-Rev. These two PCR products were then used as templates for a PCR using Aga1-Srt-For1 and Aga1-Srt-Rev1. The whole Srt piece with a little overhang of Aga1 was then inserted into pY3 vector using BsiW1 and EcorI restriction sites. Then the whole AU1-Sortase-Aga1 cassette was cut using EcoRI and XhoI restriction enzymes and inserted into pMRUS420 (pRS315 backbone) plasmid with URA3 marker (named pCUAga1Srt).

5-2-2-Generating sortase library

Sortase mutant library was made using error prone PCR and nucleotide analogs (dPTP and oxo-dGTP) (20 µM - 10 cycles) following the instruction provided in the reference (96). A construct was made with PCR with 50 bp overhangs on each side to be later used in homologous recombination with linearized vector (pCUAga1Srtm4) using HR-Aga1-Srt-For and HR-Aga1-Srt- Rev primers. The random PCR products were further amplified by a second PCR using the same primers to a total volume of 500 µl. The vector (pCUAga1Srtm4) was digested using EagI-HF and NheI-HF restriction enzymes for at least 4 hours at 37°C (total volume of 200 µl). The PCR products and the cut vector were gel purified and concentrated using pellet paint (Novagen) to 10 µl volume. The PCRs were then inserted into the cut vector by homologous recombination upon electrotransformation into the BJ5465 yeast cells that already contained LPETG substrate plasmid (pMR1). The cells were resuspended in 900 µl of cold YPD media in the cuvettes

and several transformations (5-10) were combined. After growing the cells in 30°C for 45 minutes, 100 µl of the cells were plated in dilution series (1:50 to 1:50,000) to determine the library size. The cells were pelleted and grown in SD-CAA media (pH= 4.5) overnight. The library was analyzed after three passages in selective media.

5-2-3-Sortase reaction on the surface of yeast cell

The yeast cells (BJ5465) were grown in SD-CAA media overnight. After induction in SGR-CAA media for 16-20 hours at 30°C, $6x10^6$ cells (OD_{600} of 0.6) were collected and washed with TBS (150 mM NaCl, 50 mM KCl, pH= 7.5) buffer. Then the nucleophile substrate (GGG or P0 peptide) was added to the cells in 50 µL total volume of the reaction mixture (TBS, 10 mM $CaCl_2$, pH=7.5) to the final concentration of 0.2-0.5 mM and incubated at 37°C shaker for at least 2 hours. Pilin box (P0) or GGG peptide with a biotin conjugated at the N- and C-terminus respectively (Genscript, purity > 90 %) was used as the substrate. The peptide solution was made by resuspending the lyophilized peptide in dIH2O to the concentration of 3-6 mM.

5-2-4-Flow cytometry and Immunolabeling

Cells were washed with cold PBSA (PBS and 1% BSA) and labeled with mouse anti-AU1 (Covance) (1:100 dilution) and Streptavidin Alexa fluoro 647 (Invitrogen) (1:100 dilution) for 1 hour on ice. Cells were then washed with PBSA and incubated on ice for 30 min with anti-mouse Alexa fluoro 488 for the secondary immunolabeling. After the last wash, the cells were resuspended in 750 µl of PBSA and run through flow cytometer. At least 10,000 cell events were collected per sample. Flow cytometers used included Accuri C6 (Accuri Cytometers Inc.) and FACSAria. Flow cytometric data was analyzed using Flowjo

software. The AU1 (representing sortase) display level on the surface of yeast is proportional to the background-corrected mean fluorescence intensity value

$$cMFI(AU1) = MFI_{(AU1+)} - MFI_{(AU1-)}$$

where $MFI_{(AU1-)}$ and $MFI_{(AU1+)}$ represent mean fluorescence intensity of AU1-tag calculated using Flowjo for negative and positive cell populations, respectively. The substrate (P0)-reaction level was properly calculated using background-corrected fluorescence associated fluorescence conjugated streptavidin:

$$cMFI(strep) = MFI_{(strep+)} - MFI_{(strep-)}$$

where (+) and (-) represent the mean fluorescence intensity of streptavidin calculated using Flowjo for negative and positive cell populations, respectively. cMFI of the tested mutant was calculated by dividing cMFI(AU1) by cMFI(strep).

$$cMFI = \frac{\text{cMFI(AU1)}}{\text{cMFI(strep)}}$$

Relative sortase mutant activity against P0 substrate is determined by normalizing the value for sortase mutant activity against substrate to the same value for the Srtm4, Normalized cMFI (NCMFI).

$$NCMFI = \frac{\text{cMFI(mutant)}}{\text{cMFI(Srtm4)}}$$

Normalization minimizes variability between experiments due to laser power output, detector amplification, and other cytometer parameters.

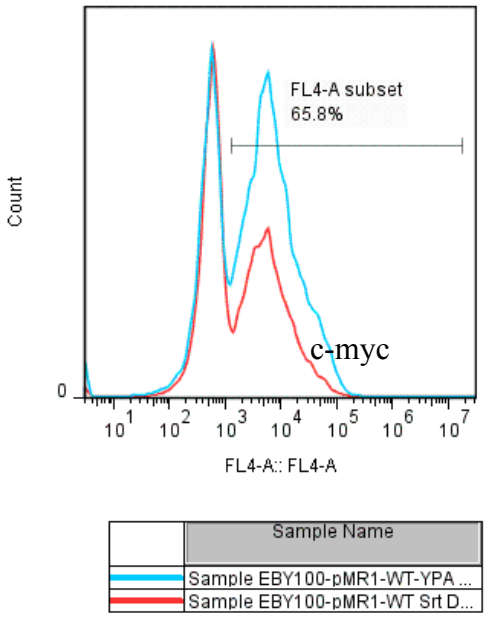
5-3- Results and Discussion

5-3-1- Sortase A site-specific peptide ligation on the yeast surface

First, purified sortase was used to immobilize GGG-biotin peptide on the surface of the yeast (EBY100) displaying LPETG substrate (pMR1) to investigate sortase-mediated peptide immobilization on the yeast surface. 30 μM of purified Srtm4 and 5 mM of GGG-biotin were used in the reaction mixture. Following growth and induction, the cells were resuspended in the reaction mixture and incubated at 37°C shaker for 3 h. Flow cytometry results showed that 3G-biotin peptide was attached to the yeast surface, and cmyc signal was reduced due to cleavage of cmyc sequence located after the displayed LPETG sequence. A control peptide (YPA) was used as a control (Figure 5-1).

For engineering sortase substrate specificity, a library of sortase mutants and LPETG peptide were displayed on the yeast surface. Plasmids containing Sortase (AU1-Srt-Aga1) and LPETG substrate (Aga2-LPETG-cmyc) were transformed into the BJ5465 yeast strain for yeast surface display. Sortase is fused to the N-terminus of Aga1, and the LPETG peptide was inserted at the C-terminus of the Aga2p protein between HA and c-myc epitope

tags. Aga1p and Aga2p proteins bind together with disulfide bonds and display sortase and



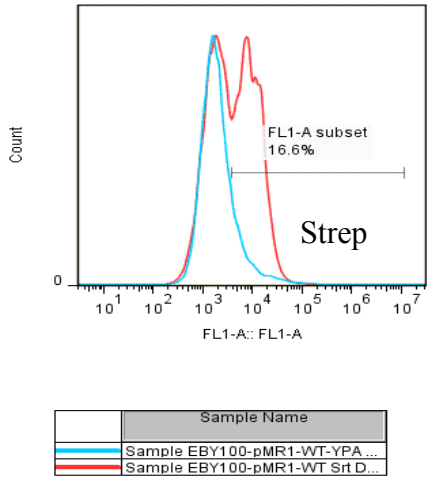


Figure 5-1. Sortase reaction on the yeast surface using purified sortase. The c-myc signal shows the display of LPETG tag. Strep is representative of GGG-biotin binding to the surface using sortase reaction. WT Srt was used either with GGG-biotin peptide (Red), or a control peptide, Biotin-YPA (Blue).

the LPETG substrate on the yeast surface. Sortase nucleophile substrate conjugated to a biotin tag (GGG-biotin) was added to the cells after induction. Sortase displayed on the yeast surface were used to bind a GGG-Biotin to the LPETG displayed on the yeast surface. The reaction was done by incubating the cells displaying LPETG substrate and Srt at 37°C shaker for 3 hr in the reaction mixture containing 1mM GGG-Biotin in the reaction buffer (TBS, 10mM $CaCl_2$, pH=7.5). Then displayed proteins (sortase and LPETG substrate) were labeled with antibodies against AU1 and HA to show the expression of sortase and LPETG on the surface of the yeast. Fluorescent-conjugated streptavidin was added to determine the extent of reaction by binding to the product (LPETGGG-biotin) (Fig. 5-2). The proteins expression and the extent of reaction were investigated by flow cytometry. WT Srt showed a signal corresponding to attachment of GGG peptide to the surface. Fluorescent intensity from cmyc signal was decreased since Srt cleaves between T and G of LPETG peptide before the c-myc sequence. WT Srt or Srtm4 attached the GGG-Biotin substrate to the LPETG sequence displayed on the yeast surface. The control sample with inactive Srt (C184G) did not show any significant GGG binding.

To engineer sortase specificity toward other known nucleophile substrates such as a primary amine in the side chain of a lysine in a pilin box sequence, P0 peptide conjugated to a biotin at the N-terminus was used as a substrate for these reactions. Srt attached P0 peptide to the LPETG substrate displayed on the yeast surface, although not as efficiently as GGG, the natural nucleophile substrate (Fig 5-6).

5-3-2- Library screening and sorting

Library of sortase mutants was made using random mutagenesis on the whole genome of Srtm4. The library size was estimated about $2*10^7$. Different concentrations of nucleotide analogs were used to generate desired mutation rate. Table 5-1 shows the effect of the nucleotide analogue (dPTP and oxo-dGTP) concentration in mutation rate of Srt library. The mutation rate was estimated by sequencing 5-10 Srt library colonies and rounded up to the closet integer number. Srt library was made with nucleotide analogue concentration of 20 μM and used in the reaction against P0 peptide as the nucleophile substrate. FACS was then used to sort the active Srt mutants against P0 peptide (Figure 5.3). Three color labeling was used for sorting the cells; antibodies against AU1 and HA epitope tags were used to label Sortase and LPETG substrate displayed on the yeast surface and to determine enzyme and substrate expression. Streptavidin showed the extent of reaction by binding to the biotin conjugated to the reaction product (Bio-P0). Biotin conjugated Ethylenediamine (EDA) was also used as a nucleophile substrate. This substrate works as an alkylamine, a poor nucleophile substrate for Srt transeptidation reaction. Srt attached this peptide to the surface although less efficiently comparing to the GGG peptide (Fig A. 4-6).

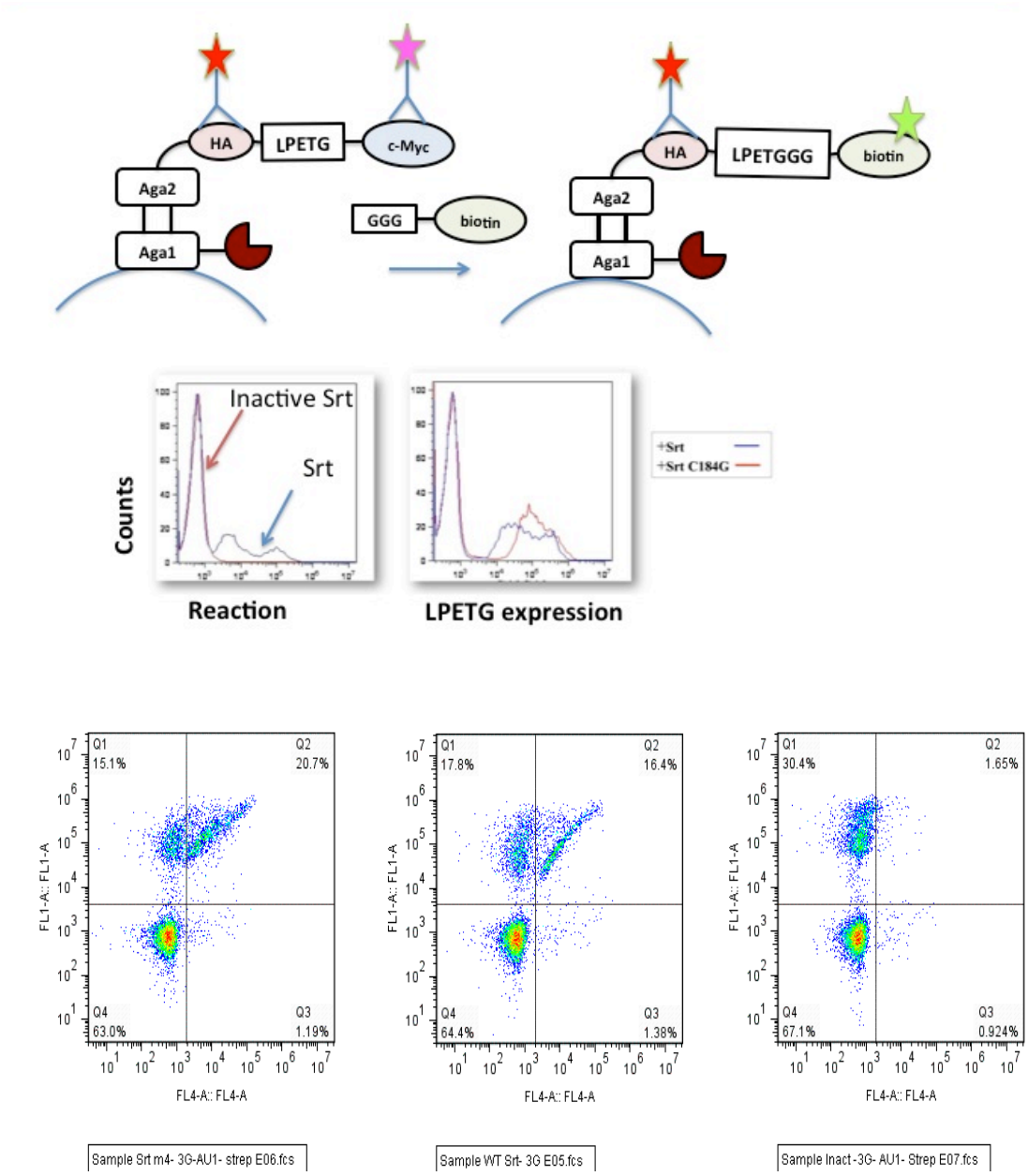


Figure 5-2. – Srt reaction on the yeast surface. a) Yeast surface display of sortase and LPETG substrate. b) Flow cytometry histograms. Left: Sreptavidin Alexa Fluoro 647 indicates the extent of GGG-biotin binding to the LPETG substrate on the yeast surface. b) Alexa fluoro 488 is representative of LPETG-myc. Blue graphs are with Srt and and red ones are with inactive Srt (C184G). FL1: AU1 expression; FL4: Streptavidin (GGG-biotin binding).

