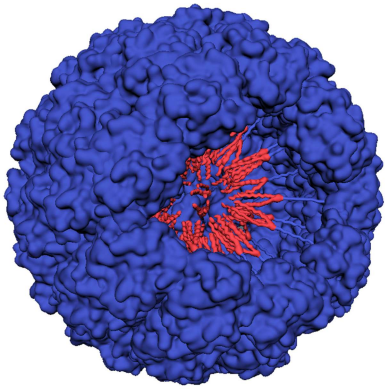
Rational Design, Synthesis and Self-assembly Studies of Facially Amphiphilic Proteins (FAPs)



A THESIS
SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT OF THE DEGREE OF THE
DOCTOR OF PHILOSOPHY
BY

MOHAN M
ID.NO. 20143333

Department of Chemistry
Indian Institute of Science Education and Research
Pune: 411008 (MH), India.

I would like to dedicate
this dissertation
to

My daughter
Aanya Reddy

and

My wife
Mrs. Sumamohan Reddy

for their love and care

Acknowledgment

First and foremost, I want to express my deepest and sincere gratitude to my supervisor **Dr. Britto S Sandanaraj**, *whose encouragement, constant support, intellectual guidance, ideas, suggestions, and affectionate behavior pushed our project in the right direction. We had to go through a tough time during initial days, and his "never give up attitude" made us succeed. I am fortunate to be his first student and was able to learn more than any later students can learn from a mentor. I have experienced several things, like setting up a lab from scratch, etc. He always took mentoring very seriously, training us rigorously, often pushing us to limits to bring the best in us. On the other side, I often find myself a part of his family too. He is exceptionally kind and helpful whenever I needed him as a friend and family member. He taught the essential lessons in life.* **"I can never forget the time I spent with him."**

I am grateful to my RAC members **Dr. Guru**, *and* **Dr. Amrita**, *who helped my research with their relevant suggestions and valuable feedback during RAC meetings. I am also grateful to the chemistry department, IISER, especially* **Dr. Amrita, Dr. Jayakannan, Dr. H N Gopi, Dr. Pramod, Dr. Pinaki, Dr. Harinath, Dr. Srivatsan** *and their groups who helped my research with their timely help during my entire Ph.D. I am also grateful to the biology department, IISER, especially* **Dr. Gayatri, Dr. Sai, Dr. Girish, Dr. Thomas,** *and their groups who also helped me. I would also like to thank all the faculties and their groups from* **IISER**. *I am also grateful to the entire staff of* **IISER** *Pune.*

I would like to thank **Dr. Radha Chauhan** *and her group (***NCCS***),* **Dr. Rahul Bambure,** *and his group (***NCL***) and our collaborators* **Dr. Sugam, Dr. Vinod** *and their group* **(BARC)** *for helping us with their SEC-MALS, MALDI-TOF AND SANS facilities respectively. I would like to extend a special thanks to* **G.E. Health care and Mr. Bruno Thomas** *for their help with the FPLC facility.*

Last but not least, in the formal list, I would like to thank the funding agencies **CSIR-UGC** *and* **Infosys** *for the fellowship and travel support;* **DST** *and* **DBT** *for the funding and* **IISER Pune** *for providing us the funding and a high-quality research environment.*

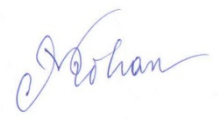
We have some of the fantastic people in my lab. With great pleasure and immense satisfaction, I would like to extend a special thanks to **Pavan**, **Jagan,** *and* **Punita** *for being by my side throughout my Ph.D. They have been not just my labmates but my friends too, and especially with* **Pavan**, *with whom I share plenty of memorable moments. I would like to thank*

my other labmates, Arjun, Rasika, Dr. Swagata, Akshay L, Akshay K, for their kind cooperation and constant encouragement.

I would like to honor and express my sincere gratitude to my M.S. (Pharm), Medicinal Chemistry, Advisor: Dr. Atul Goel. It's a great pleasure and immense satisfaction to express my heartful thanks to my old seniors Mr. Ashutosh Raghuvanshi, Ms. Amrita Parihar, Mr. Gaurav Taneja, Mr. Pankaj Nag, Mr. Ashutosh Sharma, Ms. Shahida Umar, Mr. Ajay Kumar Jha, Ms. Shachi Mishra, Mr.Chandra Prakash Gupta, Ms. Vartika Vaishya. I would always remember, respect, and appreciate them. I would also like to thank NIPER Raebareli and CSIR-CDRI as well for my previous basic training in science and research.

Finally, I would like to express my heartfelt thanks to my wife, Mrs. Sumamohan Reddy. I might not have achieved this milestone if not for her constant patience, trust, sacrifice, support, motivation, and faith in me. She is the most important person in my life who helped me through this journey. She was with me, and she has been my motivation from day one of my Ph.D. I would like to thank my daughter Aanya for her help too. She only came into my life in the middle of my Ph.D., and definitely helped me in finishing my Ph.D. She has a special place in my heart. I would like to thank Ashini, Rithanya, as well, Ashini has been our world, even before Aanya and she will always be. Rithanya is the youngest, whom I could not even meet in person because of my Ph.D. I can't wait to meet her. I want to thank my brother Mr. Vijay Reddy for his endless love and help in the toughest times I had to go through.

Last but not least, I would like to express my deep gratitude to my Parents Mr. Sambasiva Reddy, Mrs. Leelavathi, Mr. Narayan Reddy, Mrs. Sujatha, and family members, Mr. Anil Reddy, Mrs. Sushma, Mrs. Thulasi, for their care, love, and support. This dream would have remained a dream without my family members. They are my real strength. I would also like to admire my late grandparents Mr. Venakata Subba Reddy, Mrs. Kanthamma, Mr. Chinna Muni Reddy, Mrs. Eswaramma, and the Almighty for their blessings.



(Mohan M)

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the institute and can also evoke penal action from the sources which have thus not been adequately cited or from whom proper permission has not been taken when needed.

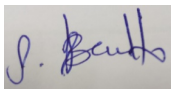


Date: 29/06/2020
Mohan M
20143333

भारतीयविज्ञानशिक्षाएिअनुसंधानसंस्थान, पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH (IISER), PUNE
(An Autonomous Institution, Ministry of Human Resource Development, Govt. of India)
900 NCL Innovation Park Dr Homi Bhabha Road Pune 411008

CERTIFICATE

*This is certifying that the work incorporated in this thesis entitled, "**Rational Design, Synthesis and Self-assembly Studies of Facially Amphiphilic Proteins (FAPs)**" submitted by **Mohan M** was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.*



Dr. Britto S Sandanaraj, Ph.D.
Assistant Professor – Chemistry
IISER Pune (M.H.), India.

Table of Contents

S.No	Title	Page No.
1.	Chapter 1. Introduction to Protein Engineering	1
1.1.	Engineering Natural Protein Cages	3
1.1.1.	Viruses	3
	1.1.1.1. Cowpea chlorotic mottle virus (CCMV)	6
	1.1.1.2. cowpea mosaic virus (CPMV)	7
	1.1.1.3. Qβ bacteriophage	7
	1.1.1.4. Tobacco mosaic virus (TMV)	9
	1.1.1.5. Bacteriophage MS2	10
	1.1.1.6. Hepatitis B virus (HBV)	11
1.1.2.	Other cages/compartments	12
	1.1.2.1. Bacterial compartments	12
	1.1.2.2. Ferritins	13
	1.1.2.3. Small heat shock proteins (sHsp)	15
	1.1.2.4. Vault proteins	15
1.2.	Engineering Natural Protein Oligomers	18
1.2.1.	Computational methods	18
1.2.2.	Symmetric oligomers-based fusion methods	20
1.2.3.	Symmetric oligomers and oligomeric coiled-coil-based fusion methods	22
1.2.4.	Coiled-coils-based protein origami	24
1.2.5.	Coiled-coil-ELP-based methods	24
1.2.6.	Coiled-coil-based other methods	26
1.2.7.	Hexameric HTHP hemeprotein based methods	26
1.3.	Engineering Monomeric Proteins	28
1.3.1.	FLP fusion-based methods	28
1.3.2.	Heme-hemeprotein-based methods	29
1.3.3.	Enzyme-inhibitor-based methods	31
1.3.4.	Other fusion-based methods	32
1.3.5.	Chemical strategies/protein-polymer conjugates	33
	1.3.5.1. "Grafting to" approaches	33
	1.3.5.2. "Grafting from" approaches	35
	1.3.5.3. Smart protein-polymer conjugates	37
	1.3.5.4. Cross-linked protein-polymer conjugates (proteinosomes)	38
2.	Chapter 2. FAPs and MAPLab Technology Development	52
2.1.	Synthesis and Self-assembly Studies of Prototype FAP	53
2.1.1.	Design	54

2.1.2.	MALDI-TOF MS optimization	58
2.1.3.	Optimization of synthesis and conjugation	59
2.1.4.	Purification optimization	69
2.1.5.	Biophysical characterization	74
2.1.6.	Self-assembly studies	76
2.1.7.	Discussion	79
2.2.	Synthesis and Self-assembly Studies of Protein Variant FAPs	82
2.2.1.	Design	83
2.2.2.	Biophysical characterization of native proteins	84
2.2.3.	Synthesis	87
2.2.4.	Conjugation and purification	87
2.2.5.	Biophysical characterization	90
2.2.6.	Self-assembly studies	93
2.2.7.	Discussion	97
2.3.	Synthesis and Self-assembly Studies of Spacer Variant FAPs	100
2.3.1.	Design	101
2.3.2.	Synthesis	102
2.3.3.	Conjugation and purification	106
2.3.4.	Biophysical characterization	107
2.3.5.	Self-assembly studies	109
2.3.6.	Discussion	112
2.4:	Synthesis and Self-assembly Studies of Tail Variant FAPs	114
2.4.1.	Design	115
2.4.2.	Synthesis	116
2.4.3.	Conjugation and purification	120
2.4.4.	Biophysical characterization	122
2.4.5.	Self-assembly studies	125
2.4.6.	Discussion	129
2.5.	Experimental Section and Methods	131
3.	Chapter 3. FAPs for the Accurate Design of Stimuli-responsive Protein Complexes	199
3.1.	Synthesis, Self-assembly and Dis-assembly Studies of Photo-responsive FAP	200
3.1.1.	Design	201
3.1.2.	Synthesis	202
3.1.3.	Conjugation and purification	206
3.1.4.	Characterization	207
3.1.5.	Self-assembly studies	207
3.1.6.	Photo-disassembly studies	210

3.1.7.	Control studies	214
3.1.8.	Discussion	215
3.2.	Synthesis, Self-assembly and Dis-assembly Studies of Redox-responsive FAP	217
3.2.1.	Design	218
3.2.2.	Synthesis	219
3.2.3.	Conjugation and purification	222
3.2.4.	Characterization	223
3.2.5.	Self-assembly studies	223
3.2.6.	Redox-disassembly studies	226
3.2.7.	Control studies	228
3.2.8.	Discussion	231
3.3.	Synthesis and Self-assembly Studies of Peptide-based FAPs	232
3.3.1.	Design	233
3.3.2.	Synthesis	235
3.3.3.	Conjugation and purification	239
3.3.4.	Biophysical characterization	239
3.3.5.	Self-assembly studies	242
3.3.6.	Discussion	245
3.4.	Experimental section and methods	247
4.	Chapter 4: Extending MAPLab Technology to N-terminus and Thiol Labelling	276
4.1.	Targeting N-terminus for Universal Methodology	277
4.1.1.	Design	278
4.1.2.	Synthesis	281
4.1.3.	Conjugation and purification	284
4.1.4.	Characterization	288
4.1.5.	Self-assembly studies	291
4.1.6.	Discussion	293
4.2.	Targeting Thiol for Universal Methodology	294
4.2.1.	Design	295
4.2.2.	Synthesis	296
4.2.3.	Conjugation and purification	298
4.2.4.	Characterization	300
4.2.5.	Self-assembly studies	300
4.2.6.	Discussion	302
4.3.	Experimental section and methods	303

5 Appendix 316
Appendix-I: Synopsis and Outlook 316
Appendix-II: List of Figures 322
Appendix-III: List of Schemes 335
Appendix-IV: List of Tables 338
Appendix-V: List of Publications and Patents 340
Appendix-Vi: Acquired Copyrights from the Publishers 341
Appendix-Vii: Spectral Data 360

List of Abbreviations

Term	:	Indicates
^{13}C	:	^{13} Carbon NMR
^{19}F	:	^{19} Fluorine NMR
^{1}H	:	^{1} Proton NMR
1T	:	1 Tail
2P	:	Di Peptide
2T	:	2 Tails
3D	:	3 Dimensional
3P	:	Tri Peptide
3T	:	3 Tails
4P	:	Tetra Peptide
AcOH	:	Acetic acid
Å	:	Armstrong
a.u.	:	Arbitrary unit
AABP	:	Amphiphilic Activity-Based Probe
ABP	:	activity-Based Probe
ABPP	:	activity-Based Protein Profiling
AMW	:	Absolute Molecular Weight
A.C.	:	Affinity Chromatography
ACN	:	Acetonitrile
Acetone-D6	:	Deuterated Acetone
AL	:	Aldolasse
Arg	:	Arginine
Asp	:	Aspartic Acid
bs	:	Broad Singlet
$(BOC)_2O$	:	Di-tert-butyl Dicarbonate
BARC	:	Bhabha Atomic Research Centre
B.D.	:	Blue Dextran
BOC	:	Tert-butyloxycarbonyl
Br	:	Bromide
BSA	:	Bovine Serum Albumin
C12	:	12 Carbon Chain
C18	:	18 Carbon Chain
C6	:	6 Carbon Chain
CA	:	Carbonic Anhydrase
Calcd	:	Calculated

CBr_4	:	Tetrabromo Methane
CD	:	Circular Dichroism
CD_3OD	:	Deuterated Methanol or Methanol-d4
$CDCl_3$	:	Deuterated Chloroform or Chloroform-D
CEG	:	Cetylethylene Glycol
$CHCl_3$	:	Chloroform
Chy	:	Chymotrypsin
CMC	:	Critical Micelle Concentration
CoA	:	Conalbumin
Cryo-EM	:	Cryo-electron Microscopy
Ctrl	:	Control
$CUSO_4$	:	Cupper Sulfate
CV	:	Column Volume
Cys	:	Cysteine
^oC	:	Degree Celsius
d	:	Doublet
D_2O	:	Deuterated Water or Deuterium oxide
Da	:	Dalton
DAHC	:	Diammoniumhydrogen Citrate
DAST	:	Diethylaminosulfur Trifluoride
DBU	:	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCC	:	N,N′-Dicyclohexylcarbodiimide
DCM/CH_2Cl_2	:	Dichloromethane
dd	:	Double Doublet
DDEG	:	Dodecaethylene Glycol
DEG	:	Diethylene Glycol
Des	:	Desalting Chromatography
D_h	:	Hydrodynamic Diameter
DHAP	:	Dihydroxyacetophenone
DIPEA	:	N,N-Diisopropylethylamine
DLS	:	Dynamic Light Scattering
DMAP	:	4-Dimethylaminopyridine
DMF	:	Dimethylformamide
DMF-D7		Deuterated N,N-Dimethylformamide-d7
DMSO	:	Dimethyl Sulfoxide
DMSO-D6		Deuterated Dimethyl Sulfoxide-d6
dRI	:	differential Refractive Index
dt	:	Doublet of Triplet
DTPA	:	Diethylene triamine pentaacetic acid
DTT	:	Dithiothreitol

δ	:	Delta (in ppm)
δ_C	:	Delta Carbon Shift
δ_F	:	Delta Fluorine Shift
δ_H	:	Delta Proton Shift
EDC.HCl	:	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide Hydrochloride
eq.	:	Equivalents
E.S.	:	Esterase
ESI	:	Electron Spray Ionization
Et_2O	:	Diethyl Ether
Et_3N	:	Triethyl Amine
EtOAc	:	Ethyl Acetate
EtOH	:	Ethanol
E.V.		Elution Volume
FT-IR	:	Fourier Transform Infrared Spectroscopy
F	:	Ferritin
FAP	:	Facially Amphiphilic Protein(s) or Protein Conjugate
Fig	:	Figure
FP	:	Flourophosphonate
FPLC	:	Fast Purification Liquid Chromatography
F.S.	:	Fluorescence Spectroscopy
g	:	Gram
GFP	:	Green Fluorescent Protein
Glu	:	Glutamic Acid
GSH	:	Glutathione
GST	:	Glutathione S-Transferase
hrs	:	Hour(s)
H	:	Protons
H_2O	:	Water
H_2SO_4	:	Sulfuric Acid
HBTU	:	N,N,N′,N′-Tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate
HCl	:	Hydrochloric Acid
HIC	:	Hydrophobic Interaction Chromatography
His	:	Histidine
HOBT	:	Hydroxybenzotriazole
HPLC	:	High Performance Liquid Chromatography
HRMS	:	High-Resolution Mass Spectrometry

Hz	:	Hertz
IEX	:	Ion Exchange Chromatography
IISER	:	Indian Institutes of Science Education and Research
I.R.	:	Infrared
J	:	Coupling Constant
KCl	:	Potassium Chloride
K_2CO_3	:	Potassium Carbonate
kDa	:	Kilo Dalton
KI	:	Potassium Iodide
KOH	:	Potassium Hydroxide
LAH	:	Lithium Aluminum Chloride
LiBr	:	Lithium Bromide
Lys	:	Lysine
L.Z.	:	Lysozyme
λem	:	Emission Wavelength
λex	:	Excitation Wavelength
m	:	Multiplet
M	:	Molar
m/z	:	Mass Per Unit (+ve) Charge
MA	:	Maleimide
MALDI-TOF MS	:	Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry
MAPLab	:	Micelle-Assisted Protein Labeling
MeOH	:	Methanol
mer	:	Oligomeric State
mg	:	Milligram
MHz	:	Megahertz
min	:	Minute(s)
mL	:	Milliliter
mM	:	Millimolar
mmol	:	Millimole(s)
mp	:	Melting Point
MS	:	Mass Spectroscopy
mV	:	milli Volts
M.W.	:	Molecular Weight
MW+H	:	Molecular Weight + Proton

MW+K	:	Molecular Weight + Potassium
MW+Na	:	Molecular Weight + Sodium
µg	:	Microgram
µL	:	Microliter
µM	:	Micromolar
µm	:	Micrometre
N	:	Normal
N.N'-DSC	:	N,N′-Disuccinimidyl Carbonate
N2	:	Nitrogen
Na Asc.	:	Sodium Ascorbate
Na_2SO_4	:	Sodium Sulfate
$NaBH_4$	:	Sodium Borohydride
NaCl	:	Sodium Chloride
NaH	:	Sodium Hydride
$NaHCO_3$	:	Sodium Bicarbonate
NaI	:	Sodium Iodide
NaN3	:	Sodium Azide
NaOH	:	Sodium Hydroxide
NB	:	Nitrobenzyl
NBS	:	N-Bromosuccinimide
NCL	:	National Chemical Laboratory
NH	:	Ninhydrine
NH_4Cl	:	Ammonium Chloride
nm	:	Nanometre
nM	:	Nanomolar
NMR	:	Nuclear Magnetic Resonance
NO	:	Nitric Oxide
NPC	:	Normal Phase Chromatography
N-ter	:	N-terminus
OvA	:	Ovalbumin
OEG	:	Octaethylene Glycol
OxCl	:	Oxalyl Chloride
PBS	:	Phosphate Buffer Saline
%	:	Per cent
$P(OEt)_3$	:	Triethylphosphite
$P(Ph)_3$	:	Triphenylphosphine
PCA	:	Pyridinecarboxaldehyde
PE	:	Hexane

Ph	:	Phenyl
pI	:	Isoelectric Point
Pka	:	Acid Dissociation Constant
PMA	:	Phosphomolybdic Acid
ppm	:	Parts Per Million
PR	:	Photo-responsive
ProBr	:	Propargyl Bromide
ProK	:	Proteinase K
Py	:	Pyridine
q	:	Quartet
R.T.	:	Room Temperature
RBF	:	Round Bottom Flask
RBr	:	Alkyl Bromide
R*f*	:	Retention Factor
R_g	:	Radius of Gyration
R_h	:	Hydrodynamic Radius
RI	:	Refractive Index
RMW	:	Relative Molecular Weight
RPC	:	Reverse Phase Chromatography
RR	:	Redox-responsive
σ	:	Rho Factor
s	:	Singlet
SANS	:	Small-Angle Neutron Scattering
SAXS	:	Small-Angle X-ray Scattering
SEC	:	Size Exclusion Chromatography
SEC-MALS	:	Size Exclusion Chromatography-Multiangle Light Scattering
Ser	:	Serine
SiO_2	:	Silica
$SOCl_2$	:	Thionyl Chloride
Sod. Phos.	:	Sodium Phosphate
SR	:	Stimuli-responsive
SS	:	Disulfide
Sub	:	Subtilisin
t	:	Triplet
TBAB	:	Tetra-n-butylammonium Bromide
TBDMSCl	:	Tert-butyldimethylsilyl Chloride
TBHP	:	Tert-butyl Hydroperoxide

tBuOH	:	Tertiary Butanol
TEG	:	Tetraethylene Glycol
TFA	:	Trifluoroacetic Acid
T.G.	:	Thyroglobulin
THF	:	Tetrahydrofuran
TLC	:	Thin Layer Chromatography
TMS	:	Tetramethylsilane
TritCl	:	Trityl Chloride
Trp	:	Tryptophan
Try	:	Trypsin
TsCl	:	Tosyl Chloride
TsOH	:	p-Toluenesulfonic Acid
Tyr	:	Tyrosine
U.V.	:	Ultraviolet
Ve	:	Elution Volume of Particular Sample
Vc	:	Geometric CV
Vo	:	Column Void Volume
Z.P.	:	Zeta Potential

Abstract

Nature has evolved to make a diverse set of protein architectures to perform the complex fundamental life processes such as transcription, translation, catalysis, metabolism, and transport. These naturally occurring protein assemblies serve as an inspiration for the design of synthetic and semi-synthetic protein assemblies. This is primarily achieved through two complementary technologies, such as genetic and chemical methods. In the past decade, genetic methods have matured as robust technology for accurate design of protein assemblies with a defined geometry. Compared to the genetic method, chemical methods are in their infancy. Most of the studies in this area are largely focused on a couple of model proteins. More importantly, there are no robust methods for purification, and therefore synthesized protein assemblies lack detailed analytical and biophysical characterization. My thesis work is directed towards developing new chemical strategies to convert monomeric proteins into protein complexes. This method is based on the generation of a simple scaffold, which we call "Facially Amphiphilic Proteins (FAPs)" from native proteins. The generation of FAP was achieved using a micelle-assisted protein labeling (MAPLab) technology.

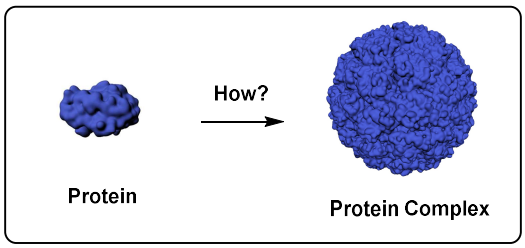


Figure1. This thesis is about a new robust chemical strategy to convert monomeric proteins into protein complexes.

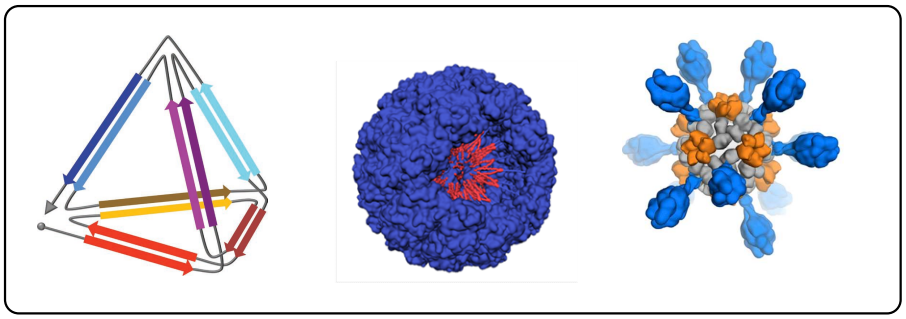
First, using serine proteases, we have demonstrated that monomeric proteins can be converted into protein complexes of bigger sizes. This was done by targeting active site Ser using fluorophosphonate chemistry. The simple design allowed us to systematically study the scaffold (FAP) with respect to individual units, i.e., protein, linker, and hydrophobic tail length/branching. This study also provided us with an opportunity to control the molecular weights, oligomeric states, and dimensions of the synthesized protein complexes precisely. Then, using a similar design and MAPLab Technology, we have also designed stimuli-responsive protein complexes. Further, we used the same method to make protein-synthetic peptide conjugates. Finally, in order to make this method more universal, we extended this technology for site-specific labeling of N-terminus. In addition, the same technology was explored for site-specific thiol bioconjugation

Keywords

Amphiphilic Activity-based Probes, Site-specific Labelling, Facially Amphiphilic Proteins, Protein Engineering, Protein Complex, MAPLab Technology.

Chapter 1

Introduction to protein engineering



Protein Engineering

Bio-molecular engineering deals with the engineering of biomolecules, like proteins, peptides, nucleic acids, and lipids for various applications (like therapeutics, biosensing, bioimaging, diagnostics) and creation of novel biomaterials. Protein engineering is at the forefront of bio-molecular engineering[1-8] and can be classified into two major areas.

i. Protein engineering for various applications: Protein engineering for various applications is mostly done by engineering natural protein cages.
ii. Protein engineering for the creation of novel biomaterials: Protein engineering for the creation of novel biomaterials or protein super-structures aims at converting protein oligomers or monomers into supramolecular structures/cages. This will be discussed in the latter part.

So, protein engineering in this thesis is discussed in the following fashion for the sake of simplicity.

1.1. Engineering natural protein cages: the application-oriented engineering
1.2. Engineering natural protein oligomers: the creation of novel biomaterials
1.3. Engineering monomeric proteins: the creation of novel biomaterials

1.1. Engineering Natural Protein Cages

The goal of protein engineering is to explore protein super-structures/cages/complexes (natural or created) for various applications. Unfortunately, very few created nano-cages/assemblies are at the application stage.[9] Most of the natural protein cages, on the other hand, are well explored for various applications using protein engineering techniques.[10] Some examples of natural protein cages, engineered for various applications are discussed below.

1.1.1. Viruses

Viruses are simple protein assemblies of various shapes and sizes from 10 nm to 1 μm, which stores, protects, and replicates their DNA or RNA by infecting living cells. Most viruses can self-assemble into stable, non-pathogenic virus-like particles (VLPs) even without their genetic material[11]. These VLPs are also amenable for modifications by genetic and chemical approaches on the interior, exterior surfaces, and even at the interfaces. Some viruses and their engineering for various applications are discussed below.

1.1.1.1. Cowpea chlorotic mottle virus (CCMV)

CCMV is perhaps the most explored virus for various applications. The CCMV is 180 subunits icosahedral virus with outer and inner diameters of 28 nm and 18 nm, respectively. The truncated form of CCMV (N-terminus deleted form) also has the ability to self-assemble into 60 (T = 1 symmetry) or 120 (T = 2 symmetry) subunit cage of outer diameters 18 nm and 22 nm respectively, where T is called triangulation number. According to the Caspar-Klug Theory, T-number represents the number of symmetrically distinct facets (unique environments) in the face of the icosahedral virus. In simple terms, T-number gives the total number of coat proteins per capsid if we multiply it by 60. For example, a T = 1 virus has 60 coat proteins, while a T = 3 virus has 180 coat proteins.[12] The N-terminus, located on the inner side of the CCMV capsids, is positively charged and helps in attracting the negatively charged RNA in the native state. They can be disassembled to dimers at pH 7.5 (in the presence of $CaCl_2$), and after removal of the viral RNA, encapsulation of negatively charged small molecules can be carried out by reassembling CCMV particles at pH 5.[13-15]

The large CCMV inner cavity of 18 nm permits the encapsulation of biomolecules like proteins and enzymes, which was successfully demonstrated in 2009, by the Cornelissen group in an efficient and controlled fashion.[16] Further, Cornelissen and co-workers also

confined two separate enzyme systems (either alone or together), i.e., glucose oxidase (GOx) and gluconokinase (GCK) in a non-covalent fashion inside an N-terminal Arg enriched CCMV particles. The confinement was achieved by covalently labeling DNA molecules on the surface of these proteins (with a heterobifunctional linker using sulfo-NHS coupling followed by maleimide-thiol chemistry). Interestingly, this encapsulation resulted in T=1 particles. Further, the DNA fragment chosen for the protein manipulation was capable of forming quadruplex, and so it has a role too. Both the encapsulated systems revealed that both k_m values and k_{cat} values were enhanced when compared with the free enzymes.[17] Similarly, the encapsulation of a large, water-soluble, aromatic macrocyclic molecule phthalocyanine (PC) into CCMV virus-like particles in a cooperative fashion was demonstrated by the same group.[18]

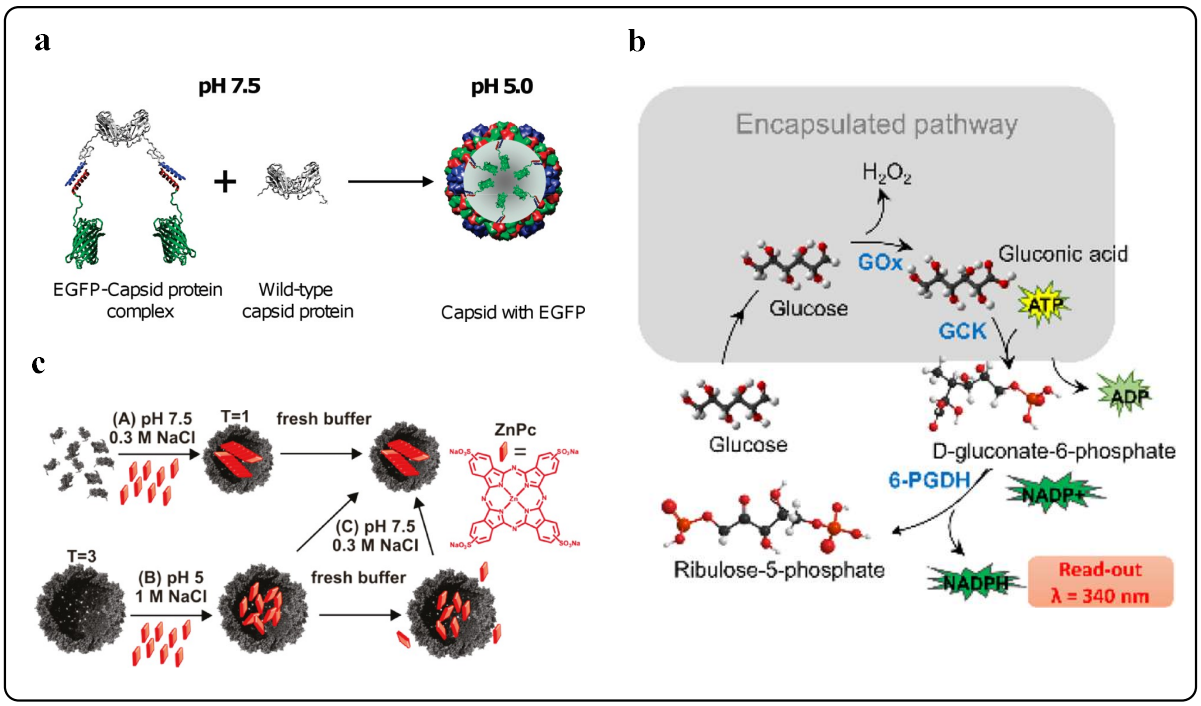


Figure 1.1. a. The EGFP-capsid is mixed with wild-type capsid at pH 7.5 and then dialyzed against pH 5.0 to induce capsid formation resulting in EGFP encapsulation. b. Cascade reaction performed by co-encapsulated GOx and GCK at the interior of the CCMV capsids (the conversion of glucose to D-gluconate-6-P). c. The encapsulation of PC stacks into CCMV VLPs. Adapted with permission from ref. 16, 17 and 18.

Constructing a CCMV-polymer biohybrid through non-covalent interaction is a fascinating idea. In 2011, CCMV-polymer biohybrid was constructed by Kostiainen et al. using

electrostatic association as a function of temperature. Using negatively charged CCMV particles and dual-functional diblock copolymers consisting of poly(diethylene glycol methyl ether methacrylate) (poly(DEGMA)) and poly((2-dimethylamino)ethyl methacrylate) (poly(DMAEMA)) blocks, micrometer-sized CCMV-polymer biohybrid was constructed. This temperature-switchable assembly of viral particles can be disassembled by reducing the solution temperature.[19]

A significant protein engineering research has been dedicated to protein-elastin-like polypeptides (ELPs) hybrids. Because of the reversible thermal responsiveness, i.e., hydrophilic-hydrophobic phase transition at a lower critical solution temperature (LCST) and polypeptide nature of elastin-like polypeptides (ELPs), they can be easily incorporated using genetic approaches to engineer or assemble monomeric proteins. The resulting protein-ELP hybrids can self-assemble in response to temperature. These hybrids are also biocompatible, which makes it even more ideal for potential in vivo applications. For example, Van Hest and co-workers in 2012, investigated the self-assembly properties CCMV, expressed as a fusion protein with an ELP. The fusion protein carried His tag, which assisted them during the purification. This protein-based block copolymer self-assembled by two mechanisms (1) pH-induced assembly (28 nm VLPs) and (2) ELP-induced assembly (18nm VLPs). The equilibrium between the dimmer state and fully assembled T1 state was studied with respect to the concentration of fusion protein, NaCl concentration, and also the temperature. Even though the pH-induced assembly was the inherent property of capsid protein, the latter was completely a result of the fusion protein.[20]

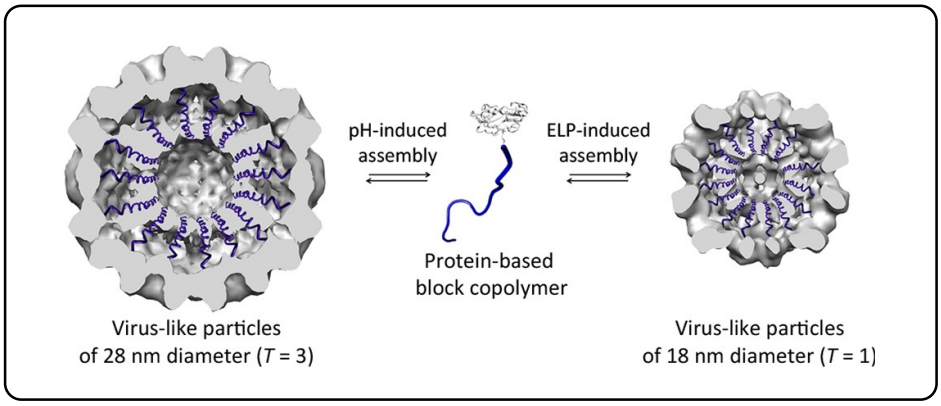


Figure 1.2. ELP-CP and it's Self- assembly: pH-induced assembly and ELP-induced assembly. Adapted with permission from ref. 20.

