Synthesis of Organic Linkers for Studying Biomolecular Interactions, Site-Specific Chemical Modification of Peptides and its Translocation Studies Through Nanopore

by

Sudipta Biswas

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2016 by the
Graduate Supervisory Committee:

Stuart Lindsay, Co-Chair
Peiming Zhang, Co-Chair
Ian Gould
Kevin Redding

ARIZONA STATE UNIVERSITY

May 2016

ABSTRACT

Biomolecules can easily recognize its corresponding partner and get bound to it, resulting in controlling various processes (immune system, inter or intracellular signaling) in biology and physiology. Bonding between two partners can be a result of electrostatic, hydrophobic interactions or shape complementarity. It is of great importance to study these kinds of biomolecular interactions to have a detailed knowledge of above mentioned physiological processes. These studies can also open avenues for other aspects of science such as drug development. Discussed in the first part of Chapter 1 are the biotin-streptavidin biomolecular interaction studies by atomic force microscopy (AFM) and surface plasmon resonance (SPR) instrument. Also, the basic working principle of AFM and SPR has been discussed.

The second part of Chapter 1 is discussed about site-specific chemical modification of peptides and proteins. Proteins have been used to generate therapeutic materials, proteins-based biomaterials. To achieve all these properties in protein there is a need for site-specific protein modification.

To be able to successfully monitor biomolecular interaction using AFM there is a need for organic linker molecule which helps one of the investigating molecules to get attached to the AFM tip. Most of the linker molecules available are capable of investigating one type of interaction at a time. Therefore, it is significant to have linker molecule which can monitor multiple interactions (same or different type) at the same time. Further, these linker molecules are modified so that biomolecular interactions can also be monitored using SPR instrument. Described in Chapter 2 are the synthesis of

organic linker molecules and their use to study biomolecular interaction through AFM and SPR.

In Chapter 3, N-terminal chemical modification of peptides and proteins has been discussed. Further, modified peptides are attached to DNA thread for their translocation through the solid-state nanopore to identify them. Synthesis of various peptide-DNA conjugates and their nanopore studies have been discussed in this chapter.

ACKNOWLEDGMENTS

First and foremost I would like to express my deepest gratitude to my advisors Dr. Stuart Lindsay and Dr. Peiming Zhang for giving me the opportunity to work under their supervision. They have been an outstanding mentor for the last five years. Their expertise, understanding, generous guidance and support made it possible for me to work on a topic that is very interesting to me. I would also like to thank them for giving me the opportunity to work in synthetic chemistry as well as in peptide, protein biochemistry. I'm hopeful that the kind of expertise I have gained in this lab will help me further in my career. I would like to thank Professor Ian Gould for his constant support from the very first day when we came here. I would also like to thank Professor Kevin Redding for his support.

I would like to thank Dr.Subhadip Seanpati, Saikat Manna, Suman Sen, Sovan Biswas, Dr.Weisi Song, Jong One, Subarna Samanta, Dr.Yanan Zhao, for their valuable suggestions whenever I have encountered any problem in my research. It was fun working with you guys.

Last but not the least I would like to thank my parents, my sisters and all other family members and friends for their constant support and faith in me. This thesis is dedicated to you guys.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF FIGURES vii

LIST OF SCHEMES x

LIST OF ABBREVIATIONS xii

CHAPTER

1. INTRODUCTION 1

   1.1 Biomolecular Interactions Monitored by Atomic Force Microscopy 1

   1.2 Atomic Force Microscopy (AFM) 6

   1.3 Surface Plasmon Resonance as a Tool to Study Biomolecular Interaction 13

   1.4 Site-Specific Chemical Modification of Peptides and Proteins 18

2. STUDYING BIOTIN-STREPTAVIDIN INTERACTION USING ORGANIC LINKER MOLECULE THROUGH AFM AND SPR 25

   2.1 Introduction 25

   2.2 Results and Discussion 31

   2.4 Experimental Procedures 44

3. SITE-SPECIFIC CHEMICAL MODIFICATION OF PEPTIDES, PROTEINS AND ITS ATTACHMENT TO DNA THREAD FOR THEIR TRANSLOCATION THROUGH SOLID-STATE NANOPORES 73

   3.1 Introduction 73

   3.2 Results and Discussion 80

CHAPTER Page

3.4 Experimental Procedures 101

REFERENCES 108

APPENDIX

A. COPYRIGHT PERMISSIONS.....116

LIST OF TABLES

Table Page

1. Kinetic Parameters of Adsorption and Desorption of Streptavidin on Biotinylated Monolayers 40
2. Measured Physical Properties of Monolayers 69
3. Data for Site-Specific Protein Modification Using PLP 74
4. Peptide Sequences Used in This Study and Their Physicochemical Properties 81
5. Effects of Reaction Conditions on the Selectivity of Azido Acetic Anhydride 84
6. Dwell Times of PolyT_{20} and its Peptide Conjugates in Different Nanopores 98

LIST OF FIGURES

Figure Page

1. Schematic Representation of LHRH or LHRH-PE40 Attached to the AFM Tip 3
2. Schematic Representation of A) Ferritin Immobilized on the Surface, B) Antiferritin Attached to the AFM Tip 4
3. Schematic Representation of Conc A Immobilized on the Gold Surface.....5
4. Working Principle of AFM. 7
5. An Agilent AFM Instrument.....8
6. Schematic Representation of Different Parts of AFM. 9
7. Force vs Distance Curve Showing Different Region of Forces During Atomic Force Microscopy (AFM) Experiment 10
8. Schematic Representation of Force Experiment Using AFM.....12
9. Schematic Diagram of Force-Distance Cycle.....13
10. Schematic Representation of a SPR Experiment 15
11. Diagram of a BI 2000 SPR Instrument 17
12. Schematic Representation of in Situ Growth of [Poly(OEGMA)].....19
13. Schematic Representation of Hydrogels Formed From Metallothioneins.....20
14. Schematic Representation of Native Amino Acid Residues in Proteins Undergoing Reaction for Attachment of Synthetic Group.....21
15. Schematic Representation of Amino Acid Residues Participating in Site-Specific Chemical Modification in Proteins 23

Figure Page

16. Schematic Representation Showing Selective Modification of Cysteine Residue by Formylglycine-Generating Enzyme 24
17. Schematic Representation Showing Attachment of Investigating Molecule to the AFM Tip Using Linker Molecule 25
18. Structure of Organic Linker Molecule Having Maleimide Group and NHS Group.26
19. Structure of Organic Linker Molecule Having Acetal Group and NHS Group 27
20. A) Biotin-streptavidin Interaction Through Hydrogen Bonding, B) Molecular Model Structure Showing Tryptophan Interaction with Biotin 30
21. A View of the Crystal Structure of a Biotin-Streptavidin Complex 31
22. Structures of Monobiotin Ligands and Spacer 36
23. Schematic Representation of SPR Experiment 37
24. Schematic Illustration of a Mixed Monolayers 38
25. A Representative Unbinding Force-Distance Curve 42
26. FTIR Spectra 70
27. HPLC and Mass Analysis of Azidoacetic Anhydride Reacting With P-1 83
28. HPLC and Mass Analysis of Acetic Anhydride Reacting With P-1 85
29. RP HPLC Profiles of Azidoacetic Anhydride Reacting With (A) P-2 and (B) P-3 .87
30. Mass analysis of Products From P-2 Reacting With Azidoacetic Anhydride.....88
31. RP HPLC Profiles of Peptide-DNA Conjugate.....90
32. CD Spectra of PolyT_{20}-Peptdie Conjugates 91
33. RP HPLC Profiles of Azidoacetic Anhydride Reacting With Proteins 93

Figure Page

34. HPLC Diagram for Synthesis of Various Conjugates.....95

35. Schematic Illustration of a Nanopore Device for Translocation Measurements 97

36. Histograms of Fractional Current Blockades for DNA-Peptide Conjugates 99

LIST OF SCHEMES

Scheme Page

1. Synthesis of Three Arm Bisbiotin Ligands 35
2. Thiolation of Three Arm Bisbiotin by Staudinger Ligation....36
3. Synthesis of Compound 2 45
4. Synthesis of Compound S6a. 47
5. Synthesis of Compound S6b 49
6. Synthesis of Compound 3. 51
7. Synthesis of Compound 4a....53
8. Synthesis of Compound 5a....54
9. Synthesis of Compound 6a....55
10. Synthesis of Compound 4b 57
11. Synthesis of Compound 5b 58
12. Synthesis of Compound 6b 59
13. Synthesis of Compound 7 60
14. Synthesis of Compound 8 62
15. Synthesis of Compound 11 63
16. Synthesis of Compound 11 66
17. N-terminal Modification of Twenty Naturally Occurring Amino Acids by PLP 74
18. Schematic Representation of Oxidative Coupling of O-amino Phelol....75
19. Schematic Representation of N-terminal Modification of Peptide or Protein by Ketene Compound 76

Scheme Page

20. Chemical Reactions for Attaching an Oligonucleotide to N-termini of Peptides 81

21. Scheme for Synthesis of Opposite Charged Molecules with Peptide 94

LIST OF ABBREVIATIONS

APCI	atmospheric pressure chemical ionization
AFM	atomic force microscopy
ARRP	atom transfer radical polymerization
aq	aqueous
BI	biosensing instrument
br s	broad singlet
^{13}C NMR	carbon-13 nuclear magnetic resonance
^{o}C	degree Celsius
cat	catalytic
$CDCl_3$	deuterated chloroform
CH_2Cl_2	methylene chloride
Conc A	concanavalin A
cm	centimeter
δ	chemical shift (ppm)
d	doublet
DFT	density functional theory
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
ESI	electrospray ionization
ELISA	enzyme-linked immune-sorbent assay

EGFR	epidermal growth factor receptors
Fab	fragment of antigen-binding
Fc	fragment crystallizable
g	gram (s)
GPCR	G protein-coupled receptor
^{1}H NMR	proton nuclear magnetic resonance
h	hour (s)
H_2O	water
HCl	hydrochloric acid
HPLC	high pressure liquid chromatography
Hz	Hertz
HSA	human serum albumin
J	coupling constant
L	liter
LHRH	Luteinizing hormone-releasing hormone
M	molar
m	multiplet
mL	milliliter
mM	millimolar
mmol	millimole(s)
Mb	myoglobin
μM	micromolar

μmol	micromole(s)
$NaHCO_3$	sodium bicarbonate
nm	nanometer
NMR	nuclear magnetic resonance
OEG	oligo[ethylene glycol]
PE40	*Pseudomonas aeruginosa* exotoxin 40
PEEK	polyether ether ketone material
PLP	pyridoxal-5-phosphate
RNA	ribonucleic acid
s	singlet
SAM	self-assembled monolayer
STM	scanning tunneling microscope
SPR	surface plasmon resonance
t	triplet
TEM	transmission electron microscopy
THF	tetrahydrofuran
TLC	thin layer chromatograpy
UV	ultraviolet

CHAPTER 1

INTRODUCTION

1.1 Biomolecular interactions monitored by atomic force microscopy (AFM)

The atomic force microscopy (AFM) is a type of scanning probe microscopy used to investigate surface properties with nanoscale resolution.[1,2] Scanning tunneling microscopy (STM) was the first of the series of probe microscopes, invented by Gerd Binnig and Heinrich Rohrer.[3] They were awarded Nobel Prize in 1986 for this discovery. STM is capable of imaging conducting and semiconducting surfaces, but unable to perform the same job in case of non-conducting surfaces. This fact led to the discovery of AFM, capable of imaging both conducting and nonconducting surfaces, by Gerd Binnig and first commercial AFM was available in 1989. Other than imaging, one of the significant applications of AFM is its ability to measure mechanical properties of the biological samples.

Studying molecular interactions is of great importance in molecular and structural biology. These molecular interactions comprise of forces which include multiple weak, non-covalent bonds formed between different interacting portions of molecules. Various techniques such as surface force apparatus,[4] optical traps, and tweezers,[5,6] have been developed for past couple of decades to investigate molecular interactions at single molecular level. Hige force sensitivities make the optical tweezers popular choice for performing experiments, but large sample length and limitation in the amount of force applied for the investigation purpose make use of this technique limited.

AFM provides a solution regarding both the features. Here are some of the important biomolecular interactions monitored by AFM listed below.

1.1.1 Receptor/ligand interactions

Biomolecules can easily recognize its corresponding partner and get bound to it, resulting in controlling various processes (immune system, inter- or intracellular signaling) in biology and physiology. Bonding between two partners can be the result of electrostatic, hydrophobic interactions or shape complementarity. It is of great importance to study these kinds of receptor/ligand interactions to have a detailed knowledge of above mentioned physiological processes. These studies can also open avenues for other aspects such as the development of new drugs. G protein-coupled receptors (GPCRs) are one such receptor which remains overexpressed in most of the cancer cells.[7] Therefore, a ligand attached to an entity having anticancer property can be used to target overexpressed receptor for drug delivery. A detailed understanding of receptor-ligand interaction is required to make this approach a successful one. One of such receptor-ligand interaction studied by AFM is shown here.

Luteinizing hormone-releasing hormone (LHRH) receptor, a GPCR, is found to be overexpressed in several cancer cell lines. But, it remains absent in healthy human visceral organs.[8] LHRH is a decapeptide hormone having a binding affinity towards LHRH receptor. Therefore, peptides having anti-cancer property can be fused to LHRH, and the whole construct can be directed towards overexpressed LHRH receptor in cancer cell lines.[9] One of such peptides having anti-cancer property is toxic *Pseudomonas aeruginosa* exotoxin 40 (PE40) which is proposed to fuse with LHRH to make anti-cancer drug conjugate .[10] Now, investigation to study the interaction between LHRH-PE40 and LHRH receptor has been carried out using AFM technique. Also, results have

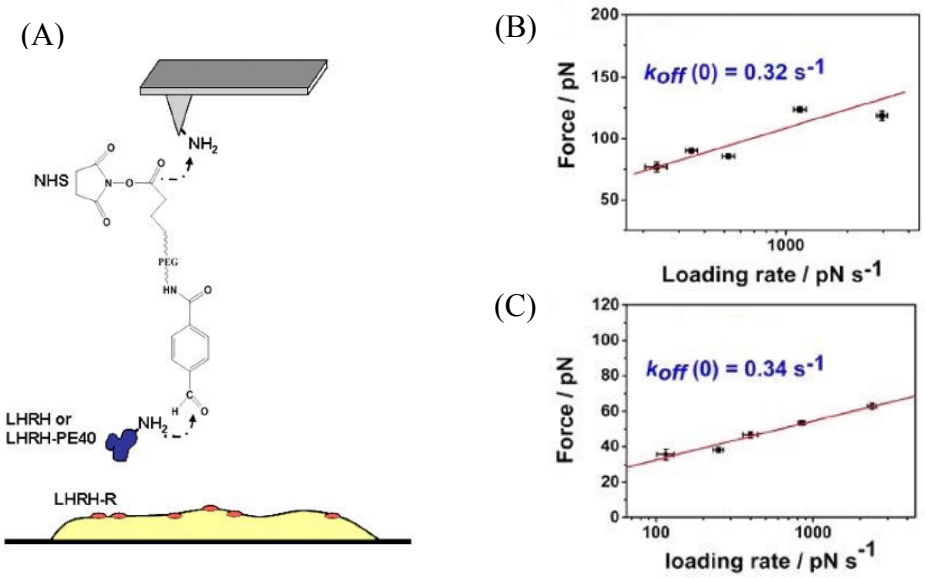


Figure 1. A) Schematic representation of LHRH or LHRH-PE40 attached to the AFM tip through linker and recptor immobilized over the surface, B) The dependence of unbinding force on loading rate for LHRH-PE40/LHRH receptor, C) The dependence of unbinding force on loading rate for LHRH/LHRH receptor.10

been compared to LHRH-LHRH receptor interaction.[10] Schematic diagram for performing experiments is shown in Figure 1.

Force spectroscopy results revealed that there is no significant difference in the dissociation rate constants between LHRH-PE40/LHRH receptor and LHRH/LHRH receptor.[10] Thus, the fusion of PE40 does not interfere with the recognition and stability of LHRH with LHRH receptors. Thus, single molecule force spectroscopy provides us with the investigations about kinetics as well as mechanical properties of the ligand.

1.1.2 Protein/protein interactions

Studying protein-protein interactions such as specific antigen-antibody interaction is of great importance. AFM has been employed to investigate forces of interactions between various antigen/antibody such as antihuman serum albumin (anti-HSA)/ HSA[11] and antifluorescyl antibody/ fluorescein[12]. One of the examples shown here is between antiferritin antibody and ferritin.[11]

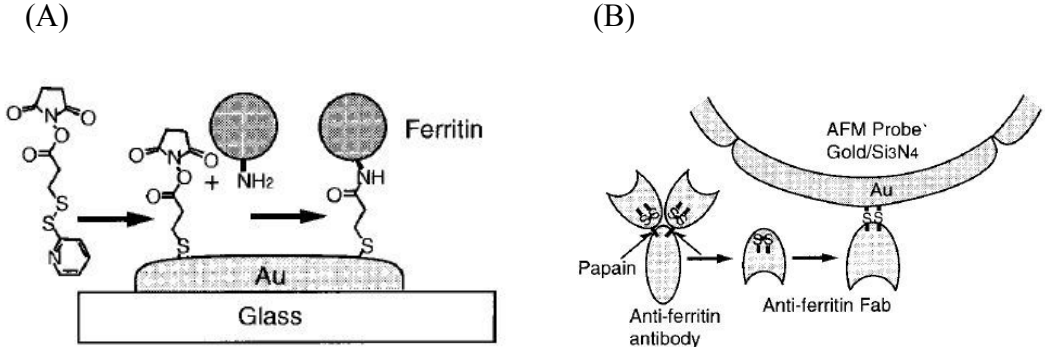


Figure 2. Schematic representation of A) ferritin immobilized over the surface, B) antiferritin attached to the AFM tip.[11]

Distribution of force histograms revealed that single binding event corresponds to 63pN force and increase in force values correspond to the multiple antigen-antibody binding events.

1.1.3 Protein/carbohydrate interactions

Affinity of carbohydrate towards lectin is a well-known phenomenon. Studying this carbohydrate/lectin interaction is of great importance for developing a new application in bioanalytical and biomedical fields.[13,14,15,16] Concanavalin A (Conc A) is such an example of plant lectins which help in agglutination of erythrocytes and precipitation of glycogen

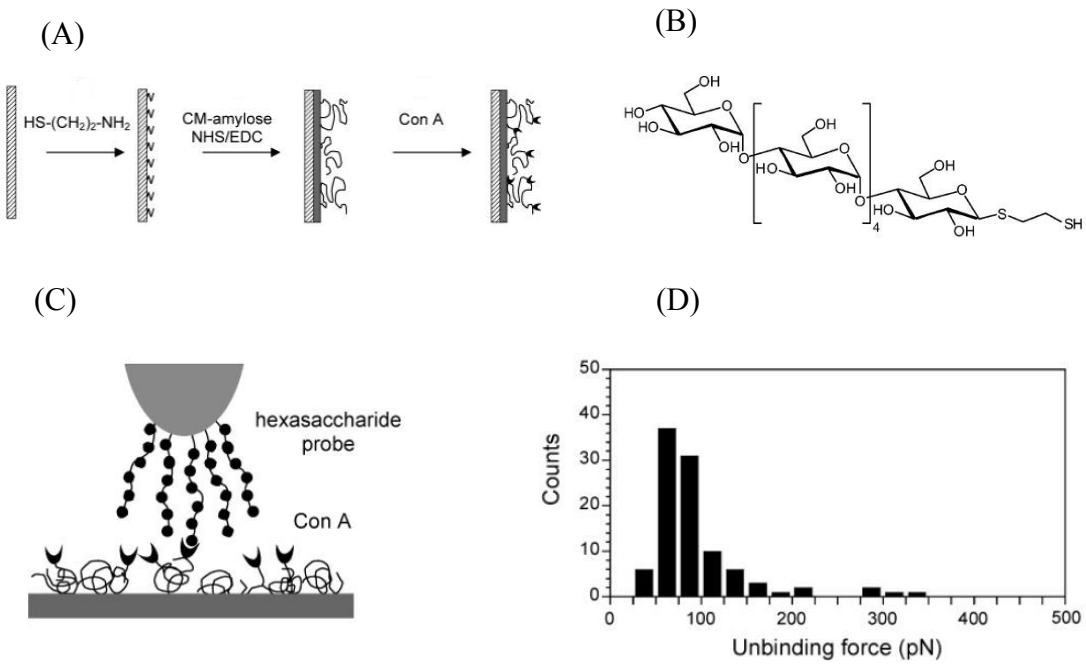


Figure 3. A) Schematic representation of Conc A immobilized over the gold surface, B) structure of thio-terminated hexasaccharide used in the study, C) schematic diagram of the experiment, D) histogram of unbinding forces.[17]

and starch from solution through its attachment to α-manosyl and α-gucosyl groups. Therefore, monitoring forces of interaction between Conc A and oligo glucose molecule would help us to understand detailed mechanisms of above-mentioned processes. Here is one such example displayed here to investigate molecular interaction between Conc A and hexasaccharide using AFM techniques.[17]

Histogram of unbinding forces reveals that mean value for forces of interaction is 55pN.[17] Two control experiments were carried out. One is monitoring forces of interaction between hydroxyl terminated polyethylene glycol molecule, and Conc A. Other one is by blocking binding site of Conc A using mannose, followed by studying force curves. Both the cases resulted in a very insignificant amount of specific forces of interactions. This, in turn, proves forces are originated from the specific Conc A-carbohydrate interaction.

1.2 Atomic force microscopy (AFM)

AFM is a powerful technique which can be used in different fields of science such as biology, nanotechnology, surface science for various purposes which include imaging, force spectroscopy, recognition imaging. There are a couple of advantages of using AFM over other imaging techniques such as Transmission Electron Microscopy (TEM) or STM.

a) For imaging, AFM can work on both conducting and non-conducting surfaces, whereas STM needs conducting surfaces.
b) Experiments can be carried out in air, liquid and vacuum medium using AFM.
c) AFM is capable of generating three-dimensional images.

Other than these above mentioned facilities AFM has some disadvantages too. Slow scanning speed and small surface area scanning are two disadvantages of AFM. Still, it's use in force spectroscopy and recognition imaging making it an indispensable tool in the field of nanoscience.

1.2 Working principle

The basic configuration of AFM is shown in Fig 4. AFM consists of a cantilever and a sharp tip attached to one end of the cantilever. When the tip is brought closer to the surface to be analyzed, the interaction between surface atoms and tip atoms cause

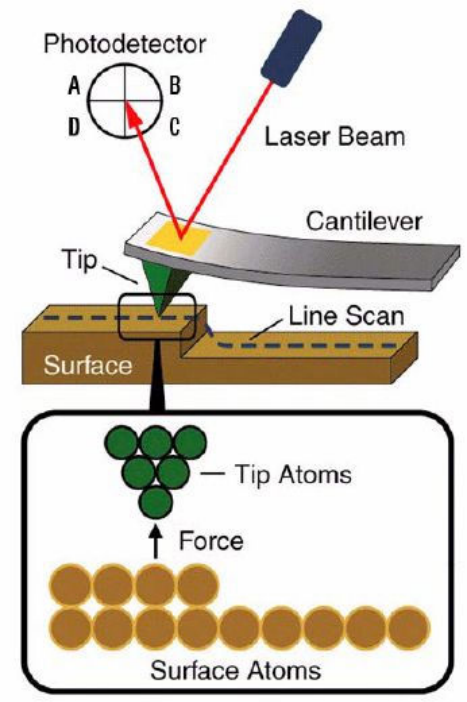


Figure 4. Working principle of AFM. (Adapted from Agilent 5500 manual)

deflection in the cantilever. This deflection is further measured by a laser beam falling on the back side of a cantilever and getting collected into the photo diode detector.

This phenomenon occurs based on Hook's law which says force (F) required to expand or contract a spring is directly proportional to the amount of displacement (x) of spring.

$$F \alpha (-x)$$

$$\text{or,} \quad F = -Kx \quad [K = \text{spring constant}]$$

[negative sign indicates direction of spring displacement is opposite to the applied force]

1.2.1 Different parts of AFM

Different parts of AFM have been described below.

AFM probe:

AFM probe consists of a holder plate carrying a cantilever and tip. The tip is made up of

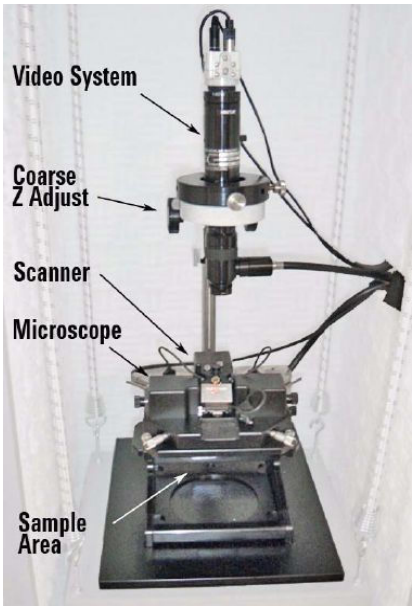


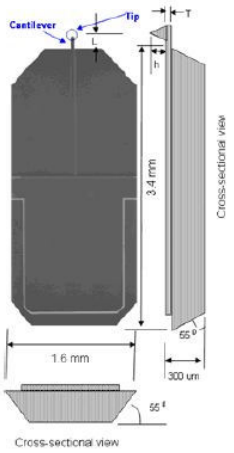
Figure 5. An Agilent AFM instrument. (Adapted from Agilent 5500 manual)

a silicon or silicon nitride. The radius of the tip is of nanometer diameter and length of the cantilever is around 15-30 µm.

AFM scanner:

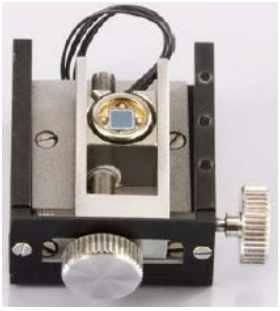
Scanner has a piezoelectric material which helps the scanner to elongate or contract upon exposure to the external electric field. This scanner movement, in turn, moves the tip along z-direction as it is attached to the scanner.

(A)



(B)

(C)



(D)

Figure 6. A) Schematic representation of AFM probe, B) AFM scanner, C) photodiode, D) sample plates. (Adapted from Agilent 5500 manual)

Detector:

Detectors consist of photodiodes which collect change in laser spot position equivalent cantilever deflection.

Sample plate:

It is equipped with a place where surfaces having immobilized sample can be kept. Different sample plates are available based on various experimental setup.

1.2.2 Modes of AFM

AFM can work in three different modes

a) Contact mode

b) Non-contact mode

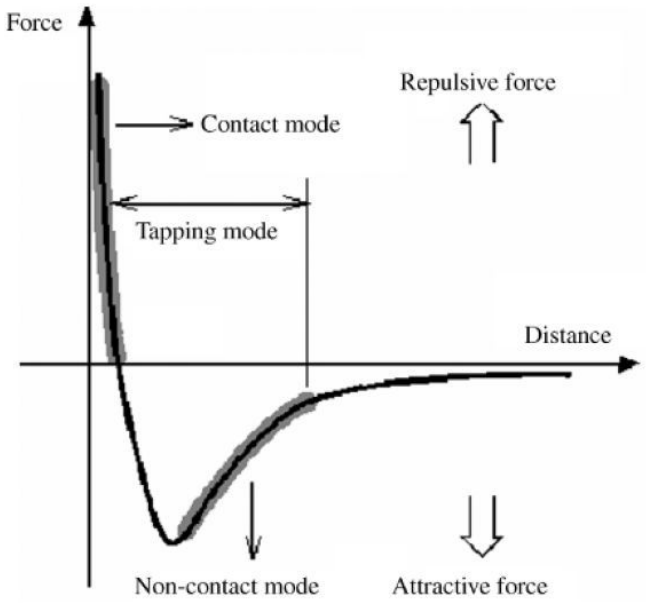


Figure 7. Force vs distance curve showing different region of forces during AFM experiment.[18]

c) Tapping mode

Contact mode

In this mode, AFM tip remains in contact with the surface. As tip always remains in contact with the surface, there is always a chance of soft biomaterial samples being destroyed. So, this mode is not suitable for soft samples.

Non-contact mode:

In this mode, the tip does not make any contact with the surface; rather it oscillates at or near resonance frequency keeping a fixed distance from the surface. As tip never touches samples, this mode is applicable for soft biological samples. But, resolution of images obtained from this mode is generally low.

Tapping Mode:

In this mode, tip oscillates at or near resonance frequency and makes intermittent contact with the surface. This mode is highly popular in taking high-resolution images of soft biomaterial sample.

1.2.3 Force spectroscopy

One of the most popular applications of AFM technique is its use to measure the force of interaction at single molecular level. To be able to investigate force of interaction between two molecules, one has to be immobilized on the surface and another one has to be attached to the AFM tip. Now, the tip is brought closer to the surface so that tip makes contact with the surface, allowing both the molecules to bind with each other. The tip is then retracted from the surface to rupture interactions between the molecules. This

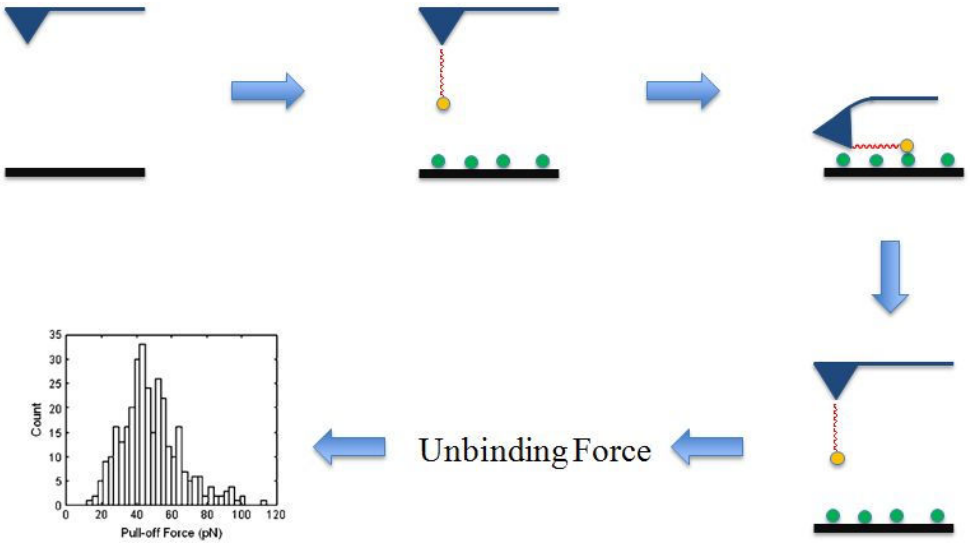


Figure 8. Force vs distance curve showing different region of forces during AFM experiment.[19]

rupturing force is called unbinding force which is finally measured by force spectroscopy. Now, approach and retraction incidents can be performed several times at a particular place or different places over the surface to collect a lot of undinding force data, followed by plotting a histogram to get a distribution of unbinding forces. In a force curve, the x-axis denotes distance of a tip from the surface and cantilever deflection is plotted along the y-axis.