Table 5-1-Effect of the nucleotide analogue (dPTP and oxo-dGTP) concentration in mutation rate of sortase library

Nucleotide analog concentration (μM)	Mutation rate per gene
5	1
20	4
100	12

Flow cytometry results for sortase naïve library and sorts are shown in figure 5-7. After three rounds of sorts the colonies have been screened for their activity against P0 peptide. P0-control peptide, in which lysine was replaced by glycine, on Naïve library and Sorts did not show any significant activity of the mutants toward this control peptide (Fig A4-5), showing that the peptide binding using Srt reaction was related to the Lys residue in the P0 peptide sequence. YPK is a common short sequence that is homologous among most pilin boxes, known as the minimal pilin box. Sortase reactions on the yeast surface was done against biotin conjugated YPK and its control peptide with the lysine replaced with an Ala (table 5-2). None of the peptides were shown any significant peptide attachment to the surface in presence of displayed Srt, Naive library, or after rounds of sorts (Fig 4A-3).

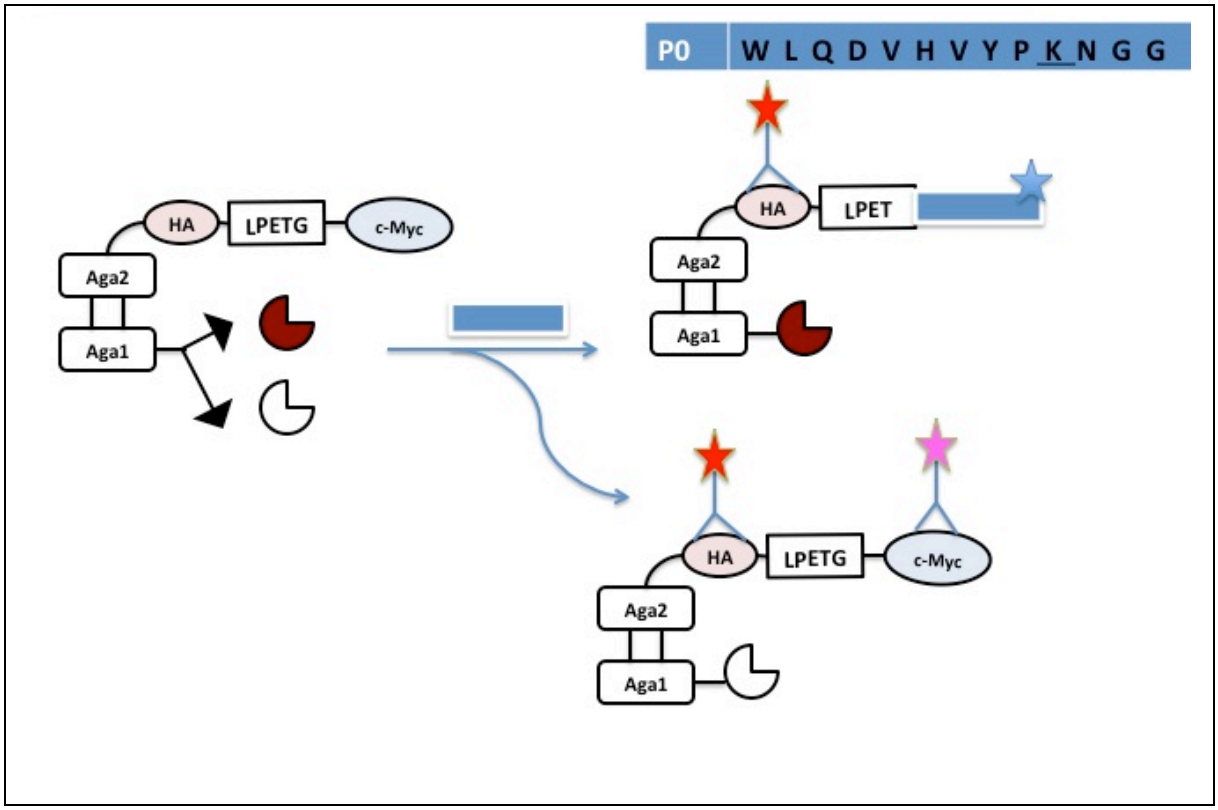


Figure 5-3. Sorting library of Srt mutants displayed on the yeast surface against P0 peptide using FACS. Active Srt (Showing with dark cut circle) will bind Biotin-P0 (the P0 peptide that has a biotin conjugated at the N-terminus) to the LPETG substrate displayed on the yeast surface.

Table 5-2- peptides used in sortase reaction on the yeast surface

Peptide	Sequence
P0	Biotin-WLQDVHVYP**K**NGG
Control P0	Biotin-WLQDVHVYP**G**NGG
3G- bio	GGGYK-Biotin
YPA	Biotin-YPA
YPK	Biotin-YPK
EDA	CH_2-CH_2-Biotin

5-3-3- Mutations

Mutations found in the colonies from round 3 of sorting the Srt library can be found in table 5-3. Among these mutations, I83T has been repeated four times, N114D three times, and K145E and N127D two times. Many of the mutations have been found from the colony screening are located close to the active site, in the β6/ β7 loop area, that is known for substrate recognition. The flow cytometry data of each mutant is shown in Fig A4-7.

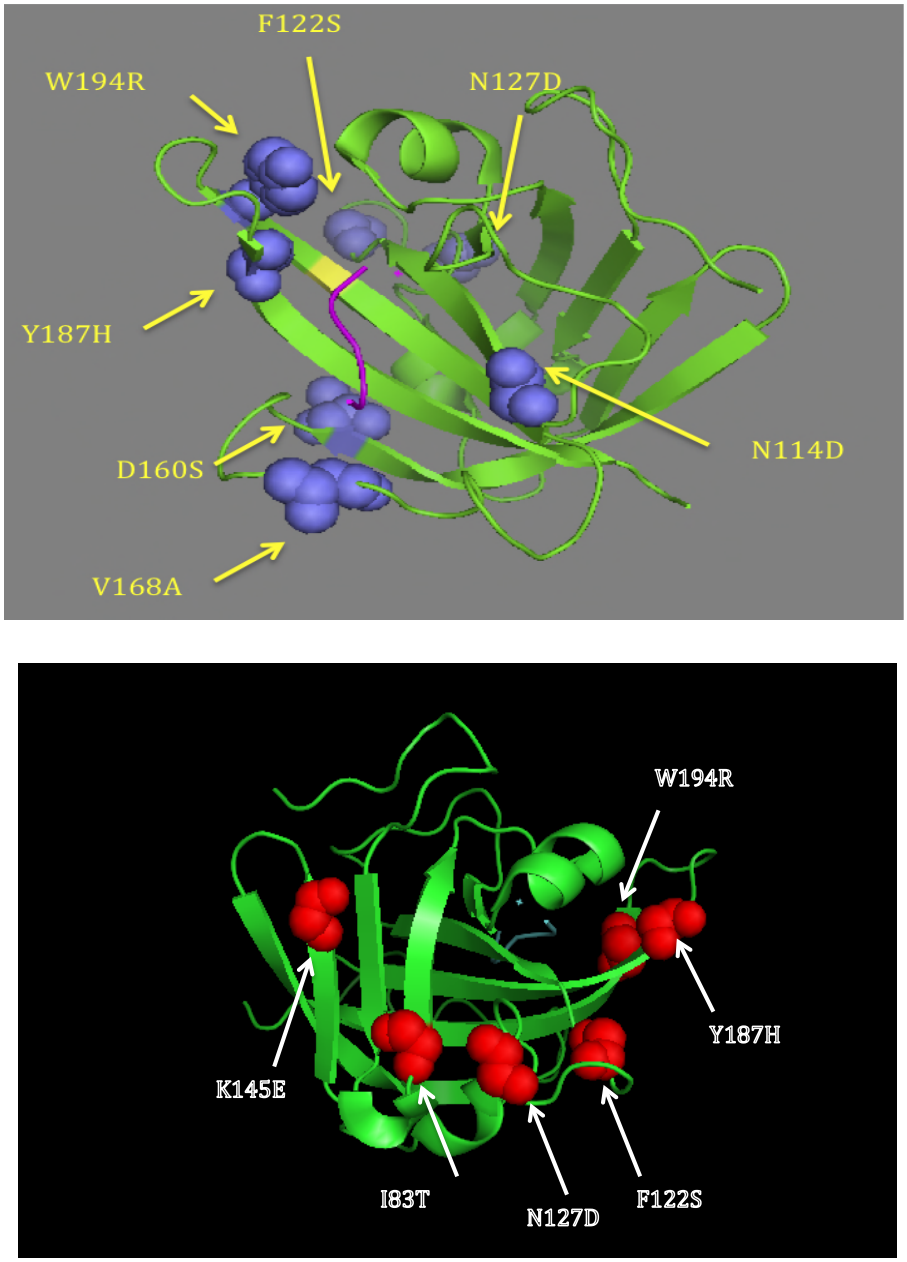


Figure 5-4 Mutations found in sortase library after 3 rounds of sorts against P0 peptide. Top: The active site (Cys184) is shown in yellow. LPETG peptide is shown in magenta. Sa Srt structure with LPETG substrate bound to it were loaded from pdb (IT2W). Pymol was used to generate the images.

Sortase mutants activity improvement for P0 peptide comparing to the WT has been determined by calculating the normalized fluorescent ratio that was corrected to the background (NCMFI) as explained in material and methods. All the experiments were done at least two times (Fig 5-4).

Y187H and W194R (in colony #7) are located close to the active site in β7/ β8 helical loop. Changing these two residues from hydrophobic to positive charge might be important in protonation mechanism of resolving the intermediate. W194 is in a close proximity to T-G bond of LPETG bound substrate to Sa Srt. W194 is also close to R197 that is known to have a role in protonation mechanism. D160 and D165 residues that have been mutated to N and A, in directed evolution of Sortase A to improve the activity, were mutated to S and V in two clones that have been screened (85). They are both located on β6/ β7 loop which is known in substrate binding and recognition. V168 that is known to bind LPETG via hydrophobic effects was also mutated to Ala in colony #7 (Fig 5-5).

Some mutations are found in the areas that are important for interaction with Ca ion and stabilizing the unstable loops in the enzyme mechanism. N114, F122, and E171 residues are shown to be important in Ca binding in Sa Srt mechanism (16). It is likely that these mutations have a role in stabilizing the structure. N114D is located close to β3/ β4 loop, in the beginning of β4 strand. The space-filling model of Sa Srt structure shows that it is exposed to the surface and located around the catalysis groove.

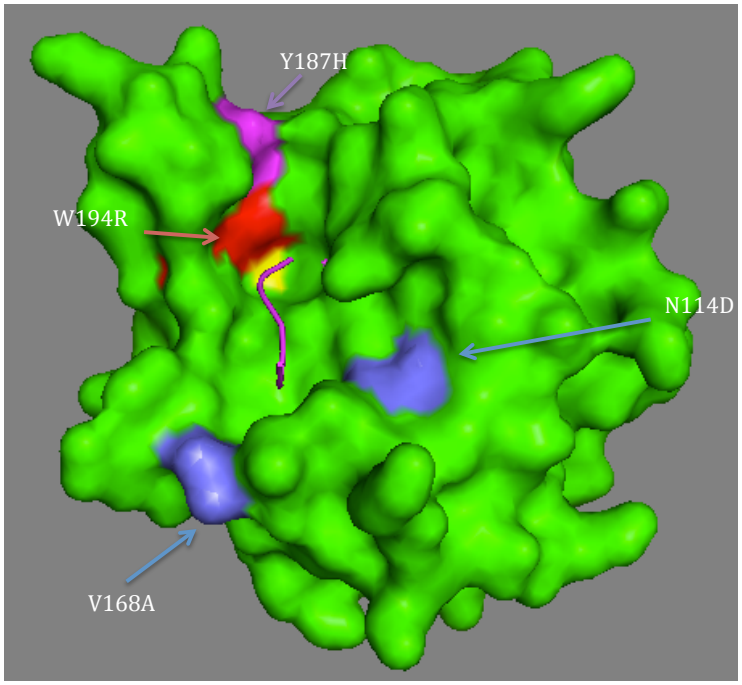


Figure 5-5. Sortase structure, space filling model, showed using Pymol. The mutations close to the active site and substrate binding site is shown in the picture.

