

**BIOMOLECULAR SCAFFOLDS FOR ENHANCED BIOMASS PROCESSING
AND TUMOR MARKER DETECTION**

by

Qing Sun

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemical Engineering

Fall 2015

© 2015 Qing Sun
All Rights Reserved

ProQuest Number: 10014775

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10014775

Published by ProQuest LLC (2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

**BIOMOLECULAR SCAFFOLDS FOR ENHANCED BIOMASS PROCESSING
AND TUMOR MARKER DETECTION**

by

Qing Sun

Approved: _____
Abraham M. Lenhoff, Ph.D.
Chair of the Department of Chemical and Biomolecular Engineering

Approved: _____
Babatunde A. Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: _____
Ann L. Ardis, Ph.D.
Interim Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Wilfred Chen, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Eleftherios T. Papoutsakis, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Kelvin H. Lee, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Brian J. Bahnson, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

The past six years of doctoral work have been life changing to say the least.

Firstly, I want to express my sincere gratitude to my PhD advisor, Dr. Wilfred Chen. Dr. Chen is tremendously passionate about science and engineering, and eagerly shares his enthusiasm with us. He spends lots of time meeting with us to discuss new ideas, new strategies, and experimental problems. He always encourages me to think innovatively and propose my own ideas. I will never forget all the discussions, debates or even arguments we had in his office, in the lab, or at the lunch table. I grew up in Dr. Chen's group and cannot thank him enough for all the things I learned from him.

Secondly, my appreciation goes to my research colleagues and labmates from the University of California, Riverside and the University of Delaware: Shen-Long Tsai, Lakshmi Cella, Miso Park, Bhawna Madan, Fang Liu, Christy Yeh, Divya Sivaraman, Payal Biswas, Heejae Kim, Daniel Balckstock, Qi Chen, Kay Siu, Rebecca Chen, Emily Hartzell, Lauren Dorsey, Vince Price, Maryam Raeeszadeh-Sarmazdeh, Andrew Gaynor, and Emily Berckman. It has been a pleasure working with everyone and building friendships. I especially appreciate Shen-Long Tsai, who introduced me to all the basic experimental and scientific details of molecular biology. I also want to specially thank Heejae Kim, Daniel Blackstock, and Qi Chen for all the research discussions we have had over lunch. These wonderful lunchtime brain-storms sessions made my Ph. D. life more fun. My appreciation also goes to Rebecca Chen and Kay Siu for our interesting discussions and thesis editing help.

I would like to acknowledge Dr. Terry Papoutsakis, Dr. Kelvin Lee, and Dr. Brian Bahnson for being a part of my dissertation committee. I greatly appreciate your expert advice and constructive feedback.

I would also like to express my gratitude for my friends from both California and Delaware. There were countless days and nights of fun with movie watching, travelling, and festival celebrating. You made my life colorful.

Lastly, I give my deepest appreciation to my entire family, who have been supporting me throughout the whole journey. I was unable to spend much time with my family in China during my Ph. D. However, they never complained and would only tell me to love myself and be happy. All my gratitude goes to my mother, father, sister, brother-in-law, brother, niece and nephew. I love you and thank you for the unconditional love you give me.

TABLE OF CONTENTS

LIST OF FIGURES	ix
ABSTRACT	xiv
Chapter	
1 INTRODUCTION	1
1.1 Co-localization of Proteins for Multiple Applications	2
1.1.1 Cellulose hydrolysis	2
1.1.2 Tumor marker detection	4
1.2 Artificial Protein Co-localization on Biomolecular Scaffolds	6
1.2.1 DNA scaffolds	7
1.2.2 Protein scaffolds	8
1.3 DNA-protein or Protein-protein Attachment.....	11
1.3.1 DNA Binding Proteins	11
1.3.2 Covalent protein-DNA linkage.....	13
1.3.3 Post-translational modification.....	15
1.4 Biomolecular Purification	16
2 CREATION OF ARTIFICIAL CELLULOSOMES ON DNA SCAFFOLDS BY ZINC FINGER PROTIEN-GUIDED ASSEMBLY FOR EFFICIENT CELLULOSE HYDROLYSIS	22
2.1 Introduction	23
2.2 Materials and Methods	25
2.2.1 Construction of expression vectors	25
2.2.2 Expression and Purification of proteins	26
2.2.3 Electrophoretic Mobility Shift Assay (EMSA)	26
2.2.4 CBM and CelA functionality.....	27
2.2.5 Dissociation constant determination.....	27
2.2.6 Artificial cellulosome efficiency characterization.....	28
2.3 Results and Discussion	28
2.3.1 ELP-Zinc Finger Protein function check.....	28
2.3.2 Characterization of CelA-E-Z and CBM-E-A.....	29

2.3.3	Artificial cellulosome assembly	30
2.4	Conclusions	31
3	HALOTAG MEDIATED ARTIFICIAL CELLULOSOME ASSEMBLY ON DNA TEMPLATE FOR EFFICIENT CELLULOSE HYDROLYSIS	38
3.1	Introduction	39
3.2	Materials and Methods	41
3.2.1	Vector construction and expression of proteins	41
3.2.2	Protein functionality	43
3.2.3	Native PAGE Mobility Shift	44
3.2.4	Rolling circle amplification	45
3.2.5	Artificial cellulosome assembly	45
3.3	Results	46
3.3.1	Protein function check	46
3.3.2	Cellulosome Assembly	47
3.4	Conclusions	49
4	POST-TRANSLATIONAL MODIFICATION OF BIONANOPARTICLES AS A MODULAR PLATFORM FOR BIOSENSOR ASSEMBLY	57
4.1	Introduction	58
4.2	Materials and Methods	60
4.2.1	Genetic manipulation.....	60
4.2.2	Protein expression	63
4.2.3	Individual components ligation onto ELP-E2 and function check.....	64
4.2.4	Sensor assembly	66
4.2.5	Characterization of protein complex	67
4.2.6	Thrombin detection	67
4.2.7	Tumor marker detection	67
4.3	Results and Discussion	69
4.3.1	Individual component ligation.....	69
4.3.2	Sensor assembled for thrombin detection with signal amplification.....	70
4.3.3	Sensor assembled for tumor marker detection with modularity..	72

4.4	Conclusions	73
5	CONCLUSIONS	87
	REFERENCES	92
Appendix		
A	LIST OF ABBREVIATIONS	104
B	REPRINT PERMISSIONS	106

LIST OF FIGURES

Figure 1.1	Three enzymes (endoglucanase, exoglucanase and β-glucosidase) needed for hydrolysis of cellulose to glucose.	19
Figure 1.2	Cellulosome with enhanced hydrolysis efficiency of cellulose and hemicellulose.	20
Figure 1.3	Different modes of artificial protein co-localization. (A) Free proteins, (B) Fusion proteins, (C) Protein clusters, (D) Scaffolds strategies.....	21
Figure 2.1	A schematic representation for the co-immobilization of two components on DNA template. Endoglucanase (red) was fused with ELP and Zinc finger protein (Zif268) while CBM (Blue) was fused with ELP and PE1A. Target template has both Zif268 and PE1A binding sites with 10 bp spacer between.	32
Figure 2.2	ELP-ZFP fusions functionality. (A) E-Z and E-A were purified by two cycles of thermal precipitation/solubilization. M: marker. The DNA-binding functionality of (B) E-Z and (C) E-A was confirmed using EMSA. In both cases, 1 nM biotinylated DNA was mixed with 500 nM protein. D: target DNA template only; D/P: target DNA template + protein; C: random DNA template + protein. The DNA–protein complexes are indicated by arrows.	33
Figure 2.3	Purification and characterization of CelA-E-Z and CBM-E-A. (A) Purification of CelA-E-Z and CBM-E-A by two cycles of thermal precipitation and solubilization. M: marker. (B) The cellulose hydrolysis activity of CelA-E-Z was measured by the reducing sugar production from PASC. (C) The cellulose-binding function of CBM-E-A. Purified proteins were mixed with avicel and the bound proteins were removed by centrifugation. The amount of CBM-E-A in the solution (B) before or (A) after binding was analyzed. The bound proteins were eluted (E) by boiling.	34
Figure 2.4	CelA-E-Z and CBM-E-A purification and ZFP binding. Determination of dissociation constants of CelA-E-Z (A) and CBM-E-A (B) by EMSA. All samples contained 1 nM DNA and 0, 10, 25, 50, 100, 300, 500, 1000 nM purified proteins. Controls contained 1 nM control DNA and 1000 nM protein. Dissociation constant calculations for CelA-E-Z and CBM-E-A (C).	35

Figure 2.5	Assembly of artificial cellulosomes on DNA template. (A) Production of reducing sugars from artificial cellulosomes. Green is DNA control; Blue is two proteins with control DNA (No binding sites); Purple is two proteins with DNA template having both binding sites. (B) EMSA analysis of all samples. Lane 1. CelA-KV8F-Zif268 with DNA; Lane 2. CelA-KV8F-Zif268 and CBM-KV8F-PE1A with DNA; Lane 3. two proteins with control template.	36
Figure 2.6	Cellulose hydrolysis by different cellulosome structures. Artificial cellulosomes containing more than one CelA (Cel-CBM-CelA) or CBM (CBM-CelA-CBM) domain were obtained using DNA templates carrying one additional binding site.	37
Figure 3.1	Schematic of artificial cellulosome on DNA template through HaloTag mediated site-specific DNA linker labeling. (A) Immobilization of cellulosome components on DNA template for cellulosome assembly; (B) Cellulose hydrolysis with cellulosome on discrete DNA template; (C) Cellulose hydrolysis with cellulosome on RCA DNA template.	51
Figure 3.2	CelA-E-H function check. SDS-PAGE (A) and Native PAGE (B) analysis of DNA labeling of CelA-E-H and activity check of CelA after labeling and purification (C). Protein (P) was purified from cell lysis (CL). Purified CelA-E-H (P) with a 3-fold excess of the CH-modified DNA (P/D) were purified by going through one cycle of precipitation and solubilization to remove the excess DNA (Pure). Activity of CelA was checked by cellulose hydrolysis (S) compared with control (C).	52
Figure 3.3	Assembly of artificial cellulosome with CelA and CBM. CBM-E-H, CelE-E-H and BglA-E-H function check. SDS-PAGE analysis (A, C, E) of DNA labeling of CBM-E-H, CelE-E- of CelA and activity check (B, D, F) of CBM, CelE and BglA after labeling and purification. In A, C and E, first lane are purified protein (P) and second lane are purified protein with DNA linker labeling (P/D). (B) The cellulose-binding function of CBM-E-H. Purified proteins (P) or DNA linker 2 modified proteins (P/D) were mixed with avicel and the bound proteins were removed by centrifugation. The amount of CBM-E-A in the solution (B) before or (A) after binding was analyzed. After wash (W), the bound proteins were eluted (E) by boiling. Activity of CelE and BglA was checked by substrate hydrolysis (S) compared with control (c).	53

Figure 3.4	Assembly of artificial cellulosome with CelA and CBM. (A) Production of reducing sugars from DNA alone (blue), free enzymes (red) and the assembly cellulosome (green). (B) EMSA analysis of protein binding. Lane 1: CelA-E-H-1 mixed with CBM-E-H-2; Lane 2: both proteins with DNA template with two binding sites. (C) Artificial cellulosome on RCA template.	54
Figure 3.5	Rolling circle amplification of DNA template. The discrete DNA template (T) was circulated (C) and rolling circle amplified to synthesize long repeating DNA sequence (R). Control was the rolling circle amplification conducted without circulated DNA template as shown (E). Nothing was detected in (E) demonstrating the synthesis of long repeating DNA template was specifically from circulated DNA template.	55
Figure 3.6	Assembly of artificial cellulosome with CelA, CBM, CelE and BglA. (A) Production of glucose from DNA alone (blue), free enzymes (red) and the assembly cellulosome (green). (B) EMSA analysis of protein binding. Lane 1, 2, 3, 4: CelA-E-H-1, CBM-E-H-2, CelE-E-H-3, BglA-E-H-4 mixed with blocker DNA; Lane 5: Mixture of Lane 1, 2, 3, and 4; Lane 12: CelA-E-H-1 and CBM-E-H-2 mixed with DNA template with four binding sites; Lane 124: CelA-E-H-1, CBM-E-H-2 and BglA-E-H-4 mixed with DNA template with four binding sites; Lane 1234: CelA-E-H-1, CBM-E-H-2, CelE-E-H-3, and BglA-E-H-4 mixed with DNA template with four binding sites. (C) Artificial cellulosome on RCA template.	56
Figure 4.1	Schematic of sortase A-mediated modification of ELP-E2 cage for sensor assembly. The LPETG tagged input and output modules were ligated onto N-terminal triglycine tag of E2 cage through sortase A mediated ligation. Signal amplification and modularity are the expected outcome advantages of this system.	75
Figure 4.2	Scheme for stepwise ligation and purification of ELP-E2-Protein No. 1 (Z-domain or HaloTag)-Protein No. 2 (Nluc or HaloTag).	76
Figure 4.3	Sortase A-mediated Z-Ct-LPETG ligation onto ELP-E2 cage. (A) SDS-PAGE analysis of reaction mixtures before (B) and after (A) ligation as well as purified products (P). (B) Transmission electron micrograph of IgG-bound E2 nanoparticles. Y-shaped IgGs were shown to bind onto intact E2 nanoparticles.....	77

Figure 4.4	E2 nanoprobe assembly with antibodies captured by the Z domain for sensing and Nluc for detection. (A) Schematic representations of IgG-Z-ELP-E2-Nluc and IgG-Z-ELP-Nluc. (B) Thrombin detection with Z-ELP-Nluc fusion proteins and IgG-Z-ELP-E2-Nluc nanoprobe; control samples contained no thrombin. (C) Signal amplification using controlled Nluc assembly by conjugating varying numbers of Nluc (5_22) onto E2 nanoparticles. (D) Detection of thrombin using E2 nanoprobe with varying Nluc decoration efficiencies.....	78
Figure 4.5	Tumor marker detection. (A) Detection of MUC1 on HeLa cells using E2 nanoprobe. (B) Detection of MUC1 with Nluc as the detection output.(C) Phase contrast and fluorescence images of HeLa cells detection with Alexa 488 as the detection output.	79
Figure 4.6	Tumor marker detection using DNA aptamer. (A) Conjugation of MUC1 aptamers onto HaloTag-ELP-E2 nanoparticles. (B) Phase contrast and fluorescence images of HeLa cells detection with using MUC1 aptamer-modified E2 nanoprobe.	80
Figure 4.7	Detailed diagram of the three-step ligation. (A) Step 1: Elastin-like-polypeptide ligation on E2 nanoparticle; (B) Step 2: Protein No. 1 (Z-Ct or HaloTag) ligation on ELP-E2 platform; (C) Step 3: Protein No. 2 (Nluc or HaloTag) ligation on ELP-E2-Protein No. 1.	81
Figure 4.8	Transmission electron micrograph of (A) unmodified E2 nanoparticles and (B) IgG- bound E2 nanoparticles.	82
Figure 4.9	Formation of IgG-binding E2 nanoprobe. (A) Expression and purification of Z-Ct-LPETG. Protein was partially purified by heating whole cell lysates at 70oC for 10 min. Most cellular proteins denatured and precipitated and was removed by centrifugation. (B) Binding of rabbit IgG to Z-E2-ELP. Rabbit IgG was mixed with Z-E2-ELP and the bound complex was recovered by one cycle of thermal precipitation and resolubilization.	83
Figure 4.10	E2 nanoprobe assembly using the Z domain for antibody capture and Nluc for detection. (A) Nluc purification and ligation onto ELP-E2 nanoparticles. (B) Detection of Nluc activity for the purified ELP-E2-Nluc nanoparticles.	84

Figure 4.11 Two-step assembly of Z-ELP-E2-Nluc nanoprobe and their use for thrombin sensing. (A) Two-step SrtA-mediated ligation of Z-Ct and Nluc onto ELP-E2 nanoparticles. (B) Dynamic light scattering of ELP-E2 and IgG-Z-ELP-E2-Nluc nanoparticles. (C) Production and purification of Z-ELP-Nluc probes. (D) Thrombin detection using IgG-Z-ELP-E2-Nluc nanoprobe. 85

Figure 4.12 Aptamer and Alexa 488 biosensor for tumor marker detection. (A) Production and ligation of Halo tag onto E2 nanoparticles. (B) Conjugation of CH-Alexa 488 to either (1) Halo-ELP-E2 or (2) Aptamer-Halo-ELP-E2. Successful conjugation was verified by detecting brightly fluorescent bands under UV light. (C) Ligation of Z-Ct and Halo tag onto ELP-E2 nanoparticles. 86

ABSTRACT

Proteins, as key components of life, are not standalone pieces. They frequently work cohesively either to form complex multi-step biochemical reactions in metabolic process or signaling pathway to receive specific inputs/signals and translate them into specified output responses. With all these protein collaborating to sustain life, nature has evolved protein co-localization on biomolecular scaffolds to enhance pathway efficiency. Inspired by the remarkable and extensive advantages provided by these scaffolds, we organized proteins on synthetic nucleotide or protein scaffolds to create protein co-localization systems for cellulose hydrolysis and cancer marker detection.

In objective one, we focused on organizing cellulases on DNA scaffolds to construct artificial cellulosomes based on zinc finger protein (ZFP)-guided assembly. Cellulosomes are naturally occurring multi-enzyme complexes with key components of cellulose binding module (CBM) and cellulases. Although artificial cellulosomes on protein scaffolds with up to 6 enzymes have been constructed with enhanced behavior, extending artificial cellulosome to more complex structures on protein scaffolds is still challenging as multi-domain large proteins tend to aggregate due to incorrect folding. Focusing on solving this problem, DNA templates were picked as scaffolds with the advantages of flexibility, easy synthesis and readily available complex structures. Zinc finger proteins as the DNA binding proteins were utilized to enable chemical modification free protein organization for artificial cellulosome assembly.

Our second objective was to improve the performance of DNA scaffold-based artificial cellulosomes by having a more stable immobilization. To achieve that, a

mutant dehalogenase enzyme (HaloTag) was used to replace the ZFP, such that fusion proteins can be covalently attached onto DNA linker, and DNA hybridization was used to immobilize cellulosome components onto a DNA scaffold. This alteration resulted in a system with increased enzyme efficiency compared with zinc finger protein-based assembly. A complex artificial cellulosome on a rolling circle amplification DNA template was achieved with 5-fold enhancement of cellulose hydrolysis efficiency.

Our final objective was to utilize three dimensional protein nanoparticles as platform for biosensor assembly for antigen and cancer cell detection. Protein-based nanoparticles have emerged as an excellent platform for biosensor assembly; however, current strategies of decorating bionanoparticles with different sensing and detection moieties often suffer from unfavorable spacing and orientation as well as bionanoparticle aggregation. To solve these problems, we use a highly modular post-translational modification approach, which enables the simultaneous modification of the *Bacillus stearothermophilus* E2 nanoparticles with different functional moieties for antibody, enzyme, DNA and dye decoration. The resulting platform offers easy purification, signal amplification and a high degree of targeting and sensing modularity. These advantages are demonstrated by the detection of both immobilized antigens and cancer cells.

Chapter 1

INTRODUCTION

Proteins are key components of life, responsible for nearly every cellular function, including cell shape and inner organization, product manufacture, waste cleanup, routine maintenance, receiving signals from outside the cell and mobilizing intracellular responses.¹

As essential components of life, proteins are not standalone pieces and usually work cohesively to achieve required functionalities. The success of many cellular processes depends on the quick communication among these partners within the same enzyme network in order to elicit more complex responses. Co-localization of these proteins through evolution has been developed with several natural systems.² The nanoscale organization has been shown to incorporate the right components into the system, to increase the local concentrations of enzymes and their substrates, to improve intermediate channeling between consecutive enzymes, and to avoid competition with other metabolites.³ With advantages provided by protein co-localization, we are going to explore artificial protein co-localization and highlight our approach to (i) organize artificial cellulosomes on DNA templates through zinc finger protein guided assembly, (ii) assemble artificial cellulosomes on DNA templates through HaloTag mediated DNA conjugation, and (iii) co-localize multi-functional components on protein nanoparticles to create biosensors for tumor marker detection.

1.1 Co-localization of Proteins for Multiple Applications

While all proteins collaborate to achieve multiple functions, protein co-localization is a common method evolved by nature to achieve improved efficiencies. Examples are naturally occurring complex enzyme clusters such as carbamoyl phosphate synthase,⁴ polyketide synthase,⁵ and tryptophan synthase⁶, all of which are parts of key metabolite synthesis pathways. These kinds of enzyme clustering form substrate channeling within the enzyme complexes to prevent loss of low-abundance intermediates, increase the effective concentrations of reactants and protect unstable intermediates, thus enhancing catalytic performance of the system.^{3,7} In addition to catalytic functions, proteins with different functions (including signaling and effector proteins) can be organized to modulate signaling pathways.² One of the most notable examples is the mitogen-activated protein kinases (MAPKs), a three-kinase regulatory cascade that is central to a highly conserved signaling mode that controls fundamental cellular processes, such as proliferation, cellular survival and differentiation.^{8,9} A protein scaffold (either ste5 protein in yeast or the kinase suppressor Ras in mammals) brings multiple signaling and effector proteins of the MAPK cascade into close proximity and thereby facilitates efficient signal amplification. Consequently, this protein scaffold acts as a signaling hub, providing intricate spatial and temporal control over MAPK signaling.⁷ The main objective in this dissertation is to develop two new strategies for creating artificial and modular scaffolds for protein co-localization and to demonstrate their usefulness in cellulose hydrolysis and tumor marker detections.

1.1.1 Cellulose hydrolysis

An increasing global energy demand along with the shrinking supply of fossil fuel, the most prevalent energy source, have made the search for alternative fuels

critical. A viable alternative is biofuel, a fuel derived from biological carbon fixation. According to a report by International Energy Agency in 2011, biofuels have the potential to meet more than a quarter of the world demand for transportation fuels by 2050.¹⁰ As an added bonus, biofuel is sustainable and has a potentially lower net carbon footprint.¹¹

Compared with food crops currently used for biofuel production, cellulose from agriculture residues and wood residues are considered more cost effective and is the most abundant biomass resources in the world.¹¹ Cellulose is a polysaccharide consisting of a linear chain of thousands of glucose monomers and is the most common form of biomass in nature. In a typical biofuel production procedure, cellulosic biomass is broken down into glucose, which is utilized by microbes to produce a variety of biofuels. Three enzymes (endoglucanase, exoglucanase and β -glucosidase)¹² are needed for the complete hydrolysis of cellulose to glucose (Figure 1.1). The significant role of cellulose hydrolysis in biofuel production has led to an increasing need for developing highly efficient cellulase systems, which is also the primary obstacle of using cellulose for biofuel production.¹³

Cellulosomes are complexes of cellulolytic enzymes naturally produced by anaerobic cellulolytic microorganisms. This multi-enzyme complex is formed through interactions between the cohesin domains on the surface scaffoldins and the dockerin domains on the enzymatic subunits (Figure 1.2).¹⁴ The enzyme cascade brings multiple enzymes in close proximity to the substrate via a cellulose-binding module (CBM). This provides a structure that ensures a high local concentration and the correct ratio and order of the enzymes.¹⁵ Enhancement of cellulose hydrolysis was observed from the cellulosome structure compared with non-organized enzymes. This enhancement is

attributed to the synergistic enzymatic action gained through substrate channeling, allowing preferential transfer of intermediates due to close proximity between active sites and a localized high concentration of substrates.^{16,17}

Inspired by nature, Fierobe and coworkers constructed artificial cellulosomes on CBM containing scaffolds with three cohesins.¹⁸ Three-fold enhancement in activity compared to free enzymes was observed from this synthetic cellulosome.¹⁹ Although artificial cellulosome with up to six enzymes have been constructed while preserving the enhanced cellulose hydrolysis efficiency,^{18,20} the possibility of extending this approach to more complex natural cellulosome structures remains a challenge as larger multi-domain proteins tend to aggregate due to incorrect folding.²⁰ Also, the right order, distance and ratio of enzymes in the cellulosome for the best synergistic effect remains to be explored while is challenging to execute with protein scaffolds.²¹

To counter the current problems associated with protein scaffolds based cellulosome, we explored the possibility of using DNA templates to organize complex artificial cellulosomes targeting at higher efficiency while simultaneously providing enough modularity to study the effect of ratio, ordering and distance of enzymes on cellulose hydrolysis efficiency.

1.1.2 Tumor marker detection

Cancer is a collection of related diseases characterized by abnormal cell growth with the potential to invade or spread to other parts of the body. It is among the leading causes of death worldwide. In 2015, an estimated 1,658,370 new cases of cancer will be diagnosed in the United States and 589,430 people will die from the disease.²² Approximately 39.6% of men and women will be diagnosed with cancer at some point during their lifetimes (based on 2010-2012 data).²³

Tumor markers, substances produced by cancer or other cells when cancer is present, can be used to confirm the presence of cancer, cancer types and the stage of cancer treatment. Most tumor markers are made by normal cells as well as by cancer cells; however, they are produced at much higher levels in cancerous conditions.²⁴ The level of tumor marker overexpression can also reflect the effectiveness of certain treatment toward cancer. To better distinguish healthy cells from cancer cells and monitor the progress during treatment, sensitive detection of tumor markers is necessary.²⁵ Meanwhile, more than 20 different tumor markers have been characterized and are in clinical use. Some are associated with only one type of cancer, whereas others are associated with two or more cancer types. Another consideration in cancer detection is that noncancerous conditions can cause the levels of certain tumor markers to increase even when no cancer exists. Considering all the possible situations associated with cancer, the detection of multiple tumor markers are necessary for cancer confirmation.^{25,26} In short, highly sensitive and modular biosensor platforms for tumor marker detection are needed for accurate cancer detection and treatment.

Biosensor functions by converting the recognition of analytes into actuated signals, making tumor marker detection possible.²⁷ To construct a functional biosensor, input modules like antibody or aptamer, output modules like fluorescence dyes or enzymes, and scaffolds, which organize input and output modules, are required.²⁸ Among all the different nanoscaffolds, protein nanoparticles are particularly attractive arising from their ability to self-assemble into simple nano-size uniform protein structures and allow precise control over material architectures.²⁹ Despite these desirable properties, current strategies of decorating bionanoparticles with different sensing and detection moieties often suffer from unfavorable spacing and orientation as

well as bionanoparticle aggregation.^{30,31} We attempt to solve these problems and expand the use of protein nanoparticles to organize highly sensitive, highly modular and easy to purify biosensor for immobilized antigen and cancer cell detection by going through a flexible post-translational sortase A-mediated modification.

1.2 Artificial Protein Co-localization on Biomolecular Scaffolds

With all the potential benefits provided by protein co-localization, artificial co-localization is desired. Several approaches have been investigated to co-localize proteins. While the protein fusion approach works well to recruit two enzymes near each other, this strategy becomes less viable for three or more enzymes (Figure 1.3a). Expression of larger fusion enzymes is often challenging since correct folding and retention of activity of all components are not guaranteed. While the effect of the latter can be ameliorated by increasing the linker length, low expression of functional proteins may lower yields despite superior synergy. Furthermore, even minute changes to the fusion enzyme require rechecking the functionality of each fusion partner. This process becomes more cumbersome as the number of fusion partners increases, which is why most fusion proteins reported in literature contain only 2–3 domains.² Enzyme clustering techniques, (Figure 1.3b) such as chemical conjugation,³² physical adsorption onto surfaces,³³ and encapsulation within a matrix,³⁴ from another aspect, has less control over the orientation, distance and orderings of all the components, which might affect substrate channeling formation. The organization of multi-protein complexes using a scaffold-based strategy is particularly attractive because of the modular nature of the design, and can be used as a flexible post-translational strategy for controlling the flow of information for complex metabolic and signaling pathways.² In this section, we will discuss the design principles of different biomolecule scaffolds and their usage for

protein co-localization. Nucleic acids and proteins, as the most abundant biomolecular materials in life, will also be explored.