CCMV particles apart from encapsulation, were also used for displaying various ligands. Douglas and co-workers in 2006, created multifunctional CCMV particles with controlled ligands (digoxigenin, biotin) display by mixing of two populations of differently labeled disassembled particles.[21] The same team also used CCMV particles as templates for targeted antimicrobial photodynamic therapy (PDT) by conjugating to a photosensitizer (Ru(bpy2)phen-IA) and targeting ligand.[22]

1.1.1.2. cowpea mosaic virus (CPMV)

CPMV is a 60 subunit icosahedral plant virus with a diameter of 28 nm. It is made of a small (24 kDa) and a large (41 kDa) asymmetric units.[23,24] CPMV virus particles are also famous for displaying various ligands, just like CCMV. In fact, CPMV was used for displaying various ligands even before CPMV. Finn and co-workers, for the first time, in 2003, used CPMV virus particles, as a display platform for mannose in a well-defined fashion, and they found that the product functions as polyvalent particles in interactions with complementary lectin depending on the density and position of functionalization.[25] Further, the same group in 2007, decorated on CPMV virus particles with polyvalent glycan ligands for cell-surface receptor recognition by using sequential enzymatic transformations.[26] Similarly, in 2004, Chatterji and co-workers displayed entire proteins by chemically cross-linking to Lys and Cys residues genetically engineered on the CPMV particles.[27]

The CPMV was used extensively as a delivery vehicle for various molecules like drugs, fluorophores, peptides, and proteins either by displaying or by encapsulation or both. For instance, CPMV was used as a delivery vehicle for the doxorubicin (DOX) by Nicole and her team. DOX was conjugated to surface carboxylates of the viral nanoparticles, and these virus-drug conjugates showed greater cytotoxicity than free DOX toward HeLa cells.[28] A non-covalent infusion technique for efficient cargo loading in RNA-encapsidated CPMV nanoparticles was reported by the same group. Infusion and retention of 130-155 fluorescent dye molecules per CPMV were achieved, and surface Lys residues were further covalent modified with dyes for making dual-functional CPMV carriers.[29] Similarly in 2009, fullerenes (C60) were covalently attached to CPMV for potential applications in photoactivated tumor therapy and the hybrid CPMV-C60/ CPMV- poly(ethylene glycol) (PEG)-C60 complexes were found to be water-soluble and biocompatible, in spite of water incompatibility of fullerenes.[30] In 2012, the same group investigated the conjugation of

fluorophores, biotin affinity tags, PEG, and various peptides through targeting reactive Cys displayed selectively on the interior surface of empty CPMV (eCPMV) particles devoid of RNA to encapsulate a wide variety of molecules.[31] In 2010, Using hydrazone ligation chemistry, CPMV nanoparticles were decorated with vascular endothelial growth factor receptors (VEGFR)-1 ligand and fluorescent PEGylated peptides. The nanoparticles demonstrated cellular imaging and tumor targeting capabilities.[32]

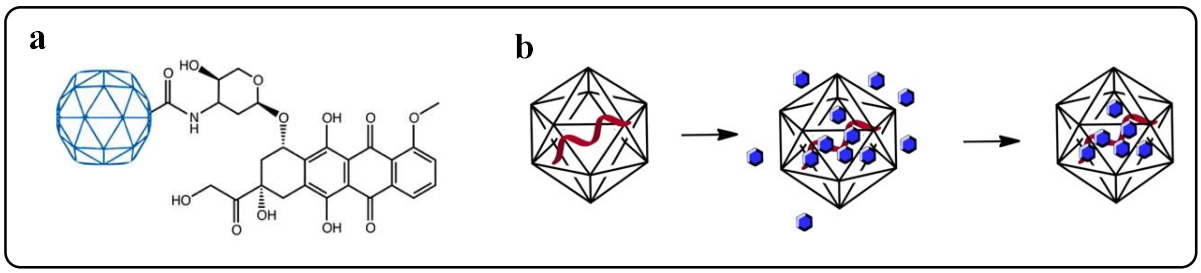


Figure 1.3. a. conjugation of the DOX to CPMV. b. CPMV, with its SS RNA molecule (in red), is exposed to a solution containing the cargo (in blue), washing, and dialysis to get intact CPMV with infused cargo. Adapted with permission from ref. 28 and 29.

1.1.1.3. Qβ bacteriophage

The Qβ is one of the most explored bacteriophages for various applications. It is an extremely stable RNA-bacteriophage of icosahedral geometry made of 180 coat protein subunits, with a diameter of 28 nm supported by inter-subunit disulfide bridges.

Qβ particles were explored for the display of cell-targeting motifs and immunogens/antigens. For example, in 2011, Qβ particles were used for the multivalent display of cell-targeting motifs, i.e., the human epidermal growth factor (EGF) on the exterior surface as a genetic fusion at C-terminus by Finn and co-workers. They found that the co-assembly of wild-type Qβ and EGF-modified subunits resulted in the formation of nanoparticles, which displayed between 5 and 12 copies of EGF. Additionally, the particles were biologically active, promoted phosphorylation of the EGF receptor and apoptosis of A431 cells.[33] The same team further explored the Qβ particles for the display of glycan (targeting ligand), along with a metalloporphyrin derivative for cells bearing the CD22 receptor for photodynamic therapy in 2012.[34] Similarly, a fragment of HIV-1 gp120 called b122a was displayed on the surface of Qβ virus-like particles, and the immune response to this structure was studied.[35]

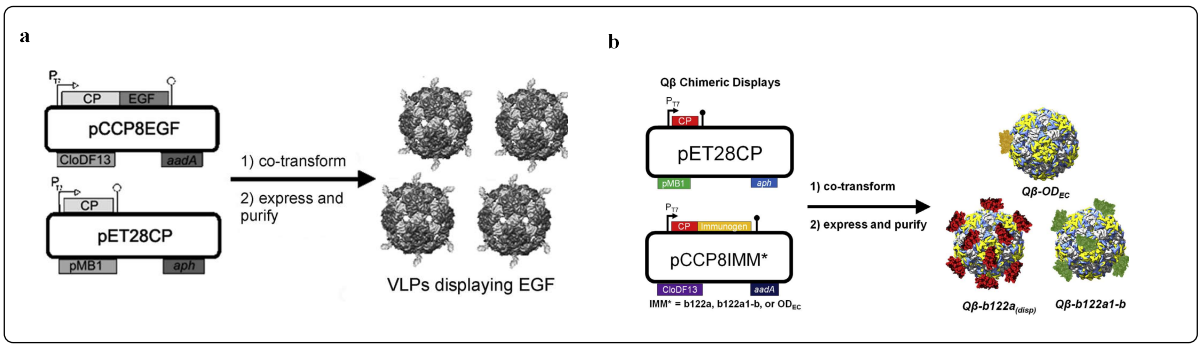


Figure 1.4. a. Qβ plasmids used for particle production and the EGF-display. b. Qβ, plasmids used for chimeric VLP display of b122a. Adapted with permission from ref. 33 and 35.

Similarly, Qβ particles were used for enhancing the stability and delivery of immune response adjuvants. For example, in 2004, by loading DNA rich in nonmethylated CG motifs (CpGs) into p33 peptides coupled to bacteriophage Qβ via a bifunctional linker to lysines, problems like stability and adverse effects of CpGs was solved with improved immunogenicity.[36]

Bacteriophage Qβ was also used for the construction of virus-polymer conjugates (polymer-coated protein nanoparticles). For example, in 2010, Finn et al. made virus-polymer conjugates using Qβ VLPs. Atom transfer radical polymerization (ATRP) initiators were installed on Qβ particles using N-hydroxysuccinimide to attach an azide group on surface-accessible amine groups of the Qβ particle. The resulting azido particle was then functionalized with a triglyme-based ATRP initiator by copper-catalyzed azide-alkyne cycloaddition to form a Qβ macroinitiator. Polymers were then grown from 2,2 0-bipyridine/CuBr/$CuBr_2$ catalyst system in water to get virus protein-polymer conjugates.[37] Further, the immune recognition properties and the effective shielding from antibody recognition of polymer conjugated Qβ particles were investigated by Pokorski and his group in 2017. They found that relative to methoxypoly(ethylene glycol) (mPEG) and ,poly(oligo(ethyleneglycol)methyletheracrylate) (POEGMEA), poly(norbornene-(oligo(ethylene glycol) ester)) (PNB) affords greater antibody shielding to the conjugate while occupying a significantly smaller hydrodynamic volume.[38]

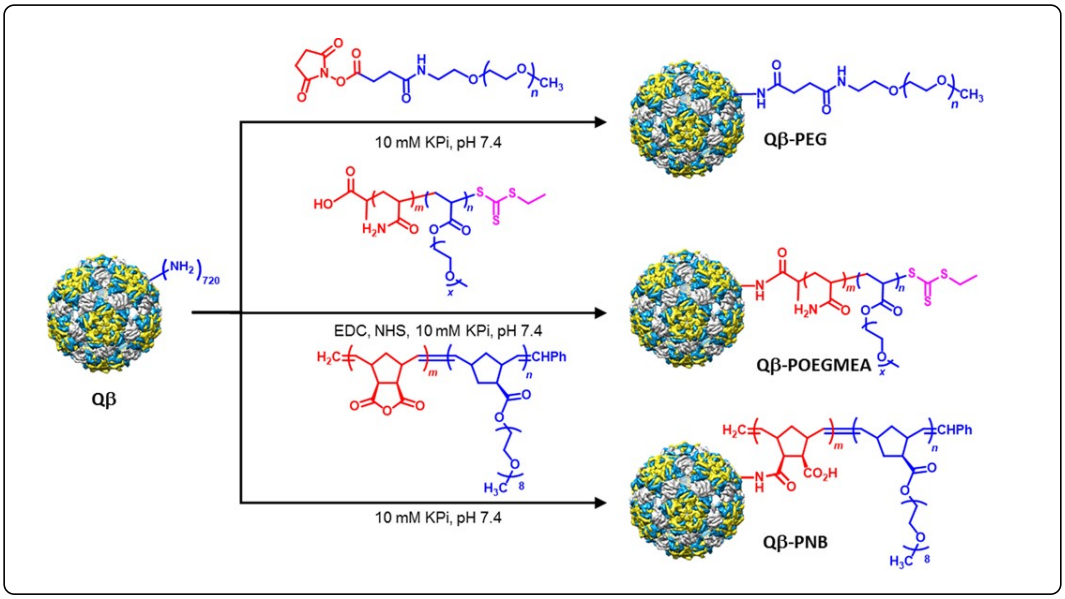


Figure 1.5. Synthesis of Qβ-Polymer conjugates to evaluate the immune recognition of polymer conjugated particles. Adapted with permission from ref. 38.

1.1.1.4. Tobacco mosaic virus (TMV)

TMV is a single-stranded RNA virus that infects a wide range of plants. Studies were done to influence and control the self-assembly of TMV using aromatic moities and DNA origami. For example, Majima et al. demonstrated the effect of the introduced aromatic stacking (pyrenes) on the formation of the tobacco mosaic virus (TMV) into a long helical rod structure. Pyrenes were selectively incorporated into the four specific amino acids (N98, Q99, A100, and N101) in the inner side of the TMV rod structure by maleimide–thiol coupling using N-(1-pyrene) maleimide reacted with the mutated cysteine residues at these positions. Pyrene-attached TMV rods were 2-8 times longer than that of the wild-type TMV rods.[39] In another report, assembly of TMV coat proteins onto DNA origami in a precisely controlled and facile manner was achieved. The protein nanotubes (TMV coat proteins) of controlled length are precisely anchored on the DNA origami at selected locations using TMV genome-mimicking RNA strands. By varying length of RNA strands, protein nanotube growth lengths were controlled, thereby achieving the stoichiometric control.[40]

TMV was also used as delivery vehicle for drugs. For instance, the Francis group reported TMV-drug conjugates. A self-assembling nanoscale disk was made from a double Arg mutant of recombinantly expressed TMV coat protein (RR-TMV). RR-TMV disks were

functionalized with DOX and further modified with PEG to improve the solubility. RR-TMV-DOX-PEG particles showed cytotoxic properties similar to DOX.[41]

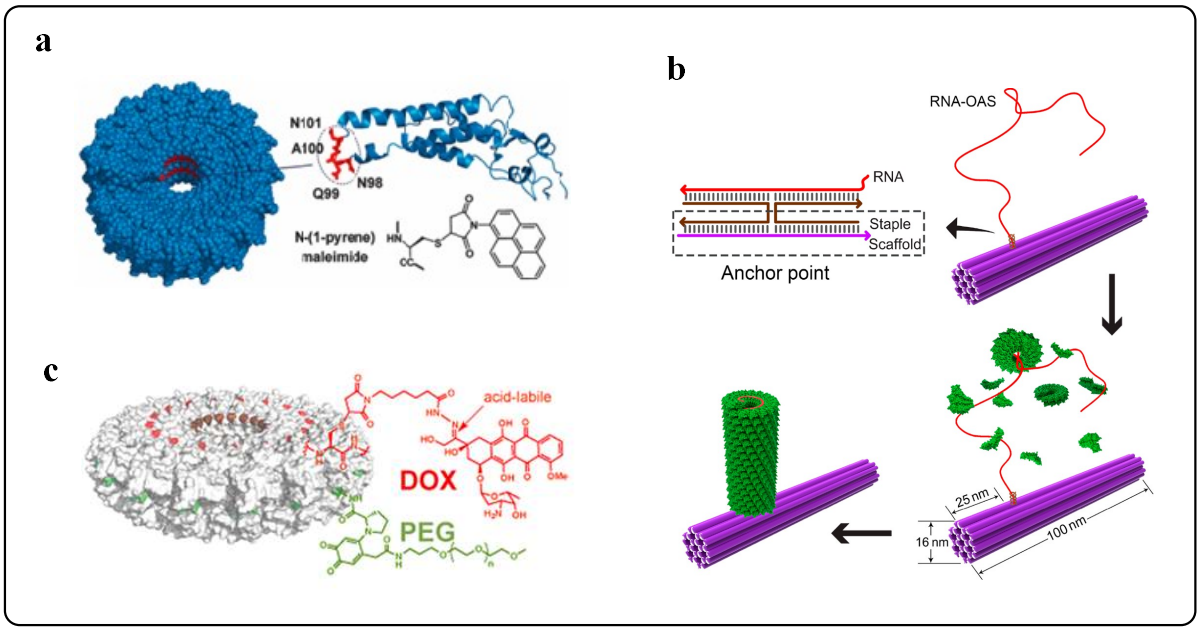


Figure 1.6. a. The representative figure to show introduced pyrenes on the formation and stability of the TMV. b. *In situ* assembly of the TMV nanotube onto a rigid DNA origami structure. c. The RR-TMV disks functionalized with DOX) and PEG for improved solubility. Adapted with permission from ref. 39, 40 and 41.

TMV was further explored for the construction of artificial nanoenzyme in 2012. TMV coat protein monomers were fabricated with the main catalytic components of glutathione peroxidase (GPx). The resulting self-assembled TMV nanoenzyme exhibited remarkable GPx activity.[42]

1.1.1.5. Bacteriophage MS2

The MS2 is 180 subunits icosahedral, RNA virus, with a diameter of 27 nm. It also comes with the 32 pores of size 2 nm on the surface.[43-45] The MS2 is well studied for the delivery of RNA, ssDNA, and porphyrin like molecules. For example, MS2 bacteriophage capsid proteins were used to deliver antisense oligodeoxynucleotides (ODNs), including some targeting p120 messenger RNAs, a biomarker overexpressed in myelogenous leukemia cells.[46] Similarly, the delivery of cargo or the encapsulated RNA was further enhanced by decorating a targeting motif. For example, In 2012, MS2 VLPs were used for the targeted delivery of a siRNA with the covalent decoration of targeting motifs on the surface.[47] In the

same year, MS2 was conjugated to HIV-1 Tat 47-57 peptide on the surface and using these VLPs, microRNA (miRNA) delivery was reported. The conjugated MS2 VLPs effectively transferred the packed pre-miR146a RNA into various cells and tissues, with higher expression of miR-146a.[48]

MS2 was also explored as a delivery vehicle for photodynamic therapy. Francis and co-workers used MS2 bacteriophage to construct a targeted, multivalent photodynamic therapy vehicle by modifying interior with 180 porphyrins and exterior with 20 copies of a Jurkat cell-specific aptamers.[49]

1.1.1.6. Hepatitis B virus (HBV)

HBV is one of the smallest enveloped icosahedral viruses with a diameter of 42 nm. HBV is a great virus for displaying various proteins and antibodies. Using HBV capsids, the fluorescent proteins (FPs) were encapsulated or surface-displayed depending on the site of genetic encoding.[50] Further, by displaying and orienting troponin antibodies Park et al. increase the sensitivity of troponin levels detection in human serum samples by fusing HBV capsids with 6x-histidine tag (His6-Tag) and by replacing the P79A80 in the HBV loop with B domains of staphylococcal protein A (SPA_B) sequences.[51]

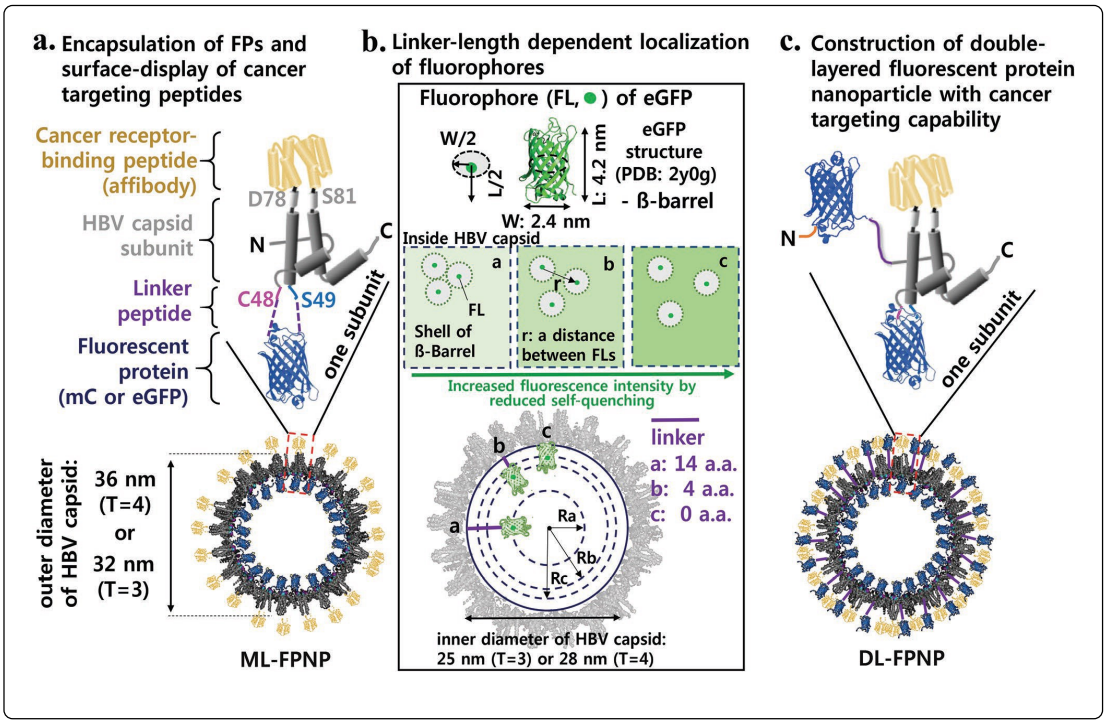


Figure 1.7. Fluorescently engineered HBV capsids. a) Encapsulation of FPs surface-display of cancer-targeting peptides. b) Linker-length dependent localization of fluorophores. c) Construction of double-layered fluorescent protein nanoparticle with cancer-targeting capability. Adapted with permission from ref. 50.

1.1.2. Other cages/compartments

1.1.2.1. Bacterial compartments

Lumazine synthase (AaLS) catalyzes the penultimate step of riboflavin biosynthesis in Aquifex aeolicus. It forms icosahedral capsids of 60 and 180 subunits with T = 1 and T = 3 state, respectively. Donald Hilvert and co-workers pioneered the engineering of AaLS. For instance, they have successfully encapsulated proteins in these capsids. AaLS-neg, a variant with 240 to 720 negative charges on the luminal surface, was created by the introduction of four glutamates R83E, T86E, T120E, and Q123E in 2011. This negatively charged capsid encapsulated proteins with a complementary positively charged peptide tag. For example, the co-expression of AaLS-neg with a green fluorescent protein (GFP) with a C-terminal deca-Arg sequence (GFP-R10) affords a cage complex containing 3 to 4 GFP molecules per T = 3 capsid. Further, using directed evolution, AaLS-13 was created, which upon co-expression can accommodate up to 7 HIV protease-R10 dimers on average, which represents a 5-10 fold improvement.[52] Further, AaLS-13, a variant of AaLS-neg with higher loading capacity, was created by directed evolution. Using GFP (+36) as a genetically encodable tag, active guest enzymes were efficiently packed in AaLS-13 protein cages. Simply mixing host and cargo under mild aqueous conditions leads to internalization. No pH or temperature changes are required. Loading is nearly quantitative up to around 45 guest enzymes per T=3 capsid, thereby affording precise control.[53]

The AaLS container structures were elucidated later in 2017 by native mass spectrometry (MS) and cryo-electron microscopy (cryo-EM) single-particle reconstruction techniques. They also found that with an increase in the number of negative charges on the lumenal surface of lumazine synthase, the native protein shell, expands to thermostable 3-MDa and 6-MDa assemblies.[54] Further, In 2017, AaLS capsids were also used for a simple and efficient system for the in vivo encapsulation of RNA molecules. Inspired by nucleotide-binding domains (NBDs) found in such small viruses, a series of simple oligoarginines (R3, R6, R9) were employed to drive the encapsulation of nucleic acids through electrostatic interactions.

These oligoarginines tags were fused to the C-terminus of cpAaLS using a seven amino acid linker, yielding cpAaLS-RX, where X specifies the number of Arg residues. By varying the number or Structure of RNA-binding peptides presented on the capsid lumen, the size-selectivity of cargo molecules can be controlled. The utility of the technology was demonstrated by trapping short-living RNAs.[55]

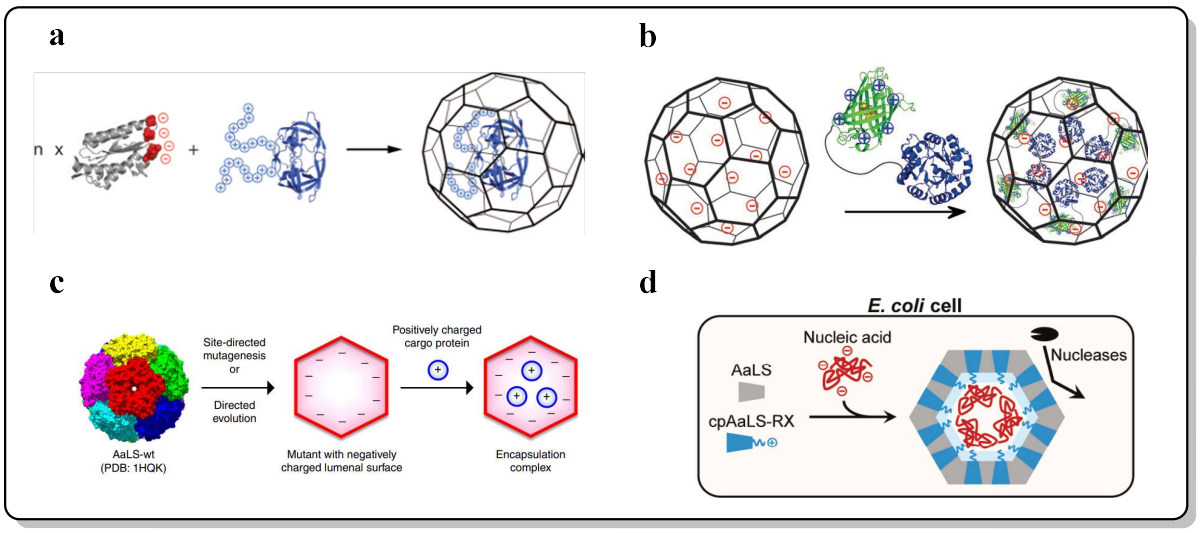


Figure 1.8. a. Encapsulation of toxic protein (HIV protease) in an engineered variant of lumazine synthase. b. Encapsulating active enzymes in a protein cage. GFP (+36) (green) serves as a directing tag to load appended cargo proteins (blue) into empty AaLS-13 capsids. c. Electrostatic interactions-based encapsulation of cargo proteins by engineered AaLS cages. d. Design and structure of a nucleic-acid packaging protein capsule cpAaLS-RX possessing an oligoarginine tag (RX). Adapted with permission from ref. 52, 53 and 55.

1.1.2.2. Ferritins

Ferritins, present in almost all living organisms, regulate iron homeostasis. They self-assemble into spherical cages of 450 kDa, made of 24 subunits with inner and outer diameters of 8 and 12 nm, respectively. These cages are stable at high pH of 9 and temperatures of 85 ^{o}C but, they broke down to subunits at pH of 2 to 3 (reversibly). These features and amenability to both genetic and chemical modifications (both the interiors and exteriors surfaces) made ferritins, an interesting scaffold.[56-58]

Ferritins were successfully used for the encapsulation of MRI contrast agents. For example, of the in 2002, gadoteridol was encapsulated in apoferritin by dissociating it at pH 2, and by

reforming it at pH 7.[59] In 2007, apoferritin was used to encapsulate anticancer drugs like cisplatin and carboplatin by two different methods, i.e., unfolding-refolding method and in situ generation method.[60] Similarly ferritins were used as a delivery vehicle for several anticancer agents like 5-fluorouracil[61] daunomycin[62,63] and doxorubicin.[64-67]

Apart from encapsulation of small molecules, peptides and antibodies were displayed successfully on ferritins. For example, antibodies were displayed in a non-covalent, controlled orientation by genetically introducing the Fc-binding peptides (FcBPs) into the loop of ferritin.[68] Similarly, in 2013, ferritin was used to display the interleukin-4receptor (IL-4R)- targeting peptide, AP-1, by fusing with ferritin-L-chain (FTL).[69] Extending peptidic display further, in 2015, a ferritin-based proteinticle delivery system for siRNA was reported. The surface of human ferritin was modified to display a cationic peptide as a fusion to the C-terminus, for interacting with siRNA, cell-targeting, cell-penetrating peptides, and peptides for enzymatic cleavage to release siRNA molecules inside the cells.[70]

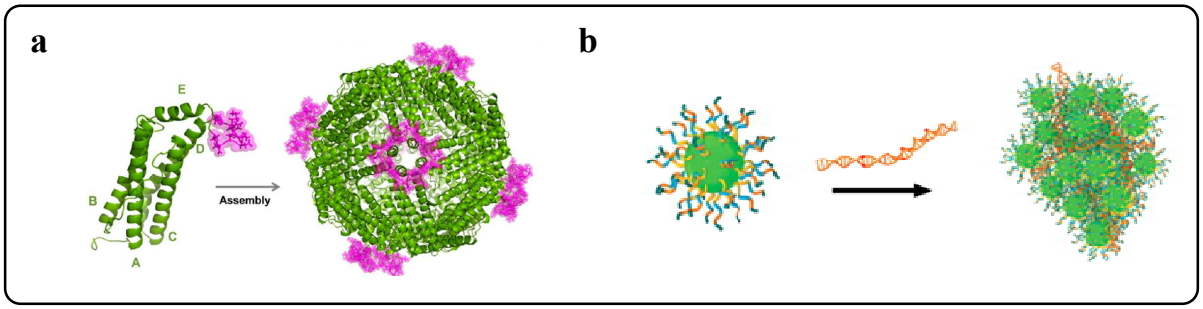


Figure 1.9. a. Representative figure of AP-1 peptide inserted ferritin subunit and assembled AP1-PBNCs. b. the formation of the condensed structure of the poly-siRNA-proteinticle complex. Adapted with permission from ref. 69 and 70.

Very recently, in 2017, ferritin-based 3D metal-organic frameworks (MOFs) was reported by the Tezcan group by building metal anchoring sites at the C3 symmetric pores, which was done by replacing Thr122 lining the outer rim of these pores with His. A library of 15 ferritin-MOFs with the expected body-centered (cubic or tetragonal) lattice arrangements was made by changing metal ions presented at the vertices of the ferritin nodes (Zn(II), Ni(II), and Co(II)) and dihydroxamate linkers.[71]

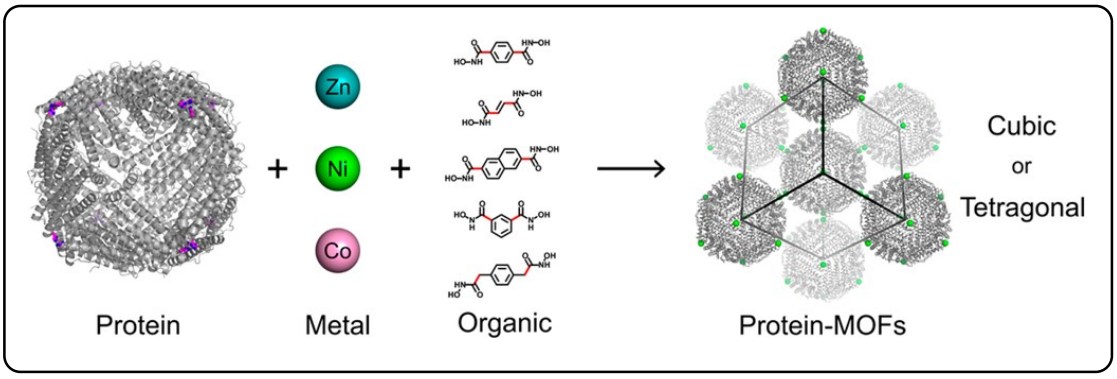


Figure 1.10. Formation of ferritin-MOFs. Adapted with permission from ref. 71.

1.1.2.3. Small heat shock proteins (sHsp)

The small heat shock proteins (sHsp) from Methanococcus jannaschii (MjHsp) is a stable, spherical age of 24 subunits. It has pores of sizes 3 nm and is stable up to 70 ºC and in pHs ranging 5-11. sHsp are usually expressed in response to stress like inflammation, hypoxia, and infection. They help in proper synthesis and folding of proteins that are expressed in response to stress. sHsps lack cysteine residues, which provides an opportunity to introduce cysteine residues for site-specific conjugation if needed.[72-75] sHsps were used to created targeted drug delivery systems. For example, a human hepatocellular carcinoma (HCC)-targetable protein nanocarrier, HspG41C-SP94, was developed by expressing the HCC-binding peptide SP94 on the surface of a naturally occurring Hsp cage. DOX-conjugated HspG41C-SP94 cages showed reduced cytotoxicity towards normal hepatocyte cell lines but maintained its cytotoxic effects against Huh-7 HCC cells.[76]

1.1.2.4. Vault proteins

Vaults are barrel-shaped, ribonucleoprotein nano-cages, and are ubiquitous in eukaryotes. They are large hollow structures, with overall dimensions of 72.5 nm (length) and 41 nm (width), and can disassemble into two halves at pH 3 to 4 irreversibly. The stability and non-immunogenicity of these cages make them attractive candidates for the delivery of drugs, immunogens, and other molecules.[77-79] Similar to protein-polymer and virus-polymer conjugates, vaults-polymer conjugates were reported by Maynard and her team. The synthetic modification of recombinant vaults with Poly(N-isopropylacrylamide) (pNIPAAm), a temperature-responsive polymer, created a stimuli-responsive vault-polymer nanocapsule,

which showed reversible aggregation above the lower critical solution temperature (LCST) of the polymer.[80]

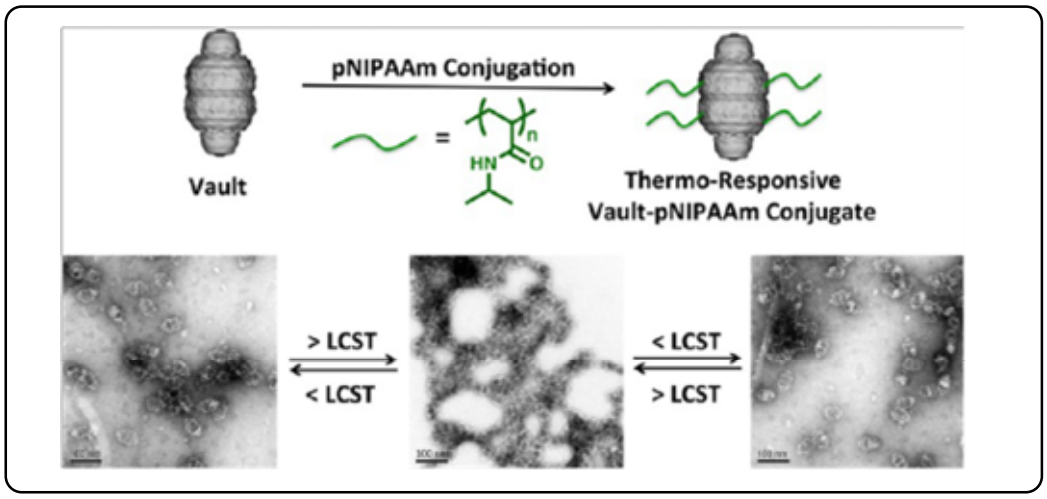


Figure 1.11. Preparation of vault-pNIPAAm conjugates and thermo-responsive behavior. Adapted with permission from ref. 80.