Mutants 13 and 16 have showed a significant increase in activity for P0 peptide. They share I83T mutation, which is the single mutation in sortase mutant 8, show a little increase in activity. Mutant 6 also contains this mutation that does not show a significant increase in the activity. I83T is located in β2 strand (Fig 5-6), far from the active site. It is likely that I83T is responsible for improving the mutant stability, rather than increasing the activity.

5-4- Conclusion

Sortase-mediated ligation and immobilization on the surface is a growing area of research. In order to make a more efficient tool for biomaterial applications, we used yeast surface display to engineer sortase activity and specificity toward a new nucleophile substrate. An interesting new nucleophile substrate is a primary amine in the side chain of lysine in a pilin box sequence. This substrate can be used in generating branched protein assemblies. Although Sortase C is responsible for pilus assembly, Sa Srt also showed a little activity toward the primary amine in the side chain of lysine in a pilin box sequence.

Sortase activity against the LPETG substrate was improved using directed evolution and yeast surface display (85). Substrate specificity of sortase against LPETG substrate has been altered using Phage display (95). However, sortase substrate specificity against the nucleophile substrate has not been engineered yet. In this work, we used YSD to display sortase and LPETG substrate fused to Aga1p and Aga2p respectively. Sortase displayed at the N-terminus of Aga1p on the yeast surface site-specifically ligated the GGG-Biotin peptide to the LPETG peptide that is displayed at the C-terminus of the Aga2p.

Protein engineering might lead to instability; i.e. it is likely that more active engineered protein mutants are less stable. Therefore, we selected the sorting gate to select a wider range of sortase (AU1) expression in sortase mutants.

Sortase substrate specificity was engineered using directed evolution. A pool of sortase mutants were made by random PCR and screened using FACS. After three rounds of sorting the sortase library, colonies have been screened and analyzed. Some mutations that were found in this study are known to be responsible for ion binding in Sa Srt mechanism.

Further studies are needed to show that mutating these Ca^{2+} binding residues make Srt mutants Ca-independent.

Srtm4, naïve library and sort3 were used in reaction mixtures containing either 3G or P0 peptide, to compare substrate specificity against these two peptides. There is a significant change in nucleophile substrate binding between Srtm4 (the template for the sortase library) and sort3 of sortase library. The substrate specificity shifted from 3G substrate in Srtm4 to P0 after three rounds of sorts (Fig 5-8).

Table 5-3- Mutations from directed evolution of sortase library for pilin peptide

Clone	**Mutations**	**Fold Change in reaction***
1	E95K, R99K, E106K, N127S, D170G, E171G, K178R	1.721936
2	E85K, F122L, N127D, K145E, T144A	1.751918
4	T76V N98S N107D N114D K145E	1.028708
7	D111N, V166A, V168A, E189A, Y187H, E189A, W194R	1.793105
8	I83T	1.163059
9	K67R, M155V, N160S, Q178R, K198R, F200S	1.683238
10	T94A, S141G, D127N	2.269908
11	G74D, N98K, S102G, I99L N127D, V142A	5.876913
13	I83T F144S	2.1678
16	I83T, N114D, T150I	6.408878
17	Y88C N114D	1.245925

*: The NCMFI was calculated as explained in the material and methods section.

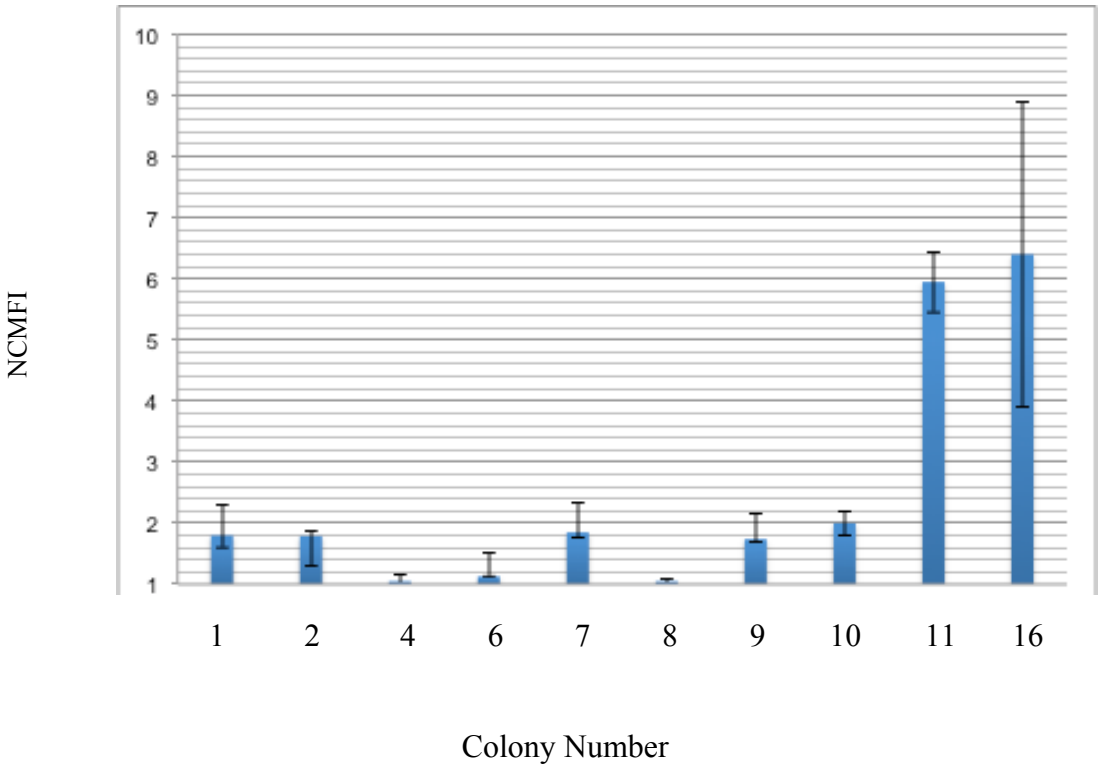


Figure 5-6. NCMFI value for different colonies. The experiments have been done at least two times.

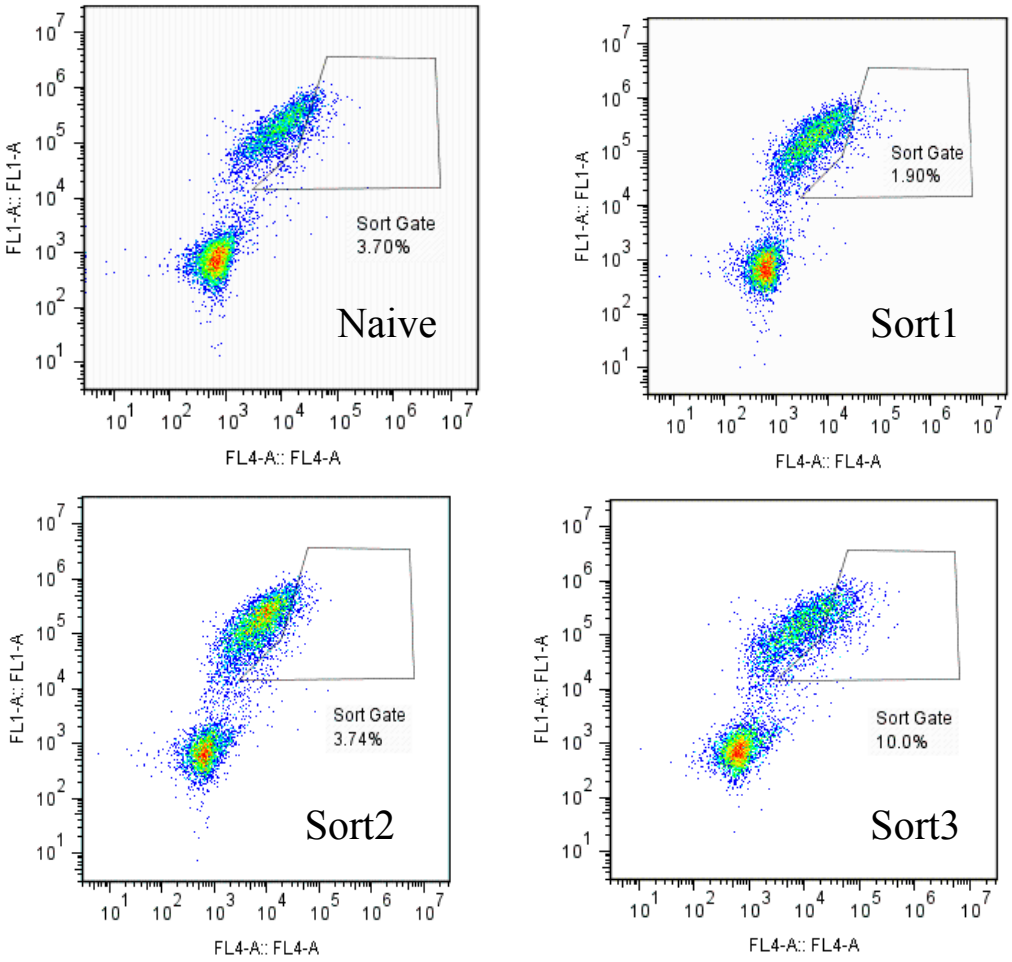


Figure 5-7. Sortase library screening against P0 peptide. Sortase library and LPETG peptide were displayed on the yeast surface. Pilin box sequence was attached to the surface using sortase reaction. The library was sorted 3 times against P0 peptide. Naïve sortase library and Sorts against P0 peptide were tested against 0.5 mM P0 peptide in the reaction mixture. FL4: Strep 647 , FL1: AU1.

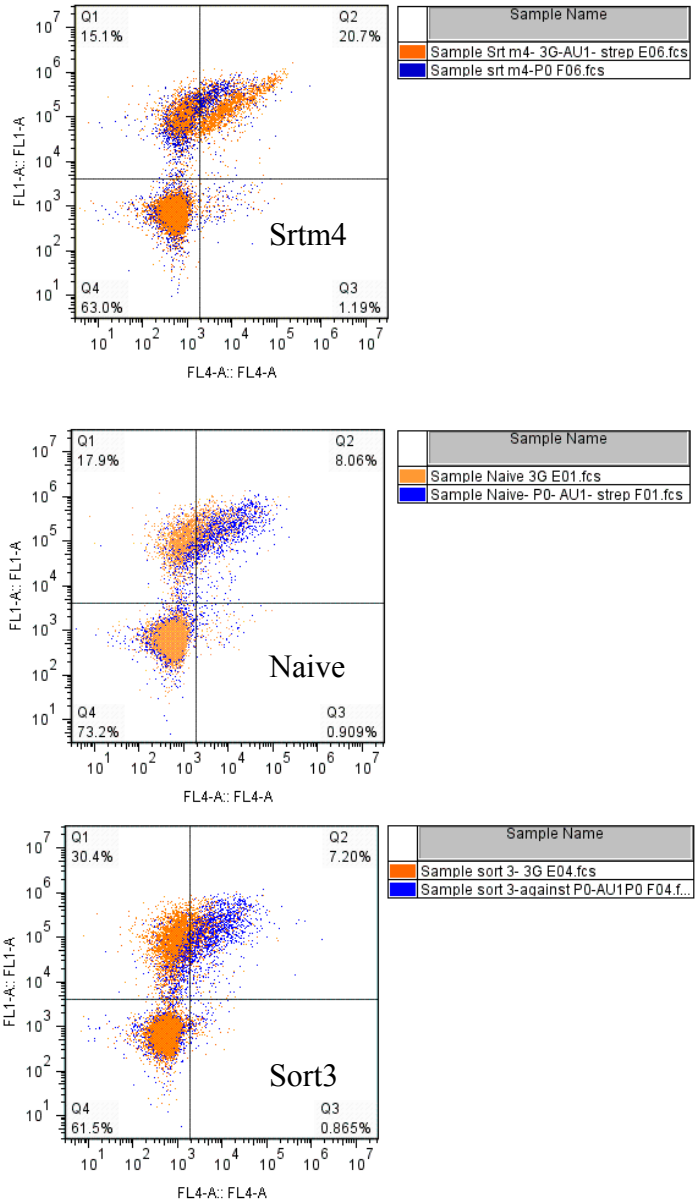


Figure 5-8. Sortase library substrate specificity against 3G-Biotin or Biotin-P0 peptide. FL4: streptavidin, FL1: AU1. The population that was reacted with P0 peptide is shown in putple, and 3G reacted population is shownin orange.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6-1- Summary and conclusion

6-1-1- Generating site-specific protein wires and 3D assemblies on the surface using Sortase A reaction

Sortase-mediate site-specific protein ligation was used to immobilize recombinant proteins on the surface. On the gold surface, a self-assembly layer of 3(G/A)C peptide was made using Au-S bond. Then recombinant fluorescent proteins with LPET(G/A) was site-specifically ligated to the immobilized peptide using Sortase A reaction. If the recombinant proteins were bifunctional (i.e., having 3(G/A) tag at the N-terminus and LPET(G/A) at the C-terminus), proteins were oligomerized upon immobilization on the surface. To have a control over immobilizing layers of proteins on the surface, we used two approaches: 1) We used bio-orthogonal sortases from two different species with slightly different substrate specificity. 2) We introduced a protecting group at the N-terminus of a bifunctional protein, which controls the protein immobilization in one direction. We immobilized layer of fluorescent protein on the surface using these methods. Each of the methods has their own advantages and disadvantages. In bio-orthogonal method, there is still some substrate specificity overlap, which prevents total control over homogeneity of each layer immobilized on the surface. The other disadvantage of this method is the limit to use enzymes with unique different substrate specificity. For instance, if we need five homogenous protein layers site-specifically immobilized on the surface, we need five

different sortases with different substrate specificity. This drawback limits generating protein molecular printboards with many layers to some extent.