1.2.1 DNA scaffolds

The emerging DNA nanotechnology has exploited nucleic acids as a promising template for immobilization of proteins. The ability to easily predict and manipulate the base-pairing property of nucleic acids along with the ease of synthesis have allowed researchers to create various DNA or RNA-based nanoscaffolds for the organization of multi-enzyme complexes.³⁵ Additionally, DNA can be replicated with polymerase especially with rolling circle amplification to generate long DNA templates with hundreds of repeating sequences, which can be used to organize individual molecules including proteins, target nucleic acids, small molecules. In one example, glucose oxidase (GOx) and horseradish peroxidase (HRP), chemically modified with DNA oligonucleotides, were spatially organized on single DNA strands generated by rolling circle amplification with amplified GOx and HRP binding sequences. The long DNA chains (up to 30 nt) enabled multi-enzyme cascades at relatively low enzyme concentrations, at which unassembled enzymes do not exhibit the full cascade activity. This demonstrated the feasibility of using DNA for ordered assembly of enzymes into hybrid composites with complex structures.³⁶

To increase the rigidity of one-dimensional nucleic acid based scaffolds, two-dimensional DNA scaffolds have also been explored. A set of pre-designed single-stranded nucleic strands with partially complementary sequences was used to form 'hexagon-like' structures with 10 bp overhanging DNA 'hinges' available for biomolecule tethering. GOx and HRP, chemically functionalized with DNA oligonucleotides, were attached onto two different hexagons by hybridization onto the

overhanging hinge areas. The overall activity of the enzyme cascades can be fine-tuned by controlling their relative position on two-hexagon and four-hexagon structures, with the two-hexagon case showing 1.2-fold higher over-all activity than that of the four-hexagon case. This is due to the expected longer diffusional path length of the intermediate using four-hexagons versus two-hexagons.³⁷ Two-dimensional DNA origami structure also looks at the effect of diffusion path lengths of the same two-enzyme system and the results were very similar to the hexagon systems.³⁸ Unfortunately, the current strategy of using single-stranded oligonucleotides when chemically conjugated to random surface residues of enzymes may lead to activity loss, decreasing the feasibility and universality of this strategy.³⁸

In this thesis, we are going to explore the organization of artificial cellulosome on DNA scaffolds through the help of zinc finger protein or HaloTag. Both methods help avoid direct chemical modification on enzymes to reduce the potential enzyme damage. Enhanced cellulose hydrolysis efficiency is also expected with this system.

1.2.2 Protein scaffolds

Other than nucleic acids, protein scaffolds have also been widely used for co-localizing proteins with different functions. Examples of protein scaffolds used for catalytic reaction enhancement exist both inside and outside of cells. Protein scaffolds organizing enzymes can guarantee the correct order and ratio of enzymes in cascade reactions, increasing communication by decreasing the diffusion path lengths.

Cellulosomes (Figure 1.2) from anaerobic cellulolytic microorganisms are one good example of protein scaffold organized multi-enzyme complexes outside cells.³⁹ For *in vivo*, evolved enzyme complexes capable of physically channeling intermediates, such as carbamoyl phosphate synthase and tryptophan synthase, are well studied.² With

orthogonal cohesin–dockerin pairs from different species cellulosomes, this system was utilized to enhance other sequential biochemical reactions beyond cellulose hydrolysis. For example, three dehydrogenases responsible for sequential conversion of methanol to carbon dioxide or three enzymes responsible for converting glyceraldehyde-3-phosphate to fructose-6-phosphate were organized with this cohesion-dockerin system. The correct assembling of enzymes resulted in more than 5-fold increase in product yield over that for unassembled enzymes.^{40,41}

In addition to in vitro biocatalysis, synthetic protein scaffolds were also used in vivo for increasing product titers. The mevalonate synthesis pathway enzymes were each fused with peptides specific to three different ligand binding proteins, PDZ, SH3, and GBD,⁴² and organized onto synthetic scaffolds composed of different numbers of binding proteins. The 77-fold enhancement in mevalonate production was achieved with binding protein ratios of GBD, PDZ, and SH3 domains of 1:2:2, which recruits a single copy of AtoB and two copies of both HMGS and HMGR.⁴² Synthetic scaffolds containing three orthogonal domains were displayed on the surface of yeast cells to assemble three different dockerin-tagged cellulases onto the scaffolds with three times ethanol production enhancement compared with free enzyme mixtures.^{12,19,43}

Other than the well-known catalytic functions, proteins also provide essential functions for sensing, intracellular signal transduction and cell-cell communication. These events require teamwork of multiple proteins organized together to enable efficient communication and since many of these protein domains are modular in function, signaling rewire can usually be achieved by exchanging or replacing the organized proteins. Mitogen-activated protein kinases (MAPKs) is one good example, with a three-kinase regulatory cascade that controls fundamental cellular processes.

They are organized on a natural protein scaffolds (Ste5p) for facilitated efficient signal propagation. By employing a simple artificial scaffold with heterologous SH3 and leucine zipper interaction pairs to dock histidine kinases (HKs) and response regulators (RRs), a 2.5-fold to 4-fold increase in activation was observed compared to the response in the absence of the scaffold. The addition of the recombinant SH3 domain also changed the specificity of the HK by direction its kinase activity toward the leucine zipper-fused response regulator that co-localized on the same artificial scaffold. On the basis of these works, we can envision using artificial scaffolds and interaction domains to connect pre-existing signaling modules in novel ways to rewire input-output combinations to generate novel response as part of synthetic signaling.

Although chimeric protein scaffolds are widely used for enhanced enzyme performance and signaling efficiency, another class of protein scaffolds that are also well studied and used are the protein nanoparticles. They offer the ability to self-assemble from simple protein building blocks and allow precise control over material architectures.

These protein nanoparticles can be used to organize multiple functional components for rewired functions as three dimensional protein scaffolds. One example is to construct biosensors, which convert the recognition of analytes into detectable signals; they have been widely used for monitoring biorecognition events.²⁹ In need of input modules to recognize analytes and output modules for detectable signals, protein nanoparticles were used to organize biosensor. An apoferritin nanoparticle with direct genetic fusion of protein G and 6xHis-tag on the particle surface was used to immobilize IgGs for input and quantum dots (QDs) as output signals. Significantly improved target interaction and detection sensitivity have been achieved by having

multiple copies of IgGs and quantum dots (QDs) on the same nanoparticles.²⁸ However, direct fusion does not work for every particle fusion partner because aggregation might happen rising from protein incorrect folding. Other ways of modifying protein nanoparticles including chemical modification or click chemistry all suffer from uncontrollable orientation or high cost problems.

In this thesis, we are going to highlight our recent progress in transforming protein nanoparticle to an ideal platform to organize protein with varieties of functions for biosensor assembly. Our platform offers easy purification, signal amplification, and modularity that could not be previously achieved.

1.3 DNA-protein or Protein-protein Attachment

In order to achieve the organization of proteins on biomolecular scaffolds (DNA and protein scaffolds), proper DNA-protein or protein-protein attachment are desired. This kind of attachment, either affinity based or covalent attachment, should serve the purpose of organizing proteins. Chemical modifications should be avoided because it might affect proteins functionality. Meanwhile, the ideal approach should guarantee proper protein functions by guaranteeing correct protein orientation and stable attachment between the scaffold and the proteins. In the following section, we will discuss our approach that will fulfill these requirements and our rationale behind its design.

1.3.1 DNA Binding Proteins

For organizing proteins on DNA templates, DNA binding proteins are good options, as they can be easily fused with target proteins to enable immobilization. There are two types of DNA binding, non-sequence specific DNA binding and sequence

specific DNA binding, non-sequence specific DNA binding and sequence specific DNA binding. Non-sequence specific DNA binding is usually used for holding chromosome structures.⁴⁴ Compared with non-sequence specific DNA binding, sequence specific DNA binding event is fundamental to many biological processes including transcription regulation and nucleases which sequence specifically cleave DNA.⁴⁵ To fulfill our requirement of organizing proteins with positional order on DNA templates, we chose to use sequence-specific DNA binding proteins.

Zinc-finger proteins (ZFP) are among the most abundant proteins in eukaryotic genomes.⁴⁶ It was first recognized 1986 as a repeated zinc-binding motif, containing conserved cysteine and histidine ligands, with an embedded zinc ion that maintains the protein's structure. The classical Cys2His2 zinc fingers are the first member of the zinc binding motif. Numerous other motifs have since been identified and designated. From the engineering angle, it was discovered that the binding affinity of zinc finger protein increases approximately 1000-fold with each finger motif added. Each of these motifs recognizes a specific three base-pair (bp) sequence and consists of 30 amino acids, with the zinc ion holding the stable $\beta\beta\alpha$ structure.⁴⁷ Zinc fingers can be redesigned to bind onto more specific three base-pair DNA sequences with high specificity. By taking advantage of the modularity of each individual zinc finger motif, multiple zinc finger protein with nanomolar affinity have been designed to extend the target sequences from 9 to 18 bp while retaining exquisite specificity.^{48,49}

Successful use of zinc finger proteins for protein organization has already been achieved. The first example is the reconstitution of a functional green fluorescent protein (GFP) in the presence of a DNA scaffold by appending two different sequence-specific ZFPs to each fragment of a split GFP.⁵⁰ The split GFP will only reconstitute

into the functional whole only when each fragment comes into close proximity with the other, thereby demonstrating that zinc finger proteins can bind onto pre-designed DNA templates to assemble their fusion partners in a position specific manner.

Zinc finger proteins have also been successfully used to enhance synthetic metabolic pathways inside *E. coli*. Using plasmid DNA as the scaffold, a variety of strains capable of self-assembling enzymes of biosynthesis pathways were developed, achieving up to fivefold enhancements of target metabolite titers.⁵¹ The real strength of this system lies in the fact that DNA is very easy to produce, allowing researchers to rapidly test different scaffolds in order to determine the best ratio of enzymes to put on the scaffold. The three metabolic pathways for resveratrol, 1,2-propanediol and mevalonate production tested in this paper⁵¹ along with L-threonine production enhancement⁵² highlight the wide applicability of this method, which makes DNA scaffolds appropriate for many diverse applications.

In this dissertation, we exploited the use of zinc finger proteins for organization of artificial cellulosomes to enhance cellulose hydrolysis efficiency. This zinc finger protein-based strategy helped us avoid chemical modifications for protein-DNA attachment and enabled predictable binding of zinc finger proteins and their fusion partners onto DNA templates in a position specific manner.

1.3.2 Covalent protein-DNA linkage

Another strategy to bind proteins onto DNA scaffolds is to covalently attach single stranded DNA onto proteins, after which DNA hybridization will be used for protein immobilization onto DNA templates. Since DNA hybridization has high affinity even with short length,⁵³ this strategy might provide higher linkage stability. Existing technology uses native cysteine or lysine residues on protein to attach functional group

modified DNA. With this method of modification, the specific position of modification cannot be controlled, because there might be multiple Cys or Lys residues present. This results in a loss of control over orientation of immobilized fusions. Additionally, loss of enzyme activity is commonly observed after chemical modifications due to the harsh conditions used during modification.

Taking advantage of the covalent linkage between protein and DNA while avoiding chemical modifications of proteins, the use of HaloTag is proposed. HaloTag is a 34 kDa monomeric haloalkane dehalogenase designed to covalently bind to synthetic chloroalkane linker attaching to useful molecules, including DNA, fluorescent dyes or solid surfaces. The reaction starts with a nucleophilic attack involving Asp106 and the formation of an ester intermediates and a point mutation in the gene clustering in a His272Phe substitution impairs the original hydrolysis step, leading to the stable covalent bond between protein and the chloroalkane modified ligands. This covalent bond formation is highly specific, occurs rapidly under mild physiological conditions, and is irreversible because it is a covalent attachment. With these characteristics, chloroalkane modified DNA should be easily attached onto proteins fused with HaloTag.⁵⁴ Since one HaloTag only binds to one ligand modified DNA, this attachment should be very specific and thus shows good control over the orientation of immobilization. Another added advantage of HaloTag is that it can help the soluble expression of fusion partners both in *E. coli*, *in vitro* and mammalian expression systems.⁵⁵

With all the advantages provided by HaloTag, DNA attachment for stable protein-DNA attachment, as well as fluorescence dyes for imaging or material surface conjugation have all been investigated.^{54,56}

In this dissertation, the use of DNA conjugated HaloTag for organization of proteins for artificial cellulosome assembly is evaluated. Higher efficiency is expected from HaloTag-based system with this stable and covalent attachment.

1.3.3 Post-translational modification

Protein-based nanoparticles are attractive platforms as they offer the ability to self-assemble from simple protein building blocks and allow precise control over material architectures.²⁹ Despite these desirable properties, protein bionanoparticles do have several drawbacks serving as platform for functional complex assembly. Most notable is the use of chemical modifications for functionalization, which can often result in random orientation and a loss of function.^{30,57-60} Direct gene fusion has been reported but is limited to relative small peptides and proteins.^{31,61} Moreover, the ability for simultaneous modifications with multiple sensing and detection modules remains challenging. A universal approach to interface different sensor components in the correct spatial orientation will be invaluable to satisfy the need to create highly modular biosensing platforms.²⁷

Sortase is a membrane-bound transpeptidase found in most Gram-positive bacteria responsible for pilin assembly and surface protein attachment. *Staphylococcus aureus* Sortase A (SrtA) recognizes the C-terminal LPXTG motif (X representing any amino acid) of a surface protein. Upon binding, a cysteine within the active site of sortase attacks the carbonyl group of threonine on the LPXTG motif, cleaving between T and G to make an acyl-enzyme intermediate. Then, a membrane-bound pentaglycine nucleophilically attacks the acyl-enzyme intermediate, forming a new native amino acid bond between the LPXT motif and the membrane-bound penta-glycine of the cell-wall peptidoglycan.^{62,63} Other than the original membrane protein ligation function, sortase

A was expressed and used to ligate two non-membrane-attached proteins together with >90% yield as long as they are LPXTG and GGG motif tagged, providing the chance of using Sortase A to post-translationally ligate proteins together.

The N and C motifs needed for ligation can guarantee the spatial orientation of the proteins, while the small size of the tags should have minimal impacts on their functionalities. With the ligation proceeding at mild pH and temperature conditions,⁶³ a series of protein fusions, alkaline phosphatase (AP), luciferase (Luc), and glucose oxidase (GOD) with the antibody-binding ZZ domain, fluorescence protein GFP, or peptides were able to be created in an orientation controlled manner.⁶⁴ Other than the direct protein post-translational ligations, Sortase A have also been exploited for site-specific labeling of proteins onto solid supports, vesicles and even capsid proteins of living M13 phages.⁵⁷

To take advantages of protein nanoparticles and Sortase A enabled post-translational modifications, we have constructed biosensors on *Bacillus stearothermophilus* E2 nanoparticles with different functional moieties, including antibody, enzyme, DNA and dye decoration. The resulting platform offers the controlled orientation of moiety decoration as well as a high degree of targeting and sensing modularity.

1.4 Biomolecular Purification

Extensive efforts have been invested in the development of simple and efficient methods for protein or protein nanoparticle purification. Use of specific protein or peptide fusion tags with reversible binding towards corresponding complementary chromatographic resins as well as multiple affinity tags such as His-Tag, FLAG, maltose binding protein (MBP), and glutathione S-transferase (GST) have long been

developed.^{65,66} Size, shape, charges, solubility, and function of proteins or protein nanoparticles were also widely explored for size-exclusion chromatography, ion-exchange chromatography, filtration or precipitation.⁶⁷ Although these purification protocols are commercially available, they are often time consuming, inefficient, and non-economical.⁶⁵ Thus, alternative methods using elastin-like polypeptides (ELPs) have been developed and can provide tremendous upside in the purification of both protein and protein nanoparticles.

ELP is a protein polymer with repeating sequence of VPGXG, where X can be any amino acid except for proline. They exhibit an inverse temperature phase transition based on a critical transition temperature. ELPs are soluble in aqueous solution below inverse temperature, but when temperature of the solution increases above the transition temperature, inverse phase transition occurs, resulting in aggregation of the ELP and forming an insoluble, polymer-rich “coacervate” phase.⁶⁸ Aggregation formed at this stage can be isolated from the surroundings using centrifugation. Since this phase transition is reversible, the ELP pellet can then be resolubilized by the addition of a cold buffer.⁶⁹

There are several factors affecting ELP transition temperatures. The guest residue (X) in the VPGXG repeat and the number of the repeats in the polymers are the two genetic factors affecting transition temperatures of the ELPs. The addition of hydrophobic guest residues lowered the transition temperature while the incorporation of ionized the polar groups raised the temperature. Increasing the number of repeats of VPGXG in the polymer also increases the transition temperature. Additionally, the transition temperature can be lowered by elevating salt concentrations, which lead to

enhanced aggregation due to the formation of salt \. With all the design principles, ELPs with response to changes in pH, redox triggers and lights were also designed.⁶⁹

The use of ELPs as partners has been reported to relay the purification frame to the whole complex system. Target proteins including thioredoxin,⁷⁰ antibodies,⁷¹ GFP,⁶⁵ OPH,⁷² and plasmids⁷³ have all been purified by going through thermal cycles either by direct fusion or interaction pairs attached onto ELP.

Due to the attractive properties of ELP, we incorporated ELP into our system. For the artificial cellulosome assembly part, ELP was fused between cellulosome components and zinc finger proteins or HaloTag as purification tag as well as linkers. For easy purification of protein nanoparticles, ELP was sortase ligated on E2 surface to enable the reversible aggregation purification of protein nanoparticles.

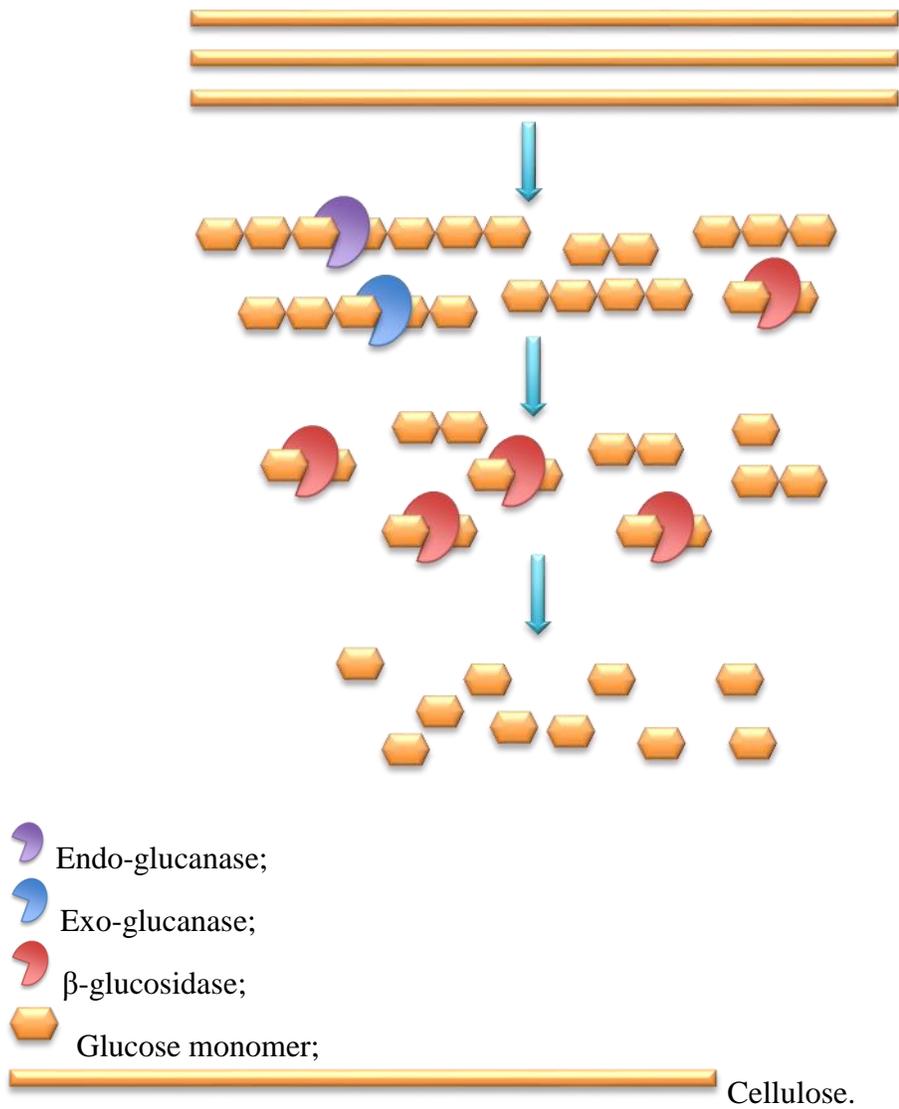


Figure 1.1 **Three enzymes (endoglucanase, exoglucanase and β -glucosidase) needed for hydrolysis of cellulose to glucose.**

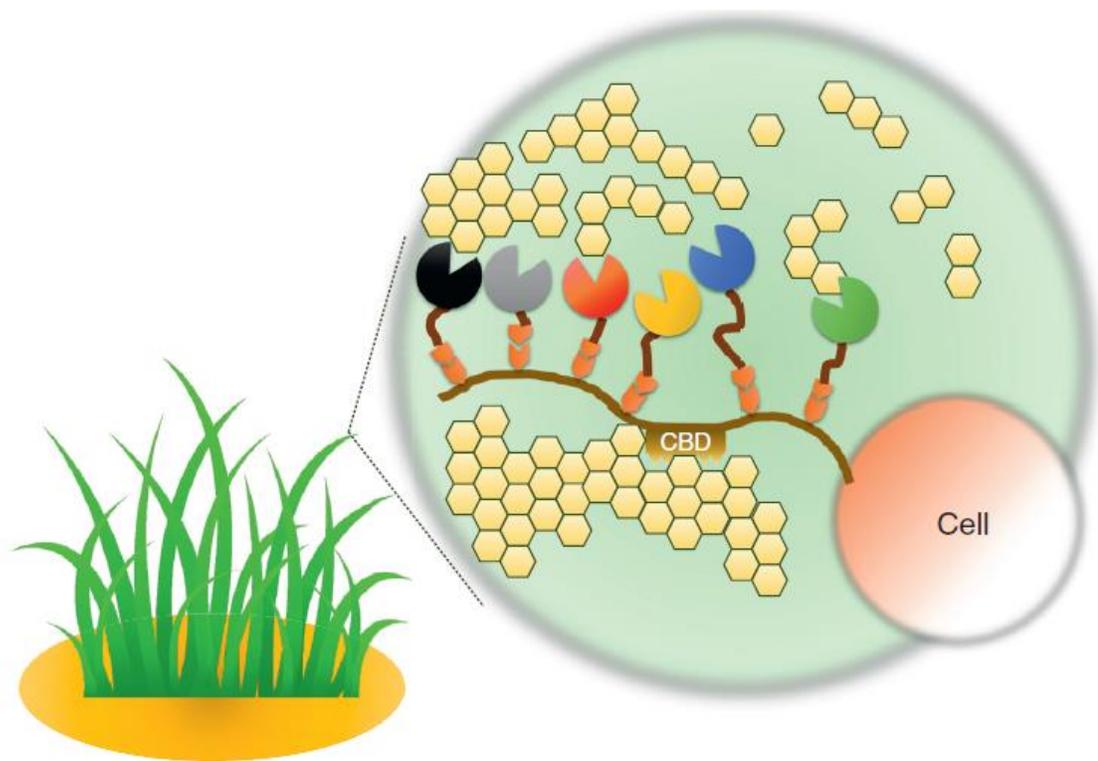


Figure 1.2 **Cellulosome with enhanced hydrolysis efficiency of cellulose and hemicellulose.**

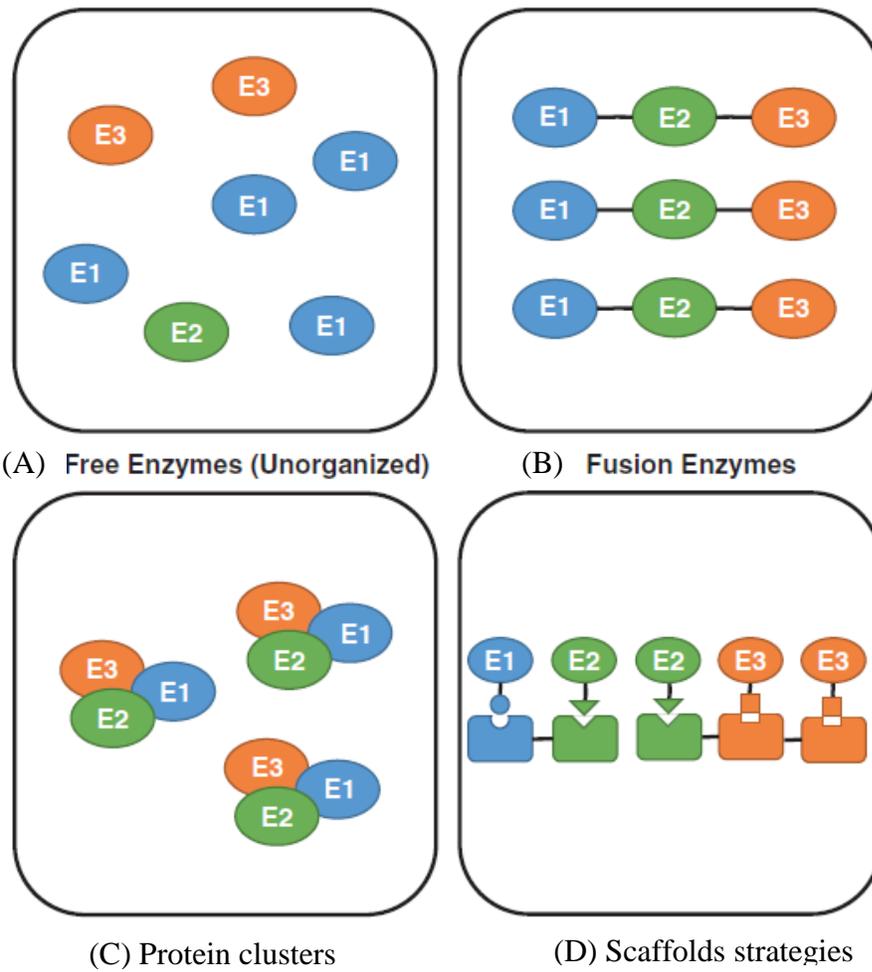


Figure 1.3 Different modes of artificial protein co-localization. (A) Free proteins, (B) Fusion proteins, (C) Protein clusters, (D) Scaffolds strategies

Chapter 2

CREATION OF ARTIFICIAL CELLULOSOMES ON DNA SCAFFOLDS BY ZINC FINGER PROTIEN-GUIDED ASSEMBLY FOR EFFICIENT CELLULOSE HYDROLYSIS

Abstract

For the first time, artificial cellulosome structures were created on DNA scaffolds based on zinc finger protein (ZFP)-guided assembly. These resulting two-component cellulosome structures exhibited enhancement in cellulose hydrolysis compared to the non-complexed mixture depending on the number of CBMs and cellulases assembled.

This Chapter is reproduced by permission of The Royal Society of Chemistry. Qing Sun, Bhawna Madan, Shen-Long Tsai, Matthew P. DeLisa and Wilfred Chen, Creation of artificial cellulosomes on DNA scaffolds by zinc finger protein-guided assembly for efficient cellulose hydrolysis, *Chemical Communication*, 50, 1423-1425, 2014.