Thus the engineering efforts based on natural protein assemblies have a huge impact in the area of vaccine design[35,81], targeted drug delivery[82,83], delivery vehicles for siRNA[70,84], targeted delivery of imaging agents[29,32,85], designing nanoreactors[17], in vivo diagnostics[32], etc.

Protein engineering for creation of novel biomaterials

The accurate design of the supramolecular protein super-structure through the bottom-up approach is at the forefront of bio-molecular engineering efforts.[6-8] The inspiration for the creation of protein-based supramolecular systems comes from the fact that proteins self-assemble into higher-order super-structures to perform the complex fundamental life processes ranging from transcription, translation to catalysis, metabolism, transport, and structural integrity. Nature has built complex protein architectures like protein cages, tubular assemblies, Ring-shaped complexes, linear assemblies, catenanes, and knot proteins to perform these functions. These proteins are brilliantly summarized in the review by Nolte and co-workers.[86] Unfortunately, the creation of precise protein nano-assemblies through the bottom-up approach is highly challenging because proteins are highly complex structures. They have multiple functionalities with covalent and non-covalent interactions/forces. They also tend to denature and aggregate under unfavorable conditions. The creation of artificial

protein super-structures/assemblies/cages can be classified into two broad categories based on the oligomeric state of protein starting material.

1.2. Engineering natural protein oligomers

1.3. Engineering monomeric proteins

The methods for creating artificial protein assemblies/cages are still limited and mostly done using know protein oligomers through genetic methods. Let us look at technologies developed for engineering natural protein oligomers to construct artificial protein assemblies/cages.

1.2. Engineering natural protein oligomers

These protein engineering methods based on natural protein oligomers are one of the major approaches in designing higher-order structures and are proven to create protein complexes with high accuracy and fidelity. The structures of protein cages/complexes created through these methods have been determined to a near-atomic resolution. These methods are in the early stages in terms of applications. Some of these protein engineering approaches use optimization of protein-protein interactions (multivalent contacts) to drive the protein assemblies like cages, polyhedral, and artificial enzymes. Other approaches are based on the rational fusion of symmetrical oligomers. Despite several advantages, these methods are prone to high failure rate and often yield structures not intended by the design[87-89]. These approaches often require extensive evaluation of a library of designs to achieve the desired products. These methods can be further classified into:

1.2.1. Computational methods

The computational approaches depend upon finding the right protein oligomers and then optimizing protein-protein interactions. The David Baker group extensively uses computational approaches for the designing of protein assemblies. The *de novo* design of protein cages from multi-component, two distinct subunits was reported by Baker and co-workers in the year 2014. Using this method, they have designed five 24-subunit cages with two different symmetric architectures.[90] Similar in a novel icosahedral protein nano-cage was designed from a single subunit (trimeric KDPG aldolase), instead of two distinct subunits. The trimeric KDPG aldolase was aligned with the three-fold axes of the octahedron for the *de novo* design of the icosahedral cage. Then the interfaces are designed along the icosahedral two-folds to make icosahedrons. 20-trimers make an icosahedron assembly here.[91]

Further, computational design of co-assembling, two-component, 120-subunit icosahedral protein cages of about 1.8 to 2.8 megadaltons, and a diameter of 24 to 40 nm was reported. Three designs for icosahedral protein cages were reported here. The first design is I53 architecture, which has 12 pentamers and 20 trimers that are aligned along the fivefold and threefold symmetry axes, respectively. The second design is I52 architecture, which has 12 pentamers and 30 dimers aligned respectively along the fivefold and twofold icosahedral symmetry axes. The last design is I32 architecture which is made of 20 trimers and 30 dimers aligned along the threefold and twofold icosahedral symmetry axes.[92]

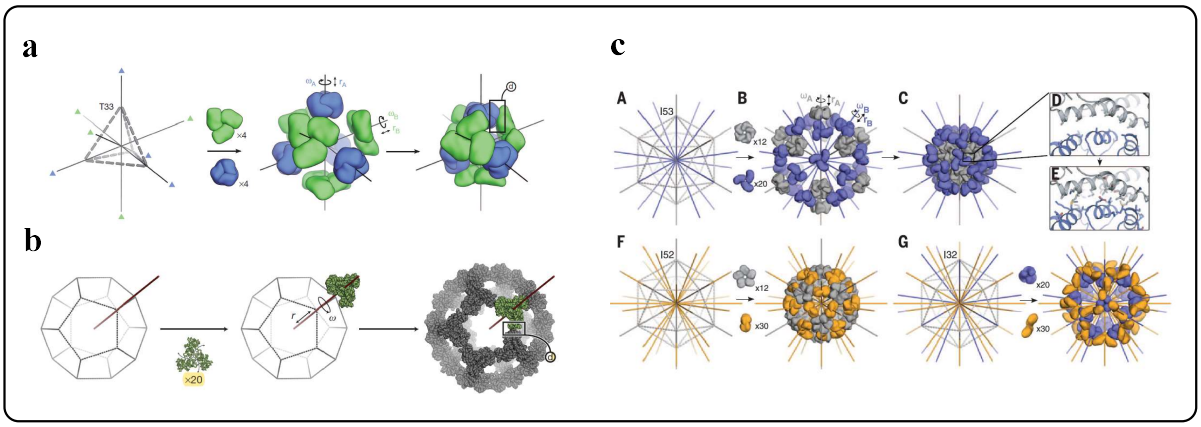


Figure 1.12. a. Overview of the computational methodology for the construction of the tetrahedral cage from two distinct trimeric building blocks. b. computational methodology for the construction of the Icosahedral cage from the trimeric building block. c. Overview of the construction of icosahedral cages in three different ways using pentamers+trimer, pentamers+dimers, and trimer+dimmers. Adapted with permission from ref. 90, 91 and 92.

Even though the methods that use natural protein oligomers to construct protein assemblies are in the early stages in terms of applications, the computational methods, in particular, showed brilliant applications in the area of vaccine design. For instance, Neil P. King and co-workers reported the structure-based design of a self-assembling two-component protein nanoparticle vaccine for the respiratory syncytial virus. This protein particle presents a prefusion-stabilized variant of the F glycoprotein trimer (DS-Cav1) in an ordered, repetitive array on the particle exterior at controllable density.[93] Later, this technique has been extended for shaping the immunogenicity of HIV-1 envelope trimers as well.[94] Further, a computationally designed protein cage, was for also redesigned for siRNA delivery applications by introducing Arg mutations to the lumenal surface.[95,96]

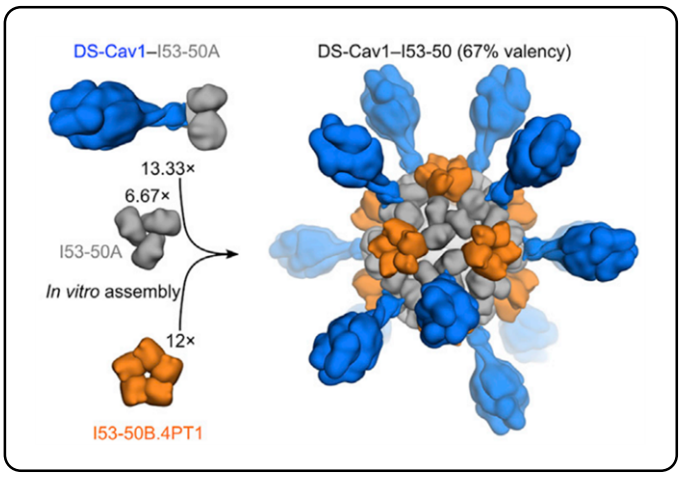


Figure 1.13. Overview of the two-component protein nanoparticle vaccine for the respiratory syncytial virus. Adapted with permission from ref. 93 and 94.

1.2.2. Symmetric oligomers-based fusion methods

The symmetry-based fusion methods are based on the fusion of proteins with known symmetries to generate supramolecular protein assemblies. Significant efforts in this direction were put forth by Todd Yeates and Ryoichi Arai groups.

Todd Yeates and co-workers described a general strategy for designing protein assemblies in 2001 by fusing two proteins (A, B), which have the natural ability to form oligomers of a different type. The fusion protein (A-B), self assembles into a designed nanohedral particle (A-B)p. This strategy is demonstrated by designing two different assemblies by fusing two proteins (The trimeric bromoperoxidase and the dimeric M1 matrix protein of influenza virus): a 49-kDa fusion protein assembled into a cage of 15 nm across, and a 44-kDa fusion protein assembled into long filaments of width 4 nm.[87] Extending the strategy, the same team designed another protein assembly in 2012, by fusing two oligomerization domains, a naturally trimeric protein (bromoperoxidase) and a dimeric protein (M1 virus matrix protein), using an a-helical linker in a geometrically controlled fashion. The designed cage assembled into a tetrahedral cage of 12 identical copies designed protein subunit ((50 kDa). This design is a variation of the above design reported in 2001, which assembled to form structures of heterogeneous size rather than a specific 12-subunit cage as intended. The current new design relied critically on making two amino acid changes to the earlier design.[97] In 2013, characterization and crystal structures were reported for the protein cages constructed by fusing two oligomerization domains in previous reports.[98]

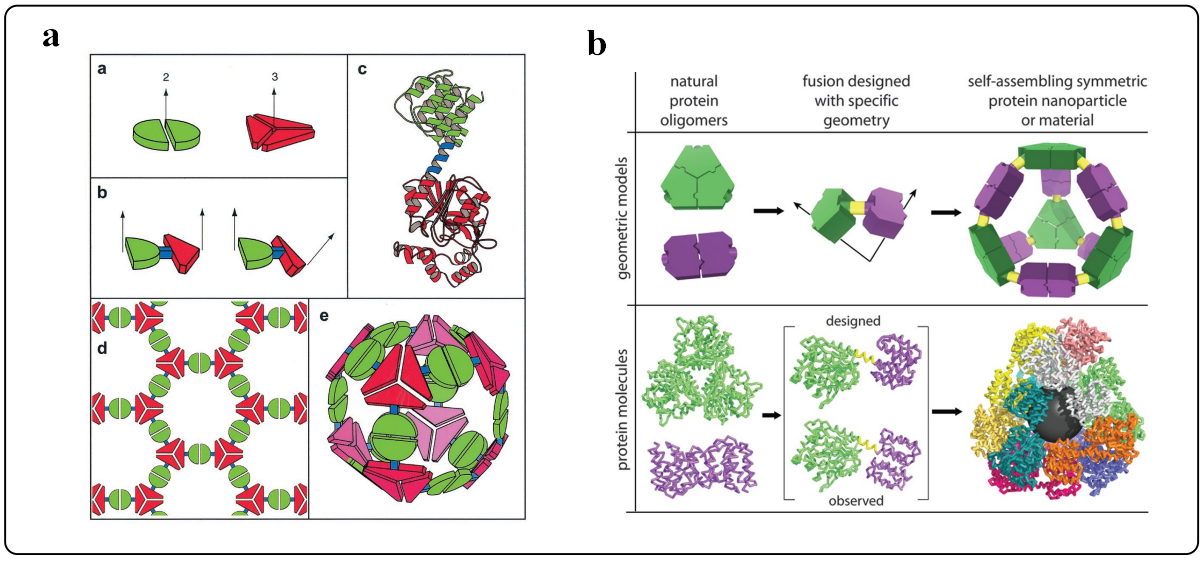


Figure 1.14. a. A general strategy for creating protein assemblies (symmetric nanostructures) by fusing dimeric protein and a trimeric protein. b. Atomic structure of a twelve subunit-tetrahedral protein cage by fusing two naturally oligomeric protein domains. Adapted with permission from ref. 87 and 97.

Further, to show the potential of symmetry-based fusion methods in creating complex architectures, an octahedral cube-shaped protein assembly was created by the same team. This octahedral protein assembly was reported from the genetic fusion of the trimeric Escherichia coli 2-keto-3-deoxy-6phosphogalactonate (KDPGal) aldolase31 (PDB ID: 2V82) and the dimeric N-terminal domain of E. coli FkpA protein32 (PDB ID: 1Q6U) using an α-helical linker. The designed novel protein self-assembled into 24 identical subunits, with molecular weight of 750 kDa, arranged in cubic (octahedral) symmetry and with central cavities of various diameters.[99]

Ryoichi Arai and co-workers, in 2015, created a protein nano building block (PNBlock) called WA20-foldon fusion protein, by fusing the dimeric Structure of WA20 to the trimeric foldon domain of fibritin from bacteriophage T4. It formed several homooligomeric forms like the small (S form), middle (M form), and large (L form) forms corresponding to hexamer (6-mer), dodecamer (12-mer), and octadecamer (18-mer), respectively. Further, the SAXS analyses showed that the S and M forms exist as a barrel-shaped hexamer and tetrahedron-like-shaped dodecamer, respectively.[88] Further, using a symmetry-based fusion method, very recently, an anionic hollow protein assembly was reported by Ryoichi Arai and his team,

using a designed fusion protein subunit. Monomers of pentamer (LSM) and dimer (MYOX-coil) forming proteins are fused and truncated icosahedral protein (TIP60) was created. It has 60 fused monomers lying at the two-fold symmetry axis. Designed TIP60 tolerates mutations and chemical modifications, on both the surfaces.[100]

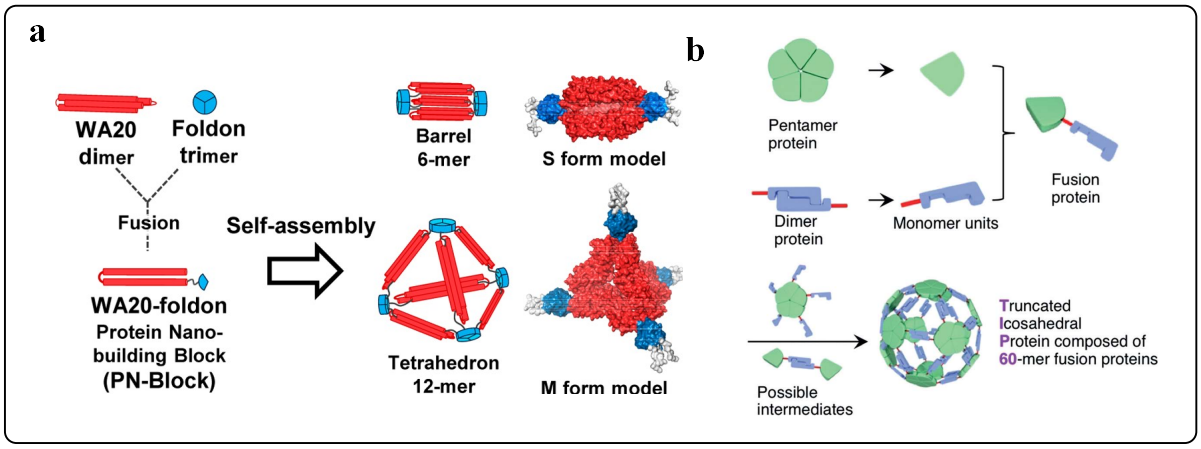


Figure 1.15. a. Schematic construction of protein assemblies construction by fusing WA20 dimer and foldon trimer. b. Schematic construction of protein nanoparticle by fusing LSm (PDB 3BY7) and MyoX-Coil (PDB 2LW9). Adapted with permission from ref. 88 and 100.

A ball and spike protein supramolecular structure was also constructed using symmetry-based fusion method by fusing C-terminal β-helix domains of gp5 from the T4 bacteriophage (gp5C) to the N terminus of the Dps from the bacterium Listeria innocua (LisDps) through a flexible linker (Both LisDps and gp5C have a threefold axis of symmetry).[101]

1.2.3. Symmetric oligomers and oligomeric coiled-coil-based fusion methods

Coiled coils are important structural motifs, often have 4 or more contiguous heptad repeats of hydrophobic (H) and polar/charged (P) residues, HPPHPPP (denoted: **abcdefg**). Coiled coils assemble primarily because of hydrophobic interactions and interhelix salt-bridging interactions.[102] Significant efforts of protein designs based on the fusion of symmetric oligomer with oligomeric coiled-coil were put forth by Neil G Marsh. In 2016, E. Neil G. Marsh and co-workers reported an approach to assemble proteins into cage-like structures using de novo designed coiled-coil domains. This method is based on the fusion of symmetric oligomer with oligomeric coiled-coil. For instance, By fusing c3-symmetric trimeric esterase with a c4-symmetric tetrameric coiled-coil domain through a short flexible linker sequence, a well-defined octahedral protein cage was reported.[89] Similarly, the same team based on the

concept of symmetric oligomer and oligomeric coiled-coils reported a well-defined tetrahedral cage by fusing trimeric enzyme with a trimeric coiled-coil domain through a flexible polyglycine linker sequence.[103]

Similarly, this methodology of protein assembly creation by fusing an oligomeric coiled-coil with the symmetric oligomer has been extended to a metal ion-dependent assembly by the same team. A tetrahedral protein assembly, containing 12 copies of the protein monomer was designed by fusing a trimeric coiled-coil domain with trimeric Tet8-M. The coiled-coil was designed to trimerize only response to the binding divalent transition metal ions like Ni^{2+}, Co^{2+}, Cu^{2+}, and Zn^{2+} ions**.** Ni2+ proved to be the most effective in inducing the tetrahedral protein assembly. This protein assembly could be reversed in the presence of ethylenediaminetetraacetic acid or acidic conditions.[104]

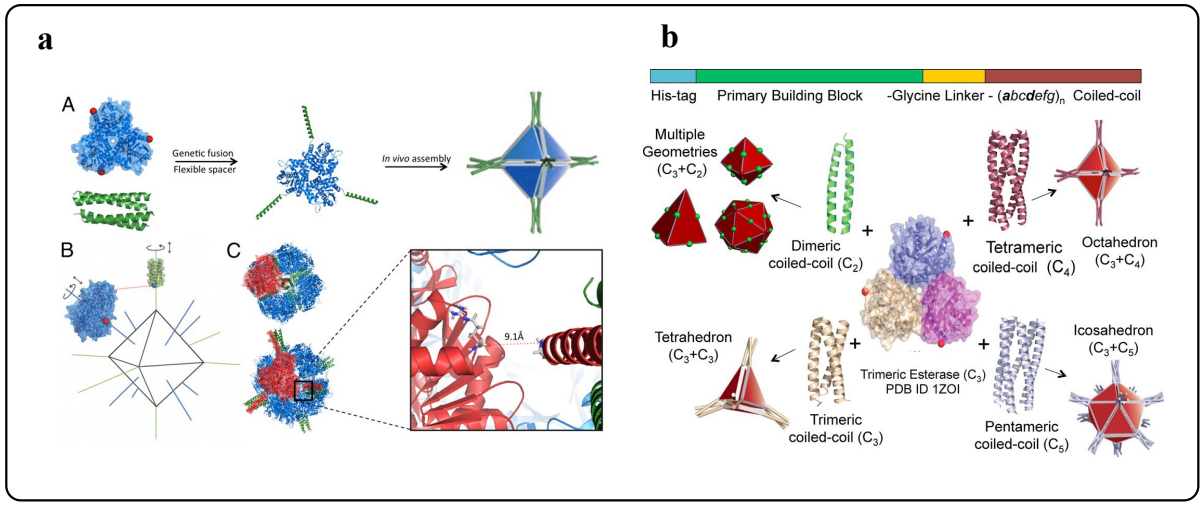


Figure 1.16. a. Design of a self-assembling octahedral protein cage from of the trimeric esterase (PDB 1ZOI) and the tetrameric coiled-coil (PDB 3R4A). b. Design of protein cages of different geometries by fusing the trimeric protein through a flexible linker to coiled coils with different oligomerization states. Adapted with permission from ref. 89 and 103.

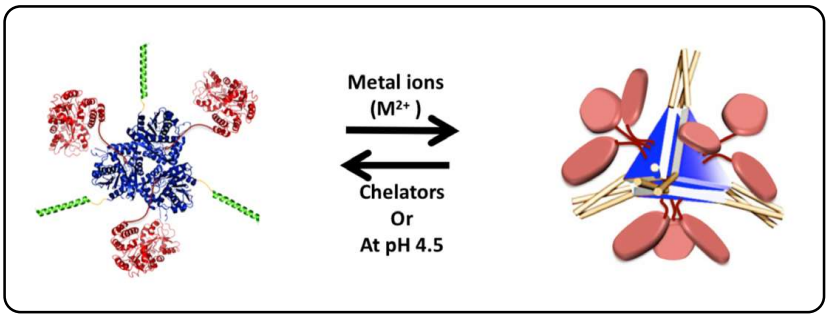


Figure 1.17. Cartoon illustrating the metal-induced trimerization of Tet8-M fusion protein. Adapted with permission from ref [104].

1.2.4. Coiled-coils-based protein origami

Protein origami is a new strategy for designing polypeptide polyhedra based on the optimization of orthogonal dimerizing segments. This is put forth by Roman Jerala and his group. In 2013, Jerala and co-workers reported a tetrahedron from a single polypeptide chain comprising of 12 concatenated coiled-coil forming segments in a defined order. The coiled-coil segments are separated by flexible peptide hinges, and they traverse each of the six edges of the tetrahedron exactly twice, driving the path for the formation of a polypeptide chain via coiled-coil dimers.[105] Similarly, in 2017, they reported more than 20 single-chain protein cages. The protein cages fall into three different shapes, a four-sided pyramid, a triangular prism, and nonsymmetrical polyhedra: a tetrahedron.[106]

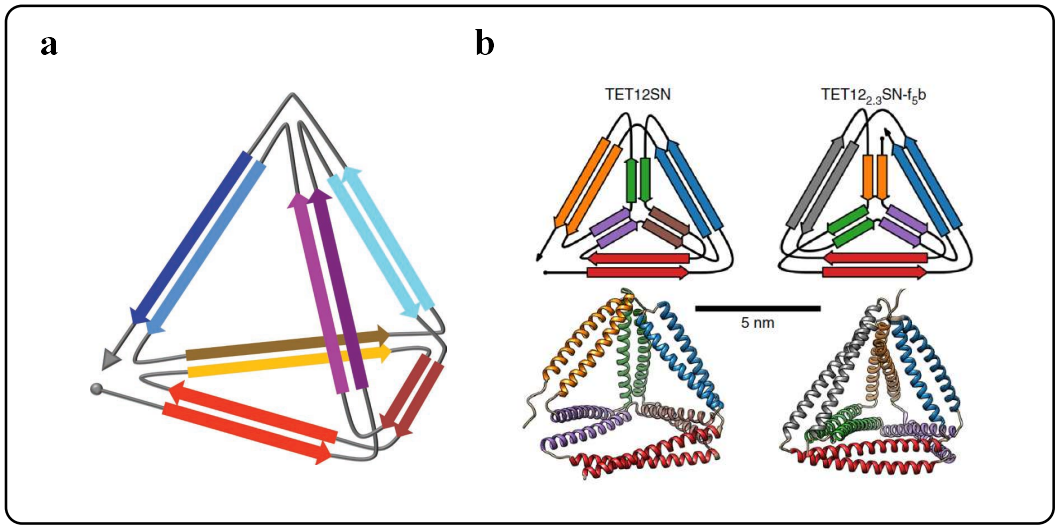


Figure 1.18. a. Design of the self-assembling polypeptide tetrahedron (Arrows denote the orientation of the coiled-coil elements in the assembly). b. Design of two tetrahedral structures based on protein origami. Adapted with permission from ref. 105 and 106.

1.2.5. Coiled-coils-ELPs-based methods

Coiled-coils and ELPs have self-assembling interactions and stimuli-responsive behavior respectively. Based on these features, coiled-coil-ELP hybrids were constructed to utilize properties of both coiled-coils and ELPs to design protein assemblies. For instance, in 2014, The Arg-rich leucine zipper motif (Z_R) forms coiled-coil complexes with its counterpart, the

Glu-rich leucine zipper motif (Z_E). Using this to drive self-assembly, Champion et al. created protein vesicles by combining three different diblock recombinant proteins: Z_R-ELP, mCherry-ZE, and EGFP-ZE using temperature as a trigger.[107,108] Similarly, in 2015, a protein-based nanoparticle with both drug encapsulation and active targeting ability was reported by the Kobatake group. This was done by fusing two peptides (known to form a heterodimer via a coiled-coil motif) to ELP-D and single-chain vascular endothelial growth factor (scVEGF). These two scaffolds formed a heterodimer coiled-coil Structure, which formed drug encapsulated protein particles upon heating in the presence of the drug.[109] In 2017, A novel helix-elastin like polymer (HELP) was created which forms highly specific protein helical zippers and short elastin segments using computational design (rosetta), which retained the specificity of coiled-coil interactions and stimuli-responsive behavior.[110]

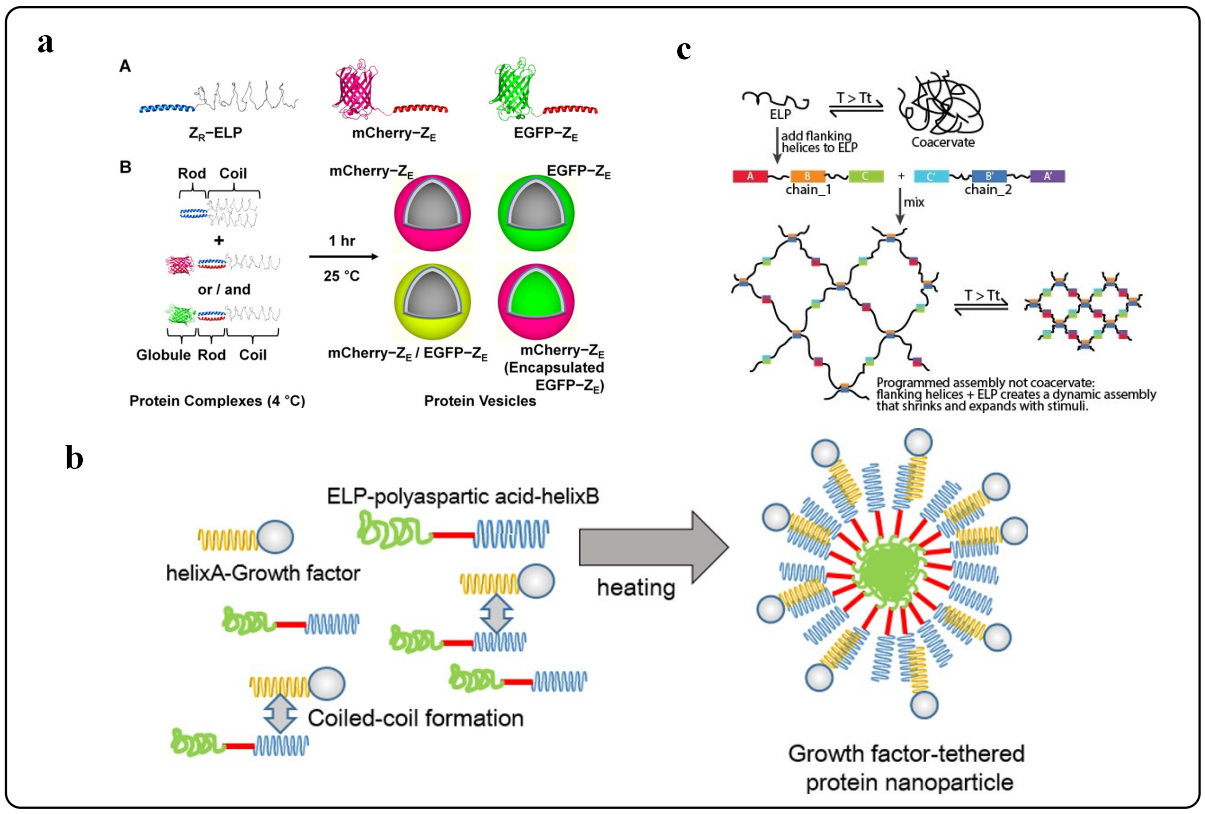


Figure 1.19. a. The self-assembly of recombinant protein amphiphiles (ZR-ELP, mCherry-ZE, and EGFP-ZE) into vesicles. b. Schematic representation of a growth factor-tethered protein nanoparticle via coiled-coil formation. c. Temperature responsive 2D networks of an elastomeric polymer created using helical zippers. Adapted with permission from ref. 107, 108, 109 and 110.

1.2.6. Coiled-coil-based other methods

Champion and her team combined An Fc-binding protein fragment with an α-helical peptide that self-assembles naturally into a hexameric coiled-coil bundle, and a hex nanocarrier was generated that efficiently internalizes into cells without cytotoxicity. The size of the hex nanocarrier ranged from 25 to 35 nm, depending on the antibody being loaded. Localization of multiple Fc-binding domains on the hexameric core allowed the hex nanocarrier to tightly bind antibody with sub-nanomolar affinity. The delivery capacity of the hex nanocarrier was demonstrated using functional antibodies like anti-β-tubulin or anti-nuclear pore complex antibody. Efficient internalization of hex nanocarrier without cytotoxicity in HeLa cells was reported.[111]

1.2.7. Hexameric HTHP hemeprotein-based methods

HTHP is a C6-symmetric, naturally hexameric tyrosine-coordinated heme protein and has been explored as a building block for the construction of higher-order assemblies. For instance, Hayashi and co-workers, in 2019, a 2D sheet assembly of an HTHP, was constructed. A single cysteine residue was introduced onto each subunit surface in the hexameric hemoprotein. This cysteine was modified with a maleimide-tethering tripeptide FGG tag to provide a host-guest interaction with one cucurbit[8]uril molecule (CB8) and two FGG tags, resulting in engineered sheet-like protein assembly.[112]

Similarly, in 2019, a supramolecular dimer was created by the same team using hexameric tyrosine coordinated hemoprotein (HTHPV44C) as a protein assembly building block. The supramolecular dimer of the hexamers was constructed using multiple pyrene-pyrene interactions by the introduction of a pyrene moiety onto each protein subunit.[113] They have also constructed a thermoresponsive protein micellar assembly, using a mutant of HTHP and poly(*N*-isopropyl acrylamide) (PNIPAAm), as the building blocks. Then, using this assembly, an artificial light-harvesting system was also created by the substitution of heme with a photosensitizer, Zn protoporphyrin IX (ZnPP).[114]

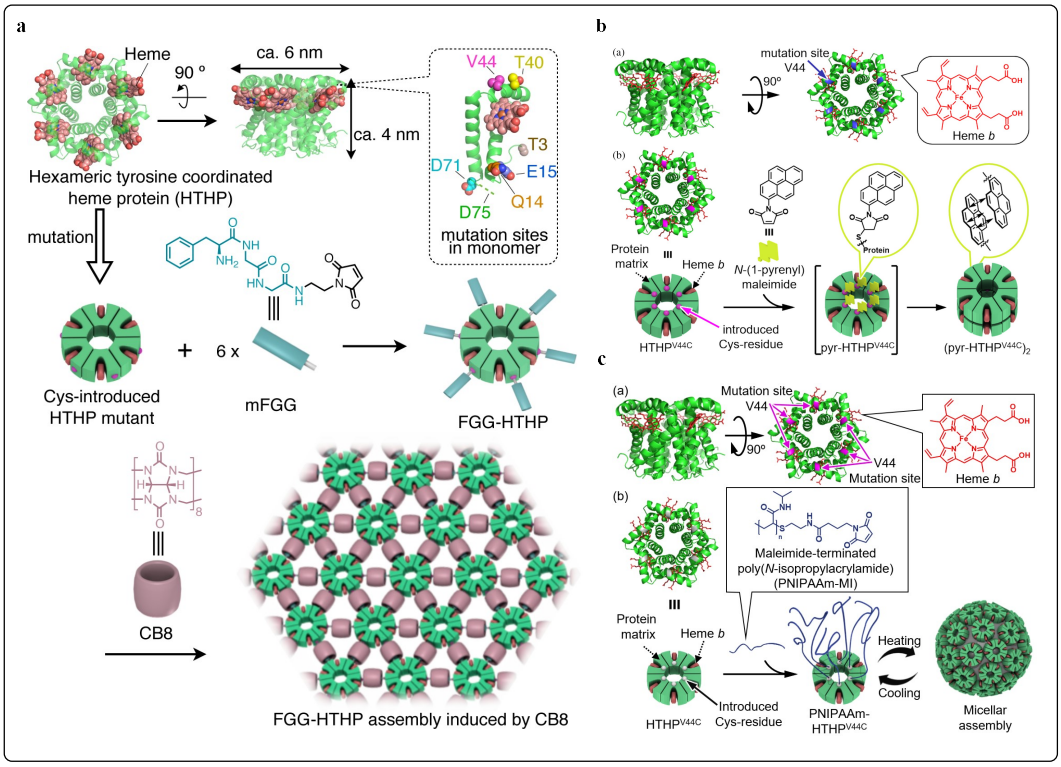


Figure 1.20. a. CB8 induced a 2D sheet assembly of a hexameric hemoprotein. b. *N*-(1-pyrenyl) maleimide induced supramolecular dimerization of the hexameric hemoprotein. c. PNIPAAm induced thermoresponsive micellar assembly formation of the hexameric hemoprotein. Adapted with permission from ref. 112, 113 and 114.