The drawback of using protecting group is that sortase added in each step of the layer-by – layer protein immobilization would partially cleave off the product from the surface, since the product is also the substrate for Srt reaction, while in orthogonal method there is a better control on this aspect.

6-1-2- Engineering Sortase A substrate specificity using directed evolution

Sortase A has broad substrate specificity for nucleophile substrate. To improve Srt substrate activity and specificity toward new substrates that are beneficial for protein immobilization and generating 3D assemblies on the surface, we engineered Sortase A substrate specificity using directed evolution. A yeast surface display system was used to display sortase and LPETG substrate on Aga1p and Aga2p on the surface of the yeast cell. Sortase reaction was done on the yeast surface to covalently attach the GGG substrate to the LPETG peptide displayed on the surface. A library of Sortase A mutants and the LPETG substrate were displayed on the yeast surface and sorted against P0 peptide (pilin box of SpaA subunit in C. Diphteriae). After three rounds of sorting against this substrate, some colonies showed an improved activity against the new substrate (P0 peptide). The colonies were screened and analyzed. Some mutations found in this study are close to the substrate binding site, or located in important loops in substrate or calcium binding.

6-2- Future works

6-2-1- Further applications of protein patterning and 3D structures

We used fluorescent proteins as model proteins for site-specific protein immobilization and generating three dimensional protein assemblies. Sortase-mediated ligation and immobilization can be used for site-specific immobilization and patterning of other proteins with applications in biosensors, or tissue engineering. Fibronectin was previously immobilized on the surface in patterns to control progenitors cell differentiation into tubular structures. Fibronection or similar proteins responsible in cell adhesion and extracellular matrix formation can be sequence specifically immobilized on the surface in specific patterns using sortase reaction to control cell morphology, differentiation and movement. Antibodies or antigens can also be site-specifically immobilized on the surface for improved biosensor applications.

6-2-2- Generating smart protein assemblies

Programming the complex protein 3D assemblies to site-specifically bind to the desired site is an important area of study. The other important feature in smart assemblies is the reconfiguration ability of the assembly. Although using non covalent bonds to generate protein assemblies have the benefit of disassembly, controlling assembly/ disassembly is not totally possible. Therefore, it is important to have a method that can control switching between these two states.

Engineering sortases that activate by an external stimulus (that can even be implemented in the assembly) can be a potential tool(s) to site specifically assemble and disassemble such

protein 3D assemblies.

6-2-3- Engineering substrate specificity of Sortase

6-2-3-1- Improving Sortase activity toward amine in the side chain of K

To engineer Srt with the ability to site-specifically cross link proteins or peptides, we screened Srt library against pilin box peptide using FACS, and selected mutants with improved activity toward this substrate.

Other similar peptide to pilin box sequence with applications in generating protein assemblies may also be used as a target substrate for next round of sorting. “FLFEFKFE” peptide has been shown to make self-assembled 3D matrix scaffold by generating a right handed helical structure (97). This peptide can be used as the next target for Sorting library of Srt mutants. The engineered sortase to be more specific toward the side chain of Lys in FLFEFKFE peptide, can be used in site-specific linking of biomolecules such as probes or drugs in the 3D matrix that has been assembled using this peptide

6-2-3-2- Engineering Sortase for LPETX substrate (X: any amino acid except G and A)

Sortase can cleave off its product (LPETGGG), since it is also the enzyme substrate. Engineering sortases to be more specific toward a novel substrate other than the natural substrate of Sortase A (LPETG) is beneficial since the engineered sortase is not able to cleave off the transpeptidase reaction product, assuming GGG was used as the nucleophile substrate (LPETGGG), and this will presumably increase the yield of enzymatic reaction. The important residues in sortase structure known in LPETG substrate binding, especially

the ones close to the Threonine in the enzyme-substrate crystal structure can be the targets for saturated mutagenesis or rational design.

β6/β7 loop has been shown to be very important in recognition and binding of LPXTG substrate in sortase A. Swapping this loop with the same loop in sortase B resulted in switching substrate recognition of sortase A to sortase B (from LPXTG to NPQTN). Crystal structure studies showed that V168 and L169 residues are important in LPXTG recognition as they possibly interact with Leu and Pro in LPXTG substrate via hydrophobic interactions. Directed evolution studies showed that mutating some residues in the β6/β7 loop can make a more relaxed enzyme toward LPXTG substrate. Thr 180, Ile 182 and Ala 118 are also known to be responsible for binding of L and T of the LPXTG substrate to sortase (17,95). Therefore, it is important to study this loop and its role in substrate recognition.

A library of enzyme mutants can be made by randomizing the amino acids that are in close proximity to the T-G bond in LPXTG binding site of the enzyme. Yeast surface display can be used to alter Srt substrate specificity to LPETV or similar substrates (17).

6-2-3-3- Engineering sortase subsatrate specificity

It has been previously shown that sortase reaction can work *in vivo;* inside the ER or in cytosol to circularize GFP proteins that have G at N-terminus and LPETG at the C-terminus (98) This methodology can be used to investigate sortase substrate specificity against different amino acids acting as nucleophile substrate. Saturation mutagenesis can be used to generate a library of sortase substrate. If a yeast surface display system is used to display the product on the surface, then sortase activity and specificity can be quantified using flow cytometry. Yeast surface display of sortase reaction product can also be used to engineer

sortase activity and specificity toward the desired peptide using rounds of positive and negative sorts.

LIST OF REFERENCES

1. Wong, L. S., Khan, F., and Micklefield, J. (2009) *Chem Rev* **109**, 4025-4053
2. Lee, Y., Lee, E. K., Cho, Y. W., Matsui, T., Kang, I. C., Kim, T. S., and Han, M. H. (2003) *Proteomics* **3**, 2289-2304
3. MacBeath, G. (2002) *Nature genetics* **32 Suppl**, 526-532
4. Guttilla, I. K., Gaspar, A. H., Swierczynski, A., Swaminathan, A., Dwivedi, P., Das, A., and Ton-That, H. (2009) *Journal of bacteriology* **191**, 5603-5612
5. Hendrickx, A. P. A., Budzik, J. M., Oh, S. Y., and Schneewind, O. (2011) *Nat Rev Microbiol* **9**, 166-176
6. Spirig, T., Weiner, E. M., and Clubb, R. T. (2011) *Molecular microbiology* **82**, 1044-1059
7. Marraffini, L. A., Dedent, A. C., and Schneewind, O. (2006) *Microbiology and molecular biology reviews : MMBR* **70**, 192-221
8. Paterson, G. K., and Mitchell, T. J. (2004) *Trends in microbiology* **12**, 89-95
9. Ton-That, H., Marraffini, L. A., and Schneewind, O. (2004) *Biochimica et biophysica acta* **1694**, 269-278
10. Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., and Ellestad, G. (2003) *Biochemistry* **42**, 11307-11315
11. Mazmanian, S. K., Ton-That, H., and Schneewind, O. (2001) *Molecular microbiology* **40**, 1049-1057
12. Ton-That, H., Mazmanian, S. K., Alksne, L., and Schneewind, O. (2002) *The Journal of biological chemistry* **277**, 7447-7452
13. Maresso, A. W., and Schneewind, O. (2008) *Pharmacological reviews* **60**, 128-141
14. Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F., and Schneewind, O. (1999) *Proceedings of the National Academy of Sciences of the United States of America* **96**, 12424-12429
15. Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O., and Narayana, S. V. (2004) *The Journal of biological chemistry* **279**, 31383-31389
16. Naik, M. T., Suree, N., Ilangovan, U., Liew, C. K., Thieu, W., Campbell, D. O., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2006) *The Journal of biological chemistry* **281**, 1817-1826
17. Bentley, M. L., Gaweska, H., Kielec, J. M., and McCafferty, D. G. (2007) *The Journal of biological chemistry* **282**, 6571-6581
18. Kruger, R. G., Otvos, B., Frankel, B. A., Bentley, M., Dostal, P., and McCafferty, D. G. (2004) *Biochemistry* **43**, 1541-1551
19. Antos, J. M., Popp, M. W., Ernst, R., Chew, G. L., Spooner, E., and Ploegh, H. L. (2009) *The Journal of biological chemistry* **284**, 16028-16036
20. Race, P. R., Bentley, M. L., Melvin, J. A., Crow, A., Hughes, R. K., Smith, W. D., Sessions, R. B., Kehoe, M. A., McCafferty, D. G., and Banfield, M. J. (2009) *The Journal of biological chemistry* **284**, 6924-6933
21. Yamamura, Y., Hirakawa, H., Yamaguchi, S., and Nagamune, T. (2011) *Chem Commun (Camb)* **47**, 4742-4744
22. Samantaray, S., Marathe, U., Dasgupta, S., Nandicoori, V. K., and Roy, R. P. (2008) *Journal of the American Chemical Society* **130**, 2132-2133
23. Gaspar, A. H., and Ton-That, H. (2006) *Journal of bacteriology* **188**, 1526-1533
24. Ton-That, H., and Schneewind, O. (2003) *Molecular microbiology* **50**, 1429-1438

25. Cozzi, R., Prigozhin, D., Rosini, R., Abate, F., Bottomley, M. J., Grandi, G., Telford, J. L., Rinaudo, C. D., Maione, D., and Alber, T. (2012) *PloS one* **7**, e49048
26. Clancy, K. W., Melvin, J. A., and McCafferty, D. G. (2010) *Biopolymers* **94**, 385-396
27. Steinhagen, M., Zunker, K., Nordsieck, K., and Beck-Sickinger, A. G. (2013) *Bioorganic & medicinal chemistry* **21**, 3504-3510
28. Theile, C. S., Witte, M. D., Blom, A. E., Kundrat, L., Ploegh, H. L., and Guimaraes, C. P. (2013) *Nature protocols* **8**, 1800-1807
29. Popp, M. W., Antos, J. M., and Ploegh, H. L. (2009) *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **Chapter 15**, Unit 15 13
30. Guimaraes, C. P., Witte, M. D., Theile, C. S., Bozkurt, G., Kundrat, L., Blom, A. E., and Ploegh, H. L. (2013) *Nature protocols* **8**, 1787-1799
31. Williamson, D. J., Fascione, M. A., Webb, M. E., and Turnbull, W. B. (2012) *Angew Chem Int Ed Engl* **51**, 9377-9380
32. Hess, G. T., Cragnolini, J. J., Popp, M. W., Allen, M. A., Dougan, S. K., Spooner, E., Ploegh, H. L., Belcher, A. M., and Guimaraes, C. P. (2012) *Bioconjugate chemistry* **23**, 1478-1487
33. Mao, H., Hart, S. A., Schink, A., and Pollok, B. A. (2004) *Journal of the American Chemical Society* **126**, 2670-2671
34. Parthasarathy, R., Subramanian, S., and Boder, E. T. (2007) *Bioconjugate chemistry* **18**, 469-476
35. Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E., and Ploegh, H. L. (2007) *Nature chemical biology* **3**, 707-708
36. Love, J. C., Estroff, L. A., Kriebel, J. K., Nuzzo, R. G., and Whitesides, G. M. (2005) *Chem Rev* **105**, 1103-1169
37. Frasconi, M., Mazzei, F., and Ferri, T. (2010) *Anal Bioanal Chem* **398**, 1545-1564
38. Nakanishi, J., Takarada, T., Yamaguchi, K., and Maeda, M. (2008) *Analytical sciences : the international journal of the Japan Society for Analytical Chemistry* **24**, 67-72
39. Ichihara, T., Akada, J. K., Kamei, S., Ohshiro, S., Sato, D., Fujimoto, M., Kuramitsu, Y., and Nakamura, K. (2006) *Journal of proteome research* **5**, 2144-2151
40. Terrettaz, S., Ulrich, W. P., Vogel, H., Hong, Q., Dover, L. G., and Lakey, J. H. (2002) *Protein science : a publication of the Protein Society* **11**, 1917-1925
41. Nakaji-Hirabayashi, T., Kato, K., Arima, Y., and Iwata, H. (2007) *Biomaterials* **28**, 3517-3529
42. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R. A., Gerstein, M., and Snyder, M. (2001) *Science* **293**, 2101-2105
43. Kazuhiro Nakanishi, T. S., Yoichi Kumada, Koreyoshi Imamura, and Hiroyuki Imanaka. (2008) *Current Proteomics* **5**, 161-175
44. Kamiya, N., Doi, S., Tominaga, J., Ichinose, H., and Goto, M. (2005) *Biomacromolecules* **6**, 35-38
45. Tanaka, Y., Tsuruda, Y., Nishi, M., Kamiya, N., and Goto, M. (2007) *Organic & biomolecular chemistry* **5**, 1764-1770