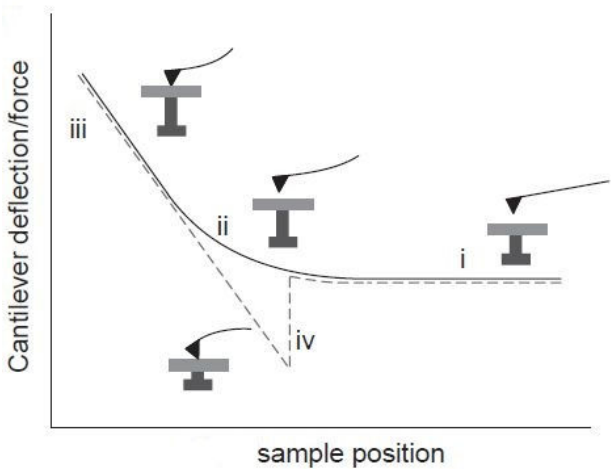


Figure 9. Schematic diagram of force-distance cycle including approach and retraction curve.[20]

1.3 Surface plasmon resonance as a tool to study biomolecular interaction

Surface plasmon resonance (SPR) is an optical biosensor technique which helps us to study kinetic analysis of biomolecular interaction by detecting refractive index change in the close vicinity of the metal surface. Surface plasmons at the metal surface oscillate to generate electric field which ranges up to 200 nm above the metal surface and sample solution junction.[21] Any changes within this range can be monitored to perform the real time kinetic analysis between investigating pair of molecules. First commercially available SPR was in the market in 1990s.[22] Since then it has been used extensively to carry out various biomolecular interactions such as protein-protein,[23,24] protein-DNA,[25,26] protein-polysaccharides[27]. SPR has a couple of advantages over other available techniques such as fluorescence or enzyme-linked immune-sorbent assay (ELISA).

a) There is no need for the analytes to be labeled by fluorescence tags for SPR studies. SPR works on detecting changes in refractive index. So, the analytes do not require to have any special characteristics.
b) SPR data can be collected in real time to get kinetic and thermodynamic parameters related to the analyte.
c) Samples tested in SPR can have a high range of molecular weights and binding affinity.

These are the features that make SPR an essential technique to monitor biomolecular interactions.

1.3.1 Working principle

Surface plasmons are delocalized conduction electrons present at the interface between any two materials (metal/air). These plasmons generate an electromagnatic wave by oscillating at a particular frequency. This wave goes in a direction parallel to the metal surface. The wave being at the boundary of a conducting surface and external medium, it gets disturbed by any changes in the close vicinity of the boundary, such as passing of any molecule over the surface. When this metal surface is irradiated with plane polarized light at a critical angle, the entire laser light undergoes total internal reflection. Under this condition resonance occurs when the frequency of incident photons matches with the one from oscillating plasmons, leading to the transfer of the energy from the photon to the plasmons. This energy transfer results in the reduction in intensities of reflected light. Now, any changes in refractive index close to the metal surface will have a different angle at which energy would be transferred.

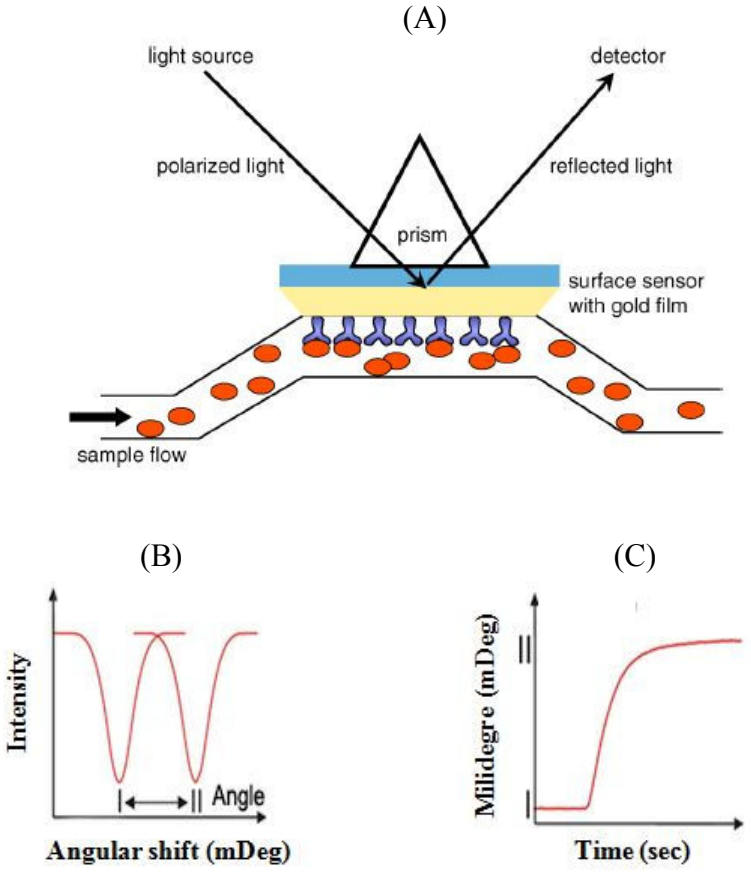


Figure 10. A) Schematic representation of a SPR experiment,[28] B) laser intensity vs angular shift graph showing dip in laser intensity at two different angles,[29] C) a typical SPR sensorgram representing ligand-analyte interaction over time.[29]

An example of these two angles is displayed in the fig.10. If changes in laser incident angle can be monitored over time, real time kinetic parameters can be derived from this.

In a typical SPR experiment, one set of molecules (ligand molecule) is immobilized on the metal thin film, and another set of molecules (analyte molecule) is flown over the metal film to study molecular interactions. When the analyte molecules reach to the metal surface, changes in the refractive index near surface cause a jump in the SPR signal

which is represented as a curve called sensorgram. An example of one such sensorgram is shown in fig.10.

1.3.1 Different parts of SPR

Diagram of a BI-2000 SPR is shown in fig 10.

It comprises of a plane polarized light source, a prism, a metal thin film and a photodetector.

Syringe pump:

For BI 2000 instrument, there is one syringe pump which controls the flow of buffer running through the SPR instrument. This syringe pump can accommodate two 10 mL syringes. Flow rate, the volume change of buffer in syringes can be adjusted through this part of an instrument. During experiments, generally, buffer flow rate through the instrument is kept between 30 μL/min - 100 μL/min.

Valve Tray

Valve tray has three component parts- mode (single and serial) controller, channel controller and sample injection area. Different injection loop sizes (100 μL - 1000 μL) can be used for SPR experiments.

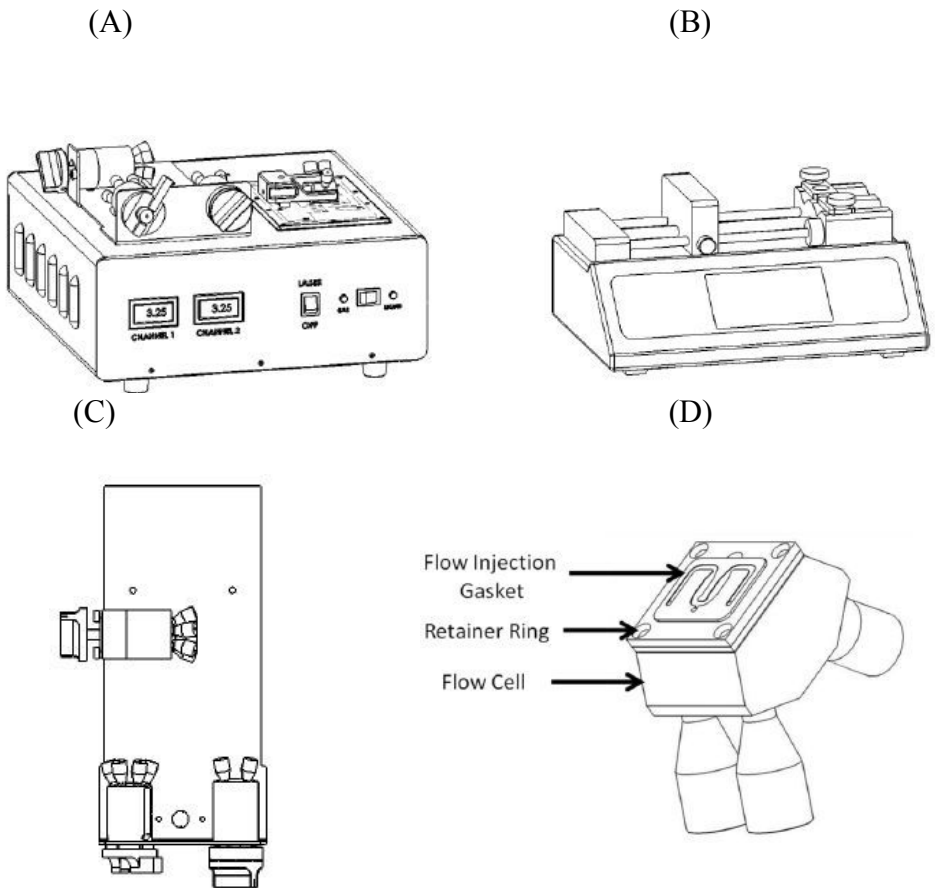


Figure 11. Diagram of a BI 2000 SPR instrument A) Diagram of different parts of SPR instrument B) syringe pump, C) valve tray, D) flow cell and retainer ring (Adapted from Biosensing Instrument BI 2000 SPR mannual)

Flow cell and retainer ring

Flow cell in SPR contains injection gasket. It is made up of polyether ether ketone material (PEEK). Dimension of flow cell is 2.8 x 2.0 x 1.4 cm.

Flow injection gasket

Injection gasket has channels through which buffer solution passes through the waste bottle. It is made up of silicone-based elastomer material.

1.4 Site-specific chemical modification of peptides and proteins

1.4.1 Importance of protein bioconjugates in biology

Proteins, class of large biomolecules comprise of long chain amino acid residues, can take part in controlling diverse functions such as catalysis,[30] storing energy,[31] responding to pathogens and neutralize them[32]. To get a detailed understanding of these functions, protein structures are needed to be studied. Recently, proteins have been used to generate therapeutic materials, proteins based biomaterials. To achieve all these functional properties in protein along with studying its structural aspect, there is a need for site-specific protein modification. Chemical modification of amino acid side chains started way back over 100 years ago and still today search is going on for simple, effective bioconjugation strategies. Some of the examples of protein-based material and protein polymer conjugates having therapeutic properties have been discussed below.

Protein-polymer drug conjugate with improved pharmacokinetics

Site-specific attachment of polymer chains to protein to improve its pharmacokinetic properties have been carried out over the past couple of decades. Use of biocompatible, non-immunogenic polyethylene glycol as a polymer chain capable of improving serum half-life and stability of a protein is very common.[33,34] One such example is shown below.

Poly(oligo(ethylene glycol)methyl ether methacrylate) [Poly(OEGMA)] was attached stoichiometrically to the N-terminus of myoglobin (Mb) protein. Mb was first site-specifically modified to a ketone functional group using biomimetic transamination

reaction, followed by attachment of atom transfer radical polymerization (ATRP) initiator through imine formation. Next, ATRP of OEGMA was carried out in situ in the buffer.

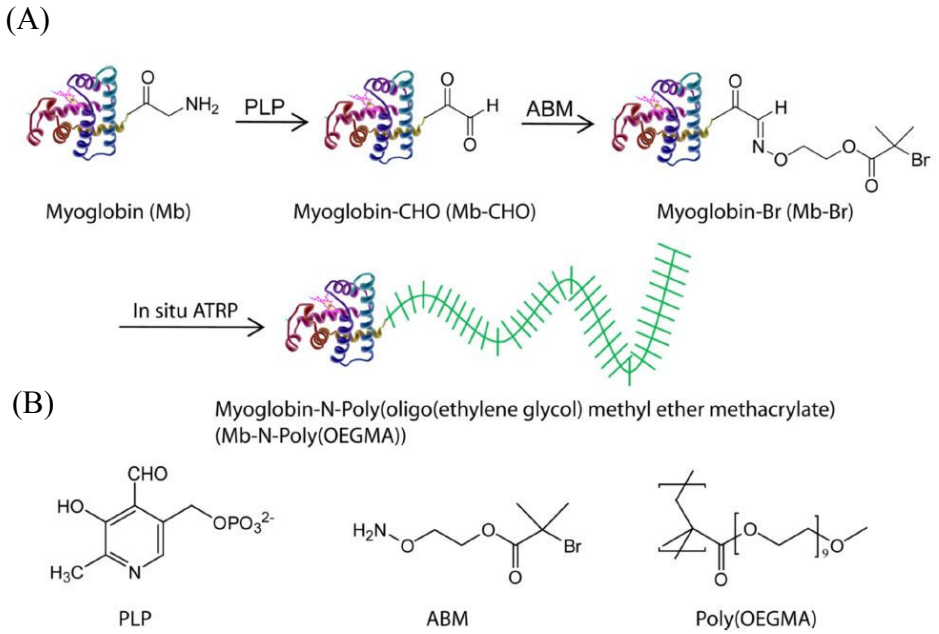


Figure 12. A) Schematic representation of in situ growth of [Poly(OEGMA)] at the N-terminus of myglobin. B) Stuructures of pyridoxal-5-phosphate (PLP), (2-(aminooxy)ethyl) 2-bromo-2-methylpropanoate (ABM) and [Poly(OEGMA)].[35]

Mb and polymer conjugated Mb have been administered intravenously into nude mice to monitor pharmacokinetic behaviors of both molecules. Results revealed that Mb has short distribution phase and rapid terminal elimination phase compared to Mb-OEGMA conjugate.[35]

Protein-based material

Attachment of synthetic groups to protein leads to a generation of protein-based materials where protein functions are still kept intact. This newly synthesized biomaterials can be used for various purposes such as building light harvesting devices

exploiting protein functions.[36,37] One example of the protein-based hydrogel is shown below.

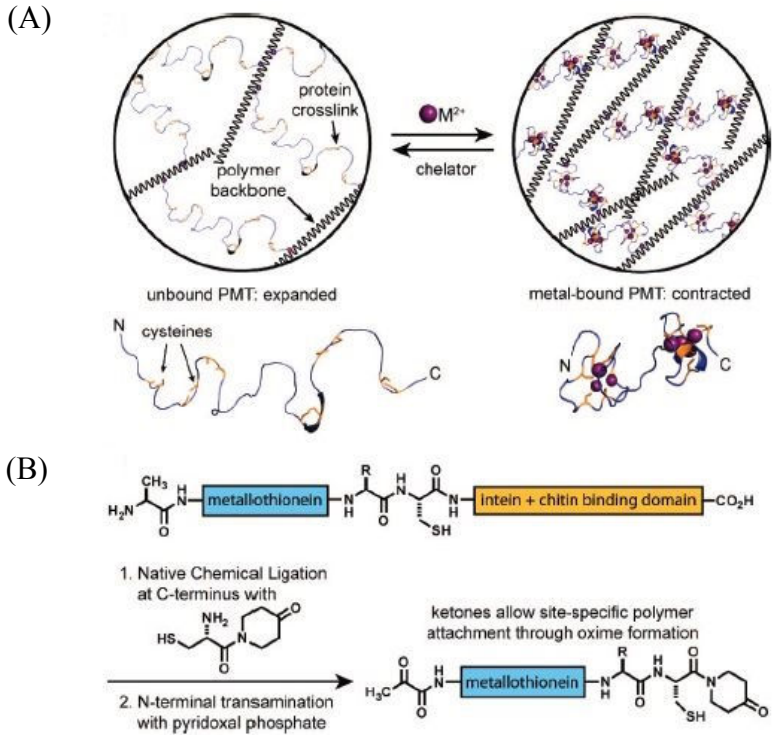


Figure 13. A) Schematic representation of hydrogels showing attachment of metals to proteins inside hydrogels. B) reaction scheme showing site-specific modification of protein at bot N- and C-terminus.[38]

Metallothioneins are conjugated to the polymer formed from polymerization of acrylamide monomer, at C- and N-terminus. Now, these conjugated polymers are allowed to form hydrogels where proteins will be incorporated inside hydrogels. These hydrogels are cut into small pieces and incubated with different metals such as cadmium, copper, mercury, zinc, manganese, calcium. Studies revealed that heavy metals cadmium, mercury help to reduce the volume of hydrogels due to its ability to bind with proteins.

But, there is no such significant amount of volume change noticed for hydrogels which were incubated with non-heavy metals. That is how, this protein-based material is capable of recognizing and sequestering toxic heavy metals in presence of large concentration of background ions.[38]

1.4.1 Traditional methods for chemical bioconjugation

The presence of unprotected polar side chains over the protein surfaces makes attachment of a synthetic group to a particular position very challenging task. Also, under the reaction conditions, proteins have to remain stable. Basically, the reaction has to be carried out in mild condition (aqueous media, near neutral pHs, mild temperature). Some of the traditional methods for protein bioconjugations are listed below.

Amino acids which are targeted for attaching chemical handles are aspartic acid, glutamic acid, cysteine and lysine residues. Lysine side chains are susceptible to reaction with N-hydroxy succinamide (NHS) ester containing reagents. Cysteines are capable of

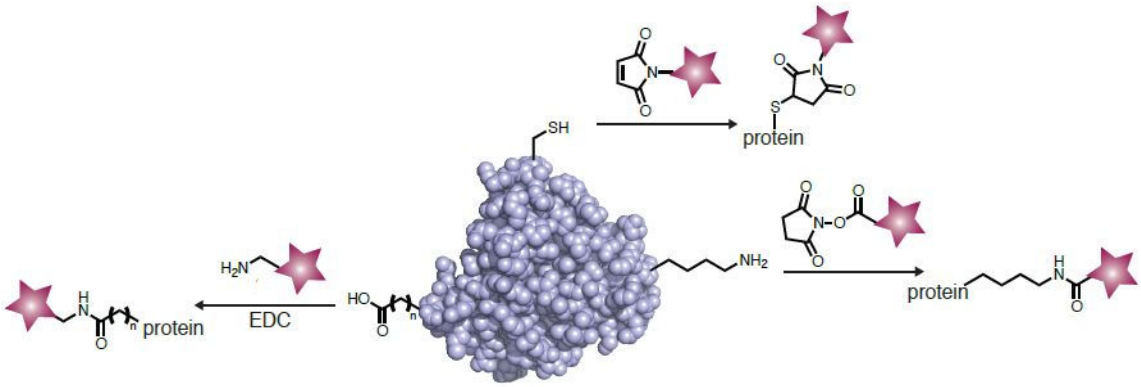


Figure 14. Schematic representation of native amino acid residues in proteins undergoing different reaction for attachment of synthetic group.[39]

reaction with maleimide functional group containing reagents. Carboxylic acid groups of aspartic acid and glutamic acid can be activated reacting with activating reagent such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC).[39]

Issues with these traditional methods are the installation of multiple copies of desired synthetic group all over the proteins. This leads to the formation of the heterogenous mixture, resulting in a detrimental effect on some of the protein-based bioconjugates such as protein-based therapeutic agents.A heterogenous mixture of protein conjugates will make pharmacokinetic analysis very challenging. Moreover, indiscriminate attachments might lead to modification of active site of a protein resulting in changes in protein functional property. To solve these problems, there is always search for chemical reactions which selectively install reactive handles at a desired location in protein.

1.4.1 Advancement in site-specific bioconjugation

Site-specific chemical modification of proteins can be carried out in a single step or two steps. A single step will have functional handle directly installed into protein. In two-step process first, a reactive handle will be site-specifically installed with proteins, followed by a bioorthogonal reaction for attaching the desired molecule to the protein. Some of the site-specific bioconjugation reactions are discussed below.

One of the techniques is to target amino acid residue which prevails in lower abundance in protein.[40] Cysteine is one such amino acid whose one solvent accessible copy can be placed into the protein. Using chemical ligation reaction, cysteine derivative can be reacted selectively with a C-terminus modified thioester.[41,42] Aromatic amino acid residues such as tyrosine, tryptophan are also found with intermediate frequency on

protein surfaces. These amino acids can be targeted for site-specific chemical modification in protein. Some of the examples have been discussed below.

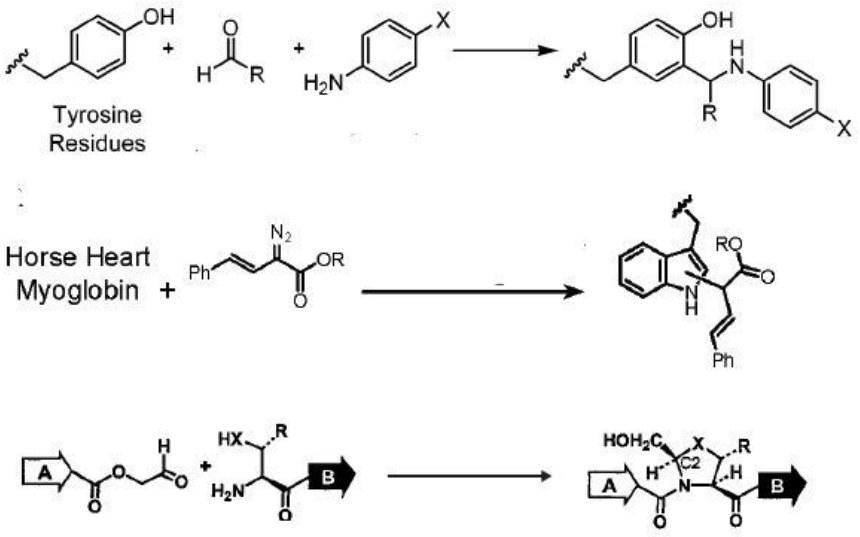


Figure 15. Schematic representation of specific amino acid residues participating in site-specific chemical modification in proteins.[43,44,46]

A tyrosine residue in chymotrypsinogen A was selectively modified by various aldehyde compounds in the presence of aniline derivatives at pH 6.5 and 37 ^{0}C temperature. Site-selectivity was confirmed by electrospray ionization mass spectrometry (ESI-MS).[43] Tryptophan residue in horse heart myoglobin was selectively modified by a vinyl diazo compound in the presence of rhodium carbenoid at pH 3.5.[44] N-terminal tryptophan residues can be modified by the pictet-spengler reaction.[45] Serin, threonine residues can be selectively reacted with a peptidyl glycoaldehyde ester.[46]

Another effective technique for modifying proteins site-selectively is using enzymatic labeling. This technique enables us to modify a particular section of protein in protein mixture. Scalability is the only problem related to this technique. One of the examples of

this technique is use of formyl glycine enzyme (FGE) to selectively identify aldehyde tag (a 6 amino acid sequence) and converts cysteine in that sequence to aldeyde. The sequence FGE recognizes is CXPXR (X= serin, threonine, alanine, glycine).[47]

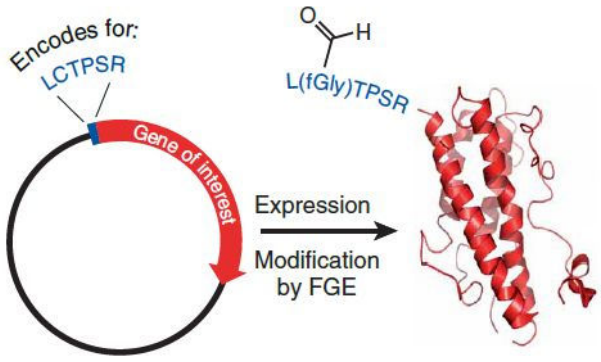


Figure 16. Schematic representation showing selective modification of cystein residue by FGE.[47]

Incorporation of unnatural amino acid into protein sequence is one of the successful and popular techniques for placing functional groups at any desired location during protein expression.[48,49] But, complexity of using genetic methods can be one of the disadvantages of this technique. That is why, our goal is to develop relatively simple and general one-step method which can be performed in the aqueous medium at near neutral pH and mild temperature to modify protein site selectively. Detailed of this process has been discussed in chapter 3.

CHAPTER 2

2 Studying streptavidin-biotin interaction using organic linker molecule through AFM and SPR

2.1 Introduction

Organic linkers used to study biomolecular interaction through AFM

Studying molecular interaction between small organic molecule and biomolecules is an important area of research. Most of the targeted drug delivery systems are developed based on the interaction between the small molecule and transmembrane receptor.[50] AFM is one technique which provides us with the opportunities to monitor force of interaction between these pairs in piconewton force level.

…As already mentioned in the first chapter, one of the two investigating molecules has to be attached to the AFM tip during force spectroscopy experiment. That can be achieved by using organic linker molecule is shown below.

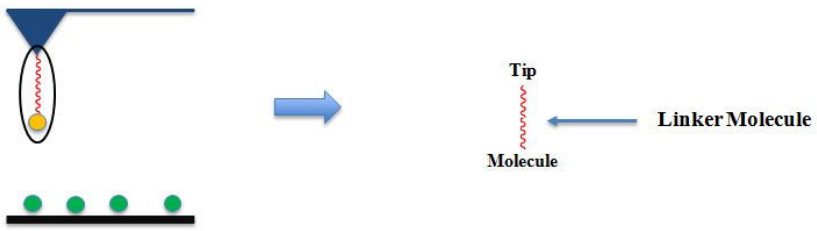


Figure 17. Schematic representation showing attachment of investigating molecule to the AFM tip using linker molecule

One end of the linker molecule will get attached to the AFM tip, and another end will have investigating molecule attached covalently to the linker. The role of linker molecule is paramount to get a high resolution and highly reproducible data. Linker molecule has

to have a particular criterion to be effective for force spectroscopy. Firstly, there are two types of forces recorded, non-specific and specific forces of interaction during force spectroscopy measurement. Non-specific originates from the interaction between AFM tip and surface, whereas the interaction between two investigating molecules contributes to the specific force of interaction. Recording of unbinding specific forces is desired. To be able to have distinguished specific forces, linker molecules have to be longer than certain lengths. Therefore, the length of the linker molecule is very important during force spectroscopy measurement. Secondly, linker molecule has to be stable, rigid and robust so that it can maintain its structure during the experiment. Lastly, the flexibility of linker is another important criteria because it allows investigating molecule to freely orient itself and get bound to the molecule immobilized on the surface. Some of the linker molecules used in the AFM studies have reported here. The structure of an organic linker molecule having maleimide functional group at one end and N-hydroxyl succinamide ester (NHS) group at the other end has been shown below. Maleimide group has been used to attach thiolated DNA aptamer, and NHS group has been reacted with its amine

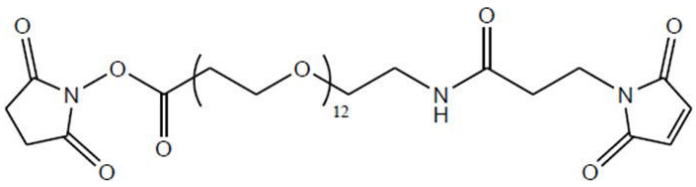


Figure 18. Structure of organic linker molecule having maleimide group and NHS group.[21]

functional group on the AFM tip. Force spectroscopy studies have been carried out

between DNA aptamer and immunoglobin E (IgE), resulting in 49+18 pN unbinding force.[21]

Another linker having NHS group on one end and acetal group on the other end has been shown below. NHS group has been attached to the AFM tip and acetal group has been reacted with biotin modified immunoglobin G (IgG). Force spectroscopy studies

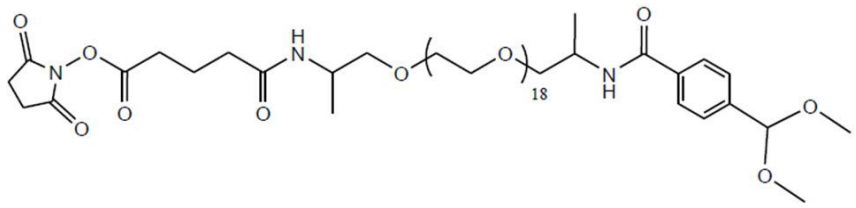


Figure 19. Structure of organic linker molecule having acetal group and NHS group.[51]

have been carried out between avidin and biotin, resulting in 51 pN of unbinding force.[51]

Issues with these linkers are only one type of biomolecular interaction can be monitored at a time because only one functional group site is available for the biomolecule attachment to the linker. Therefore, there is a need for a linker system through which more than one type of biomolecular interaction (same or different type) can be monitored. Three arm linker molecule has been proposed to serve this purpose.

Other than using three arm linker to improve biomolecular interactions investigation through AFM, its structural resemblance to anibody leads us to think of its use as antibody mimic.

Three arm linker as antibody mimic

"Magic bullets", such as small molecule inhibitors and monoclonal antibodies for targeting epidermal growth factor receptors (EGFR), have been developed for the treatment of cancer.[52] Although great progress in the targeted therapy has been made, there is still an absence of effective agents to tackle those proteins encoded by *TP53*, *RAS*, and *MYC* genes, where the most common mutations in human cancer take place.[53,54,55] For comparison, the small molecule drugs can easily access intracellular disease targets, but they have notorious off-target toxicities and lack effectiveness to block protein-protein interactions[56] whereas the antibodies possess high specificity and binding affinities to their protein targets, but exhibit poor extravasation and limited tissue penetration due to their large size and complexity.[57] Based on current knowledge, one can envision a new molecular entity that possesses merits of both large and small drugs for the discovery of cancer drugs. Recently, McEnaney *et al* have synthesized a class of antibody mimics that is composed of targeting and effector moieties, which has approximately 5% of the molecular weight of an antibody, demonstrating that the antibody analogy can selectively elicit an immune response for cellular phagocytosis through bivalent interactions.[58] This may imply a new direction in drug discovery through the creation of miniaturized antibody mimics with small molecules.[59]

We are interested first in mimicking the antibody's structure to understand effects of molecular architecture on functions of a molecule. The antibody is generally viewed as a Y-shaped molecule, which is composed of two Fab (fragment of antigen-binding) moieties connected to the Fc (fragment crystallizable) subunit through flexible linkers so

it can interact with its antigens bivalently. It is known that a bivalent interaction increases the affinity (i.e. avidity) and residence time of a drug molecule to its target.[60] These two parameters are paramount for the development of effective therapeutic agents.[61]

A Y-shaped three-arm linker has been designed in order to build an antibody mimic. Such a scaffold would allow us to functionalize its two arms each with an affinity molecule to form a homo or hetero bivalent ligand, and leave the third arm for attaching a ligand that binds to a receptor on immune cells or a fluorescent dye for monitoring the bivalent interactions. The Y-shaped structure may have a more favourable geometry for a bivalent ligand binding to two separated binding sites of a receptor than a linear linker. To demonstrate the effectiveness of an antibody mimic for bivalent binding, the biotin-streptavidin pair has been chosen as a model for the proof of principle study. The reason for choosing this pair as a model system is the fact that biotin-streptavidin pair is an example of one of strongest ligand-receptor interaction and the system has been well studied over the past decades.

Biotin-streptavidin interaction

Streptavidin is homotetrameric protein isolated from *Streptomyces avidini*.[62] Each monomer consists of eight antiparallel beta strands which in turn fold to form a beta barrel structure. Biotin is bound to the open end of the beta barrel structure. The affinity of this bonding is exceptionally high (dissociation constant 10^{-15} M).[63] This high biotin-binding affinity of streptavidin arises from the following reasons.

a) Biotin fits really well into the streptavidin pocket due to its high shape complementarity.

b) Hydrogen bonding involving Asp23, Tyr43, Ser27, Asn49, Ser88, Thr90 and Asp128 residues of streptavidin build up a strong network of interaction with biotin.

c) Streptavidin has lots of aromatic contacts which help to make a hydrophobic network with biotin molecule.