1.3. Engineering monomeric proteins

Even though the protein engineering strategies which use protein oligomers are emerging as powerful methods for the design of synthetic protein complexes with high accuracy and fidelity, precise construction protein complexes/assemblies from monomeric proteins has also evolved in recent years, on par with other methods. Assembling monomeric proteins is completely different from assembling oligomers. This strategy relies on the addition of new components (small molecules/polymers/peptides/dendrimers, etc.), which can significantly induce/influence the self-assembly.

1.3.1. ELP fusion-based methods

The Chilkoti group extensively studied the ELPs. A range of ELP block copolymers with varying hydrophilic-to-hydrophobic block ratios was made genetically, and their self-assembly behavior was studied, and the factors responsible for the critical micelle concentration (CMC) and critical micelle temperature (CMT) were determined.[115] Temperature triggered collapsing of ELPs is the main driving force these methods to assemble the fusion protein of ELPs. For example, temperature triggered reversible immobilization of thioredoxin-ELP (Trx-ELP) fusion protein onto the ELP nanopattern above the LCST was reported in 2004. The orientation of Trx-ELP was found to be accessible as the Trx-ELP can bind to an anti-thioredoxin monoclonal antibody (anti-Trx) in the nanoarray.[116]

Similarly, temperature triggered, spherical micelles formation was reported by the Chilkoti group using thioredoxin (Trx) and a fibronectin type iii domain (Fn3) by the genetic incorporation of ELP block copolymers (ELP_{BC}s). The hydrodynamic radius of the spherical micelles formed was in the range of 24 to 37 nm, depending on ELP_{BC}s.[117] In 2016, Olsen and co-workers studied hydrophobicity and molar mass of ELPs on self-assembly of ELP fusion proteins. Two different ELP blocks, ELP0 and ELP1, were fused with mCherry to get different ELP0-based and ELP1-based mCherry-ELP fusions and they found that the hydrophobicity and molar mass both promotes self-assembly.[118]

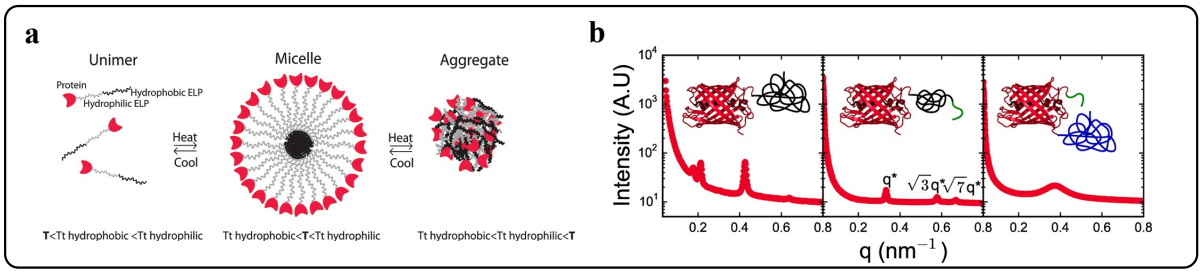


Figure 1.21. a. Three state behavior of protein-ELPBC fusion, unimer, micelle, and a polydisperse micrometer-sized aggregate. b. designed mCherry-ELP fusion block copolymers using two different ELP sequences (ELP0 and ELP1), showing different SAXS profiles. Adapted with permission from ref. 117 and 118.

1.3.2. Heme-hemeprotein-based methods

Hemeproteins, are a family of cofactor heme-containing proteins. Examples of hemeproteins include hemoglobin, myoglobin, cytochromes, and P450 peroxidases. The hemeproteins (apoproteins) and their high affinity for heme (cofactor) were explored to design novel protein assemblies. Hayashi and this team, for instance, designed a unique supramolecular linear assembly of hemoprotein (cytochromeb562) using the high affinity of apoprotein for heme moieties. The linear assembly was achieved by designing a monomeric heme protein (H63C single mutant) unit with cofactor moiety at the back of the heme pocket of the protein.[119] Using a similar approach, in 2009, the same group designed a two-dimensional network of hemoprotein by the introduction of the heme triad 3 molecule.[120]

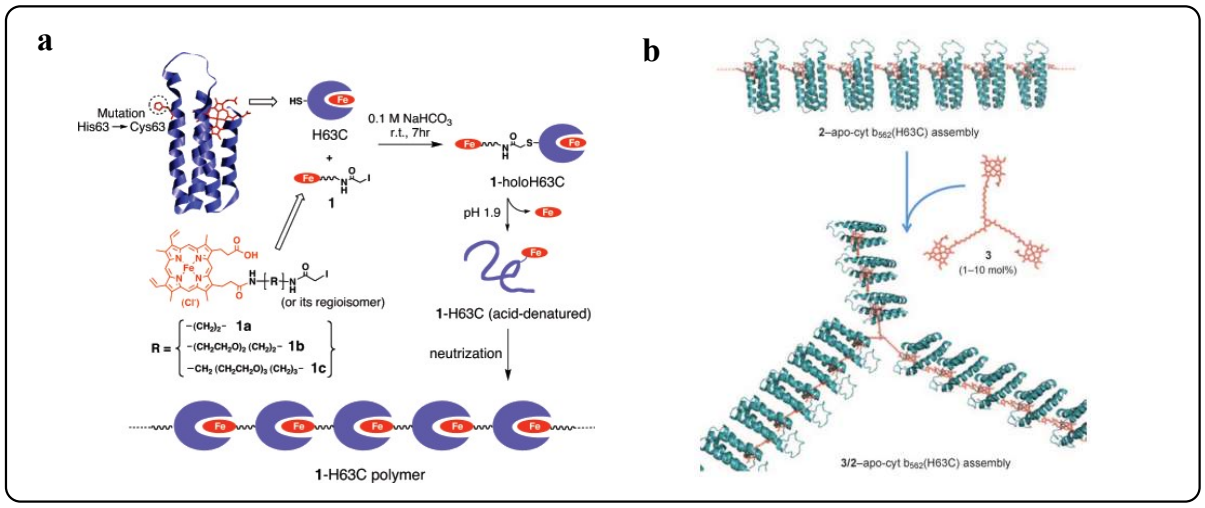


Figure 1.22. a. Schematic representation of the supramolecular hemoprotein linear polymerization. b. Schematic representation of 2D network of hemoprotein. Adapted with permission from ref. 119 and 120.

Apart from cytochromeb562, myoglobin was also employed to create supramolecular protein nanostructures. For example, In 2011, myoglobin (Mb, A125C), was used to construct supramolecular linear protein nanostructure, by attaching the synthetic heme moiety using thiol-maleimide conjugation.[121] Using a similar approach, heterotropic protein-protein complexation between homotetrameric streptavidin and dimeric apomyoglobin (MbA125C) was demonstrated by using an artificial prosthetic group with two biotin units tethered through a flexible linker to one heme-propionate side-chain.[122]

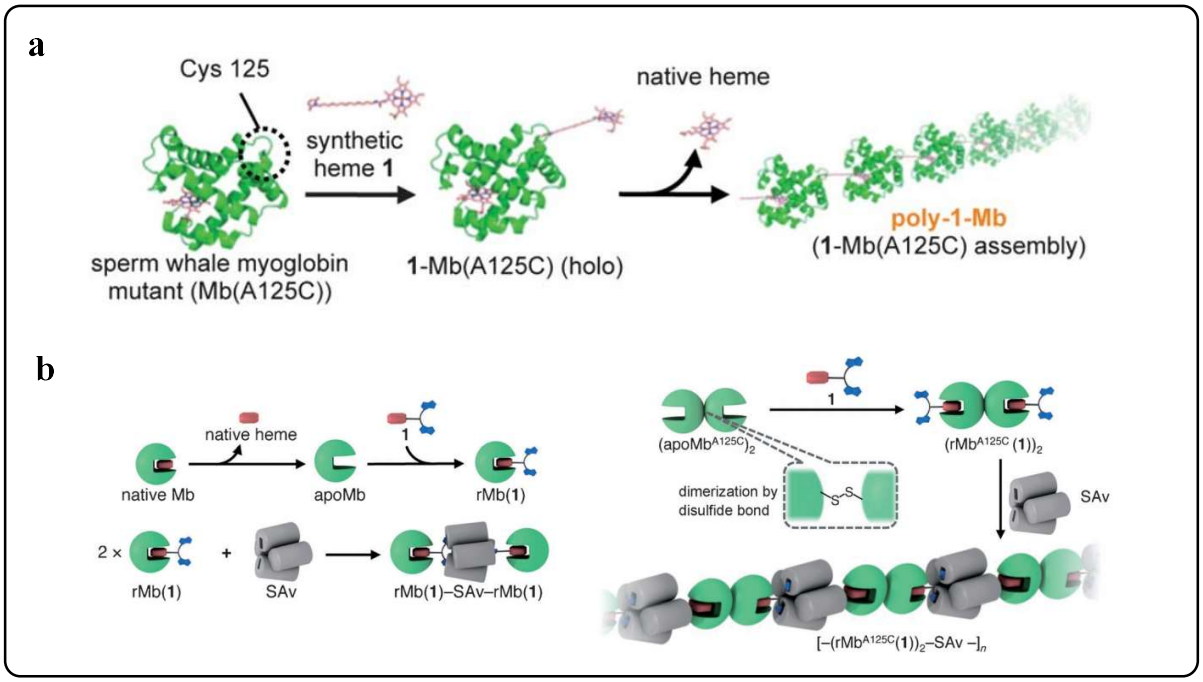


Figure 1.23. a. Schematic representation of the supramolecular myoglobin linear polymerization. b. Schematic representation of heterotropic protein-protein complexation between homotetrameric streptavidin and dimeric apomyoglobin. Adapted with permission from ref. 121 and 122.

Cytochrome b562 was also converted into a thermodynamically stable, ring-shaped trimer in 2018 by attaching heme moiety externally using an appropriate linker.[123] Further, bioorganometallic catalysts was demonstrated by the reconstitution of hemoprotein with unnatural cofactors.[124]

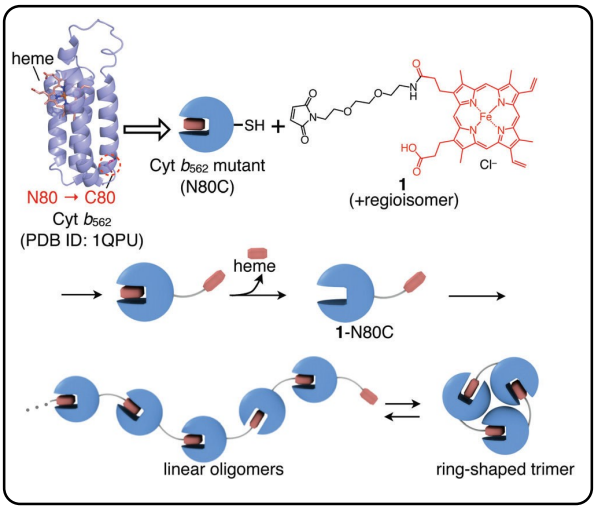


Figure 1.24. a. Cartoon representation of maleimide-tethered heme showing self-assembled states: linear oligomer and ring-shaped trimer. Adapted with permission from ref. 123.

1.3.3. Enzyme-inhibitor-based methods

Enzyme-inhibitor chemistry played a crucial role in protein engineering. For example, cyclic protein mega molecules were reported by Wagner using enzyme-inhibitor reactions. They have designed dihydrofolate reductase (DHFR) fusion proteins with varying linkers and achieved cyclic self-assembled structures by using a DHFR inhibitor.[125] Similarly, in 2010, the synthesis of cyclic protein oligomers, having diameters of 10-20 nm, was reported by Mrksich and his team. N-terminal cutinase and C-terminal SnapTag domains react irreversibly with p-nitrophenyl phosphonate (pNPP) and benzylguanine (BG) groups, respectively. Using these molecules and CS fusion protein, cyclic protein oligomers were generated.[126]

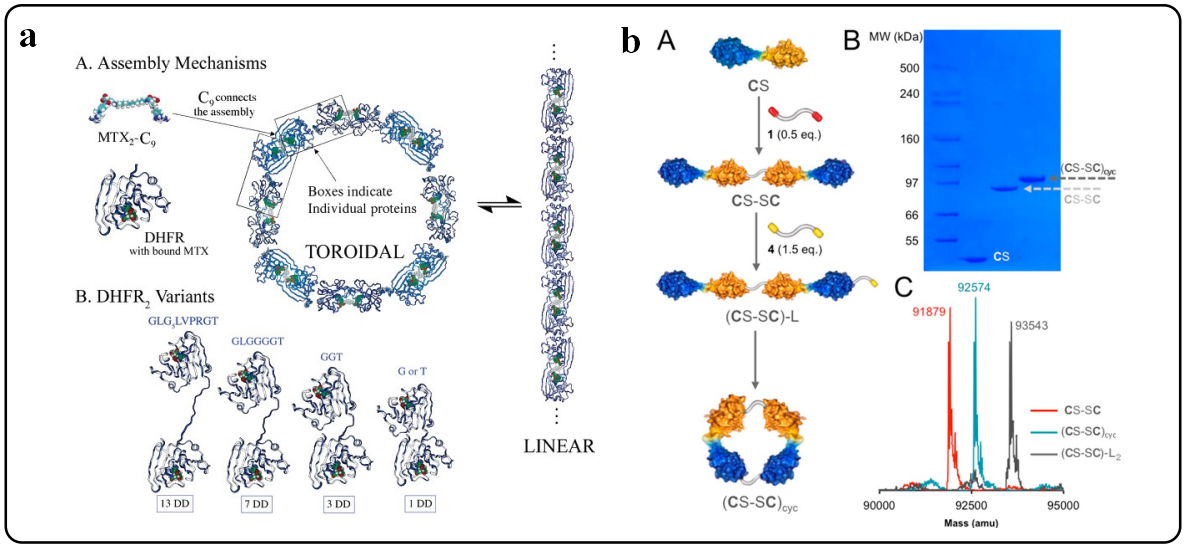


Figure 1.25. a. cyclic protein assembly of ecDHFR2 and MTX2-C9. b. synthetic scheme of linear and cyclic forms of the CS dimer, CS-SC. Adapted with permission from ref. 125 and 126.

1.3.4. Other fusion-based methods

Yongwon Jung and his team used spontaneous assembly of 11th β strand of green fluorescent protein (GFP) to produce monodispersed protein assemblies in 2D fashion inside the cells by introducing a spacer peptide sequence between GFP 1-10 and GFP 11. They even showed that the assemblies could be purified by gel-based electroelution.[127] In 2018, a new kind of fusion proteins that can self-assemble into micelles of diameter 27 nm was reported. The protein sequences is based on intrinsically disordered proteins (IDPs), which has been appended with a hydrophobic segment. The micelles were shown to solubilize hydrophobic metal complexes and organic molecules.[128]

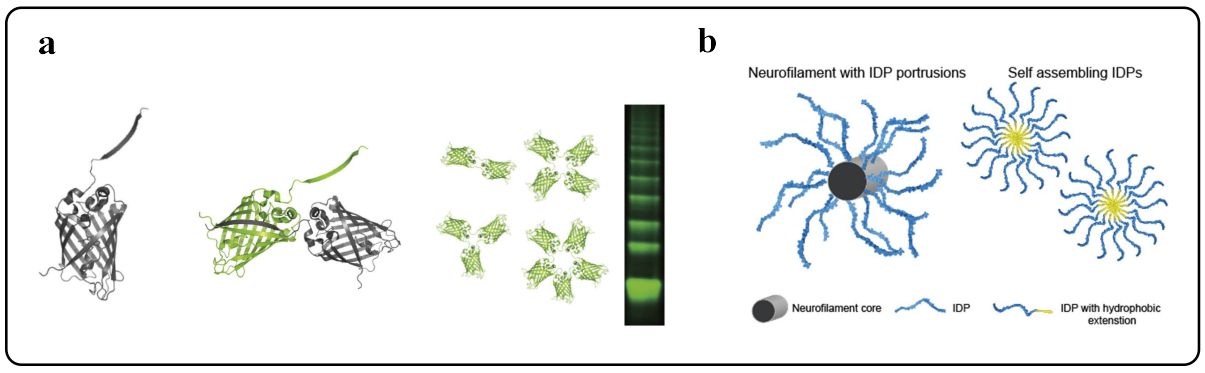


Figure 1.26. a. GFP (nano) polygons were achieved by connecting the 11th β strand of GFP to the N terminus of the rest of GFP (1–10 β strands) using a short peptide linker. b. Cartoon representation of genetically incorporation of a hydrophobic domain at the C-terminus of the IDP, and self-assembly around an artificial core. Adapted with permission from ref. 127 and 128.

1.3.5. Chemical strategies/protein-polymer conjugates.

Most chemical protein engineering strategies are based on the generation of protein conjugates using site-specific chemistries like thiol-maleimide; they lack precise site-specific engineering, characterization, and often yield polydisperse samples which may limit their potential applications and therefore lack high-resolution structural information. The underlying design rules are still elusive, and methods are restricted to few proteins and lack a modular approach. However, Protein engineering utilizing chemical technologies offer remarkable advantages over genetic method as they are not restricted to standard 20 canonical amino acids. For example, Protein engineering using chemical strategy provides an opportunity for the design of stimuli-responsive protein complexes because of access to innumerable building blocks.

The majority of the chemical protein engineering strategies are based on protein-polymer conjugates generated either by "grafting to" or "grafting from" approaches.

1.3.5.1. "Grafting to" approaches

In the first method, the polymer is attached to the protein by chemical method. This is a popular conjugation strategy as it permits the controlled polymerization, molecular characterization of the polymer before conjugation. However, the conjugation of a synthetic polymer to a protein can be challenging (large excess of the polymer and organic solvents are needed). Another challenge would be the separation of the unreacted polymer from the bioconjugate.

The synthetic or chemical protein engineering reports date back to the early 2000s. One such report is from the Nolte group, in the year 2001. A lipase-polystyrene amphiphile (enzyme-polymer hybrid, a giant amphiphile) was reported using thiol-maleimide chemistry by functionalizing the disulfide bridge (cys293-cys311). They studied its self-assembling behavior in water.[129] In another study, using thiol-maleimide conjugation, Nolte et al. reacted

thiol (Cys-34 residue) of BSA with an alkyne bearing maleimide. Then biohybrid amphiphiles were made by reacting with polystyrene block via "click" chemistry.[130] Another early report of giant amphiphilic protein-polymer hybrids was reported using affinity as a driving force by the same group. This biohybrid was constructed by the association of two biotinylated polystyrene molecules with streptavidin. Streptavidin homo-tetramers have a high affinity for biotin.[131]

Nolte et al., in the same year, reported the amphiphilic protein-polymer hybrid, synthesized by the cofactor reconstitution method. The modification of an HRP enzyme with macromolecular polystyrene formed aggregate in aqueous media, and this conjugate retained some catalytic activity.[132] Similarly, the same team in 2006, using the cofactor reconstitution method, protein-polymer giant amphiphiles were made, and their aggregation behavior was studied. In this report, the horseradish peroxidase (HRP) and myoglobin (MB) were modified with an apolar polystyrene polymer chain, and the resulting conjugates formed spherical aggregates (vesicles) in aqueous solution.[133] In 2007, using diblock copolymer, the synthesis and self-assembly behavior of protein-polymer biohybrid was reported using myoglobin (MB) or horseradish peroxidase (HRP). The synthetic diblock copolymer polystyrene b-polyethylene glycol (PSm-b-PEG113), is functionalized with the heme cofactor and then by reconstitution, protein-polymer biohybrids (triblock copolymers) were made. The obtained amphiphilic biohybrids self-assembled into a variety of aggregates in aqueous solution.[134]

In 2011, using the thiol-maleimide conjugation, protein-polymer conjugate (block copolymer) was reported using a mCherry (S131C) and end-functionalized poly(n-isopropyl acrylamide) (PNIPAM).[135] In 2012, Olsen et al., using thiol-maleimide bioconjugation, studied the self-assembly of mCherry-PNIPAM block copolymers with respect to the polymer coil fraction and the solvent. These conjugates self-assembled into cylinders, perforated lamellae, lamellae, or hexagonal and disordered micellar phases depending on polymer coil fraction and the solvent.[136]

The same group systematically studied the effect of small molecule osmolytes,[137] effects of polymer chemistry, [138], and the nature of protein interactions[139] on self-assembly behavior of mCherry-PNIPAM conjugates. In another study, in 2015, Olsen et al. synthesized myoglobin (Mb)-PNIPAM bioconjugates using a thiol-maleimide coupling. The resulting bioconjugate shows the enhanced activity of enzyme when devised as film.[140]

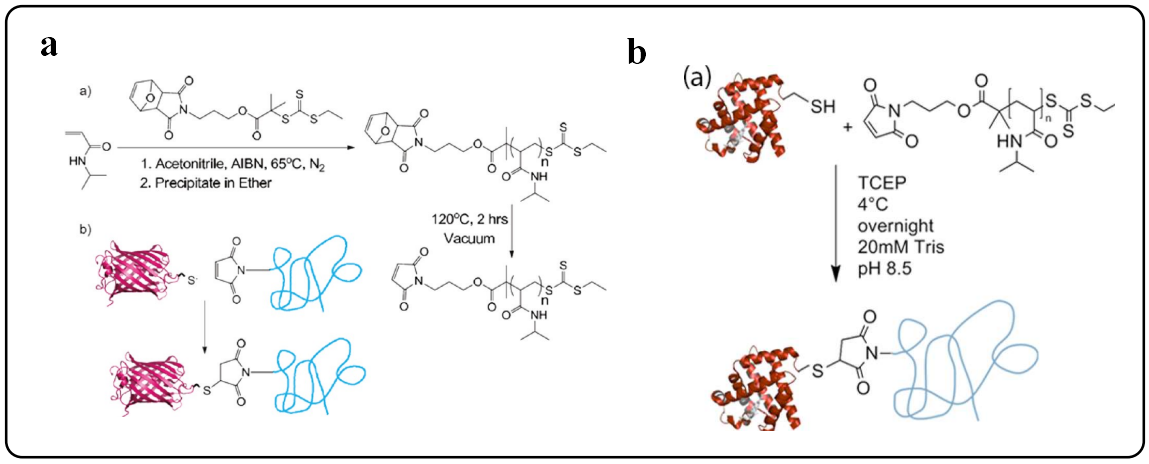


Figure 1.27. a. Synthesis of maleimide-functionalized pNIPAAm and its bioconjugation to mCherry S131C to create a protein-polymer diblock copolymer. b. hMb-PNIPAM bioconjugates synthesized using a thiol-maleimide coupling. Adapted with permission from ref. 135 and 140.

Yingqin et al. reported protein-poly (amino acid) (PAA) conjugates in 2016. With in situ installation of two "chemical handles," a thioester for native chemical ligation, and a polyglycine nucleophile for sortase A-mediated ligation, this method avoids pre-functionalization of initiator or monomer units, nor post-polymerization modification of the polymers.[141]

In a very recent study, Ju et al. reported covalently connected polymer–protein nanostructures fabricated through a reactive self-assembly approach. Synthesis of the polymer, poly(tert-butyl methacrylate-co-pyridyl disulfide methacrylamide) (PtBMA-co-PPDSMA) was done by RAFT polymerization. The thiol-disulfide exchange reaction between pyridyl disulfide groups on the polymer chains and thiol groups on the protein molecules yielded covalently connected nanostructures (CCNs). [142]

1.3.5.2."Grafting from" approaches

In this second method ("grafting from" approach), the polymer is generated directly, in situ, on the protein itself (protein as a macroinitiator) using Atom transfer radical polymerization (ATRP) or reversible addition-fragmentation chain transfer (RAFT) polymerization techniques. It has several disadvantages, like the polymerization conditions that need to be done in aqueous media and often result in high polydispersities. The characterization of the polymer is extremely difficult. The purification is also challenging in this strategy.

As discussed earlier, "grafting from" uses proteins as a macroinitiator. Maynard et al. prepared protein-polymer conjugate by using protein as a macroinitiator. A known biotinylated ATRP initiator was prepared and interacted with streptavidin to generate protein macroinitiator. Then by in situ polymerizations of NIPAAm, protein conjugate was made.[143]

Using NIPAAm, two different protein-polymer conjugates were also reported by the Maynard group in 2005. The free Cys-34 of bovine serum albumin (BSA) and Cys-131 of lysozyme were modified with initiators for atom transfer radical polymerization (ATRP). Then the respective thermosensitive protein-polymer conjugates were made by polymerization of NIPAAm from the protein macroinitiators.[144]

Armando et al. reported in situ polymerizations of a water-soluble monomer from a protein with an initiator, which yielded amphiphilic protein conjugates of a water-insoluble polymer. HSA was Site-specifically modified with an ATRP initiator to form a macroinitiator (HSA-Br), followed by polymerization of 2-hydroxypropyl methacrylate (HPMA) from HSA-Br to yield HSA conjugates of a hydrophobic polymer. These conjugates assembled into nanostructures with tunable morphologies from spheres to vesicles.[145]

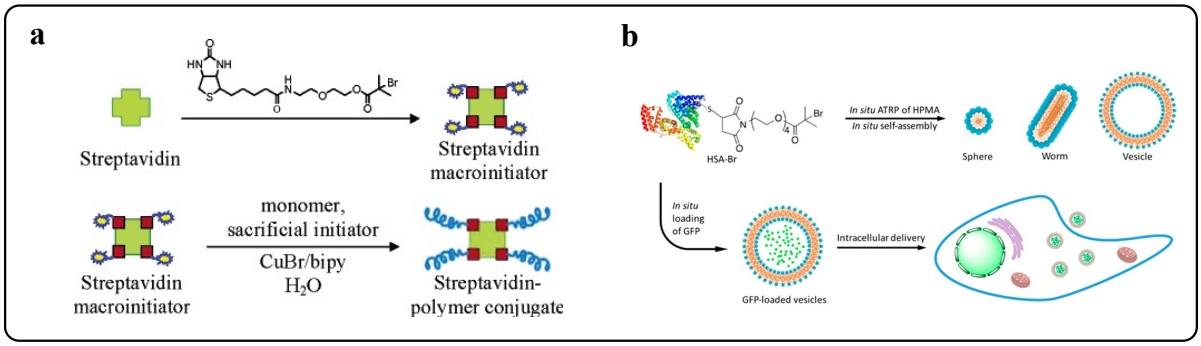


Figure 1.28. a. Streptavidin-initiated polymerization of NIPAAm. b. Schematic representation of in situ growth of HSA-PHPMA nanostructures. Adapted with permission from ref. 143 and 145.

Xin Huang group efficiently synthesized protein-based nanoparticles is by taking advantage of photoinitiated reversible addition-fragmentation chain transfer (RAFT) polymerization. They induced self-assembly by using multi-RAFT modified bovine serum albumin (BSA) as a macro-RAFT agent. The particles showed encapsulation ability toward both hydrophobic and hydrophilic components. They also demonstrated the protease-mediated release of the encapsulants after loading cancer drug DOX or biomacromolecule DNA.[146]

1.3.5.3. Smart protein-polymer conjugates

As discussed previously, protein engineering using chemical strategy provides an opportunity for the design of stimuli-responsive protein complexes because of access to innumerable building blocks. Most of the reported strategies still could not take full advantage of this scope.

Most of the reported stimuli-responsive protein complexes are thermo-responsive in nature. For example, in 2010, by using a combination of ATRP and a click reaction, PNIPAM bearing porphyrin moiety (PPIXZn-PNIPAM) was synthesized. Then by subsequent cofactor reconstitution process between apomyoglobin and PPIXZn-PNIPAM, a thermo-responsive myoglobin-b-PNIPAM protein-polymer conjugate was made.[147] The Olsen group demonstrated site-specific conjugation of RAFT homopolymers and blocked copolymers to GFP, via Expressed Protein Ligation (EPL) to generate linear conjugate. By EPL followed by thiol–maleimide ligation, a Y-shaped conjugate with two different polymers was synthesized. Both linear and Y-shaped GFP–PDMA–PNIPAM BCP conjugates formed stable, thermo-responsive micelles.[148]

Another example of temperature-responsive protein-polymer bio-conjugate was synthesized by Reilly et al. in 2015 using a super folder green fluorescent protein (sfGFP). The azide-containing noncanonical amino acid p-azidophenylalanine (pAzF) was site-specifically incorporated into sfGFP, and alkyne terminated PEGMA was synthesized by RAFT polymerization. Then using copper-catalyzed "click" chemistry, sfGFP was conjugated to the PEGMA.[149]

Similarly, a simple bioconjugate composed of green fluorescent protein (GFP) molecules and dialkyne poly(ethyleneoxide) (PEO) that self-assembles into extended fibrous structures reversible was reported.[150] Caruso et al. in 2015 reported a strategy to convert proteins into pH cleavable particles that retain protein functionality after disassembly. An acid-labile reversible cross-linker based on amine-maleic anhydride chemistry was used to cross-link proteins with 8-arm-PEG. The protein particles can be rapidly disassembled at a mildly acidic pH (<6.5).[151]

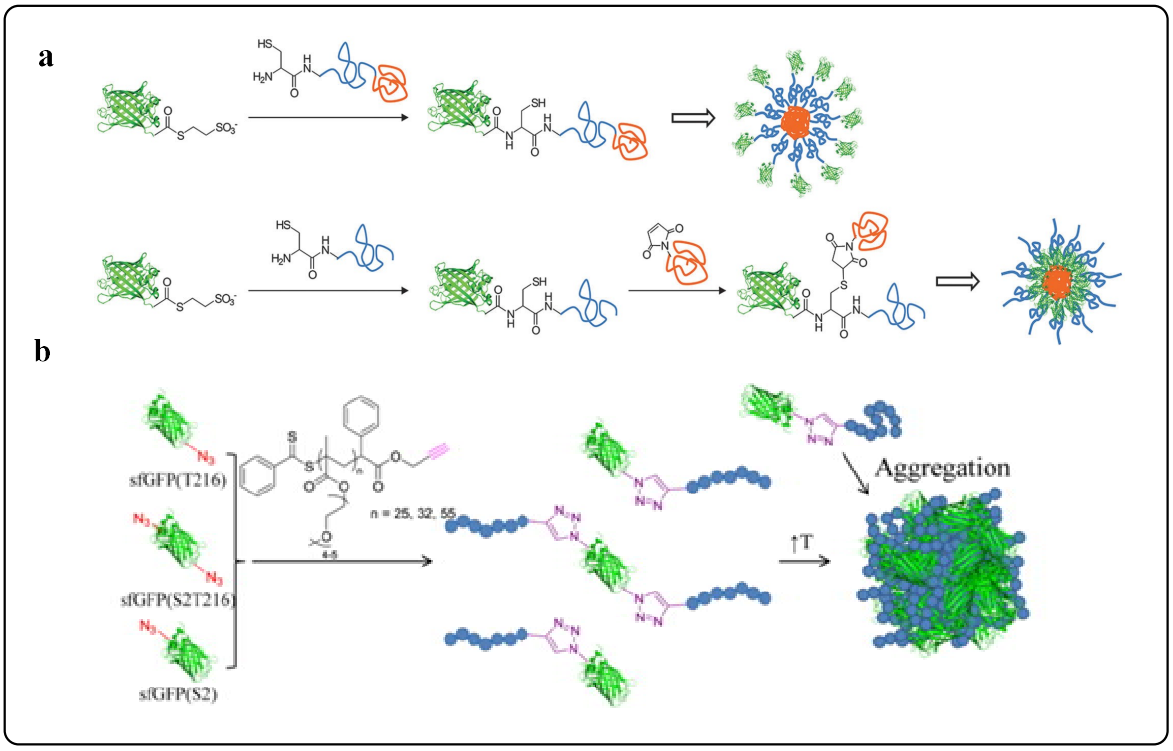


Figure 1.29. a. Synthesis of linear (top) and Y-shaped (bottom) GFP–block copolymer conjugates and their proposed micellar assembly. b. synthetic strategy for sfGFP-PEGMA bioconjugates and its temperature-dependent aggregation. Adapted with permission from ref. 148 and 149.

1.3.5.4. Cross-linked protein-polymer conjugates (proteinosomes)

Proteinosomes are a different, cross-linked class of protein-polymer conjugates. In 2013, Stephen Mann and co-workers reported an approach to interfacially assemble protein-based micro-compartments (proteinosomes) using protein-polymer nano-conjugates (BSA-NH_2/PNIPAAm). Protein-polymer conjugates were synthesized with a mercaptothiazoline-activated terminal amide containing PNIPAAm. The formed protocells exhibited various protocellular properties like encapsulation, selective permeability, gene-mediated protein synthesis, and internalization of enzyme and its catalysis.[152] Han et al. reported another novel approach to form proteinosomes based on protein-polymer conjugates and their supramolecular self-assembly. BSA with multiple thiols on the surface was generated, and BSA-βCD was obtained after an exchange reaction between thiols on BSA and βCD-s-s-Py. The supramolecular protein-polymer conjugate was prepared by mixing βCD modified bovine serum albumin (BSA) and adamantine terminated poly (N-isopropylamide) (Ad-

PNIPAM) in aqueous solution. This dual-responsive proteinosomes dissociated in the presence of trypsin or glutathione.[153]

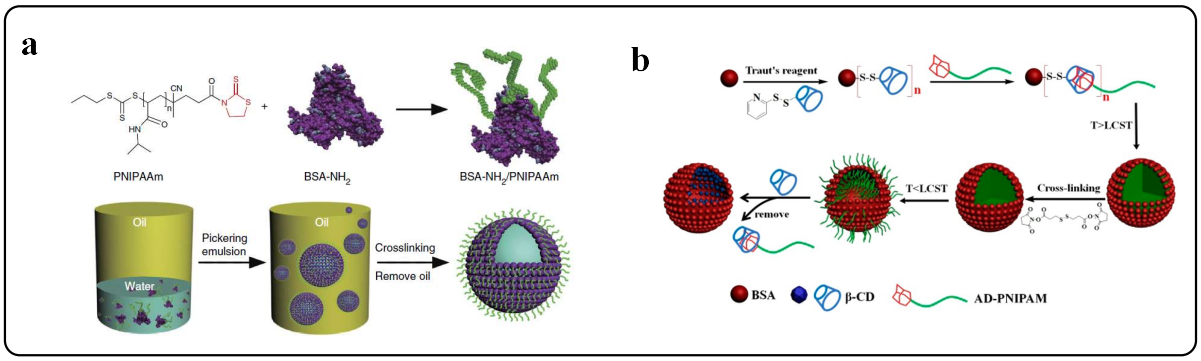


Figure 1.30. a. protein-polymer nano-conjugates (BSA-NH2/PNIPAAm) base proteinosomes generation protocol. b. Proteinosomes based on self-assembly of a supramolecular bioconjugate of βCD modified BSA and Ad-PNIPAM. Adapted with permission from ref. 152 and 153.