46. Rashidian, M., Song, J. M., Pricer, R. E., and Distefano, M. D. (2012) *Journal of the American Chemical Society* **134**, 8455-8467
47. Tsukiji, S., and Nagamune, T. (2009) *Chembiochem : a European journal of chemical biology* **10**, 787-798
48. Chan, L., Cross, H. F., She, J. K., Cavalli, G., Martins, H. F., and Neylon, C. (2007) *PloS one* **2**, e1164
49. Clow, F., Fraser, J. D., and Proft, T. (2008) *Biotechnology letters* **30**, 1603-1607
50. Ito, T., Sadamoto, R., Naruchi, K., Togame, H., Takemoto, H., Kondo, H., and Nishimura, S. (2010) *Biochemistry* **49**, 2604-2614
51. Yap, F. L., and Zhang, Y. (2007) *Biosensors & bioelectronics* **22**, 775-788
52. Hultschig, C., Kreutzberger, J., Seitz, H., Konthur, Z., Bussow, K., and Lehrach, H. (2006) *Current opinion in chemical biology* **10**, 4-10
53. Kam, L., and Boxer, S. G. (2001) *Journal of biomedical materials research* **55**, 487-495
54. Kung, L. A., Kam, L., Hovis, J. S., and Boxer, S. G. (2000) *Langmuir* **16**, 6773-6776
55. Volcke, C., Gandhiraman, R. P., Basabe-Desmonts, L., Iacono, M., Gubala, V., Cecchet, F., Cafolla, A. A., and Williams, D. E. (2010) *Biosensors & bioelectronics* **25**, 1295-1300
56. Shen, K., Thomas, V. K., Dustin, M. L., and Kam, L. C. (2008) *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7791-7796
57. Sinisi, A., Popp, M. W., Antos, J. M., Pansegrau, W., Savino, S., Nissum, M., Rappuoli, R., Ploegh, H. L., and Buti, L. (2012) *Bioconjugate chemistry*
58. de la Rica, R., and Matsui, H. (2010) *Chemical Society reviews* **39**, 3499-3509
59. Petrov, A., and Audette, G. F. (2012) *Wiley interdisciplinary reviews. Nanomedicine and nanobiotechnology* **4**, 575-585
60. Muller, M. K., Petkau, K., and Brunsveld, L. (2011) *Chem Commun (Camb)* **47**, 310-312
61. Jones, E. Y., and Yeates, T. O. (2012) *Current opinion in structural biology* **22**, 197-199
62. Timpl, R., and Brown, J. C. (1996) *BioEssays : news and reviews in molecular, cellular and developmental biology* **18**, 123-132
63. Ludden, M. J., Mulder, A., Tampe, R., Reinhoudt, D. N., and Huskens, J. (2007) *Angew Chem Int Ed Engl* **46**, 4104-4107
64. Gonzalez-Campo, A., Hsu, S. H., Puig, L., Huskens, J., Reinhoudt, D. N., and Velders, A. H. (2010) *Journal of the American Chemical Society* **132**, 11434-11436
65. Crespo-Biel, O., Jan Ravoo, B., Huskens, J., and Reinhoudt, D. N. (2006) *Dalton Trans*, 2737-2741
66. Boder, E. T., Raeeszadeh-Sarmazdeh, M., and Price, J. V. (2012) *Archives of biochemistry and biophysics* **526**, 99-106
67. Boder, E. T., and Jiang, W. (2011) *Annual review of chemical and biomolecular engineering* **2**, 53-75
68. Boder, E. T., and Wittrup, K. D. (1997) *Nature biotechnology* **15**, 553-557
69. Pepper, L. R., Cho, Y. K., Boder, E. T., and Shusta, E. V. (2008) *Combinatorial chemistry & high throughput screening* **11**, 127-134

70. Liu, Y. C., Wang, C. M., and Hsiung, K. P. (2001) *Analytical biochemistry* **299**, 130-135
71. Rusmini, F., Zhong, Z., and Feijen, J. (2007) *Biomacromolecules* **8**, 1775-1789
72. Frolov, L., Wilner, O., Carmeli, C., and Carmeli, I. (2008) *Adv Mater* **20**, 263-+
73. Park, K., Lee, J. M., Jung, Y., Habtemariam, T., Salah, A. W., Fermin, C. D., and Kim, M. (2011) *The Analyst* **136**, 2506-2511
74. Carmeli, I., Frolov, L., Carmeli, C., and Richter, S. (2007) *Journal of the American Chemical Society* **129**, 12352-12353
75. Lee, J. M., Park, H. K., Jung, Y., Kim, J. K., Jung, S. O., and Chung, B. H. (2007) *Analytical chemistry* **79**, 2680-2687
76. Soellner, M. B., Dickson, K. A., Nilsson, B. L., and Raines, R. T. (2003) *Journal of the American Chemical Society* **125**, 11790-11791
77. Jiang, R., Weingart, J., Zhang, H., Ma, Y., and Sun, X. L. (2012) *Bioconjugate chemistry* **23**, 643-649
78. Wammes, A. E., Fischer, M. J., de Mol, N. J., van Eldijk, M. B., Rutjes, F. P., van Hest, J. C., and van Delft, F. L. (2013) *Lab on a chip* **13**, 1863-1867
79. Wasserberg, D., Nicosia, C., Tromp, E. E., Subramaniam, V., Huskens, J., and Jonkheijm, P. (2013) *Journal of the American Chemical Society* **135**, 3104-3111
80. Seo, M. H., Han, J., Jin, Z., Lee, D. W., Park, H. S., and Kim, H. S. (2011) *Analytical chemistry* **83**, 2841-2845
81. Sijbrandij, T., Cukkemane, N., Nazmi, K., Veerman, E. C. I., and Bikker, F. J. (2013) *Bioconjugate chemistry* **24**, 828-831
82. Matsumoto, T., Tanaka, T., and Kondo, A. (2012) *Langmuir* **28**, 3553-3557
83. Ton-That, H., Mazmanian, S. K., Faull, K. F., and Schneewind, O. (2000) *The Journal of biological chemistry* **275**, 9876-9881
84. Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999) *Science* **285**, 760-763
85. Chen, I., Dorr, B. M., and Liu, D. R. (2011) *Proceedings of the National Academy of Sciences of the United States of America* **108**, 11399-11404
86. Liu, Z., Zu, Y., Fu, Y., Zhang, Z., and Meng, R. (2008) *Microscopy research and technique* **71**, 802-809
87. Piluso, S., Cassell, H. C., Gibbons, J. L., Waller, T. E., Plant, N. J., Miller, A. F., and Cavalli, G. (2013) *Soft matter* **9**, 6752-6756
88. Liu, F., Luo, E. Y., Flora, D. B., and Mezo, A. R. (2014) *The Journal of organic chemistry*
89. Levary, D. A., Parthasarathy, R., Boder, E. T., and Ackerman, M. E. (2011) *PloS one* **6**, e18342
90. Feinberg, A. W., and Parker, K. K. (2010) *Nano letters* **10**, 2184-2191
91. Bloom, J. D., Meyer, M. M., Meinhold, P., Otey, C. R., MacMillan, D., and Arnold, F. H. (2005) *Current opinion in structural biology* **15**, 447-452
92. Chica, R. A., Doucet, N., and Pelletier, J. N. (2005) *Current opinion in biotechnology* **16**, 378-384
93. Lutz, S. (2010) *Current opinion in biotechnology* **21**, 734-743
94. Dasgupta, S., Samantaray, S., Sahal, D., and Roy, R. P. (2011) *The Journal of biological chemistry* **286**, 23996-24006

95. Piotukh, K., Geltinger, B., Heinrich, N., Gerth, F., Beyermann, M., Freund, C., and Schwarzer, D. (2011) *Journal of the American Chemical Society* **133**, 17536-17539
96. Chao, G., Lau, W. L., Hackel, B. J., Sazinsky, S. L., Lippow, S. M., and Wittrup, K. D. (2006) *Nature protocols* **1**, 755-768
97. Zhang, S., Marini, D. M., Hwang, W., and Santoso, S. (2002) *Current opinion in chemical biology* **6**, 865-871
98. Strijbis, K., Spooner, E., and Ploegh, H. L. (2012) *Traffic* **13**, 780-789

APPENDIX

APPENDIX 1

ATOMIC FORCE MICROSCOPY IMAGES

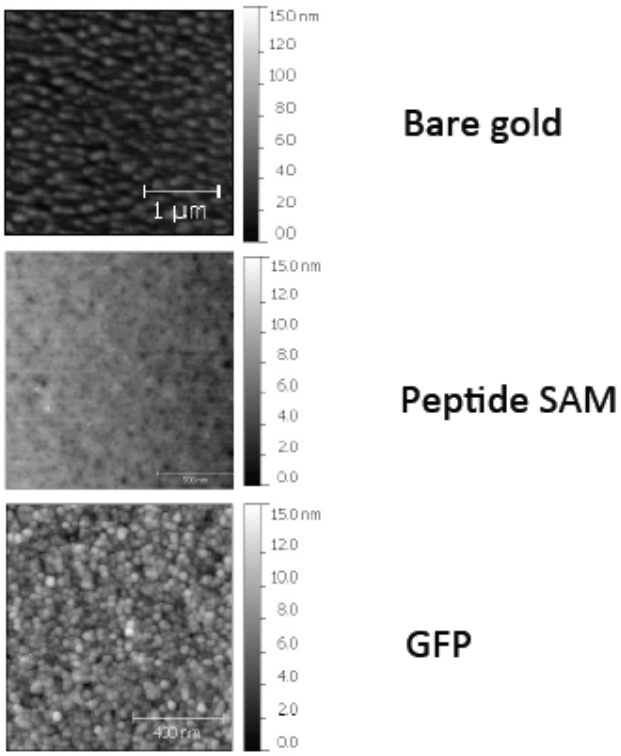


Figure A1-1- Atomic force microcopy for bare gold, peptide self assembly, and GFP-LPETG layer on the gold surface.