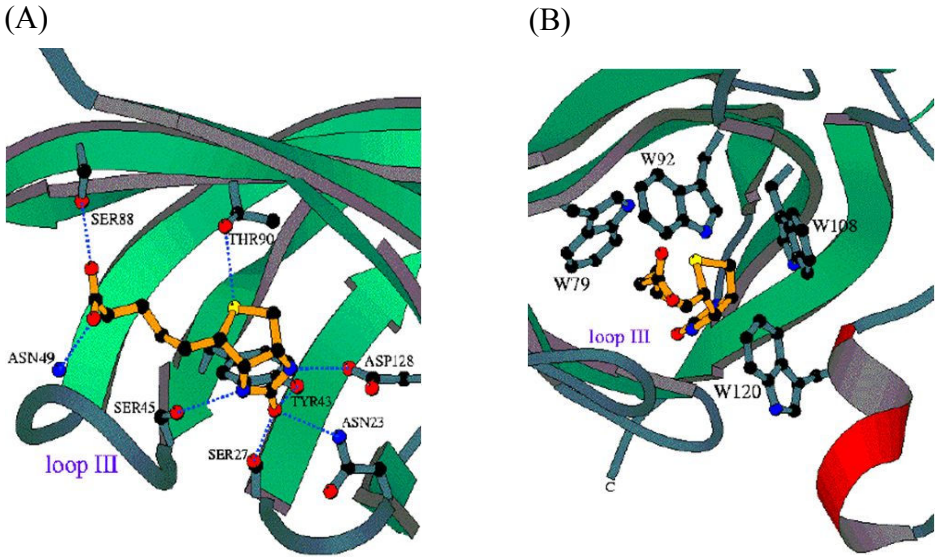


Figure 20. A) Biotin-streptavidin interaction through hydrogen bonding. B) molecular model structure showing tryptophan interaction with biotin.[63]

2.2. Results and discussion

2.2.1 Design of three arm bisbiotin

Thermodynamically, if two ligands of a bivalent molecule are unstrained and independent, the change of enthalpy in the bivalent association (ΔH^{Bi}) is twice as much as that in the monovalent binding (ΔH^{mono}).[64]

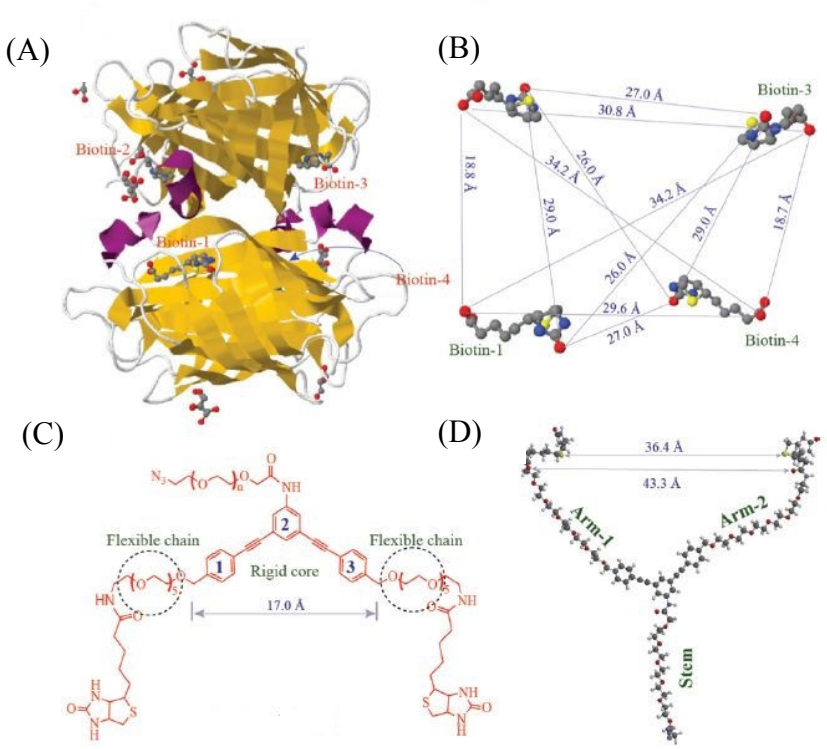


Figure 21. (**A**) A view of the crystal structure of a biotin-streptavidin complex from Protein Data Bank (PDB ID: 3RY2); (**B**) Distances between biotins in the complex, determined by measuring the distance between the ureido oxygen atoms, and between oxygen atoms of the carboxylates in the crystal structure; (**C**) Chemical structure of a three arm bisbiotin ligand; (**D**) A DFT model of three arm bisbiotin, calculated by B3LYP in combination with 6-31G* basis sets in a vacuum using software Spartan'14.

We have implemented these thermodynamic principles in our design of the three arm linker, managing (a) to make a bivalent structure with sufficient flexibility so that its two ligands can bind to the receptor in much the same way as they are in the free state; (b) to maximize the rigidity of the structure to reduce the possible loss in conformational entropy upon binding; (c) to avoid non-specific interactions of the scaffold with the receptor and other non-targeting molecules, which are often driven by cooperative hydrogen bonding, desolvation, or counter-ion release. For example, when DNA is used as a linker, it can interact with proteins through electrostatic attraction, which may become an entropically favorable process due to the release of the counterions,[65,66] resulting in strong nonspecific binding.

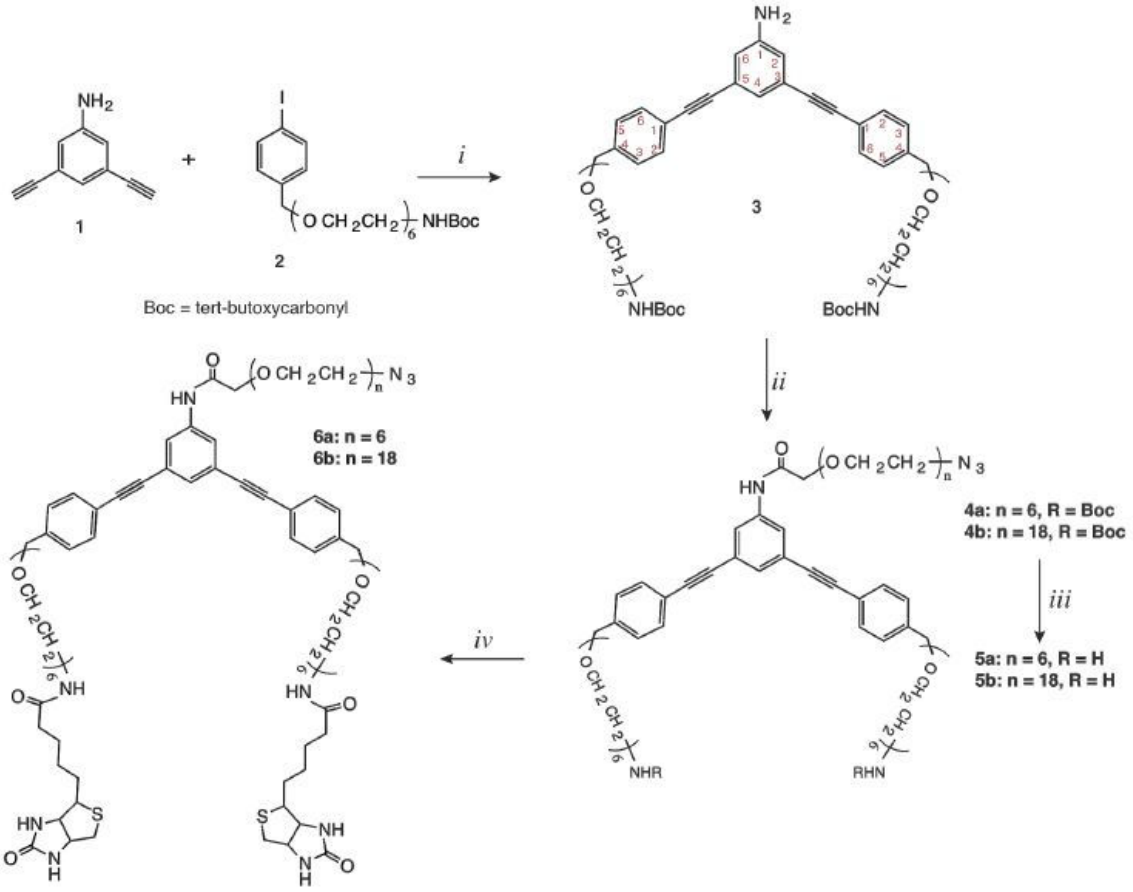
In the present study, we chose the biotin-streptavidin interaction as a model because it is one of the most studied system, which has a well-defined crystal structure (Resolution: 0.95 Å).[67] The streptavidin is a tetrameric molecule, each subunit bearing a biotin binding site.[62] From the projection of the X-ray structure shown in Figure 21-A, three bound biotin molecules can be clearly seen in the streptavidin-biotin complex, indicating that these binding sites are accessible simultaneously by a bi- or tri-valent ligand. This has been demonstrated by Taylor *et al* using trisbiotinylated oligonucleotides with the length of > 20 nucleotides to occupy three of streptavidin's four biotin-binding sites.[68] As shown in Figure 21-B, those biotin molecules in the complex are separated by a range from 19 to 35 Å. Based on these data, we designed the three arm bisbiotin molecule, shown in Figure 21-C. It is composed of a rigid core of 3,5-bis(phenylethynyl)aniline with a width of ~17 Å, which is surrounded with flexible oligo[ethylene glycol] (OEG) linkers. In

turn, two biotin molecules are respectively attached to those OEG linkers situated at the para positions of the phenyl ring 1 and 3, and the third function (azide in the figure) to the aniline fragment also through a flexible OEG linker. The DFT (Density Functional Theory) modeling shows that the structure can display as a Y shape with its two biotins separated by a distance of about 34 Å (Figure 21-D), which matches the distance between biotin 1 and biotin 2 in the crystal structure (see Figure 21-B). We designate the two branches connecting to biotins as Arm 1 and Arm 2, and the third one as the Stem (see Figure 21-D). The short OEG (6 ethylene glycol units) in the structure should allow the biotin moiety to have sufficient freedom for the binding and also improves solubility of the three arm bisbiotin in water.

We believe that the span between the two biotins in the Y-shape structure should be wide enough for them to bind the streptavidin simultaneously. In the present study, we used the stem (Figure 21-D) to immobilize the three arm bisbiotin on an SPR chip for studying the kinetics of the bivalent interaction. Similarly, we attached the three arm bisbiotin molecule to an atomic force microscopy (AFM) tip for studying the bivalent interactions at a single molecule level.

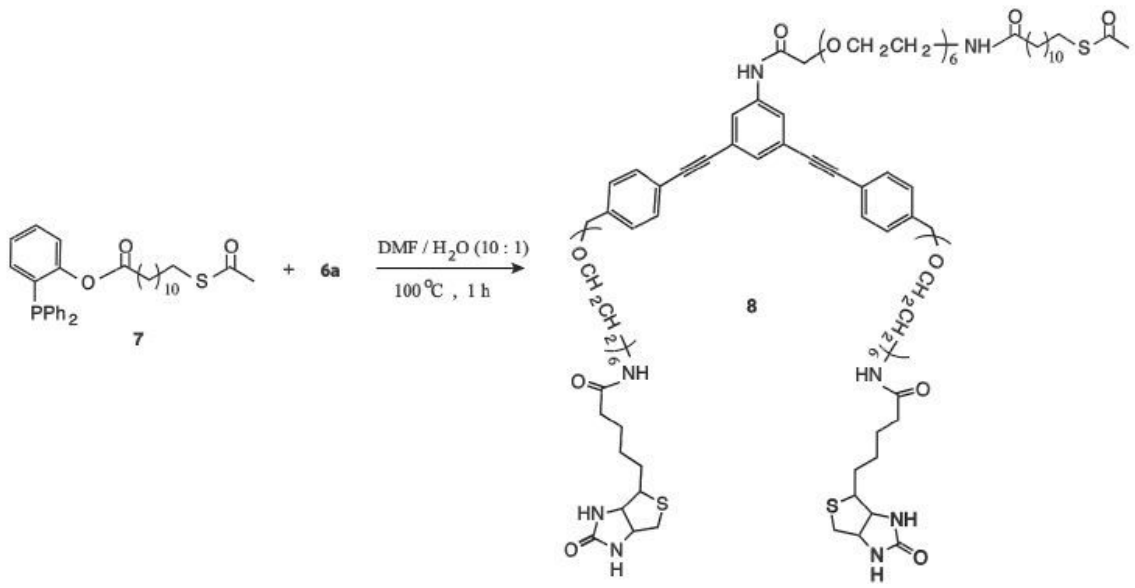
2.2.2 Synthesis of three arm bisbiotin

We have developed a synthetic route to preparing three arm bisbiotin (Scheme 2). Under the Sonogashira cross-coupling conditions,[69,70] 3,5-diethynylaniline (**1**)[71] reacted with *p*-iodobenzyloxy-$(OEG)_6$-NHBoc (**2,** see experimental section for its synthesis), producing compound **3** in a 75% yield. In the presence of pyridine, the arylamine of compound **3** reacted with α-OEGylated acetyl chloride **S7a** and **S7b** (see experimental section for their synthesis) respectively, generating **4a** (59% yield) and **4b** (69% yield), which were subsequently treated with trifluoroacetic acid to remove the Boc protecting groups, resulting in the scaffold molecules **5a** (69% yield) and **5b** (70% yield). The desired three arm bisbiotin ligands **6a** and **6b** were produced via **5a** and **5b** reacting with biotin N-hydroxysuccinimide ester in a yield of ~ 62% and 67%, respectively. The molecules **6a** and **6b** differ in the length of their stems. We studied their interactions with streptavidin by AFM and SPR, respectively.



Scheme 1. Synthesis of three arm bisbiotin ligands
Reagents and conditions: (*i*) $Pd(PPh_3)_2Cl_2$, CuI, THF : Et_3N = 1:1, rt, 5 h; (*ii*) **S7a** or **S7b**, pyridine in dichloromethane, rt, 12 h; (*iii*) trifluoroacetic acid, rt, 10 min; (*iv*) biotin N-hydroxysuccinimide ester, triethylamine, DMF, rt, 3 h

To study the interactions of three arm bisbiotin with streptavidin by Surface Plasmon Resonance (SPR), a ω-thiolated alkyl chain was added to the azido end of **6a** (Scheme 3). We synthesized a triphenyl phosphane 11-(acetylthio)undecanoate reagent **7** (see experimental section) to react with **6a,** furnishing compound **8** containing an acetylated thiol via a traceless Staudinger ligation.[72] In addition, both monobiotin **9** (purchased from Quanta Biodesign) and **10** (see experimental section for its synthesis) were used as



Scheme 2. Thiolation of three arm bisbiotin by Staudinger ligation

controls, and compound **11** (see experimental section for its synthesis) as a spacer for formation of mixed monolayers (Figure 21).

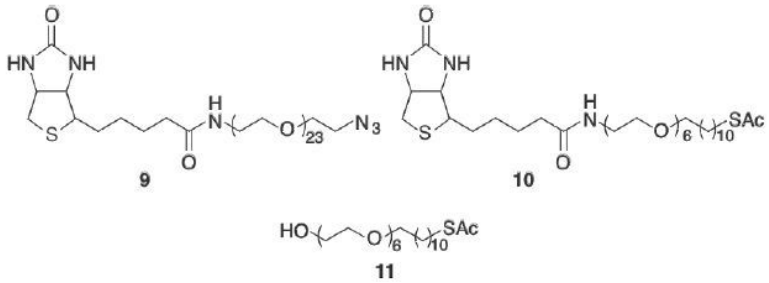


Figure 22. Structures of monobiotin ligands and spacer used in formation of mixed monolayers (Ac = acetyl)

2.2.3 Kinetics of three arm bisbiotin binding to streptavidin

In turn, we have studied the kinetics of three arm bisbiotin interacting with streptavidin by means of SPR. For comparison, we constructed two mixed self-assembled monolayers

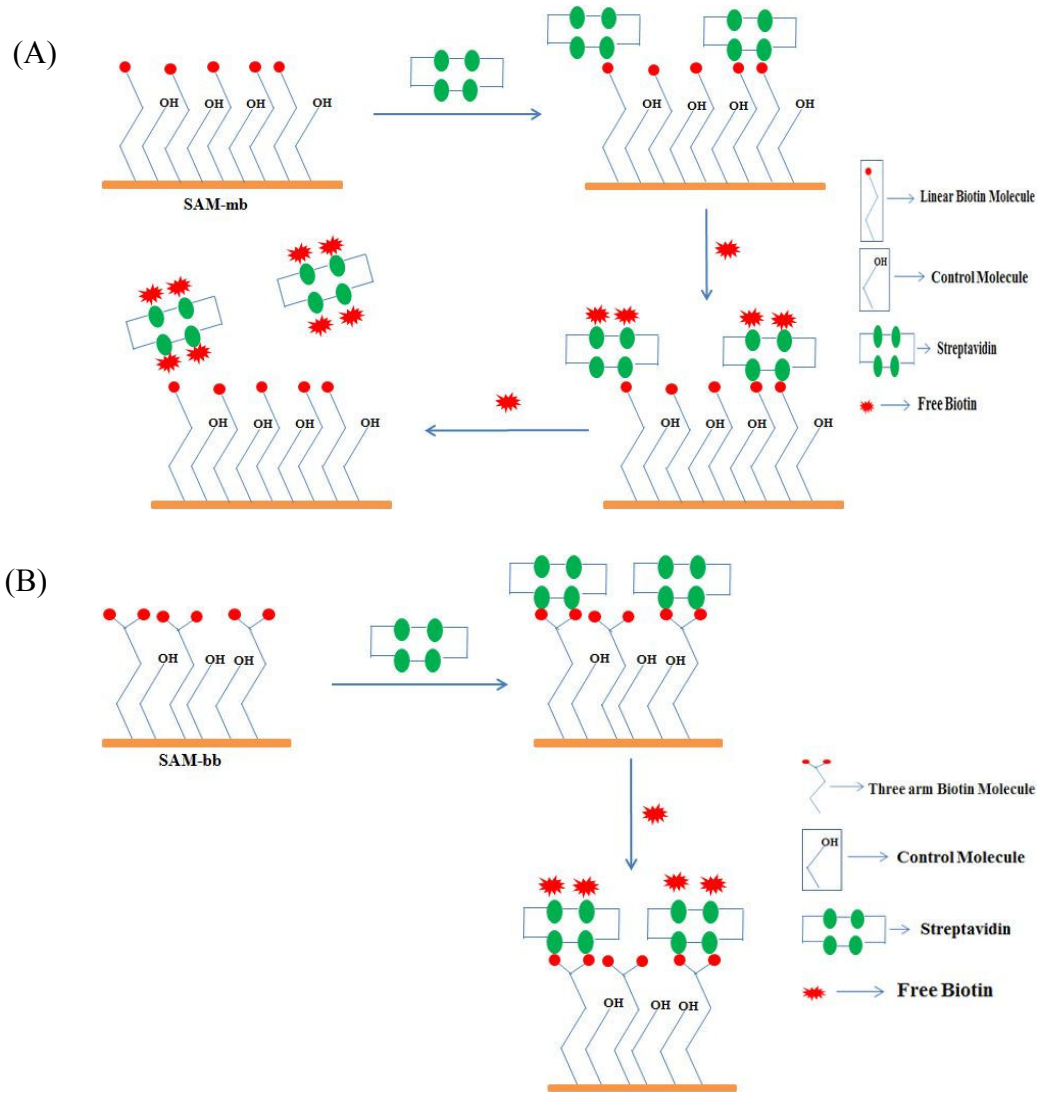


Figure 23. Schematic representation of SPR experiment performed on A) SAM-mb, B) SAM-bb. On both cases, first streptavidin was run over the mixed monolayers to from the biotin-streptavidin complex, followed by flowing excess biotin solution over the complex to check its stability.

on gold substrates (Figure 23): one was composed of the monobiotin **10** with spacer **11** (SAM-mb, illustrated in Figure 23-A), and the other one composed of three arm bisbiotin **8** also with spacer **11** (SAM-bb, illustrated in Figure 23-B). It should be noted that we made these monolayer components with a mercaptoundecyl (11 carbons) chain at their

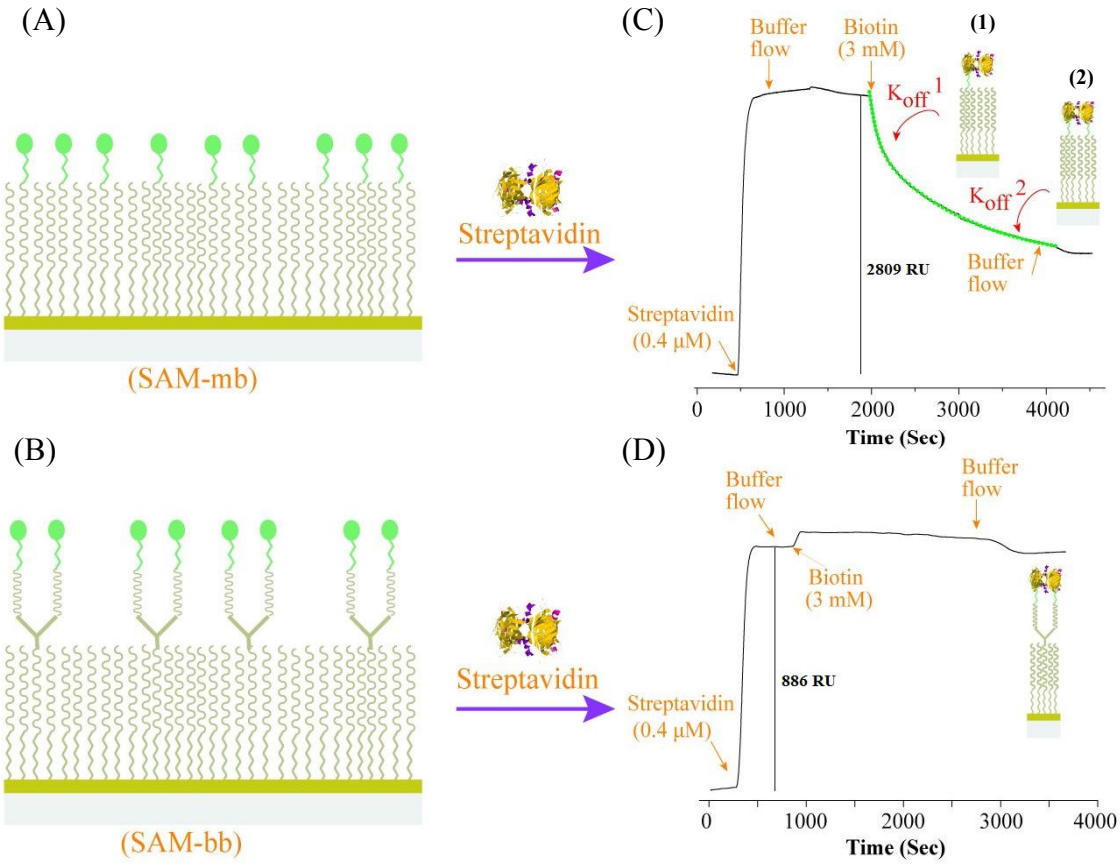


Figure 24. (A) Schematic illustration of a mixed monolayer composed of the monobiotinylated ligand and a molecular spacer; (B) Schematic illustration of a mixed monolayer composed of three arm bisbiotin and a molecular spacer; (C) SPR sensorgram of adsorption and desorption of streptavidin on monobiotinyl formed monolayer; (D) SPR sensorgram of adsorption and desorption of streptavidin on the bisbiotinyl formed monolayer. Red and green dot lines: fitting curves.

ends. Presumably, the odd number of CH_2 units in these molecules would give the monolayer better orientation and higher packing density in comparison to an even number of CH_2 units, due to the odd-even effect.[73] The monolayers were prepared in ethanolic solution, and characterized with contact angle goniometry, ellipsometry, and FTIR (see experimental section). The ratio of biotin to spacer in the solution was adjusted to 1:9, which gave the highest SPR signals on the monolayers than other ratios. First, we examined non-specific absorption on these two monolayers by flowing a solution of streptavidin blocked by native biotins, finding that neither of the two monolayers had measurable non-specific adsorption (data not shown). Then, we measured SPR sensorgrams of streptavidin on these two monolayers. Figure 24-C shows a sensorgram generated on the SAM-mb by flowing a streptavidin solution, followed by a buffer solution, a biotin solution, and then the buffer solution again. Figure 24-D shows the sensorgram generated on the SAM-bb in the same fashion. By comparing between these two sensorgrams, one can notice that the SPR response generated from the streptavidin absorption was about two-thirds stronger on the SAM-mb than on the SAM-bb. This could be because less three arm bisbiotin was incorporated into the mixed monolayer in comparison to the monobiotin under the same reaction conditions and/or the three arm bisbiotin was less reactive to the gold substrate than the monobiotin or the spacer. From the sensorgrams, we were able to extract the kinetic data. First, we determined the association rate constants of the monobiotin at 30 μL/min flow rate by fitting association curve to the Langmuir model, finding that the association rate constant k_{on} 4.9 x 10^4 $M^{-1}s^{-}$

[1]. k_{on} for streptavidin binding to monobiotin is almost as same as to the value for binding to the three arm bisbiotin.

Table 1. Kinetic parameters of adsorption and desorption of streptavidin on biotinylated monolayers

	Flow rate (μL/min)	SAM-mb	SAM-bb
k_{off}^1 (s^{-1})	30	$9.1 \pm 1.5 \times 10^{-3}$	ND
k_{off}^2 (s^{-1})		$7.1 \pm 1.2 \times 10^{-4}$	ND

ND: not able to determine. Green dot line: fitting curve for dissociation rate. Each rate constant is an average of three independent experiments.

Since, desorption of streptavidin from the biotinylated surface was unobservable in a pure buffer in an SPR time frame, we utilized a competition method to accelerate the dissociation process by flowing a 3 mM biotin solution through the monolayer. More than 50% of streptavidin was desorbed from the SAM-mb in 30 minutes. The dissociation curve was fitted into a bi-exponential decay function (green dot line in Figure 24-C),[74] resulting in two dissociation rate constants k_{off}^1 and k_{off}^2 (Table 1). We employ a model proposed by Pérez-Luna et al (see insert **1** and **2** in Figure 24-C) to explain the result.[75] On the SAM-mb, streptavidin formed both monovalent and bivalent complexes with the monobiotin ligand. The monovalent complex would desorb from the monolayer faster than the bivalent one. Accordingly, we assign the k_{off}^1 to streptavidin dissociating from the monovalent complex and k_{off}^2 from the bivalent complex. However, we should point out that the streptavidin formed an intermolecular bivalent complex with the mono biotin ligand on the SAM-mb. Strikingly, there was no desorption of streptavidin on the SAM-bb, even after flowing a biotin solution (3 mM). We believe that the streptavidin formed

an extremely stable intramolecular bivalent complex with three arm bisbiotin on the SAM-bb (see insert in Figure 24-D) because the three arm bisbiotin fits well with the streptavidin structure, resulting in a synergetic interaction.

There is always chance for the surface bound biotin ligand to get rebound to streptavidin pocket in the flow cell. But, for this experiment we were using 3mM of free biotin which was very high concentration compared to the biotin concentration on the surface. Therefore, it was highly unlikely that surface bound three arm bis biotin will be able to rebind to streptavidin after its detachment from one pocket. Also, if we could predict time scale for three arm bis biotin streptavidin from monobiotin dissociation time scale we would come up with fairly large time needed for streptavidin biotin complex dissociation from bis biotin. To check this part, dissociation experiment was performed for couple of hours without any dissociation. Also, excess biotin was run for three times over the streptavidin-biotin complex to see if there was any dissociation due to multiple run. But, no dissociation was observed for three arm bis biotin streptavidin complex. This reinforces the fact that three arm bis biotin form far more stable complex than linear biotin.

2.2.4 AFM measurement of forces to break the three arm bisbiotin-streptavidin complex

Gomez-Casado *et al* has demonstrated that the AFM technique is an effective tool to study the interactions of multivalent adamantane ligands with cyclodextrins immobilized on a surface.[76] Accordingly, we have attempted to measure the force needed to break the "indissociable" three arm bisbiotin-streptavidin complex using AFM at a single molecule

level. As illustrated in Figure 25-A, three arm bisbiotin was attached to AFM tips through a spontaneous reaction of azido function with a cyclooctyne functionalized AFM tip, a process we previously developed for recognition imaging.[77] The streptavidin molecules were immobilized on a mica substrate functionalized with 3-aminopropyltriethoxysilane (APTES)/glutaraldehyde—a process developed in our laboratory for immobilization of proteins (Figure 25-B).[78] To have a better insight into the bivalent interactions, we measured AFM force spectra of both three arm bisbiotin **6b** and the control molecule monobiotin **9** interacting with streptavidin under the same conditions. The AFM force measurements were carried out in a phosphate buffer, pH 7.4, and ~ 1100 force-distance curves collected for each measurement. As shown in Figure 25-C, there are two main rupture events occurring in the force-distance curves, one of which is located within a range of distance from 2 to 7 nm and the other around 12-15 nm. We confirmed that the short distance rupture was due to non-specific interactions of the AFM tip with the surface by blocking. In a typical blocking experiment (performed after collecting force-distance curves), we injected a streptavidin solution into the flow cell, which blocked those

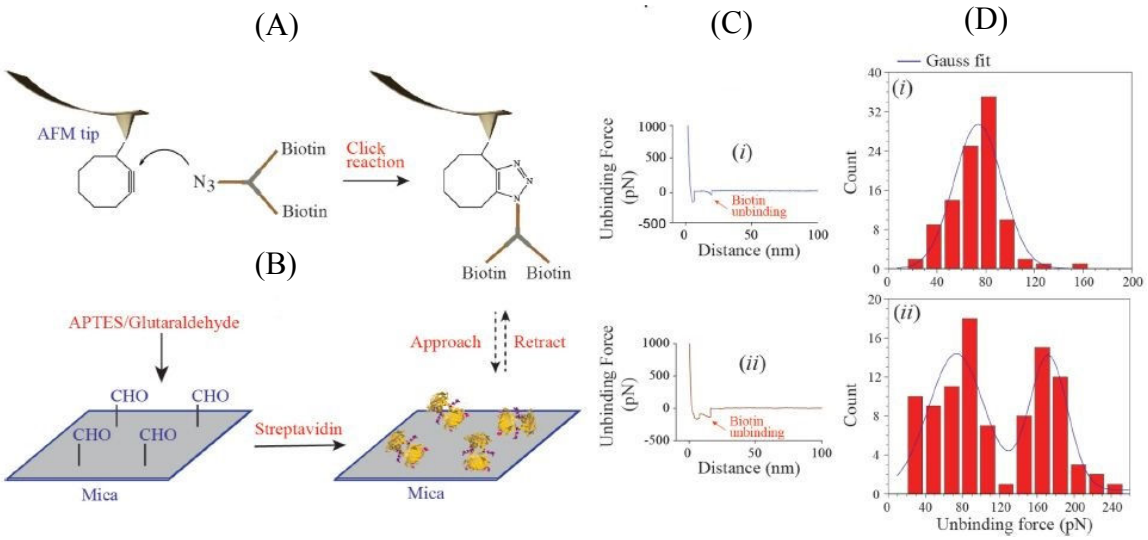


Figure 25. (A) Process of functionalizing AFM tips with a bisbiotin ligand, and (B) immobilizing streptavidin to a mica substrate; (C) A representative unbinding force-distance curve of *i*: monobiotin **9**, and *ii*: three arm bisbiotin **6b** withdrawing from the streptavidin coated mica surface with a loading rate of 1.62 x 10^4 pN/s; (D) Force histogram of *i*: monobiotin **9** and *ii*: three arm bisbiotin **6b** unbinding from streptavidin with their Gaussian fitting curves

biotin molecules on the tip from interacting with the streptavidin on the surface, and then collected another set of force-distance curves. We found that the rupture around the distance of 2-7 nm remained and those around the 12-15 nm disappeared (data not shown). Thus, we assign the ruptures in the range of 12-15 nm, which is close to the contour length of the linker (~11 nm estimated from the computer modeling), as the biotin unbinding from streptavidin. We also found that the ruptures of both monobiotin and three arm bisbiotin from streptavidin appeared as a single step event. On average, about 31% of the force-distance curves generated with a monobiotin-functionalized tip and 27% of those generated with a three arm bisbiotin-functionalized tip contained the specific ruptures, from which we determined the unbinding forces. Figure 25-D shows

the histogram distribution of the unbinding forces, which was analysed by first smoothing with the Savitzky-Golay method[79] and then fitting with a Gaussian distribution. As a result, we obtained a mean value of 73.8±0.4 pN for unbinding of the monobiotin at the loading rate of 1.62 x 10^4 pN/s. This value is close to the most probable force (MPF) of one biotin unbinding from the streptavidin.[80] In contrast, the histogram of the three arm bisbiotin was well fit to two Gaussians (Figure 25-**D**-*ii*). The first peak had a mean value of 73.6±1.7 pN. Therefore, we assign this peak to a monovalent unbinding of the three arm bisbiotin from streptavidin. Although this peak has a much wider force distribution than the monobiotin, it has the same mean value as the monobiotin. The discrepancy between the widths of the distributions may reflect the difference in structure between the three arm bisbiotin and the monobiotin. Notably, the second peak is situated at a mean value of 172.3±1.4 pN, which is twice as large as the rupture force of unbinding one biotin from the streptavidin. The second peak is well separated from the first one. This result indicates that the three arm bisbiotin did bind to streptavidin cooperatively without causing any significant strains.