References

1. Luo, Q., Dong, Z., Hou, C. & Liu, J. Protein-based supramolecular polymers: progress and prospect. *Chem. Commun.***2014**, 50, 9997-10007.
2. Estrada, L. H. & Champion, J. Protein nanoparticles for therapeutic protein delivery. *Biomater. Sci.***2015**, 3, 787-799.
3. Luo, Q., Hou, C., Bai, Y., Wang, R. & Liu, J. Protein Assembly: Versatile Approaches to Construct Highly Ordered Nanostructures. *Chem. Rev.***2016**, 116, 13571-13632.
4. Obermeyer, A. C. & Olsen, B. D. Synthesis and application of protein-containing block copolymers. *ACS Macro Lett.***2015**, 4, 101-110.
5. Wen, A. M. & Steinmetz, N. F. Design of virus-based nanomaterials for medicine, biotechnology, and energy. *Chem. Soc. Rev.* **2016**, 45, 4074-4126.
6. Lee, E. J., Lee, N. K. & Kim, I.-S. Bioengineered protein-based nanocage for drug delivery. *Adv. Drug Delivery Rev.* **2016**, 106, 157-171.
7. Bhaskar, S. & Lim, S. Engineering protein nanocages as carriers for biomedical applications. *NPG Asia Mater.* **2017**, 9, e371-e371.
8. Wilson, C. J. *et al.* Biomolecular Assemblies: Moving from Observation to Predictive Design. *Chem. Rev.* **2018**, 118, 11519−11574.
9. Morris, C. *et al.* A Modular Vaccine Platform Combining Self-Assembled Peptide Cages and Immunogenic Peptides. *Adv. Funct. Mater.* **2019**, 1807357.
10. Lee, L. A. & Wang, Q. Adaptations of nanoscale viruses and other protein cages for medical applications. *Nanotechnol. Biol. Med.* **2006**, 2, 137-149.
11. Douglas, T. & Young, M. Viruses: making friends with old foes. *Science* **2006**, 312, 873-875.
12. Caspar, D. L. & Klug, A. in *Cold Spring Harbor symposia on quantitative biology*. 1-24 (Cold Spring Harbor Laboratory Press).
13. Speir, J. A., Munshi, S., Wang, G., Baker, T. S. & Johnson, J. E. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* **1995**, 3, 63-78.
14. Brumfield, S. *et al.* Heterologous expression of the modified coat protein of Cowpea chlorotic mottle bromovirus results in the assembly of protein cages with altered architectures and function. *J. Gen. Virol.* **2004**, 85, 1049-1053.

15. Tang, J. *et al.* The role of subunit hinges and molecular "switches" in the control of viral capsid polymorphism. *J. Struct. Biol.* **2006**, 154, 59-67.

16. Minten, I. J., Hendriks, L. J., Nolte, R. J. & Cornelissen, J. J. Controlled encapsulation of multiple proteins in virus capsids. *J. Am. Chem. Soc.* **2009**, 131, 17771-17773.

17. Brasch, M. *et al.* Assembling enzymatic cascade pathways inside virus-based nanocages using dual-tasking nucleic acid tags. *J. Am. Chem. Soc.* **2017**, 139, 1512−1519.

18. Brasch, M. *et al.* Encapsulation of phthalocyanine supramolecular stacks into virus-like particles. *J. Am. Chem. Soc.* **2011**, 133, 6878-6881.

19. Kostiainen, M. A., Pietsch, C., Hoogenboom, R., Nolte, R. J. & Cornelissen, J. J. Temperature-Switchable Assembly of Supramolecular Virus–Polymer Complexes. *Adv. Funct. Mater.* **2011**, 21, 2012-2019.

20. van Eldijk, M. B. *et al.* Designing two self-assembly mechanisms into one viral capsid protein. *J. Am. Chem. Soc.* **2012**, 134, 18506-18509.

21. Gillitzer, E., Suci, P., Young, M. & Douglas, T. Controlled ligand display on a symmetrical protein-cage architecture through mixed assembly. *Small* **2006**, 2, 962-966.

22. Suci, P. A., Varpness, Z., Gillitzer, E., Douglas, T. & Young, M. Targeting and photodynamic killing of a microbial pathogen using protein cage architectures functionalized with a photosensitizer. *Langmuir* **2007**, 23, 12280-12286.

23. Porta, C. *et al.* Cowpea mosaic virus-based chimaeras: Effects of inserted peptides on the phenotype, host range, and transmissibility of the modified viruses. *Virology* **2003**, 310, 50-63.

24. Ochoa, W. F., Chatterji, A., Lin, T. & Johnson, J. E. Generation and structural analysis of reactive empty particles derived from an icosahedral virus. *Chem. Biol.***2006**, 13, 771-778.

25. Raja, K. S., Wang, Q. & Finn, M. Icosahedral virus particles as polyvalent carbohydrate display platforms. *ChemBioChem* **2003**, 4, 1348-1351.

26. Kaltgrad, E. *et al.* On-virus construction of polyvalent glycan ligands for cell-surface receptors. *J. Am. Chem. Soc.* **2008**,130, 4578-4579.

27. Chatterji, A. *et al.* Chemical conjugation of heterologous proteins on the surface of cowpea mosaic virus. *Bioconjugate Chem.* **2004**, 15, 807-813.

28. Aljabali, A. A., Shukla, S., Lomonossoff, G. P., Steinmetz, N. F. & Evans, D. J. CPMV-Dox Delivers. *Mol. Pharmaceutics* **2012**, 10, 3-10.
29. Yildiz, I., Lee, K. L., Chen, K., Shukla, S. & Steinmetz, N. F. Infusion of imaging and therapeutic molecules into the plant virus-based carrier cowpea mosaic virus: cargo-loading and delivery. *J. Controlled Release* **2013**, 172, 568-578.
30. Steinmetz, N. F. *et al.* Buckyballs meet viral nanoparticles: candidates for biomedicine. *J. Am. Chem. Soc.* **2009**, 131, 17093-17095.
31. Wen, A. M. *et al.* Interior engineering of a viral nanoparticle and its tumor homing properties. *Biomacromolecules* **2012**, 13, 3990-4001.
32. Brunel, F. M. *et al.* Hydrazone ligation strategy to assemble multifunctional viral nanoparticles for cell imaging and tumor targeting. *Nano Lett.* **2010**, 10, 1093-1097.
33. Pokorski, J. K., Hovlid, M. L. & Finn, M. Cell targeting with hybrid Qβ virus-like particles displaying epidermal growth factor. *ChemBioChem* **2011**, 12, 2441-2447.
34. Rhee, J.-K. *et al.* Glycan-targeted virus-like nanoparticles for photodynamic therapy. *Biomacromolecules* **2012**, 13, 2333-2338.
35. Purwar, M. *et al.* Design, display and immunogenicity of HIV1 gp120 fragment immunogens on virus-like particles. *Vaccine* **2018**, 36, 6345-6353.
36. Storni, T. *et al.* Nonmethylated CG motifs packaged into virus-like particles induce protective cytotoxic T cell responses in the absence of systemic side effects. *J. Immunol.* **2004**, 172, 1777-1785.
37. Pokorski, J. K., Breitenkamp, K., Liepold, L. O., Qazi, S. & Finn, M. Functional virus-based polymer–protein nanoparticles by atom transfer radical polymerization. *J. Am. Chem. Soc.* **2011,** 133, 9242-9245,
38. Lee, P. W. *et al.* Polymer Structure and Conformation Alter the Antigenicity of Virus-like Particle–Polymer Conjugates. *J. Am. Chem. Soc.* **2017**, 139, 3312-3315.
39. Endo, M., Wang, H., Fujitsuka, M. & Majima, T. Pyrene-Stacked Nanostructures Constructed in the Recombinant Tobacco Mosaic Virus Rod Scaffold. *Chem. - Eur. J.* **2006**, 12, 3735-3740.
40. Zhou, K., Ke, Y. & Wang, Q. Selective in situ assembly of viral protein onto DNA origami. *J. Am. Chem. Soc.* **2018**,
41. Finbloom, J. A. *et al.* Stable disk assemblies of a tobacco mosaic virus mutant as nanoscale scaffolds for applications in drug delivery. *Bioconjugate Chem.* **2016**, 27, 2480-2485.

42. Hou, C. *et al.* Construction of GPx active centers on natural protein nanodisk/nanotube: a new way to develop artificial nanoenzyme. *ACS nano* **2012**, 6, 8692-8701.

43. Cargile, B. J., McLuckey, S. A. & Stephenson, J. L. Identification of bacteriophage MS2 coat protein from E. coli lysates via ion trap collisional activation of intact protein ions. *Anal. Chem.***2001**, 73, 1277-1285.

44. Johnson, H. R., Hooker, J. M., Francis, M. B. & Clark, D. S. Solubilization and stabilization of bacteriophage MS2 in organic solvents. *Biotechnol. Bioeng.* **2007**, 97, 224-234.

45. Valegård, K., Liljas, L., Fridborg, K. & Unge, T. The three-dimensional structure of the bacterial virus MS2. *Nature* **1990**, 345, 36-41.

46. Wu, M., Sherwin, T., Brown, W. L. & Stockley, P. G. Delivery of antisense oligonucleotides to leukemia cells by RNA bacteriophage capsids. *Nanotechnol. Biol. Med.* **2005**, 1, 67-76.

47. Galaway, F. A. & Stockley, P. G. MS2 viruslike particles: a robust, semisynthetic targeted drug delivery platform. *Mol .Pharmaceutics* **2012**, 10, 59-68.

48. Pan, Y. *et al.* Development of a microRNA delivery system based on bacteriophage MS2 virus-like particles. *FEBS J.* **2012**, 279, 1198-1208.

49. Stephanopoulos, N., Tong, G. J., Hsiao, S. C. & Francis, M. B. Dual-surface modified virus capsids for targeted delivery of photodynamic agents to cancer cells. *ACS Nano* **2010**, 4, 6014-6020.

50. Kim, S. E., Jo, S. D., Kwon, K. C., Won, Y. Y. & Lee, J. Genetic Assembly of Double-Layered Fluorescent Protein Nanoparticles for Cancer Targeting and Imaging. *Adv. Sci.* **2017**, 4, 1600471.

51. Park, J.-S. *et al.* A highly sensitive and selective diagnostic assay based on virus nanoparticles. *Nat. Nanotechnol.***2009**, 4, 259-264.

52. Wörsdörfer, B., Woycechowsky, K. J. & Hilvert, D. Directed evolution of a protein container. *Science* **2011**, 331, 589-592.

53. Azuma, Y., Zschoche, R., Tinzl, M. & Hilvert, D. Quantitative packaging of active enzymes into a protein cage. *Angew. Chem. Int. Edn. Engl.* **2016**, 55, 1531-1534.

54. Sasaki, E. *et al.* Structure and assembly of scalable porous protein cages. *Nat. commun.* **2017**, 8, 14663.

55. Azuma, Y., Edwardson, T. G., Terasaka, N. & Hilvert, D. Modular Protein Cages for Size-Selective RNA Packaging in Vivo. *J. Am. Chem. Soc.* **2018**, 140**,** 566.

56. Stefanini, S. *et al.* Thermal stability of horse spleen apoferritin and human recombinant H apoferritin. *Arch. Biochem. Biophys.***1996**, 325, 58-64.

57. Bryce, C. F. & Crichton, R. R. The catalytic activity of horse spleen apoferritin. Preliminary kinetic studies and the effect of chemical modification. *Biochem. J* **1973**, 133, 301-309.

58. Kim, M. *et al.* pH-dependent structures of ferritin and apoferritin in solution: disassembly and reassembly. *Biomacromolecules* **2011**, 12, 1629-1640.

59. Aime, S., Frullano, L. & Geninatti Crich, S. Compartmentalization of a gadolinium complex in the apoferritin cavity: a route to obtain high relaxivity contrast agents for magnetic resonance imaging. *Angew. Chem. Int. Edn. Engl.* **2002**, 41, 1017-1019.

60. Yang, Z. *et al.* Encapsulation of platinum anticancer drugs by apoferritin. *Chem. Commun.* **2007**, 3453-3455.

61. Liu, X. *et al.* Bio-inspired protein–gold nanoconstruct with core–void–shell structure: beyond a chemo drug carrier. *J. Mater. Chem.* **2013**, 1, 3136-3143.

62. Ma-Ham, A. *et al.* Apoferritin-based nanomedicine platform for drug delivery: equilibrium binding study of daunomycin with DNA. *J. Mater. Chem.* **2011**, 21, 8700-8708.

63. Luo, Y., Wang, X., Du, D. & Lin, Y. Hyaluronic acid-conjugated apoferritin nanocages for lung cancer targeted drug delivery. *Biomater. Sci.* **2015**, 3, 1386-1394.

64. Zhen, Z. *et al.* RGD-modified apoferritin nanoparticles for efficient drug delivery to tumors. *ACS Nano* **2013**, **7**, 4830-4837.

65. Blazkova, I. *et al.* Apoferritin modified magnetic particles as doxorubicin carriers for anticancer drug delivery. *Int. J. Mol. Sci.* **2013**, 14, 13391-13402.

66. Liang, M. *et al.* H-ferritin–nanocaged doxorubicin nanoparticles specifically target and kill tumors with a single-dose injection. *Proc. Natl Acad. Sci. USA* **2014**, 111, 14900-14905.

67. Bellini, M. *et al.* Protein nanocages for self-triggered nuclear delivery of DNA-targeted chemotherapeutics in Cancer Cells. *J. Controlled Release* **2014**,196, 184-196.

68. Kang, H. J. *et al.* Developing an antibody-binding protein cage as a molecular recognition drug modular nanoplatform. *Biomaterials* **2012**, 33, 5423-5430.

69. Jeon, J. O. *et al.* Designed nanocage displaying ligand-specific Peptide bunches for high affinity and biological activity. *ACS Nano* **2013**, 7, 7462-7471.

70. Lee, E. J. *et al.* Engineered proteinticles for targeted delivery of siRNA to cancer cells. *Adv. Funct. Mater.* **2015**, 25, 1279-1286.

71. Bailey, J. B., Zhang, L., Chiong, J. A., Ahn, S. & Tezcan, F. A. Synthetic Modularity of Protein–Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2017**, 139, 8160−8166.

72. Sun, Y. & MacRae, T. H. Small heat shock proteins: molecular structure and chaperone function. *Cell. Mol. Life Sci.* **2005**, 62, 2460-2476.

73. Kim, K. K., Kim, R. & Kim, S.-H. Crystal structure of a small heat-shock protein. *Nature* **1998**, 394, 595-599.

74. Bova, M. P., Huang, Q., Ding, L. & Horwitz, J. Subunit Exchange, Conformational Stability, and Chaperone-like Function of the Small Heat Shock Protein 16.5 fromMethanococcus jannaschii. *J. Biol. Chem.***2002**, 277, 38468-38475.

75. Flenniken, M. L., Willits, D. A., Brumfield, S., Young, M. J. & Douglas, T. The small heat shock protein cage from Methanococcus jannaschii is a versatile nanoscale platform for genetic and chemical modification. *Nano Lett.* **2003**, 3, 1573-1576.

76. Toita, R. *et al.* A nanocarrier based on a genetically engineered protein cage to deliver doxorubicin to human hepatocellular carcinoma cells. *Chem. Commun.* **2013**, 49, 7442-7444.

77. Stephen, A. G. *et al.* Assembly of vault-like particles in insect cells expressing only the major vault protein. *J. Biol. Chem.* **2001**, 276, 23217-23220.

78. Goldsmith, L. E., Yu, M., Rome, L. H. & Monbouquette, H. G. Vault nanocapsule dissociation into halves triggered at low pH. *Biochemistry* **2007**, 46, 2865-2875.

79. Esfandiary, R., Kickhoefer, V. A., Rome, L. H., Joshi, S. B. & Middaugh, C. R. Structural stability of vault particles. *J. Pharm. Sci.* **2009**, 98, 1376-1386.

80. Matsumoto, N. M., Prabhakaran, P., Rome, L. H. & Maynard, H. D. Smart vaults: thermally-responsive protein nanocapsules. *ACS Nano* **2013**, 7, 867-874.

81. Smith, M. L. *et al.* Modified tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. *Virology* **2006**, 348, 475-488.

82. Ma, Y., Nolte, R. J. & Cornelissen, J. J. Virus-based nanocarriers for drug delivery. *Adv. Drug Delivery Rev.* **2012**, 64, 811-825.

83. Yan, F., Zhang, Y., Yuan, H.-k., Gregas, M. K. & Vo-Dinh, T. Apoferritin protein cages: a novel drug nanocarrier for photodynamic therapy. *Chem. Commun.* **2008**, 4579-4581.

84. Lam, P. & Steinmetz, N. F. Delivery of siRNA therapeutics using cowpea chlorotic mottle virus-like particles. *Biomater. Sci.* **2019**, 7, 3138−3142.

85. Lin, X. *et al.* Chimeric ferritin nanocages for multiple function loading and multimodal imaging. *Nano Lett.* **2011**, 11, 814-819.

86. Pieters, B. J., van Eldijk, M. B., Nolte, R. J. & Mecinović, J. Natural supramolecular protein assemblies. *Chem. Soc. Rev.* **2016**, 45, 24-39.

87. Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl Acad. Sci. USA* **2001**, 98, 2217-2221.

88. Kobayashi, N. *et al.* Self-assembling nano-architectures created from a protein nano-building block using an intermolecularly folded dimeric de novo protein. *J. Am. Chem. Soc.* **2015**, 137, 11285-11293.

89. Sciore, A. *et al.* Flexible, symmetry-directed approach to assembling protein cages. *Proc. Natl Acad. Sci. USA* **2016**, 113, 8681−8686.

90. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **2014**, 510, 103-108.

91. Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein icosahedron. *Nature* **2016**, 535, 136-139.

92. Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **2016**, 353, 389-394.

93. Marcandalli, J. *et al.* Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. *Cell* **2019**, 176, 1420-1431. e1417.

94. Brouwer, P. J. *et al.* Enhancing and shaping the immunogenicity of native-like HIV-1 envelope trimers with a two-component protein nanoparticle. *Nat. Commun.* **2019**, 10, 1-17.

95. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **2012**, 336, 1171-1174.

96. Edwardson, T. G., Mori, T. & Hilvert, D. Rational Engineering of a Designed Protein Cage for siRNA Delivery. *J. Am. Chem. Soc.* **2018**, 140, 10439-10442.

97. Lai, Y.-T., Cascio, D. & Yeates, T. O. Structure of a 16-nm cage designed by using protein oligomers. *Science* **2012**, 336, 1129-1129.

98. Lai, Y.-T., Tsai, K.-L., Sawaya, M. R., Asturias, F. J. & Yeates, T. O. Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. *J. Am. Chem. Soc.* **2013**,135, 7738-7743.

99. Lai, Y.-T. *et al.* Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem.* **2014**, 6, 1065-1071.

100. Kawakami, N. *et al.* Design of hollow protein nanoparticles with modifiable interior and exterior surfaces. *Angew. Chem. Int. Edn. Engl.* **2018**, 57, 12400-12404.

101. Sugimoto, K., Kanamaru, S., Iwasaki, K., Arisaka, F. & Yamashita, I. Construction of a ball-and-spike protein supramolecule. *Angew. Chem. Int. Edn. Engl.* **2006**, 45, 2725-2728.

102. Lupas, A. N. & Gruber, M. in *Advances in protein chemistry* Vol. 70 37-38 (Elsevier, 2005).

103. Badieyan, S. *et al.* Symmetry-directed Self-assembly of a Tetrahedral Protein Cage Mediated by De Novo-designed Coiled Coils. *ChemBioChem* **2017**, 18, 1888–1892.

104. Cristie-David, A. S. & Marsh, E. N. G. Metal-dependent assembly of a protein nano-cage. *Biopolymers* **2019**, 28, 1620-1629.

105. Gradišar, H. *et al.* Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat. Chem. Biol.* **2013**, 9, 362−366.

106. Ljubetič, A. *et al.* Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. *Nat. Biotechnol.* **2017**, 11, 1094−1101.

107. Park, W. M. & Champion, J. A. Thermally triggered self-assembly of folded proteins into vesicles. *J. Am. Chem. Soc.* **2014**, 136, 17906-17909.

108. Jang, Y. *et al.* Engineering globular protein vesicles through tunable self-assembly of recombinant fusion proteins. *Small* **2017**, 13, 1700399.

109. Assal, Y., Mizuguchi, Y., Mie, M. & Kobatake, E. Growth factor tethering to protein nanoparticles via coiled-coil formation for targeted drug delivery. *Bioconjugate chem.* **2015**, 26, 1672-1677.

110. Fazelinia, H. *et al.* Genetically Engineered Elastomeric Polymer Network through Protein Zipper Assembly. *ChemistrySelect* **2017**, 2, 5008-5012.

111. Lim, S. I., Lukianov, C. I. & Champion, J. A. Self-assembled protein nanocarrier for intracellular delivery of antibody. *J. Controlled Release* **2017**, 249, 1-10,

112. Oohora, K., Hirayama, S., Uchihashi, T. & Hayashi, T. Construction of a Hexameric Hemoprotein Sheet and Direct Observation of Dynamic Process of its Formation. *Chem. Lett.* **2020**, 49, 186-190.

113. Oohora, K., Hirayama, S., Mashima, T. & Hayashi, T. Supramolecular dimerization of a hexameric hemoprotein via multiple pyrene–pyrene interactions. *J. Porphyrins Phthalocyanines* **2019**, 1-9.

114. Hirayama, S., Oohora, K., Uchihashi, T. & Hayashi, T. A Thermoresponsive Micellar Assembly Constructed from a Hexameric Hemoprotein Modified with Poly (N-isopropylacrylamide) toward an Artificial Light-harvesting System. *J. Am. Chem. Soc.* **2020**, 142, 1822–1831.

115. Dreher, M. R. *et al.* Temperature triggered self-assembly of polypeptides into multivalent spherical micelles. *J. Am. Chem. Soc.* **2008**, 130, 687-694.

116. Hyun, J., Lee, W.-K., Nath, N., Chilkoti, A. & Zauscher, S. Capture and release of proteins on the nanoscale by stimuli-responsive elastin-like polypeptide "switches". *J. Am. Chem. Soc.* **2004**, 126, 7330-7335.

117. Hassouneh, W. *et al.* Unexpected multivalent display of proteins by temperature triggered self-assembly of elastin-like polypeptide block copolymers. *Biomacromolecules* **2012**, 13, 1598-1605.

118. Qin, G., Perez, P. M., Mills, C. E. & Olsen, B. D. Effect of ELP sequence and fusion protein design on concentrated solution self-assembly. *Biomacromolecules* **2016**, 17, 928-934.

119. Kitagishi, H. *et al.* Supramolecular Hemoprotein Linear Assembly by Successive Interprotein Heme- Heme Pocket Interactions. *Journal J. Am. Chem. Soc.* **2007**, 129, 10326-10327.

120. Kitagishi, H. *et al.* Self-Assembly of One-and Two-Dimensional Hemoprotein Systems by Polymerization through Heme–Heme Pocket Interactions. *Angew. Chem. Int. Edn. Engl.* **2009**, 48, 1271-1274.

121. Oohora, K. *et al.* A chemically-controlled supramolecular protein polymer formed by a myoglobin-based self-assembly system. *Chem. Sci.* **2011**, 2, 1033-1038.

122. Oohora, K. *et al.* Chemically programmed supramolecular assembly of hemoprotein and streptavidin with alternating alignment. *Angew. Chem. Int. Edn. Engl.* **2012**, 51, 3818-3821.

123. Oohora, K., Kajihara, R., Fujimaki, N., Uchihashi, T. & Hayashi, T. A ring-shaped hemoprotein trimer thermodynamically controlled by the supramolecular heme–heme pocket interaction. *Chem. Commun.* **2019**, 55, 1544-1547.

124. Oohora, K., Onoda, A. & Hayashi, T. Hemoproteins Reconstituted with Artificial Metal Complexes as Biohybrid Catalysts. *Acc. Chem. Res.* **2019,** 52, 945-954.

125. Carlson, J. C. *et al.* Chemically controlled self-assembly of protein nanorings. *J. Am. Chem. Soc.* **2006**,128, 7630-7638.

126. Modica, J. A., Lin, Y. & Mrksich, M. Synthesis of Cyclic Megamolecules. *J. Am. Chem. Soc.* **2018**, 140, 6391−6399.

127. Kim, Y. E., Kim, Y.-n., Kim, J. A., Kim, H. M. & Jung, Y. Green fluorescent protein nanopolygons as monodisperse supramolecular assemblies of functional proteins with defined valency. *Nat. Commun.* **2015**, 6, 7134−7142.

128. Klass, S. H. *et al.* Self-Assembling Micelles Based on an Intrinsically Disordered Protein Domain. *J. Am. Chem. Soc.* **2015**, 6, 7134−7142.

129. Velonia, K., Rowan, A. E. & Nolte, R. J. Lipase polystyrene giant amphiphiles. *J. Am. Chem. Soc.* **2002**, 124, 4224-4225.

130. Dirks, A. T. *et al.* Preparation of biohybrid amphiphiles via the copper catalysed Huisgen [3+ 2] dipolar cycloaddition reaction. *Chem. Commun.* **2005**, 4172-4174.

131. Hannink, J. M. *et al.* Protein–polymer hybrid amphiphiles. *Angew. Chem. Int. Edn. Engl.* **2001**, 40, 4732-4734.

132. Boerakker, M. J. *et al.* Giant amphiphiles by cofactor reconstitution. *Angew. Chem. Int. Edn. Engl.* **2002**, 41, 4239-4241.

133. Boerakker, M. J. *et al.* Aggregation behavior of giant amphiphiles prepared by cofactor reconstitution. *Chem. - Eur. J.* **2006**, 12, 6071-6080.

134. Reynhout, I. C., Cornelissen, J. J. & Nolte, R. J. Self-assembled architectures from biohybrid triblock copolymers. *J. Am. Chem. Soc.* **2007**, 129, 2327-2332.

135. Thomas, C. S., Glassman, M. J. & Olsen, B. D. Solid-state nanostructured materials from self-assembly of a globular protein–polymer diblock copolymer. *ACS Nano* **2011**, 5, 5697-5707.

136. Thomas, C. S., Xu, L. & Olsen, B. D. Kinetically Controlled nanostructure formation in self-assembled globular protein–polymer diblock copolymers. *Biomacromolecules* **2012**, 13, 2781-2792.

137. Thomas, C. S., Xu, L. & Olsen, B. D. Effect of Small Molecule Osmolytes on the Self-Assembly and Functionality of Globular Protein–Polymer Diblock Copolymers. *Biomacromolecules* **2013**, 14, 3064-3072.

138. Chang, D., Lam, C. N., Tang, S. & Olsen, B. D. Effect of polymer chemistry on globular protein–polymer block copolymer self-assembly. *Biopolymers* **2014**, 5, 4884-4895.

139. Lam, C. N. *et al.* The Nature of Protein Interactions Governing Globular Protein–Polymer Block Copolymer Self-Assembly. *Biomacromolecules* **2014**, 15, 1248-1258.

140. Huang, A., Qin, G. & Olsen, B. D. Highly Active Biocatalytic Coatings from Protein–Polymer Diblock Copolymers. *ACS Appl. Mater. Interfaces* **2015**, 7, 14660-14669.

141. Hou, Y., Yuan, J., Zhou, Y., Yu, J. & Lu, H. A Concise Approach to Site-Specific Topological Protein–Poly (amino acid) Conjugates Enabled by in Situ-Generated Functionalities. *J. Am. Chem. Soc.* **2016**, 138, 10995-11000.

142. Ju, Y. *et al.* Covalently Connected Polymer–Protein Nanostructures Fabricated by a Reactive Self-Assembly Approach. *Chem. - Eur. J.* **2017**, 23, 3366-3374.

143. Bontempo, D. & Maynard, H. D. Streptavidin as a macroinitiator for polymerization: in situ protein- polymer conjugate formation. *J. Am. Chem. Soc.* **2005**, 127, 6508-6509.

144. Heredia, K. L. *et al.* In situ preparation of protein-"smart" polymer conjugates with retention of bioactivity. *J. Am. Chem. Soc.* **2005**, 127, 16955-16960.

145. Liu, X. & Gao, W. In Situ Growth of Self-Assembled Protein–Polymer Nanovesicles for Enhanced Intracellular Protein Delivery. *ACS Appl. Mater. Interfaces* **2017**, 9, 2023-2028.

146. Ma, C. *et al.* Efficient Way to Generate Protein-Based Nanoparticles by in-Situ Photoinitiated Polymerization-Induced Self-Assembly. *ACS Macro Lett.* **2017**, 6, 689-694.

147. Wan, X. & Liu, S. Fabrication of a thermoresponsive biohybrid double hydrophilic block copolymer by a cofactor reconstitution approach. *Macromol. Rapid Commun.* **2010**, 31, 2070-2076.

148. Xia, Y., Tang, S. & Olsen, B. D. Site-specific conjugation of RAFT polymers to proteins via expressed protein ligation. *Chem. Commun.* **2013**, 49, 2566-2568.

149. Moatsou, D. *et al.* Self-assembly of temperature-responsive protein–polymer bioconjugates. *Bioconjugate Chem.* **2015**, 26, 1890-1899.

150. Averick, S. *et al.* Cooperative, reversible self-assembly of covalently pre-linked proteins into giant fibrous structures. *Angew. Chem. Int. Edn. Engl.* **2014**, 53, 8050-8055.

151. Suma, T. *et al.* Generalizable Strategy for Engineering Protein Particles with pH-Triggered Disassembly and Recoverable Protein Functionality. *ACS Macro Lett.* **2015**, 4, 160-164.

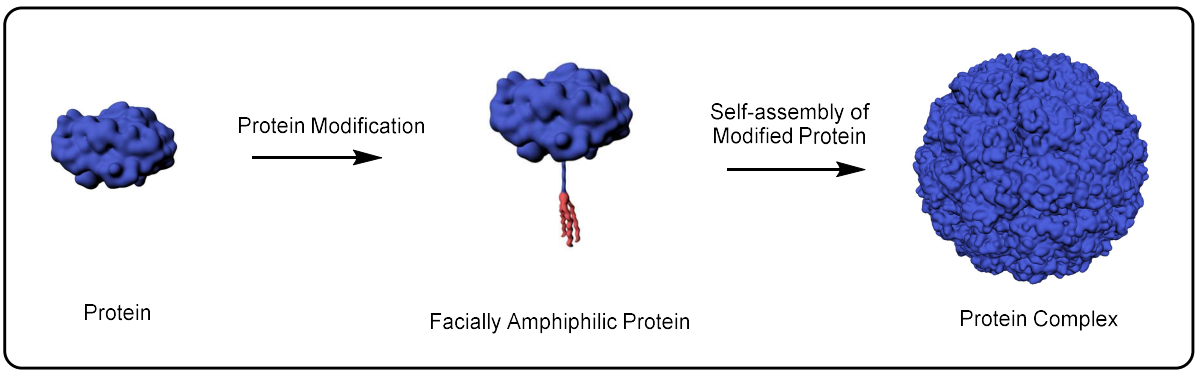
152. Huang, X. *et al.* Interfacial assembly of protein-polymer nano-conjugates into stimulus-responsive biomimetic protocells. *Nat. Commun.* **2013**, 4, 2239.

153. Han, G., Wang, J.-T., Ji, X., Liu, L. & Zhao, H. Nanoscale Proteinosomes Fabricated by Self-Assembly of a Supra-molecular Protein-Polymer Conjugate. *Bioconjugate Chem.* **2017**, 28, 636-641.

Chapter 2

Facially Amphiphilic Proteins (FAPs) and Micelle Assisted Protein Labelling (MAPLab) Technology Development

2.1. Synthesis and Self-assembly Studies of Prototype FAP



2.1.1. Design

Design and construction of protein super-structures has gained enormous interest in the recent years[1-5] which is mainly done by genetic engineering[6-8] and chemical strategies.[9-12] Chemical strategies are still in infancy. They are restricted to a few proteins and often yield polydisperse samples. To address these drawbacks, we wanted to develop a methodology to build protein super-structures/complex through the bottom-up approach in precise fashion from monomeric protein, which can be applied to the majority of the proteins (Fig. 2.1).