2.1 Introduction

Efforts to improve the efficiency of cellulose hydrolysis are gaining momentum because of its significant role in biofuel production. Cellulosomes are naturally occurring multi-enzyme complexes with a substantially enhanced ability for cellulose hydrolysis. The major component of cellulosome is a surface-displayed scaffoldin composed of at least one cellulose binding module (CBM) and repeating cohesin domains that are decorated individually with a cellulase tagged with a dockerin domain.¹⁴ This self-assembled enzyme cascade brings multiple cellulases in close proximity to the substrate via CBM binding, and provides a structure that ensures a high local concentration of enzymes for efficiency cellulose hydrolysis.^{16,17} Synthetic minicellulosomes consisting of three to six cohesin/dockerin domains from different species have been constructed while preserving the enhanced cellulose hydrolysis efficiency.^{18,74} Although most natural scaffoldins contain 9 to 11 cohesin domains,¹⁵ the possibility of extending this synthetic approach to larger, more complex cellulosome structures remains elusive since these large multi-domain proteins tend to degrade due to incorrect folding. This was clearly demonstrated by the extensive degradation of even a smaller hexavalent scaffoldin in a recent report.⁷⁴ DNA is another attractive biological scaffold for multi-enzyme organization.³⁷ Semisynthetic conjugates of enzymes tagged with a short, single-stranded DNA oligo have been generated for the position-specific assembly of a two-enzyme complex consisting of NAD(P)H:FMN oxidoreductase (NFOR) and luciferase (Luc) based on hybridization onto a complementary DNA. This artificial bienzyme complex was shown to generate 3-fold more light compared with the free enzyme system.⁷⁵ Although this result provides the framework of multi-enzyme organization on a DNA scaffold, the high cost and loss of enzyme activity due to chemical conjugation has made it less than ideal.³⁸ However, DNA affords many

advantages since longer, rigid, and more complex one-, two- or three-dimensional DNA scaffolds can be easily synthesized for enzyme organization.^{37,51,76} Recently, artificial cellulosomes were constructed by conjugating multiple copies of the endoglucanase, Cel5A, onto a double-stranded DNA template. These DNA-Cel5A conjugates were prepared by using transglutaminase-mediated cross-linking between a lysine-based K tag on Cel5A and benzyloxycarbonyl-L-glutaminyglycine (Z-QG) motifs on the DNA template. Although the resulting Z-QG-modified DNA templates can be easily synthesized, this random conjugation method lacks the ability to organize cellulases in a position-specific manner,⁷⁷ which has been shown to be important in improving the overall performance of artificial cellulosomes.²¹ In contrast, proteins can be site-specifically localized onto a doublestranded DNA scaffold using DNA-binding proteins. Zinc-finger proteins (ZFPs) are DNA binding proteins that are composed of three subunits with each recognizing a specific three base-pair (bp) sequence.⁴⁷ By taking advantage of the modularity of the individual zinc finger protein, multiple zinc finger motifs with nanomolar affinity have been designed to sketch the target sequences from 9 to 45 bp while retaining the exquisite specificity.^{78,79} Successful reconstitution of a functional green fluorescent protein (GFP) has been demonstrated by appending two different sequence-specific ZFPs to each fragment of a split GFP, allowing binding onto two adjacent sites on the same DNA template.^{50,80} This result suggests that synthetic DNA scaffolds can be used to create artificial cellulosomes with the desired ordering and spacing based on ZFP-guided assembly. In this communication, a synthetic DNA scaffold was used for the site-specific organization of two ZFP-appended proteins (an endoglucanase CelA and a cellulose binding module CBM from *Clostridium*

thermocellum)⁸¹ into a bifunctional cellulosome structure for enhanced cellulose hydrolysis (Figure 2.1).

2.2 Materials and Methods

2.2.1 Construction of expression vectors

Escherichia coli strain NEB 5-alpha (*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) was used as the host for all genetic cloning. The ELP[KV8F-40] polypeptide was constructed by overlapping oligonucleotides (Integrated DNA Technologies, Coralville, IA) as previously described.⁶⁵ The fragment was inserted into pET24a to generate pET24a-ELP[KV8F]40 with several restriction sites at the C- terminus for inserting genes of interest. DNA coding for Zif268 was ordered from Genescript (Piscataway, NJ) and PCR amplified with forward primer 5'- CCG GAATTC GGATCC GGC GGC AGC AGCCC and reverse primer 5'- CCG CTC GAG CTG GTC TTC TTC AGA GAT AAG CTT -3'. The PCR product was inserted into ELP[KV8F]40 at *Bam*HI and *Xho*I restriction sites to form pE-Z. DNA coding for PE1A was ordered from Genescript (Piscataway, NJ). It was inserted into pET24a-ELP[KV8F]40 at *Sal*I and *Xho*I restriction sites to form pE-A. CelA-E-Z was constructed by first PCR amplifying the CelA fragment⁸¹ with forward primer: 5'-AGTCC CCCGGG CCGCGG CATATGGCTAGC GCGGCC and reverse primer: 5'- GGG AAT TCC ATA TGG GCG GCC GCC CGC GGG CTG CCG -3'. The amplified CelA fragment was digested and ligated into *Nde*I digested pE-Z to form pCelA-E-Z. Orientation of CelA was confirmed by DNA sequencing. CBM-E-A was constructed by PCR amplifying CBM with primers 5'-AGTCC CCCGGG

CCGCGG CATATGGCTAGC GCGGCC and 5'- GGG AAT TCC ATA TGG GCG GCC GCC CGC GGG CTG CCG -3' and inserting into *NdeI* digested pE-A.

2.2.2 Expression and Purification of proteins

All proteins were expressed in *E. coli* BLR [*F- ompT hsdSB (r-B m-B) gal dcm(DE3) Δ (srl-recA)306::Tn10(TetR)*; Novagen, Madison, WI]. Strains with plasmids were inoculated in 20 mL Terrific broth (TB) medium supplemented with 100 µg/mL kanamycin and incubated at 37°C until OD600 reached 1. The cultures were moved to 25°C for overnight expression. Cells were harvested by centrifugation, resuspended in DNA binding buffer (100 mM Tris base, 90 mM KCl, 1 mM MgCl₂, 100 µM ZnCl₂, 5 mM DTT, pH 7.5), and lysed by sonication. Purification of the ELP fusion protein was achieved by two cycles of inverse phase transitions. NaCl was added to the cell lysates to a final concentration of 1 M and the mixture was incubated at 37 °C for 10 min before centrifuging for 30 min at 15,000 rpm at the same temperature. The pellet was resuspended in ice-cold binding buffer and centrifuged for 30 min at 15,000 rpm at 4 °C to remove the insoluble cellular proteins. This precipitation and resolubilization process was repeated a second time and the purity of the protein was determined by 10% SDS-PAGE electrophoresis followed by Coomassie blue staining.

2.2.3 Electrophoretic Mobility Shift Assay (EMSA)

EMSA was performed with the Lightshift Chemiluminescent EMSA kit (Pierce Chemical Co., Rockford, IL) according to manufacturer's recommendations. All DNA templates were 5' biotinylated (Integrated DNA Technologies, Coralville, IA). Purified protein and biotinylated DNA were incubated at 37°C for 1 h. The complexes run on a

10% native acrylamide gel at 90 volts for 30 min. Membrane transferring and chemiluminescence signaling were done with protocols suggested by the manufacture.

2.2.4 CBM and CelA functionality

CBM fusion proteins were incubated with Avicel for 1 h at room temperature. Avicel was then removed by centrifugation and washed once with the binding buffer. The bound proteins were eluted by boiling for 10 min in the elution buffer (1% SDS, 0.1M NaOH).

CelA activity was assessed by mixing 80 μ M of CelA with 1% CMC. Samples were collected after one hour and immediately mixed with 0.5 mL of DNS reagent (10g/liter dinitrosalicylic acid, 10 g/L sodium hydroxide, 2 g/L phenol, 0.5 g/L sodium sulfite). After incubation at 95°C for 10 min, 1 mL of 40% Rochelle salts was added to fix the color before measuring the absorbance at 575 nm.

2.2.5 Dissociation constant determination

The DNA concentration was held constant at 1 nM while the protein concentration was varied from 0 nM to 1000 nM. The binding affinities were calculated as follows. Let θ represent the fraction of total DNA bound to protein, [P] represent unbound protein, [D] represent unbound DNA, [PD] represent the bound protein and DNA complex, [D]_{total} represent the total DNA:

$$K_D = \frac{[P][D]}{[PD]}$$

$$\theta = \frac{[PD]}{[D]_{total}} = \frac{[PD]}{[D] + [PD]} = \frac{\frac{[P][D]}{K_D}}{[D] + \frac{[P][D]}{K_D}} = \frac{[P]}{K_D + [P]}$$

The dissociation constant K_D is equal to the free protein concentration when the DNA bound fraction $\theta=50\%$. In this case, the DNA concentration is very low compared

with the protein concentration, thus $[P] \approx [P]_{\text{total}}$. K_D is then measured as concentration of total protein concentration $[P]$ when $\theta=0.5$ (Fig. S3).

2.2.6 Artificial cellulosome efficiency characterization

Artificial cellulosome assembly and phosphoric acid-swollen cellulose (PASC) for reducing sugars assay were prepared as previous described.¹⁹ Enzyme activity was assayed in the presence of a 0.1% (wt/vol) concentration of cellulose at 30°C in 20 mM Tris-HCl buffer (pH 6.0). Samples were collected periodically and immediately mixed with 0.5 mL of DNS reagents (10g/liter dinitrosalicylic acid, 10 g/L sodium hydroxide, 2 g/L phenol, 0.5 g/L sodium sulfite). After incubation at 95°C for 10 min, 1 mL of 40% Rochelle salts was added to fix the color before measuring the absorbance at 575 nm.

2.3 Results and Discussion

2.3.1 ELP-Zinc Finger Protein function check

Two well-characterized ZFPs, Zif268 and PE1A,^{82,83} each with high affinity for a unique 9 bp sequence, were chosen for spatial organization of the two cellulosomal components onto the DNA scaffold. Although ZFP fusions can be purified by adding a flanking His-tag, the residual imidazole must be removed by repeated dialysis to preserve the ZFP binding function because of the complexation with Zn^{2+} .⁸⁴ To bypass this problem, an elastin-like-polypeptide (ELP) tag was added for simple purification based on its thermally triggered, reversible aggregation property.⁶⁵ To ensure that this thermal purification strategy has no effect on the DNA-binding function, Zif268 (E-Z) and PE1A (E-A) were first fused to ELP[KV8F]40 and purified by two cycles of precipitation/solubilization (Figure 2.2A).^{65,85} The DNA-binding functionality was investigated using the electrophoretic mobility shift assay (EMSA), which takes

advantage of the differences in mobility between protein–DNA complexes and free DNA.⁸⁶ Neither the ELP tag nor the thermal cycles has any effect on the ZFPs as both purified fusions retained their individual DNA-binding functionality (Figure 2.2B and C). A slower mobility band corresponding to the DNA–protein complex was detected for each ZFP fusion upon mixing with a DNA template containing the target sequence, while no binding was observed when a control template containing a random sequence was added.

2.3.2 Characterization of CelA-E-Z and CBM-E-A

To construct the bifunctional cellulosome structure, CelA and CBM were inserted into the N-terminus of ELP in order to maintain both the function of ZFPs as well as the inserted domains. The resulting CelA-E-Z and CBM-E-A fusions were again purified by two cycles of thermal precipitation and resolubilization (Figure 2.3A). The functionality of CelA was confirmed by demonstrating the ability of CelA-E-Z to hydrolyze cellulose (Figure 2.3B). More importantly, the ability of Zif268 to bind DNA was not compromised as demonstrated by EMSA (Figure 2.4A and B). This result is significant as we demonstrated for the first time that ELP can serve both as a purification tag and a peptide linker separating two functional moieties. Similarly, CBM-E-A was expressed and purified using the ELP property. Purified proteins were fully functional retaining the ability to bind both avicel (Fig. 2.3C). The binding affinity for each protein was determined using a fixed amount of DNA and an increasing amount of proteins (Figure 2.4A and B). The calculated apparent binding constant for CelA-E-Z was 63 nM while that for CBM-E-A was 190 nM; both are consistent with values reported in the literature (Figure 2.4C).^{82,83}

2.3.3 Artificial cellulosome assembly

Next, the feasibility of ZFP-guided assembly of a two-component artificial cellulosome onto the same DNA template was demonstrated. A double-stranded DNA template (50 biotin-GGACCTTA TGCAATCACAGTG C G T G G G C GCCAGTACTGT A T A A A T A A CGGCTC GAGGTGT) with a 10 bp spacing between two binding sites was used (Figure 2.1). Binding of CelA-E-Z and/or CBM-E-A onto the DNA scaffold was examined using a 1 : 1 protein to DNA ratio (Figure 2.5A). While binding of CelA-E-Z was close to 100% under this condition, only around 50% of the DNA template was found to contain both proteins due to the lower affinity of PE1A. To assess the effect of the complex on cellulose hydrolysis, the assembled mixtures were added directly to a solution containing phosphoric acid-swollen cellulose (PASC). CelA, an endoglucanase, cuts cellulose randomly to produce smaller oligomers, and the hydrolysis efficiency was quantified by measuring the number of reducing ends from the smaller oligomers (reducing sugars) using the dinitrosalicylic method (Figure 2.5B). As expected, no reducing sugar production was observed in the presence of only the DNA template. However, when a mixture of CelA-E-Z and CMB-E-A was added without the DNA template, a significantly higher level of reducing sugar was produced. The level of reducing sugar was further enhanced by 1.7-fold when both proteins were mixed with the DNA template. The level of enhancement is consistent with other protein-based artificial cellulosomes containing only CelA and CBM,^{81,87} indicating that our DNA-based scaffold affords the same level of enzyme synergy and enhancement in cellulose hydrolysis.

One of the most powerful aspects of using our DNA-based assembly is the ability to increase the number of binding sites simply by employing a longer DNA template. To demonstrate this flexibility, two new DNA templates containing either an

additional Zif268 or PE1A binding site were synthesized. Cellulose hydrolysis experiments were repeated as described above. In all cases, the amounts of CelA-E-Z and CBM-E-A were kept constant. A 25% increase in reducing sugar production was observed using a template with two CelA binding sites instead of only one. Although the level of increase is smaller than expected, this is likely due to the lower percentage of scaffolds carrying all three required components. In contrast, increasing the number of CBM on the cellulosome structure has a smaller effect on cellulose hydrolysis (Figure 2.6). This may explain why the natural cellulosomal scaffolding CipA contains nine cohesin domains but only one CBM³⁹. It is also clear that a more efficient design will require the use of ZFPs with binding affinities below 10 nM as in the case of the cohesin–dockerin interaction⁸⁸. This can be achieved either by isolating 9-bp binders with higher affinities or by employing more complex ZFPs that recognize 15 or 18 bp.

2.4 Conclusions

In summary, we have constructed a series of artificial cellulosomes based on ZFP-guided assembly using DNA as a scaffold. The site-specific docking of CelA and CBM onto a single DNA template was achieved by exploiting the specificity of the two tethering ZFPs. The modular nature of the design allows easy alteration of the number, spacing, and ordering of enzymes assembled, leading to the virtually unlimited combination of artificial cellulosome structures optimized for a given target cellulosic substrate.

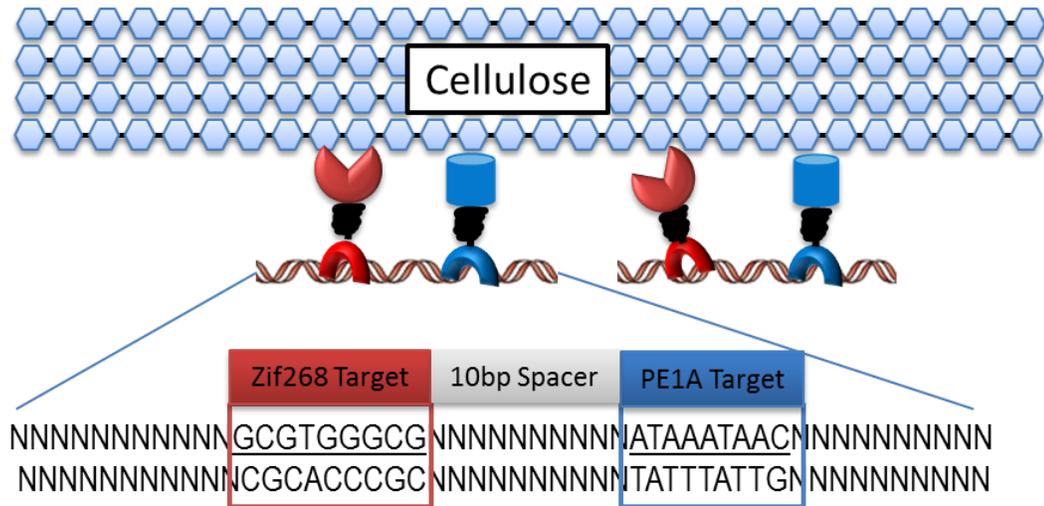


Figure 2.1 A schematic representation for the co-immobilization of two components on DNA template. Endoglucanase (red) was fused with ELP and Zinc finger protein (Zif268) while CBM (Blue) was fused with ELP and PE1A. Target template has both Zif268 and PE1A binding sites with 10 bp spacer between.

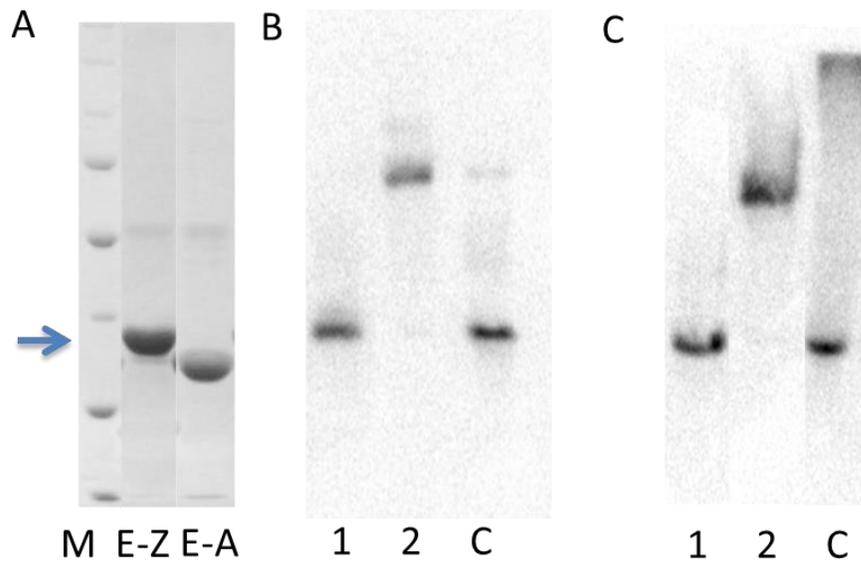


Figure 2.2 ELP-ZFP fusions functionality. (A) E-Z and E-A were purified by two cycles of thermal precipitation/solubilization. M: marker. The DNA-binding functionality of (B) E-Z and (C) E-A was confirmed using EMSA. In both cases, 1 nM biotinylated DNA was mixed with 500 nM protein. D: target DNA template only; D/P: target DNA template + protein; C: random DNA template + protein. The DNA-protein complexes are indicated by arrows.

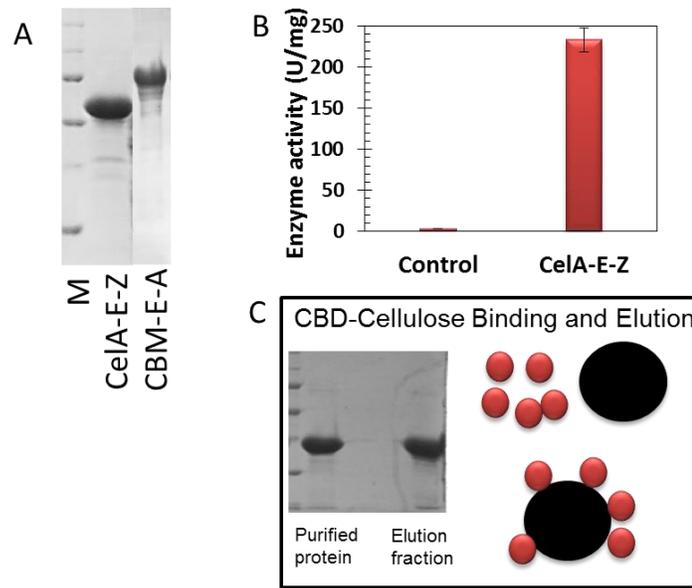


Figure 2.3 Purification and characterization of CelA-E-Z and CBM-E-A. (A) Purification of CelA-E-Z and CBM-E-A by two cycles of thermal precipitation and solubilization. M: marker. (B) The cellulose hydrolysis activity of CelA-E-Z was measured by the reducing sugar production from PASC. (C) The cellulose-binding function of CBM-E-A. Purified proteins were mixed with avicel and the bound proteins were removed by centrifugation. The amount of CBM-E-A in the solution (B) before or (A) after binding was analyzed. The bound proteins were eluted (E) by boiling.

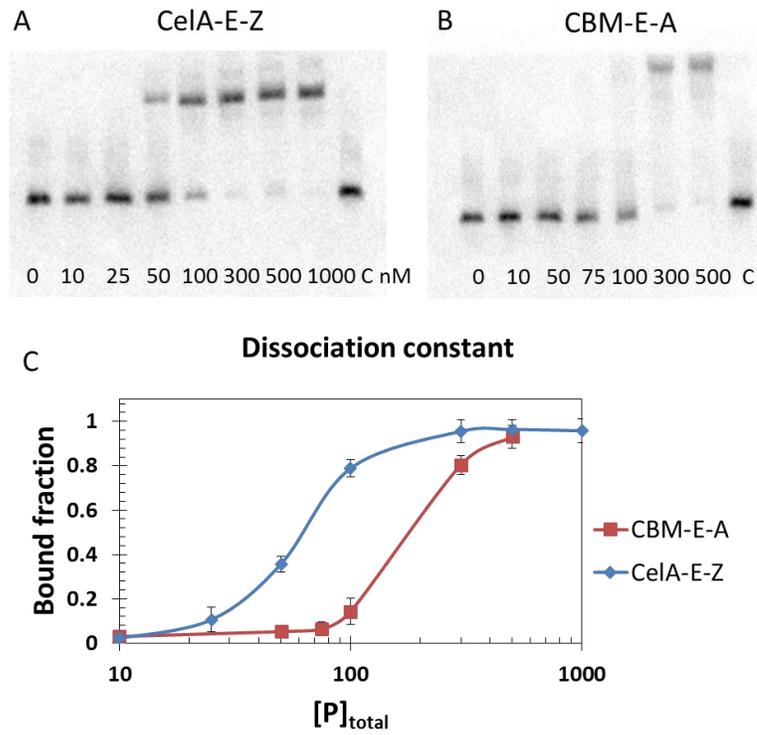


Figure 2.4 CelA-E-Z and CBM-E-A purification and ZFP binding. Determination of dissociation constants of CelA-E-Z (A) and CBM-E-A (B) by EMSA. All samples contained 1 nM DNA and 0, 10, 25, 50, 100, 300, 500, 1000 nM purified proteins. Controls contained 1 nM control DNA and 1000 nM protein. Dissociation constant calculations for CelA-E-Z and CBM-E-A (C).

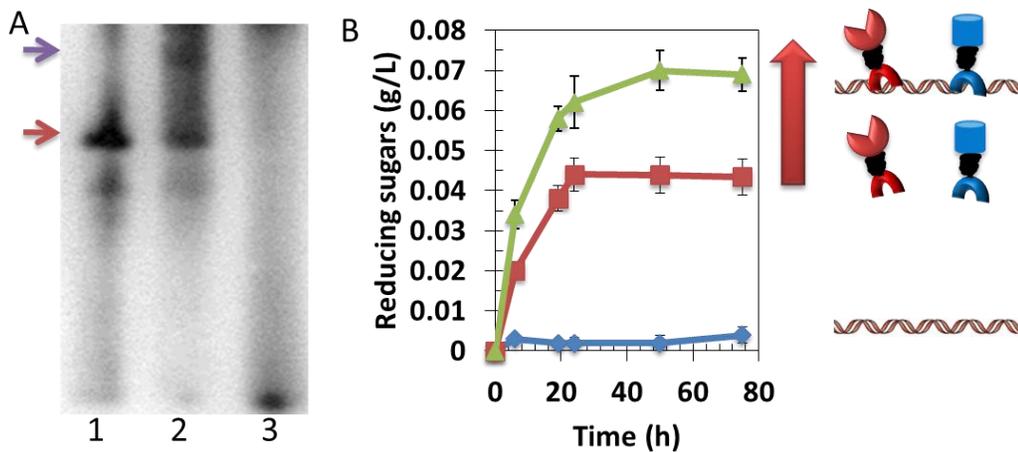


Figure 2.5 Assembly of artificial cellulosomes on DNA template. (A) Production of reducing sugars from artificial cellulosomes. Green is DNA control; Blue is two proteins with control DNA (No binding sites); Purple is two proteins with DNA template having both binding sites. (B) EMSA analysis of all samples. Lane 1. CelA-KV8F-Zif268 with DNA; Lane 2. CelA-KV8F-Zif268 and CBM-KV8F-PE1A with DNA; Lane 3. two proteins with control template.

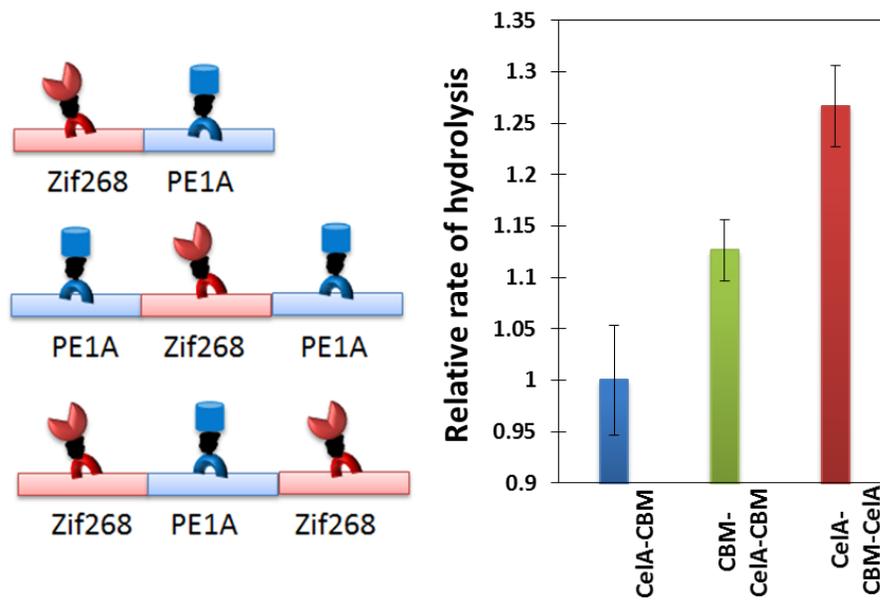


Figure 2.6 **Cellulose hydrolysis by different cellulosome structures.** Artificial cellulosomes containing more than one CelA (Cel-CBM-CelA) or CBM (CBM-CelA-CBM) domain were obtained using DNA templates carrying one additional binding site.