2.3. Experimental procedures

2.3.1 General information

All chemicals were purchased from commercial suppliers (Sigma-Aldrich, Fluka, Santa Cruz Biotechnology, Alfa Aesar). Anhydrous organic solvents were Sure/Seal™ from Aldrich. Streptavidin and Dulbecco's phosphate buffered saline solution (PBS, pH 7.4) were purchased form Sigma-Aldrich., Azido-dPEG®$_{23}$-biotin was purchased from Quanta Biodesign. All the synthetic reactions were carried out under nitrogen atmosphere. Thin layer chromatography (TLC) was used to monitor progress of organic reactions. An automated flash chromatography system (CombiFlash Rf, Teledyne Isco, Inc.) was used to separate the organic compounds with silica gel columns. FTIR data were collected using Thermo Scientific Nicolet™ 6700 FT-IR spectrometer. The HPLC purification was carried out in Agilent 1100 series equipped with a UV detector and a fraction collector. All the proton NMR (^{1}H) spectra were recorded on a Varian 400 MHz instrument. ^{1}H chemical shifts were referenced relative to the residual solvent peak (such as $CDCl_3$: δ_H = 7.26 ppm). MALDI-TOF analysis was performed on Voyager-DE STR instrument. We used water from Millipore's Milli-Q water purification system with a real time monitor of total of carbon (TOC) connected to a BioPak Polisher to remove biological contaminates. TOC level was strictly maintained below 5 ppb.

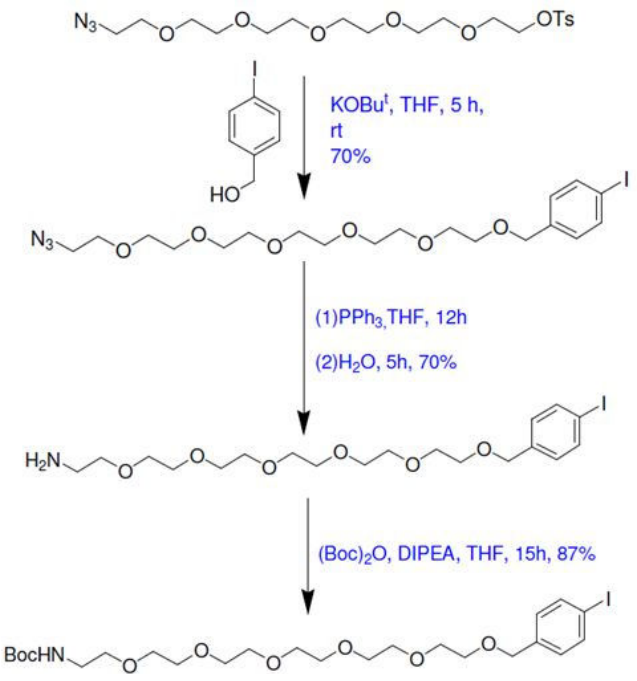
2.3.2 Synthesis

Special note on nomenclature:

To avoid excessive use of a long series of numbers, the mathematical shorthand for expressing arithmetic progressions is used to denote the positions of oxygen atoms in the elongated PEG chains, as proposed by Hill and coworkers.[81]

Synthesis of 1-(tert-butoxycarbonylamino)-17-p-iodobenzyloxy-3n$_{15}$[3] pentaoxaheptadecane (2)

The compound was synthesized following a route as described in Scheme 3.



Scheme 3. Synthesis of compound **2**

1-Azido-17-*p*-iodobenzyloxy-3n$_{15}$^{3}-pentaoxaheptadecane (S2) Potassium *tert*-butoxide (1.26 g, 11.27 mmol) was added to a solution of 4-iodobenzyl alcohol (2.23 g, 9.5 mmol) in anhydrous THF (25 mL) followed by addition of **S1**[82] (4.0 g, 8.6 mmol) and the mixture was stirred at room temperature for 5 h, concentrated by rotary evaporation. The residue was purified by flash chromatography using a gradient of methanol (0-2.5% over 4 h) in dichloromethane to give **S2** as colorless oil (3.17 g, 70%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 3.37 (t, 2H, J = 5.2 Hz), 3.58-3.68 (m, 22H), 4.50 (s, 2H), 7.10 (d, 2H, J = 8 Hz), 7.66 (d, 2 H, J = 8 Hz). ^{13}C NMR (100 MHz, $CDCl_3$): δ 40.3, 69.5, 70.0, 70.5-70.6, 72.4, 129.5, 129.8, 137.4, 138.0. HRMS (APCI) (M+H): found m/z 524.1253, calcd for $C_{19}H_{30+1}IN_3O_6$ 524.1254.

1-Amino-17-*p*-iodobenzyloxy-3n$_{15}$^{3}-pentaoxaheptadecane (S3) To a solution of **S2** (3 g, 5.7 mmol) in anhydrous THF (20 mL), triphenylphosphine (2.25 g, 8.6 mmol) was added and stirred at room temperature for 12 h. Water (1 mL) was added to the mixture, which was stirring for another 4 h, concentrated by rotary evaporation. The crude product was purified by flash chromatography using a gradient of methanol (0-20% over 4 h) in dichloromethane to give **S3** as colorless oil (2 g, 70%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 2.12 (s, 2H), 2.87 (t, 2H, J = 5 Hz), 3.37 (t, 2H, J = 5.2 Hz), 3.58-3.68 (m, 20H), 4.50 (s, 2H), 7.10 (d, 2H, J = 8 Hz), 7.66 (d, 2H , J = 8 Hz); ^{13}C NMR (100MHz, $CDCl_3$): δ 40.4, 69.5, 70.1, 70.5-70.6, 72.4, 129.5, 129.8, 137.4, 138.0. HRMS (APCI) (M+H): found m/z 498.1331, calcd for $C_{19}H_{32+1}INO_6$ 498.1335.

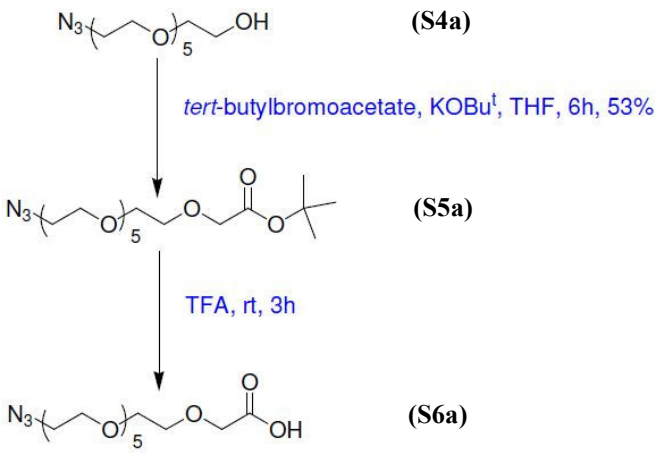
1-(*tert*-butoxycarbonylamino)-17-*p*-iodobenzyloxy-3n$_{15}$^{3}-pentaoxaheptadecane (2) To a solution of **S3** (2.0 g, 4 mmol) in anhydrous THF (15 mL), di-*tert*-butyl dicarbonate

(2.52 mL, 11 mmol) was added followed by addition of diisopropylamine (0.67 mL, 4.8 mmol). The solution was stirred at room temperature for 15 h, concentrated by rotary evaporation. The residue was purified by flash chromatography using a gradient of methanol (0-5% over 4 h) in dichloromethane to give **2** as colorless oil (2.1 g, 87%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.44 (s, 9H), 3.28-3.32 (m, 2H), 3.48 (t, 2H, J = 5.2 Hz), 3.49-3.68 (m, 20H), 4.50 (s, 2H), 5.11 (s, 1H), 7.10 (d, 2H, J = 8 Hz), 7.66 (d, 2 H, J = 8 Hz); ^{13}C NMR (100 MHz, $CDCl_3$): δ 28.4, 40.3, 68.9, 69.5, 70.1, 70.5-70.6, 72.4, 129.5, 129.8, 137.4, 138.0, 155.9. HRMS (APCI) (M+H): found m/z 598.1863, calcd for $C_{24}H_{40+1}INO_8$ 598.1867.

Synthesis of 20-azido-3n$_{18}^{3}$-hexaoxaicosan-1-oyl chloride (S7a)

The compound was synthesized following a route as described in Scheme 4.

***Tert*-butyl 20-azido-3n$_{18}^{3}$-hexaoxaicosan-1-oate (S5a)** To a solution of compound **S4a**[82] (2 g, 6.5 mmol) in anhydrous THF (10 mL), potassium *tert*-butoxide (0.54 g, 7.8 mmol)



Scheme 4. Synthesis of compound **S6a**

was added followed by addition of *tert*-butyl bromoacetate (0.65 mL, 8.8 mmol). The mixture was stirred at room temperature for 5 h, and concentrated by rotary evaporation. The residue was purified by flash chromatography using a gradient of methanol (0-3.3% over 4 h) in dichloromethane to give **S5a** as colorless oil (1.2 g, 52%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.43 (m, 9H), 3.36 (t, 2 H, J = 5.2 Hz), 3.53-3.68 (m, 22H), 3.90 (s, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 28.03, 50.6, 68.9, 69.5, 70.0, 70.5-70.6, 72.4, 164.5. HRMS (APCI) (M+H): found m/z 422.3463, calcd for $C_{18}H_{35+1}N_3O_8$ 422.3465.

20-azido-3n$_{18}^{3}$-hexaoxaicosan-1-oic acid (S6a) Compound **S5a** (1.0 g, 2.4 mmol) was dissolved in 10 mL of trifluoroacetic acid and stirred for 4 h at room temperature, and concentrated by rotary evaporation. The residue was purified by flash chromatography using a gradient of methanol (0-6% over 3 h) in dichloromethane to give **S6a** as colorless oil (0.62 g, 73%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 3.37 (t, 2 H, J = 5.2 Hz), 3.53-3.69 (m, 22H), 3.93 (s, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 50.6, 69.1, 69.5, 70.0, 70.5-70.6, 72.4, 166.8. HRMS (APCI) (M+H): found m/z 366.2754, calcd for $C_{14}H_{27+1}N_3O_8$ 366.2759.

Synthesis of 20-azido-3n$_{18}^{3}$-hexaoxaicosan-1-oyl chloride (S7a) S6a (0.5 g, 1.2 mmol) was dissolved in thionyl chloride (2 mL, 2.7 mmol), stirred for 6 h at room temperature, and concentrated by rotary evaporation to furnish **S7a** as yellowish oil.

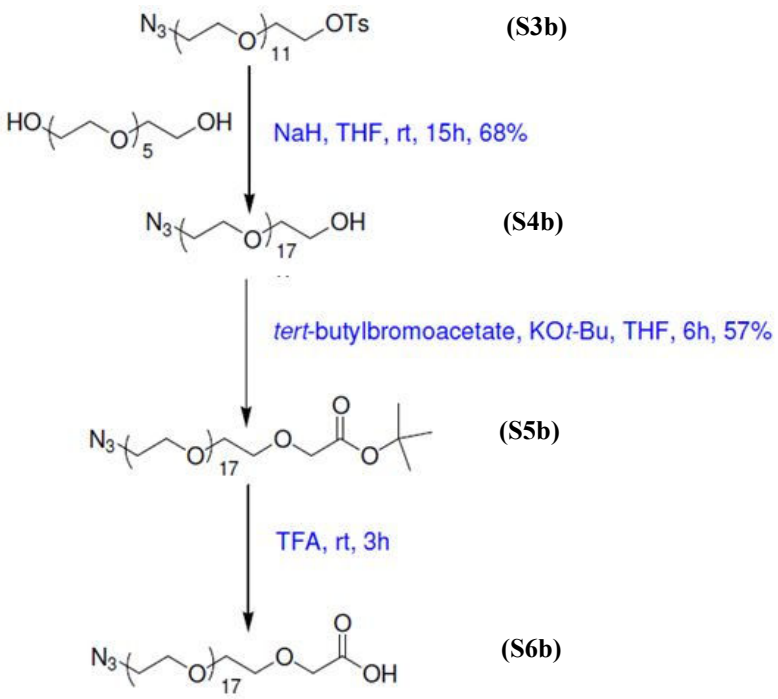
56-azido-3n$_{54}^{3}$-octadecaoxahexapentacontan-1-oyl chloride (S7b)

The compound was synthesized following a route as described in Scheme 6.

53-azido-3n$_{51}^{3}$-heptadecaoxatripentacontan-1-ol (S4b)

The starting material **S3b** was synthesized following a procedure we previously reported.[82]

To a solution of hexaethylene glycol (6.72 g, 23.8 mmol) in anhydrous THF (40 mL), Sodium hydride (0.74g, 30.8 mmol) was added and stirred at 0 °C for 1 h. Compound **S3b** (3.5 g, 7.5 mmol) in anhydrous THF (20 mL) was added to the mixture and allowed to warm to room temperature, stirred for another 15 h. Methanol (5 mL) was added



Scheme 5. Synthesis of compound **S6b**

dropwise to stop the reaction. The mixture was concentrated by rotary evaporation, purified in a silica gel column by flash chromatography using a gradient of methanol (0-3.5% over 4 h) in dichloromethane. Compound **S4b** was obtained as a colorless liquid (2.9 g, 72%). ^{1}H NMR (400 MHz, CDCl3): δ 2.69 (s, 1 H, broad), 3.36 (t, 2 H, J = 5.2

Hz), 3.53- 3.68 (m, 70 H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 69.5, 70.1, 70.5-70.6, 72.4. HRMS (APCI) (M+H): found m/z 836.5330, calcd for $C_{36}H_{73+1}N_3O_{18}$ 836.5332.

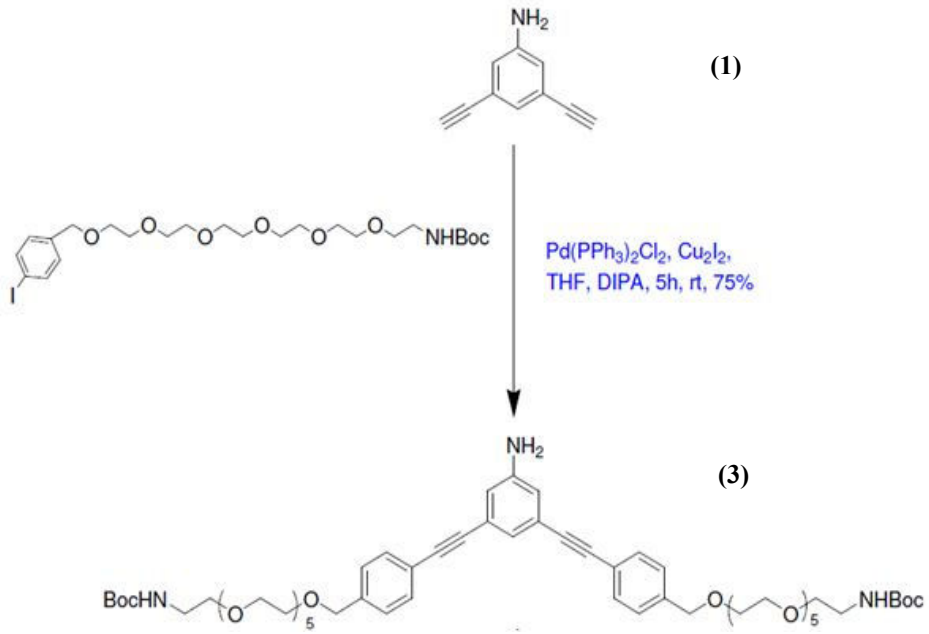
***Tert*-butyl 56-azido-3n$_{54}^3$-octadecaoxahexapentacontan-1-oate (S5b)** To a solution of compound **S4b** (2 g, 2.3 mmol) in anhydrous THF (10 mL), potassium *tert*-butoxide (0.34 g, 3.1 mmol) was added followed by addition of *tert*-butyl bromoacetate (0.45 mL, 3.1 mmol). The mixture was stirred at room temperature for 5 h, and concentrated by rotary evaporation. The residue was purified by flash chromatography using a gradient of methanol (0-3.3% over 4 h) in dichloromethane to give **S5b** as colorless oil (1.3 g, 57%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.43 (m, 9H), 3.36 (t, 2 H, J = 5.2 Hz), 3.54-3.68 (m, 70H), 3.91 (s, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 28.4, 50.6, 69.0, 69.5, 70.0, 70.5-70.6, 72.4, 164.5. HRMS (APCI) (M+H): found m/z 950.6191, calcd for $C_{42}H_{83+1}N_3O_{20}$ 950.6192.

56-Azido-3n$_{54}^3$-octadecaoxahexapentacontan-1-oic acid (S6b) Compound **S5b** (1.0 g, 1 mmol) was dissolved in 10 mL of trifluoroacetic acid and stirred for 4 h at room temperature, and concentrated by rotary evaporation. The residue was purified by flash chromatography using a gradient of methanol (0-6% over 3 h) in dichloromethane to give **S6b** as colorless oil (0.65 g, 69%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 3.37 (t, 2 H, J = 5.2 Hz), 3.54-3.69 (m, 70H), 3.92 (s, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 50.6, 68.9, 69.5, 70.0, 70.5-70.6, 72.4, 166.7. HRMS (APCI) (M+H): found m/z 894.5636, calcd for $C_{38}H_{75+1}N_3O_{20}$ 894.5637

56-azido-3n$_{54}^{3}$-octadecaoxahexapentacontan-1-oyl chloride (S7b) S6b (0.5 g, 0.55 mmol) was dissolved in thionyl chloride (2 mL, 2.7 mmol), stirred for 6 h at room temperature, concentrated by rotary evaporation to furnish **S7b** as yellowish oil.

3,5-Bis{4-[19-*tert*-butoxycarbonylamino-(3n$_{18}^{3}$-1) hexaoxanonadecyl]phenylethynyl}aniline (3)

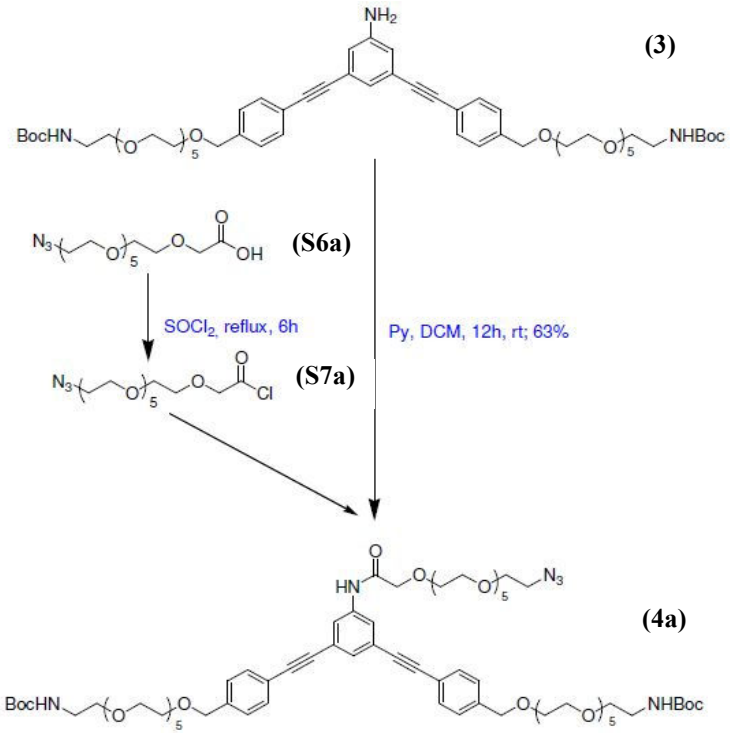
To a solution of compound **2** (2.6 g, 4.4 mmol) in THF (20 mL), 3,5-diethynylaniline **1**[71] (0.3 g, 2.1 mmol), bis(triphenylphosphine)palladium(II)dichloride (0.074 g, 0.1 mmol) and copper(I) iodide (0.040 g, 0.2 mmol) were added. The mixture was stirred at room temperature for 5 h, filtered to remove the precipitate, concentrated by rotary



Scheme 6. Synthesis of compound **3**.

evaporation. The residue was purified by flash column chromatography on silica gel using a gradient of methanol (0-2.8% over 4 h) in dichloromethane as an eluent to give **3** as a colorless oil (1.72 g, 75%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.43 (s, 18H), 3.27-3.32 (m, 4H), 3.52 (t, 4H, J = 5.2 Hz), 3.58-3.69 (m, 44 H), 4.56 (s, 4H), 5.13 (s, 2H), 6.81 (s, 2H), 7.10 (s, 1H), 7.32 (d, 4H, J = 8 Hz), 7.47 (d, 4H, J = 8 Hz); ^{13}C NMR (100 MHz, $CDCl_3$): δ 28.3, 40.3, 69.5, 70.1, 70.3-70.6, 72.7, 88.9, 90.0, 117.7, 122.2, 124.0, 124.9, 127.4, 131.5, 138.6, 146.5, 155.9; HRMS (APCI) (M+H): found m/z 1080.6002, calcd for $C_{58}H_{85+1}N_3O_{16}$ 1080.6001.

N-(3,5-Bis{4-[19-*tert*-butoxycarbonylamino-(3n$_{18}^{3}$-1)-hexaoxanonadecyl]phenylethynyl}aniline)-20-azido-3n$_{18}^{3}$-icosan-1-amide (4a)



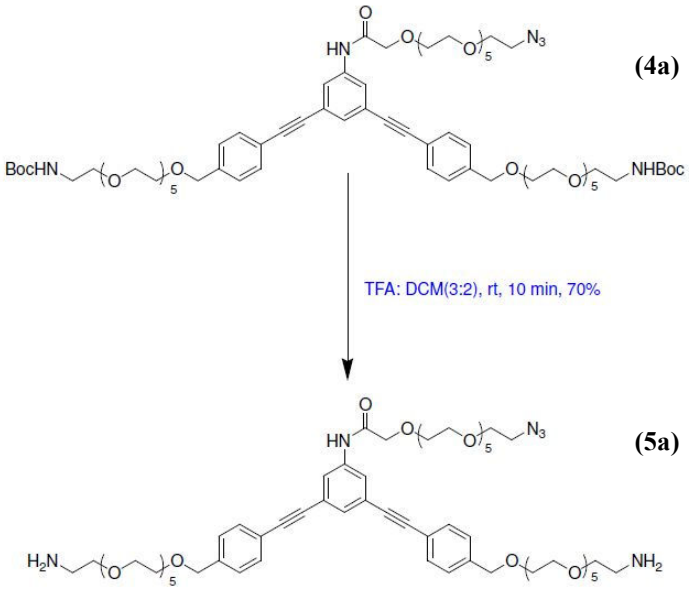
Scheme 7. Synthesis of compound **4a**

Compound **S7a** dissolved in anhydrous dichloromethane (1 mL) was added to a solution of compound **3** (1.4 g, 1.27 mmol) in anhydrous dichloromethane (10 mL) and pyridine (98 μL, 1.2 mmol). The solution was stirred for 12 h at room temperature, concentrated by rotary evaporation. The residue was purified by flash column chromatography on silica gel using a gradient of methanol (0-4% over 4 h) in dichloromethane as an eluent to give **4a** as a colorless oil (1.1 g, 59%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.43 (s, 18H), 3.21-3.25 (m, 4H), 3.34 (t, 6H, J = 5.2 Hz), 3.51-3.78 (m, 62 H), 4.13 (s, 2H), 4.56 (s, 4H), 5.13 (s, 2H), 6.81 (s, 2H), 7.12 (s, 1H), 7.33 (d, 4H, J = 8 Hz), 7.47 (d, 4H, J = 8

Hz), 8.85 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 28.3, 40.3, 50.6, 69.4, 69.7, 69.9, 70.0, 70.4-70.6, 72.7, 88.9, 90.0, 117.7, 122.3, 124.1, 124.9, 127.4, 131.6, 138.6, 146.5, 155.9, 168.8. MALDI-MS (M+H): found m/z 1427.65, calcd for $C_{72}H_{110+1}N_6O_{23}$ 1427.67.

N-(3,5-Bis{4-[19-amino-(3n$_{18}^3$-1)-hexaoxanonadecylphenyl]ethynyl}aniline)-20-Azido-3n$_{18}^3$-icosan-1-amide (5a)

Compound **4a** (0.50 g, 0.35 mmol) was dissolved in a solution of TFA (0.3 mL, 3.3 mmol) and anhydrous dichloromethane (0.2 ml). The solution was stirred for 10 minutes at room temperature, and concentrated by rotary evaporation. The residue was purified by flash column chromatography using a gradient of methanol (0-13% over 4 h) in dichloromethane as an eluent to give **5a** as colorless oil (0.30 g, 69%). ^{1}H NMR (400

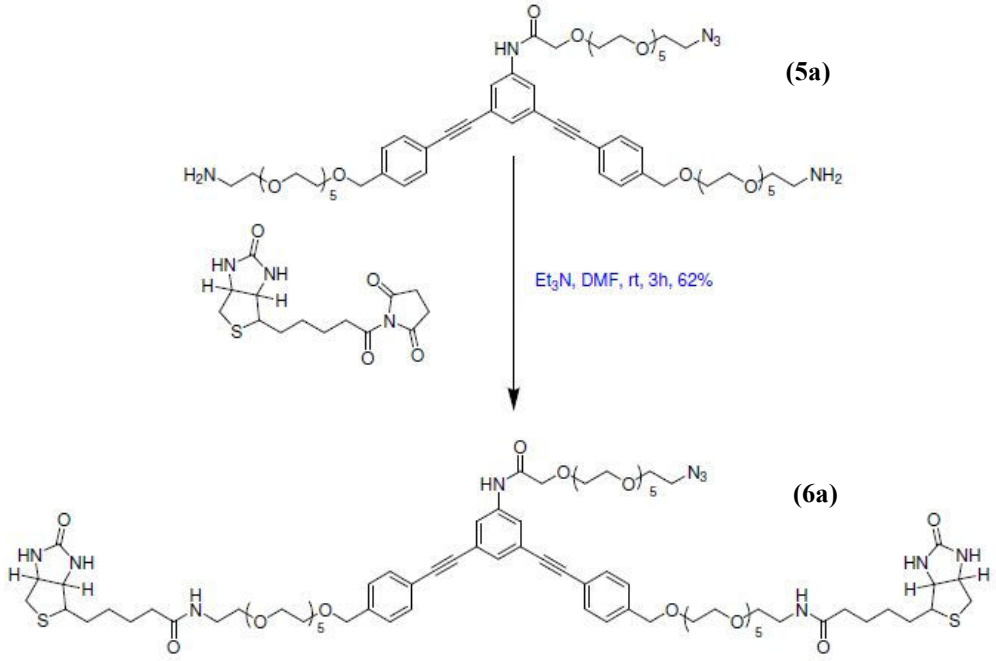


Scheme 8. Synthesis of compound **5a**

MHz, $CDCl_3$): δ 2.5 (s, 4H), 3.20-3.25 (m, 4H), 3.34 (t, 6H, J = 5.2 Hz), 3.50-3.77 (m, 62 H), 4.13 (s, 2H), 4.56 (s, 4H), 6.80 (s, 2H), 7.11 (s, 1H), 7.32 (d, 4H, J = 8 Hz), 7.47 (d, 4H, J = 8 Hz), 8.89 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 40.3, 50.6, 69.4, 69.7, 69.9, 70.0, 70.4-70.6, 72.7, 88.9, 90.0, 117.7, 122.3, 124.1, 124.9, 127.4, 131.6, 138.6, 146.5, 168.8; MALDI-MS (M+H): found m/z 1227.40, calcd for $C_{62}H_{94+1}N_6O_{19}$ 1227.44.

N-(3,5-Bis{4-[19-biotinylated-amino-(3n$_{18}^3$-1)-hexaoxanonadecyl]phenylethynyl}aniline))-20-Azido-3n$_{18}^3$-icosan-1-amide (6a)

A solution of **5a** (0.20 g, 0.16 mmol), NHS-Biotin (0.116 g, 0.34 mmol), and

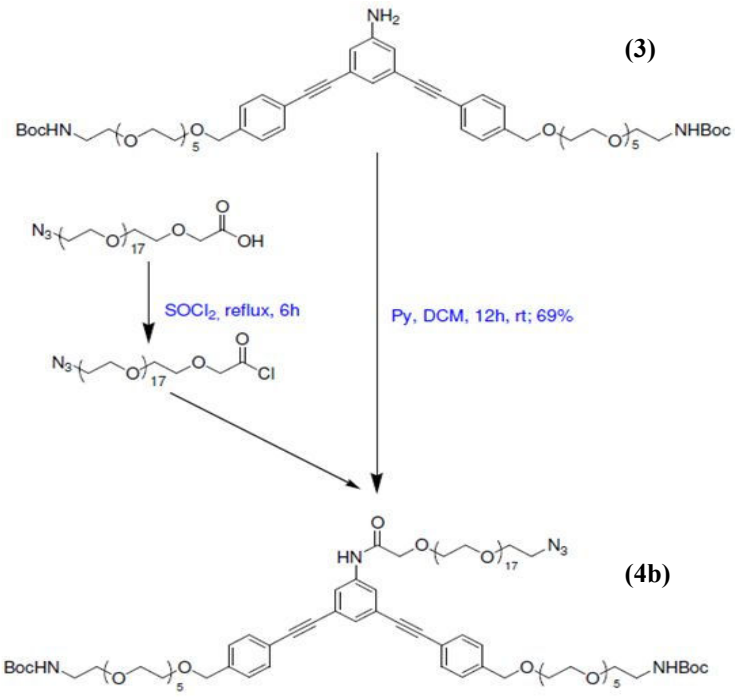


Scheme 9. Synthesis of compound **6a**

triethylamine (0.047 mL, 0.34 mmol) in anhydrous DMF (2 mL) was stirred at room temperature for 2.5 h, and concentrated by rotary evaporation. The residue was purified

by flash column chromatography on silica gel using a gradient of methanol (0-5% over 4 h) in dichloromethane as an eluent to give **6a** as a colorless oil (0.17 g, 62%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.30-1.80 (m, 12H), 2.21 (t, 4H, J = 7.1 Hz), 2.72 (d, 2H, J = 12.5 Hz), 2.86 (dd, 2H, J = 4.8 Hz, J = 12.5 Hz), 3.13-3.17 (m, 2H), 3.21-3.26 (m, 4H), 3.33 (t, 6H, J = 5.2 Hz), 3.52-3.78 (m, 62 H), 4.13 (s, 2H), 4.31-4.35 (m, 2H), 4.49-4.53 (m, 2H), 4.58 (s, 4H), 5.75 (s, 2H), 6.45 (s, 2H), 6.50 (s, 2H), 6.81 (s, 2H), 7.10 (s, 1H), 7.32 (d, 4H, J = 8 Hz), 7.47 (d, 4H, J = 8 Hz), 8.91 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 25.5, 28.0, 28.1, 35.8, 39.1, 40.3, 50.6, 55.5, 60.1, 61.7, 69.3, 69.7, 69.9, 70.0, 70.4-70.6, 72.7, 88.9, 90.0, 117.7, 122.3, 124.1, 124.8, 127.4, 131.6, 138.6, 146.5, 163.9, 168.7, 173.5; MALDI-MS (M+H): found m/z 1679.76, calcd for $C_{82}H_{122+1}N_{10}O_{23}S_2$ 1679.81.