APPENDIX 2

PLASMID MAPS

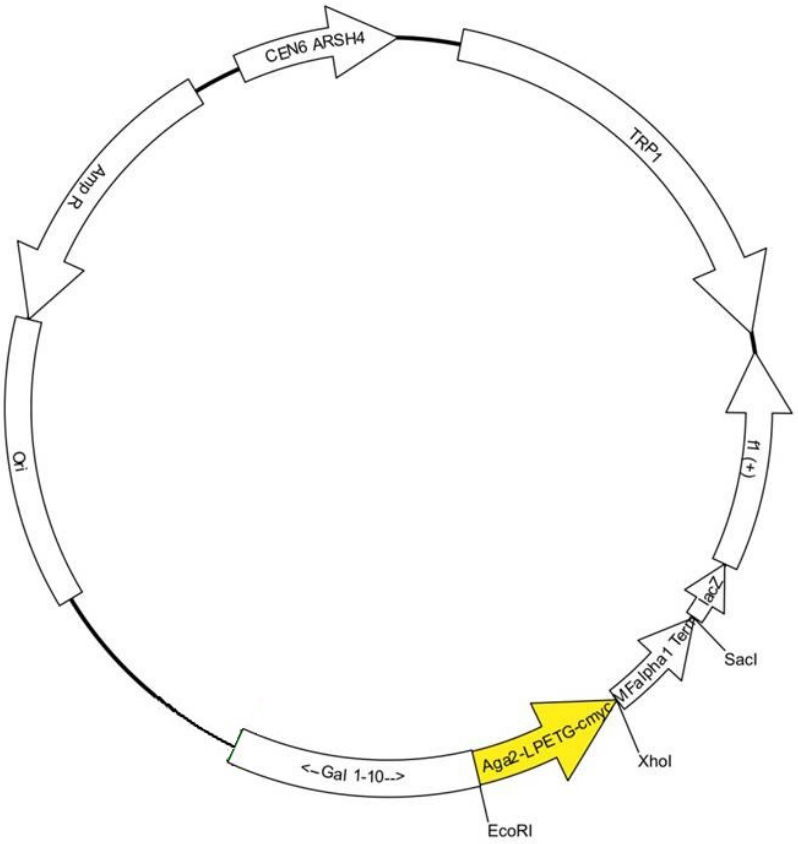


Figure A2-1- Map of pMR2 plasmid

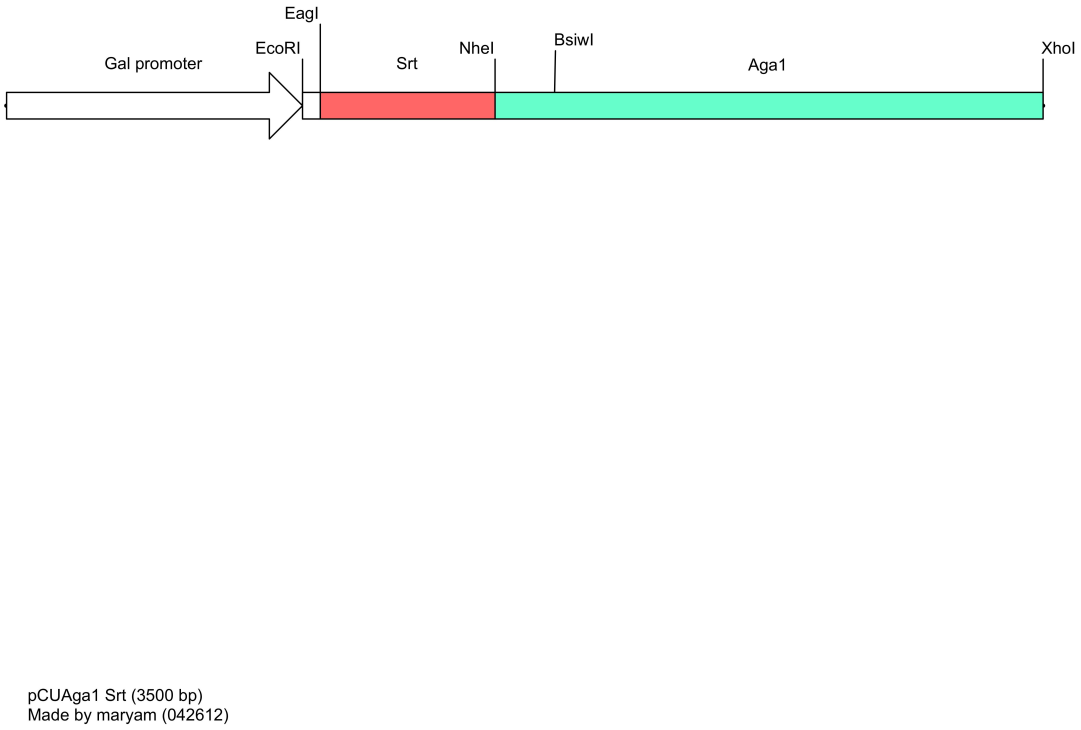


Figure A2-2- pCUAga1Srt construct

APPENDIX 3

SEQUENCES

Sortase A sequence ---pMR5- cloned between NdeI and XhoI—backbone pET vector

== Linear Map of Sequence:

```
             DraI            XbaI         BarI                      NdeI
             \               \            \                         \
      1   gtkwaaasgggtcattccctctagaataatttgtttaactttaagaaggagatatacat     60
          camwtttscccagtaagggagatcttattaaaacaaattgaaattcttcctctatatgta
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         X  X  X  G  H  S  L  *  N  N  F  V  *  L  *  E  G  D  I  H

                  BarI                 BamHI
                  \                    \
     61   atgagaggatcgcatcaccatcaccatcacggatcccaagctaaacctcaaattccgaaa    120
          tactctcctagcgtagtggtagtggtagtgcctagggttcgatttggagtttaaggcttt
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         M  R  G  S  H  H  H  H  H  H  G  S  Q  A  K  P  Q  I  P  K

                                                AlwNI
                                                \
    121   gataaatcgaaagtggcaggctatattgaaattccagatgctgatattaaagaaccagta    180
          ctatttagctttcaccgtccgatataactttaaggtctacgactataatttcttggtcat
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         D  K  S  K  V  A  G  Y  I  E  I  P  D  A  D  I  K  E  P  V

            PfoI                                  HindIII
            \                                     \
    181   tatccaggaccagcaacacctgaacaattaaatagaggtgtaagctttgcagaagaaaat    240
          ataggtcctggtcgttgtggacttgttaatttatctccacattcgaaacgtcttctttta
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         Y  P  G  P  A  T  P  E  Q  L  N  R  G  V  S  F  A  E  E  N

                     BclI     SspI MfeI
                     \        \    \
    241   gaatcactagatgatcaaaatatttcaattgcaggacacactttcattgaccgtccgaac    300
          cttagtgatctactagttttataaagttaacgtcctgtgtgaaagtaactggcaggcttg
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         E  S  L  D  D  Q  N  I  S  I  A  G  H  T  F  I  D  R  P  N

                                                            DraI
                                                            \
    301   tgtcaatttacaaatcttaaagcagccaaaaaaggtagtatggtgtactttaaagttggt    360
          acagttaaatgtttagaatttcgtcggtttttccatcataccacatgaaatttcaacca
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         C  Q  F  T  N  L  K  A  A  K  K  G  S  M  V  Y  F  K  V  G
```

```
    361   aatgaaacacgtaagtataaaatgacaagtataagagatgttaagcctacagatgtagga    420
          ttactttgtgcattcatattttactgttcatattctctacaattcggatgtctacatcct
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         N  E  T  R  K  Y  K  M  T  S  I  R  D  V  K  P  T  D  V  G

            XbaI                                AseI
            \                                   \
    421   gttctagatgaacaaaaaggtaaagataaacaattaacattaattacttgtgatgattac    480
          caagatctacttgtttttccatttctatttgttaattgtaattaatgaacactactaatg
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         V  L  D  E  Q  K  G  K  D  K  Q  L  T  L  I  T  C  D  D  Y


    481   aatgaaaagacaggcgtttgggaaaaacgtaaaatctttgtagctacagaagtcaaataa    540
          ttacttttctgtccgcaaacccttttttgcattttagaaacatcgatgtcttcagtttatt
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         N  E  K  T  G  V  W  E  K  R  K  I  F  V  A  T  E  V  K  *

                   XhoI
          BamHI    PspXI
          \        \
    541   ggatcctaactcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccga    600
          cctaggattgagctcgtggtggtggtggtggtgactctaggccgacgattgtttcgggct
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         G  S  *  L  E  H  H  H  H  H  H  *  D  P  A  A  N  K  A  R
```

The sequence for 3A-mcherry-LPETA (between NdeI and XhoI)

```
                         XbaI         BarI                      NdeI
                         \            \                         \
      1   taaaacggtcattccctctagaataattttgtttaactttaagaaggagatatacatatg     60
          attttgccagtaagggagatcttattaaaacaaattgaaattcttcctctatatgtatac
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         *  N  G  H  S  L  *  N  N  F  V  *  L  *  E  G  D  I  H  M

                                                MscI
                BarI   BtsI                     BseRI              Start codon
                \      \                        \
     61   gcagcagcagtgagcaagaagggcgaggaggataacatggccatcatcaaggagttcatg    120
          cgtcgtcgtcactcgttcttcccgctcctcctattgtaccggtagtagttcctcaagtac
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         A  A  A  V  S  K  K  G  E  E  D  N  M  A  I  I  K  E  F  M
          |AAA-RFP-→
                   ApaLI                      BssSI
                   \                          \
    121   cgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgag    180
          gcgaagttccacgtgtacctcccgaggcacttgccggtgctcaagctctagctcccgctc
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         R  F  K  V  H  M  E  G  S  V  N  G  H  E  F  E  I  E  G  E

                                                     BstEII
                                                     \
    181   ggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggc    240
          ccgctcccggcggggatgctcccgtgggtctggcggttcgacttccactggttcccaccg
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         G  E  G  R  P  Y  E  G  T  Q  T  A  K  L  K  V  T  K  G  G

          AcuI                   AhdI                           StuI
          \                      \                              \
    241   cccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctac    300
          ggggacgggaagcggaccctgtaggacaggggagtcaagtacatgccgaggttccggatg
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         P  L  P  F  A  W  D  I  L  S  P  Q  F  M  Y  G  S  K  A  Y

                                                                 FalI
                                                                 \
    301   gtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaag    360
          cacttcgtggggcggctgtaggggctgatgaacttcgacaggaaggggctcccgaagttc
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         V  K  H  P  A  D  I  P  D  Y  L  K  L  S  F  P  E  G  F  K

                                            BstEII BseRI     AloI
                                            \      \         \
    361   tgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctcc    420
          accctcgcgcactacttgaagctcctgccgccgcaccactggcactgggtcctgaggagg
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         W  E  R  V  M  N  F  E  D  G  G  V  V  T  V  T  Q  D  S  S

              PstI                          BsaXI
              SbfI                      AloI
              \                         \   \
    421   ctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgac    480
          gacgtcctgccgctcaagtagatgttccacttcgacgcgccgtggttgaaggggaggctg
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         L  Q  D  G  E  F  I  Y  K  V  K  L  R  G  T  N  F  P  S  D

                                      BseRI
```

```
                                    GsaI
                              NcoI  PflMI   StuI
                BsaXI            BbsI    BseYI          BsrBI    BplI
                \               \ \  \\   \   \          \        \
    481   ggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtacccc    540
          ccggggcattacgtcttcttctggtacccgaccctccggaggaggctcgcctacatgggg
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         G  P  V  M  Q  K  K  T  M  G  W  E  A  S  S  E  R  M  Y  P

                  SfoI
                 NarI                        AlwNI
                KasI                       AcuI                    AcuI
                \\\                        \ \                     \
    541   gaggacggcgccctgaagggcgagatcaagcagaggctgaagctgaaggacggcggccac    600
          ctcctgccgcgggacttcccgctctagttcgtctccgacttcgacttcctgccgccggtg
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         E  D  G  A  L  K  G  E  I  K  Q  R  L  K  L  K  D  G  G  H

                    Bpu10I                                       SfoI
                    BbvCI                                       NarI
             AcuI  DrdI                                  PvuII  KasI
             \     \\                                    \      \\\
    601   tacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcc    660
          atgctgcgactccagttctggtggatgttccggttcttcgggcacgtcgacgggccgcgg
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         Y  D  A  E  V  K  T  T  Y  K  A  K  K  P  V  Q  L  P  G  A

             BsgI  BsaXI                            BsaXI    AleI
             \     \                                \        \
    661   tacaacgtcaacatcaagttggacatcacctcccacaacgaggactacaccatcgtggaa    720
          atgttgcagttgtagttcaacctgtagtggagggtgttgctcctgatgtggtagcacctt
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         Y  N  V  N  I  K  L  D  I  T  S  H  N  E  D  Y  T  I  V  E

                                                          BsgI
                                          XcmI                HindIII
                                      SgrAI   NmeAIII    BsrGI
                                      \    \  \          \\   \
    721   cagtacgaacgcgccgagggccgccactccaccggcggcatggacgagctgtacaagctt    780
          gtcatgcttgcgcggctcccggcggtgaggtggccgccgtacctgctcgacatgttcgaa
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         Q  Y  E  R  A  E  G  R  H  S  T  G  G  M  D  E  L  Y  K  L
                                                          RFP--
                      XhoI
                      PspXI
                      \
    781   cctgagactgcactcgagcaccaccaccaccaccactgagatccggctgctaacaaagcc    840
          ggactctgacgtgagctcgtggtggtggtggtggtgactctaggccgacgattgtttcgg
              ^    *    ^    *    ^    *    ^    *    ^    *    ^    *
1         P  E  T  A  L  E  H  H  H  H  H  H  *  D  P  A  A  N  K  A
           -- LPETA |      |His 6 tag-------|
                                        BlpI  Stop codon
```

APPENDIX 4

FLOW CYTOMETRY DATA

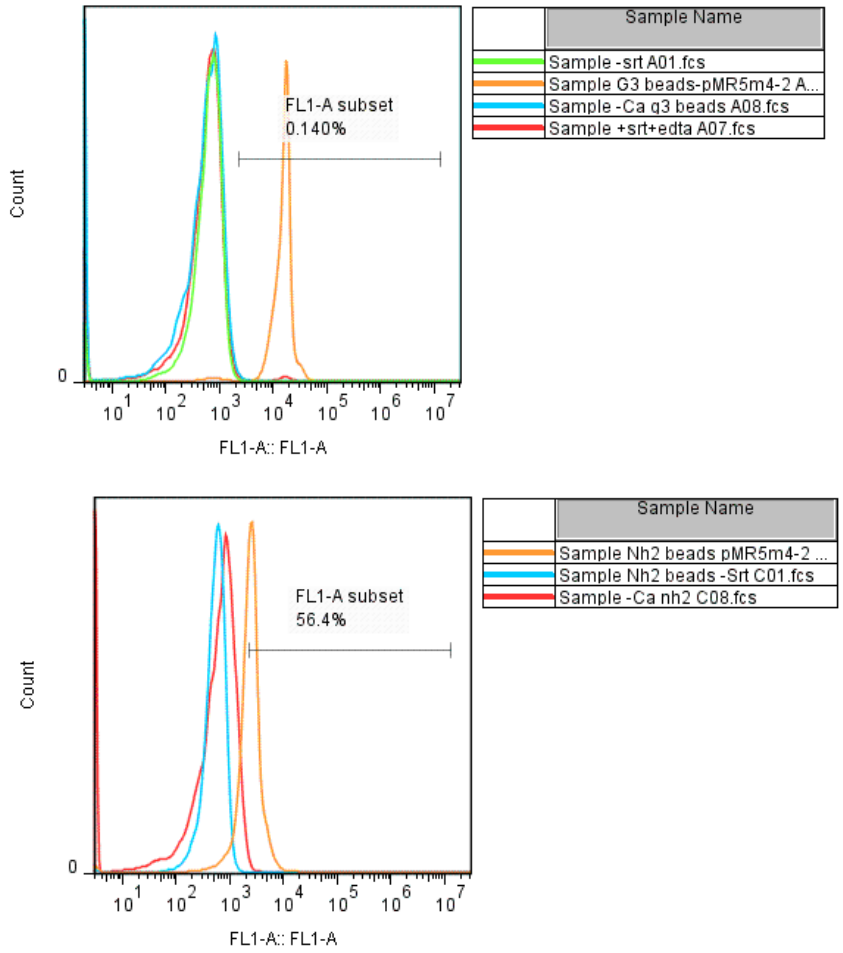


Figure A4-1. GFP immobilization on polystyrene beads.