Chapter 3

HALOTAG MEDIATED ARTIFICIAL CELLULOSOME ASSEMBLY ON DNA TEMPLATE FOR EFFICIENT CELLULOSE HYDROLYSIS

Abstract

In this chapter, we report the generation of complex artificial cellulosomes assembly on DNA template through DNA hybridization, which was achieved through HaloTag site specific oligo conjugation. The high-affinity interaction associated with DNA hybridization improves the assembly efficiency of DNA template based artificial cellulosome compared with a zinc finger protein guided strategy, leading to 2.5-fold enhancement of cellulose hydrolysis efficiency. For the first time, an artificial cellulosome with four components (endoglucanase, carbohydrate binding module, exoglucanase and β -glucosidase) for complete cellulose hydrolysis to glucose was assembled. More importantly, the use of rolling circle amplification enables the simple amplification of binding sites for improved enzyme density on a single scaffold structure. A 5-fold enhancement in glucose release was observed with this complex cellulosome structure. The possibility to create more complex structure either to incorporate more varieties of cellulases to mimic enzyme diversity of natural cellulosome or create more complex two or three dimensional structures with this strategy are the advantages of this system.

3.1 Introduction

Cellulosic biomass found in many agriculture or wood residues are considered particularly well suited for biofuel production as they are fairly inexpensive and widely available.¹³ The primary obstacle of using cellulose for biofuel production is the absence of a low-cost technology for releasing glucose monomers from cellulosic biomass.¹³ This significant hurdle in biomass processing has attracted attention to the naturally occurring cellulosome structures found in many anaerobic bacteria, which are complexes of cellulases with significantly enhanced performance.¹⁴ Cellulosomes are organized through interaction of cohesin domains on the surface scaffold subunits and dockerin domains on the enzymatic subunits.¹⁴ The synergistic action of a cocktail of three different enzymes, endoglucanase, exoglucanase, and β -glucosidase, is used for complete hydrolysis of cellulose to glucose monomers.¹⁹ Presence of a carbohydrate binding module (CBM) on the surface scaffold enhances cellulose binding and brings the enzymes in close proximity to the substrate.¹⁶ The very high local enzyme and substrate concentrations afforded by this elaborate structure is the main driving force for the significantly faster hydrolysis rate.¹⁷ As a result, the number of enzymatic subunits that are incorporated into the cellulosome complex are known to effectively increase the overall enzyme density and hydrolytic efficiency.²⁰

In order to mimic natural cellulosome systems, artificial cellulosomes have been assembled on protein scaffolds with up to six enzymes immobilized on the same scaffolds.^{74,89} However, attempts to design protein-based scaffolds for artificial cellulosome with higher complexity^{15,90} have been challenging because larger multi-domain protein tend to aggregate due to incorrect folding.⁷⁴ Moreover, the effect of enzyme ordering, ratio and distance on cellulose hydrolysis efficiency was complicated to study because a new protein scaffold must be designed for every combination.^{91,92}

To bypass these limitations, DNA has been proposed as a potential scaffold for constructing artificial cellulosomes.⁹³ Several characteristics of DNA have made it an ideal scaffold for enzyme organization. First, single stranded DNA can hybridize into double stranded DNA in a sequence-specific manner, and programmable hybridization has been used to turn single stranded DNA into predictable, stable and complex one, two or three dimensional structures.³⁵ Second, DNA can be replicated using DNA polymerase, and rolling circle amplification (RCA) has been used to generate long DNA templates with repeating sequences, which can be used to organize up to several hundred copies of proteins, nucleic acids, and nanoparticles onto the same template.⁹⁴

Artificial cellulosomes based on DNA scaffolds have already been achieved using zinc finger protein (ZFP) guided assembly.⁵⁹ By fusing CBM and the endoglucanase CelA to two different zinc finger proteins, a two-component mini-cellulosome was created using a double stranded DNA template based on position-specific enzyme assembly. Although the resulting 1.7-fold enhancement in cellulose hydrolysis was in line with that achieved using protein scaffolds, the major drawback of this strategy is the low binding affinity of zinc finger proteins toward DNA, leading to only 50% cellulosome assembly efficiency.⁵⁹ Another parallel effort to assemble artificial cellulosomes on a DNA template is through transglutaminase-mediated cross-linking between a lysin-based K tag on Cel5A and the benzyloxycarbonyl-L-glutaminyglycine (Z-QG) motifs on the DNA template. Although multiple copies of Cel5A were immobilized onto DNA template for cellulosome synthesis, this random conjugation method lacks the ability to organize cellulases in a position-specific manner, which has been shown to be important for the overall performance of artificial cellulosomes.⁹⁵ More importantly, none of these reports were able to assemble artificial

cellulosomes containing all three cellulases and CBM in order to achieve complete liberation of glucose from cellulose.

To assemble a complex four-component artificial cellulosome in a position-specific manner, we proposed to conjugate single stranded DNA onto each protein component to enable cellulosome assembly through DNA hybridization. Since direct chemical modification of DNA linker on proteins usually leads to functional loss,³⁸ the self-labeling HaloTag was used for conjugation. HaloTag is a 34kDa monomeric mutant haloalkane dehalogenase designed to covalently bind to synthetic chlorohexane (CH) linker attaching useful molecules, including DNA, fluorescent dyes, and solid supports.⁵⁴ The covalent attachment is highly specific and occurs rapidly under mild physiological conditions.⁵⁴ By genetically tethering HaloTag to the different cellulosome components, CH-modified DNA oligonucleotides can be site-specifically conjugated for cellulosome assembly. By combining sequence specific DNA hybridization with RCA, we demonstrated up to 5.6-fold enhancement in cellulose hydrolysis using more complex cellulosome structures generated using the longer DNA templates from RCA.

3.2 Materials and Methods

3.2.1 Vector construction and expression of proteins

Escherichia coli strain NEB 5-alpha (NEB #C2987I) was used as the host for cloning of expression vectors. ELP-HaloTag-His6 was constructed starting with PCR amplifying HaloTag with HaloTag-Forward, HaloTag-Reverse. PCR amplified HaloTag was inserted into pET24(a)-ELP[KV8F]40⁶⁵ with *Bam*HI and *Xho*I to form pET24(a)-ELP-HaloTag.

CBM was then PCR amplified with CBM-Forward and CBM-Reverse and inserted into pET24(a)-ELP-HaloTag at *NdeI* to form CBM-ELP-HaloTag. CelA, CelE, and BglA were then individually PCR amplified using CelA-Forward, CelA-Reverse, CelE-Forward, CelE-Reverse, BglA-Forward, BglA-Reverse and inserted into pET24(a)-CBM-ELP-HaloTag at *SacII* and *SpeI* restriction sites to form CelA-ELP-HaloTag, CelE-ELP-HaloTag and BglA-ELP-HaloTag, respectively.

All proteins were expressed in host *E. coli* BLR [*F-ompT hsdSB (r-B m-B) gal dcm(DE3) Δ (srl-recA)306::Tn10(TetR)*; Novagen, Madison,WI] in TB medium. The four plasmid-containing strains were inoculated in TB medium supplemented with 50 µg/mL kanamycin at 37 °C until OD reached 1. Then the culture was moved to 25 °C shaker for overnight leaky expression.

After protein expression, cells were harvested by centrifugation at 4200 g, resuspended in PBS and lysed by sonication. The cell debris was removed by centrifugation at 16,100 g for 10 min at 4 °C. ELP purification of all components were then conducted to purify the proteins from cell lysate.

Lists of primers used for gene manipulations are:

HaloTag-Forward	5'- CCG CAT GGA TCC GGC GGC AGC AGC CCG AGC ACC CCG CCG ACC CCG AGC CCG AGC ACC CCG CCG GGC GGC AGC GCA GAA ATC GGT ACT GGC TTT CCA T -3'
HaloTag-Reverse	5'- TTG GCC CAA GCT TGC CGG AAA TCT CGA GCG TCG -3'
CBM-Forward	5'- GGG AAT TCC ATA TGC CGC GGG ATC CGA CCA AGG GAG CAA CAC -3'
CBM-Reverse	5'- GGG AAT TCC ATA TGG CTG CCG CCC GGC GGG GTG CTC GGG CTC GGG GTC GGC GGG GTG CTC GGG CTG CTG CCG CCA CTA GTG TCG ACT ACT ACA CTG CCA CCG G -3'
CelE-Forward	5'- AAA TCC CCG CGG ATG CTT GTT GGG GCA GGA GAT TTG -3'
CelE-Reverse	5'- TTC TAG ACT AGT TGC AGC GAA GTC CAA TGC ATC C -3'

BglA-Forward	5'- AAA TCC CCG CGG TCA AAG ATA ACT TTC CCA AAA GAT TTC ATA TGG GGT -3'
BglA-Reverse	5'- TTC TAG ACT AGT AAA ACC GTT GTT TTT GAT TAC TTC TTT GTA CCA GTA G -3'
CelA-Forward	5'- AAA TCC CCG CGG GCA GGT GTG CCT TTT AAC ACA AAA TAC CC -3'
CelA-Reverse	5'- TTC TAG ACT AGT GTT TCC TGT TAT GTA CAA CAA AGT GAG CAG TCT C -3'

3.2.2 Protein functionality

ELP functionality was confirmed by purifying all the proteins from cell lysate after expression. The protein was collected by inverse phase transition with the addition of 1M Na₂SO₄, incubated at 37°C for 10 min, and centrifuged at 16,000g for 10 min at 37 °C. After discarding the supernatant, the protein pellets were resolubilized in 4 °C cold buffer. This thermal cycling was repeated for better purity. DNA linkers modified with a 5' amine group were ordered from Integrated DNA Technologies (Coralville, IA) Upon arrival, they were modified with the HaloTag Succinimidyl (O4) Chlorohexane (CH) ligand (Promega P6751) for HaloTag attachment. The CH ligand was mixed with the DNA linker at a molar ratio of 30:1 and incubated at room temperature for 4 hours. A 3,000 Da ultrafiltration column (Vivaspin 500, Sartorius Stedim Biotech) was used for dialysis to remove excess CH ligand. The purified DNA linkers were then mixed with the purified Halo fusions, using a 3x molar excess of DNA linker and incubated overnight at 4 °C. Then, utilizing the ELP tag, unconjugated DNA linker were removed by pelleting the protein-DNA linker and washing away any residual DNA linker. The protein only, protein plus excess DNA linker and purified protein-DNA linker samples were loaded onto a 10% SDS-PAGE and stained with coomassie blue for analysis of labeling efficiency.

CelA activity was assessed by mixing 80 μ M of CelA with 1% CMC. Samples were collected after one hour and immediately mixed with 0.5 mL of DNS reagents (10g/liter dinitrosalicylic acid, 10 g/L sodium hydroxide, 2 g/L phenol, 0.5 g/L sodium sulfite). After incubation at 95°C for 10 min, 1 mL of 40% Rochelle salts was added to fix the color before measuring the absorbance at 575 nm.

CelE activity was assessed by mixing 200 μ M of CelE with 0.3% PASC, prepared from Avicel PH101 (Sigma) as previously described.¹⁹ Samples were collected after one hour and reducing sugars checked like for CelA.

CBM fusion proteins were incubated with Avicel for 1 h at room temperature. Avicel was then removed by centrifugation and washed once with the binding buffer. The bound proteins were eluted by boiling for 10 min in the elution buffer (1% SDS, 0.1M NaOH).

To assay the functional secretion of β -glucosidase BglA, 100 μ L of a 10 mM of fluorescent substrate p-4-methylumbellifery- β -D-glucopyranoside was mixed with 5 μ M of purified and labeled BglA and incubated at 37°C for 10 min. The activity was confirmed by detecting the fluorescence under UV light.

3.2.3 Native PAGE Mobility Shift

Mobility shift experiments were carried out in a 4.5% non-denaturing acrylamide gel. The protein only, protein with DNA linker, and purified protein-DNA linkers were loaded and run for 45 min at 90 V. The gel was then stained with ethidium bromide for 5 min before imaging.

3.2.4 Rolling circle amplification

The single stranded DNA template with four binding sites (5'-GAGAGTCAGTCAGGAATT TTTAAAGGAGGGAGGGGAATTTTACAGCGAGCGTCTACATTTTACACCA GCCAGC CAACTTTTTTTTATGCATCGTCAGTTAG-3') at a concentration of 1×10^{-6} M was ligated following a quick circular ligation protocol with CircLigase ssDNA Ligase (epicentre). Then, 5×10^{-8} M of the ligated oligonucleotide was treated with 0.5×10^{-3} M dNTPs, and 0.4 U/ μ l of Phi29 DNA polymerase (NEB) in 50 μ l of Phi29 buffer for 1hr, at 30°C, the enzymes were then deactivated at 65°C for 10 min. Single stranded DNA template, circulated DNA template, RCA product and purified RCA product was run on 4.5% native PAGE gels to check on the RCA template synthesis.

3.2.5 Artificial cellulosome assembly

For artificial cellulosome assembly, equal concentrations of linker-labeled cellulosome components were mixed with DNA template in DNA hybridization buffer. The mixture was heated to 37 °C and incubated for 10 min. Then 5X detergent buffer (25mM DTT, 50% glycerol and 0.5% IGEPAL) was added into the mixture to prevent potential association of all the components. After sitting at 4 °C overnight, the mixture were taken out and mixed with 0.1% cellulose buffer. Samples were collected periodically to check reducing sugars or glucose concentration, which was determined with a glucose HK assay kit from Sigma.

3.3 Results

3.3.1 Protein function check

To enable complete hydrolysis of cellulose to glucose, four different components: CelA (endoglucanase from *Clostridium Thermocellum*), CBM (carbohydrate binding module from *Clostridium Thermocellum*), CelE (exoglucanase from *Clostridium Cellulolyticum*) and BglA (β -glucosidase from *Clostridium Thermocellum*),^{43,81} were employed to assemble the cellulosome structures. Each cellulosomal component was genetically fused at the N-terminus to HaloTag to enable the self-specific conjugation of CH-modified DNA oligonucleotides. An elastin like polypeptide (ELP) tag was incorporated between the cellulosome component and HaloTag to simplify purification by taking advantage of the inverse phase transition property associated with ELP.^{65,85}

To co-localize all the cellulosome components onto the same DNA template, a single stranded DNA was designed (target DNA scaffold) with four unique 20bp hybridization domains, I, II, III and IV, that are complementary to four different DNA linkers 1, 2, 3 and 4 (Figure 3.1A) on discrete DNA template (Figure 3.1B) or RCA template (Figure 3.1C) for cellulose hydrolysis.

To enable the self-conjugation of linker 1, 2, 3 and 4 onto fusion proteins, a CH ligand was incorporated to the 5' end as described previously.⁵⁶ Full length CelA-ELP-HaloTag (CelA-E-H) fusions was purified using two cycles of thermal precipitation and solubilization. To incorporate the CH-modified linker 1, 3x molar excess oligo was incubated with the purified proteins. Close to 100% conjugation was demonstrated by detecting a single, larger band corresponding to the CelA-E-H-linker 1 complex (Figure 3.2A). Formation of the CelA-E-H-linker 1 (CelA-E-H-1) complex was further

confirmed by detecting a slower mobility protein-DNA band with the electromobility shift assay (EMSA). After labeling, excess linker 1 was removed by another cycle of phase transition to avoid potential interference with cellulosome assembly (Figure 3.2A and B). The functionality of the CelA domain was investigated by measuring the release of reducing sugar from cellulose hydrolysis (Figure 3.2C). Conjugation with DNA has no impact on the functionality of CelA as the same activity was detected after purification and conjugation (data not shown). The same procedure was executed to conjugate DNA linker 2, 3 and 4 onto purified CBM-ELP-HaloTag (CBM-E-H), CelE-ELP-HaloTag (CelE-E-H) and BglA-ELP-HaloTag (BglA-E-H), respectively, as shown in Figure 3.3. The functionality of other components were confirmed either by cellulose binding (CBM) (Figure 3.3A, C and E) or by activity assays (CelE and BglA) (Figure 3.3B, D and F).

3.3.2 Cellulosome Assembly

After successful DNA conjugation, we next investigated the capability of cellulosome assembly. A simplified cellulosome containing only CelA and CBM was first constructed by mixing CelA-E-H-1 and CBM-E-H-2 with a same amount of the target DNA scaffold. As a control, CelA-E-H-1 and CBM-E-H-2 were first pre-blocked with their complementary oligo (I or II) before mixing together for cellulose hydrolysis. As depicted in Figure 3.4A, a 2.1-fold more reducing sugars was released using the cellulosome structure as compared with the control, indicating the functional assembly of cellulosome with this strategy. To confirm cellulosome assembly, the resulting mixtures were evaluated by EMSA. Compared with the control, a slower migration band was detected with the cellulosome mixture demonstrating the successful assembly of the two proteins onto the target DNA scaffold (Figure 3.4B). From the band intensity,

over 90% of the added proteins were assembled, which is significantly better than the 50% assembly efficiency with the zinc finger protein guided approach. This improved assembly efficiency may explain the higher enhancement in cellulose hydrolysis (2.1-fold vs 1.5-fold) because of the higher degree of protein promixity.⁵⁹

A key benefit of the strategy is the ease of increasing the cellulosome complexity by simply employing a longer DNA scaffold. To achieve this, rolling circle amplification (RCA) was used to prepare templates with up to 200 repeats of the four binding sites.³⁶ Successful template amplification by RCA was demonstrated using a native gel and the presence of large DNA band was detected only with the RCA sample (Figure 3.5). We proceeded to investigate whether a complex cellulosome structure created using the RCA template could result in further enhancement in cellulose hydrolysis. CelA-E-H-1 and CBM-E-H-2 were organized onto the RCA template as described above, and the amount of reducing sugars released was compared. Although the amount of proteins used in all cases were the same, the RCA-derived cellulosome structure produced 3.8-fold more reducing sugars than the control as compared to only 2.1-fold using the single binding site template. This level of enhancement is in line with those reported for multi-Cel5A conjugated onto a single DNA template through transglutaminase-mediated cross-linking⁹⁵ or artificial cellulosome structures created by clustering cellulases onto quantum dots,⁹⁶ all supporting a direct correlation between local enzyme density and the efficiency of cellulose hydrolysis (Figure 3.4C).

To further increase enzyme density for complete cellulose hydrolysis to glucose, all four components, CelA-E-H-1, CBM-E-H-2, CelE-E-H-3, and BglA-E-H-4, were co-localized onto the same DNA scaffold. Again, a mixture of CelA-E-H-1, CBM-E-H-2, CelE-E-H-3, BglA-E-H-4 pre-blocked with the corresponding oligo I, II, III and

IV was used as a control. While only individual bands were detected for the pre-blocked proteins or the mixture, a significantly larger band was observed when all four proteins are assembled onto the DNA scaffold (Lane 12). The presence of all four components is further supported by the slower mobility protein-DNA complex for the four-component system as compared to the two-component system (Figure 3.6A). The ability of the assembled cellulosome structures to produce glucose from cellulose was further assessed. Compared with a mixture of free enzymes, a 3.1-fold enhancement in glucose production was achieved using the four-component cellulosome structure assembled on the DNA scaffold (Figure 3.6B). The level of enhancement is better than the two-component system, consistent with other reports indicating increased enzyme synergy with increasing number of cellulases recruited to the cellulosome structures. More importantly, the same four-component cellulosome structure when assembled onto the RCA template resulted in more than 5-fold enhancement in glucose release (Figure 3.6C). This represents the highest level of enhancement reported for artificial cellulosomes designed for complete liberation glucose from cellulose. This may be attributed to the flexibility of the DNA scaffold allowing the cellulosome structure to potentially wrap around a single cellulose fiber for enhanced hydrolysis.

3.4 Conclusions

In conclusion, we presented a new approach to create complex cellulosome structures based on DNA hybridization. The use of HaloTag fusions enables site-specific assembly without any effect on enzyme activity. The high-affinity interaction (typically in the range of 0.5nM) afforded by site-specific DNA hybridization over cohesin-dockerin pairs greatly improves the assembly efficiency and results in improved enzyme synergy. The levels of enhancement with our two-component and

four-component cellulosome structures are 50% higher than those achieved using protein scaffolds.⁸¹ More importantly, the use of RCA enables the simple amplification of binding sites for improved enzyme density on a single scaffold structure. The resulting 5-fold enhancement in glucose release using the complex RCA-derived cellulosome structure is the highest synergy reported so far. However, the true benefit of the strategy is the flexibility to embed additional hybridization sites to assemble a large array of cellulases to better mimic the enzyme diversity of the naturally occurring cellulosomes. This in combination with the possible to create more complex two-dimensional or three-dimensional DNA scaffolds structures for cellulosome assembly should pave the way to generate new cellulosome structures for improved cellulose processing.

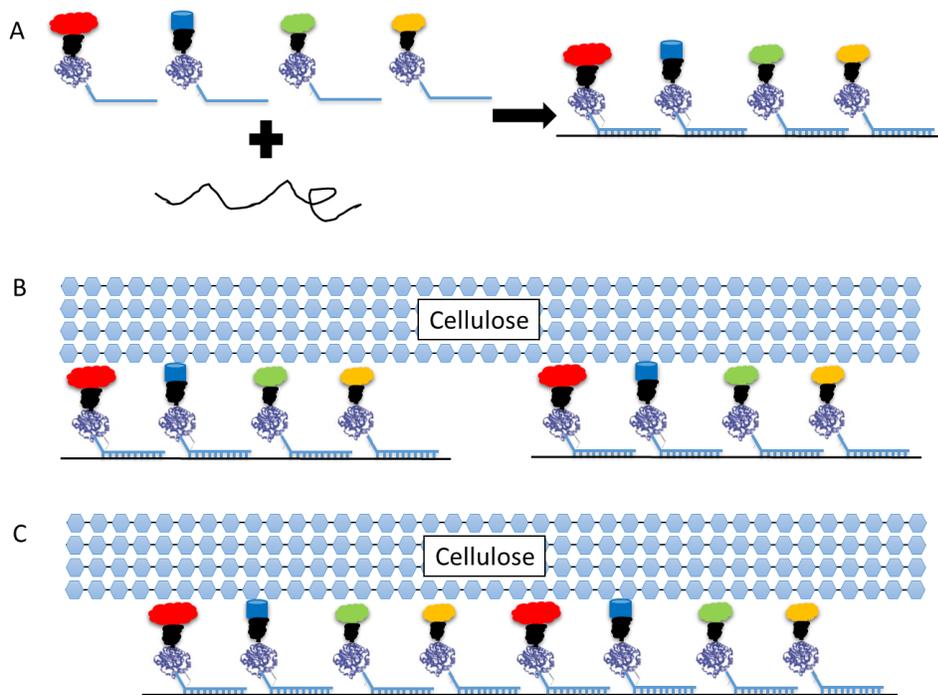


Figure 3.1 Schematic of artificial cellulosome on DNA template through HaloTag mediated site-specific DNA linker labeling. (A) Immobilization of cellulosome components on DNA template for cellulosome assembly; (B) Cellulose hydrolysis with cellulosome on discrete DNA template; (C) Cellulose hydrolysis with cellulosome on RCA DNA template.

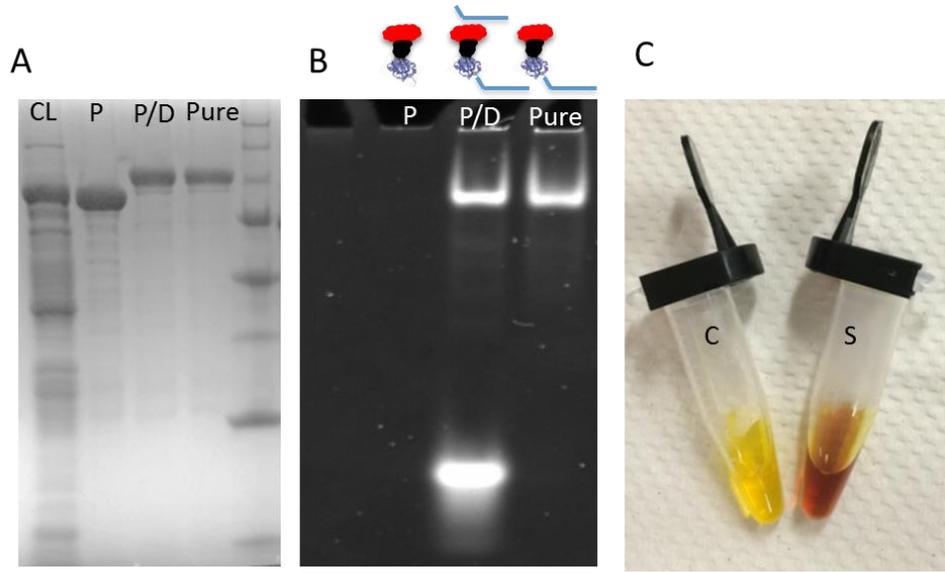


Figure 3.2 CelA-E-H function check. SDS-PAGE (A) and Native PAGE (B) analysis of DNA labeling of CelA-E-H and activity check of CelA after labeling and purification (C). Protein (P) was purified from cell lysis (CL). Purified CelA-E-H (P) with a 3-fold excess of the CH-modified DNA (P/D) were purified by going through one cycle of precipitation and solubilization to remove the excess DNA (Pure). Activity of CelA was checked by cellulose hydrolysis (S) compared with control (C).

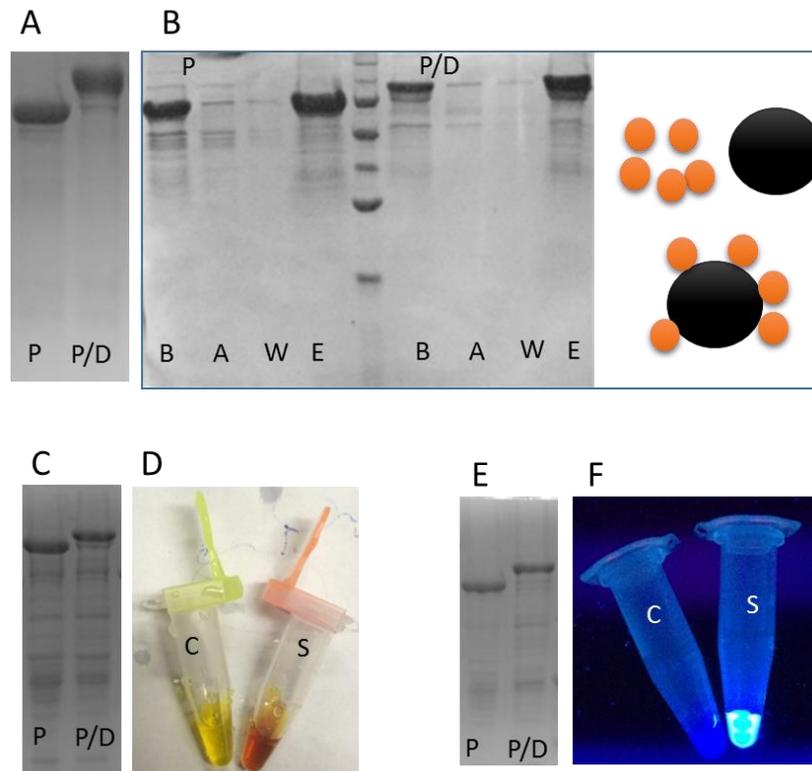


Figure 3.3 CBM-E-H, CelE-E-H and BglA-E-H function check. SDS-PAGE analysis (A, C, E) of DNA labeling of CBM-E-H, CelE-E-H, CelA-E-H and activity check (B, D, F) of CBM, CelE and BglA after labeling and purification. In A, C and E, first lane was purified protein (P) and second lane was purified protein with DNA linker labeling (P/D). (B) The cellulose-binding function of CBM-E-H. Purified proteins (P) or DNA linker 2 modified proteins (P/D) were mixed with avicel and the bound proteins were removed by centrifugation. The amount of CBM-E-A in the solution (B) before or (A) after binding was analyzed. After wash (W), the bound proteins were eluted (E) by boiling. Activity of CelE and BglA was checked by substrate hydrolysis (S) compared with control (c).