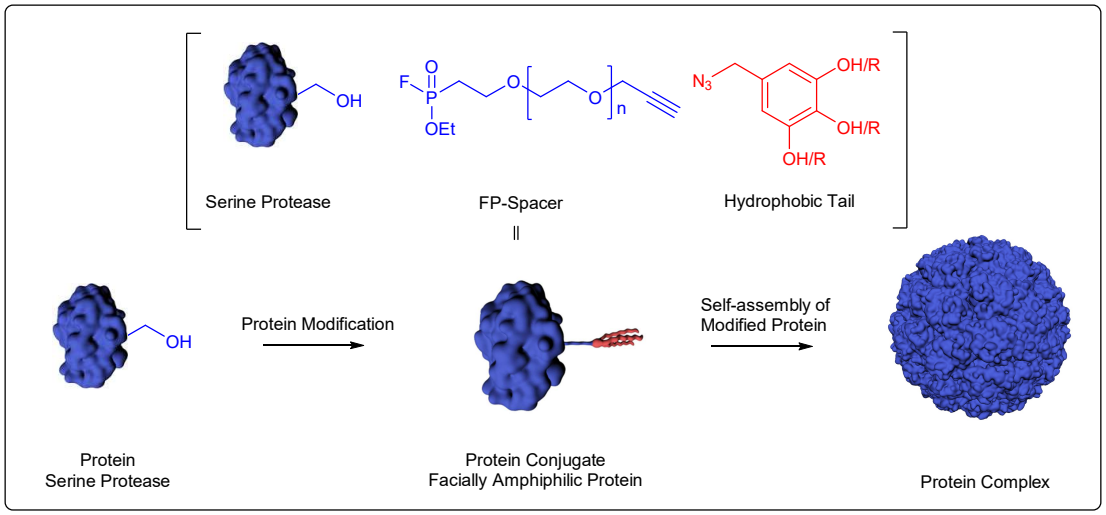


Figure 2.1. Process of converting monomeric protein into a complex.

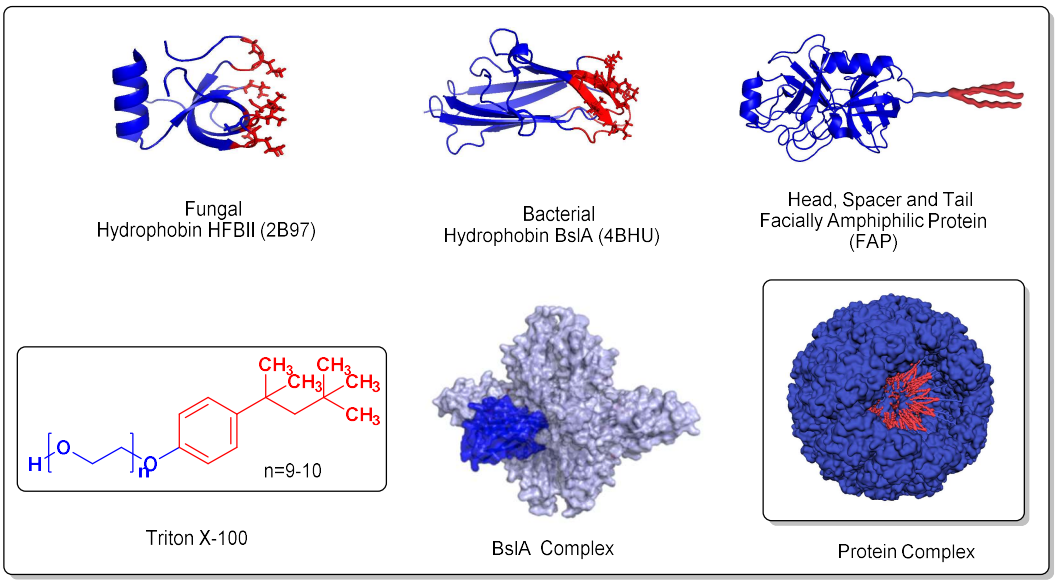


Figure 2.2. Crystal structure of fungal and bacterial hydrophobins (PDB accession: 2B97 and 4BHU) and designed FAP consists of globular water-soluble protein and hydrophobic tail

connected by the hydrophilic linker (hydrophilic and hydrophobic faces are shown in blue and red, respectively).

The inspiration for our design came from structural, biophysical and self-assembly studies of naturally occurring facially amphiphilic hydrophobins[13] (Fig. 2.2), in particular, a recent report of the high-resolution decameric structure of a bacterial hydrophobin[14] where the hydrophobic face of the protein is shielded away from the water molecules by forming a dense hydrophobic core and hydrophilic face is presented outside, which mimics protein cages[15].

Hydrophobins are 8-14 kDa proteins, in which most of the hydrophobic amino acids are presented on one face, making them facially amphiphilic. They are produced in fungi and play a vital role in the life cycle of filamentous fungi, during the formation of aerial hyphae, during the attachment of spores, germination, and infection of spores and prevent the recognition by host defense systems. A recent study revealed that hydrophobins are also present in bacteria, plays an important role in the survival of the same. For example, *Bacillus subtilis* biofilms are resistant to liquid wetting and gas penetration and show a broad-spectrum resistance to antimicrobial agents. In a study by kazuo et al., BslA was identified as a major contributor to the surface repellency of *B. subtilis* biofilms.[16]

There are two classes of fungal hydrophobins, divided solely based on their sequence and distribution of cysteines, hydrophilic and hydrophobic residues, and this classification distinguishes their physicochemical properties. Even though both the classes have four disulfide bonds that are formed between the eight cysteines, they have different inter cysteine spacing with the class II having short and regular spacing. All the hydrophobin structures contain a β-barrel consisting of four antiparallel β-strands. Class II hydrophobins like HFBI and HFBII exhibit a closed barrel core structure, whereas the class I hydrophobins like DewA and EAS have a relatively open or half-barrel core. In class II, the barrel is associated with a loop and a short α-helix, whereas the class I have very different associations. In addition, the class II hydrophobins are closely related to each other when compared with class I. Class I hydrophobins form a monolayer with fibrillar substructures known as rodlets, which can only be depolarized using strong acids whereas class II assemblies generally lack clear structural elements, which can be disrupted by using detergents and hot alcohols. Just encountering interface is sufficient to trigger the polymerization of these hydrophobins. During the

formation of rodlets of class I hydrophobins, the formation of β-sheet and hydrogen bonding is involved, whereas in the case of class II, except for the increase in helical content, there are no significant conformational changes.

Hydrophobins are extremely useful and are used as fabricators for nanodevices and medical implants, as emulsifiers in food and personal care industries, as biocompatibility enhancers for hydrophobic surfaces and as a generic intermediate layer for secondary protein immobilization. They are also used in drug delivery and as cell attachment enhancers.[13] These studies on hydrophobins gave us a clue for the construction of protein complexes. We envisioned that by converting protein into facially amphiphilic proteins (FAPs), a monomeric protein could be converted into a protein complex. The macromolecular design of FAP has three core structural elements: (i) hydrophilic globular protein, (ii) flexible hydrophilic linker, and (iii) hydrophobic tail (Fig. 2.2).

We envisioned that the designed FAP would self-assemble into 3D complex (Fig. 2.2), primarily driven by strong hydrophobic interaction, which is one of the major driving forces for the formation of supramolecular protein assemblies in natural[17] and artificial systems[18]. The basic tenet of our hypothesis is that the attributes of a protein complex (size, oligomeric state, surface charge, and shape) would be dictated by the choice of a core structural element of macromolecular design.[19] We have designed this scaffold in a modular fashion that should allow us to study the effects of head, spacer, and tail on self-assembly systematically. Since the FAP has three components, we need two site-specific chemistries to make the scaffold. Choosing the right site-specific method is a bottleneck in chemical protein engineering strategies. So, even though we have plenty of proteins, amino acids, and chemistries to choose from, we cannot choose random proteins. We choose serine proteases, Flouorophonate (FP) chemistry, and click chemistry.

Hydrophilic globular protein: Serine proteases

For the sake of systematically studying the facially amphiphilic proteins (FAPs) / conjugates, we have made a brilliant decision to choose serine proteases as our target proteins. Almost one-third of all proteases can be classified as serine proteases, and it is the largest family in humans; the method can be extended to as many as 200 proteins (Fig. 2.4).[20,21] To modify the serine proteases, we chose activity-based probes[22] (FP chemistry), which targets the active

site of these proteins with extremely fast kinetics. The site-specificity of this FP group toward the active site Ser comes from the catalytic triad of the active site (Asp, Ser, His) (Fig. 2.4).

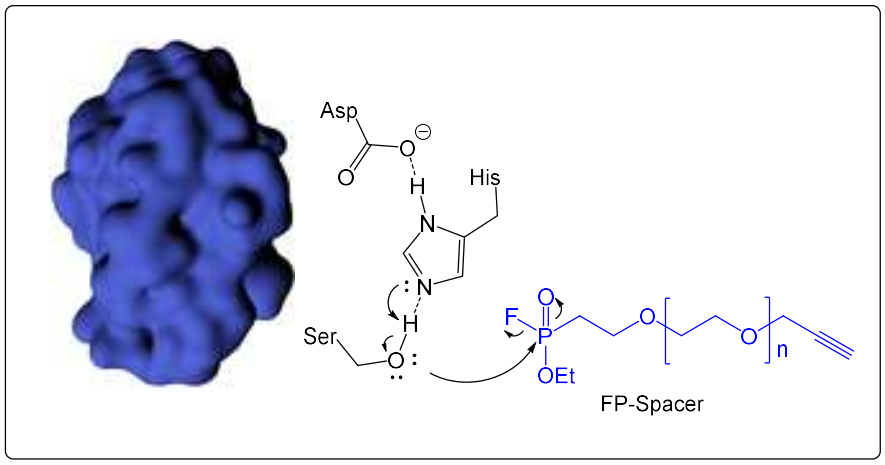


Figure 2.3. The catalytic triad of a serine protease and its reacting mechanism with the FP probe.

Flexible hydrophilic linker: Monodisperse oligoethylene glycol

Monodisperse oligoethylene glycol (defined repeating units) is used as spacer moiety because of its water solubility and protein repellent properties, and most importantly, spacer length could be systematically engineered to generate a unique library of protein conjugates (Fig. 2.4).

Hydrophobic tail

Finally, to drive the self-assembly, we choose hydrophobic tail; hydrophobicity is the universal driving force for self-assembly. Hydrophobicity of a tail can be systematically varied by modifying tail length/branching (Fig. 2.4).[23] We decided to use click chemistry to join the spacer and the hydrophobic tail (Fig. 2.4). We this basic hypothesis, we started making prototype FAP. During this process, we have optimized MALDI-TOF MS, synthesis, conjugation, purification, characterization, and did self-assembly studies. These steps are presented in detail in this sub-chapter.

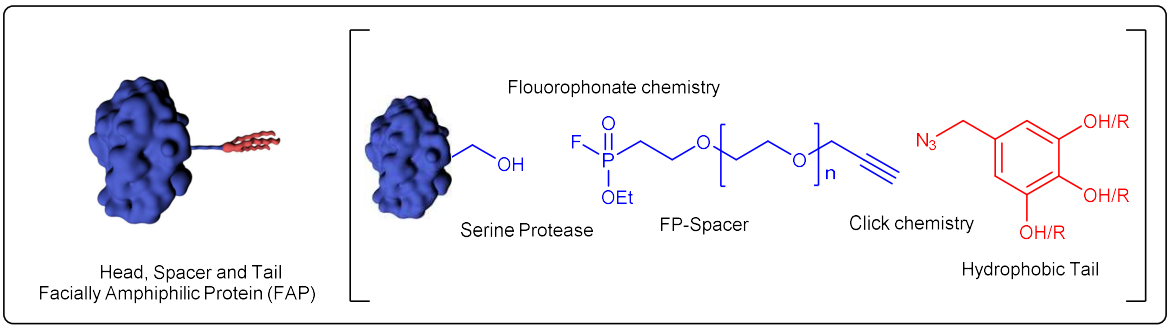


Figure 2.4. Process of converting monomeric protein into a complex.

2.1.2. MALDI-TOF MS optimization

This is the first and foremost requirement for the project. Molecular weight determination prior to modification is important to track the molecular weight changes accurately during conjugation. MALDI-TOF MS was optimized. We chose trypsin (Try) (23 kDa), *i.e.*, serine proteins for making prototype FAP. Try was purchased from a commercial vendor. The molecular weight of native Try was determined as per the optimized protocol (Fig. 2.5). The detailed procedure is mentioned in the methods section of this chapter.

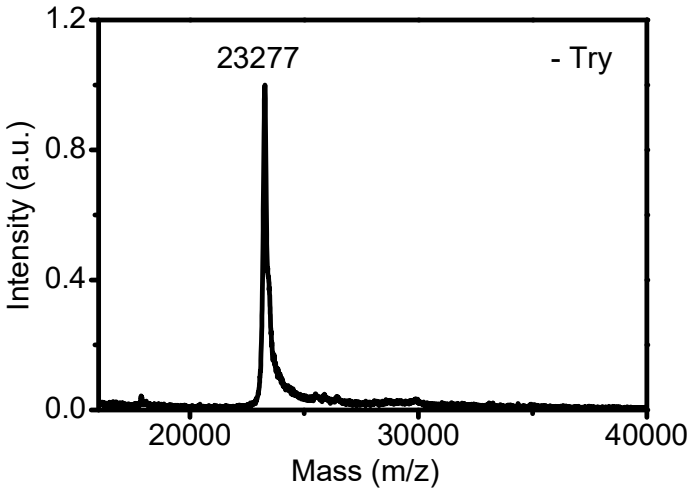


Figure 2.5. The molecular weight of native Try with the optimized protocol.

Optimized MALDI-TOF MS procedure

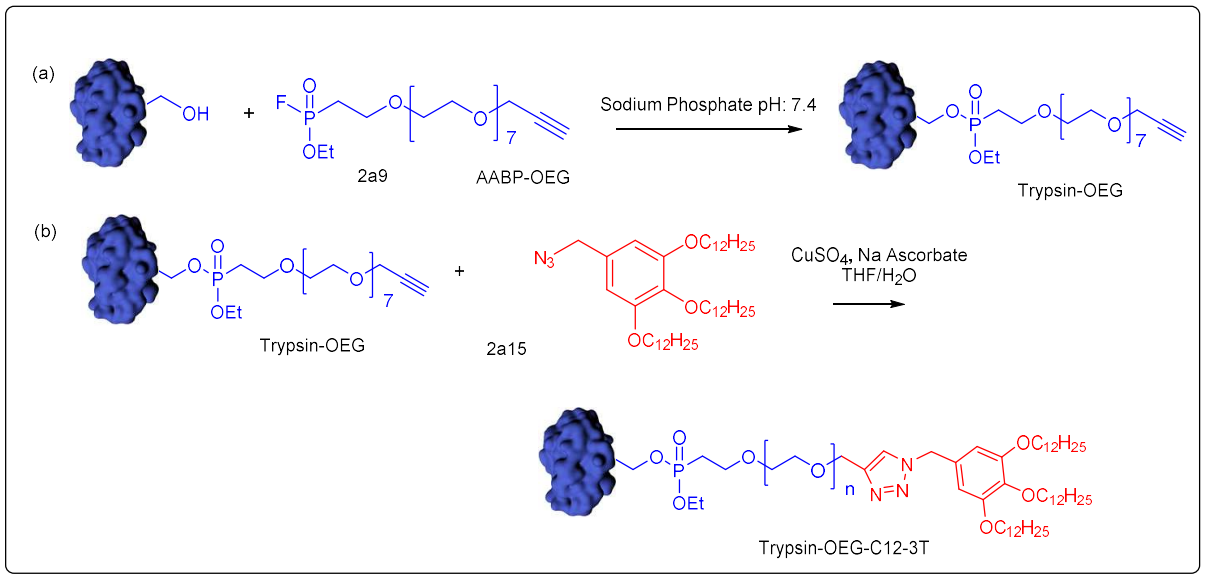
The samples were analyzed in Linear High Mass mode in AB Sciex 4800 plus MALDI-TOF/TOF analyzer with 4000 Series Explorer as software. Mass was scanned between 10,000 Da and 40,000 Da with a focus mass at 25,000-29,000 Da, depending on the protein

analyzed. The matrix used is a combination of Dihydroxy acetophenone (DHAP) and Diammonium hydrogen citrate (DAHC) in ethanol and milli-Q water. The sample preparation was carried out in a 1:1:1 ratio of a protein sample, 2 % trifluoroacetic acid (TFA), and matrix mixture. 100 µM protein concentration was found to be optimum for MALDI-TOF MS analysis. Detailed matrix mixture preparation is given at the end of this chapter.

2.1.3. Optimization of synthesis and conjugation

The FAP design seems simple, but to make it, we need to bring the hydrophobic group into the aqueous medium without denaturing the protein. This scaffold can be made in two ways. In the first approach, the spacer will be added to the protein, and then the installation of hydrophobic tails will be done. This is an extremely difficult approach. The second approach, however, is relatively easier in which we planned to add spacer and tail first, which results in an amphiphile, then this amphiphile will be added to the protein. With these approaches in our mind, we have attempted making this prototype FAP in the following ways.

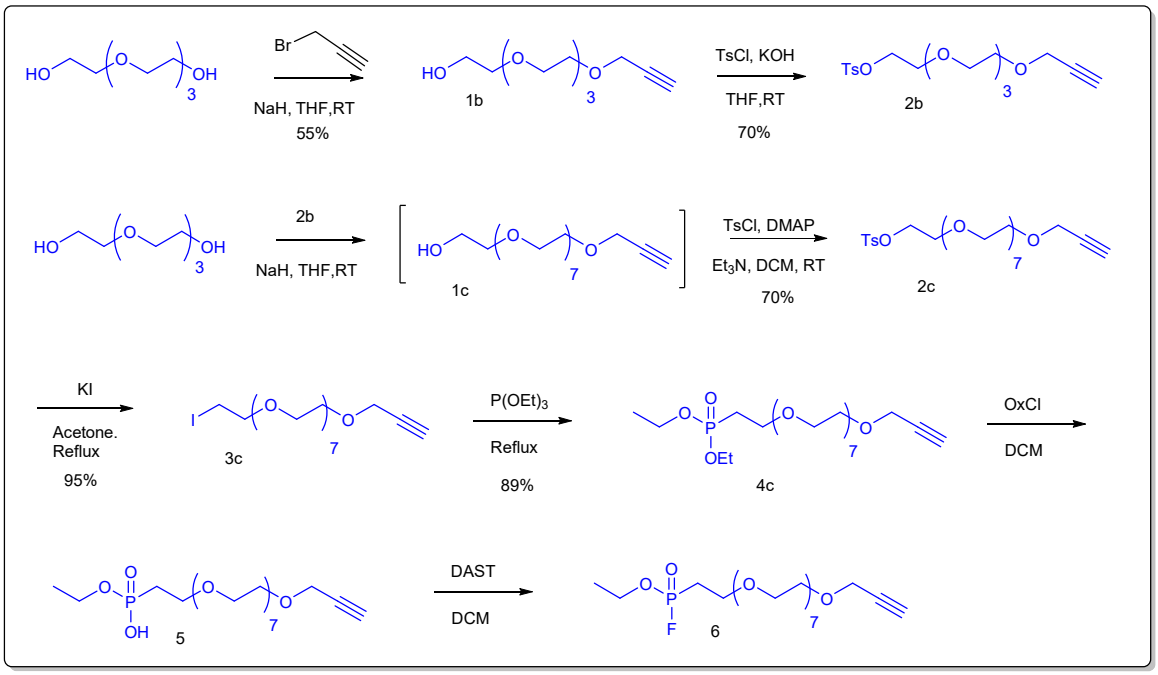
The approach i: (synthesis of spacer and tail, step by step conjugation)



Scheme 2.1. Proposed step by step conjugation to make FAP as per approach i.

The first approach is a step by step approach where the hydrophilic spacer will be added to protein followed by addition of hydrophobic tail (scheme 2.1). The spacer we choose is octaethylene glycol (OEG), and the tail is C12-3T. The spacer and tail molecules were made separately, as shown below.

Synthesis of the hydrophilic spacer



Scheme 2.2. Synthesis of the octaethylene glycol hydrophilic spacer.

In an oven-dried RBF, tetraethylene glycol (TEG) (1.0 eq) was dissolved with stirring in Tetrahydrofuran (THF). Sodium hydride (NaH) (1.0 eq) was added to the flask in small portions at 0 ºC. After 1 hr, propargyl bromide (ProBr) (0.7 eq) was added dropwise, maintaining the reaction at the same temperature. Then, the reaction was stirred for 12 hrs at RT. Upon completion, excess NaH was quenched with dropwise addition of water. The resulting reaction mixture was extracted thrice with Dichloromethane (DCM). The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified by Normal Phase chromatography (NPC) using methanol (MeOH) / dichloromethane (DCM) as eluent. In an oven-dried RBF, monopropargylate (**1b)** (1.0 eq) was dissolved with stirring in THF at 0 ºC. Aqueous potassium hydroxide (KOH) (4 eq) was added to the flask in small portions immediately. After 10 mins, tosyl chloride (TsCl) (1.2 eq) solution in THF was added dropwise and stirred for 12 hrs. Upon completion, the reaction was quenched with aqueous ammonium chloride and extracted with DCM thrice to get the crude product (**2b),** which was purified by NPC using MeOH / DCM as eluent.

In an oven-dried RBF, a mixture of TEG (1.0 eq) and **2b** (0.5 eq) was taken and dissolved in THF under stirring. The reaction mixture was cooled to 0 ºC and NaH (1.0 eq) was added in

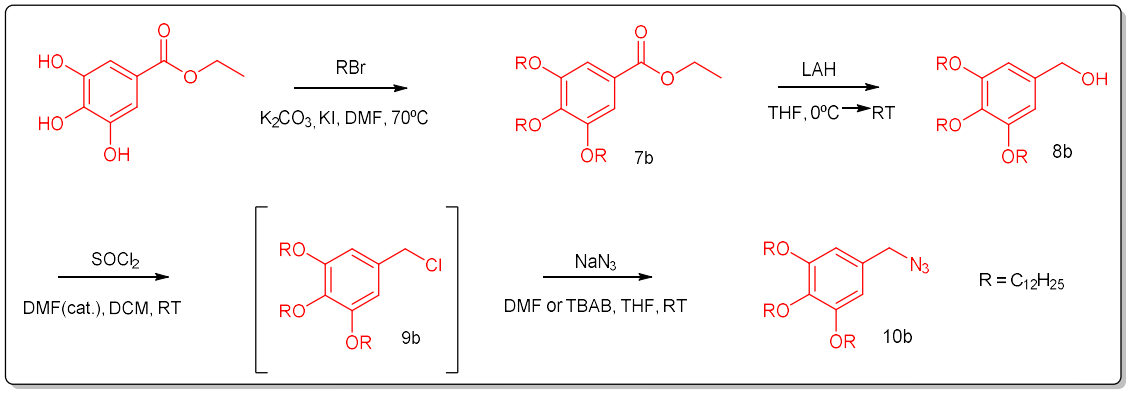
small portions. The resultant mixture was stirred at RT for 12 hrs. Upon completion of the reaction, excess of NaH was quenched by the dropwise addition of water. The resultant mixture was concentrated under reduced pressure. To the obtained residue, water was added and washed with EtOAc thoroughly. Now, the aqueous layer was concentrated under reduced pressure, and to the obtained residue, fresh EtOAc was added and filtered to remove salts. Finally, the collected EtOAc layer was dried over Na_2SO_4 and concentrated under reduced pressure to get the crude product, which was purified by NPC using MeOH / DCM as eluent. In an oven-dried RBF, monopropargylate (**1c)** (1.0 eq) was dissolved with stirring in THF at 0 ºC. Aqueous KOH (4 eq) was added to the flask in small portions immediately. After 10 mins, TsCl (1.2 eq) solution in THF was added dropwise and stirred for 12 hrs. Upon completion, the reaction was quenched with aqueous ammonium chloride (NH_4Cl) and extracted with DCM thrice to get the crude product, which was purified by NPC using MeOH / DCM as eluent.

A mixture of above-obtained tosylate (**2c**) (1.0 eq) and Potassium Iodide (KI) (4.0 eq) was refluxed in acetone for 18 h. Upon completion, excess KI was filtered and washed thrice with acetone. Collected acetone fraction was evaporated under vacuum to get residue, which was then washed with water and extracted with DCM. The combined organic layer was washed with aqueous Na_2CO_3 and then concentrated under vacuum to get the crude product, which was purified by NPC using MeOH / DCM as eluent. In an oven-dried RBF, iodide (**3c**) (1.0 eq) and triethylphosphite ($P(OEt)_3$) (4.0 eq) were taken and refluxed for 1 h. Upon completion of the reaction, the excess $P(OEt)_3$ was removed under vacuum, and the reaction mixture was directly loaded onto a silica gel column, and the crude mixture was purified using MeOH / DCM as eluent.

Diphosphonate ester (**4c**) (1 eq) was dissolved in DCM with stirring. Then, oxalyl chloride (OxCl) (4 eq) was added dropwise at RT and allowed to react for 18 hrs under stirring. Upon completion, excess of OxCl and DCM were removed under vacuum. Then, water was added to the residue and stirred for 5 mins. The resulting mixture was extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get crude product which was used for next step without further purification. To the stirring solution of monophosphonate ester (**5**) (1 eq) in DCM, Diethylaminosulfur trifluoride (DAST) (4 eq) was added dropwise at RT and allowed to react for 4 hrs. Excess of DAST and DCM were evaporated under reduced pressure. To the obtained residue, water was added

and stirred for 2 more mins to quench any residual DAST. The reaction mixture was then extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product (**6**). This final probe as used for protein modification without further purification.

Synthesis of the hydrophobic tail

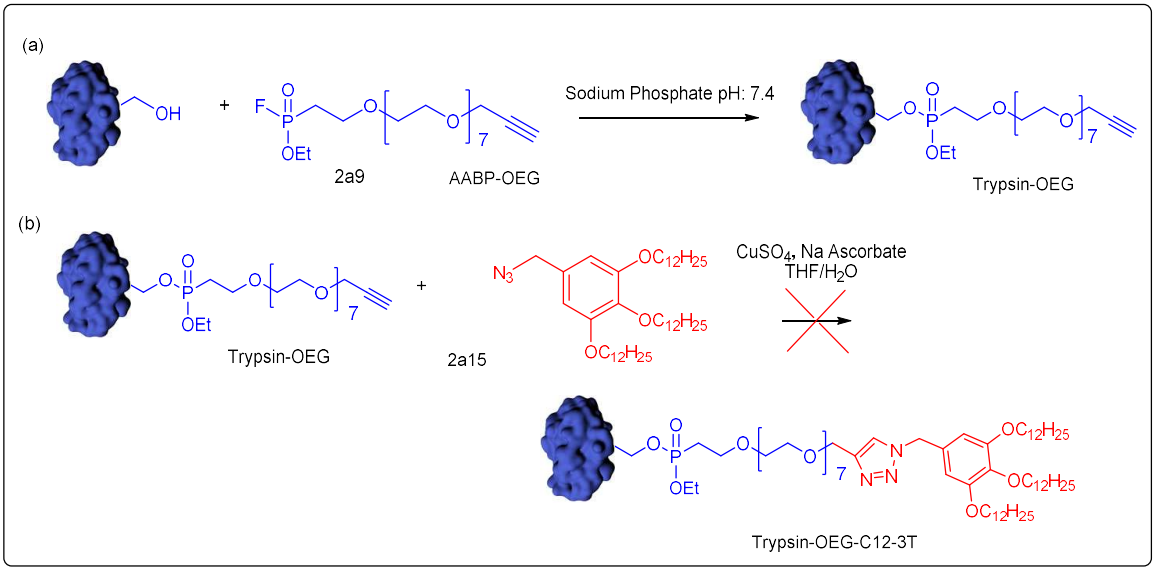


Scheme 2.3. Synthesis of the C12-3T hydrophobic tail.

Ethyl 3,4,5-trihydroxybenzoate (1.0 eq), alkyl bromide (RBr) (3.5 eq), K_2CO_3 (3.5 eq), KI (0.05 eq) were taken in an oven-dried RBF. The flask was then purged with nitrogen, and DMF was added under stirring to dissolve the mixture and heated at 70 ºC for 12 hrs. Upon completion, the reaction mixture was neutralized with acidic water and extracted thrice with EtOAc. The combined organic layer was dried over Na_2SO_4 and evaporated under vacuum to get the crude product, which was purified using NPC using EtOAc / PE. Then, in an oven-dried RBF, above ester (1.0 eq) was taken and dissolved in THF with stirring at 0 ºC. Then lithium aluminum hydride (LAH) (3 eq) was added in portions, maintaining the reaction temperature at 0 ºC and after 30 mins, allowed to react at RT for 12 hrs. Upon completion, LAH was quenched by the dropwise addition of EtOAc at 0 ºC (transesterification was observed if the temperature was raised). Water was added to the reaction mixture and stirred until color changes to off-white. The resulting solution was filtered using Buchner funnel, and the residue was washed with EtOAc. The EtOAc layer was separated from water, and the residual water layer was again extracted thrice with EtOAc. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified by NPC using EtOAc / PE.

In an oven-dried RBF, the above-obtained alcohol (1.0 eq) was dissolved in DCM, and a catalytic amount of dimethylformamide (DMF) was added. Then, thionyl chloride ($SOCl_2$) (1.6 eq) was added dropwise and allowed to react under stirring for 30 mins at RT. Upon completion of the reaction, DCM and excess of $SOCl_2$ were evaporated under vacuum. The residue was dissolved in diethyl ether (Et_2O) and washed thrice with water. The Et_2O layer was then dried over Na_2SO_4 and concentrated under vacuum. The resulting crude product was used for next step without any purification. In an oven-dried RBF, a mixture of above-obtained chloride (1.0 eq), sodium azide (NaN_3) (1.5 eq) was taken, and DMF was added. The resulting mixture was stirred for 12 hrs at RT. Upon completion, water was added to the reaction mixture and extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified using NPC using EtOAc / PE.

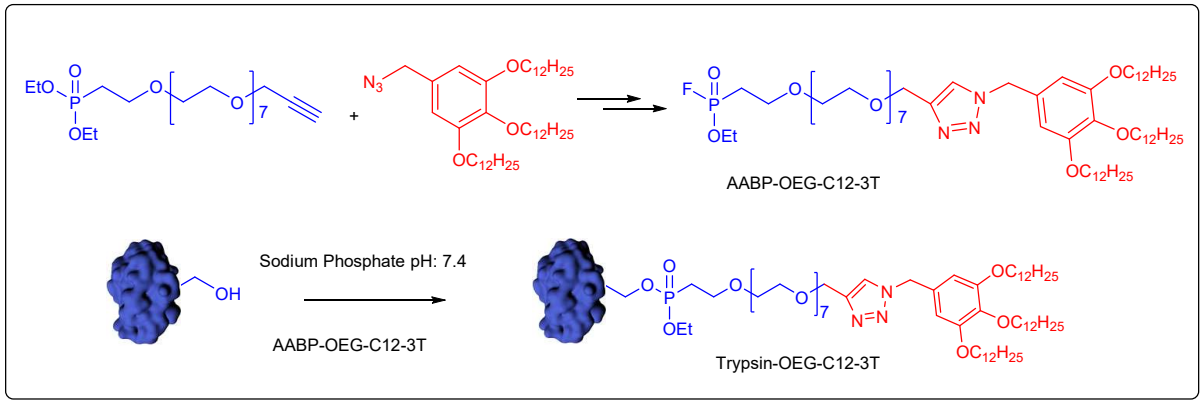
Conjugation with approach i



Scheme 2.4. Attempted step by step conjugation to make FAP as per approach i.

With the hydrophilic spacer and hydrophobic tail made and purified, we have added the spacer to the protein, i.e., Try. MALDI-TOF MS analysis of reaction mixtures was done. The conjugation went well without any problem. Then, we have attempted the installation of a hydrophobic tail (scheme. 2.4). Conjugation didn't work as per MALDI-TOF MS because of orthogonal solubility protein and tail. We have observed protein precipitation, and MALDI-TOF MS results were noisy.

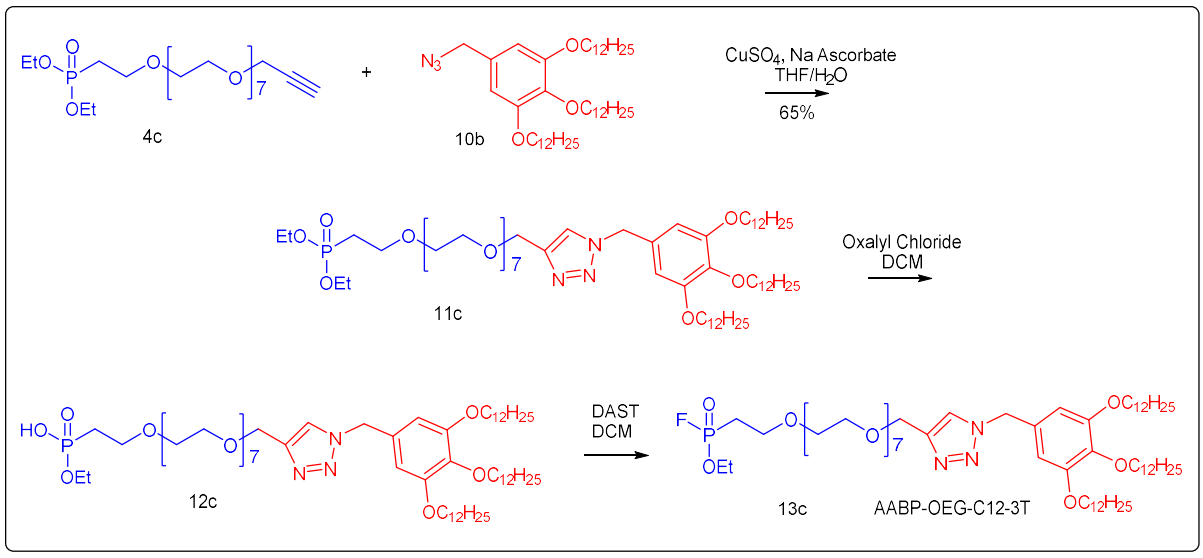
Approach ii: (Synthesis of amphiphilic probe and conjugation)



Scheme 2.5. Proposed conjugation to make FAP as per approach ii.