N-(3,5-Bis{4-[19-*tert*-butoxycarbonylamino-(3n$_{18}^{3}$-1)-hexaoxanonadecyl]phenylethynyl}aniline)-56-azido-3n$_{54}^{3}$-octadecaoxahexapentacontan-1-amide (4b)

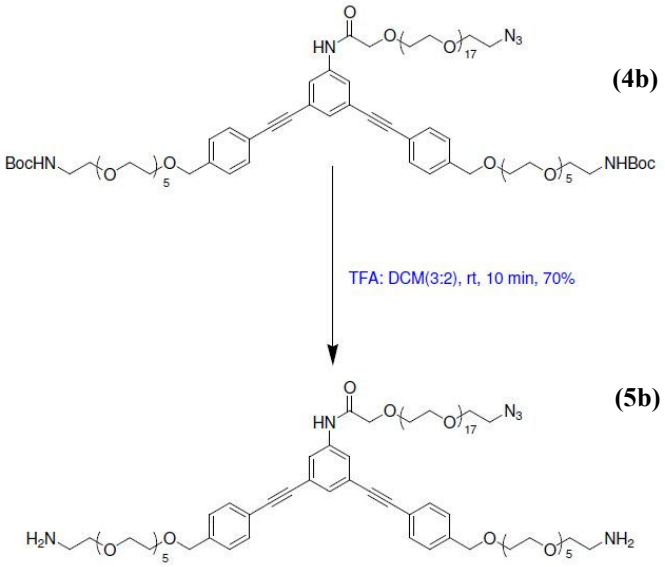


Scheme 10. Synthesis of compound **4b**

Compound **S7b** (see S2 in Supporting Information for its synthesis) dissolved in anhydrous dichloromethane (0.5 mL) was added to a solution of compound **3** (0.66 g, 0.6 mmol) in anhydrous dichloromethane (0.5 mL) and pyridine (98 μL, 1.2 mmol). The solution was stirred for 12 h at room temperature, and then concentrated by rotary evaporation. The residue was purified by flash column chromatography on silica gel using a gradient of methanol (0-4% over 4 h) in dichloromethane as an eluent to give **4b** as a colorless oil (0.76 g, 69%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.43 (s, 18H), 3.20-3.25

(m, 4H), 3.34 (t, 6H, J = 5.2 Hz), 3.52-3.78 (m, 110 H), 4.13 (s, 2H), 4.56 (s, 4H), 5.13 (s, 2H), 6.80 (s, 2H), 7.12 (s, 1H), 7.32 (d, 4H, J = 8 Hz), 7.47 (d, 4H, J = 8 Hz), 8.85 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 28.3, 40.3, 50.6, 69.4, 69.7, 69.9, 70.0, 70.3-70.6, 72.7, 88.9, 90.0, 117.7, 122.3, 124.1, 124.9, 127.4, 131.6, 138.6, 146.5, 155.9, 168.9. MALDI-MS (M+H): found m/z 1957.36, calcd for $C_{96}H_{158+1}N_6O_{35}$ 1957.31.

N-(3,5-Bis{4-[19-amino-(3n$_{18}^3$-1)-hexaoxanonadecyl]phenylethynyl}aniline)-56-Azido-3n$_{54}^3$-octadecaoxahexapentacontan-1-amide (5b)



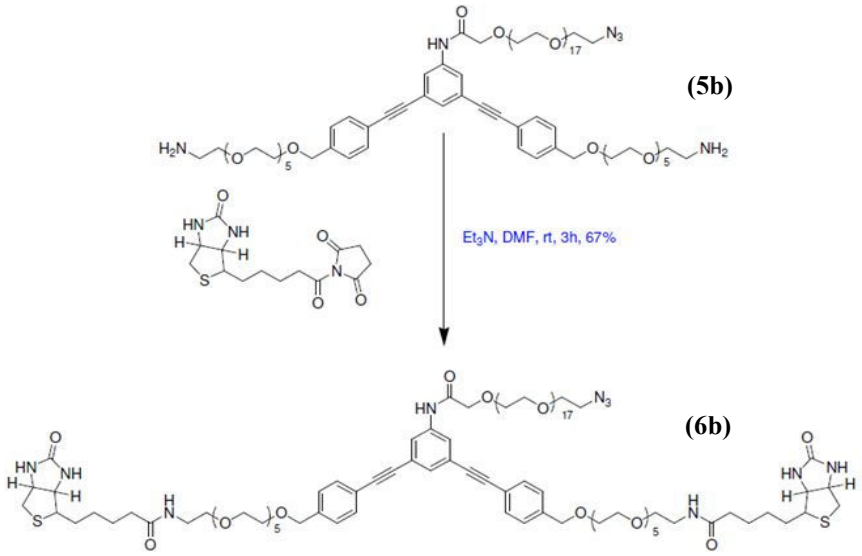
Scheme 11. Synthesis of compound **5b**

Compound **4b** (0.7 g, 0.35 mmol) was dissolved in a solution of TFA (0.3 ml) and anhydrous dichloromethane (0.2 ml). The solution was stirred for 10 minutes at room temperature, and then concentrated by rotary evaporation. The residue was purified by flash column chromatography using a gradient of methanol (0-13% over 4 h) in dichloromethane as an eluent to give **5b** as colorless oil (0.44 g, 70%). ^1H NMR (400

MHz, $CDCl_3$): δ 2.5 (s, 4H), 3.20-3.25 (m, 4H), 3.34 (t, 6H, J = 5.2 Hz), 3.50-3.77 (m, 110 H), 4.13 (s, 2H), 4.56 (s, 4H), 6.81 (s, 2H), 7.11 (s, 1H), 7.32 (d, 4H, J = 8 Hz), 7.47 (d, 4H, J = 8 Hz), 8.89 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 40.3, 50.6, 69.4, 69.7, 69.9, 70.0, 70.3-70.6, 72.7, 88.9, 90.0, 117.7, 122.3, 124.1, 124.9, 127.4, 131.6, 138.6, 146.5, 168.9; MALDI-MS (M+H): found m/z 1757.16, calcd for $C_{86}H_{142+1}N_6O_{31}$ 1757.07.

N-(3,5-Bis{4-[19-biotinylated-amino-(3n$_{18}^3$-1)-hexaoxanonadecyl]phenylethynyl}aniline)-56-Azido-3n$_{54}^3$-octadecaoxahexapentacontan-1-amide (6b)

To a solution of **5b** (0.2 g, 0.11 mmol), NHS-Biotin (0.116 g, 0.34 mmol), and triethylamine (0.047 mL, 0.34 mmol) in anhydrous DMF (2 mL) was stirred at room

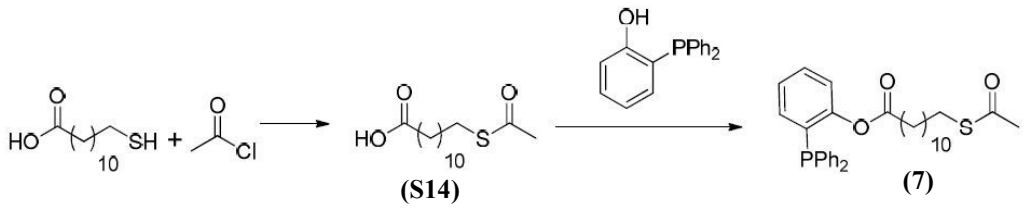


Scheme 12. Synthesis of compound **6b**

temperature for 2.5 h, and concentrated by rotary evaporation. The residue was purified by flash column chromatography on silica gel using a gradient of methanol (0-5% over 4 h) in dichloromethane as an eluent to give **6b** as a colorless oil (0.17 g, 67%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.30-1.80 (m, 12H), 2.21 (t, 4H, J = 7.1 Hz), 2.72 (d, 2H, J = 12.5 Hz), 2.86 (dd, 2H, J = 4.8 Hz, J = 12.5 Hz), 3.13-3.17 (m, 2H), 3.21-3.26 (m, 4H), 3.33 (t, 6H, J = 5.2 Hz), 3.52-3.78 (m, 110 H), 4.13 (s, 2H), 4.30-4.35 (m, 2H), 4.49-4.52 (m, 2H), 4.58 (s, 4H), 5.75 (s, 2H), 6.45 (s, 2H), 6.50 (s, 2H), 6.81 (s, 2H), 7.09 (s, 1H), 7.32 (d, 4H, J = 8 Hz), 7.47 (d, 4H, J = 8 Hz), 8.90 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 25.5, 28.0, 28.1, 35.8, 39.1, 40.3, 50.6, 55.5, 60.1, 61.7, 69.3, 69.7, 69.9, 70.0, 70.3-70.6, 72.7, 88.9, 90.0, 117.7, 122.3, 124.1, 124.8, 127.4, 131.6, 138.6, 146.5, 163.9, 168.8, 173.4; MALDI-MS (M+H): found m/z 2209.52, calcd for $C_{106}H_{170+1}N_{10}O_{35}S_2$ 2209.66.

Synthesis of triphenyl phosphane 11-(acetylthio)undecanoate (7) Compound **7** was synthesized following the route in scheme **13**.

S14 was synthesized following literature procedure.[83] ^{1}H NMR (400 MHz, $CDCl_3$): δ



(S14) **(7)**

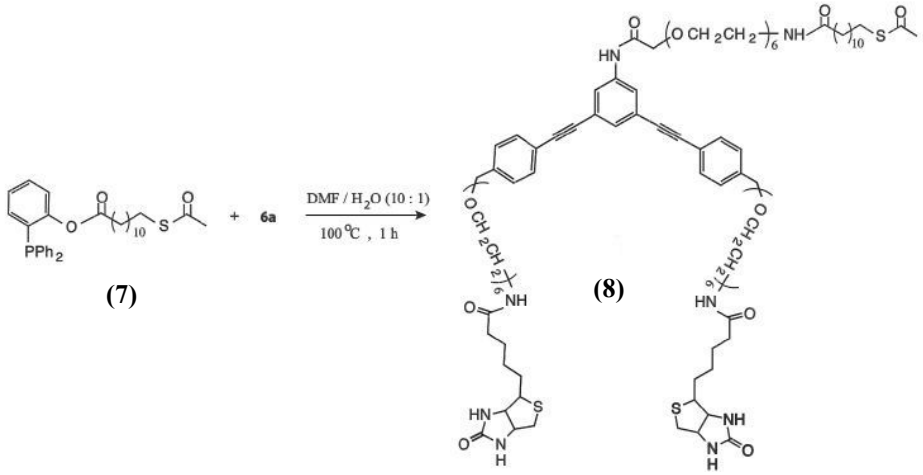
Scheme 13. Synthesis of compound **7**

1.22-1.47 (m, 14H), 1.52-1.63 (m, 4H), 2.31 (s, 3H), 2.34 (t, 2H, J = 5.2 Hz), 2.84 (t, 2H, J = 7.6 Hz). ^{13}C NMR (100 MHz, $CDCl_3$): δ 26.4, 28.9, 29.2, 29.6, 29.8-29.9, 32.5, 34.2, 39.5, 165.3, 195.0. HRMS (APCI) (M+H): found m/z 274.1346, calcd for $C_{13}H_{25+1}O_3S$ 274.1355.

To a stirred solution of **S14** (0.5 g, 1.8 mmol) in 5 mL anhydrous DMF, EDC (0.5 g, 2.5 mmol) and catalytic amount of DMAP (0.010 g, 0.01 mmol) were added and stirred at 0 °C for 30 minutes. To this solution, (2-Hydroxyphenyl)diphenylphosphine (0.52 g, 1.8 mmol) in 5 mL anhydrous DMF was added and stirred overnight. Then the solvent was evaporated and the residue was purified by flash column chromatography on silica gel using a gradient of methanol (0 - 5% over 4 h) in dichloromethane as an eluent to give compound **7** (0.7 g, 63%)**.** ^{1}H NMR (400 MHz, CDCl3): δ 1.22-1.47 (m, 14H), 1.52-1.63 (m, 4H), 2.31 (s, 3H), 2.34 (t, 2H, J = 7.6 Hz), 2.86 (t, 2H, J = 7.2 Hz), 6.76-6.84 (m, 1H), 7.06-7.16 (m, 2H), 7.23-7.40 (m, 11H). ^{13}C NMR (100 MHz, CDCl3): δ 26.4, 28.9, 29.2, 29.6, 29.8-29.9, 32.5, 34.2, 39.5, 122.4, 126.3, 128.4, 128.7, 129.1, 130.2, 130.3, 133.5, 133.8, 134.2, 135.6, 135.7, 164.3, 195.1. HRMS (APCI) (M+H): found m/z 535.6918, calcd for $C_{32}H_{39+1}O_3PS$ 535.6915.

N-(3,5-Bis{4-[19-biotinylated-amino-(3n$_{18}^{3}$-1)-hexaoxanonadecyl]phenylethynyl}aniline)-20-(11-acetylthioundecanamido)-3n$_{18}^{3}$-hexaoxaicosan-1-amide (8)

A solution of **7** (0.065 g, 0.13 mmol) and **6a** (0.2 g, 0.1 mmol) in DMF (1.5 mL) and water (500 μL) was stirred at 100 °C for 1 h, and concentrated by rotary evaporation. The residue was separated by flash column chromatography on silica gel using a gradient of methanol (0-4.5% over 4h) in dichloromethane as an eluent to give **8** as a colorless oil (35 mg, 87 %). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.26-1.79 (m, 30H), 2.21 (t, 6H, J = 7.1 Hz), 2.26 (s, 3H), 2.32 (t, 2H, J = 5.2 Hz), 2.72 (d, 2H, J = 12.5 Hz), 2.86 (dd, 2H, J = 4.8 Hz, J = 12.5 Hz), 3.13-3.17 (m, 2H), 3.20-3.25 (m, 6H), 3.34 (t, 6H, J = 5.2 Hz), 3.52-

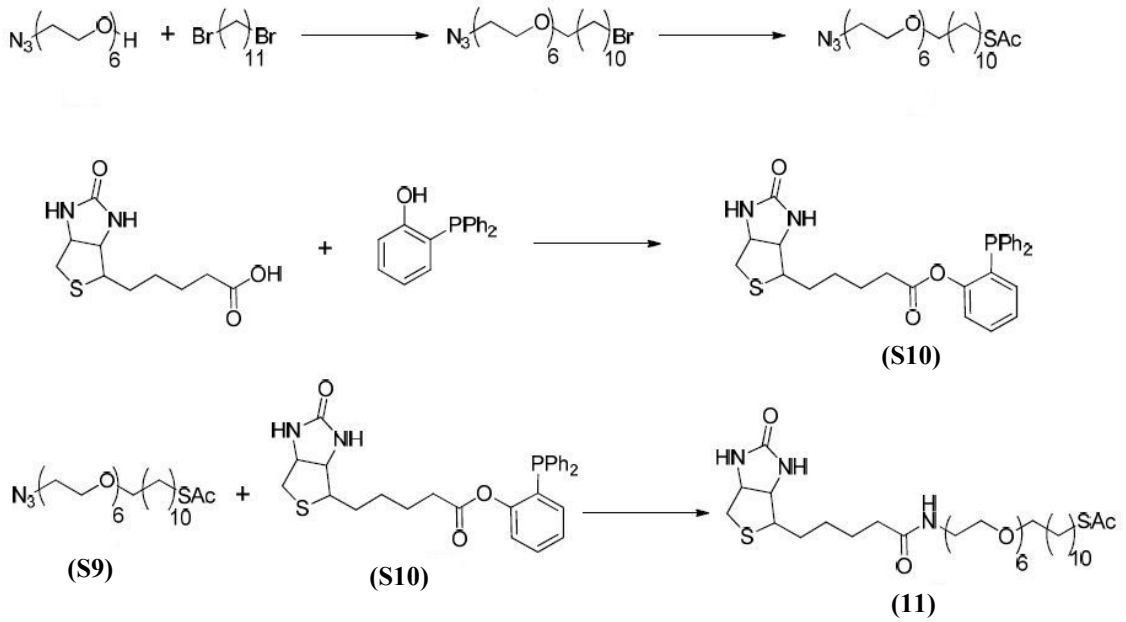


Scheme 14. Synthesis of compound **8**

3.78 (m, 60 H), 4.13 (s, 2H), 4.30-4.35 (m, 2H), 4.49-4.52 (m, 2H), 4.58 (s, 4H), 5.76 (s, 3H), 6.45 (s, 2H), 6.50 (s, 2H), 6.81 (s, 2H), 7.09 (s, 1H), 7.32 (d, 4H, J = 8 Hz), 7.47 (d, 4H, J = 8 Hz), 8.91 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 25.5, 26.3, 28.0, 28.1, 28.9,

29.2, 29.5, 29.8-29.9, 32.5, 35.8, 39.1, 40.3, 50.6, 55.5, 60.1, 61.7, 69.3, 69.7, 69.9, 70.0, 70.3-70.6, 72.7, 88.9, 90.0, 117.7, 122.3, 124.1, 124.8, 127.4, 131.6, 138.6, 146.5, 163.9, 168.8, 173.4, 195.1; MALDI-MS (M+H): found m/z 1909.87, calcd for $C_{96}H_{148+1}N_8O_{25}S_3$ 1909.97.

Synthesis of N-(biotinylatedamido)-18-(11-acetylthioundecano)-3n$_{18}^{3}$-hexaoxaoctadecan-1-amide (10) The biotin functionalized linker **10** was synthesized as described in scheme 16.



Scheme 15. Synthesis of compound **11**

Compound **S4a**[82] was synthesized as we reported earlier. Compound **S4a** (1 g, 3.3 mmol) and NaH (96 mg, 4 mmol) were dissolved in dry DMF and stirred for 30 min at 0 °C. The resulting solution was added drop wise to 1,11-dibromoundecane (0.93 g, 0.7 mL, 3 mmol) in dry DMF and subsequently stirred for 12 h. The reaction was quenched with methanol and the solvent was evaporated in a rotavapor. The resulting oil was dissolved

in 100 mL dichloromethane, washed thrice with water, and dried over $MgSO_4$. Subsequently, the solvent was evaporated and the residue was purified by column chromatography using 6% methanol inn dichloromethane over 4 h yielding 1.2 g of compound **S8** (68%), as verified by NMR and mass spectrometry. ^{1}H NMR (400 Hz, $CDCl_3$): δ 1.19-1.24 (m, 14H), 1.4 (t, 2H, J = 6.8 Hz), 1.55 (q, 2H, J = 6.8 Hz), 1.83 (q, 2H, J = 6.8 Hz), 3.37 (t, 2H, J = 5.2 Hz), 3.42 (t, 2H, J = 6.8 Hz), 3.55 (t, 2H, J = 5.2 Hz), 3.59-3.68 (m, 20H). ^{13}C NMR (100 MHz, CDCl3): δ characteristic peaks for PEG and PE were observed. HRMS (APCI) (M+H): found m/z 540.2581, calcd for $C_{23}H_{46+1}BrN_3O_6$ 540.2650.

To a stirred solution of compound **S8** (1.1 g, 2.2 mmol) in dry DMF, potassium thioacetate (0.28 g, 2.4 mmol) was added and the resulting mixture was stirred overnight at 40 °C. Then the solvent was evaporated and the residue was purified by column chromatography using 5% DCM in methanol to yield 0.9 g of compound **S9** (82%). ^{1}H NMR (400 MHz, CDCl3): δ 1.21-1.4 (m, 14H), 1.46-1.59 (m, 4H), 2.29 (s, 3H), 2.83 (t, 2H, J = 7.2 Hz), 3.35 (t, 2H, J= 5.2 Hz), 3.41 (t, 2H, J = 6.8 Hz), 3.54 (t, 2H, J= 5.2 Hz), 3.6-3.72 (m, 20H). ^{13}C NMR (100 MHz, CDCl3): δ characteristic peaks for PEG and PE were observed. Characteristic peak for carbonyl was observed at δ= 195.9. HRMS (APCI) (M+H): found m/z 536.3379, calcd for $C_{25}H_{49+1}N_3O_7S$ 536.3371.

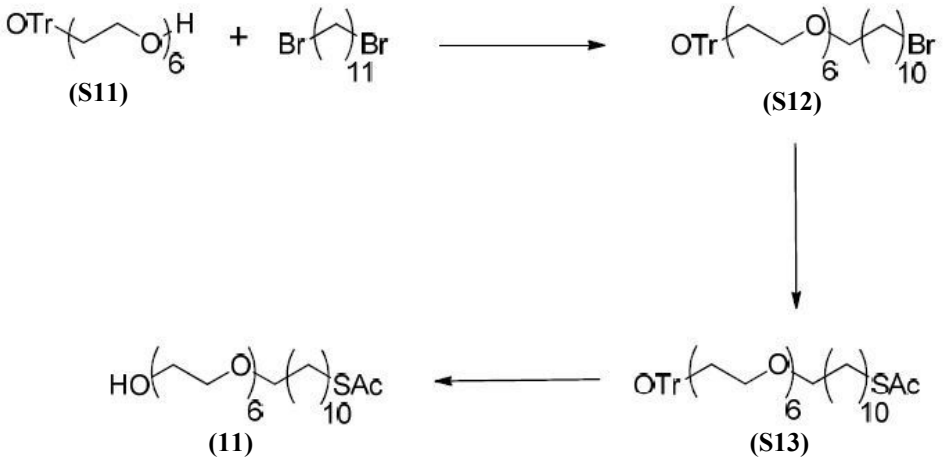
To a stirred solution of D-biotin (0.50 g, 2 mmol) in 5 mL dry DMF, EDC (0.51 g, 2.7 mmol) and catalytic amount of DMAP (10 mg) were added and stirred at 0 °C for 30 minutes. To this solution (2-Hydroxyphenyl)diphenylphosphine (0.57 g, 2 mmol) in 5 mL dry DMF was added and stirred overnight. Then the solvent was evaporated and the

residue was purified using 5% methanol in DCM to yield compound **S10** (0.61 g, 59%)**.** ^{1}H NMR (400 MHz, CDCl3): δ 1.34 (quin, 2H, J = 7.6 Hz), 1.44-1.70 (m, 4H), 2.27 (t, 2H, J = 7.6 Hz), 2.67 (d, 1H, J = 12.8 Hz), 2.74-2.87 (m, 1H), 3.0-3.10 (m, 1H), 4.18-4.25 (m, 1H), 4.37-4.44 (m, 1H), 5.81 (s, 1H), 6.33 (s, 1H), 6.76-6.84 (m, 1H), 7.06-7.16 (m, 2H), 7.23-7.40 (m, 11H). ^{13}C NMR (100 MHz, CDCl3): δ 24.3, 28.2, 33.6, 40.5, 55.5, 60.1, 61.9, 122.5, 126.1, 128.5, 128.6, 129.0, 129.9, 130.0, 130.2, 133.7, 133.8, 134.0, 135.5, 135.5, 152.7, 152.8, 164.0, 171.5. HRMS (APCI) (M+H): found m/z 505.1725, calcd for $C_{28}H_{29+1}N_2O_3PS$ 505.1717.

To a stirred solution of compound **S10** (0.5 g, 1 mmol) in 5 mL DMF:water (10:1) was added solution of compound **S9 (**0.53 g, 1 mmol) in 5 mL DMF:water (10:1) and reaction mixture was stirred at 100 °C for 1 h. Then the solvent was evaporated and the resulting residue was purified using 10% methanol-DCM to give compound **8** (0.51 g, 74%). ^{1}H NMR (400 MHz, CDCl3): δ 1.12-1.72 (m, 21H), 2.16 (t, 2H, J = 7.2 Hz), 2.25 (s, 3H), 2.46 (d, 1H, J =12.8 Hz), 2.74-2.84 (m, 4H), 3.01-3.11 (m, 1H), 3.30-3.41 (m, 4H), 3.47-3.53 (m, 4H), 3.53-3.63 (m, 20H), 4.21-4.26 (m, 1H), 4.4-4.47 (m, 1H), 5.9 (s, 1H, broad), 6.76 (s, 1H, broad), 6.88 (s, 1H, broad); ^{13}C NMR (100 MHz, CDCl3): δ characteristic peaks for PEG and PE were observed. 3 carbonyl peaks were observed at δ = 164.2, 173.4, 195.97; HRMS (APCI) (M+H): found m/z 736.4217, calcd for $C_{35}H_{65+1}N_3O_9S_2$ 736.4242.

Synthesis of 18-(11-acetylthioundecano)-3n$_{18}^{3}$-hexaoxaoctadecan-1-ol (5) The control molecule **11** was synthesized as followed 17.

Compound **S11** (1 g, 2 mmol) and NaH (0.057 g, 2.4 mmol) were dissolved in dry DMF and stirred for 30 min at 0 °C. The resulting solution was added drop wise to 1,11-



Scheme 16. Synthesis of compound **11**

dibromoundecane (0.57 g, 0.42 mL, 1.8 mmol) in dry DMF and subsequently stirred for 12 h. The reaction was quenched with methanol and the solvent was evaporated. The resulting oil was dissolved in 250 mL dichloromethane, washed thrice with water, and dried over MgSO4. Subsequently, the solvent was evaporated and the residue was purified by column chromatography using 6% methanol inn dichloromethane over 4 h yielding 1.29 g of compound **S12** (68%), as verified by NMR and mass spectrometry. ^{1}H NMR (400 MHz, CDCl3): δ 1.21-1.46 (m, 14H), 1.55 (q, 2H, J = 6.8 Hz), 1.83 (q, 2H, J = 7.2 Hz), 3.22 (t, 2H, J = 5.2 Hz), 3.38 (t, 2H, J = 6.8 Hz), 3.42 (t, 2H, J = 6.8 Hz), 3.54 (t, 2H, J = 5.2 Hz), 3.58-3.7 (m, 20H), 7.16-7.48 (m, 15H, aromatic). ^{13}C NMR (100

MHz, CDCl3): δ characteristic peaks for PEG and PE were observed. Peaks for aromatic carbons were at δ= 126.89, 127.73, 128.69, 144.11. HRMS (APCI) (M+H): found m/z 756.3668, calcd for $C_{42}H_{61+1}BrO_7$ 757.3679.

To a stirred solution of compound **S12** (1 g, 1.3 mmol) in dry DMF, potassium thioacetate (0.162 g, 1.4 mmol) was added and the resulting mixture was stirred overnight at 40 °C. Then the solvent was evaporated and the residue was purified by column chromatography using 5% DCM in methanol to yield 0.8 g of compound **S13** (81%). ^{1}H NMR (400 MHz, CDCl3): δ 1.18-1.36 (m, 14H), 1.47-1.61 (m, 4H), 2.29 (s, 3H), 2.83 (t, 2H, J = 7.2 Hz), 3.21 (t, 2H, J = 5.2 Hz), 3.41 (t, 2H, J = 6.8 Hz), 3.54 (t, 2H, J = 5.2 Hz), 3.57- 3.72 (m, 20 H), 7.15-7.49 (m, 15H, aromatic); ^{13}C NMR (100 MHz, CDCl3): δ characteristic peaks for PEG and PE were observed. Other peaks were δ= 126.88, 127.73, 128.69, 144.10 (all aromatic) and 196.02 (carbonyl). HRMS (APCI) (M+H): found m/z 753.4415, calcd for $C_{44}H_{64+1}O_8S$ 753.4402.

A solution of 50% trifluoroacetic acid (1 mL) was added to a mixture of **S13** (0.75g, 1.0 mmol) in DCM (10 mL) at 0 °C, and stirred 1 h. The reaction was poured into saturated NaHCO3 and extracted with DCM. The combined organic layers were washed with brine and dried over Na_2SO_4. After filtration, the filtrate was evaporated in vacuo and purified by column chromatography on silica gel (elution; 10% MeOH in DCM). Production fractions were collected and evaporated in vacuo to give compound **11** (light yellow oil, 0.39 g, 76%). %). ^{1}H NMR (400 MHz, CDCl3): δ 1.14-1.38 (m, 14H), 1.45-1.59 (m, 4H), 2.28 (s, 3H), 2.62 (broad, 1H), 2.82 (t, 2H, J = 7.2 Hz), 3.40 (t, 2H, J = 6.8 Hz), 3.49-3.66 (m, 24 H); ^{13}C NMR (100 MHz, CDCl3): δ characteristic peaks for PEG and PE were

observed. Carbonyl peak was observed at δ= 196.05 (carbonyl). HRMS (APCI) (M+H): found m/z 511.3310, calcd for $C_{25}H_{50+1}O_8S$ 511.3305.

2.3.3 Preparation and characterization of mixed monolayers

Both monolayer SAM-mb and SAM-bb was prepared by the following procedure. To a mixed solution of the monobiotinyl ligand **10** or three arm bisbiotin **8** (10 μM) and the spacer **11** (90 μM) in an absolute ethanol (3 mL) was pyrrolidine (30 μL, 0.36 mM) added in a Petri dish covered with a lid. After 15 min, thin layer chromatography (TLC) showed that the acetyl group was removed to free the thiol group. A gold chip was immersed in the solution for 24 h, followed by rinsing with absolute ethanol and drying with a nitrogen flow, and used immediately. The chip was characterized by contact angle goniometry, ellipsometry, and FTIR.

Characterization of mixed monolayers

These monolayers were characterized with contact angles, thicknesses, and FTIR spectra. Contact angles were measured using Kruss Easy Drop by injecting a water droplet (2 μL) on a modified gold chip placed on the sample plate and contact angles were measured in the video window of manufacturer's DSA software. Data shown in the contact angle column in Table 2 is an average of those five measurements.

Thicknesses of the monolayers were measured using Gaertner-Stokes ellipsometer. For thickness calculation, refractive indices of organic layers were assumed to be 1.46.[84] Thickness of the bare gold chip was measured first, followed by measurement of modified gold chips. Measurements were taken on five different arbitrary positions on the

substrate and finally the average value was taken. The thickness of the monolayer was determined by subtracting the bare gold thickness from the measured one.

Table 2. Measured physical properties of monolayers

Monolayer	*Length of Molecule (calc.)(Å)	*Length of Molecule **11** (calc.)(Å)	Measured Thickness (Å)	Measured Contact angle (deg)
SAM-mb	48 (mol. **10**)	38	32.8 ± 0.9	38 ± 1
SAM-bb	69 (mol. **8**)	38	48.2 ± 0.8	40 ± 1

* Molecular lengths calculated using ChemDraw 3D

The spectra were acquired using Thermo Nicolet 6700. FTIR Spectra of monolayer and powder samples were recorded with a surface grazing angle accessory (SAGA) and attenuated total reflection (ATR) accessory (Smart Orbit, Thermo Electron Corporation) with 256 scans respectively. Area of 8mm diameter circle of the modified gold chip was used as an active spot for the monolayer spectra measurement. A cluster of the band appears in between 2800-3000 cm^{-1} in monolayers and solid samples are assigned to CH_2 symmetric and asymmetric stretches from backbone and ethylene glycol moieties. Band near 1700 cm^{-1} corresponds to the presence of carbonyl group in solid and monolayer sample. Also, Broad band near 3300 cm^{-1} is assigned to the stretches of both amide (-NH) group and hydroxyl (-OH) groups present in all the samples.