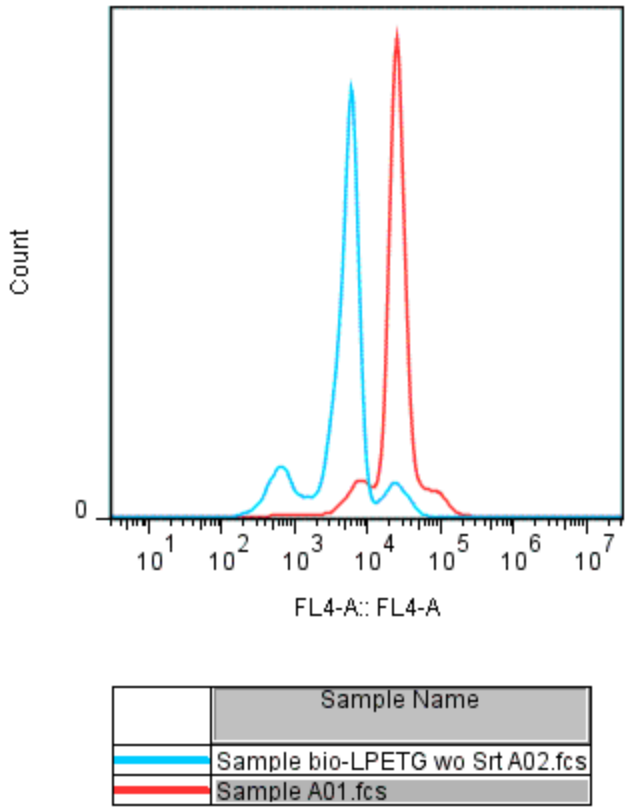


Figure A4-2 . Biotin-LPETG peptide immobilization on 3G polystyrene beads.

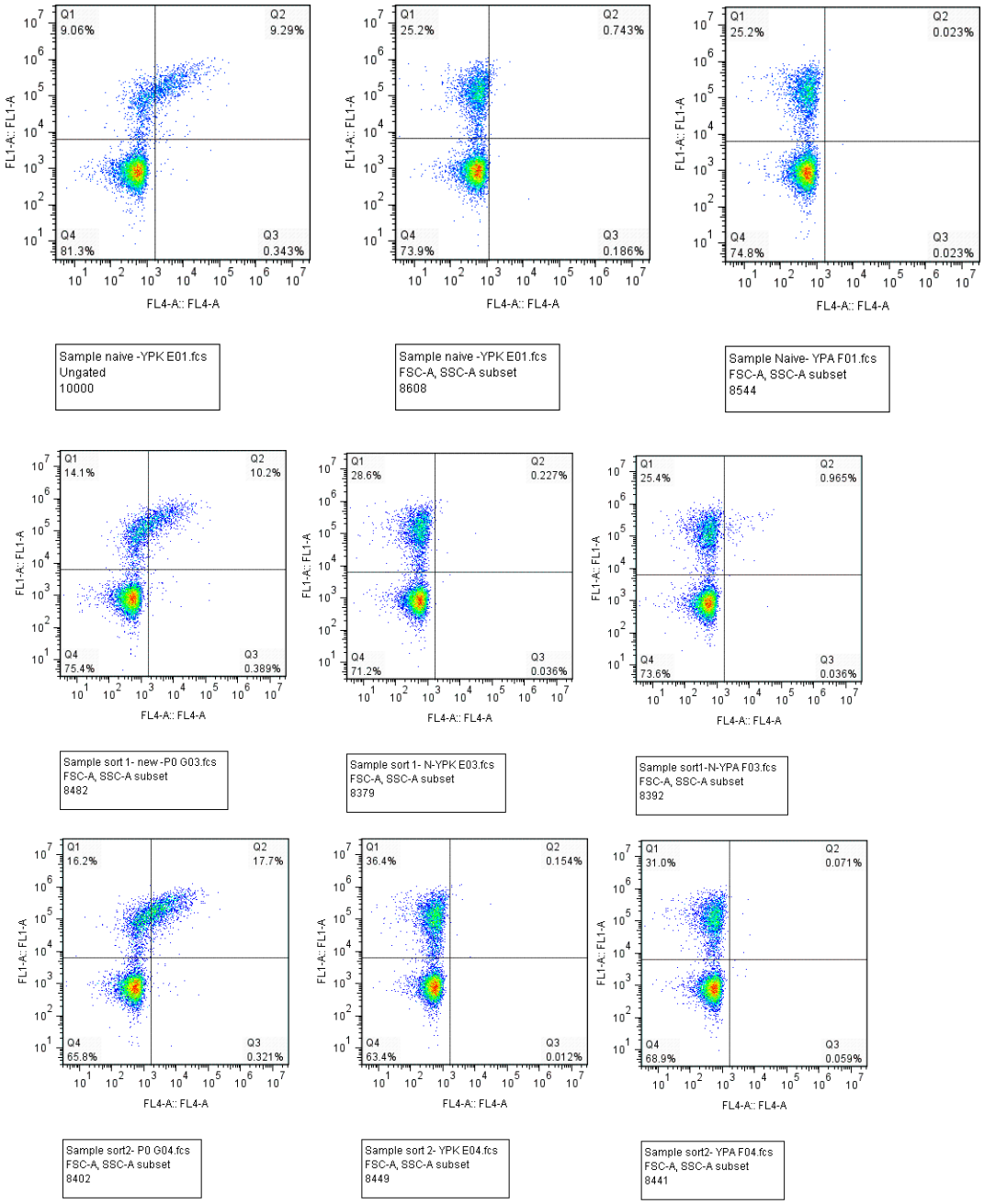


Figure A4-3. Srt library sorts against P0 peptide comparing to YPA and YPK peptides. FL1: AU1 expression, FL4: peptide binding ; streptavidin 647

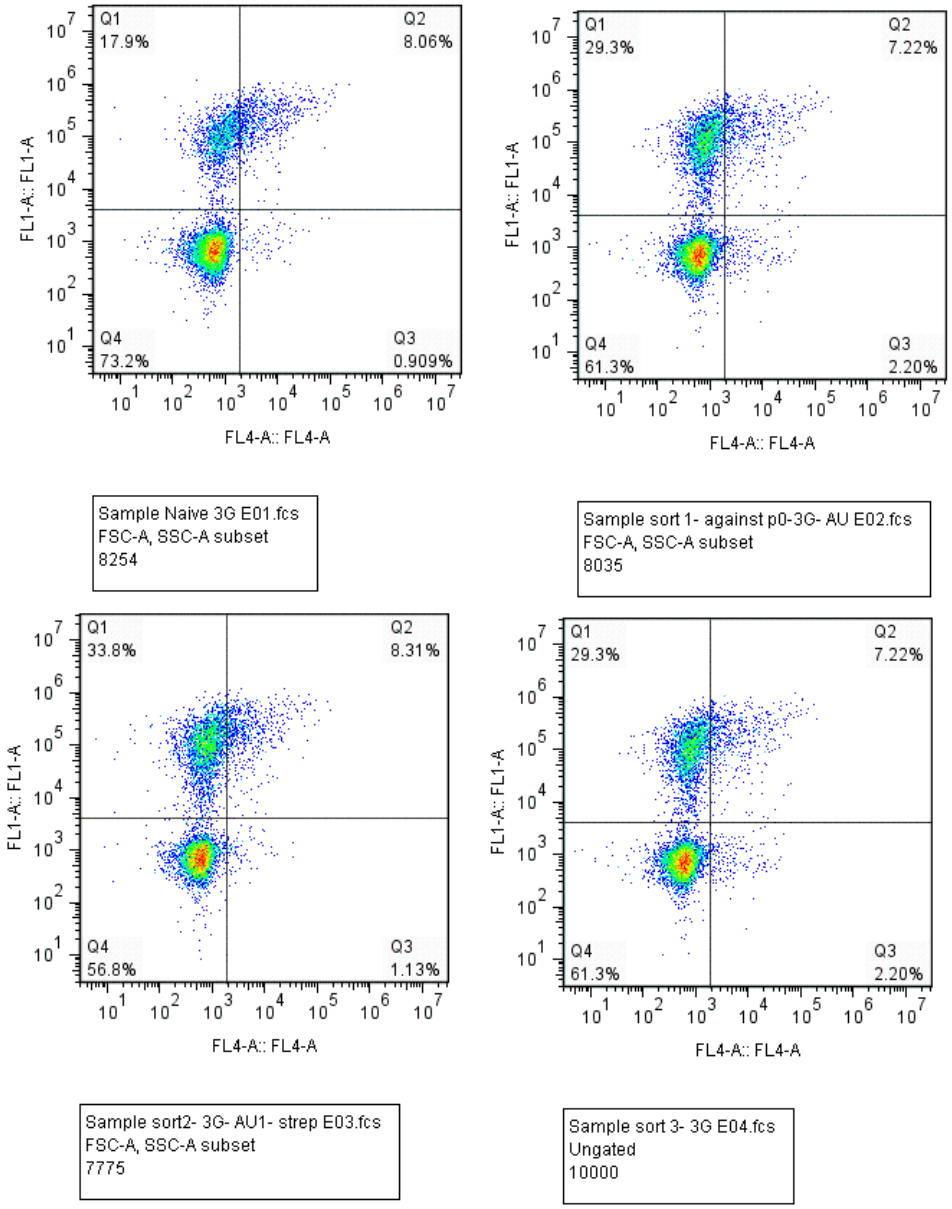


Figure A4-4. Sortase library and sorts against 3G-biotin peptide. FL1: AU1 expression, FL4: Streptavidin binding to the 3G-biotin bound to the surface.

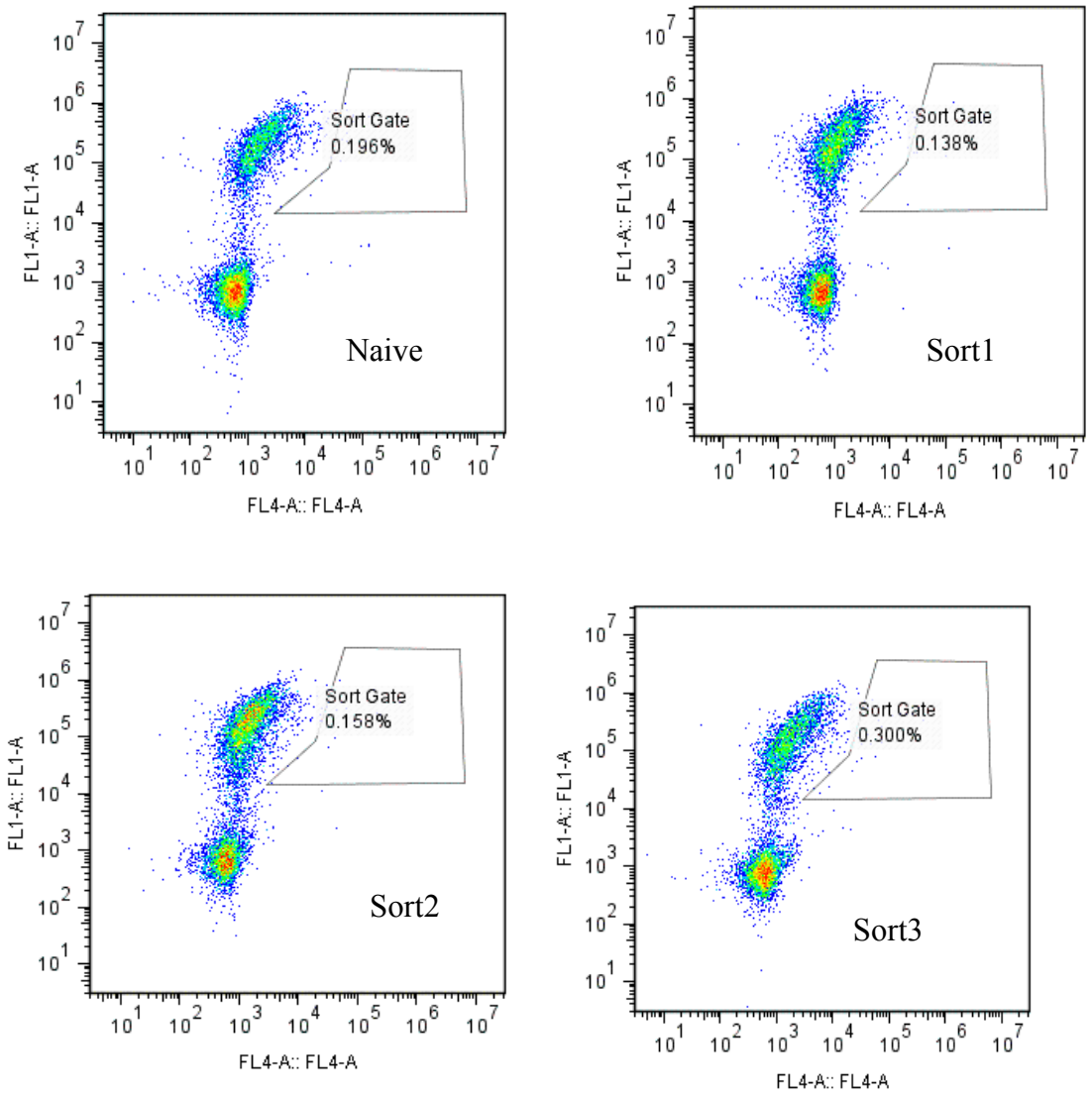


Figure A4-5. Control-P0 (K to G) peptide reaction tests. Naïve sortase library and sorts were tested against P0 control peptide, that has its lysine replaced by a glycine. FL1: AU1, FL4: streptavidin.