This second approach is based on the hypothesis that adding spacer and tail first should lead to an amphiphilic molecule with improved aqueous solubility and should react directly without any aid. This amphiphile will be added to the protein in one step (scheme. 2.5). With this hypothesis in our mind, we started this second approach. Amphiphilic activity-based probe (AABP) was made from diphosphonate spacer and tail using click chemistry, and then in 2 steps, we made active AABP as discussed below.

Synthesis of AABP



Scheme 2.6. Synthesis of the amphiphilic probe.

The **AABP** was synthesized by [2 + 3] dipolar cycloaddition (click reaction) of the hydrophilic alkynes (diphosphonate esters) with hydrophobic azides in the presence of sodium Ascorbate (Na Asc.) and copper (II) sulfate ($CuSO_4$) in 50 % THF / H_2O. Then, the resulting product was treated with OxCl in DCM to get monophosphonate ester and finally fluorinated using diethylaminosulfur triflouride (DAST) in DCM. The detailed procedure is given below.

Hydrophobic azide (**10b**) (1 eq) and hydrophilic alkyne (**4c**) (1 eq) were dissolved in degassed THF and stirred until a clear solution was obtained, then degassed H_2O was added and stirred vigorously for 10 more mins. Freshly prepared 1M Na Asc. (0.05 eq) and 1M $CuSO_4$ (0.1 eq) were added to the reaction mixture thrice in intervals of 45 mins and allowed to react for 16 hrs at RT. Upon completion, the reaction mixture was extracted with DCM and the combined organic layer was dried over Na_2SO_4 and concentrated under reduced pressure to get crude product which was purified using reverse-phase chromatography (RPC) using ACN / H_2O system first (to remove unreacted diphosphonate ester) and chloroform ($CHCl_3$) later (to elute clicked product) followed by NPC using MeOH / DCM solvent system.

Note

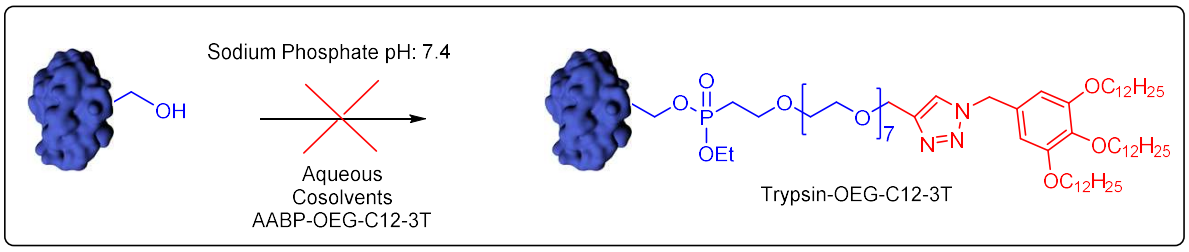
MALDI-TOF MS analysis of the above reaction showed the presence of unreacted diphosphonate ester and click product. Both the components were at the same Retention Factor (R_f) in normal-phase TLC with different solvent systems. However, a huge R_f difference was observed in reverse-phase TLC with 30 % ACN / H_2O. Hence, their separation was achieved using reverse phase chromatography. MALDI-TOF MS analysis of purified fractions showed the complete absence of unreacted diphosphonate ester.

Diphosphonate ester (**11c**) (1 eq) was dissolved in DCM with stirring. Then, OxCl (4 eq) was added dropwise at RT and allowed to react for 18 hrs under stirring. Upon completion, excess of OxCl and DCM were removed under vacuum. Then, water was added to the residue and stirred for 5 mins. The resulting mixture was extracted thrice with DCM, and the combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get crude product which was used for next step without further purification.

To the stirring solution of monophosphonate ester (**12c**) (1 eq) in DCM, DAST (4 eq) was added dropwise at RT and allowed to react for 4 hrs. Excess of DAST and DCM were

evaporated under reduced pressure. To the obtained residue, water was added and stirred for 2 more mins to quench any residual DAST. The reaction mixture was then extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product. This final probe was used for protein modification without further purification.

Conjugation with approach ii



Scheme 2.7. Attempted conjugation of the amphiphilic probe with co-solvents.

The AABP, though amphiphilic, was not aqueous soluble. So, we tried the labeling with the help of co-solvents, even in the presence of co-solvents, the probe solution was turbid and precipitating (scheme. 2.7). We observed very turbid and precipitated samples even in the reaction mixtures, probably due to precipitation of trypsin because of organic solvents. When monitored in MALDI-TOF MS we saw noisy results (could not identify the protein peak) and often multiple labeling. Then we came up with another approach, which is Micelle-Assisted Protein Labeling (MAPLab).

Approach iii: Micelle Assisted Protein Labelling (MAPLab)

As expected and discussed before, the solubility of the probe (FP-OEG-C12-3T) posed a challenge in terms of choosing an optimum condition to carry out protein labeling reaction in 100 % aqueous medium and even in co-solvent systems. Several reaction conditions were attempted, but none of them was successful. Considering the amphiphilic nature of the probe, we hypothesized that it could be solubilized in the presence of a small-molecule micelle.

Surfactant screening for MAPLab

To test this hypothesis that the probe could be solubilized in the presence of a small-molecule micelle, extensive surfactant screening was carried out. We choose neutral surfactants like tritons and tweens for testing (Fig. 2.6). Tritons are typical non-ionic detergents, derived from

polyoxyethylene and contain an alkylphenyl hydrophobic group. Among the members of triton family, triton X-100 (n = 9-10, CMC = 0.2-0.9 mM (20-25 ºC, cloud point = 65 ºC), triton X-114 (triton X-114, n = 7-8, CMC = 0.2 mM (20-25 ºC, cloud point 23 ºC), were screened. The CMC values of these detergents are low, and therefore these detergents cannot be easily removed by dialysis. Tweens are polysorbate surfactants with a fatty acid ester moiety and a long polyoxyethylene chain. The CMC values of these detergents are also low and form micellar assemblies just like tritons.

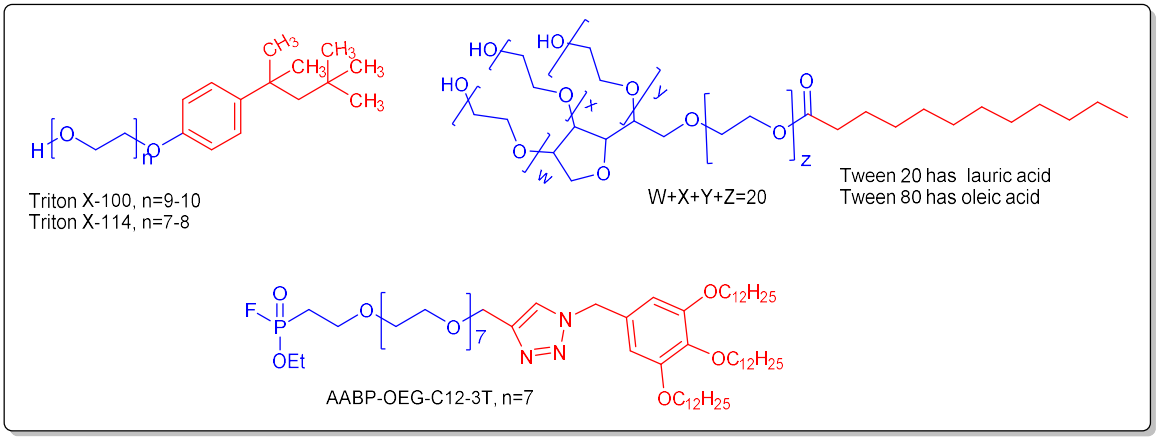


Figure 2.6. Structure comparison of tritons and tweens with respect to AABP.

MAPLab mechanism

To our delight, both triton X-100 and Triton X-114 were found to assist the labeling. Among the members of the tween family, tween 20 and tween 80 were screened, and to our surprise, we couldn't see the labeling in both the cases. A detailed investigation is still needed regarding why these surfactants did not help in conjugation. The structure of tritons and tweens in comparison with the probe is given below. A reasonable assumption is that the amphiphilic probe would make a mixed micelle in the presence of another suitable surfactant, where the tail component of a probe would be part of the hydrophobic micellar interior, and a hydrophilic end (oligoethylene glycol with a terminal fluorophosphonate group) would be available for active-site labeling of a serine protease[32] (Fig. 2.7).

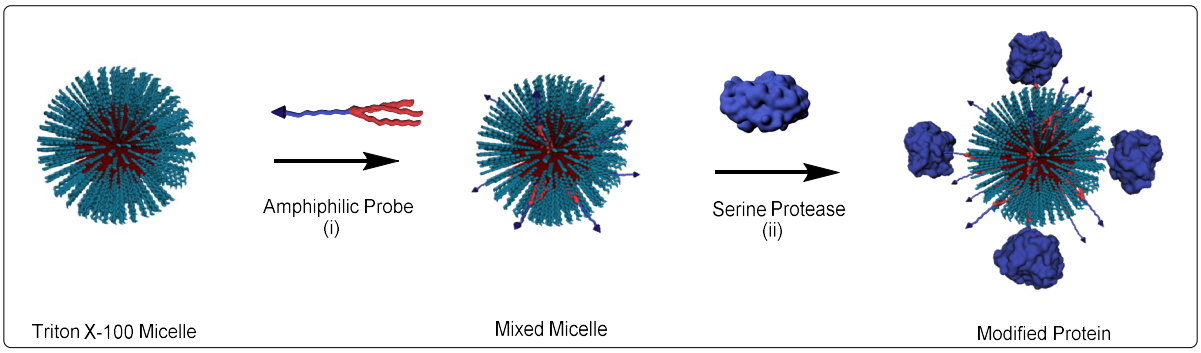
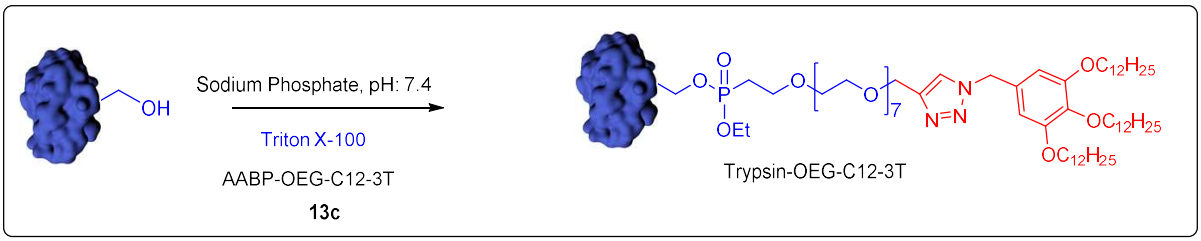


Figure 2.7. The proposed scheme of MAPLab technology (approach iii). The first step involves solubilization or mixed micellization of the probe in triton X-100. In this state, the FP group of the probe will be available for labeling. In the next step, we are adding a protein solution to this micellar solution to do the conjugation. After surfactant screening, we have optimized MAPLab technology; Triton X-100 was found to be best, as it satisfies all the requirements for protein labeling reaction, and we did the test scale reaction.

Test scale conjugation



Scheme 2.8. Proposed conjugation strategy to make FAP (MAPLab Technology, approach iii).

As a test case, micelle-assisted labeling of Try (serine protease) by FP-OEG-C12-3T was first attempted. Try modification was carried out at the concentration of 100 µM, which was found to be optimum for MALDI-TOF MS monitoring. Triton X-100 was used to solubilize the AABP at concentrations 100 times (20 mM) more than CMC or 2 % of the total volume of the reaction mixture. Typically, for test reactions, the final volume of the reaction mixture is 1 mL. Try was weighed (2.3 mg) in microcentrifuge tubes and 500 µL of 50 mM sodium phosphate (Sod. Phos.) pH 7.4 was added and mixed gently with a pipette to make 200 µM solutions. Then AABP (1 or 2 equivalents) was weighed in a different microcentrifuge tube, followed by the addition of 20 µL triton X-100 and 480 µL of 50 mM Sod. Phos. pH 7.4 and vortexed for 15 mins. When the AABP solution becomes clear, the protein solution was

added into AABP solution to get 100 µM (1 mL) protein solutions and allowed to react for 24 hrs on rotospin at 20 rpm at 25 ºC (scheme 2.8) Reaction mixture was clear and MALDI-TOF MS analysis became easier and confirmed conjugate formation.

Scale-up conjugation

Try modification was carried out in a falcon tube at 200 mg scale following the linear scale-up of the above-mentioned procedure for understanding the self-assembly behavior.

Monitoring the conjugation

To monitor the extent of protein modification, the samples were directly withdrawn from the reaction mixture using a pipette and analyzed. In brief, 2 µL of the reaction mixture was mixed with 2 µL of 2 % TFA and 2 µL of matrix mixture as previously stated, vortexed and spotted on MALDI-TOF MS plate. Please note that the matrix preparation and sample preparation procedures remained the same here (Fig. 2.8).

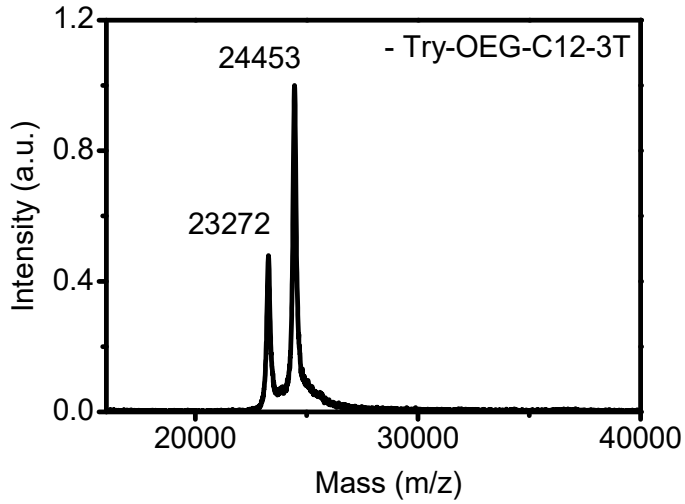


Figure 2.8. MALDI-TOF MS analysis of reaction mixture (MAPLab Technology, approach iii).

2.1.4. Purification optimization

The purification of this facially amphiphilic protein is the next biggest challenge we have encountered in this work as the made FAP resembles membranes proteins (Fig. 1a) in terms of hydropathy index[24]. There are 4 components in the reaction mixture, 1. Conjugate, 2.

Unreacted native protein, 3. Excess probe, and 4. Triton X-100. We have explored several strategies; some of them are briefed below.

Hydrophobic interaction chromatography

Hydrophobic interaction chromatography (HIC) works on the fact that the proteins have varying degrees of surface hydrophobicity. HIC resin is made of hydrophobic ligands and interacts with proteins via a reversible interaction. This interaction varies for proteins with different degrees of surface hydrophobicity. This interaction is influenced by salt concentration in the running buffer. A high salt concentration in the running buffer enhances the interaction and helps in the capture. Lowering the salt concentration weakens the interaction and helps in elution. The elution of proteins will be in the order of least hydrophobicity to the highest. Based on the structural differences between surface hydrophobicities of the native protein, and conjugate, we tried HIC but, triton seems to hamper the binding interaction of proteins. We observed both native and FAP/conjugate eluting through the run because of probable interference of triton X-100.

We have also performed HIC after removing triton X-100, which also did not help. Elution in reduced salts does not yield FAP even here. But when washed with NaOH, we have observed FAP eluting. This might be because of the strong interaction of FAP with the HIC column. Another issue is that NaOH wash gives a very large volume of samples often in liters, which need to be concentrated and buffer exchanged to monitor. The conjugate also shows native protein in this process.

Cloud point protein extraction

While the cloud point of Triton X-100 is 64 ºC, the cloud point of Triton X-114 is 23 ºC. This allows for membrane protein extraction and solubilization in Triton X-114 without bringing the samples up to warmer temperatures, which may denature many proteins. We tried separating the native protein from conjugate based on cloud point protein extraction (CPPE). To do this, we did conjugation in Triton X-114, which worked, and when we did the separation of the reaction mixture, native stayed in aq. layer predominantly and conjugate got into the triton X-114 layer during CPPE (Fig. 2.9)

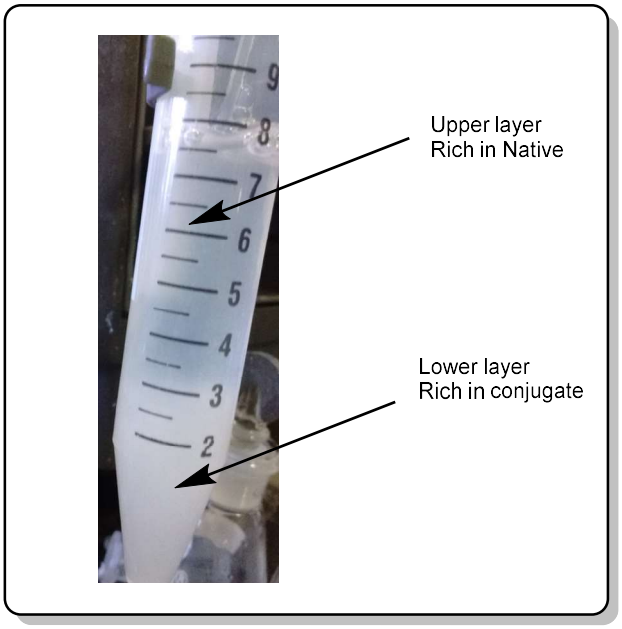


Figure 2.9. Cloud point-based separation of native and conjugate.

3-step purification

Finally, we have purified the Try conjugate (prototype FAP) by three-step purification, *i.e.,* IEX, SEC, and desalting performed using FPLC.

Ion-exchange chromatography

The first step is ion-exchange chromatography (IEX), in which neutral triton X-100 and excess unreacted probe were removed as they do not bind to the column. IEX was performed on SP sepharose resin, a cation-exchange resin at pH 7.4.

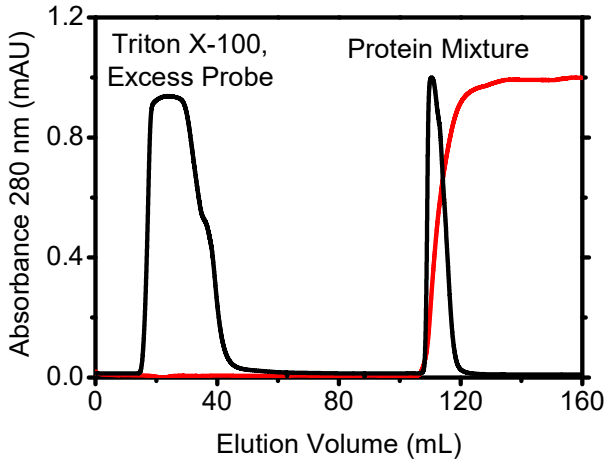


Figure 2.10. The first step of three-step purification for separating protein conjugates from reaction mixtures. Representative IEX chromatogram showing the separation of triton X-100 from protein mixture (native and modified protein).

During this cation-exchange chromatography (CEX), the column was pre-equilibrated with the same buffer (50 mM Sod. Phos. pH 7.4) which was used for modification and then the sample was injected followed by post-injection equilibration for at least 2 Column Volumes (CVs) or until the complete removal of triton X-100 for large scale reactions. The elution of native protein and its corresponding conjugate together as a mixture was later achieved using 50 mM Sod. Phos. pH 7.4, 1 M NaCl as elution buffer (Fig. 2.10).

Following a similar procedure as state above, Triton X-114 used is used in place of Triton X-100, IEX worked only for Try-OEG-C6-3T and not for higher conjugates. This needs to be investigated. Even After the successful separation of native and conjugate using CPPE, we tried separating the triton-X 114 from conjugate using IEX. Unfortunately, IEX did not work, and we could not separate triton-X 114 from conjugate (Try-OEG-C12-3T). We observed elution of triton-X 114 and conjugate together (conjugate is supposed to bind to the column). A detailed investigation is needed regarding this aspect.

Size-exclusion chromatography

The second step is based on the hypothesis that the FAP is amphiphilic, and should result in complex formation and consequently elute at less volume than the native proteins in Size exclusion chromatography, which would aid in purification by the Size-exclusion chromatography (SEC) either at 200mM or increased NaCl concentrations.[33]

As expected, when the obtained IEX fractions were subjected to SEC, protein conjugate Try-OEG-C12-3T got eluted in void volume suggesting the formation of higher-order complex, while monomeric native Try eluted at 65 mL or later. For the separation of native protein from protein conjugate, 50 mM Sod. Phos. pH 7.4, 1 M NaCl was used as buffer using either Sephacryl S-100 HR 16/60 or Sephacryl S-200 HR 16/60 or Sephacryl S-300 HR 16/60 (Fig. 2.11).

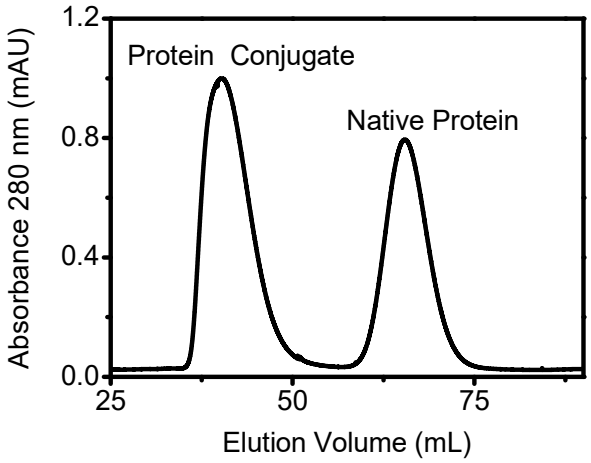


Figure 2.11. Representative SEC chromatogram showing for separation of protein conjugate from native. The modified protein appears larger in size as evident from elution volume, because of complex formation.

Desalting

Finally, NaCl was removed by the standard desalting method using a Sephadex-G25 column (Fig. 2.12). After three-step purification, a single peak for the purified conjugate at 24,453 Da with very high purity was obtained with an overall yield of 20 %. The desalted fractions were quickly lyophilized and later dissolved in the required buffer when needed.

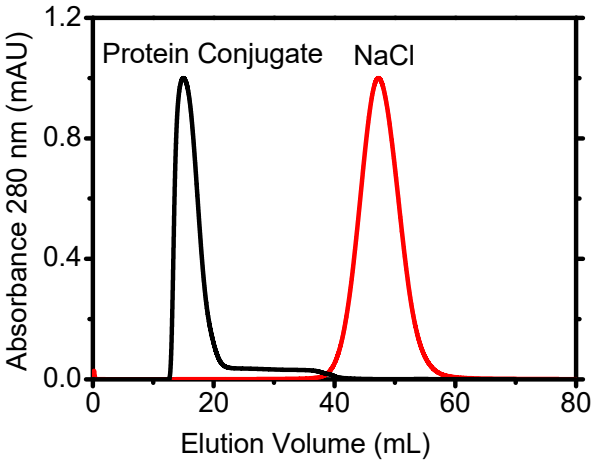


Figure 2.12. The third step of three-step purification for separating salt from protein conjugate. The representative desalting chromatogram showing the removal of salt.

Monitoring purification

Monitoring of IEX, SEC, and desalted fractions were carried out using the same procedure mentioned for native proteins and reaction mixtures except for the addition of triton X-100 in protein fractions. In short, 98 µL of samples from each FPLC fractions were mixed with 2 µL of triton X-100 (2 % or 100 times the CMC) in a separate microcentrifuge tube and vortexed for 4 h. The samples were then analyzed using the same procedure as mentioned above. Please note that the matrix and sample preparation procedure remained the same. Best results in MALDI-TOF MS obtained during the SEC stage as desalting resulted in further dilution of protein samples. Desalting shouldn't affect the purity of the samples.

2.1.5. Biophysical characterization

Before understanding the self-assembly behavior of the conjugates, we characterized the prototype FAP using MALDI-TOF MS, Circular Dichroism (CD) spectroscopy, fluorescence spectroscopy, and zeta potential measurements. The detailed procedures for characterization are given in the experimental section. The molecular weights of conjugates obtained in MALDI-TOF MS analysis are in agreement with the calcd ones (Fig. 2.13). In order to probe whether the secondary and tertiary structure of made protein conjugate is intact, detailed biophysical studies were carried out using fluorescence.[36] Fluorescence studies reveal there was no shift in tryptophan emission maximum (295 nm), which indicates the 3D structure of a protein is preserved in the complex state. However, we observed an increase in the intensity of the emission for the conjugate. This could be because of the increased order/folding of the conjugate when compared with native (Fig. 2.14). CD spectroscopy[37] (195-250 nm) also indicated no perturbation in the secondary structure of the conjugate (Fig. 2.15) except for a slight decrease in ellipticity. More discussion is presented in the next sub-chapter. However, a zeta potential measurement was carried out to investigate the surface charge of protein conjugate. The obtained value +16.5 mV is different from that of the native protein (+30.4 mV). This needs further investment. Soon after characterization, we started self-assembly studies of these conjugates (Fig. 2.16).

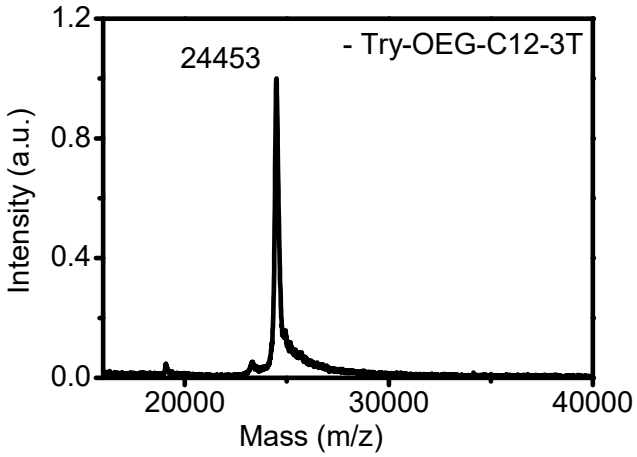


Figure 2.13. MALDI-TOF MS characterization of Try-OEG-C123T conjugate after purification.

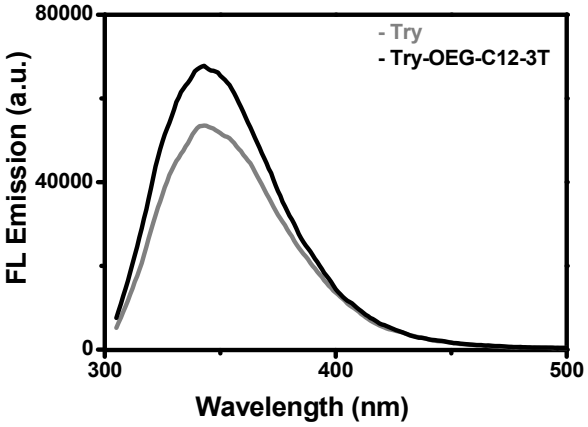


Figure 2.14. Fluorescence characterization of Try-OEG-C123T conjugate after purification.

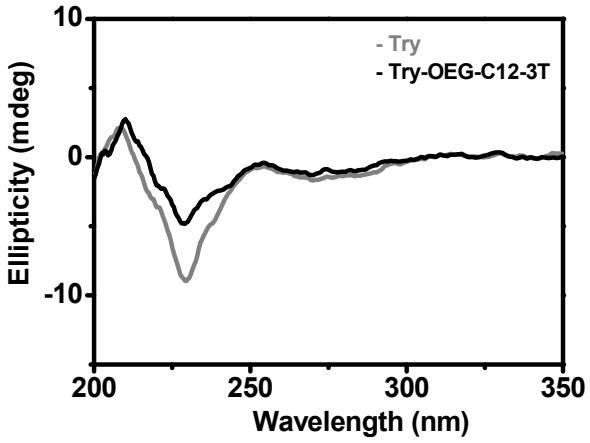


Figure 2.15. CD characterization of Try-OEG-C123T conjugate after purification.

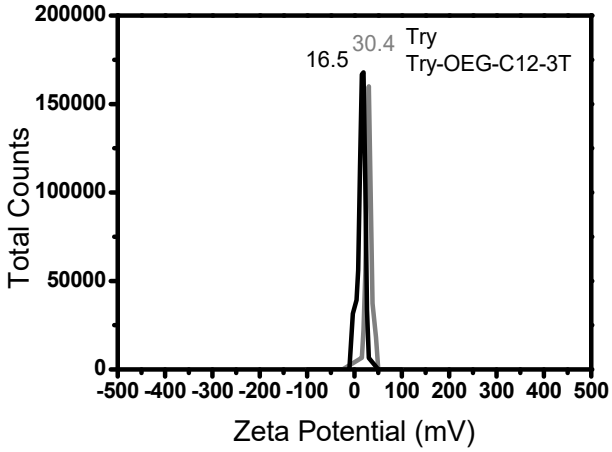


Figure 2.16. Zeta potential characterization of Try-OEG-C123T conjugate after purification.

Table 2.1. Summary of characterization data of prototype FAP.

Protein	MW (Da)	Zeta Potential (+mV)
Try native	**23,272**	**30.4**
Try conjugate (FAP)	**24,445**	**19.5**

2.1.6. Self-assembly studies

Soon after purification, we have initiated the self-assembly studies of prototype FAP using a variety of techniques, which include analytical size-exclusion chromatography (SEC), SEC coupled with multi-angle laser light scattering (SEC-MALS) dynamic light scattering (DLS) and small-angle neutron scattering (SANS). The procedure for measuring SEC, SEC-MALS, DLS, and SANS is given in the methods section at the end of this chapter. To determine the relative molecular weight (RMW) of soluble protein complexes, the protein conjugate was dissolved individually in Milli Q water, and a series of SEC runs were performed. The conjugate eluted at 11.4 mL corresponding to 240 kDa (Fig. 2.17). The mean size of the protein complexes was measured using DLS using Zetasizer Nano 2590 (Malvern, UK). DLS showed that the conjugate could form a higher-order complex of hydrodynamic diameter (D_h) of 11 nm (Fig. 2.18). To further understand the oligomeric states of the complex formed by

prototype FAP, we did SANS studies, which showed that the complex is 10 mer (Fig. 2.19). SEC-MALS analyses of the protein complexes were performed on a Superdex-200 10/300 GL column (GE Healthcare) connected to an Agilent HPLC system equipped with the 18-angle light scattering detector (Wyatt Dawn HELIOS II) and a refractive index detector (Wyatt OptilabTrEX) calibrated with BSA to determine the absolute molecular weight (AMW). SEC MALS showed that the conjugate is forming a higher-order complex of about 250 kDa. (Fig. 2.20). The obtained data in DLS, SEC, SANS, and SEC-MALS studies are in agreement with each other.

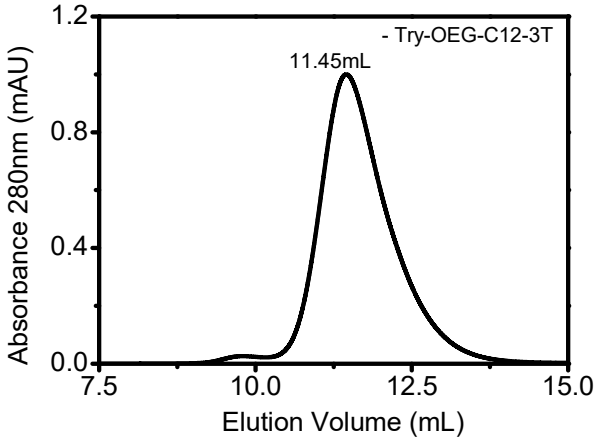


Figure 2.17. SEC data of Try-OEG-C123T conjugate.

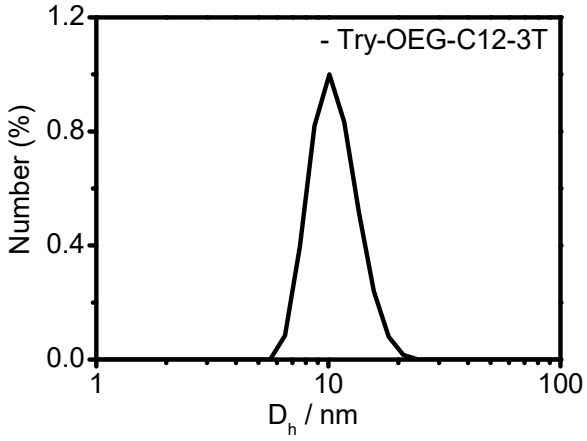


Figure 2.18. DLS data of Try-OEG-C123T conjugate.

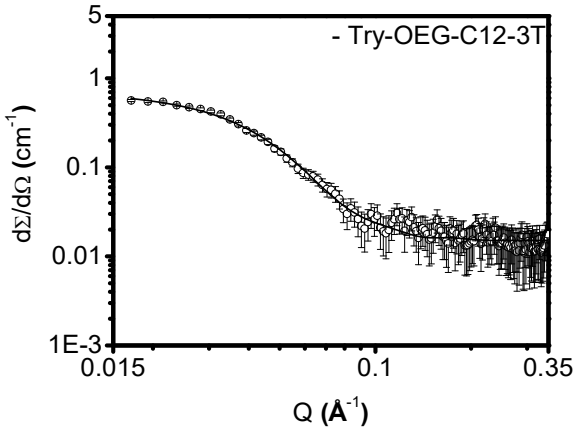


Figure 2.19. SANS data of Try-OEG-C123T conjugate.

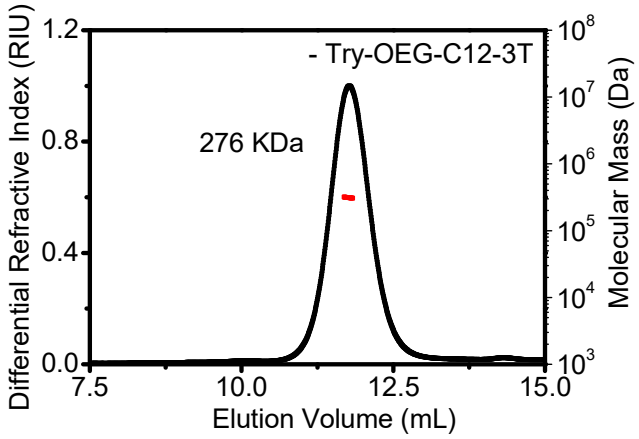


Figure 2.20. SEC-MALS data of Try-OEG-C123T conjugate.