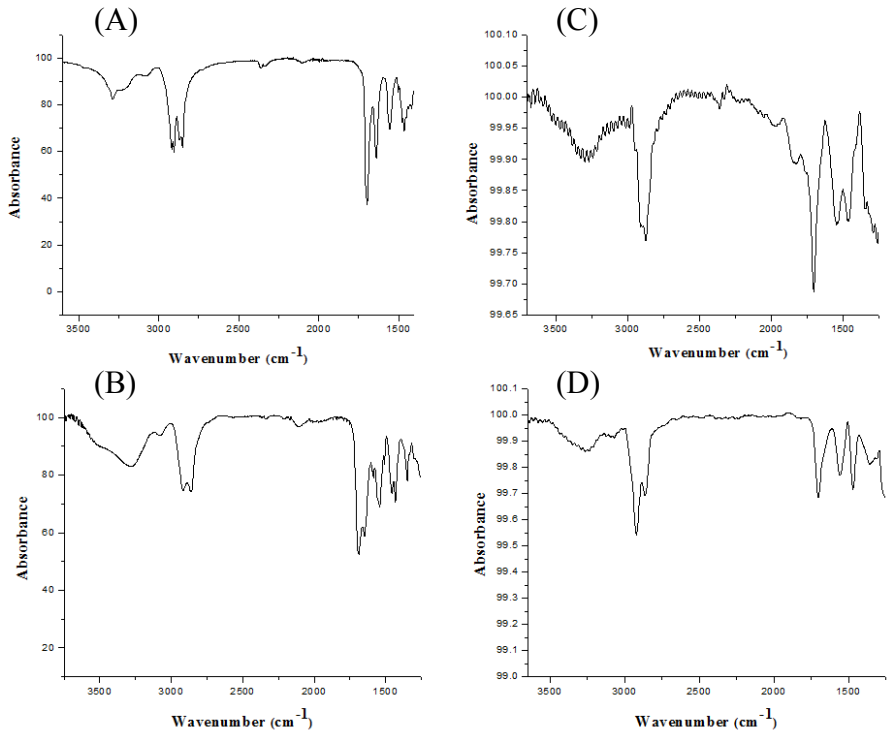


Figure 26. FTIR spectra of (A) compound **10**, (B) compound **8**; (C) mixed monolayer SAM-mb; (D) mixed monolayer SAM-bb.

2.3.4 SPR experiments

The instrument Bi 2000 from Biosensing Instrument was used for SPR measurements. A gold chip coated with a monobiotinyl monolayer (SAM-mb) was mounted on the instrument, and calibrated with 1% ethanol in a PBS buffer, pH 7.4. A solution of streptavidin in PBS buffer (0.4 μM) was flowed onto the chip at a rate of 30 ul/min over a period of 2.5 min, followed by flowing the PBS buffer until a steady-state response was observed, and then a solution of D-biotin in PBS buffer (3 mM) was flowed onto the chip over a period of 25 min at the same rate. The SPR experiments on the SAM-bb were carried out in the same procedure. Association rate constants (k_{on}) were determined by fitting the curve into most commonly used Langmuir 1:1 interaction model in OriginPro

2015 and competitive dissociation rate constants (k_{off}) were determined by fitting the curve into an equation reported in literature[74] in OriginPro 2015:

$$y = y_0 + A_1 exp\left(-k_{off}^1(x - x_0)\right) + A_2 exp\left(-k_{off}^2(x - x_0)\right)$$

2.3.5 Summaries

Departing from a benzene ring, one of the most commonly used scaffolds to construct multivalent molecules,[58] we extended the structure with phenylethynyl groups to create a larger rigid structure, on which small sized antibody mimics can be constructed. The phenylacetylene chemistry used in this present study can be extended to synthesizing even more complicated and larger scaffolds.[85] Using the biotin-streptavidin interaction as a model, we have demonstrated that the three arm bisbiotin ligand: (1) formed an intramolecular bivalent complex more stable than its intermolecular counterpart on the surface; (2) bound to streptavidin cooperatively. We should point out that three arm bisbiotin formed a very stable complex with streptavidin mainly because it reduced the dissociation rate significantly. This implies that an antibody mimic would increase the residence time (smaller k_{off}) of a ligand interacting with its receptor, which is of importance in applications for drug discovery. Further studies are needed to get detailed information on using three arm linker as antibody mimic.

CHAPTER 3

3 Site-specific chemical modification of peptides, proteins and its attachment to DNA thread for their translocation through solid-state nanopores

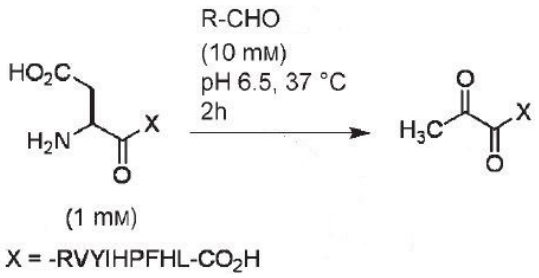
3.1 Introduction

3.1.1 Advancement in N-terminal modification of peptides and proteins

Development of protein modification strategies is an important field of research. Placing a synthetic group at a desired location in protein make this task even more challenging. Some of the methods targeting specific amino acids have already been discussed in chapter 1. The presence of multiple copies of same native amino acid all over the protein makes this strategy less effective to achieve the goal of site-specific modification. Cysteine being one of the rarest amino acids can be targeted to solve this problem.[40] But, the introduction of new cysteine moiety by genetic engineering may sometimes affect the structure and function of proteins. For example, antibodies held their structure by disulfide bond formation[86] and metallothioneins have multiple cysteine residues in their active site[87]. Incorporation of new cysteine residue may affect these properties. Sometimes there is more than one modification needed at different places of protein. In those cases, other than cysteine modifications new strategies are still required for the second modification.

One of the strategies that can be followed is targeting native functionalities that occur only once in peptide and proteins i.e. N-, C- terminus. N-terminus being nucleophilic provides us with the opportunities to carry out the reaction with different reagents. Here, some of the reagents targeting N-terminus are listed below.

Pyridoxal-5-phosphate (PLP) is such a reagent which has been used to install a reactive ketone functional group at the N-terminus of peptides and proteins.[88] The reaction scheme is shown below.



Scheme 17. N-terminal modification of twenty naturally occurring amino acids by PLP.[88]

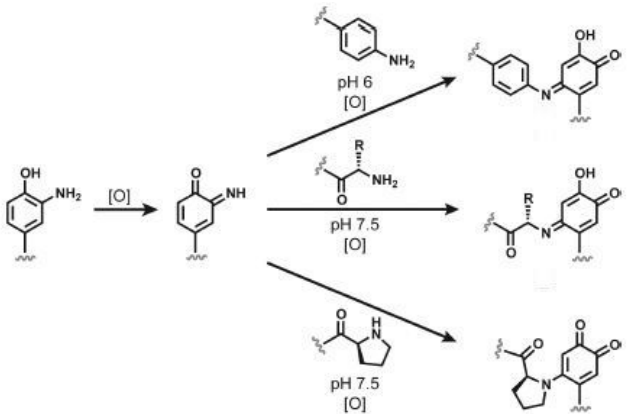
First, the reaction has been optimized for different peptides and further applied to some of the proteins. One table is shown below for showing data for protein modification by PLP.

This PLP mediated reaction would not be effective for the N-terminal residues having serine, threonine, cysteine and tryptophan because they can have side reactions with aldehydes. Also, N-terminal proline would remain unreactive towards PLP.

Table 3. Data for site-specific protein modification using PLP.[88]

Entry	Protein	N-terminal amino acid	Conc. [μM]	*T* [°C]	Conv. [%]
1	myoglobin	glycine	50	37	69
2	GFP-1V	valine	10	55	67
3	GFP-1G	glycine	10	37	41
4	GFP-1G	glycine	10	55	80
5	RNase A	lysine	50	37	50
6	thioredoxin	glycine	94	37	50
7	protein G′	methionine	33	41	30

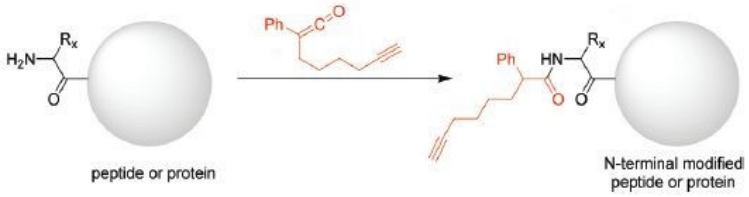
O-amino phenol is another reagent used for N-terminal peptide, protein modification through oxidative coupling strategy.[89] The reaction scheme is shown below.



Scheme 18. Schematic representation of oxidative coupling of O-amino phelol with A) aniline moiety, B) twenty naturally occurring amino acids, C) prolin residue.[89]

O-amino phenol has been tested for a series of peptides and proteins showing its effectiveness for N-terminal modification. Data showed that oxidative coupling strategy is very much effective for proline residue at the N-terminal of peptides and proteins compared to the other 20 naturally occurring amino acids. This makes use of this reagent very limited.

Ketens have also been used for site-specific N-terminal peptides and proteins modification.[90] The reaction scheme is shown below.



Scheme 19. Schematic representation of N-terminal modification of peptide or protein by ketene compound.[90]

Great site-specificity has been achieved effectively with ketene reagent. But, conversions of peptides and proteins are poor.

Defficiencies in above mentioned examples lead us to think of a simple and effective strategy which should be capable of site-selectively modifying any N-terminus amino acid in peptides and proteins in mild condition. The progress of this strategy has been discussed in details later in this chapter.

3.2 Peptide and protein sequencing using nanopores

Nanopores—orifices with nanometer diameters— can function as nanofluidic channels for the flow of ions and the transport of biomolecules. When a charged molecule is electrophoretically driven through the nanopore, it partially obstructs the passage of ions and modulates the current through the pore. Parameters derived from the current blockade can be used to identify the molecule and even identify structural subunits. Nanopore techniques are emerging as a single molecule tool for sequencing DNA,[91] detecting proteins,[92] polysaccharides,[93] viruses,[94] with possible clinical applications[95]. As the protein nanopore DNA sequencing technology makes inroads into genomic research,[96,97,98] a legitimate question naturally arises: can the nanopore sequence proteins

as well? Given the fact that even a MspA protein nanopore, which has a finer nanopore (0.5 nm thickness and 1.2 nm diameter) than α-hemolysin, only demonstrates a four-nucleotide resolution[91] with ~ 85% accuracy,[99] it seems unlikely that ion-current blockade measurements will resolve individual amino acids because the amino acid calling will otherwise depend on sorting $20^4 = 160{,}000$ different signals. A new reading mechanism has to be developed to achieve single amino acid residue resolution for protein sequencing. Electron tunneling detection has been shown to have this capability. Recently, we have demonstrated that individual amino acids can be identified and two different peptides distinguished at a single molecule level by a technique we call recognition tunneling, which measures tunneling currents of analytes in a ~ 2.5 nm nanogap with its two electrodes functionalized with recognition molecules.[100] Kawai and coworkers have also reported the identification of amino acids and phosphorylated peptides by means of electron tunneling currents with either 0.5 or 0.7 nm nanogaps.[101] Thus, one can conceive of a device that integrates a tunneling gap with a solid-state nanopore for analyzing protein sequences.

Here, another primary issue has been addressed in developing the nanopore technology for analysis of proteins in a proteome, which is the translocation of proteins and peptides. While DNA is always negatively charged with uniform distribution along its phosphate backbone under physiological conditions, a protein can be zero, positive or negative in its net charge, however, it bears both positively and negatively charged side chains randomly distributed on its amide backbone. This makes translocating the protein electrophoretically challenging. Due to lack of a PCR-like techniques for analysis of

proteins, it is very difficult to even acquire direct evidences to prove protein translocation.[102] Recently, Akeson and his coworkers first demonstrated that a recombinant ubiquitin-like protein Smt3 bearing a polyanionic peptide at its C-terminus was unfolded and pulled through a α-hemolysin (α-HL) nanopore by the AAA+ unfoldase ClpX.[103] Almost at the same time, Bayley's team reported that a thioredoxin protein tethered to a negatively charged oligonucleotide could also be unfolded and translocated through the α-HL nanopore by an applied voltage.[104] These studies suggest a new approach to translocating proteins using a charged 'pulling-string' to draw the protein into a nanopore.

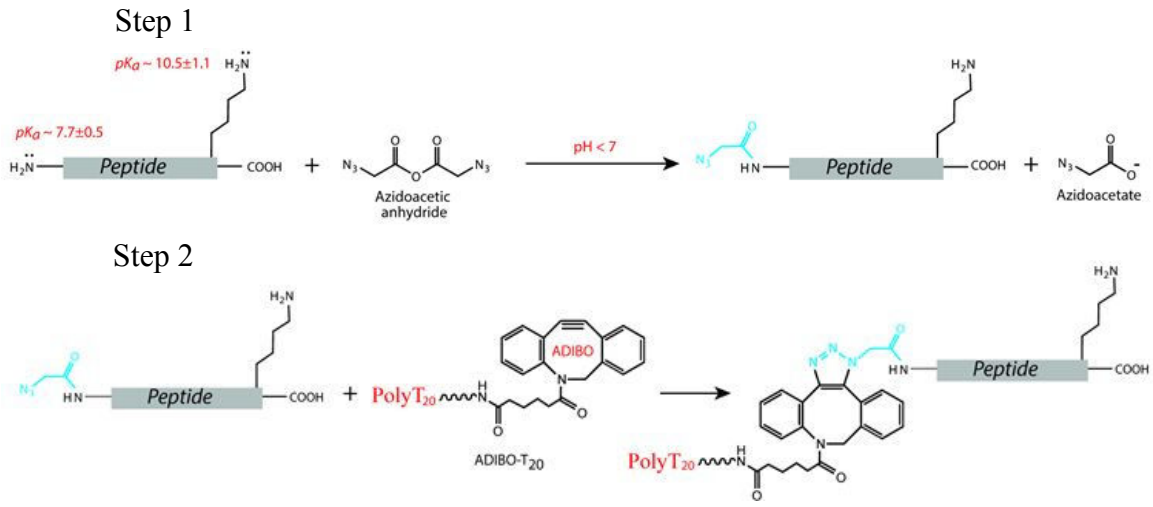
With the ultimate goal of sequencing proteins, our initial objective has focused on using a recognition tunneling nanopore to identify peptides. As a matter of fact, the most commonly used method in proteomics is the shotgun mass spectrometry, in which proteins are first digested into peptides with enzymes, such as trypsin–that generates peptides containing only one lysine or arginine residue at their C-terminus, separated with liquid chromatography, and injected as charged ions into mass spectrometer for identification.[105] While working on fabricating a fixed-gap tunnel junction in solid-state nanopores,[106] a molecular threading strategy has been developed to facilitate translocation of peptides: tethering a chain molecule with a large net charge, which functions as a molecular thread, to the termini of peptides to make them all have the same sign of charge so that they can be carried over from one side to another of a nanopore by the threading molecule under a voltage bias. Thus, a basic requirement for the threading molecule is that it can readily translocate through nanopores. We chose an

oligonucleotide composed of 20 thymidines (referred to as $PolyT_{20}$) as a negatively charged thread because a PolyT molecule has a well-defined structure and charge distribution, its translocation has been well studied,[107,108] and it is also immediately available from commercial sources. In this present study, we have focused on developing a simple and effective chemistry for attaching the oligonucleotide to the N-termini of peptides, and demonstrating the translocation of the DNA-peptide conjugates through solid-state nanopores.

3.3 Results and Discussion

3.3.1 Addition of PolyT$_{20}$ to N-termini of peptides

We chose to functionalize the N-terminal α-amine of peptides because it is more nucleophilic than the carboxylate at C-terminus, susceptible of chemical modification with electrophilic reagents in the physiological conditions. Then, the challenge is to find a reagent and conditions that selectively react with the α-amine in the presence of other nucleophilic groups, such as ε-amine of lysine. Although there are a plethora of chemical methods for modification of proteins reported in the literature,[110] there is a lack of universal chemistry to specifically modify the N-termini of peptides. To a certain degree, the N-terminal reactions vary with peptide sequences[90] and amino acid residues at the end [88,111] in terms of their selectivity and efficiencies. For our purpose, it is particularly important to have a general method to modify any N-terminal amino acid with high selectivity no matter what sequence it is connected to. Acylation is a commonly used chemical reaction in the modification of proteins, which utilizes chemical reagents such as carboxylic halide, carboxylic anhydride, or active ester.[110] It has been reported that the carboxylic anhydrides can selectively react with the N-terminal amines of peptides at slightly basic conditions.[112,113] Accordingly, we devised a process that harnesses the anhydride reaction to effectively introduce a bioorthogonal function to N-termini of peptides for the DNA attachment. As shown in Scheme 21, an azidoacetic anhydride reagent first reacts with a peptide bearing a lysine residue at its C-terminus, generating an N-azidoacetylated peptide under a slightly acidic condition (Step 1), and then the azide group reacts with



Scheme 20. Chemical reactions for attaching an oligonucleotide to N-termini of peptides.[109]

aza-dibenzocyclooctyne (ADIBO) functionalized $PolyT_{20}$ through a click reaction, resulting in a desired DNA-peptide conjugate (Step 2). The ADIBO function—a strained cycloalkyne—specifically reacts with azide at a high reaction rate (k = 0.3 $M^{-1}s^{-1}$) without need of a copper catalyst[114] (so called copper free click chemistry[115]). Given these advances in click chemistry, we paid more attention to the reaction of azidoacetic anhydride with peptide in high selectivity. Besides ε-amine of lysine, the acyl anhydride

Table 4. Peptide sequences used in this study and their physicochemical properties[109]*

	Sequence	Mass [a]	Net Charge at pH 7 [a]
P-1	YLGEEYVK	999.49	-1
P-2	DRVYIHPFHL	1295.68	+0.2
P-3	EAIYAAPFAKKK	1335.76	+2

* [a] Calculated using the peptide property calculator in http://www.innovagen.com/custom-peptide-synthesis/peptide-property-calculator/peptide-property-calculator.asp;

can react with the phenolate ion of tyrosine, sulfhydryl group of cysteine, aliphatic hydroxyl of serine and threonine, and imidazolyl ring of histidine as well.[116] However, the intrinsic reactivity of these groups to an electrophile and stability of their acyl derivatives are all different. In most cases, the reactive groups involved in the acylation are the α- and ε-amine, imidazolyl ring, and to a lesser extent, –SH and –OH. The thioester and ester from acylation of cysteine, tyrosine, serine and threonine residues can be reversed to the original groups.[117] In an aqueous solution, these functional groups have distinguishable acid dissociation constants (pK_a): for example, an average pK_a value for ε-amine of lysine in proteins is 10.5, for α-amine of the N-terminus is 7.7.[118] Because a protonated amine is not reactive, this discrepancy gives us a room to tune the selectivity of the acylation reaction by changing the pH and, in turn, the protonated states of these amines (Step 1 in Scheme 21). To achieve high selectivity for the α-amine, indeed, we chose a slightly acidic condition for the acylation reaction, instead of the basic.

We have studied three representative short peptides (designated as **P-1**, **P-2**, and **P-3**, respectively, in Table 4) for the reactions shown in Scheme 21.[109] The sequences and calculated physicochemical properties of these peptides are listed in Table 4. In particular, **P-1** and **P-3** contain different numbers of lysine (K) and no histidine (H), and P-**2** has two histidines and no lysine. Each of them carries a different net charge at neutral pH. Circular Dichroism (CD) showed that they took a random conformation in aqueous solution, whereas $PolyT_{20}$ adopted an organized right-handed helical conformation so that it may provide an entropic advantage for threading into a nanopore. We began with **P-1,** a peptide that mimics a trypsin digest. First, we compared the anhydride with another

commonly used acylating reagent N-hydroxysuccinimidyl (NHS) ester for their selectivity. The NHS azidoactate reacted with **P-1** at pH 6.7, but produced two products that were characterized as a peptide modified by one and two azidoacetyl (N_3CH_2CO-) groups by MALDI mass spectrometry. In general, the NHS ester may preferentially react

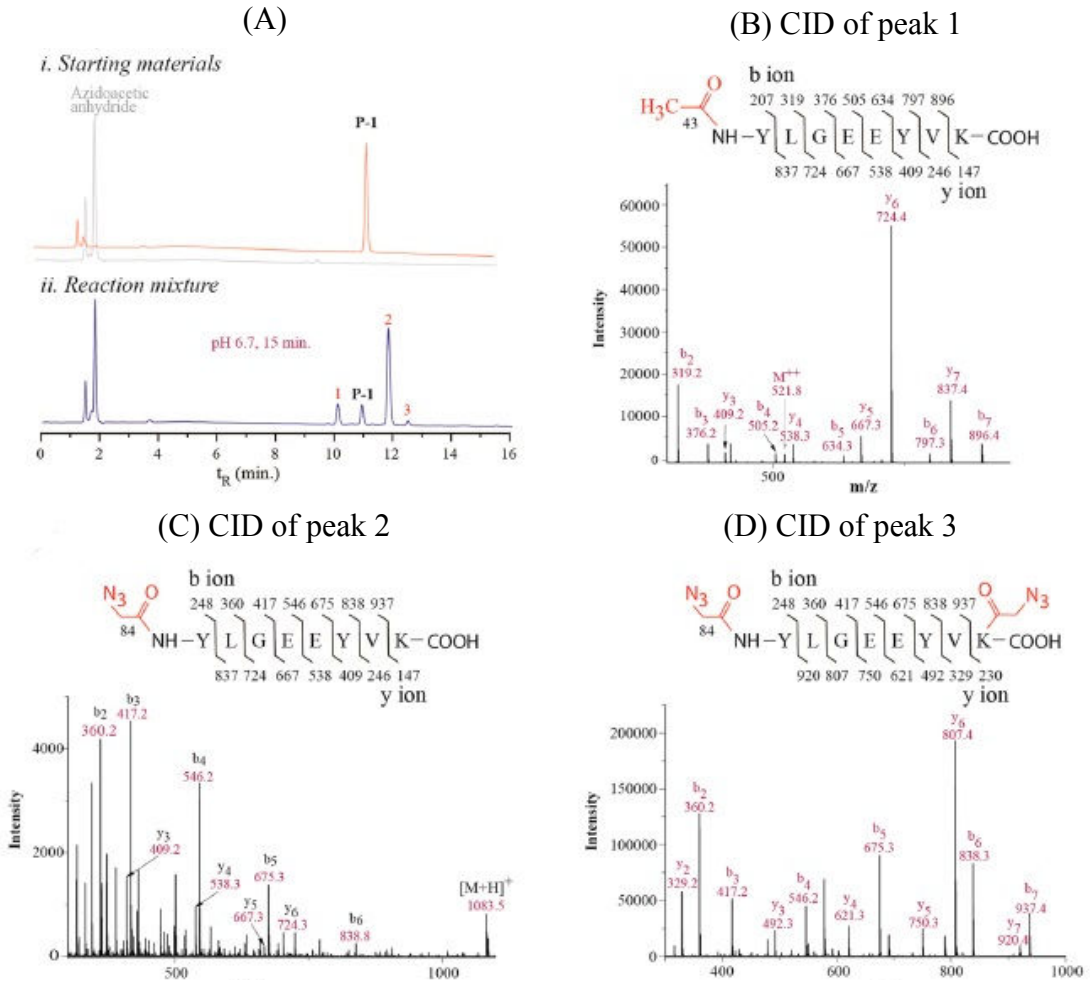


Figure 27. HPLC and Mass analysis of azidoacetic anhydride reacting with **P-1**: (A) RP HPLC chromatograms of (*i*) starting materials and (*ii*) the reaction mixture at pH 6.7; (B) Tandem mass spectrum of peak 1 in (A); (C) Tandem mass spectrum of peak 2 in (A); (D) Tandem mass spectrum of peak 3 in (A); Inserts in B, C, and D are calculated fragment ions of the corresponding peptides.[109]

with the lysine amine. Mentinova et al have showed the preference for the N-terminal amine only when pH was ~ 5.[119] based on these results, we chose azidoacetic anhydride as an acylating reagent (see Experimental Methods for its synthesis). Initially, the acylation reaction was carried out with **P-1** at a concentration of 0.4 mM and azidoacetic anhydride at 1.2 mM in a sodium acetate buffer, pH 6.7 at 0 °C, and monitored by reverse phase (RP) HPLC. Retention time (t_R) of each starting material was determined by a separate HPLC run (Figure 27, A-*i*). After 15 min, three new peaks appeared in the HPLC chromatogram, labeled as 1, 2 and 3 in red (Figure 21, A-*ii*). The peak 1 has a shorter retention time (t_R) than **P-1**; in contrast, both peak 2 and 3 have longer retention times compared to **P-1**. The ratio of these three peaks was 5:89:6, determined by their chromatographic peak areas. We separated these individual products and characterized them with MALDI and tandem mass spectrometry. Peak 1 was determined as a product

Table 5. Effects of reaction conditions on the selectivity of azidoacetic anhydride[109]

pH	Time (min)	Conversion of **P-1**	Acylating ratio* (α to ε amine)
5.5	15	15%	100 : 0
6.1	15	61%	98.9 : 1.1
6.7	15	85%	96 : 4
6.7	60	97%	91 : 9

* calculated based on areas of peak 2 and 3

resulting from adding a 43 Da mass unit to **P-1**. In its CID spectrum (Figure 27, B), the mass of observed y ions (y3 to y7) matches those calculated from **P-1** without modification from its C-terminus to one next to the N-terminus, indicating that the addition took place at the N-terminus. The N-terminal modification was further confirmed by observed b ions (Figure 27, B), every of them matching up with the mass derived from a **P-1** fragment plus an additional 42 Da. The 42 Da mass may be explained by substituting an acetyl group for one hydrogen of the α-amine (CH_3CO - H). To prove this substitution, we carried out a reaction of acetic anhydride with **P-1** under the same conditions, finding out that the major product (98%) had the same t_R and mass as the peak 1 (Figure 28). This acetyl byproduct was unexpected, albeit only ~ 5 % in the product mixture (and not reactive in the following reaction). Further investigation is required to determine its origin. Peak 2, which was the major product, has a mass of

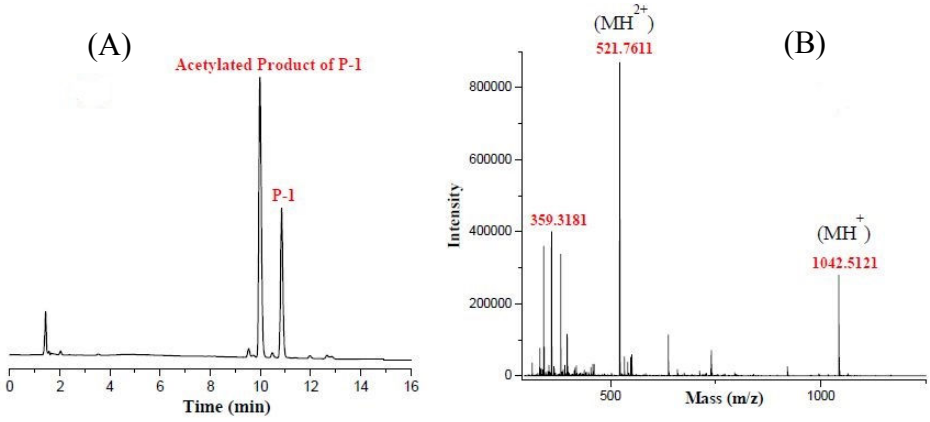


Figure 28. HPLC and Mass analysis of acetic anhydride reacting with **P-1:** (A) RP HPLC chromatogram of reaction mixture at pH 6.7; (B) MALDI mass spectrum of the acetylated product [calculated mass for $C_{49}H_{72}N_9O_{16}$ (MH^+): 1042.50, found: 1042.51].[109]

1083.5 Da, corresponding to mono-azidoacetylated **P-1**. By analyzing its CID spectrum, peak 2 was identified as a product of α-amine azidoacetylated **P-1** (Figure 27, C). Peak 3 has a mass of 1166.5, corresponding to the addition of two azidoacetyl groups to **P-1**, and its CID spectrum indicates that it is a product of both N-terminal and ε-lysine amine azidoacetylated **P-1** (Figure 27, D). This study also showed that the tyrosine residue did not react with azidoacetic anhydride below neutral pH. Next, we studied effects of pH and reaction time on products by means of the HPLC analysis, and the results are listed in Table 5. First of all, at pH below 7, azidoacetic anhydride reacted with the N-terminal amine with high selectivity (> 90%). It increased by decreasing the pH, but the conversion of **P-1** to products was reduced as well. Extending the reaction time increased the conversion rate of peptide to products but also reduced the selectivity and more byproducts were produced. Overall, the reaction of azidoacetic anhydride with **P1** can achieve > 90% selectivity and a > 90% conversion rate of starting material to product.

In turn, we examined the possible reaction of azidoacetic anhydride with histidine. **P-2** was a peptide adopted from hormone angiotensin II, carrying two histidine residues and no lysine in its sequence. At pH 6.7, it reacted with azidoacetic anhydride, resulting in three products (Figure 29, A-*i*, labeled as 1, 2, 3 in red). Their ratio was determined as 91 : 5 : 4, in which peak 1 was a major product. The CID mass analysis confirmed that the peak 1 was a product of N-terminal azidoacetylated **P-2** (Figure 30, A), both peak 2 and 3 were products of **P-2** with one of its histidines residue azidoacetylated. When lowering the reactant ratio to 3:1 between azidoacetic anhydride and **P-2**, the conversion rate was reduced to 17%, but the product ratio between peak 1 and 2 was 91:9 and peak 3 did not

appear

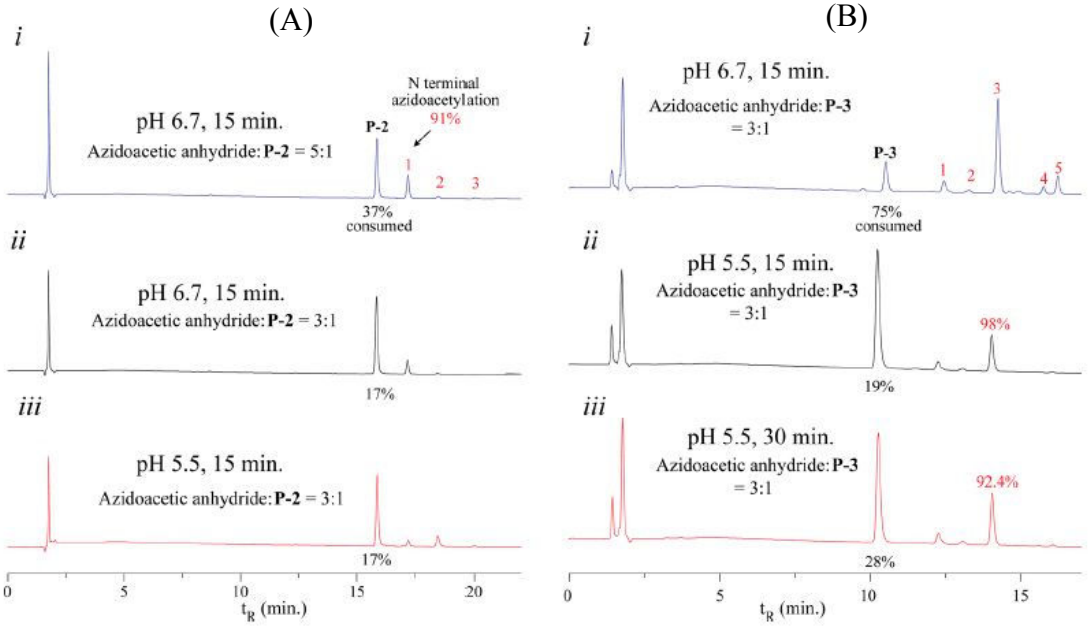


Figure 29. RP HPLC profiles of azidoacetic anhydride reacting with (A) **P-2** and (B) **P-3** at different pH and reaction time.[109]

(Figure 29,A-*ii*). At pH 5.5, the peak 2 then became a major product (Figure 29, B-*iii*). This is probably because the imidazolyl ring of histidine has a lower pK_a than the N-terminus amine so that it is more reactive at the low pH. Unexpectedly, the azidoacetyl group on the imidazolyl ring could not be removed by the base treatment (even with concentrated ammonia).