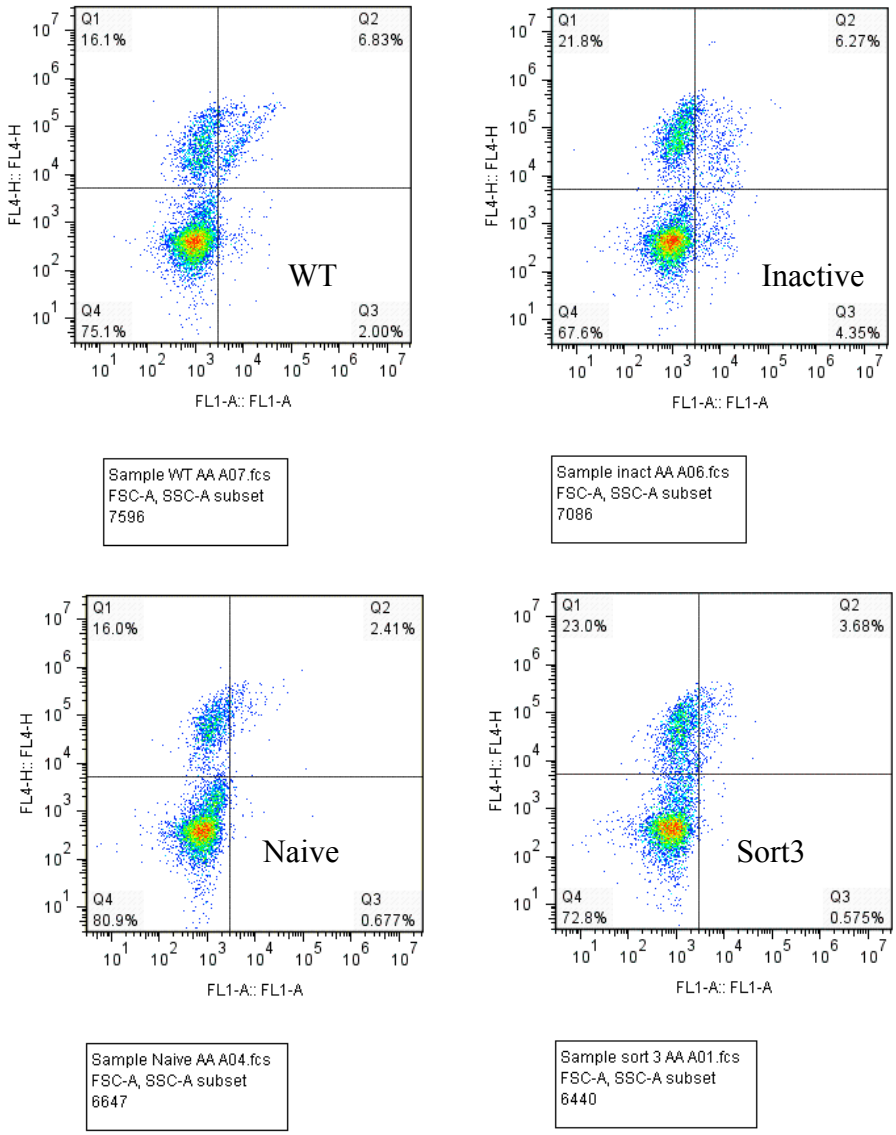
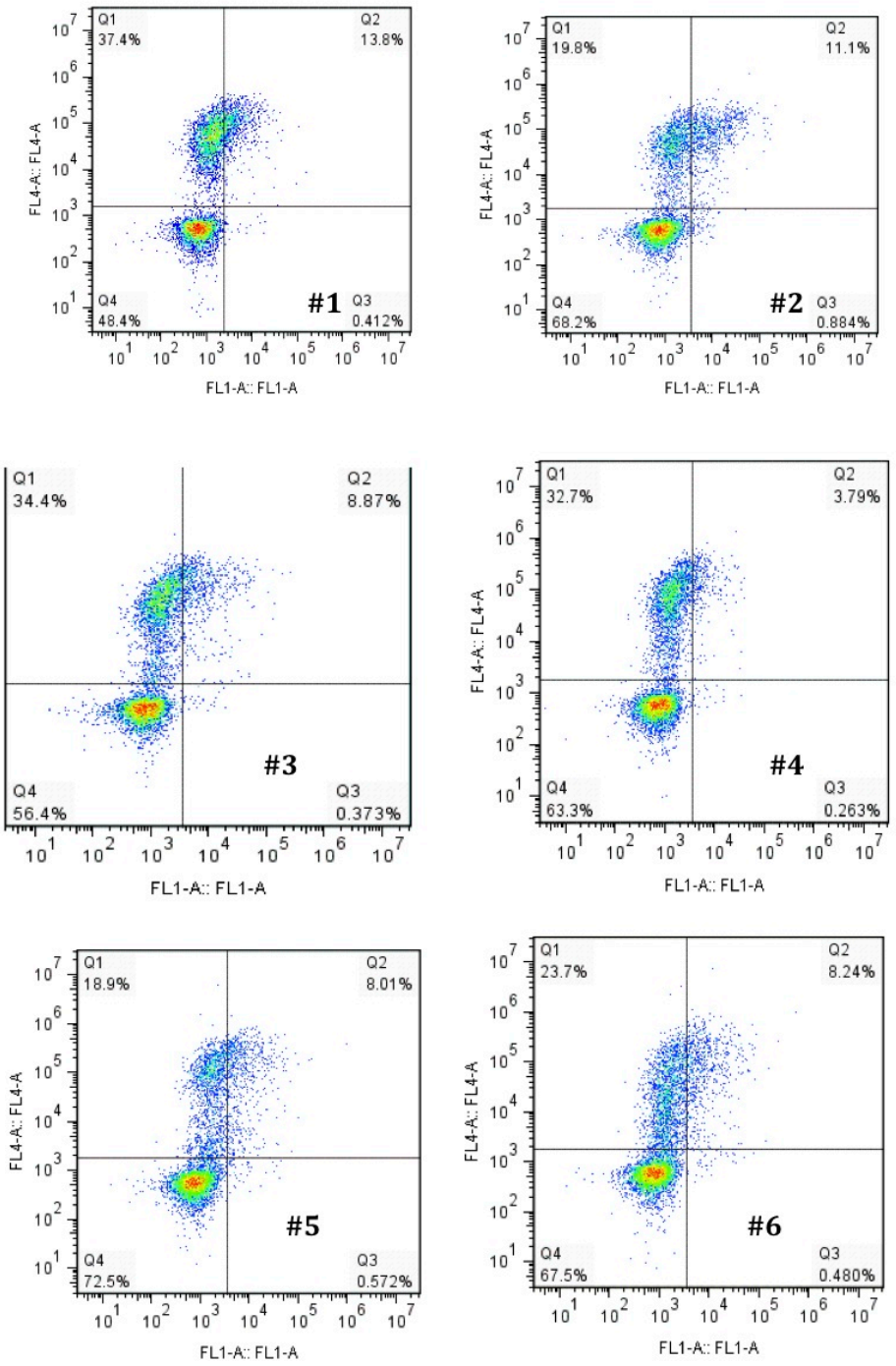


Figure A4-6. Tests against Ethylenediamine-biotin conjugated (EDA) peptide for substrate specificity against alkylamines. WT, Inactive, Naïve library and sort 3 (against P0 peptide) were tested against EDA peptide in a reaction mixture (TBS, $CaCl_2$ 10mM, pH= 7.5). FL4: AU1, FL1: streptavidin.

Figure A4-7- Flow cytometry data for individual colonies after three round of sort against P0 peptide. FL1 : Streptavidin, FL4: AU1.


#1
Q1 37.4%
Q2 13.8%
Q4 48.4%
Q3 0.412%
#2
Q1 19.8%
Q2 11.1%
Q4 68.2%
Q3 0.884%
#3
Q1 34.4%
Q2 8.87%
Q4 56.4%
Q3 0.373%
#4
Q1 32.7%
Q2 3.79%
Q4 63.3%
Q3 0.263%
#5
Q1 18.9%
Q2 8.01%
Q4 72.5%
Q3 0.572%
#6
Q1 23.7%
Q2 8.24%
Q4 67.5%
Q3 0.480%
FL1-A:: FL1-A
FL4-A:: FL4-A

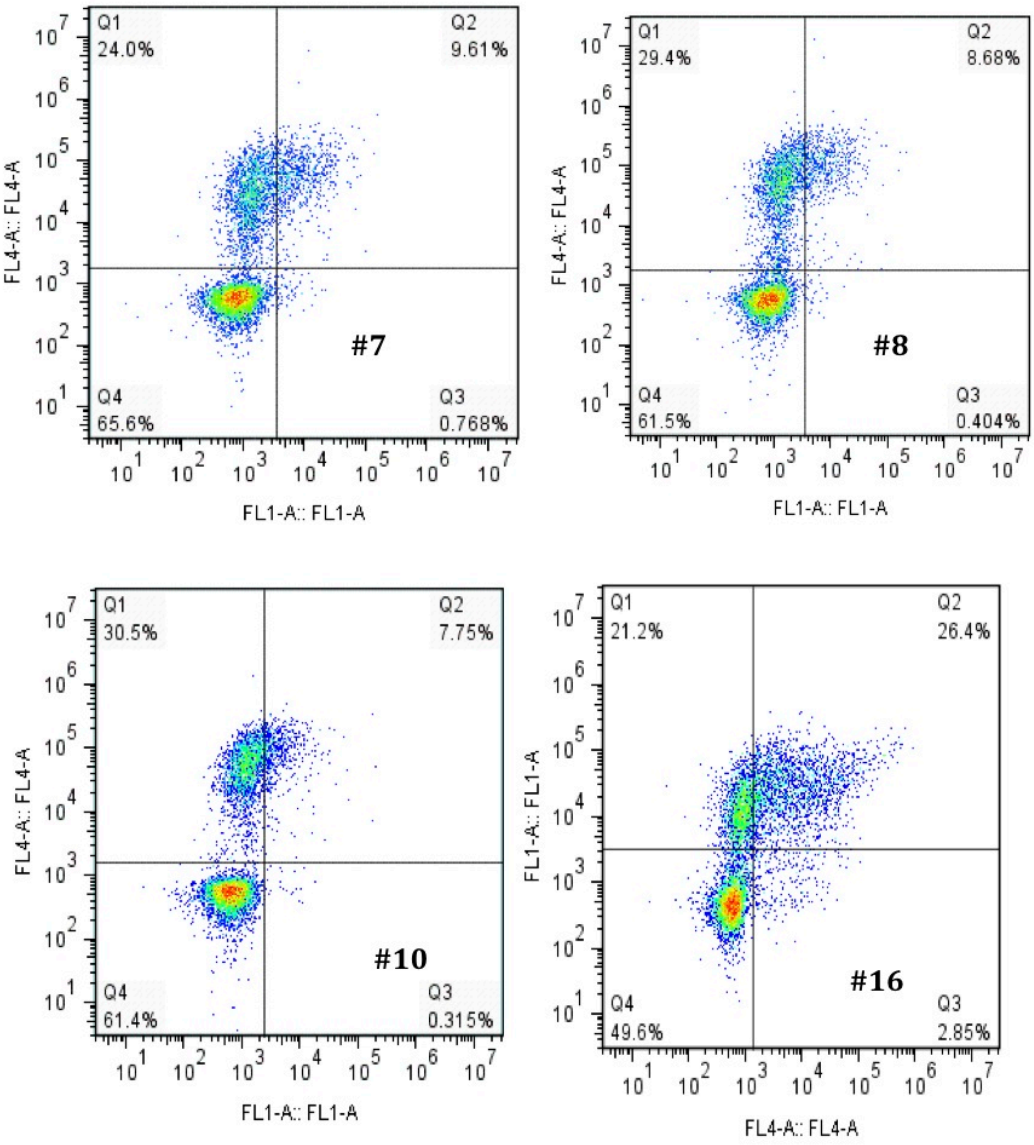


Figure A4-7- Continued

VITA

Maryam Raeeszadeh Sarmazdeh was born in 19 June 1983 in Mashhad, Iran. She has had a passion for science since childhood, which led her to continue her path in science and math at high school. She was accepted in the first step of Chemistry Olympiad at high school. She got her diploma in "math and physics" with distinction from NODET (National Organization for Development of Exceptional Talent) high school in Mashhad, Iran in 2000. Because of her interest in chemistry and math, she decided to continue Chemical Engineering at Sharif University of Technology in Tehran, Iran. Toward the end of her B.S. in Chemical Engineering, she became interested in bioengineering. Therefore, she continued her graduate studies in biotechnology at the same university and got a MS degree. After that, she moved to Knoxville,TN at 2009 to work on her PhD in Chemical and Biomolecular Engineering at UTK.