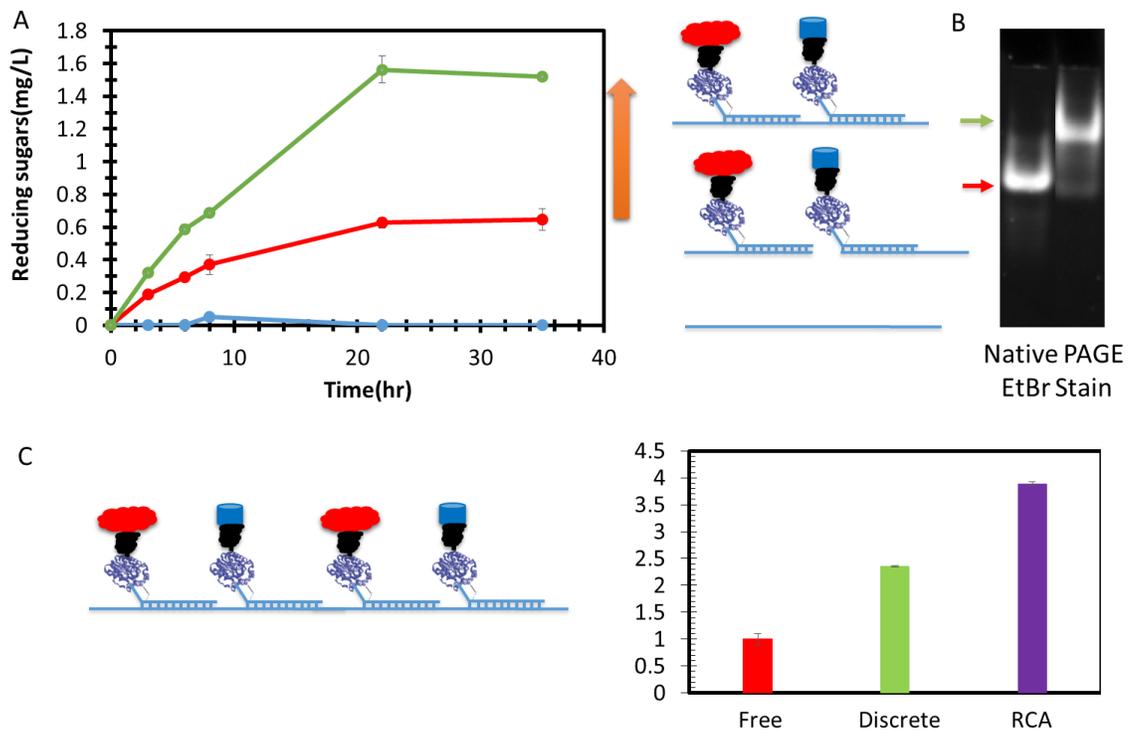


Figure 3.4 Assembly of artificial cellulosome with CelA and CBM. (A) Production of reducing sugars from DNA alone (blue), free enzymes (red) and the assembly cellulosome (green). (B) EMSA analysis of protein binding. Lane 1: CelA-E-H-1 mixed with CBM-E-H-2; Lane 2: both proteins with DNA template with two binding sites. (C) Artificial cellulosome on RCA template.



Figure 3.5 Rolling circle amplification of DNA template. The discrete DNA template (T) was circulated (C) and rolling circle amplified to synthesize long repeating DNA sequence (R). Control was the rolling circle amplification conducted without circulated DNA template as shown (E). Nothing was detected in (E) demonstrating the synthesis of long repeating DNA template was specifically from circulated DNA template.

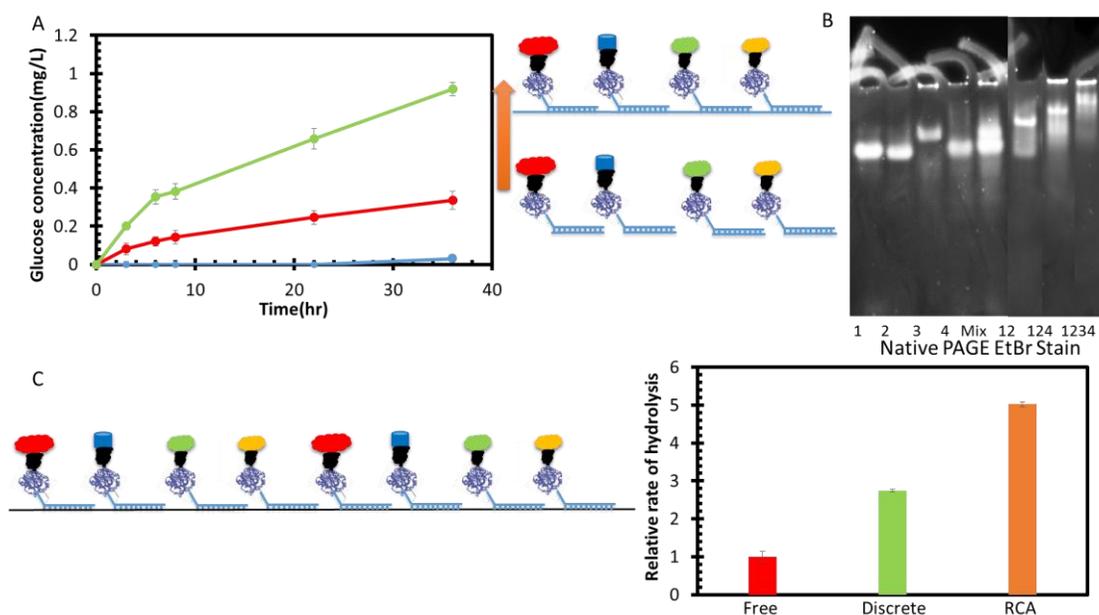


Figure 3.6 Assembly of artificial cellulosome with CelA, CBM, CelE and BglA. (A) Production of glucose from DNA alone (blue), free enzymes (red) and the assembly cellulosome (green). (B) EMSA analysis of protein binding. Lane 1, 2, 3, 4: CelA-E-H-1, CBM-E-H-2, CelE-E-H-3, BglA-E-H-4 mixed with blocker DNA; Lane 5: Mixture of Lane 1, 2, 3, and 4; Lane 12: CelA-E-H-1 and CBM-E-H-2 mixed with DNA template with four binding sites; Lane 124: CelA-E-H-1, CBM-E-H-2 and BglA-E-H-4 mixed with DNA template with four binding sites; Lane 1234: CelA-E-H-1, CBM-E-H-2, CelE-E-H-3, and BglA-E-H-4 mixed with DNA template with four binding sites. (C) Artificial cellulosome on RCA template.

Chapter 4

POST-TRANSLATIONAL MODIFICATION OF BIONANOPARTICLES AS A MODULAR PLATFORM FOR BIOSENSOR ASSEMBLY

Abstract

Context driven biosensor assembly with modular targeting and detection moieties is gaining significant attentions. Although protein-based nanoparticles have emerged as an excellent platform for biosensor assembly, current strategies of decorating bionanoparticles with targeting and detection moieties often suffer from unfavorable spacing and orientation as well as bionanoparticle aggregation. Herein, we report a highly modular post-translational modification approach for biosensor assembly based on sortase A-mediated ligation. This approach enables the simultaneous modifications of the *Bacillus stearothermophilus* E2 nanoparticles with different functional moieties for antibody, enzyme, DNA aptamer, and dye decoration. The resulting easy-purification platform offers a high degree of targeting and detection modularity. This flexibility is demonstrated for the detection of both immobilized antigens and cancer cells.

This Chapter is reproduced by permission of American Chemical Society. Qing Sun, Qi Chen, Daniel Blackstock and Wilfred Chen, Post-translational modification of bio-nanoparticles as a modular platform for biosensor assembly, *ACS Nano*, 9, 8554-8561, 2015.

4.1 Introduction

Analytes, including pathogens,^{97,98} toxins,^{99–101} tumor markers,^{28,102} and metabolites,¹⁰³ have attracted substantial attentions due to their impact on human health and daily lives. Biosensors, which employ biological elements to convert the recognition of analytes into detectable signals, have been widely used for monitoring biorecognition events.²⁹ The most common biosensor format employs a target-specific primary antibody for recognition and the use of secondary antibodies modified with either fluorescence dyes or enzymes to reflect the initial binding events.²⁷ Direct labeling of primary antibodies has also been reported, however, this method suffers from tedious chemical modifications, reduced binding affinity, and limited sensitivity.¹⁰⁴ Ideally, a universal adaptor platform that is able to link any recognition motif of interest to a detection module is of great interest, especially if signal amplification can be achieved to detect analytes with required sensitivity.²⁸

Recently, a bifunctional adaptor with the ability to link Immunoglobulin G (IgGs) to different DNA-modified output domains was constructed to achieve a high degree of modularity and selectivity.²⁷ Significantly improved target interaction and detection sensitivity have also been achieved by using complex three-dimensional nanoscaffolds, allowing dual modifications with IgGs and quantum dots (QDs).²⁸ This improved sensitivity is the result of signal amplification achieved by linking multiple QDs to a single binding event. A similar enhancement in detection sensitivity has been reported by immobilizing multiple biotin-conjugated antibodies and invertases onto streptavidin magnetic beads. Binding of a single bead onto targets resulted in the conversion of sucrose to glucose by multiple invertases for highly sensitive detection.¹⁰⁵

Among different nanoscaffolds, protein-based nanoparticles are the most attractive as they offer the ability to self-assemble from simple uniform-size protein

building blocks and allow precise control over material architectures.²⁹ Despite these desirable properties, protein bionanoparticles do have several drawbacks. Most noticeably is the use of chemical modifications for functionalization, which can result in random orientation and loss of functions.^{30,57–60} Direct genetic fusion has been reported previously to fuse protein G and 6xHis-tag on apoferritin surface to capture IgGs and nickel-nitriolotriacetic acid (Ni-NTA) coated quantum dots, gold particles or magnetic particles for sensor assembly.²⁸ However, genetic fusion is limited to relative small peptides and proteins, restricting the varieties of sensing and detection moieties that can be decorated.^{31,61} As a result, the ability for simultaneous modifications with multiple sensing and detection modules remains challenging. Ideally, a universal approach to interface different sensor components in the correct spatial orientation will be invaluable to satisfy the need to create highly modular biosensing platforms.²⁷

At the same time, purification of protein-based nanoparticles from cellular contaminants has been one obstacle for wider application of protein-based nanoparticles. Traditional ways including sucrose gradient, ultracentrifugation or chromatography are time and expense consuming, leading to the needs of exploiting an easy purification method of the protein-based nanoparticles.^{106,107}

Staphylococcus aureus Sortase A (SrtA) is a bacterial transpeptidase that catalyzes the condensation reaction between a C-terminal LPXTG recognition motif and an N-terminal tri-glycine tag to generate a native amide bond.^{62,108} The N to C ligation can be used to guarantee the spatial orientation of targeted moieties, while the small size of the tags has a minimal impact on functionalities. Since the ligation can proceed at mild pH and temperature conditions,⁶³ we argue that SrtA can be exploited to attach

multiple proteins and peptides onto protein nanoparticles without any effect on folding and functionalities.

Our group has recently demonstrated this feasibility of attaching multiple proteins onto the E2 core of the pyruvate dehydrogenase complex from *Bacillus stearothermophilus*, a genetically-modifiable 60-mer protein bionanoparticle of 24nm diameter.^{109–111} Simple recovery of modified E2 nanoparticles was achieved using two cycles of thermal precipitation and resolubilization by taking advantage of the tethered thermal-responsive elastin-like-polypeptide (ELP) moiety.^{65,85,112} This thermal based purification scheme enables recovery of bio-nanoparticles within 1 h without the use of conventional sucrose gradient, ultracentrifugation or chromatography that are time-consuming and expensive.^{106,107} Because of the versatility in decorating E2 with proteins of different sizes and properties and the ease in purification, this approach is ideal in transforming E2 into a modular sensing platform capable of facile exchange and integration of different targeting and detection components. To illustrate this modularity, we modified the E2 nanoparticles with a combination of the antibody-binding Z domain,¹¹³ the DNA- or fluorescent dye-conjugating HaloTag,^{54,56} and nanoluciferase (NLuc)¹¹⁴ to form highly specific and sensitive biosensors for the detection of antigens and cancer cells (Figure 4.1, Figure 4.2 and Figure 4.7).

4.2 Materials and Methods

4.2.1 Genetic manipulation

Escherichia coli strain NEB 5-alpha (NEB #C2987I) was used as the host for genetic manipulations. BL21 (DE3) (Novagen) and BLR (DE3) (Novagen) were used for protein expressions. Details of DNA manipulation and expression procedures can be

found in the following session for GGG-E2, His6-Sortase A, ELP-LPETG, Nluc-LPETG, Z-Ct-LPETG, HaloTag-LPETG.

E2 monomer was expressed using plasmid pGGGE2, constructed by amplifying E2(158) [E2 encoding amino acids 158-427] fragment using the E2-Forward and E2-Reverse primer. The PCR product was inserted into pET11(a) at *NdeI* and *BamHI* restriction sites. ELP-LPETG-His6 was constructed by first overlapping 5 oligonucleotides: Oligo 1, Oligo 2, Oligo 3, Oligo 4 and Oligo 5 to form the fragment of (G4S)₃-LPETG-His6 with *BamHI* and *XhoI* overhangs. The overlapping oligo was ligated into *BamHI* and *XhoI* digested pET24(a)-ELP[KV8F- 40] (Liu, Tsai, Madan, & Chen, 2012) to form pET24(a)-ELP-LPETG.

For HaloTag-LPETG and Nluc-LPETG, HaloTag and Nluc were first individually PCR amplified with HaloTag-Forward, HaloTag-Reverse and Nluc-Forward, Nluc-Reverse. These two PCR amplified fragments were inserted into pET24(a)-ELP-LPETG using *NdeI* and *BamHI* restriction sites to form HaloTag-LPETG and Nluc-LPETG.

Z-Ct-LPETG was completed through three-step cloning. GST (Glutathione S-transferase) was first PCR amplified with primers GST-Forward and GST-Reverse. The PCR product was inserted into pET24(a)-ELP-LPETG at *NdeI* and *BamHI* restriction sites to form GST-LPETG. Z-domain was PCR amplified with Z-Forward and Z-Reverse and inserted into GST-LPETG at *NdeI* restriction site to form Z-GST-LPETG. Cohesin Ct was then PCR amplified with Ct-Forward and Ct-Reverse and inserted into Z-GST-LPETG at *HindIII* and *BamHI* to form Z-Ct-LPETG.

The control fusion protein Z-ELP-Nluc was constructed by first inserting PCR amplified Z domain into *NdeI* digested pET24(a)-ELP-LPETG to form Z-ELP-LPETG.

Then Nluc was PCR amplified with Nluc-Forward and Nluc-Reverse and inserted into *Bam*HI digested Z-ELP-LPETG to form Z-ELP-Nluc.

Lists of primers used for gene manipulations are:

E2- 5'-
Forward GCGCCATATGGGCGGTGGAGCTAGCGTGCTGAAAGAAGAC-3'
E2- 5'- GCGCGGATCCTTAAGCTTCCATCAGC-3'
Reverse
Oligo 1 5'- GAT CCC CGC GGG GCG GCG GAG GGT CTG GAG GAG GCG
GTT CAG GTG GTG GGG GCT -3'
Oligo 2 5'- CCC TAC CAG AAA CCG GAG GAT CTG GCC ACC ATC ACC
ATC ACC ACT GAT AAC -3'
Oligo 3 5'- TCG AGT TAT CAG TGG TGA TGG TGA TGG -3'
Oligo 4 5'- TGG CCA GAT CCT CCG GTT TCT GGT AGG GAG CCC CCA
CCA CCT GAA CCG CCT -3'
Oligo 5 5'- CCT CCA GAC CCT CCG CCG CCC CGC GGG -3'
HaloTag- 5'- CCT CCA GAC CCT CCG CCG CCC CGC GGG -3'
Forward
HaloTag- 5'- TGG CGC GGA TCC GCC GGA AAT CTC GAG CGT CG -3'
Reverse
Nluc- 5'- GGG AAT TCC ATA TGG TCT TCA CAC TCG AA -3'
Forward
Nluc- 5'- TTC CGC GGG GAT CCG G -3'
Reverse
GST- 5'- GGG AAT TCC ATA TGA TGC ATG GCA AAA CCC AGG CGA
Forward CCA GCG GCA CCA TTC AGA GCA TGC ATG GCA AAA CCC
AGG CGA CCA GCG GCA CCA TTC AGA GCA TGC ATG GCA
AAA CCC AGG CGA CCA GCG GCA CCA TTC AGA GCA AGC
TTG GCG GCG GTG GTA GCT CCC CTA TAC TAG GTT ATT GGA
AAA TTA AGG GCC -3'
GST- 5'- TAT TGG CGC GGA TCC CAG GGG CCC C -3'
Reverse
Z- 5'- GGG AAT TCC ATA TGA GCG GCA GCG GCA GC -3'
Forward
Z- 5'- TCC CAA GCT TAC TGC CAC CGC TCC CGC CTC CGC TAC
Reverse CGC CTC CTT TCG GCG CCT GAG CAT -3'
Ct- 5'- GGG AAT TCC ATA TGC CAT CAA CAC AGC CTG TAA CAA
Forward CAC C -3'
Ct- 5'- TAT TGG CGC GGA TCC TAT ATC TCC AAC ATT TAC TCC
Reverse ACC GTC AAA GAA C -3'

4.2.2 Protein expression

Plasmid pMR5-His6-Sortase A was a gift from Dr. Eric T. Boder, University of Tennessee, Knoxville and used as a source of Sortase A. SrtA, GGG-E2, Nluc-LPETG, HaloTag-LPETG, Z-Ct- LPETG were expressed in BL21(DE3) (*F2 ompT gal dcm lon hsdSB(rB2 mB2) kDE3*).

For SrtA expression, overnight SrtA cultures were sub-inoculated into 25 mL Luria-Bertani (LB) medium supplemented with 50 µg/mL kanamycin. The culture was grown at 37°C until OD600 reached 1. Protein expression was induced at 37 °C for 4 hours with 1 mM IPTG.

GGG-E2 was inoculated with LB medium supplemented with 100ug/mL ampicillin, induced with 0.2 mM IPTG when OD600 reached 0.5 and grown at 20 °C overnight for protein expression.

HaloTag-LPETG was inoculated in LB medium supplemented with 50 µg/mL kanamycin, induced at 30°C for 4 hours by 0.25 mM IPTG at OD 0.5.

Z-Ct-LPETG was cultured in LB medium supplemented with 1.5% glycerol and 50 µg/mL kanamycin at 37 °C until OD600 reached 1. Protein expression was induced by 0.2mM IPTG at 25 °C for overnight. Nluc-LPETG was cultured in Terrific Broth (TB) medium until OD reached 1. 0.2mM IPTG was used to induce protein expression for overnight at 25 °C.

ELP and Z-ELP-Nluc was expressed in *E. coli* BLR [*F- ompT hsdSB (r-B m-B) gal dcm(DE3) Δ (srl-recA)306::Tn10(TetR)*]; Novagen, Madison,WI]in TB medium.

After protein expression, cells were harvested by centrifugation at 4200 g, resuspended in SrtA ligation buffer (50 mM Tris, 150 mM NaCl, pH 8) and lysed by

sonication. The cell debris was removed by centrifugation at 16.1 k rcf for 10 min at 4 °C.

Nluc-LPETG, Z-Ct-LPETG and GGG-E2 were partially purified based on their thermostable character. Nluc-LPETG was incubated at 55 °C for 10 min, while Z-Ct-LPETG, GGG-E2 was incubated at 70 °C for 10 min to denature most of the E. coli contaminant proteins. The aggregated contaminants were removed by centrifugation at 15,000 rpm at 4 °C for 10 min. Purification of the ELP fusion protein was achieved by two cycles of inverse phase transition. NaCl was added to the cell lysates to a final concentration of 1 M and the mixture was incubated at 37 °C for 10 min before centrifugation for 30 min at 15,000 rpm at 37 °C. The pellet was resuspended in ice-cold buffer and centrifuged for 30 min at 15,000 rpm at 4 °C to remove insoluble cellular proteins. This precipitation and resolubilization process was repeated a second time. The purity of the purified protein was determined by 10% SDS-PAGE electrophoresis followed by coomassie blue staining. The molar concentration of SrtA, GGG-E2, Nluc-LPETG, HaloTag-LPETG, Z-Ct-LPETG and ELP-E2 cage were checked by Bradford protein assay from Bio- Rad (Hercules, CA). The molar concentration of ELP was calculated according to absorbance at 215 nm.⁶⁵

4.2.3 Individual components ligation onto ELP-E2 and function check

To prepare ELP-E2, 1 μM of E2 cage, 12 μM of ELP-LPETG and 12 μM of Sortase A were used in a reaction volume of 400 μL reaction buffer (50 mM Tris, 150 mM NaCl, 60 mM CaCl₂, pH8). Reaction mixtures were incubated at 37 °C for 4 h. The ligated products were collected by inverse phase transition with addition of 1 M Na₂SO₄, 37 °C incubation for 10min and centrifugation at 160,000g for 10min at 37 °C. Then after supernatant removal, pellets were resolubilized in 4 °C cold buffer. This thermal

cycle was repeated for better purity. To ligate Z-Ct-LPETG onto ELP-E2 (10% E2 monomers ligated with ELP), 1 μ M of ELP-E2 cage, 12 μ M of Z-Ct-LPETG and 12 μ M of Sortase A were used in 400 μ L reaction volume. The precipitation and resolubilization procedures were the same as ELP-E2 as mentioned above. The resulting products were analyzed with 10% SDS-PAGE. Nluc-LPETG and HaloTag-LPETG ligation onto ELP-E2 were achieved with similar procedures.

Z-domain functionality was checked by capturing and purifying rabbit antibody. 1 μ M Z- ELP-E2 cage was incubated with 20 μ M RIgG for 3 hours at room temperature. Then 1M NaCl was added into the solution and the mixture was incubated at 28 $^{\circ}$ C for 10 min before 160,000 g centrifugation for 10 min at 15,000 rpm and 28 $^{\circ}$ C. The pellet was resuspended in ice-cold binding buffer for resolubilization. Full complex assemblies were visualized by transmission electron microscopy. Samples were prepared in DDI water and stained with 2% uranyl acetate on carbon-coated copper grids (Electron Microscopy Science). Zeiss Libra 120 Transmission Electron Microscope was used to visualize the samples with voltage of 120 kV.

Nluc activity was checked using Nano-Glo Luciferase assay system from Promega (Madison, WI). Assay substrate and Nluc cell lysate was allowed to equilibrate to room temperature (25 $^{\circ}$ C). Then one volume of Nano-Glo luciferase assay reagent was mixed with one volume of diluted Nluc cell lysate. After Nluc catalyzing the transfer of Furimazine to Furimamide with light emitting as product, luminescence was captured by Synergy H4 hybrid multi-mode microplate reader from BioTek (Winooski, VT).

HaloTag functionality on HaloTag-ELP-E2 was checked by chlorhexoane modified anti-MUC1 aptamer, which was ordered from Integrated DNA Technologies

(Coralville, IA). The Anti-MUC1 ordered was modified with a 5' amine group with a sequence of CTT CTC TCT TCC TCT CTC CTT GCA GTT GAT CCT TTG GAT ACC CTG G. Upon arrival, the aptamer was modified with the Chlorohexane (CH) ligand from Promega (Madison, WI). The CH ligand was mixed with the MB at a molar ratio of 30:1 and incubated at room temperature for 4 hours and 4°C for overnight. The mixture was then purified by removing excess CH ligand using a 3,000 Da ultrafiltration column (Vivaspin 500, Sartorius Stedim Biotech). The modified and purified aptamer was then mixed with the HaloTag-ELP-E2 at 3 aptamer to 1 HaloTag-E2 ratio and incubated overnight at 4 °C. Extra aptamer was removed by going through ELP thermal cycle purifications. SDS-PAGE and coomassie blue was used to visualize the labeling efficiency.

The hydrodynamic diameters of the complexes were measured by dynamic light scattering using Zetasizer Nano ZS (Malvern Instruments).

4.2.4 Sensor assembly

Biosensor assembly was achieved by three step ligations. ELP-LPETG was first ligated onto GGG-E2 to form ELP-E2 with 10% ligation efficiency. Then No. 1 protein was ligated onto ELP-E2 to form ELP-E2-No. 1 protein and purified as described. No. 2 protein was then ligated onto ELP-E2-No. 1 protein to achieve ELP-E2-No. 1 protein-No. 2 protein. The No. 1 protein refers to Z-Ct-LPETG or HaloTag-LPETG and No. 2 protein refers to Nluc-LPETG or HaloTag-LPETG. During the three-step ligation, No. 1 and No. 2 protein conjugation efficiency can be easily adjusted by changing No. 1 or No. 2 protein concentration ranging from 12 μ M to 60 μ M.

4.2.5 Characterization of protein complex

The ligation products were analyzed by SDS-PAGE. The ratio of ligated products onto E2 was estimated by densitometry, in which GelQuant (BiochemLabSolutions) was used to quantify the band densities. Five repeats were used to obtain the standard deviation associated with the decoration efficiency of each component. The hydrodynamic diameters of the complexes were measured by dynamic light scattering using Zetasizer Nano ZS (Malvern Instruments). The samples were in buffer of 50 mM Tris and 150 mM NaCl at pH8 for dynamic light scattering. Full complex assemblies were visualized by transmission electron microscopy. Samples were prepared in DDI water and stained with 2% uranyl acetate on carbon-coated copper grids (Electron Microscopy Science). Zeiss Libra 120 Transmission Electron Microscope was used to visualize the samples with voltage of 120 kV.

4.2.6 Thrombin detection

Varying concentration of thrombin was coated onto 96 well plates by incubation at room temperature for 2 hours. 200uL 5% Milk was then used to block the wells by incubating at 4 °C for overnight. 100 µL 15 nM thrombin antibody was added into the wells and incubated for 1 hour. Cage decorated with Z domain and Nluc was added into the wells. After 2 hours' interaction, cages were removed and 100 µL room temperature Nano-Glo luciferase assay reagent were added. Luminescence were checked by microplate reader. Wells were washed 4 times with TPBS between every step to wash out extra reagent.

4.2.7 Tumor marker detection

Hela cells were obtained from American Type Culture Collection (ATCC). The cells were grown and cultured at 37 °C in a 5% CO₂ atmosphere with 1x autoclavable

minimum essential medium (Life technologies: 11700-077) containing 1 vol% of 7.5% NaHCO₃, 2 vol% of 1M Hepes, 1 vol% of non-essential amino acids (Fisher: SH3023801), 2.5 vol% of Pen/Strep (Fisher: SV30010), 2 vol% of L-glutamine (Fisher: SH3003401) and 10 vol% of FBS (Sigma-Aldrich: F6178).

Hela cells were seeded into 96 well plates and grown at 37 °C in 5% CO₂ overnight, until the well cell density reached around 90% confluency. The growth media was then removed and the cells were washed twice with 1x TBS. 100uL 4% formaldehyde was used to fix Hela cells for 10min and 100uL 100mM NH₄Cl was used to incubate with fixed Hela cells to minimize background fluorescence. 5% milk was used to block unspecific binding. Then 100uL 5ug/mL Anti-MUC1 antibody was used to incubate with Hela cells for 1 hours to bind onto surface MUC1 tumor marker, and 100uL 10nM Z-ELP-E2-Nluc was used to bind onto antibody. Nluc activity was checked with luminescence by detecting luminescent light with microplate reader. 4 times washes with TPBS between every step were used to remove extra reagent. To use Z-ELP-E2-HaloTag-Alexa 488 for cancer detection, microscope with filter was used to detect cancer rather than Nluc activity.