To further explore limitation of the acylation reaction, we studied the reaction of azidoacetic anhydride with **P-3**, which has three lysine residues at its C-terminus. As a result, **P-3** yielded a mixture of five new products at pH 6.7, labeled as 1, 2, 3, 4, 5 in red, respectively (Figure 29, B-*i*). We have assigned these peaks to their corresponding products in the same way as was done for **P-1**. The peak 1 is a product resulting from

acetylating **P-3** at the N-terminus, the peak 2 is a product of **P-3** with one lysine

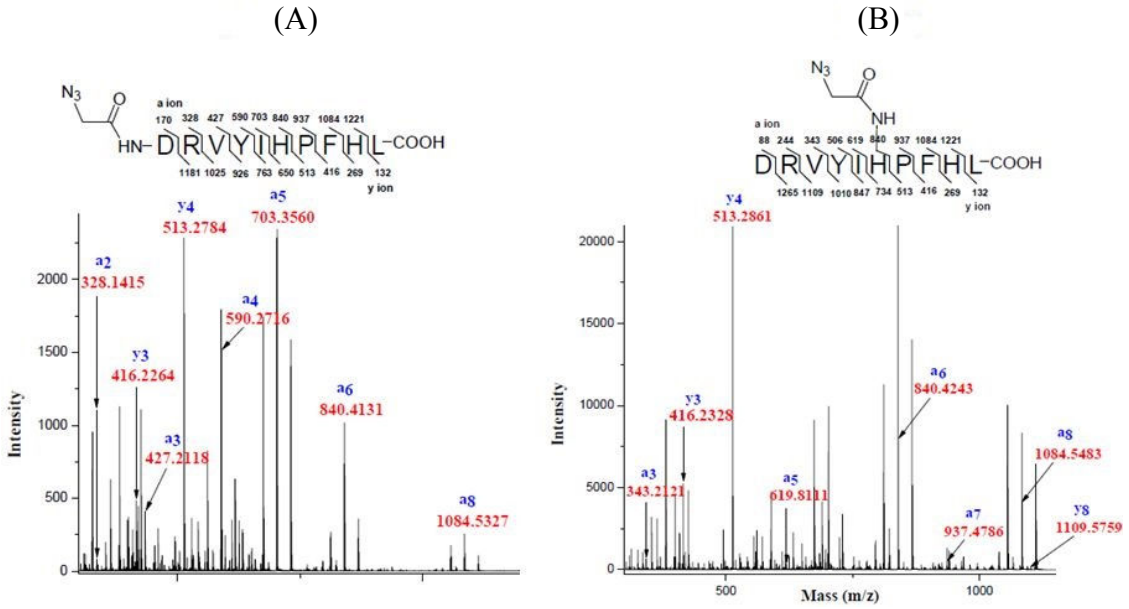


Figure 30. Mass analysis of products from **P-2** reacting with azidoacetic anhydride: (A) Tandem mass spectrum of peak 1[calculated mass for $C_{64}H_{91}N_{20}O_{15}$ (MH^+): 1379.696, found: 1379.687,]; (B) Tandem mass spectrum of peak 2 [calculated mass for $C_{64}H_{91}N_{20}O_{15}$ (MH^+): 1379.696, found: 1379.708,].[109]

azidoacetylated, peak 3 is **P-3** with N-terminus azidoacetylated, peak 4 and 5 are **P-3** with two lysine azidoacetylated. We determined the ratio of these peaks as 1:2:3:4:5 = 7.1 : 2.2 : 73.0 : 5.0 : 12.7. Again, the N-terminal azidoacetylated product (peak 3) was the major product. This result shows that the azidoacetylating reaction can selectively take place at the N-terminal amine of a peptide even when the peptide contains multiple lysine residues. The selectivity can be improved by lowering pH. As shown in Figure 29, B-*ii*, the selectivity increased to ~ 98% at pH 5.5, but the conversion rate of the peptide-to-product was reduced to 19% from 75% at pH 6.7. By comparing the results shown in B-*ii* with B-*iii* of Figure 29, it indicates that extending the reaction time increased the

conversion rate of peptide-to-product, but also reduced the selectivity. The above-discussed studies indicate that azidoacetic anhydride can rapidly react with α-amine with high selectivity at a slightly acidic condition, suitable for labeling the N-termini of trypsin digests.

Next, we studied the reaction of azidoacetylated peptides with $PolyT_{20}$ (Step 2 in Scheme 21).[109] ADIBO-T_{20} was synthesized by reacting ADIBO-NHS ester with $PolyT_{20}$ bearing a C12 amino modifier at its 5' end in a phosphate buffer pH 8, and purified by RP HPLC. It spontaneously reacted with each of N-azidoacetylated peptides when they were mixed in a TEAA buffered solution (pH 7), resulting in desire peptide-$PolyT_{20}$ conjugates, designated as **P-1-T_{20}**, **P-2-T_{20}**, and **P-3-T_{20}** (characterized by MALDI mass spectrometry). The HPLC analysis indicated that these peptides were quantitatively converted to the corresponding peptide-$PolyT_{20}$ conjugates with no detectable byproducts. Nonetheless, the times to complete these reactions were different among these peptides, 45 min for reacting with N-azidoacetylated **P-1**, 20 min with

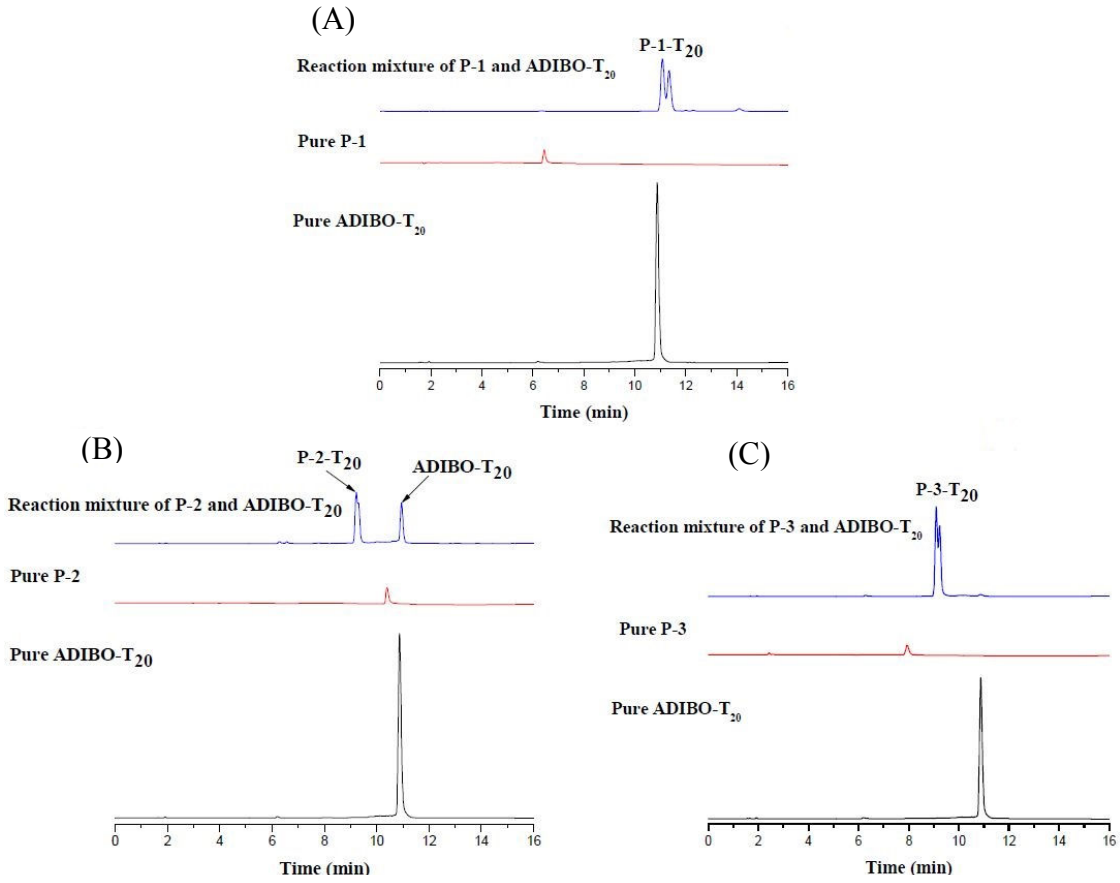


Figure 31. RP HPLC profiles of peptide-DNA conjugates monitored at 260 nm wavelength: (A) reaction mixture of **P-1-T_{20}** (B) reaction mixture of **P-2-T_{20}** (C) reaction mixture of **P-3-T_{20}**.[109]

azidoacetylated **P-2,** 10 min with azidoacetylated **P-3**. The reaction rates seem to correlate with net charges of these peptides (see Table 4). The more positively charged **P-3** reacted with negatively charged PolyT_{20} fastest, the negatively charged **P-1** the slowest, and with **P-2** (which has a smaller positive charge) intermediate. We also noticed that the product peaks are split into two or broadened in the HPLC chromatograms. This is

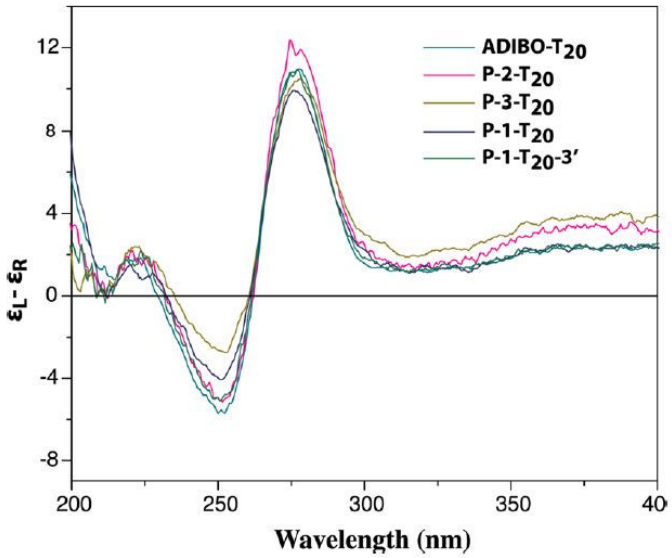


Figure 32. CD spectra of PolyT$_{20}$-peptdie conjugates. The measurement was carried out with each analyte of a 100 μm concentration in a sodium acetate buffer, pH 6.7. Each curve was an average of 4 scans with the buffer as a reference.[109]

because the ADIBO-azide reaction produces a regioisomeric mixture of the triazole connection. In our case, the peptides were connected to either position 1 or 3 of the triazole ring (see Scheme 21. We also characterized conformations of these conjugates with CD spectroscopy (Figure 32). At the first glimpse, these CD spectra are dominated by signature of DNA. Compared to polyT$_{20}$ modified with ADIBO, the negative peaks (at ~ 250 nm) of the conjugates are reduced, with an order of their intensities: **ADIBO-T$_{20}$** > **P-2-T$_{20}$** ≈ **P-1-T$_{20}$-3'** > **P-1-T$_{20}$** > **P-3-T$_{20}$**. This may reflect increase in the lysine residue at the C-termini interacting with the phosphate backbone when the peptide is conjugated to the 5' end of DNA, resulting in diminished helical structure. **P-3-T$_{20}$** that contains three lysine residues at its C-termini has the strongest interaction with PolyT$_{20}$.

3.3.2 Site-specific protein modification using azidoacetic anydride

We also studied effectiveness of azidoacetic anhydride in site-specific modification of couple of proteins such as lysozyme and insulin. The acylation reaction was carried out with lysozyme at a concentration of 0.4 mM and azidoacetic anhydride at 1.2 mM in a sodium acetate buffer, pH 6.7 at 0 °C, and monitored by reverse phase (RP) HPLC. After 15 min, two new peaks appeared in the HPLC chromatogram, labeled as 1 and 2 (Figure 33, A). The ratio of these two peaks was 83:17, determined by their chromatographic peak areas. We separated these individual products and characterized them with MALDI. Peak 1 was determined as a product and peak 2 as doubly modified product. But, further studies need to be carried out to confirm modification occurred at N-terminus of protein.

To further explore effectiveness of the acylation reaction, we studied the reaction of azidoacetic anhydride with insulin protein. As a result, insulin yielded a mixture of three new products at pH 6.7, labeled as 1, 2, 3 respectively (Figure 23, B). We have assigned these peaks to their corresponding products in the same way as was done for lysozyme. The peak 1 and peak 2 are a product resulting from mono acetylation insulin and peak 3 corresponds to double modification. This reaction needs to be further optimized to get a better selectivity. Also, tandem mass spectrometry studies need to be peroformed to indentify modification of insulin occurs at N-terminus.

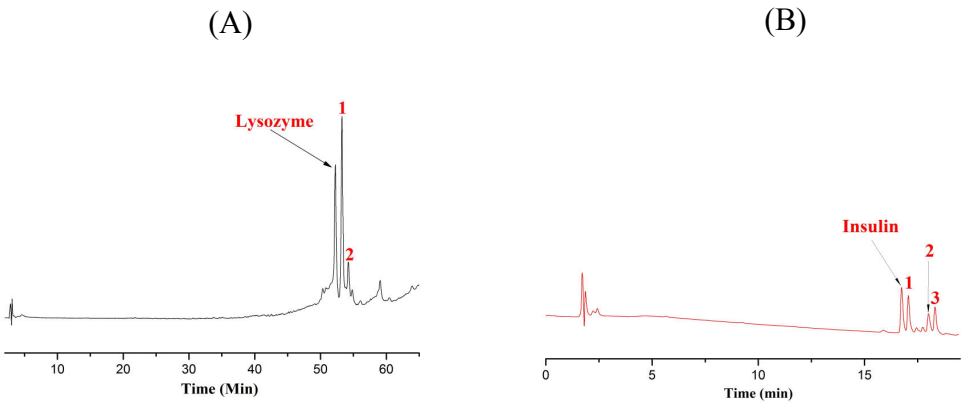
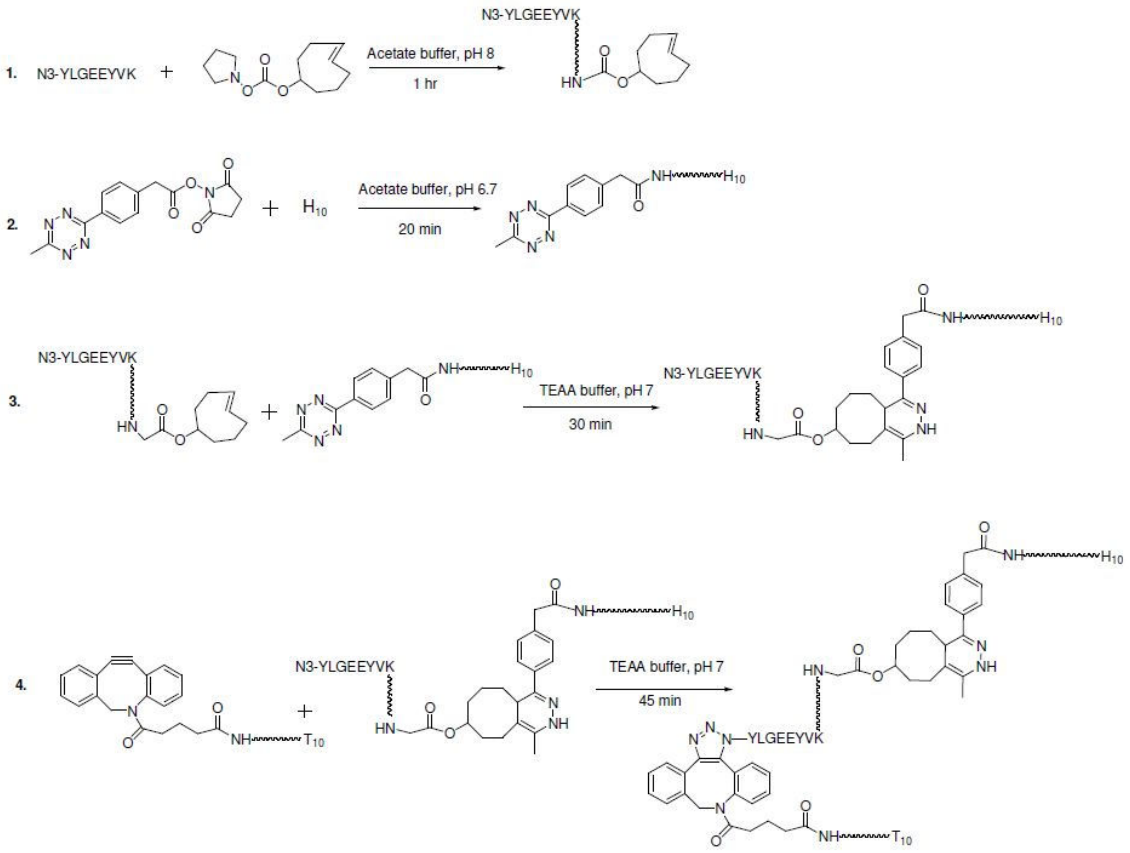


Figure 33. RP HPLC profiles of azidoacetic anhydride reacting with (A) lysozyme and (B) insulin at pH 6.7.

3.3.3 Developing chemistry to ligate different charged molecules on N- and C-termini of peptide

One new conjugate having opposite charges on N- and C-termini of peptide has been synthesized. DNA thread ($\mathbf{T_{10}}$) was attached to the N-termini and C-termini had a peptide having 10 histidine residues ($\mathbf{H_{10}}$). Histidines are expected behave as neutral molecule near pH 7 and is supposed to have positive charges below pH 6. Therefore, a pH gradient within the nanopore will lead to a construct ($\mathbf{T_{10}}$**-P-1-**$\mathbf{H_{10}}$) where two opposite charges will be generated on both the end of peptide.



Scheme 21. Scheme for synthesis of opposite charged molecules with peptide (1) **P-1** modification by NHS-transcyclooctene at the lysine side chain; (2) Modification of N-termini of peptide having 10 histidine residues (**H_{10}**) with NHS-methyl tetrazine; (3) Synthesis of **P-1-H_{10}**; (4) Synthesis of **T_{10}-P-1-H_{10}**

Synthesis of **T_{10}-P-1-H_{10}** is shown in scheme 21. First, **P-1** was modified by NHS-ester of trans cyclooctene at the lysine side chain. This reaction was carried out in acetate buffer (pH 8) over a period of 1hr. Starting material (**P-1**) was consumed and one new peak was generated with 6.8 min retention time which was characterized to be the peptide modified product (**P-1-Transcyclooctene**) (Figure 34). Then, **H_{10}** was modified at the N-

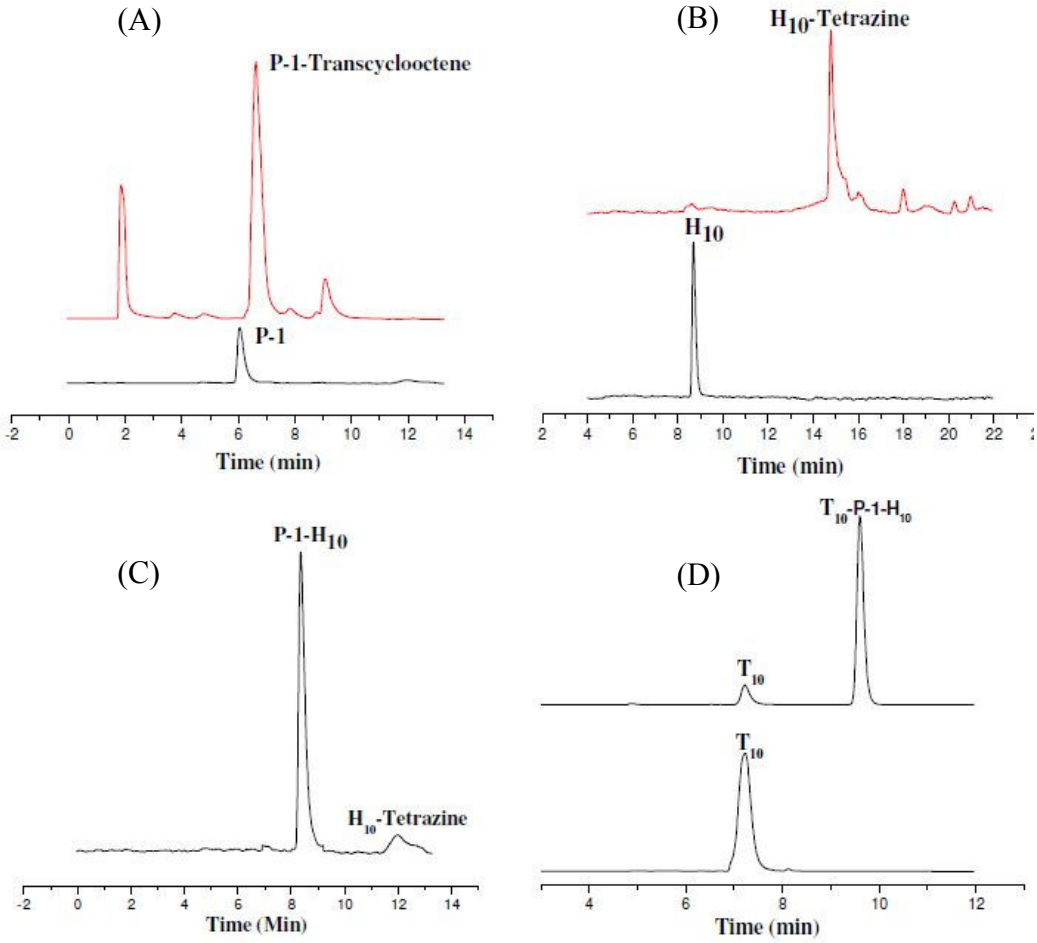


Figure 34. HPLC diagram for synthesis of various conjugates (A) **P-Transcyclocotene**; (B) **H_{10}-Tetrazine**; (C) **P-1-H_{10}**; (D) **T_{10}-P-1-H_{10}**

terminus by reacting **H_{10}** with NHS ester of methyl tertrazine. HPLC diagram showed that a new peak at 15 min retention was **H_{10}-Tetrazine**.

Next, **P-1** was coupled to **H_{10}** to give **P-1-H_{10}** conjugate. This reaction was performed in TEAA buffer (pH 7) for 30 min. **H_{10}-Tetrazine** was used in excess amount in the reaction so that **P-1-Transcyclooctene** got consumed totally to give the new conjugate. Finally, P-1-H_{10} was reacted to T_{10}-ADIBO to give **T_{10}-P-1-H_{10}**. T_{10}-ADIBO was used in excess in the reaction and it took 45 min to get the reaction completed. Product peak was

found to have 9.8 min of retention time. All the products were characterized by MLADI-TOF.

3.3.4 Translocation of peptide-PolyT_{20} conjugates through solid-state nanopores

We devised a setup as illustrated in Figure 33-A to measure the molecular translocation, following a procedure we previously reported on measurements of DNA translocation through solid-state nanopores.[120] Final concentration of analyte solution ~ 1.0 μM. All of measurements were carried out in a 0.4 M KCl electrolyte solution buffered with 1.0 mM phosphate, pH 7.4. Figure 33-B shows three typical nanopores we used for the the translocation measurement. Their sizes and shapes are slightly different from one another, determined by TEM imaging. With each individual nanopore, we were able to finish measurements on translocation of a peptide-PolyT_{20} conjugate as well as its parent peptide and oligonucleotide before it became clogged. In a typical translocation experiment, we began with measuring following an order of PolyT_{20}, peptide-PolyT_{20} conjugate, and then peptide. Between the measurements, the nanopore was rinsed with

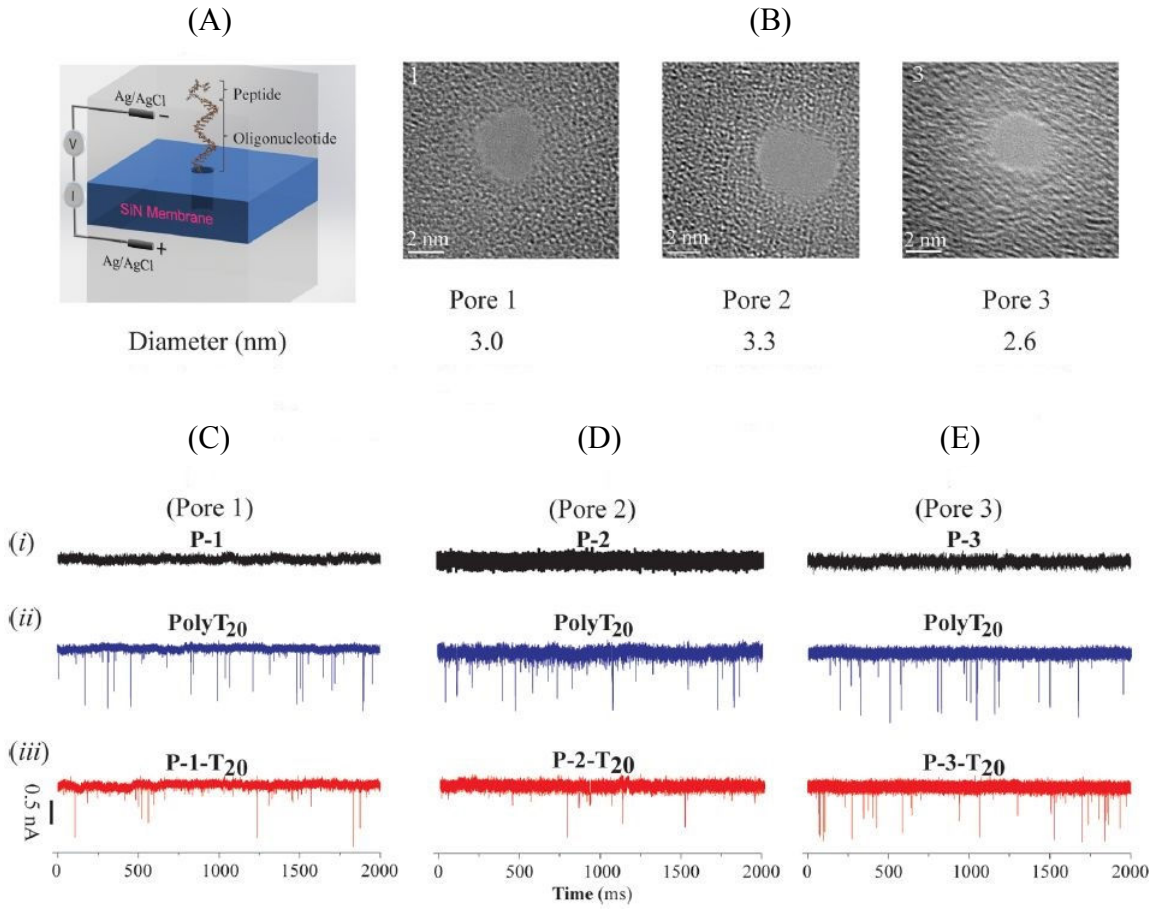


Figure 35. (A) Schematic illustration of a nanopore device for translocation measurements; (B) TEM images of the nanopores used for translocation and their physical parameters; (C) Ionic current traces of *i*: **P-1**, *ii:* **PolyT$_{20}$,** *iii*: **P-1-T$_{20}$** in Pore-1;(D) Ionic current traces of *i*: **P-2**, *ii*: **PolyT$_{20}$,** *iii*: **P-2-T$_{20}$** in Pore-2; (E) Ionic current traces of *i*: **P-3**, *ii*: **PolyT$_{20}$,** *iii*: **P-3-T$_{20}$** in Pore-3 (Bias: 500 mV; Analyte concentration: ~ 1.0 μM).[109]

the electrolyte solution to remove any analyte residue and restore it to the original conductive state. The raw data generated by these nanopores are shown in Figure 33-C (Pore-1), D (Pore-2), and E (Pore-3), respectively. We immediately notice that there was no translocation of peptides because neither negatively charged **P-1** nor positively charged **P-2** and **P-3** created spikes of current blockades (Figure 33, C-*i*, D-*i*, and E-*i*)

under our measurement conditions. As expected, PolyT$_{20}$ was readily translocated through these nanopoes (Figure 33, C-*ii*, D-*ii*, and E-*ii*), and so do the peptide-PolyT$_{20}$ conjugates (Figure 33, C-*iii*, D-*iii*, and E-*iii*). These current spikes were analyzed by means of the OpenNanopore software and each spike was assigned a dwell time and a current-blockade value. Median values of dwell time for all the samples (PolyT$_{20}$ and conjugates) are reported in the Table 6. The distributions change from pore to pore, as can be seen by comparing values for PolyT$_{20}$ between the three pores. However, by most

Table 6. Dwell times of PolyT$_{20}$ and its peptide conjugates in different nanopores.[109]

	Pore 1		Pore 2		Pore 3	
Sample	**P-1-T$_{20}$**	**PolyT$_{20}$**	**P-2-T$_{20}$**	**PolyT$_{20}$**	**P-3-T$_{20}$**	**P-1-T$_{20}$**
Median (μS)	60	70	60	90	70	70

measures, we see the somewhat surprising result that the conjugates translocate more rapidly, or on about the same time scale, as PolyT$_{20}$ alone (data for Pores 1 to 3).

Data for the current blockades give a clearer picture of the differences between PolyT$_{20}$ and its peptide conjugates than the widely distributed translocation times. In order to compare data for the various samples, each blockade-current data set was normalized by its maximum value and plotted into a normalized histogram, as shown in Figure 34. We consider first the PolyT$_{20}$ data. All the data sets were fitted by a double-peaked Gaussian function with $R^2 > 0.90$ (Figure 34), with a major peak at $\Delta I/I_0$ in the range of 0.15 to 0.2 and a minor peak in the range of 0.5 to 0.6. Similar features have been reported for DNA translocation through α-hemolysin pores where they were explained by the so-called

Christmas-tree effect.[121,122] When a single stranded DNA is translocated, it can thread either via its 5'- or its 3'-end into a nanopore. Meller and co-workers demonstrated by all-atom molecular dynamics (MD) that DNA bases in a stretched conformation preferably tilt toward the 5'-end in a confined pore.[121] This is because all of nucleosides in DNA have a β configuration, in which a nucleobase stays at the same side with the 5'-hydroxyl group of a nucleoside on the deoxyribose ring. As a result,

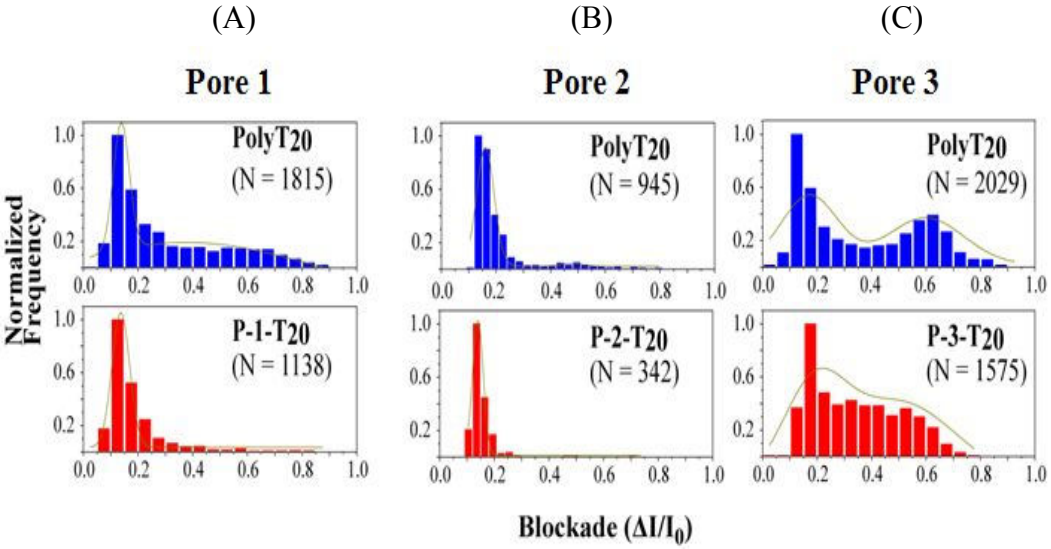


Figure 36. Histograms of fractional current blockades for PolyT20 and its peptide conjugates at bias 500 mV. (A) Translocation through Pore 1; (B) Translocation through Pore 2; (C) Translocation through Pore 3. Blue: $PolyT_{20}$, Red: peptide conjugate; $\Delta I = I_0$ (open pore current) – I (blocked pore current). The dark yellow lines in histograms are the Gaussian fitting curves and N = number of events.[109]

DNA translocation with threading through the 3'-end should be more frequent and less blockaded than through the 5'-end (just as a Christmas tree can be moved into a door more easily from its trunk end than from its tip). Based on this hypothesis, we may assign the major blockade peaks (at smaller current values) for translocation events through the 3'-end threading. From the literature, we found a similar trend in the translocation of

$PolyA_{20}$ through a α-HL nanopore.[123] Thus, when a short peptide is tethered to the 5'-end of $PolyT_{20}$, it reduces the probability of threading through the 5'-end. This accounts for the observation that the large current blockades were diminished in the translocation of the peptide-$PolyT_{20}$ conjugates compared to the corresponding $PolyT_{20}$. **P-3-T_{20}** displays a large tail following its major peak (Figure 34-C) and also translocates more slowly (Table 6) than the other two conjugates just mentioned above. This might be explained by the conformational structure of the peptide-$PolyT_{20}$ conjugate. The CD spectrum (Figure 32) shows that the peptide **P-3** has stronger interactions with the backbone of $PolyT_{20}$ in the conjugate than other two peptides, which may result in looped structures. When a conjugate with such a structure translocates through the nanopore, it would create larger blockade currents, as observed.