To use aptamer for cancer detection, after fixing Hela cells and blocking, Anti-MUC1 aptamer and Alexa 488 modified cage was used to incubated with Hela cells to bind and reflect surface cancer marker. All imaging was performed on a Zeiss AxioObserver Z1 inverted fluorescence microscope.

4.3 Results and Discussion

4.3.1 Individual component ligation

The E2 core is composed of 60 identical monomers that self-assemble into a highly stable cage-like structure (Figure 4.8A).¹⁰⁹ A tri-glycine tag was added to the N-terminus of the E2 monomer and up to 60 different protein moieties can be ligated onto a single E2 nanoparticle. Antibody, one of the most commonly used binding motifs, was chosen as the initial recognizing module. Since it is difficult to modify each target antibody individually, an antibody capturing Z domain, a shorter synthetic domain derived from the *S. aureus* protein A, was used for antibody immobilization. The Z domain has a reported binding affinity of 10 nM to the Fc region and can be used to properly orient IgGs without affecting accessibility of the Fab domains.¹¹³ To achieve high level expression, the Z domain was fused to a thermally stable carrier protein, cohesin Ct from *Clostridium thermocellum*.¹⁹ A LPETG tag was added to the C-terminal of Ct for SrtA ligation. Expression of Z-Ct-LPETG fusion was confirmed by SDS-PAGE analysis (Figure 4.9A). One of the added benefits of using a thermally stable partner is the ability to purify the fusion protein by incubating at 70°C for 10 min to denature most other cellular proteins (Figure 4.9A). The flexibility to choose any fusion partner of interest without size limitation is another advantage of post-translational modification compared with direct fusion.

ELP-E2 conjugates were first prepared by StrA ligation as reported previously (Figure 4.7A).^{63,111} Next, Z-Ct-LPETG was further ligated onto purified ELP-E2 to form Z-ELP-E2 conjugates and purified again using two cycles of thermal precipitation and resolubilization. The final purified product was verified by SDS-PAGE analysis and a new band corresponding to Z-Ct-E2 was detected (Figure 4.3A). From the band

density analysis, roughly 5 Z-Ct-LPETG were ligated per E2 cage for antibody capture. To check the function of ligated Z-domains, rabbit IgGs were incubated with Z-ELP-E2 nanoparticles at a 3:1 ratio and thermal purification was used to remove unbound antibodies. Successful capturing of IgGs was confirmed by detecting bands corresponding to the heavy chain fragments (Figure 4.9B) after binding and purification. Transmission electron micrograph (TEM) images further confirmed the capture of Y-shaped IgGs by the ligated Z domain (Figure 4.3B and 4.8B).

In many sensitive immunoassays, enzyme is a popular choice as the sensor output due to the ability to amplify signals through enzyme activity. NanoLuc (Nluc) luciferase is a small (19.1kDa) monomeric luminescent reporter enzyme engineered from deep sea *Oplophorus gracilirostris*.¹¹⁴ Nluc is an ATP-independent luciferase, about 150-fold brighter than either firefly (*Photinus pyralis*) or *Renilla reniformis* luciferase. Since it is a thermally stable enzyme with a melting temperature of 58°C, we were able to partially purify Nluc-LPETG by incubating cell lysis at 55°C (Figure 4.10A). Nluc was ligated onto ELP-E2 nanoparticle and purified as described above (Figure 4.10A). The Nluc conjugates were fully functional as verified by the detection of luminescence (Figure 4.10B).

4.3.2 Sensor assembled for thrombin detection with signal amplification

With the individual sensing and detection components expressed and ligated onto ELP-E2 nanoparticle, the first generation of biosensor was assembled using the Z-domain for antibody capturing and Nluc as the detection module. To generate this multifunctional particle, a sequential ligation procedure was employed to control the decoration efficiency of the individual component in each step. Z-Ct-LPETG was first ligated onto ELP-E2 to form Z-ELP-E2, followed by the ligation of Nluc-LPETG to

assemble the Z-ELP-E2-Nluc conjugates (Figure 4.4A and Figure 4.11A). This sequential ligation strategy was utilized to provide a decoration efficiency of 5 Z domains and 5 Nluc per nanoparticle, which could provide up to 5 times signal amplification upon antibody binding (Figure 4.11A). Dynamic light scattering confirmed the correct nanoprobe assembly as the particle size increased from 32 nm for Z-ELP-E2 to 40 nm for the IgG-Z-ELP-E2-Nluc conjugate (Figure 4.11B).

To test the functionality of assembled biosensors, thrombin, a human protease involved in the coagulation cascade for anti-clotting therapeutics, thrombosis and hemostasis, was used as the analyte.^{115,116} As a control, a fusion protein with the Z domain in the N-terminus, Nluc in the C-terminus, and ELP in between as a linker for purification, was constructed (Figure 4.4A and Figure 4.11C). As expected, both constructs were successful in detecting 10 nM thrombin. However, the signal was 5-fold higher with the E2 nanoprobe, a result consistent with the 5 times higher ratio of Nluc per nanoparticle (Figure 4.4B).

To further investigate whether we can adjust the level of signal amplification, we took advantage of the two-step ligation procedure to fix the amount of Z-Ct ligated on ELP-E2 while varying the decoration efficiency of Nluc in the second step by changing the reactant concentration (Figure 4.7C). Densitometry analysis revealed that up to 22 Nluc per nanocage was achieved (Figure 4.4C). With the use of this series of E2 nanoprobe for thrombin detection, a corresponding increase in the signal amplification was detected with increasing ratios of Nluc per nanoprobe (Figure 4.4D). This result highlights the flexibility in fine-tuning the signal amplification by changing the detection module decoration efficiency. Using the nanoprobe containing 22 Nluc, we further investigate the sensitivity of the assay. As low as 2.5nM thrombin was easily

detected (Figure 4.11D); this detection limit is in line with other reported enzyme-based ELISA assay for thrombin.¹¹⁷

4.3.3 Sensor assembled for tumor marker detection with modularity

We next demonstrated the utility of E2 nanoprobe for the detection of cancer cells. MUC1, a cell surface associated glycoprotein, was chosen as the target, whose overexpression was associated with a variety of breast, ovarian, lung, and pancreatic cancers.^{118,119} A well-characterized anti-MUC1 antibody was loaded onto our nanoprobe to detect the surface-exposed MUC1 markers on fixed HeLa cells (Figure 4.5A). A significant Nluc activity was detected only with antibody-loaded nanoprobe, while no signal was observed for the control nanoprobe without antibodies (Figure 4.5B).

While very sensitive, this Nluc-based assay cannot be easily adapted to traditional microscopy imaging for cancer detection.¹²⁰ To address this problem, we replaced Nluc with Alexa 488 as the output while keeping the Z-domain for antibody capture. HaloTag, a modified haloalkane dehalogenase capable of linking a suicide chlorohexane (CH) ligand covalently to the protein, was employed to conjugate CH-modified Alexa 488.⁵⁴ The expression of HaloTag-LPETG was confirmed by SDS-PAGE (Figure 4.12A), and the successful ligation of HaloTag onto ELP-E2 was achieved (Figure 4.12A). The functionality of the ligated HaloTag was further demonstrated by the ability to conjugate with CH-modified Alexa 488 (Figure 4.12B). By simply replacing Nluc with HaloTag in the second step of the sequential ligation reaction, a new nanoprobe composed of the MUC1 antibody and Alexa 488 was constructed to visualize fixed HeLa cells (Figure 4.12C). Again, brightly fluorescent cells were detected only with nanoprobe loaded with antibodies (Figure 4.5C). This

example highlights the modularity of our approach because of the flexibility and the ease of exchanging a new detection module using SrtA-mediated ligation.

In addition to antibody, DNA aptamers have gained considerable attention as a sensitive sensing component since they can be selected to bind many target antigens with specificity and affinity rivaling that of antibodies.^{121,122} In addition, DNA can also function as a detection module through hybridization to DNA-modified reporter molecules such as DNA nanobarcodes.¹²³ To provide a universal platform to incorporate DNA into the E2 nanoprobe, HaloTag was again exploited to conjugate to CH-modified DNA. A MUC1-specific DNA aptamer¹¹⁹ was conjugated to HaloTag, and the DNA-HaloTag-ELP-E2 conjugates were confirmed by detecting a slower mobility band compared with free HaloTag-ELP-E2 protein (Figure 4.6A). Alexa 488 was further conjugated onto the nanoprobe for detection purpose (Figure 4.12B). Similar to the results with antibody, HeLa cells were detected only with nanoprobe loaded with DNA aptamers (Figure 4.6B). Since DNA aptamers and Alexa 488 shared the same HaloTag, the fluorescence signal was less compared with that for the antibody and Alexa 488 combination. This can be easily addressed by ligating more HaloTag onto the nanoprobe for improved sensitivity in a fashion similar to the case with Nluc. This is precisely the modularity of our approach in tuning the required ratio and function of individual sensing and detection modules for each unique application.

4.4 Conclusions

In summary, we have presented a modular approach to develop highly sensitive nanoprobe by decorating E2 protein nanoparticles with a wide range of sensing and detection functionalities using SrtA-mediated ligation. The use of Z domain allows not only directional immobilization of antibody by targeting the Fc region, but also the

virtually unlimited number of commercially available antibodies for sensing specificities. Similarly, an unlimited number of target-specific DNA aptamers can be conjugated to the nanoprobe using the HaloTag. However, the flexibility is beyond just the sensing module. Our approach also provides the feasibility to select either enzymes or fluorescent dyes for detection. The sensitivity can be easily fine-tuned by controlling the number of enzymes or dyes conjugated onto the nanoprobe and the use of more than one detection module is possible because of the modular nature of the approach. The multistep nanoprobe assembly process is further simplified by the ELP-based purification scheme. In addition to the sensing and detection modalities reported, there are many other examples such as single chain variable fragments, fibronectin type III domains, and fluorescence proteins that can be ligated onto the E2 nanoprobe. Even inorganic gold nanoparticles and quantum dots can be easily coupled by employing specific metal binding peptides to further expand the range of applications that can be addressed using this modular platform.

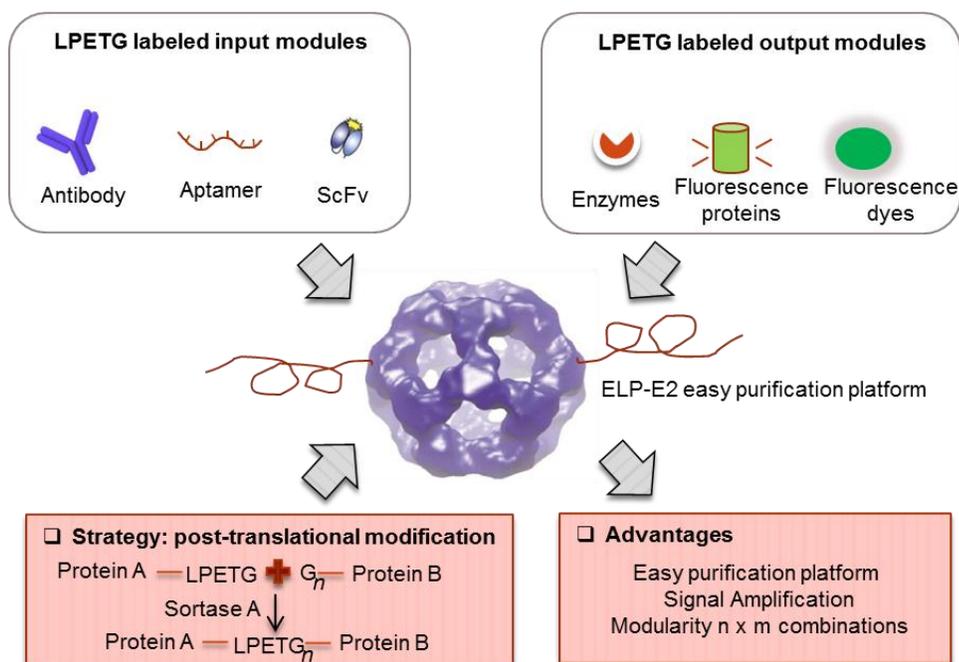


Figure 4.1 **Schematic of sortase A-mediated modification of ELP-E2 cage for sensor assembly.** The LPETG tagged input and output modules were ligated onto N-terminal triglycine tag of E2 cage through sortase A mediated ligation. Signal amplification and modularity are the expected outcome advantages of this system.

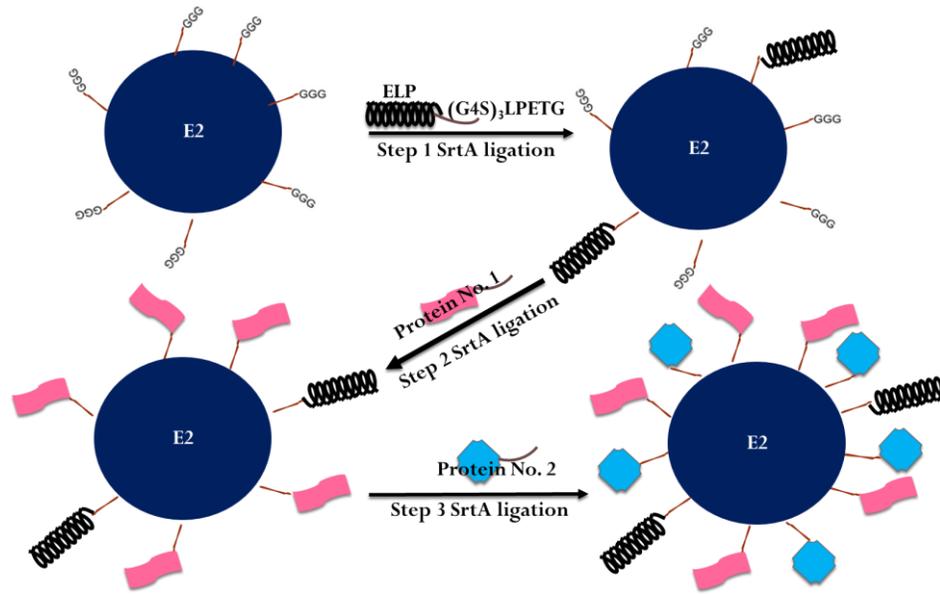


Figure 4.2 Scheme for stepwise ligation and purification of ELP-E2-Protein No. 1 (Z-domain or HaloTag)-Protein No. 2 (Nluc or HaloTag).

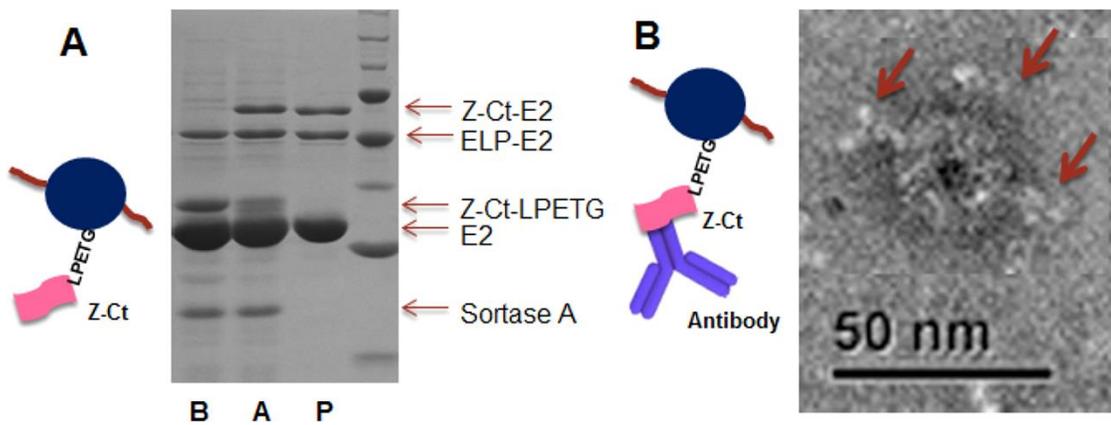


Figure 4.3 Sortase A-mediated Z-Ct-LPETG ligation onto ELP-E2 cage. (A) SDS-PAGE analysis of reaction mixtures before (B) and after (A) ligation as well as purified products (P). (B) Transmission electron micrograph of IgG-bound E2 nanoparticles. Y-shaped IgGs were shown to bind onto intact E2 nanoparticles.

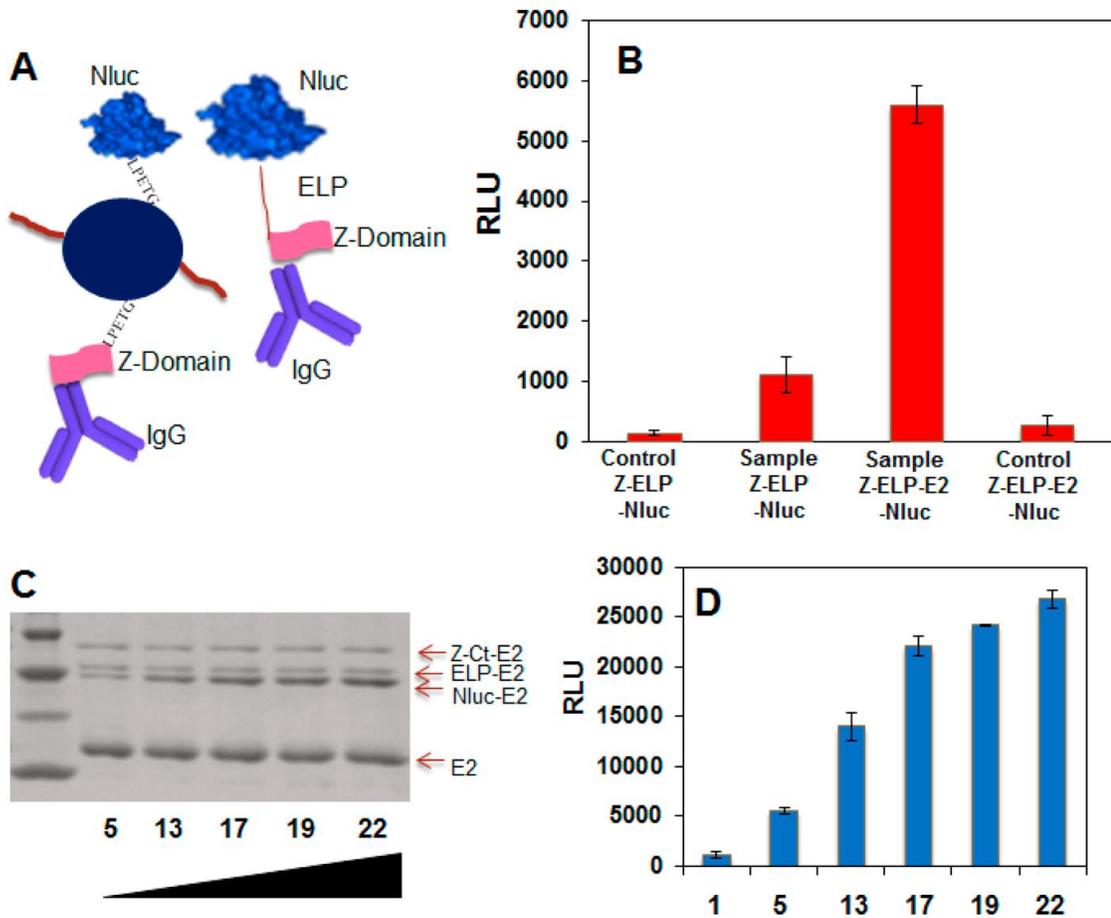


Figure 4.4 E2 nanoprobe assembly with antibodies captured by the Z domain for sensing and Nluc for detection. (A) Schematic representations of IgG-Z-ELP-E2-Nluc and IgG-Z-ELP-Nluc. (B) Thrombin detection with Z-ELP-Nluc fusion proteins and IgG-Z-ELP-E2-Nluc nanoprobe; control samples contained no thrombin. (C) Signal amplification using controlled Nluc assembly by conjugating varying numbers of Nluc (5_22) onto E2 nanoparticles. (D) Detection of thrombin using E2 nanoprobe with varying Nluc decoration efficiencies.

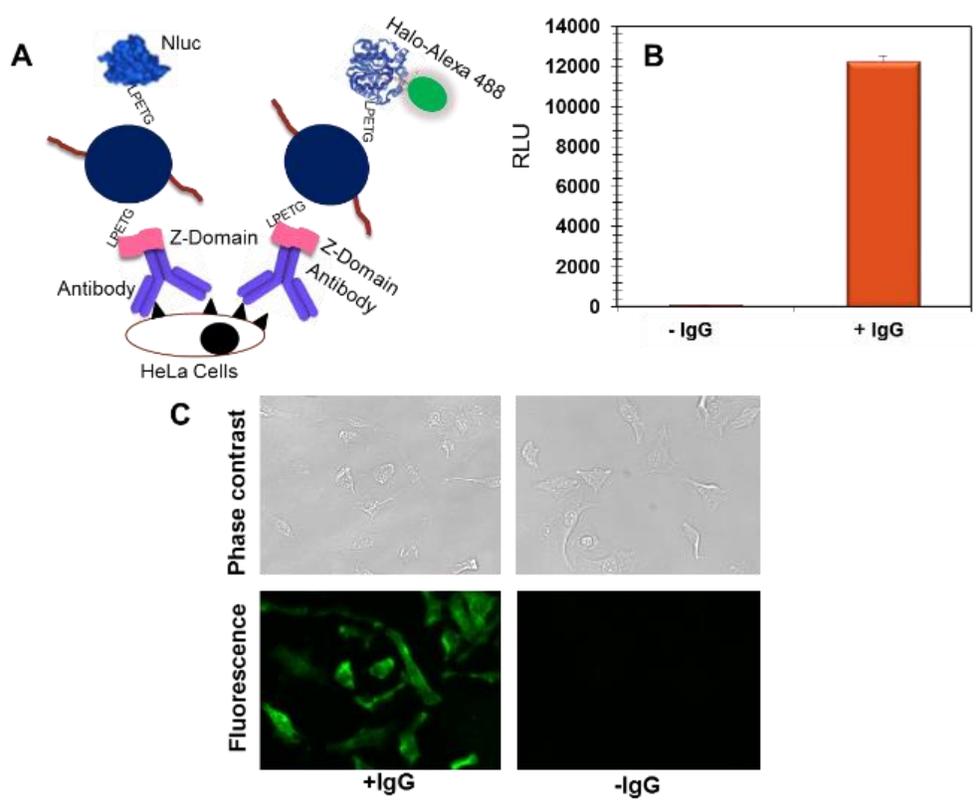


Figure 4.5 **Tumor marker detection.** (A) Detection of MUC1 on HeLa cells using E2 nanoprobes. (B) Detection of MUC1 with Nluc as the detection output. (C) Phase contrast and fluorescence images of HeLa cells detection with Alexa 488 as the detection output.

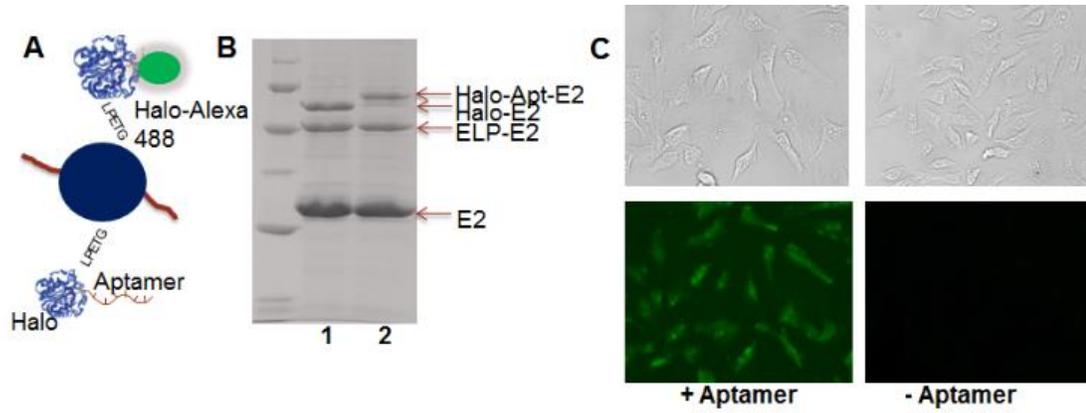
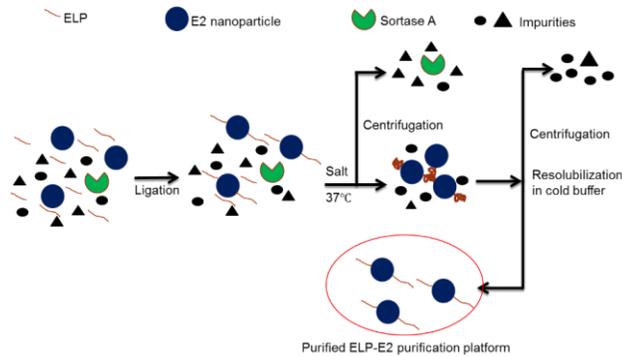
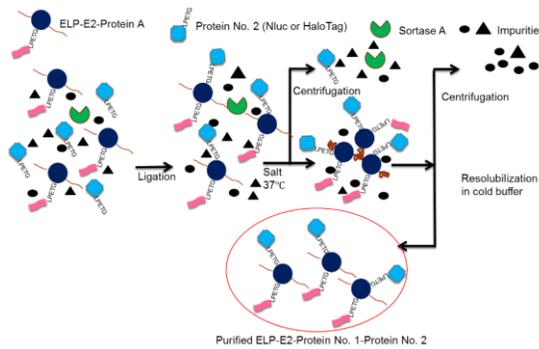


Figure 4.6 **Tumor marker detection using DNA aptamer.** (A) Conjugation of MUC1 aptamers onto HaloTag-ELP-E2 nanoparticles. (B) Phase contrast and fluorescence images of HeLa cells detection with using MUC1 aptamer-modified E2 nanoprobes.

A Step 1: ELP ligation on E2 nanoparticle



B Step 2: Protein No. 1 (Z-Ct or HaloTag) ligation



C Step 3: Protein No. 2 (Nluc or HaloTag) ligation

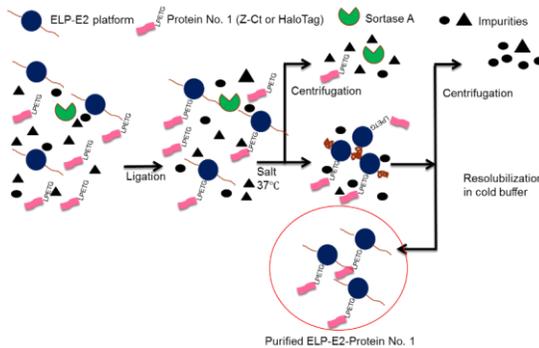


Figure 4.7 Detailed diagram of the three-step ligation. (A) Step 1: Elastin-like-polypeptide ligation on E2 nanoparticle; (B) Step 2: Protein No. 1 (Z-Ct or HaloTag) ligation on ELP-E2 platform; (C) Step 3: Protein No. 2 (Nluc or HaloTag) ligation on ELP-E2-Protein No. 1.

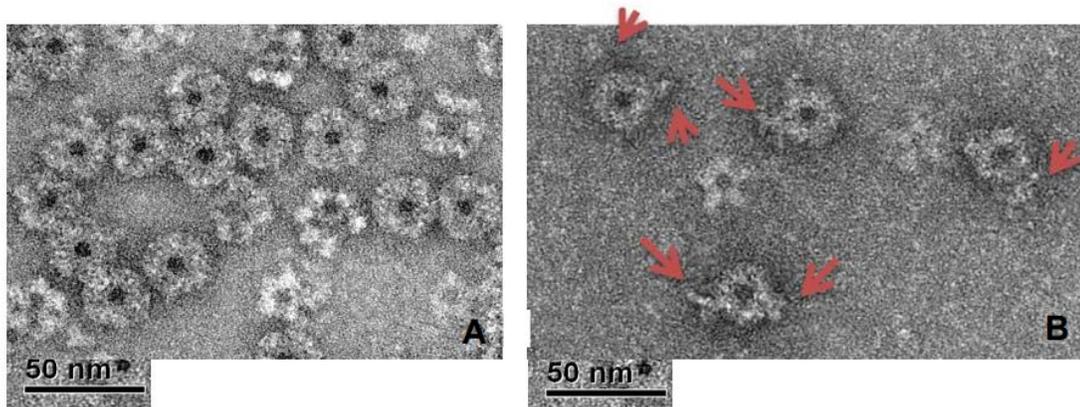


Figure 4.8 Transmission electron micrograph of (A) unmodified E2 nanoparticles and (B) IgG- bound E2 nanoparticles.

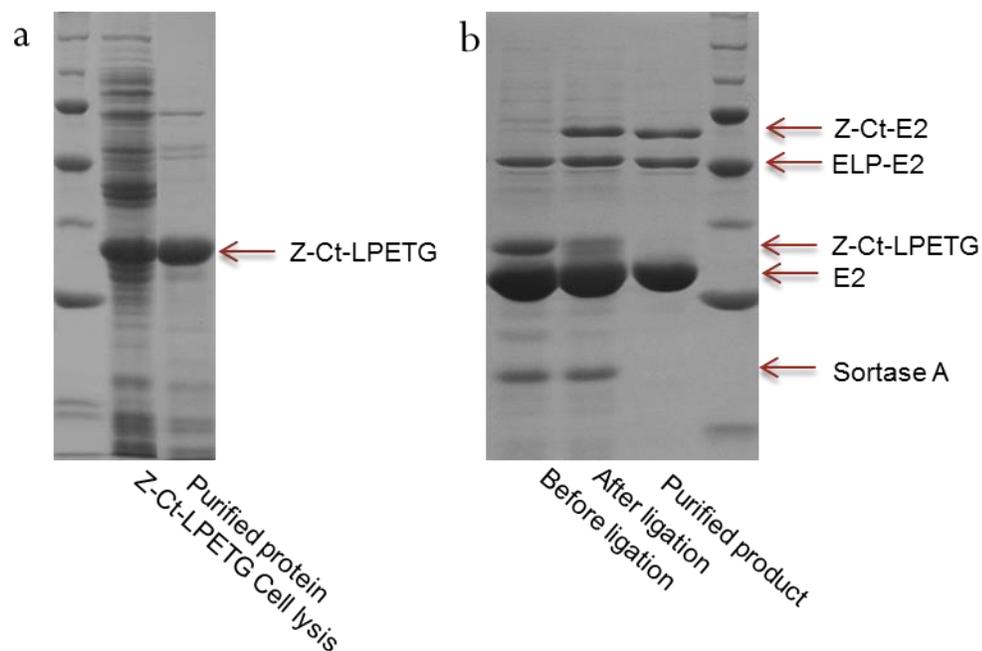


Figure 4.9 Formation of IgG-binding E2 nanoprobes. (A) Expression and purification of Z-Ct-LPETG. Protein was partially purified by heating whole cell lysates at 70°C for 10 min. Most cellular proteins denatured and precipitated and was removed by centrifugation. (B) Binding of rabbit IgG to Z-E2-ELP. Rabbit IgG was mixed with Z-E2-ELP and the bound complex was recovered by one cycle of thermal precipitation and resolubilization.

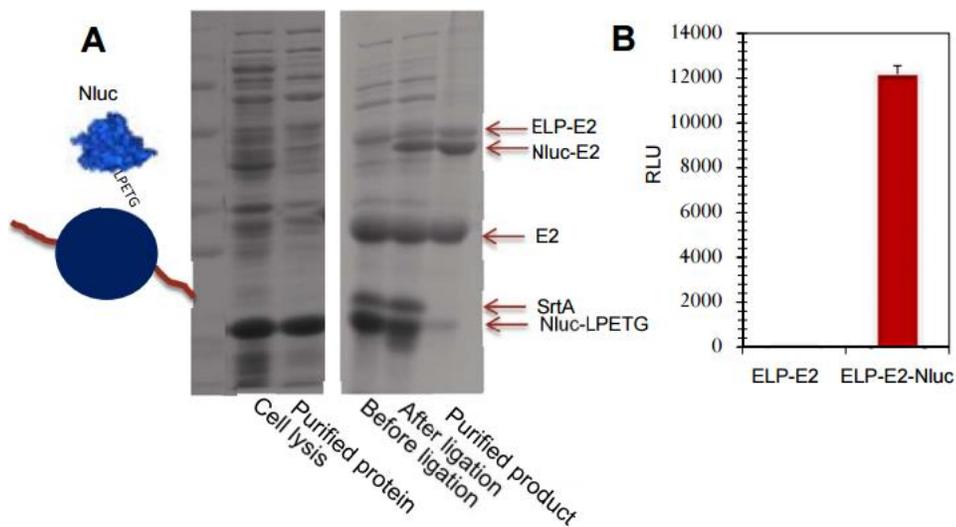


Figure 4.10 E2 nanoprobe assembly using the Z domain for antibody capture and Nluc for detection. (A) Nluc purification and ligation onto ELP-E2 nanoparticles. (B) Detection of Nluc activity for the purified ELP-E2-Nluc nanoparticles.

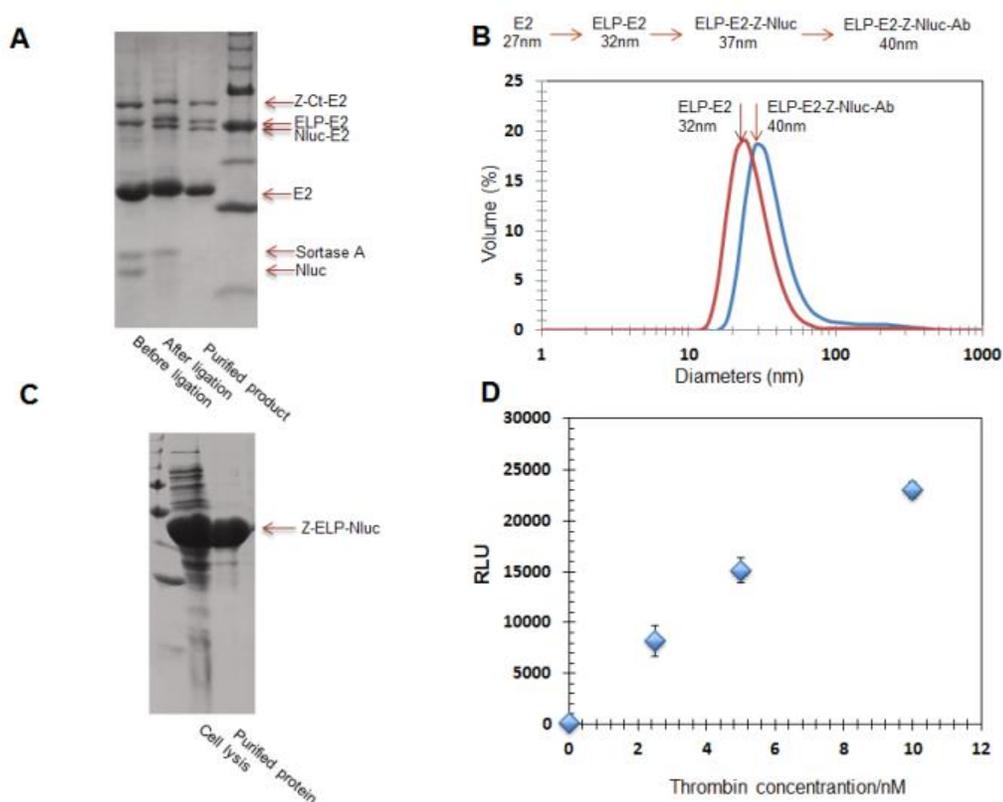


Figure 4.11 Two-step assembly of Z-ELP-E2-Nluc nanoprobe and their use for thrombin sensing. (A) Two-step SrtA-mediated ligation of Z-Ct and Nluc onto ELP-E2 nanoparticles. (B) Dynamic light scattering of ELP-E2 and IgG-Z-ELP-E2-Nluc nanoparticles. (C) Production and purification of Z-ELP-Nluc probes. (D) Thrombin detection using IgG-Z-ELP-E2-Nluc nanoprobe.

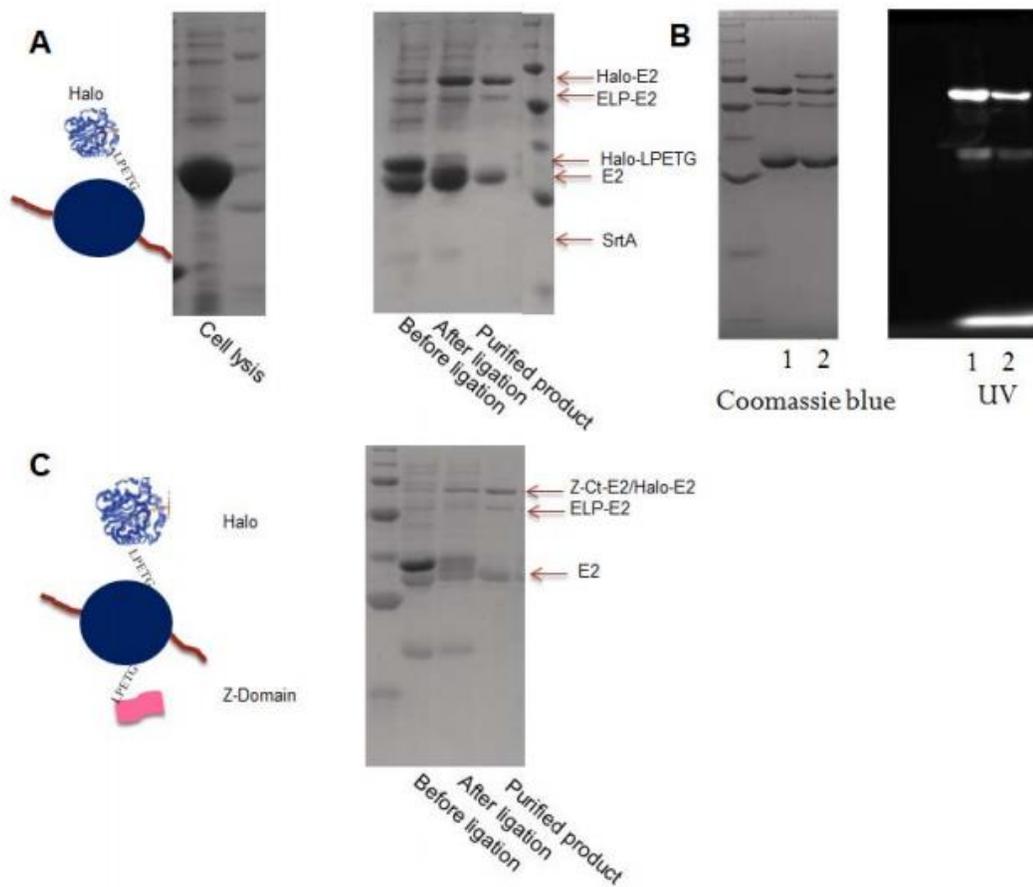


Figure 4.12 **Aptamer and Alexa 488 biosensor for tumor marker detection.** (A) Production and ligation of Halo tag onto E2 nanoparticles. (B) Conjugation of CH-Alexa 488 to either (1) Halo-ELP-E2 or (2) Aptamer-Halo-ELP-E2. Successful conjugation was verified by detecting brightly fluorescent bands under UV light. (C) Ligation of Z-Ct and Halo tag onto ELP-E2 nanoparticles.

Chapter 5

CONCLUSIONS

With all the advantages associated with natural and artificial protein co-localization, DNA and protein scaffolds were utilized to co-localize multiple proteins for cellulose hydrolysis and tumor marker detection. Various strategies were implemented to assess and to enhance the performance of these systems.

In the first section, we constructed artificial cellulosomes based on Zinc Finger Protein (ZFP) -guided assembly using DNA as a scaffold. The site-specific docking of CelA and CBM onto a single DNA template was achieved by exploiting the specificity of the two tethering ZFPs. The resulting two-component cellulosome structures exhibited 1.7-fold enhancement in cellulose hydrolysis compared to the non-complexed protein mixture and further enhancement was achieved by varying the number of CBMs and cellulases assembled. The modular nature of the design allows easy alteration of the number, spacing, and ordering of enzymes assembled, leading to virtually unlimited combinations of artificial cellulosome structures that can be optimized for any given target cellulosic substrate.

Although ZFPs, for the first time to our knowledge, provide the possibility of organizing artificial cellulosomes on DNA templates in position-specific manner, this design suffered from low assembly efficiency as a result of low binding affinity of ZFPs toward target DNA sequences. To enhance the performance of artificial cellulosomes, we replaced zinc finger protein with a HaloTag enzyme in the second chapter; the HaloTag enzyme enables covalent linkages of cellulosome components to single stranded linker DNAs, such that we can use DNA hybridization to assemble artificial cellulosome. The high-affinity interaction (typically in the range of 0.5nM) afforded by site-specific DNA

hybridization over cohesin-dockerin pairs greatly improves assembly efficiency and results in improved enzyme synergy of 2.5-fold enhancement of cellulose hydrolysis. With this HaloTag approach, we were able to construct cellulosomes with four necessary component (endoglucanase, exoglucanase, β -glucosidase and CBM) to release glucose monomers from cellulose. More importantly, rolling circle amplification were used to amplify the number of binding sites to increase enzyme density on a single scaffold structure. The resulting 5-fold enhancement in glucose release using the complex RCA-derived cellulosome structure is the highest synergy reported so far. With this novel strategy of covalently attaching single stranded DNA linkers onto cellulosome components, more complex DNA scaffolds with additional hybridization sites for immobilizing a large array of cellulases to better mimic the enzyme diversity of naturally occurring cellulosomes or higher order two- or three-dimensional DNA scaffolds structures should enable the generation of new cellulosome architecture for improved cellulose processing.

Other than increasing enzyme complex efficiency, multi-functional proteins were co-localized to achieve biosensing functions in the last section of this thesis. A modular approach was utilized to develop highly sensitive, modular and easy to purify nanoprobe by decorating E2 protein nanoparticles with a wide range of sensing and detection functionalities using SrtA-mediated ligation. First, an easy purification scheme was created by post-translationally ligating Elastin-Like Polypeptide (ELP) onto E2 protein nanoparticle surface, thereby enabling the E2 particles to be purified by going through ELP-mediated thermal cycle purification. Second, modularity was built into our system. Biosensors need input modules to recognize analytes and output modules to reflect the analyte recognition events. The use of Z domains as input

modules not only allows directional immobilization of antibody, but also allows any commercially available antibodies for unlimited sensing specificities. Similarly, an unlimited number of target-specific DNA aptamers can be conjugated to the nanoprobe as input modules using the HaloTag. This flexibility is extended beyond just the sensing module; our approach also provides the feasibility to use either enzymes or fluorescent dyes as output modules. In addition to the input and output modules reported, many other examples such as single chain variable fragments, fibronectin type III domains, and fluorescence proteins can also be ligated onto the E2 nanoprobe. Even inorganic gold nanoparticles and quantum dots can be easily coupled to our nanoprobe by employing specific metal binding peptides to further expand the range of applications of this modular platform. Thirdly, the sensitivity of these biosensors can be fine-tuned with ease by controlling the number of enzymes or dyes conjugated onto the nanoprobe, and the use of more than one output module is also possible because of the modular nature of this approach. Overall, this biosensor platform offers easy purification, signal amplification, and modularity on protein scaffolds with post-translational modifications.

Future research opportunities

So far we have used biomolecular scaffolds to co-localize proteins for cellulose hydrolysis and tumor marker detection. In the future, we can either optimize the existing systems to further enhance performance or employ biomolecular scaffolds to organize other proteins to achieve novel functionalities.

1. For artificial cellulosome assembled on DNA templates, we were able to organize all four cellulosome components necessary for complete cellulose hydrolysis. With the ease to synthesis DNA templates, we should be able to employ longer DNA

templates with more binding sites to incorporate more diverse cellulases and hemi-cellulases to mimic natural cellulosome. Also, DNA nanotechnology is available to create complex and predictable one-, two-, or three- dimensional structures with precise control over the distance of all the components. With our system, we should be able to create artificial cellulosome with complex structures to increase cellulose hydrolysis efficiency.

2. For enzyme complexes, DNA templates show promise in organizing enzymes in a position specific manner. Since DNA template is highly modular, this strategy may be used to enhance other sequential biochemical reactions beyond cellulose hydrolysis. A good potential system is the three dehydrogenases responsible for sequential conversion of methanol to carbon dioxide. The correct assembly of enzymes on protein scaffolds already resulted in more than 5-fold increase in product yield over that of unassembled enzymes⁴⁰, and DNA templates can potentially be used to organize and optimize this enzyme reaction cascade beyond what is currently achieved by protein scaffolds. The same hypothesis can be proposed on other enzyme systems.
3. For biosensor constructed from protein nanoparticles, we already tested antibody and aptamer as input modules and fluorescence dyes and nanoluciferase as output modules. To enhance the modularity of this system, more sensor components can be incorporated, including single chain variable fragments, fibronectin type III domains, fluorescence proteins, inorganic gold nanoparticles and quantum dots. These options could provide the modularity needed for real context-driven sensor assembly.

4. Since the bionanoparticle platform we constructed can be easily purified and is capable of incorporating diverse functional domains, we can easily extend this system for other applications. One of the potential application is targeted drug delivery. To achieve this, multiple components need to be immobilized onto our protein scaffolds: i) ELP as purification tag; ii) targeting motif to direct nanoparticles to diseased cells by recognizing and binding onto disease cell surface markers; iii) markers including fluorescence proteins or fluorescence dyes to indicate location of particles during the process; iv) drug carrying and releasing motif so drugs can be loaded *in vitro* and released upon getting inside diseased cells. With the high modularity of our platform, drug delivery is particularly applicable and should be carefully explored.

REFERENCES

- (1) Ouzounis, C. a; Coulson, R. M. R.; Enright, A. J.; Kunin, V.; Pereira-Leal, J. B. Classification Schemes for Protein Structure and Function. *Nat. Rev. Genet.* **2003**, *4*, 508–519.
- (2) Chen, R.; Chen, Q.; Kim, H.; Siu, K. H.; Sun, Q.; Tsai, S. L.; Chen, W. Biomolecular Scaffolds for Enhanced Signaling and Catalytic Efficiency. *Curr. Opin. Biotechnol.* **2014**, *28*, 59–68.
- (3) Huang, X.; Holden, H. M.; Raushel, F. M. Channeling of Substrates and Intermediates in Enzyme-Catalyzed Reactions. *New York* **2001**, *70*, 149–180.
- (4) Thoden, J. B.; Holden, H. M.; Wesenberg, G.; Raushel, F. M.; Rayment, I. Structure of Carbamoyl Phosphate Synthetase: A Journey of 96 ?? From Substrate to Product. *Biochemistry* **1997**, *36*, 6305–6316.
- (5) Aparicio, J. F.; Fouces, R.; Mendes, M. V.; Olivera, N.; Martín, J. F. A Complex Multienzyme System Encoded by Five Polyketide Synthase Genes Is Involved in the Biosynthesis of the 26-Membered Polyene Macrolide Pimaricin in *Streptomyces Natalensis*. *Chem. Biol.* **2000**, *7*, 895–905.
- (6) Miles, E. W. The Tryptophan Synthase ?????? Complex: A Model for Substrate Channeling, Allosteric Communication, and Pyridoxal Phosphate Catalysis. *J. Biol. Chem.* **2013**, *288*, 10084–10091.
- (7) Good, M. C. Scaffold Proteins : Hubs for Controlling the Flow of Cellular Information. *Science (80-.)*. **2011**, *332*, 680–686.
- (8) Meister, M.; Tomasovic, A.; Banning, A.; Tikkanen, R. Mitogen-Activated Protein (MAP) Kinase Scaffolding Proteins: A Recount. *Int. J. Mol. Sci.* **2013**, *14*, 4854–4884.
- (9) Kolch, W. Coordinating ERK/MAPK Signalling through Scaffolds and Inhibitors. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 827–837.
- (10) Tanaka, N. Technology Roadmap - Biofuels for Transport. *Int. Energy Agency* **2011**, 1–56.
- (11) Rajagopal, D.; Zilberman, D. Review of Environmental , Economic and Policy Aspects of Biofuels. *Polyicy Res. Work. Pap.* **2007**, 109.

- (12) Tsai, S.; Dasilva, N. A.; Chen, W. Functional Display of Complex Cellulosomes on the Yeast Surface via Adaptive Assembly. *ACS Synth. Biol.* **2013**, *2*, 14–21.
- (13) Lynd, L. R.; Laser, M. S.; Bransby, D.; Dale, B. E.; Davison, B.; Hamilton, R.; Himmel, M.; Keller, M.; McMillan, J. D.; Sheehan, J.; *et al.* How Biotech Can Transform Biofuels. *Nat. Biotechnol.* **2008**, *26*, 169–172.
- (14) Bayer, E. a.; Lamed, R.; White, B. a.; Flints, H. J. From Cellulosomes to Cellulosomics. *Chem. Rec.* **2008**, *8*, 364–377.
- (15) Schwarz, W. H. The Cellulosome and Cellulose Degradation by Anaerobic Bacteria. *Appl. Microbiol. Biotechnol.* **2001**, *56*, 634–649.
- (16) Lynd, L. R.; Weimer, P. J.; Zyl, W. H. Van; Pretorius, I. S. Microbial Cellulose Utilization : Fundamentals and Biotechnology. *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 506–577.
- (17) Fontes, C. M. G. a; Gilbert, H. J. Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. *Annu. Rev. Biochem.* **2010**, *79*, 655–681.
- (18) Fierobe, H.-P.; Mingardon, F.; Mechaly, A.; Bélaïch, A.; Rincon, M. T.; Pagès, S.; Lamed, R.; Tardif, C.; Bélaïch, J.-P.; Bayer, E. a. Action of Designer Cellulosomes on Homogeneous versus Complex Substrates: Controlled Incorporation of Three Distinct Enzymes into a Defined Trifunctional Scaffoldin. *J. Biol. Chem.* **2005**, *280*, 16325–16334.
- (19) Tsai, S.; Oh, J.; Singh, S.; Chen, R.; Chen, W. Functional Assembly of Minicellulosomes on the *Saccharomyces Cerevisiae* Cell Surface for Cellulose Hydrolysis and Ethanol Production. *Appl. Environ. Microbiol.* **2009**, *75*, 6087–6093.
- (20) Sarah Moraïs, Ely Morag, Yoav Barak, Dan Goldman, Yitzhak Hadar, Raphael Lamed, Yuval Shoham, David B. Wilson, and E. A. B. Deconstruction of Lignocellulose into Soluble Sugars by Native and. *MBio* **2012**, *3*, 1–11.
- (21) Srikrishnan, S.; Chen, W.; Da Silva, N. a. Functional Assembly and Characterization of a Modular Xylanosome for Hemicellulose Hydrolysis in Yeast. *Biotechnol. Bioeng.* **2013**, *110*, 275–285.
- (22) Cancer Statistics - National Cancer Institute.

- (23) National Cancer Institute. What Is Cancer? - National Cancer Institute.
- (24) Tumor Markers - National Cancer Institute.
- (25) Duffy, M. J. Tumor Markers in Clinical Practice: A Review Focusing on Common Solid Cancers. *Med. Princ. Pract.* **2012**, *22*, 4–11.
- (26) Sharma, S. Tumor Markers in Clinical Practice: General Principles and Guidelines, 2009, *30*, 1–8.
- (27) Tran, T. N. N.; Cui, J.; Hartman, M. R.; Peng, S.; Funabashi, H.; Duan, F.; Yang, D.; March, J. C.; Lis, J. T.; Cui, H.; *et al.* A Universal DNA-Based Protein Detection System. *J. Am. Chem. Soc.* **2013**, *135*, 14008–14011.
- (28) Hwang, M. P.; Lee, J.-W.; Lee, K. E.; Lee, K. H. Think Modular: A Simple Apoferritin-Based Platform for the Multifaceted Detection of Pancreatic Cancer. *ACS Nano* **2013**, *7*, 8167–8174.
- (29) Xu, J.-J.; Zhao, W.-W.; Song, S.; Fan, C.; Chen, H.-Y. Functional Nanoprobes for Ultrasensitive Detection of Biomolecules: An Update. *Chem. Soc. Rev.* **2014**, *43*, 1601–1611.
- (30) Kamiya, N.; Doi, S.; Tanaka, Y.; Ichinose, H.; Goto, M. Functional Immobilization of Recombinant Alkaline Phosphatases Bearing a Glutamyl Donor Substrate Peptide of Microbial Transglutaminase. *J. Biosci. Bioeng.* **2007**, *104*, 195–199.
- (31) Domingo, G. J.; Orru, S.; Perham, R. N. Multiple Display of Peptides and Proteins on a Macromolecular Scaffold Derived from a Multienzyme Complex. *JMB* **2001**, 259–267.
- (32) Sletten, E. M.; Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angewandte Chemie - International Edition*, 2009, *48*, 6974–6998.
- (33) Sokic-Lazic, D.; Arechederra, R. L.; Treu, B. L.; Minteer, S. D. Oxidation of Biofuels: Fuel Diversity and Effectiveness of Fuel Oxidation through Multiple Enzyme Cascades. *Electroanalysis*, 2010, *22*, 757–764.
- (34) Kim, J.; Grate, J. W.; Wang, P. Nanostructures for Enzyme Stabilization. *Chem. Eng. Sci.* **2006**, *61*, 1017–1026.

- (35) Pinheiro, A. V.; Han, D.; Shih, W. M.; Yan, H. Challenges and Opportunities for Structural DNA Nanotechnology. *Nat. Nanotechnol.* **2011**, *6*, 763–772.
- (36) Wilner, O. I.; Shimron, S.; Weizmann, Y.; Wang, Z.; Willner, I. Self-Assembly of Enzymes on DNA Scaffolds : En Route to Biocatalytic Nanowires. *Nano Lett.* **2009**, *9*, 2040–2043.
- (37) Wilner, O. I.; Weizmann, Y.; Gill, R.; Lioubashevski, O.; Freeman, R.; Willner, I. Enzyme Cascades Activated on Topologically Programmed DNA Scaffolds. *Nat. Nanotechnol.* **2009**, *4*, 249–254.
- (38) Fu, J.; Liu, M.; Liu, Y.; Woodbury, N. W.; Yan, H. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *J. Am. Chem. Soc.* **2012**, *134*, 5516–5519.
- (39) Doi, R. H.; Kosugi, A. Cellulosomes: Plant-Cell-Wall-Degrading Enzyme Complexes. *Nat. Rev. Microbiol.* **2004**, *2*, 541–551.
- (40) Liu, F.; Banta, S.; Chen, W. Functional Assembly of a Multi-Enzyme Methanol Oxidation Cascade on a Surface-Displayed Trifunctional Scaffold for Enhanced NADH Production. *Chem. Commun. (Camb).* **2013**, *49*, 3766–3768.
- (41) You, C.; Myung, S.; Zhang, Y.-H. P. Facilitated Substrate Channeling in a Self-Assembled Trifunctional Enzyme Complex. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 8787–8790.
- (42) Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D. Synthetic Protein Scaffolds Provide Modular Control over Metabolic Flux. *Nat. Biotechnol.* **2009**, *27*, 753–759.
- (43) Tsai, S.-L.; Goyal, G.; Chen, W. Surface Display of a Functional Minicellulosome by Intracellular Complementation Using a Synthetic Yeast Consortium and Its Application to Cellulose Hydrolysis and Ethanol Production. *Appl. Environ. Microbiol.* **2010**, *76*, 7514–7520.
- (44) Sandman, K.; Pereira, S. L.; Reeve, J. N. Diversity of Prokaryotic Chromosomal Proteins and the Origin of the Nucleosome. *Cell. Mol. Life Sci.* **1998**, *54*, 1350–1364.
- (45) Kadonaga, J. T. Regulation of RNA Polymerase II Transcription by Sequence-Specific DNA Binding Factors. *Cell* **2004**, *116*, 247–257.