3.4 Experimental procedure

3.4.1 General information

Chemicals were purchased from Sigma-Aldrich and anhydrous organic solvents were Aldrich's Sure/Seal^{TM}. ADIBO-NHS ester was purchased from Click Chemistry Tools. Peptides were custom synthesized by CPC Scientific (San Jose, C, USA) and oligonucleotides ($PolyT_{20}$) with a C12 amino modifier at its 5' end by IDT (Integrated DNA Technologies). ^{1}H and ^{13}C NMR spectra were recorded at 400 MHz (^{1}H), 100 MHz (^{13}C), respectively. Chemical shifts are given in parts per million (ppm) on the delta scale (δ) and are referred to the solvent residual peak. HPLC analysis and purification were carried out in Agilent 1100 series equipped with a UV detector and a fraction collector. A Zorbax Eclipse Plus C18 column (4.6 x 150 mm, particle size 5 μm) from Agilent was used for the reversed phase HPLC. MALDI-TOF analysis was performed on Voyager-DE STR instrument.[109]

Azidoacetic anhydride synthesis

The synthesis was carried out following a method reported in literature with modification.[124] *N,N'*-Dicyclohexylcarbodiimide (DCC, 204 mg, 0.98 mmol) was added to a solution of 2-azidoacetic acid (200 mg, 1.98 mmol) in anhydrous tetrahydrofuran (2 ml). The solution was stirred for 2 h and 15 min, during which the precipitate was gradually produced, and filtered. The filtrate was concentrated by rotary evaporation, giving azidoacetic anhydride (150 mg, 42%) as a colorless liquid. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 3.85 (s, 4H); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 168.3, 50.0.

General procedure for reaction of azidoacetic anhydride with peptide (P-1, P-2, and P-3)

A solution of azidoacetic anhydride in acetonitrile (5 mM, 3.0 μL) was added to a solution of peptide in a 50 mM sodium acetate buffer (0.5 mM, 10 μL) with a predefined pH value in an eppendorf tube. The reaction was kept at 0 °C for 15 min, followed by the addition of water (10 μL), and remaining the reaction in ice for another 10 min for completely terminating the reaction. RP HPLC, with a gradient of 5 to 45% B in 25 min (solvent A: 0.1 M TEAA buffer, pH 7.0; solvent B: acetonitrile), was used to monitor the reactions and separate products.

Functionalization of PolyT_{20} with ADIBO (ADIBO-T_{20})

A solution of PolyT_{20} containing a C12 amino modifier at its 5'-end (1 mM, 10 μL) in water and DBCO-NHS ester in DMSO (15 mM, 80 μL) were mixed in a phosphate buffer (30 μL, pH 8). The solution was shaken for 20 min at room temperature. The product was purified by RP HPLC in a Zorbax Eclipse Plus C18 column (4.6 x 150 mm, particle size 5 μm) with a gradient of 10 to 60% B in 25 min (solvent A: 0.1 M TEAA buffer, pH 7.0; solvent B: acetonitrile). MALDI-TOF-MS calc. for $[C_{233}H_{304}N_{42}O_{143}P_{20}](M+H)$: *m/z* 6600.58; found: 6603.34.

Reaction of N-azidoacetylated P-1 with ADIBO-T_{20} (P-1-T_{20})

A solution of N-azidoacetylated **P-1** (30 μM, 15 μL) in a TEAA buffer (50 mM, pH 7) was mixed with ADIBO-T_{20} (30 μM, 15 μL) in the TEAA buffer (50 mM, pH 7), and shaken at room temperature for 45 min. RP HPLC analysis showed the starting material was consumed. The product was purified by RP HPLC in a Zorbax Eclipse Plus C18

column (4.6 x 150 mm, particle size 5 μm) with a gradient of 10 to 60% B in 25 min (solvent A: 0.1 M TEAA buffer, pH 7.0; solvent B: acetonitrile). After lyophilization, the product was given as a white powder. MALDI-TOF-MS calc. for [$C_{282}H_{374}N_{54}O_{159}P_{20}$] (M+H): *m/z* 7683.12; found: 7684.72.

Reaction of N-azidoacetylated P-2 with ADIBO-T_{20} (P-2-T_{20})

A solution of N-azidoacetylated **P-2** (30 μM, 15 μL) in a TEAA buffer (50 mM, pH 7) was mixed with ADBCO-T_{20} (30 μM, 25 μL) in the TEAA buffer (50 mM, pH 7), and shaken at room temperature for 20 min, purified by RP HPLC under the same conditions as did **P-1-T_{20}**. After lyophilization, the product was given as a white powder. MALDI-MS calc. for [$C_{297}H_{394}N_{62}O_{158}P_{20}$] (M+H): 7979.31; found: 7978.64).

Reaction of N-azidoacetylated P-3 with ADIBO-T_{20} (P-3-T_{20})

A solution of N-azidoacetylated P-3 (30 μM, 15 μL) in a TEAA buffer (50 mM, pH 7) was mixed with ADIBO-T_{20} (30 μM, 15 μL) in the TEAA buffer (50 mM, pH 7), and shaken at room temperature for 10 min, purified by RP HPLC under the same conditions as did for **P-1-T_{20}**. After lyophilization, the product was given as a white powder. MALDI-MS calc. for [$C_{299}H_{406}N_{60}O_{160}P_{20}$] (M+H): *m/z* 8019.39; found: 8020.58.

Reaction of azidoacetic anhydride with lysozyme

A solution of azidoacetic anhydride in acetonitrile (5 mM, 3.0 μL) was added to a solution of peptide in a 50 mM sodium acetate buffer (0.5 mM, 10 μL) with a predefined pH value in an eppendorf tube. The reaction was kept at 0 °C for 15 min, followed by the addition of water (10 μL), and remaining the reaction in ice for another 10 min for completely terminating the reaction. RP HPLC, with a gradient of 25 to 55% B in 25 min

(solvent A:0.1% TFA in water,; solvent B: 0.09% TFA in 80% acetonitrile/ 20% water), was used to monitor the reactions and separate products.

Reaction of azidoacetic anhydride with insulin

A solution of azidoacetic anhydride in acetonitrile (5 mM, 3.0 μL) was added to a solution of peptide in a 50 mM sodium acetate buffer (0.5 mM, 10 μL) with a predefined pH value in an eppendorf tube. The reaction was kept at 0 $^{\circ}$C for 15 min, followed by the addition of water (10 μL), and remaining the reaction in ice for another 20 min for completely terminating the reaction. RP HPLC, with a gradient of 10 to 55% B in 25 min (solvent A:0.1% TFA in water,; solvent B: 0.09% TFA in 80% acetonitrile/ 20% water), was used to monitor the reactions and separate products.

Synthesis of P-1-Trans cyclooctene

A solution of NHS-Trans cyclooctene in acetonitrile (5 mM, 3.0 μL) was added to a solution of peptide in a 50 mM sodium acetate buffer (0.5 mM, 10 μL) with pH 8 in an eppendorf tube. The reaction was kept at room temperature for 1 h. RP HPLC, with a gradient of 10 to 45% B in 25 min (solvent A: 0.1 M TEAA buffer, pH 7.0; solvent B: acetonitrile), was used to monitor the reactions and separate products. MALDI-TOF-MS calc: *m/z* 1235.51; found: 1236.38.

Synthesis of H_{10}-Tetrazine

A solution of NHS-Tetrazine in acetonitrile (5 mM, 3.0 μL) was added to a solution of peptide in a 50 mM sodium acetate buffer (0.5 mM, 10 μL) with pH 6.7 in an eppendorf tube. The reaction was kept at room temperature for 20 min. RP HPLC, with a gradient of 10 to 40% B in 25 min (solvent A:0.1% TFA in water,; solvent B: 0.09% TFA in 80%

acetonitrile/ 20% water), was used to monitor the reactions and separate products. . MALDI-TOF-MS calc: *m/z* 1601.23; found: 1602.44.

Synthesis of P1-H_{10}

A solution of **P-1-Trans cyclooctene** (20 μM, 20 μL) in a TEAA buffer (50 mM, pH 7) was mixed with **H_{10}-Tetrazine** (20 μM, 25 μL) in the TEAA buffer (50 mM, pH 7), and shaken at room temperature for 30 min. The product was purified by RP HPLC in a Zorbax Eclipse Plus C18 column (4.6 x 150 mm, particle size 5 μm) with a gradient of 10 to 50% B in 25 min (solvent A: 0.1 M TEAA buffer, pH 7.0; solvent B: acetonitrile). After lyophilization, the product was given as a white powder. . MALDI-TOF-MS calc: *m/z* 2836.18; found: 2837.20.

Synthesis of T_{10}-P1-H_{10}

A solution of **T_{10}-ADIBO** (25 μM, 25 μL) in a TEAA buffer (50 mM, pH 7) was mixed with **H_{10}-Tetrazine** (25 μM, 15 μL) in the TEAA buffer (50 mM, pH 7), and shaken at room temperature for 30 min. The product was purified by RP HPLC in a Zorbax Eclipse Plus C18 column (4.6 x 150 mm, particle size 5 μm) with a gradient of 10 to 50% B in 25 min (solvent A: 0.1 M TEAA buffer, pH 7.0; solvent B: acetonitrile). After lyophilization, the product was given as a white powder. . MALDI-TOF-MS calc: *m/z* 5978.50; found: 5980.38.

3.4.2 Tandem mass spectrometry

Collisional-induced dissociation (CID) and subsequent mass analysis of modified peptides were carried out on a Bruker MaXis 4G quadrupole-time-of-flight (Q-TOF) mass spectrometer equipped with a microflow nebulizer electrospray ionization (ESI)

source operated in positive ion mode. HPLC-purified peptide samples were diluted to a final concentration of approximately 20 μM in a 50/50 mixture of acetonitrile/water containing 0.1% (v/v) formic acid. Prior to analysis, the TOF mass analyzer was externally calibrated with a tuning mixture supplied by Agilent containing 10 species of varying but equally spaced masses from 118 to 2722 Da. Peptide solutions were infused into the ion source by syringe pump at a rate of 1.0 μL/min. The end plate offset and capillary were set to potentials of 500 V and 4,500 V, respectively. The nebulizer gas and the dry gas (both N_2) were set to 1.4 Bar and 4.0 L/min, respectively, and the dry gas temperature was set to 220 °C. The RF amplitude on ion funnel 1 and the multipole were both set to 400 Vpp. No in-source CID energy was used. Quadrupole ion energy was set to 3.0 eV. Precursor ions were selected with a *m/z* width of 2 and imparted enough energy (usually 10 – 40 eV) to diminish the relative abundance of the precursor ion to less than 5%. Spectral digitzation was set to a rate of 4 GHz and individual TOF transients were summed and recorded at a rate of 1 Hz. After equilibration and spray stabilization, each mass spectrum was recorded for 1.5 minutes then averaged into a single spectrum.

3.5 Summaries

The objective of this present study was to develop simple and effective chemistry to functionalize peptides with a charged threading molecule to facilitate their translocation through nanopores. We have harnessed an acylation reaction for rapid introduction of an orthogonal azido function to N-termini of peptides, which allowed us to quantitatively attach charged oligonucleotides to peptides using a click reaction without the need of separating the intermediate product. At pH ~ 6.7 and 0 °C, azidoacetic anhydride quickly

reacted with the α amine of a peptide containing one lysine residue with > 90% selectivity, and the side reaction with imidazolyl ring of histidine can also be reduced to minimum. Thus, our chemistry should be practical for use in preparing peptides sample obtained from trypsin digests for nanopore analysis. The selectivity is less (~ 80%) in a peptide containing three lysine residues, so further improvements are required if the method is to be applied more widely to peptides from other sources.

Furthermore, we have shown that the peptide-PolyT_{20} conjugates can effectively translocate through solid-state nanopores, which lays down a foundation for us to develop a technique for analysis of proteins using nanopores. Our data indicate that the peptide-PolyT_{20} conjugates preferably enter the nanopore from the 3' end of PolyT_{20}. As a result, we conclude that an oligonucleotide can be used as a molecular thread to carry peptides through solid-state nanopores. While the sensitivity of ion-current measurements is unlikely to be adequate for de novo sequencing, it is clear that structural aspects of the conjugates can be probed. We hope to integrate this technology with recognition tunneling in order to explore the possibility of sequencing with single amino acid resolution in the future.[100]

References

(1) Cheng, S.; Bryant, R.; Doerr, S. H.; Rhodri, W. P.; Wright, C. J. *J Microsc* **2008**, *231*, 384.

(2) Mueller, D. J.; Schabert, F. A.; Bueldt, G.; Engel, A. *Biophys. J.* **1995**, *68*, 1681.

(3) Binning, G.; Rohrer, H.; Gerber, C.; Weibel, E. *Phys. Rev. Lett.* **1982**, *49*, 57.

(4) Israelachvili, J. N. *Intermolecular and Surface Forces*; McGraw-Hill Publishing Co. Japan, Ltd., 1991.

(5) Ashkin, A.; Dziedzic, J. M.; Yamane, T. *Nature* **1987**, *330*, 769.

(6) Ashkin, A. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4853.

(7) Lundstrom, K. *Methods Mol. Biol. (Totowa, NJ, U. S.)* **2009**, *552*, 51.

(8) Dharap, S. S.; Minko, T. *Pharm. Res.* **2003**, *20*, 889.

(9) Dharap, S. S.; Wang, Y.; Chandna, P.; Khandare, J. J.; Qiu, B.; Gunaseelan, S.; Sinko, P. J.; Stein, S.; Farmanfarmaian, A.; Minko, T. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12962.

(10) Zhang, J.; Wu, G.; Song, C.; Li, Y.; Qiao, H.; Zhu, P.; Hinterdorfer, P.; Zhang, B.; Tang, J. *J. Phys. Chem. B* **2012**, *116*, 13331.

(11) Harada, Y.; Kuroda, M.; Ishida, A. *Langmuir* **2000**, *16*, 708.

(12) Stuart, J. K.; Hlady, V. *Langmuir* **1995**, *11*, 1368.

(13) Sharon, N.; Lis, H. *Science* **1989**, *246*, 227.

(14) Glynn, L. E. *Cell Biochemistry and Function* **1990**, *8*, 190.

(15) Macholz, R. *Food / Nahrung* **1988**, *32*, 212.

(16) Singh, R. S.; Tiwary, A. K.; Kennedy, J. F. *Crit. Rev. Biotechnol.* **1999**, *19*, 145.

(17) Touhami, A.; Hoffmann, B.; Vasella, A.; Denis, F. A.; Dufrene, Y. F. *Langmuir* **2003**, *19*, 1745.

(18) Leite, L. F.; Mattoso, C. H.; Oliveira Jr. N.O; Herrmann Jr, S. P. P. *Modern Research and Educational Topics in Microscopy.* **2007**, A. Mendez-Vilas and J.Diaz (Eds), 747.

(19) Lin, L.; Wang, H.; Liu, Y.; Yan, H.; Lindsay, S. *Biophys. J.* **2006**, *90*, 4236.

(20) Heinz, W. F.; Hoh, J. H. *Trends Biotechnol.* **1999**, *17*, 143.

(21) Homola, J.; Yee, S. S.; Gauglitz, G. *Sens. Actuators, B* **1999**, *54*, 3.

(22) Owen, V. M. *Biosens. Bioelectron.* **1997**, *12*, v.

(23) Kim, M.; Park, K.; Jeong, E.-J.; Shin, Y.-B.; Chung, B. H. *Anal. Biochem.* **2006**, *351*, 298.

(24) Madeira, A.; Vikeved, E.; Nilsson, A.; Sjogren, B.; Andren, P. E.; Svenningsson, P. *Curr Protoc Protein Sci* **2011**, *Chapter 19*, Unit19.21.

(25) Majka, J.; Speck, C. *Adv. Biochem. Eng./Biotechnol.* **2007**, *104*, 13.

(26) Teh, H. F.; Peh, W. Y. X.; Su, X.; Thomsen, J. S. *Biochemistry* **2007**, *46*, 2127.

(27) Beccati, D.; Halkes, K. M.; Batema, G. D.; Guillena, G.; Carvalho de Souza, A.; van Koten, G.; Kamerling, J. P. *ChemBioChem* **2005**, *6*, 1196.

(28) Bakhtiar, R. *J. Chem. Educ.* **2013**, *90*, 203.

(29) Nguyen, H. H.; Park, J.; Kang, S.; Kim, M. *Sensors* **2015**, *15*, 10481.

(30) Albery, W. J.; Knowles, J. R. *Biochemistry* **1976**, *15*, 5627.

(31) Nelson, N.; Ben-Shem, A. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 971.

(32) Honjo, T.; Habu, S. *Annu Rev Biochem* **1985**, *54*, 803.

(33) Zalipsky, S. *Bioconjugate Chem.* **1995**, *6*, 150.

(34) Baker, D. P.; Lin, E. Y.; Lin, K.; Pellegrini, M.; Petter, R. C.; Chen, L. L.; Arduini, R. M.; Brickelmaier, M.; Wen, D.; Hess, D. M.; Chen, L.; Grant, D.; Whitty, A.; Gill, A.; Lindner, D. J.; Pepinsky, R. B. *Bioconjugate Chem.* **2006**, *17*, 179.

(35) Gao, W.; Liu, W.; Mackay, J. A.; Zalutsky, M. R.; Toone, E. J.; Chilkoti, A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 15231.

(36) Miller, R. A.; Presley, A. D.; Francis, M. B. *J. Am. Chem. Soc.* **2007**, *129*, 3104.

(37) Dedeo, M. T.; Duderstadt, K. E.; Berger, J. M.; Francis, M. B. *Nano Lett.* **2010**, *10*, 181.

(38) Esser-Kahn, A. P.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2008**, *130*, 15820.

(39) Tilley, S. D.; Joshi, N. S.; Francis, M. B.; John Wiley & Sons, Inc.: **2009**; Vol. 4, p 158.

(40) R. F. Doolittle *in Prediction of Protein Structure and the Principles of Protein Conformation* (Ed.: G. D. Fasman), Plenum, NewY ork, **1989**.

(41) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science (Washington, D. C.)* **1994**, *266*, 776.

(42) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249.

(43) Joshi, N. S.; Whitaker, L. R.; Francis, M. B. *J. Am. Chem. Soc.* **2004**, *126*, 15942.

(44) Antos, J. M.; McFarland, J. M.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2009**, *131*, 6301.

(45) Li, X.; Zhang, L.; Hall, S. E.; Tam, J. P. *Tetrahedron Lett.* **2000**, *41*, 4069.

(46) Tam, J. P.; Miao, Z. *J. Am. Chem. Soc.* **1999**, *121*, 9013.

(47) Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. *Nat. Chem. Biol.* **2007**, *3*, 321.

(48) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science.* **2001**, *292*, 498.

(49) Link, A. J.; Mock, M. L.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2003**, *14*, 603.

(50) Drews, J. *Science.* **2000**, *287*, 1960.

(51) Wildling, L.; Unterauer, B.; Zhu, R.; Rupprecht, A.; Haselgrubler, T.; Rankl, C.; Ebner, A.; Vater, D.; Pollheimer, P.; Pohl, E. E.; Hinterdorfer, P.; Gruber, H. J. *Bioconjugate Chem.* **2011**, *22*, 1239.

(52) Imai, K.; Takaoka, A. *Nat. Rev. Cancer* **2006**, *6*, 714.

(53) Yu, X.; Narayanan, S.; Vazquez, A.; Carpizo, D. R. *Apoptosis* **2014**, *19*, 1055.

(54) Cox, A. D.; Fesik, S. W.; Kimmelman, A. C.; Luo, J.; Der, C. J. *Nat. Rev. Drug Discovery* **2014**, *13*, 828.

(55) Dang, C. V. *Cell.* **2012**, *149*, 22.

(56) London, N.; Raveh, B.; Schueler-Furman, O. *Curr. Opin. Chem. Biol.* **2013**, *17*, 952.

(57) Chan, A. C.; Carter, P. J. *Nat. Rev. Immunol.* **2010**, *10*, 301.

(58) McEnaney, P. J.; Fitzgerald, K. J.; Zhang, A. X.; Douglass, E. F., Jr.; Shan, W.; Balog, A.; Kolesnikova, M. D.; Spiegel, D. A. *J. Am. Chem. Soc.* **2014**, *136*, 18034.

(59) Rader, C. *Nature.* **2015**, *518*, 38.

(60) Vauquelin, G.; Charlton, S. J. *Br. J. Pharmacol.* **2013**, *168*, 1771.

(61) Hill, Z. B.; Perera, B. G. K.; Maly, D. J. *J. Am. Chem. Soc.* **2009**, *131*, 6686.

(62) Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R. *Science (Washington, D. C., 1883-)* **1989**, *243*, 85.

(63) Stayton, P. S.; Freitag, S.; Klumb, L. A.; Chilkoti, A.; Chu, V.; Penzotti, J. E.; To, R.; Hyre, D.; Le Trong, I.; Lybrand, T. P.; Stenkamp, R. E. *Biomol. Eng.* **1999**, *16*, 39.

(64) Choi, S.-K. *Synthetic Multivalent Molecules: Concepts and Biomedical Applications*; John Wiley & Sons, Inc.: Hoboken, New Jersey, 2004

(65) Mascotti, D. P.; Lohman, T. M. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 3142.

(66) Harries, D.; May, S.; Ben-Shaul, A. *Soft Matter* **2013**, *9*, 9268.

(67) Le Trong, I.; Wang, Z.-Z.; Hyre, D. E.; Lybrand, T. P.; Stayton, P. S.; Stenkamp, R. E. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 813.

(68) Taylor, S. K.; Wang, J.; Kostic, N.; Stojanovic, M. N. *Angew. Chem., Int. Ed.* **2013**, *52*, 5509.

(69) Chinchilla, R.; Najera, C. *Chem Rev* **2007**, *107*, 874.

(70) Chinchilla, R.; Najera, C. *Chem Soc Rev* **2011**, *40*, 5084.

(71) Pollock, J. B.; Cook, T. R.; Stang, P. J. *J. Am. Chem. Soc.* **2012**, *134*, 10607.

(72) Mamat, C.; Flemming, A.; Koeckerling, M.; Steinbach, J.; Wuest, F. R. *Synthesis* **2009**, 3311.

(73) Tao, F.; Bernasek, S. L. *Chem. Rev. (Washington, DC, U. S.)* **2007**, *107*, 1408.

(74) Duan, X.; Li, Y.; Rajan, N. K.; Routenberg, D. A.; Modis, Y.; Reed, M. A. *Nat. Nanotechnol.* **2012**, *7*, 401.

(75) Perez-Luna, V. H.; O'Brien, M. J.; Opperman, K. A.; Hampton, P. D.; Lopez, G. P.; Klumb, L. A.; Stayton, P. S. *J. Am. Chem. Soc.* **1999**, *121*, 6469.

(76) Gomez-Casado, A.; Dam, H. H.; Yilmaz, M. D.; Florea, D.; Jonkheijm, P.; Huskens, J. *J. Am. Chem. Soc.* **2011**, *133*, 10849.

(77) Senapati, S.; Manna, S.; Lindsay, S.; Zhang, P. *Langmuir* **2013**, *29*, 14622.

(78) Wang, H.; Bash, R.; Yodh, J. G.; Hager, G. L.; Lohr, D.; Lindsay, S. M. *Biophys. J.* **2002**, *83*, 3619.

(79) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627.

(80) Guo, S.; Ray, C.; Kirkpatrick, A.; Lad, N.; Akhremitchev, B. B. *Biophys. J.* **2008**, *95*, 3964.

(81) Loiseau, F. A.; Hii, K. K.; Hill, A. M. *J. Org. Chem.* **2004**, *69*, 639.

(82) Manna, S.; Senapati, S.; Lindsay, S.; Zhang, P. *J. Am. Chem. Soc.* **2015**, *137*, 7415.

(83) Swartz, J. D.; Gulka, C. P.; Haselton, F. R.; Wright, D. W. *Langmuir* **2011**, *27*, 15330.

(84) Nguyen, T.; Byrd, E.; Bentz, D. *J. Adhes.* **1995**, *48*, 169.

(85) Kawaguchi, T.; Walker, K. L.; Wilkins, C. L.; Moore, J. S. *J. Am. Chem. Soc.* **1995**, *117*, 2159.

(86) Alley, S. C.; Okeley, N. M.; Senter, P. D. *Curr. Opin. Chem. Biol.* **2010**, *14*, 529.

(87) Esser-Kahn, A. P.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2008**, *130*, 15820.

(88) Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. *Angew. Chem., Int. Ed.* **2006**, *45*, 5307.

(89) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. *J. Am. Chem. Soc.* **2014**, *136*, 9572.

(90) Chan, A. O.-Y.; Ho, C.-M.; Chong, H.-C.; Leung, Y.-C.; Huang, J.-S.; Wong, M.-K.; Che, C.-M. *J. Am. Chem. Soc.* **2012**, *134*, 2589.

(91) Laszlo, A. H.; Derrington, I. M.; Ross, B. C.; Brinkerhoff, H.; Adey, A.; Nova, I. C.; Craig, J. M.; Langford, K. W.; Samson, J. M.; Daza, R.; Doering, K.; Shendure, J.; Gundlach, J. H. *Nat. Biotechnol.* **2014**, *32*, 829.

(92) Oukhaled, A.; Bacri, L.; Pastoriza-Gallego, M.; Betton, J.-M.; Pelta, J. *ACS Chem. Biol.* **2012**, *7*, 1935.

(93) Fennouri, A.; Przybylski, C.; Pastoriza-Gallego, M.; Bacri, L.; Auvray, L.; Daniel, R. *ACS Nano* **2012**, *6*, 9672.

(94) McMullen, A.; de Haan, H. W.; Tang, J. X.; Stein, D. *Nat. Commun.* **2014**, *5*, 4171.

(95) Reiner, J. E.; Balijepalli, A.; Robertson, J. W. F.; Campbell, J.; Suehle, J.; Kasianowicz, J. J. *Chem. Rev. (Washington, DC, U. S.)* **2012**, *112*, 6431.

(96) Pennisi, E. *Science.* **2014**, *343*, 829.

(97) Mikheyev, A. S.; Tin, M. M. Y. *Mol. Ecol. Resour.* **2014**, *14*, 1097.

(98) Ashton, P. M.; Nair, S.; Dallman, T.; Rubino, S.; Rabsch, W.; Mwaigwisya, S.; Wain, J.; O'Grady, J. *Nat. Biotechnol.* **2015**, *33*, 296.

(99) Jain, M.; Fiddes, I. T.; Miga, K. H.; Olsen, H. E.; Paten, B.; Akeson, M. *Nat. Methods* **2015**, *12*, 351.

(100) Zhao, Y.; Ashcroft, B.; Zhang, P.; Liu, H.; Sen, S.; Song, W.; Im, J.; Gyarfas, B.; Manna, S.; Biswas, S.; Borges, C.; Lindsay, S. *Nat. Nanotechnol.* **2014**, *9*, 466.

(101) Ohshiro, T.; Tsutsui, M.; Yokota, K.; Furuhashi, M.; Taniguchi, M.; Kawai, T. *Nat Nanotechnol* **2014**, *9*, 835.

(102) Krasniqi, B.; Lee, J. S. *PLoS One* **2014**, *9*, e88004/1.

(103) Nivala, J.; Marks, D. B.; Akeson, M. *Nat. Biotechnol.* **2013**, *31*, 247.

(104) Rodriguez-Larrea, D.; Bayley, H. *Nat. Nanotechnol.* **2013**, *8*, 288.

(105) Steen, H.; Mann, M. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 699.

(106) Pang, P.; Ashcroft, B. A.; Song, W.; Zhang, P.; Biswas, S.; Qing, Q.; Yang, J.; Nemanich, R. J.; Bai, J.; Smith, J. T.; Reuter, K.; Balagurusamy, V. S. K.; Astier, Y.; Stolovitzky, G.; Lindsay, S. *ACS Nano* **2014**, *8*, 11994.

(107) Venta, K.; Shemer, G.; Puster, M.; Rodriguez-Manzo, J. A.; Balan, A.; Rosenstein, J. K.; Shepard, K.; Drndic, M. *ACS Nano* **2013**, *7*, 4629.

(108) Si, W.; Sha, J. J.; Liu, L.; Qiu, Y. H.; Chen, Y. F. *Sci. China: Technol. Sci.* **2013**, *56*, 2398.

(109) Biswas, S.; Song, W.; Borges, C.; Lindsay, S.; Zhang, P. *ACS Nano* **2015**, *9*, 9652.

(110) Lundblad, R. L. *Chemical Reagents for Protein Modification*, 4th ed.; CRC Press: Boca Raton, FL, **2014**.

(111) Witus, L. S.; Netirojjanakul, C.; Palla, K. S.; Muehl, E. M.; Weng, C.-H.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2013**, *135*, 17223.

(112) Song, J.; Kim, H.-J. *Anal. Biochem.* **2012**, *423*, 269.

(113) Koehler, C. J.; Arntzen, M. O.; Strozynski, M.; Treumann, A.; Thiede, B. *Anal. Chem. (Washington, DC, U. S.)* **2011**, *83*, 4775.

(114) Debets, M. F.; van Berkel, S. S.; Schoffelen, S.; Rutjes, F. P. J. T.; van Hest, J. C. M.; van Delft, F. L. *Chem. Commun. (Cambridge, U. K.)* **2010**, *46*, 97.

(115) Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16793.

(116) Gounaris, A. D.; Perlmann, G. E. Succinylation of Pepsinogen. *J. Biol. Chem.* **1967**, *242*, 2739–2745.

(117) Hermanson, G. T. Bopconjugate Techniques, 2nd ed.; Elsevier: Amsterdam, 2008.

(118) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. A. *Protein Sci.* **2009**, *18*, 247–251.

(119) Mentinova, M.; Barefoot, N. Z.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom*. **2012**, *23*, 282–289.

(120) Krishnakumar, P.; Gyarfas, B.; Song, W.; Sen, S.; Zhang, P.; Krstic, P.; Lindsay, S. *ACS Nano* **2013**, *7*, 10319.

(121) Mathe, J.; Aksimentiev, A.; Nelson, D. R.; Schulten, K.; Meller, A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12377.

(122) Muzard, J.; Martinho, M.; Mathe, J.; Bockelmann, U.; Viasnoff, V. *Biophys. J.* **2010**, *98*, 2170.

(123) Ying, Y.-L.; Li, D.-W.; Liu, Y.; Dey, S. K.; Kraatz, H.-B.; Long, Y.-T. *Chem. Commun. (Cambridge, U. K.)* **2012**, *48*, 8784.

(124) Grajkowski, A.; Cieslak, J.; Gapeev, A.; Schindler, C.; Beaucage, S. L. *Bioconjugate Chem.* **2010**, *21*, 2147.

APPENDIX A

COPYRIGHT PERMISSION