- (46) Laity, J. H.; Lee, B. M.; Wright, P. E. Zinc Finger Proteins : New Insights into Structural and Functional Diversity. *Curr. Opin. Struct. Biol.* **2001**, *11*, 39–46.
- (47) Choo, Y.; Isalan, M. Advances in Zinc Finger Engineering. *Curr. Opin. Struct. Biol.* **2000**, *10*, 411–416.
- (48) Rebar, E.; Miller, J. Design and Applications of Engineered Zinc Finger Proteins. *Bio. Tech. Int.* **2004**, 1–4.
- (49) Tan, S.; Guschin, D.; Davalos, A.; Lee, Y.-L.; Snowden, A. W.; Jouvenot, Y.; Zhang, H. S.; Howes, K.; McNamara, A. R.; Lai, A.; *et al.* Zinc-Finger Protein-Targeted Gene Regulation: Genomewide Single-Gene Specificity. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 11997–12002.
- (50) Stains, C. I.; Porter, J. R.; Ooi, A. T.; Segal, D. J.; Ghosh, I. DNA Sequence-Enabled Reassembly of the Green Fluorescent Protein. *J. Am. Chem. Soc.* **2005**, *127*, 10782–10783.
- (51) Conrado, R. J.; Wu, G. C.; Boock, J. T.; Xu, H.; Chen, S. Y.; Avbelj, M.; Gaber, R.; Lebar, T.; Turns, J.; Hodnik, V.; *et al.* DNA-Guided Assembly of Biosynthetic Pathways Promotes Improved Catalytic Efficiency. *Nucleic Acids Res.* **2012**, *40*, 1879–1889.
- (52) Lee, J. H.; Jung, S.; Bui, L. M.; Kang, K. H.; Song, J.; Kim, S. C. Improved Production of L -Threonine in Escherichia Coli by Use of a DNA Scaffold System. *Appl. Environ. Microbiol.* **2013**, *79*, 774–782.
- (53) Kukol, A.; Li, P.; Estrela, P.; Ko-Ferrigno, P.; Migliorato, P. Label-Free Electrical Detection of DNA Hybridization for the Example of Influenza Virus Gene Sequences. *Anal. Biochem.* **2008**, *374*, 143–153.
- (54) Los, G. V.; Encell, L. P.; Mcdougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; *et al.* Halo Tag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **2008**, *3*, 373–382.
- (55) N. Peterson, S. The HaloTag: Improving Soluble Expression and Applications in Protein Functional Analysis. *Curr. Chem. Genomics* **2012**, *6*, 8–17.
- (56) Blackstock, D.; Chen, W. Halo-Tag Mediated Self-Labeling of Fluorescent Proteins to Molecular Beacons for Nucleic Acid Detection. *Chem. Commun.* **2014**, *50*, 13735–13738.

- (57) Hess, G. T.; Cragolini, J. J.; Popp, M. W.; Allen, M. A.; Dougan, S. K.; Spooner, E.; Ploegh, H. L.; Belcher, A. M.; Guimaraes, C. P. M13 Bacteriophage Display Framework That Allows Sortase-Mediated Modification of Surface-Accessible Phage Proteins. *Bioconjug. Chem.* **2012**, *23*, 1478–1487.
- (58) Parolo, C.; Escosura-mun, A. De; Polo, E.; Grazu, V.; Fuente, M. De. Design, Preparation, and Evaluation of a Fixed-Orientation Antibody / Gold-Nanoparticle Conjugate as an Immunosensing Label. *ACS Appl. Mater. Interfaces* **2013**, *5*, 10753–10759.
- (59) Sun, Q.; Madan, B.; Tsai, S.-L.; DeLisa, M. P.; Chen, W. Creation of Artificial Cellulosomes on DNA Scaffolds by Zinc Finger Protein-Guided Assembly for Efficient Cellulose Hydrolysis. *Chem. Commun. (Camb)*. **2014**, *50*, 1423–1425.
- (60) Blackstock, D.; Sun, Q.; Chen, W. Fluorescent Protein-Based Molecular Beacons by Zinc Finger Protein-Guided Assembly. *Biotechnol. Bioeng.* **2015**, *112*, 236–241.
- (61) Patel, K. G.; Swartz, J. R. Surface Functionalization of Virus-like Particles by Direct Conjugation Using Azide-Alkyne Click Chemistry. *Bioconjug. Chem.* **2011**, *22*, 376–387.
- (62) Ton-That, H.; Mazmanian, S. K.; Alksne, L.; Schneewind, O. Anchoring of Surface Proteins to the Cell Wall of *Staphylococcus Aureus*. Cysteine 184 and Histidine 120 of Sortase Form a Thiolate-Imidazolium Ion Pair for Catalysis. *J. Biol. Chem.* **2002**, *277*, 7447–7452.
- (63) Proft, T. Sortase-Mediated Protein Ligation: An Emerging Biotechnology Tool for Protein Modification and Immobilisation. *Biotechnol. Lett.* **2010**, *32*, 1–10.
- (64) Levary, D. a; Parthasarathy, R.; Boder, E. T.; Ackerman, M. E. Protein-Protein Fusion Catalyzed by Sortase A. *PLoS One* **2011**, *6*, e18342.
- (65) Liu, F.; Tsai, S.-L.; Madan, B.; Chen, W. Engineering a High-Affinity Scaffold for Non-Chromatographic Protein Purification via Intein-Mediated Cleavage. *Biotechnol. Bioeng.* **2012**, *109*, 2829–2835.
- (66) Terpe, K. Overview of Tag Protein Fusions : From Molecular and Biochemical Fundamentals to Commercial Systems. *Appl. Microbiol. Biotechnol.* **2003**, *60*, 523–533.

- (67) Saraswat, M.; Musante, L.; Ravidá, A.; Shortt, B.; Byrne, B.; Holthofer, H. Preparative Purification of Recombinant Proteins : Current Status and Future Trends. *Biomed Res. Int.* **2013**, *2013*, 1–18.
- (68) Urry, D. W. Physical Chemistry of Biological Free Energy Transduction As Demonstrated by Elastic Protein-Based Polymers. *J. Phys. Chem. B* **1997**, *101*, 11007–11028.
- (69) MacEwan, S. R.; Chilkoti, A. Elastin-like Polypeptides: Biomedical Applications of Tunable Biopolymers. *Biopolymers* **2010**, *94*, 60–77.
- (70) Meyer, D. E.; Trabbic-Carlson, K.; Chilkoti, a. Protein Purification by Fusion with an Environmentally Responsive Elastin-like Polypeptide: Effect of Polypeptide Length on the Purification of Thioredoxin. *Biotechnol. Prog.* **2001**, *17*, 720–728.
- (71) Madan, B.; Chaudhary, G.; Cramer, S. M.; Chen, W. ELP-Z and ELP-Zz Capturing Scaffolds for the Purification of Immunoglobulins by Affinity Precipitation. *J. Biotechnol.* **2013**, *163*, 10–16.
- (72) Shimazu, M.; Mulchandani, A.; Chen, W. Thermally Triggered Purification and Immobilization of Elastin – OPH Fusions. *Biotechnol. Bioeng.* **2003**, *81*, 74–79.
- (73) Kostal, J.; Mulchandani, A.; Chen, W. Affinity Purification of Plasmid DNA by Temperature-Triggered Precipitation. *Nat. Protoc.* **2007**, *2*, 1263–1268.
- (74) Moraís, S.; Morag, E.; Barak, Y. Deconstruction of Lignocellulose into Soluble Sugars by Native and. *MBio* **2012**, *3*, 1–11.
- (75) Niemeyer, C. M.; Koehler, J. DNA-Directed Assembly of Bienzymic Complexes from In Vivo Biotinylated NAD (P) H : FMN Oxidoreductase and Luciferase. *ChemBioChem* **2002**, *3*, 242–245.
- (76) He, Y.; Ye, T.; Su, M.; Zhang, C.; Ribbe, A. E.; Jiang, W.; Mao, C. Hierarchical Self-Assembly of DNA into Symmetric Supramolecular Polyhedra. *Nature* **2008**, *452*, 198–201.
- (77) Mori, Y.; Ozasa, S.; Kitaoka, M.; Noda, S.; Tanaka, T.; Ichinose, H.; Kamiya, N. Aligning an Endoglucanase Cel5A from *Thermobifida Fusca* on a DNA Scaffold: Potent Design of an Artificial Cellulosome. *Chem. Commun. (Camb)*. **2013**, *49*, 6971–6973.

- (78) Negi, S.; Imanishi, M.; Matsumoto, M.; Sugiura, Y. New Redesigned Zinc-Finger Proteins: Design Strategy and Its Application. *Chemistry* **2008**, *14*, 3236–3249.
- (79) Liu, Q.; Xia, Z.; Case, C. C. Validated Zinc Finger Protein Designs for All 16 GNN DNA Triplet Targets. *J. Biol. Chem.* **2002**, *277*, 3850–3856.
- (80) Ooi, A. T.; Stains, C. I.; Ghosh, I.; Segal, D. J. Sequence-Enabled Reassembly of Beta-Lactamase (SEER-LAC): A Sensitive Method for the Detection of Double-Stranded DNA. *Biochemistry* **2006**, *45*, 3620–3625.
- (81) Tsai, S.-L.; Oh, J.; Singh, S.; Chen, R.; Chen, W. Functional Assembly of Minicellulosomes on the *Saccharomyces Cerevisiae* Cell Surface for Cellulose Hydrolysis and Ethanol Production. *Appl. Environ. Microbiol.* **2009**, *75*, 6087–6093.
- (82) Elrod-erickson, M.; Rould, M. A.; Nekludova, L.; Pabo, C. O. Zif268 Protein – DNA Complex Refined at 1.6 Å : A Model System for Understanding Zinc Finger – DNA Interactions. *Structure* **1996**, *4*, 1171–1180.
- (83) Segal, D. J. The Use of Zinc Finger Peptides to Study the Role of Specific Factor Binding Sites in the Chromatin Environment. *Methods* **2002**, *26*, 76–83.
- (84) Bird, A. J.; Mccall, K.; Kramer, M.; Blankman, E.; Winge, D. R.; Eide, D. J. Zinc Fingers Can Act as Zn²⁺ Sensors to Regulate Transcriptional Activation Domain Function. *EMBO J.* **2003**, *22*, 5137–5146.
- (85) Lim, D. W.; Trabbic-carlson, K.; Mackay, J. A.; Chilkoti, A. Improved Non-Chromatographic Purification of a Recombinant Protein by Cationic Elastin-like Polypeptides. *Biomacromolecules* **2007**, 1417–1424.
- (86) Hellman, L. M.; Fried, M. G. Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein-Nucleic Acid Interactions. *Nat. Protoc.* **2007**, *2*, 1849–1861.
- (87) Morais, S.; Barak, Y.; Caspi, J.; Hadar, Y.; Lamed, R.; Shoham, Y.; Wilson, D. B. Cellulase-Xylanase Synergy in Designer Cellulosomes for Enhanced Degradation of a Complex Cellulosic Substrate. *MBio* **2010**, *1*, 3–10.
- (88) Mechaly, A.; Fierobe, H. P.; Belaich, A.; Belaich, J. P.; Lamed, R.; Shoham, Y.; Bayer, E. a. Cohesin-Dockerin Interaction in Cellulosome Assembly: A Single Hydroxyl Group of a Dockerin Domain Distinguishes between Nonrecognition and High Affinity Recognition. *J. Biol. Chem.* **2001**, *276*, 9883–9888.

- (89) Scaffoldin, T.; Fierobe, H.; Mingardon, F.; Mechaly, A.; Be, A.; Rincon, M. T.; Lamed, R.; Tardif, C.; Be, J.; Bayer, E. A. Action of Designer Cellulosomes on Homogeneous versus Complex Substrates: Controlled Incorporation of Three Distinct Enzymes into a Defined Trifunctional Scaffoldin. *J. Biol. Chem.* **2005**, *280*, 16325–16334.
- (90) Miles, E. W.; Rhee, S.; Davies, D. R. The Molecular Basis of Substrate Channeling. *J. Biol. Chem.* **1999**, *274*, 12193–12196.
- (91) Stern, J.; Kahn, A.; Vazana, Y.; Shamsoum, M.; Morais, S.; Lamed, R.; Bayer, E. A. Significance of Relative Position of Cellulases in Designer Cellulosomes for Optimized Cellulolysis. *PLoS One* **2015**, *10*, e0127326.
- (92) Caspi, J.; Barak, Y.; Haimovitz, R.; Irwin, D.; Lamed, R.; Wilson, D. B.; Bayer, E. A. Effect of Linker Length and Dockerin Position on Conversion of a *Thermobifida Fusca* Endoglucanase to the Cellulosomal Mode. *Appl. Environ. Microbiol.* **2009**, *75*, 7335–7342.
- (93) Tsai, S. L.; Park, M.; Chen, W. Size-Modulated Synergy of Cellulase Clustering for Enhanced Cellulose Hydrolysis. *Biotechnol. J.* **2013**, *8*, 257–261.
- (94) Ali, M. M.; Li, F.; Zhang, Z.; Zhang, K.; Kang, D.-K.; Ankrum, J. a; Le, X. C.; Zhao, W. Rolling Circle Amplification: A Versatile Tool for Chemical Biology, Materials Science and Medicine. *Chem. Soc. Rev.* **2014**, *43*, 3324–3341.
- (95) Online, V. A.; Mori, Y.; Ozasa, S.; Kitaoka, M.; Noda, S.; Tanaka, T.; Ichinose, H.; Kamiya, N. Aligning an Endoglucanase Cel5A from *Thermobifida Fusca* on a DNA Scaffold: Potent Design of an Artificial Cellulosome. *Chem. Commun. (Camb)*. **2013**, *49*, 6971–6973.
- (96) Kim, D.-M.; Nakazawa, H.; Umetsu, M.; Matsuyama, T.; Ishida, N.; Ikeuchi, A.; Takahashi, H.; Asano, R.; Kumagai, I. A Nanocluster Design for the Construction of Artificial Cellulosomes. *Catal. Sci. Technol.* **2012**, *2*, 499–503.
- (97) Lazcka, O.; Del Campo, F. J.; Muñoz, F. X. Pathogen Detection: A Perspective of Traditional Methods and Biosensors. *Biosens. Bioelectron.* **2007**, *22*, 1205–1217.
- (98) Byrne, B.; Stack, E.; Gilmartin, N.; Kennedy, R. O. Antibody-Based Sensors: Principles, Problems and Potential for Detection of Pathogens and Associated Toxins. *Sensors* **2009**, *9*, 4407–4445.

- (99) Rucker, V. C.; Havenstrite, K. L.; Herr, A. E. Antibody Microarrays for Native Toxin Detection. *Anal. Biochem.* **2005**, *339*, 262–270.
- (100) Bianco, A.; Yasumoto, T.; Prato, M.; Poli, M.; Tubaro, A. Sandwich ELISA Assay for the Quantitation of Palytoxin and Its Analogs in Natural Samples. *Environ. Sci. Technol.* **2013**, *47*, 2034–2042.
- (101) Cella, L. N.; Sanchez, P.; Zhong, W.; Myung, N. V.; Chen, W.; Mulchandani, A. Nano Aptasensor for Protective Antigen Toxin of Anthrax. *Anal. Chem.* **2010**, *82*, 2042–2047.
- (102) Xiang, Y.; Lu, Y. Using Personal Glucose Meters and Functional DNA Sensors to Quantify a Variety of Analytical Targets. *Nat. Chem.* **2011**, *3*, 697–703.
- (103) Menon, S.; Manning, B. Nutrient Sensing Lost in Cancer. *Nature* **2013**, *498*, 444–445.
- (104) Tong, S.; Ren, B.; Zheng, Z.; Shen, H.; Bao, G.; Engineering, B.; States, U. Tiny Grains Give Huge Gains : Nanocrystal-Based Signal Amplification for Biomolecule. *ACS Nano* **2013**, *7*, 5142–5150.
- (105) Su, J.; Xu, J.; Chen, Y.; Xiang, Y.; Yuan, R.; Chai, Y. Personal Glucose Sensor for Point-of-Care Early Cancer Diagnosis. *Chem. Commun. (Camb)*. **2012**, *48*, 6909–6911.
- (106) Wenger, M. D.; Dephillips, P.; Price, C. E.; Bracewell, D. G. An Automated Microscale Chromatographic Purification of Virus-like Particles as a Strategy for Process Development. *Biotechnol. Appl. Biochem.* **2007**, *47*, 131–139.
- (107) Vicente, T.; Roldão, A.; Peixoto, C.; Carrondo, M. J. T.; Alves, P. M. Large-Scale Production and Purification of VLP-Based Vaccines. *J. Invertebr. Pathol.* **2011**, *107*, S42–S48.
- (108) Swee, L. K.; Guimaraes, C. P.; Sehrawat, S.; Spooner, E.; Barrasa, M. I.; Ploegh, H. L. Sortase-Mediated Modification of α DEC205 Affords Optimization of Antigen Presentation and Immunization against a Set of Viral Epitopes. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *110*, 1428–11433.
- (109) Milne, J. L. S.; Shi, D.; Rosenthal, P. B.; Sunshine, J. S.; Domingo, G. J.; Wu, X.; Brooks, B. R.; Perham, R. N.; Henderson, R.; Subramaniam, S. Molecular Architecture and Mechanism of an Icosahedral Pyruvate Dehydrogenase Complex: A Multifunctional Catalytic Machine. *EMBO J.* **2002**, *21*, 5587–5598.

- (110) Izard, T.; Aevansson, A.; Allen, M. D.; Westphal, A. H.; Perham, R. N.; de Kok, A.; Hol, W. G. Principles of Quasi-Equivalence and Euclidean Geometry Govern the Assembly of Cubic and Dodecahedral Cores of Pyruvate Dehydrogenase Complexes. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 1240–1245.
- (111) Mizuno, T.; Murao, K.; Tanabe, Y.; Oda, M.; Tanaka, T. Metal-Ion-Dependent GFP Emission in Vivo by Combining a Circularly Permutated Green Fluorescent Protein with an Engineered Metal-Ion-Binding Coiled-Coil. *J. Am. Chem. Soc.* **2007**, *129*, 11378–11383.
- (112) Chen, Q.; Sun, Q.; Molino, N. M.; Wang, S.-W.; Boder, E.; Chen, W. Sortase A-Mediated Multi-Functionalization of Protein Nanoparticles. *Chem. Commun.* **2015**, *51*, 12107–12110.
- (113) Braisted, A. C.; Wells, J. A. Minimizing Binding Protein from Protein A. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 5688–5692.
- (114) Hall, M. P.; Unch, J.; Binkowski, B. F.; Valley, M. P.; Butler, B. L.; Wood, M. G.; Otto, P.; Zimmerman, K.; Vidugiris, G.; Machleidt, T.; *et al.* Engineered Luciferase Reporter from a Deep Sea Shrimp Utilizing a Novel Imidazopyrazinone Substrate. *ACS Chem. Biol.* **2012**, *16*, 1848–1857.
- (115) Wang, X.; Zhao, Q. A Fluorescent Sandwich Assay for Thrombin Using Aptamer Modified Magnetic Beads and Quantum Dots. *Microchim. Acta* **2012**, *178*, 349–355.
- (116) Choi, D.; Jeong, H.; Kim, K. Electrochemical Thrombin Detection Based on the Direct Interaction of Target Proteins and Graphene Oxide as an Indicator. *Analyst* **2014**, *139*, 1331–1333.
- (117) Shimada, J.; Maruyama, T.; Kitaoka, M.; Kamiya, N.; Goto, M. Microplate Assay for Aptamer-Based Thrombin Detection Using a DNA-Enzyme Conjugate Based on Histidine-Tag Chemistry. *Anal. Biochem.* **2012**, *421*, 541–546.
- (118) Mcauley, J. L.; Linden, S. K.; Png, C. W.; King, R. M.; Pennington, H. L.; Gendler, S. J.; Florin, T. H.; Hill, G. R.; Korolik, V.; McGuckin, M. A. MUC1 Cell Surface Mucin Is a Critical Element of the Mucosal Barrier to Infection. *J. Clin. Invest.* **2007**, *117*, 2313–2324.

- (119) Ferreira, C. S. M.; Matthews, C. S.; Missailidis, S. DNA Aptamers That Bind to MUC1 Tumour Marker: Design and Characterization of MUC1-Binding Single-Stranded DNA Aptamers. *Tumour Biol.* **2006**, *27*, 289–301.
- (120) Painter, J. T.; Clayton, N. P.; Herbert, R. a. Useful Immunohistochemical Markers of Tumor Differentiation. *Toxicol. Pathol.* **2010**, *38*, 131–141.
- (121) Deng, B.; Lin, Y.; Wang, C.; Li, F.; Wang, Z.; Zhang, H.; Li, X. F.; Le, X. C. Aptamer Binding Assays for Proteins: The Thrombin Example-A Review. *Anal. Chim. Acta* **2014**, *837*, 1–15.
- (122) Germer, K.; Leonard, M.; Zhang, X. RNA Aptamers and Their Therapeutic and Diagnostic Applications. *Int. J. Biochem. Mol. Biol.* **2013**, *4*, 27–40.
- (123) Li, Y.; Cu, Y. T. H.; Luo, D. Multiplexed Detection of Pathogen DNA with DNA-Based Fluorescence Nanobarcodes. *Nat. Biotechnol.* **2005**, *23*, 885–889.

Appendix A

LIST OF ABBREVIATIONS

ZFP: zinc finger protein

CBM: cellulose binding module

MAPK: mitogen-activated protein kinases

GOx: glucose oxidase

HRP: horseradish peroxidase

TCS: tow-component system,

HKs: histidine kinases

RRs: response regulators

CCMV: Cowpea Chlorotic Mottle virus

QDs: quantum dots

Bp: base pair

GFP: green fluorescence protein

SrtA: Sortase A

AP: alkaline phosphatase

Luc: luciferase

GOD: glucose oxidase

MBP: maltose binding protein

GST: glutathione S-transferase

ELP: elastin-like polypeptide

NFOR: NAD(P)H:FMN oxidoreductase

Z-QG: benzyloxycarbonyl-L-glutaminyglycine

EMSA: Electrophoretic mobility shift assay

E-Z: ELP-Zif268

E-A: ELP-PE1A

CelA-E-Z: CelA-ELP-Zif268

CBM-E-A: CBM-ELP-PE1A

RCA: rolling circle amplification

CelA-E-H: CelA-ELP-HaloTag

CelA-E-H-1: CelA-ELP-HaloTag-linker 1

CBM-E-H: CBM-ELP-HaloTag

CBM-E-H-2: CBM-ELP-HaloTag-linker 2

CelE-E-H: CelE-ELP-HaloTag

CelE-E-H-3: CelE-ELP-HaloTag-linker 3

BglA-E-H: BglA-ELP-HaloTag

BglA-E-H-4: BglA-ELP-HaloTag-linker 4

IgGs: Immunoglobulin G

Ni-NTA: nickel-nitriolotriacetic acid

NLuc: nanoluciferase

TEM: Transmission electron micrograph

CH: chlorohexane

PCR: polymerase chain reaction

Appendix B

REPRINT PERMISSIONS

Reprint permission from The Royal Society of Chemistry.

Qing Sun, Bhawna Madan, Shen-Long Tsai, Matthew P. DeLisa and Wilfred Chen, Creation of artificial cellulosomes on DNA scaffolds by zinc finger protein-guided assembly for efficient cellulose hydrolysis, *Chemical Communication*, 50, 1423-1425, 2014.

Author reusing their own work published by the Royal Society of Chemistry

You do not need to request permission to reuse your own figures, diagrams, etc, that were originally published in a Royal Society of Chemistry publication. However, permission should be requested for use of the whole article or chapter except if reusing it in a thesis. If you are including an article or book chapter published by us in your thesis please ensure that your co-authors are aware of this.

Reuse of material that was published originally by the Royal Society of Chemistry must be accompanied by the appropriate acknowledgement of the publication. The form of the acknowledgement is dependent on the journal in which it was published originally, as detailed in 'Acknowledgements'.

Reprint permission from American Chemical Society.

Qing Sun, Qi Chen, Daniel Blackstock and Wilfred Chen, Post-translational modification of bio-nanoparticles as a modular platform for biosensor assembly, *ACS Nano*, 9, 8554-8561, 2015.

Title: Post-Translational Modification of Bionanoparticles as a Modular Platform for Biosensor Assembly
Author: Qing Sun, Qi Chen, Daniel Blackstock, et al
Publication: ACS Nano
Publisher: American Chemical Society
Date: Aug 1, 2015
Copyright © 2015, American Chemical Society

Logged in as:

Qing Sun

LOGOUT

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

Reprint permission from Elsevier.

Rebecca Chen, Qi Chen, Heejae Kim, Ka-Hei Siu, Qing Sun, Shen-Long Tsai and Wilfred Chen, Biomolecular Scaffolds for Enhanced Signaling and Catalytic Efficiency, *Current Opinion in Biotechnology*, 28, 59-68, 2014.

9/25/2015

Rightslink® by Copyright Clearance Center



RightsLink®

Home

Account Info

Help



Title: Biomolecular scaffolds for enhanced signaling and catalytic efficiency

Author: Rebecca Chen, Qi Chen, Heejae Kim, Ka-Hei Siu, Qing Sun, Shen-Long Tsai, Wilfred Chen

Publication: Current Opinion in Biotechnology

Publisher: Elsevier

Date: August 2014

Logged in as:

Qing Sun

Account #:

3000955256

LOGOUT

Copyright © 2013 Elsevier Ltd. All rights reserved.

Order Completed

Thank you very much for your order.

This is a License Agreement between Qing Sun ("You") and Elsevier ("Elsevier"). The license consists of your order details, the terms and conditions provided by Elsevier, and the [payment terms and conditions](#).

License Number	3716231008446
License date	Sep 25, 2015
Licensed content publisher	Elsevier
Licensed content publication	Current Opinion in Biotechnology
Licensed content title	Biomolecular scaffolds for enhanced signaling and catalytic efficiency
Licensed content author	Rebecca Chen, Qi Chen, Heejae Kim, Ka-Hei Siu, Qing Sun, Shen-Long Tsai, Wilfred Chen
Licensed content date	August 2014
Licensed content volume number	28
Licensed content issue number	n/a
Number of pages	10
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Biomolecular scaffolds for enhanced biomass processing and tumor marker detection
Expected completion date	Oct 2015
Estimated size (number of pages)	120
Elsevier VAT number	GB 494 6272 12
Customer Tax ID	USDelaware
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD

ORDER MORE...

CLOSE WINDOW