Table 2.2. Summary of self-assembly data of prototype FAP.

Protein Conjugate	Elution Vol SEC (mL)	Mol Wt MALS (kDa)	Mol Wt SANS (kDa)	Oligomeric State (mer)	D_h DLS (nm)
Try-OEG-C12-3T	11.4	276	245	10	11

CMC estimation

We have attempted pyrene encapsulation-based CMC estimation but failed to encapsulate pyrene. We have also attempted to measure the CMC using SEC (Fig. 2.21). By injecting as little as possible in the SEC, we have measured the concentration of eluted protein complex using nanoDrop. The concentration was found to be 250 nM. Therefore the CMC seems to be less than 250 nM. Detailed investigation regarding CMC is needed. The particles are stable over 4 months, even after 5 free thaw cycles.

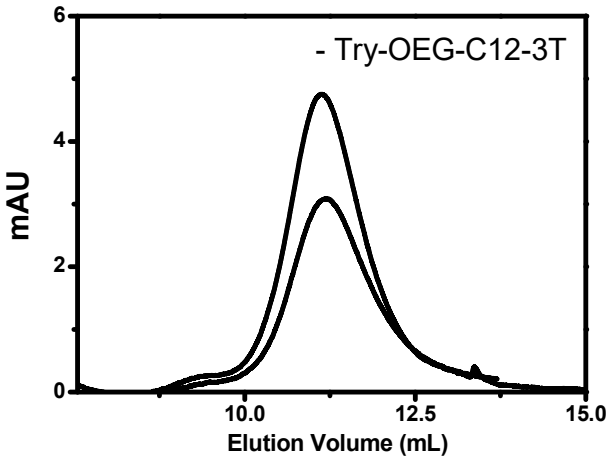


Figure 2.21. SEC based CMC estimation of Try-OEG-C123T.

2.1.7. Discussion

We have designed a new scaffold called FAP inspired by structural, biophysical, and self-assembly studies of naturally occurring facially amphiphilic proteins called hydrophobins. To construct this scaffold, we have tried optimizing the conjugation protocol and arrived at MAPlab technology, which uses FP equipped AABP and Triton X-100. FP equipped AABP was conjugated to serine protease, i.e., Try. The purification was then optimized and finally done in three steps. All the stages were MALDI-TOF MS monitored, and the conjugate was obtained in more than 95 % purity. The designed protein conjugate, i.e., prototype FAP, driven by strong hydrophobicity, assembled to form the complex as expected (Fig. 2.22).

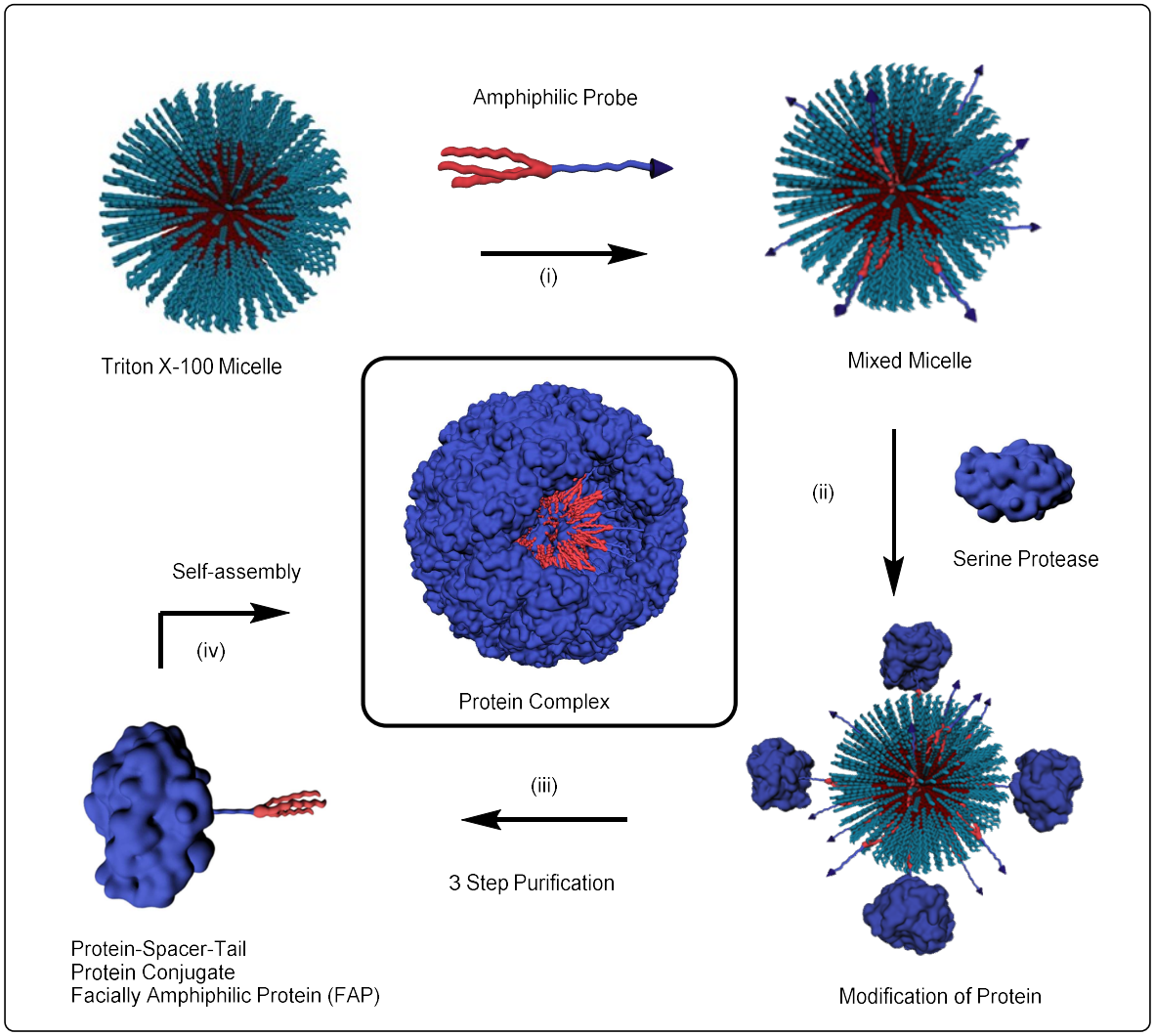


Figure 2.22. Summary of chemical methodology to make FAPs. i. The AABP was solubilized in the presence of triton X-100, ii. MAPLab. iii. Three-step purification to remove triton X-100 and the unreacted protein using IEX, SEC, and desalting. iv. Self-assembly of FAP.

Library plan.

At this state, after establishing the MAPLab technology, we wanted to study the FAP systematically by making a library of FAPs. The FAP library was planned in the fashion shown below. We wanted to explore 4 proteins, 5 spacers, and 9 tail groups. In each library, we varied 1 component at a time, keeping the remaining 2 components constant. For example, in the head library, proteins were varied, keeping the spacer and tail constant, and vice versa, which is discussed next three sub-chapters 2.2, 2.3, 2.4 (Fig. 2.23).

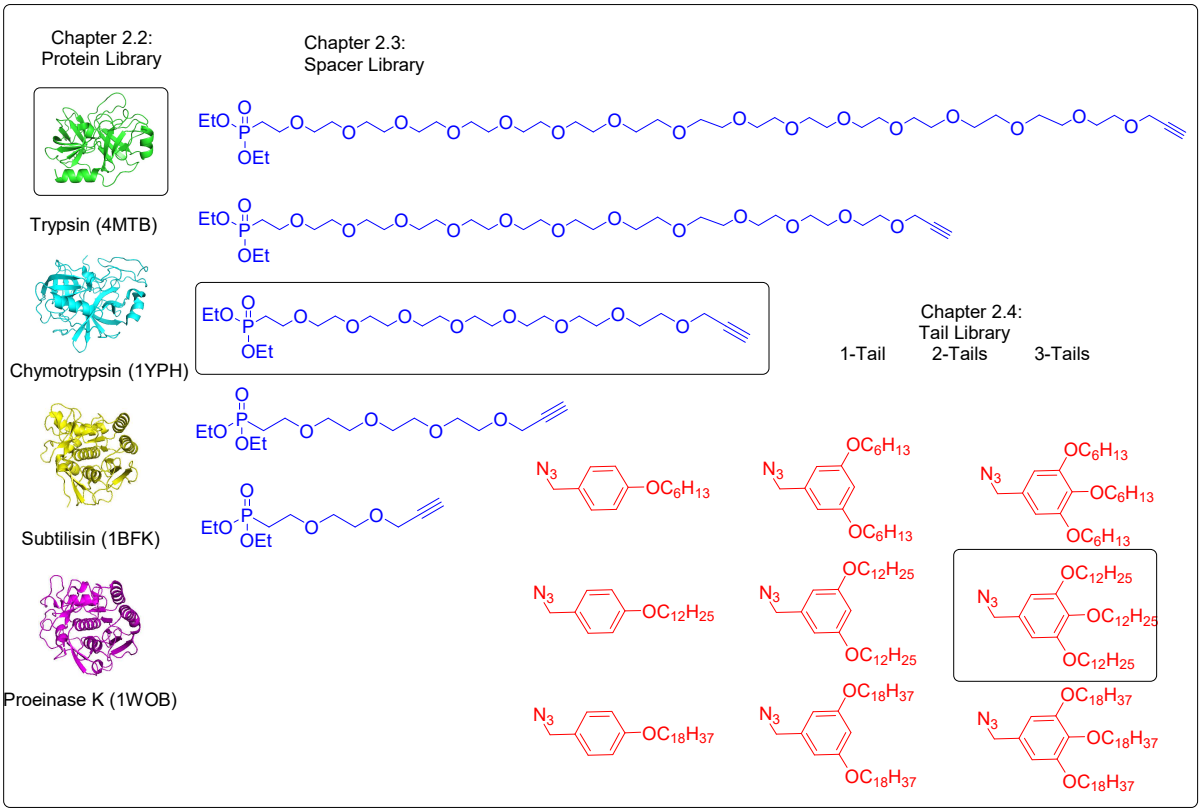
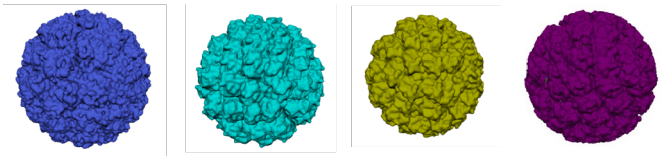


Figure 2.23. Crystal structures of serine proteases used in this study (Try, Chy, Sub, ProK; PDB: 4MTB, 1GL0, 4C3V, 5B1D respectively) and chemical structures of the hydrophilic linker and hydrophobic tails.

2.2. Synthesis and Self-assembly Studies of Protein Variant FAPs



2.2.1. Design

This sub-chapter is about studying the influence of protein groups in protein complex formation by FAP. Motivated by the preliminary results we got for prototype FAP in sub-chapter 2.1, we have made a protein variant library to further validate the method and to investigate the effect of protein group on complex attributes. Serine hydrolases consist of at least 200 enzymes in humans, but we chose only four proteins to study the effect of protein head group, i.e., i. Try, ii. Chymotrypsin (Chy), iii. Proteinase K (ProK) and iv. Subtilisin (Sub), which are close in terms of MW. The crystal structures of these proteins given below. Spacer and Tail groups were kept constant as octaethylene glycol and C12-3T, respectively (Fig. 2.24). The structure of conjugates is given below (Fig. 2.25).

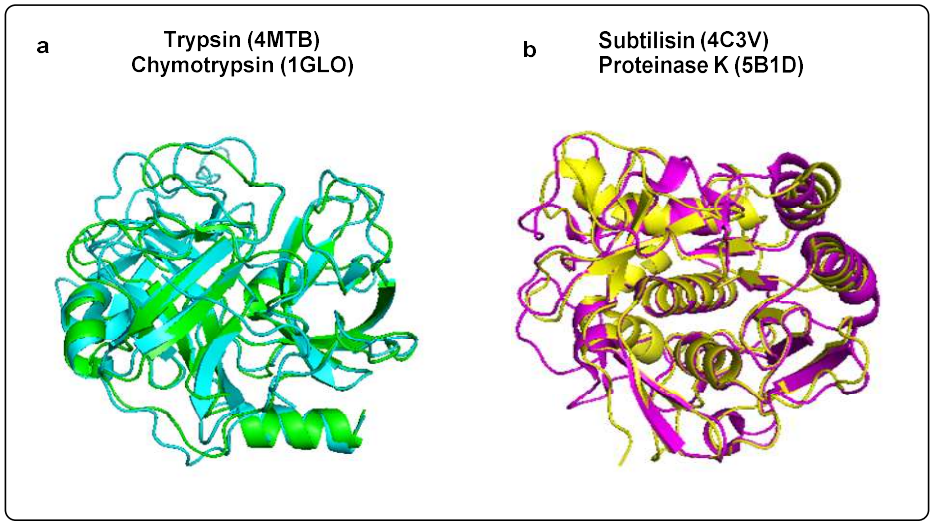


Figure 2.24. Overlaid crystal structures of serine proteases used in this study. a. Try, and Chy (4MTB, 1GLO, respectively), b. Sub and ProK (4C3V, 5B1D respectively)

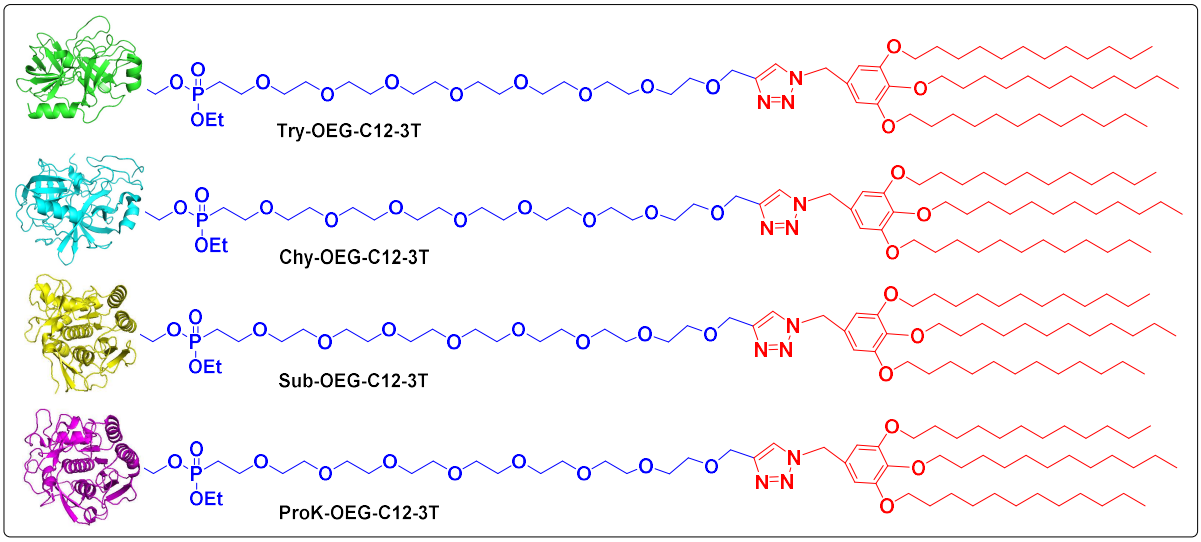


Figure 2.25. Structures of protein conjugates (head variant library).

2.2.2. Biophysical characterization of native proteins

Before doing the conjugation of proteins, we characterized the proteins using MALDI-TOF MS, SANS, and zeta potential measurements. The detailed procedures for characterization are given in the experimental section. The very first thing we did is getting the molecular weights using MALDI-TOF MS because we have to monitor the molecular weight changes accurately during conjugation and purification. The molecular weights obtained are shown below. The molecular weights of native proteins obtained in MALDI-TOF MS analysis are in agreement with the reported ones (Fig. 2.26). SANS analysis was performed to make sure that the native proteins exist as monomers and to know the size information of the same. As expected, these proteins are monomeric with a diameter of less than 2 nm (Fig. 2.27 and Fig. 2.28). As part of characterization, zeta potential measurements were also carried out to investigate the surface charge of selected proteins to design complexes. The surface charge analysis showed that both Try and Chy complexes are highly cationic (+30.4 mV and +19.5 mV), ProK and Sub are anionic (-6.23 mV and -8.4 mV) (Fig. 2.29 and Fig. 2.30). Soon after the characterization of native proteins, we started the synthesis of spacer, tail, and AABP for conjugation with these proteins.

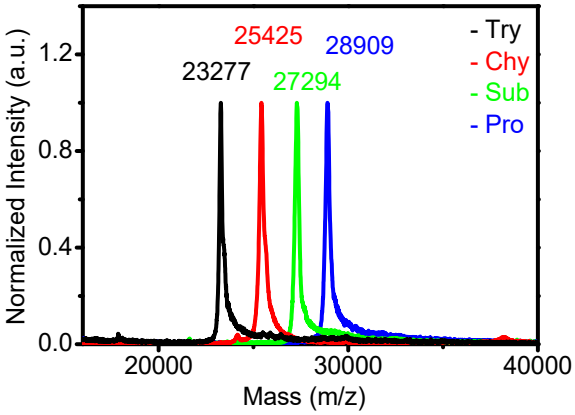


Figure 2.26. MALDI-TOF MS characterization of native proteins (Try, Chy, ProK and Sub).

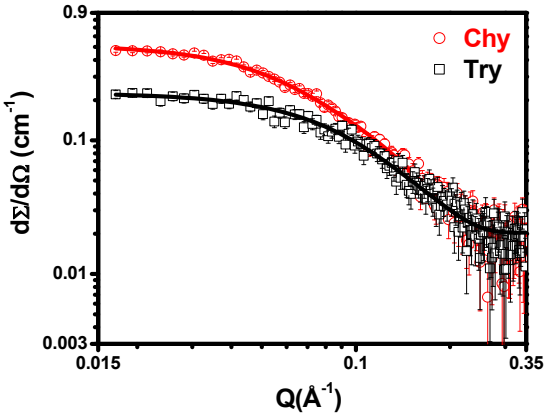


Figure 2.27. SANS for native proteins (Try and Chy).

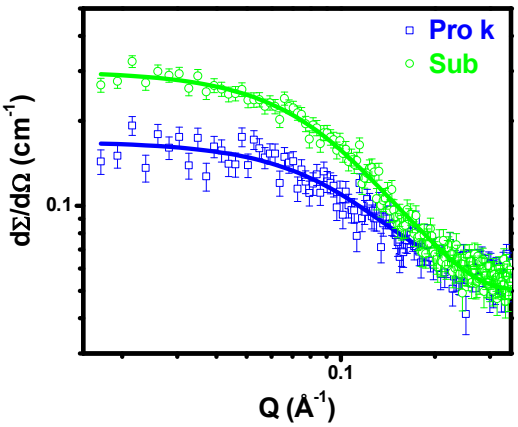


Figure 2.28. SANS for native proteins (ProK and Sub).

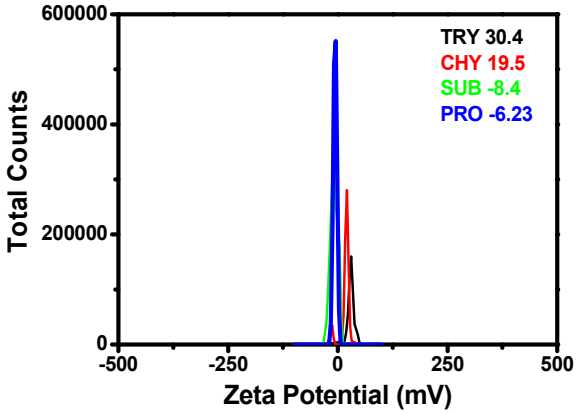


Figure 2.29. Zeta potential measurement of native proteins (Try, Chy, ProK and Sub).

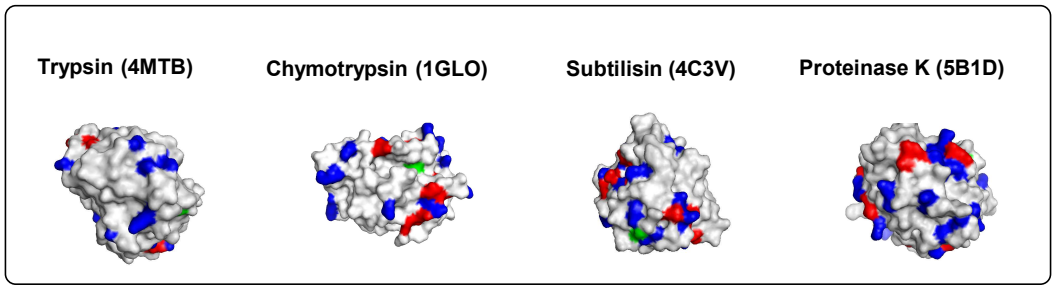


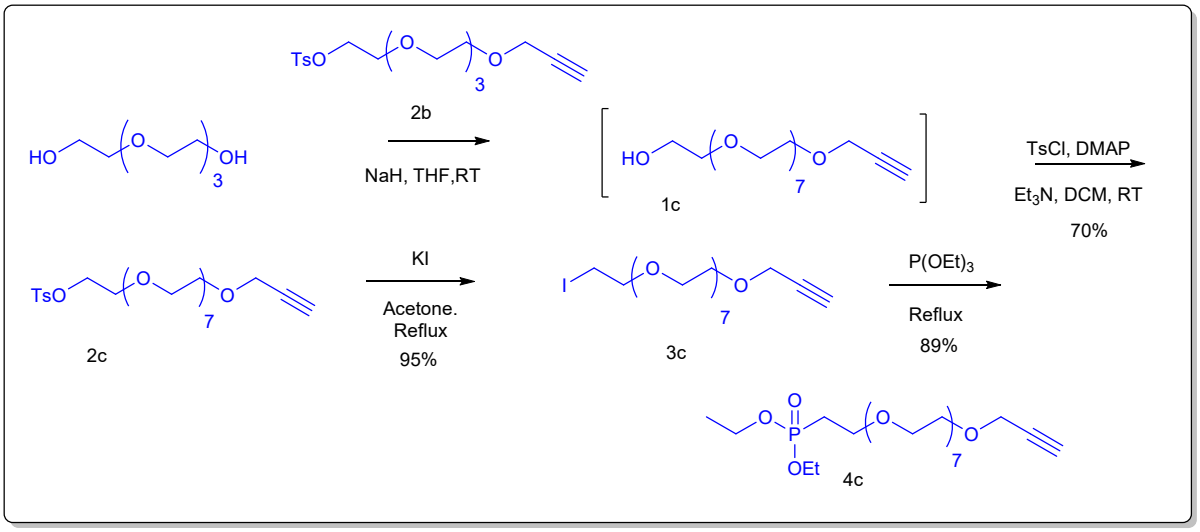
Figure 2.30. The electrostatic surface potential of native proteins. Negatively charged residues like Asp and Glu are colored red. Positively charged residues like Arg and Lys are colored blue. His residues are colored green, and active site Ser alone is represented in magenta (site of conjugation).

Table 2.3. Summary of characterization data for native proteins.

Native Protein	Mw (Da)	Zeta Potential (mV)	R_g from P(r) (Å)	Model Used
Try	**(23,272)**	**+30.4**	**16**	
Chy	**(25,425)**	**+19.5**	**18**	**Prolate**
ProK	**(28,909)**	**-6.2**	**13.1**	**Ellipsoidal**
Sub	**(27,294)**	**-8.4**	**14.4**	

2.2.3. Synthesis

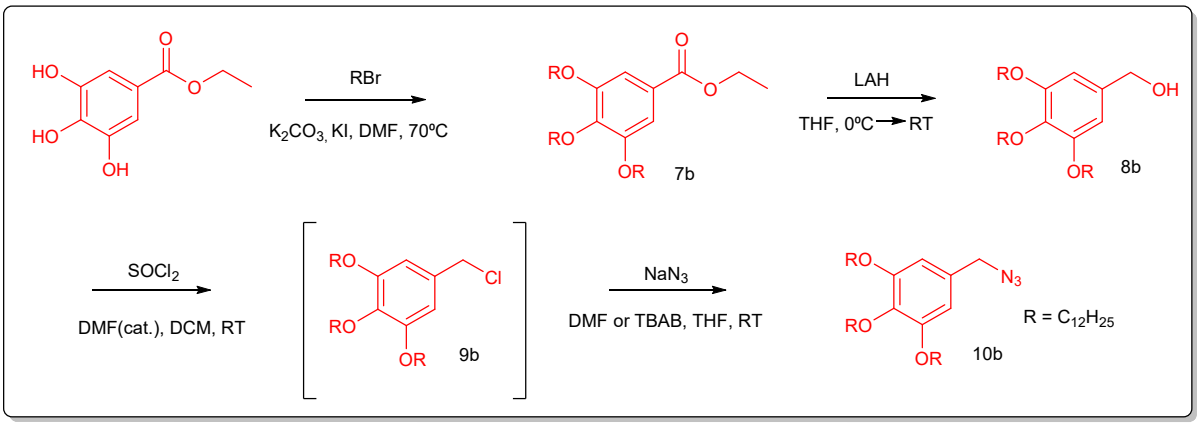
Synthesis of the hydrophilic spacer



Scheme 2.9. Synthesis of the OEG spacer.

As discussed in chapter 2.1, the required diphosphonate ester (**4c**) was made in the following steps. Synthesis of the hydrophilic spacer (**1c)** was made from tetraethylene glycol (TEG) and the tosylate (**2b**) in the presence of NaH to get mono propargyl derivative (**1c**). This monopropargylate was tosylated in the presence of KOH. The obtained tosylate (**2c**) was refluxed in acetone for with KI to get the corresponding iodide. Finally, in the presence of $P(OEt)_3$, the iodide was converted into diphosphonate ester (**4c**) (scheme 2.9).

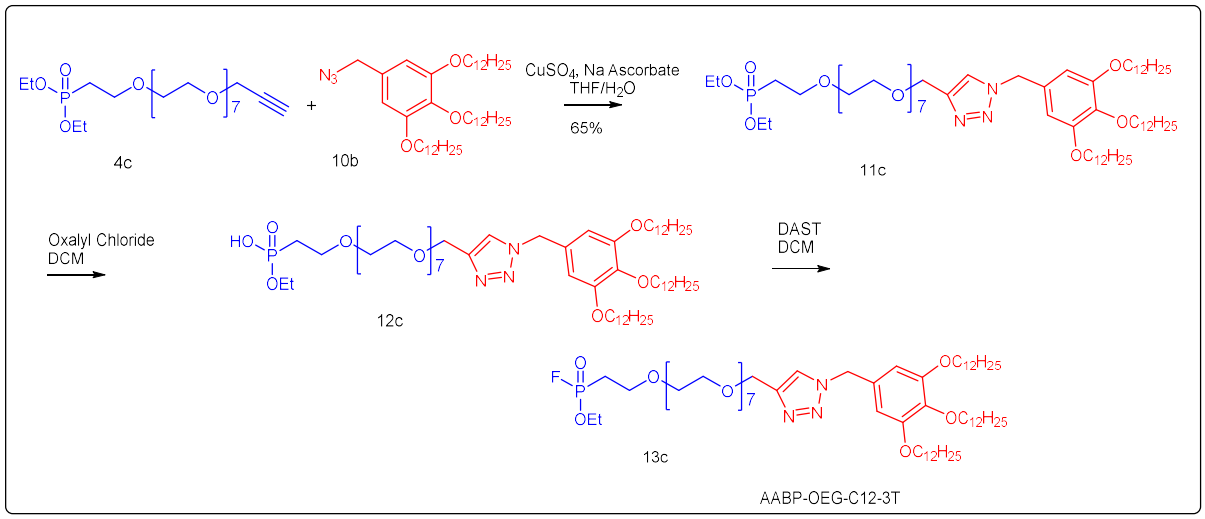
Synthesis of hydrophobic azide



Scheme 2.10. Synthesis of C12-3T hydrophobic azide.

The synthesis of hydrophobic azide was discussed in chapter 2.1. To brief, Ethyl 3,4,5-trihydroxybenzoate was alkylated in the presence of RBr, K_2CO_3, KI to get alkylated ester (**7b**). Then in the presence of LAH, the ester was reduced to its corresponding alcohol (**8b**). This was then converted into chloride (**9b**) using $SOCl_2$ in the presence of DMF. Finally above obtained chloride (1.0 eq) in the presence of NaN_3 (1.5 eq) was converted into azide (**10b**) (scheme 2.10).

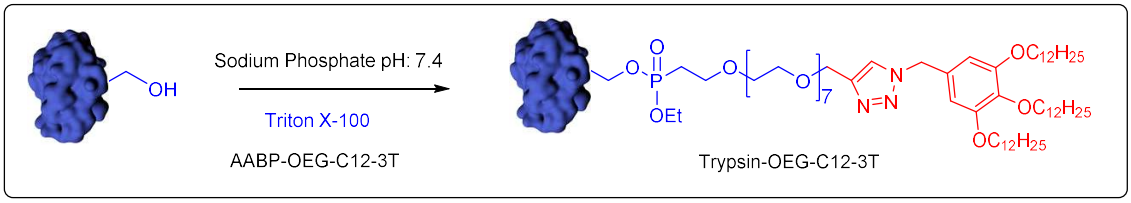
Synthesis of the AABP



Scheme 2.11. Synthesis of the amphiphilic probe for protein variant library.

The synthesis of this AABP is discussed in sub-chapter 2.1. To brief, All the AABPs in this library were synthesized by [2 + 3] dipolar cycloaddition of the hydrophilic alkyne (**4c**) with hydrophobic azide (**10b**) in the presence of Na Asc. and $CuSO_4$ in 50 % THF / H_2O. Then, the resulting diphosphonate ester (**11c**) was treated with OxCl in DCM to get monophosphonate ester (**12c**) and finally fluorinated using DAST in DCM. This final AABP (**13c**) was used for protein modification without further purification (Scheme 2.11).

2.2.4. Conjugation and purification



Scheme 2.12. General conjugation scheme for making protein variant library.

The modification of proteins was carried out at the concentration of 100 µM, which was found to be optimum for MALDI-TOF MS monitoring. Triton X-100 was used to solubilize the AABPs at concentrations 100 times (20 mM) more than CMC or 2 % of the total volume of the reaction mixture. Typically, for test reactions, the final volume of the reaction mixture is 1 mL. Proteins were weighed (2.3 to 2.9 mg, depending on protein) in microcentrifuge tubes and 500 µL 50 mM Sod. Phos. pH 7.4 was added and mixed gently with a pipette to make 200 µM solutions. Then AABPs (1 or 2 eq) were weighed in a different microcentrifuge tube, followed by the addition of 20 µL triton X-100 and 480 µL of 50 mM Sod. Phos. pH 7.4 and vortexed for 15 mins. When the AABP solution becomes clear, the protein solution was added into AABP solution to get 100 µM (1 mL) protein solutions and allowed to react for 24 hrs on rotospin at 20 rpm at 25 ºC (scheme 2.12).

FP-OEG-C12-3T labeled all the serine proteases as evident from MALDI-TOF MS results. We saw more than 90 % conversion with just 2 eq on the test scale. The required molecular weight changes can be seen in the reaction mixtures MALDI-TOF MS (Fig. 2.31). To monitor the extent of protein modification, the samples were directly withdrawn from the reaction mixture using a pipette and analyzed. In brief, 2 µL of the reaction mixture was mixed with 2 µL of 2 % TFA and 2 µL of matrix mixture as previously stated, vortexed and spotted on MALDI-TOF MS plate. Please note that the matrix preparation and sample preparation procedures remained the same here. Protein modifications were carried out in falcon tubes at 200 mg scale following the linear scale-up of the above-mentioned procedure for understanding the self-assembly behavior.

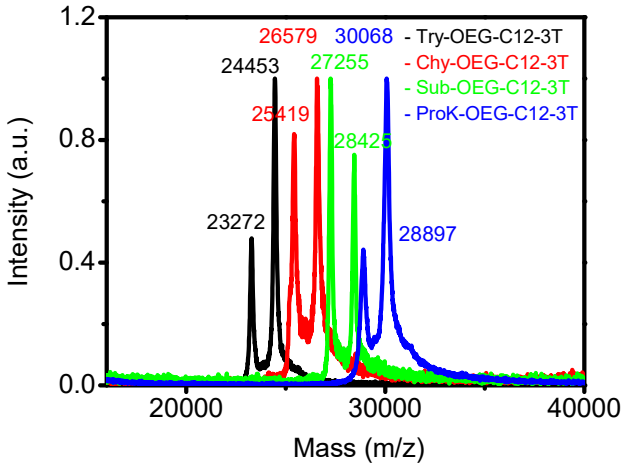


Figure 2.31. MALDI-TOF MS characterization of protein variant library before purification (reaction mixture).

All the conjugates were purified by three-step purification, as described in chapter 2.1. *i.e.,* IEX, SEC, and desalting performed using FPLC. IEX was performed to remove triton X-100 using either SP sepharose or Q sepharose resins depending on isoelectric point (pI) and surface charges of proteins (Try: 30.4, Chy: 19.5, ProK: -8.4, Sub: -6.2). For example, to purify the reaction mixture of Try or Chy, we used SP sepharose, a cation-exchange resin at pH 7.4, and to purify Sub and ProK, we used Q sepharose, an anion-exchange resin at pH 10. The obtained IEX fractions were subjected to SEC immediately to remove the native proteins from the protein conjugates in 50 mM Sod. Phos. pH 7.4, 1 M NaCl using either Sephacryl S-100 HR 16/60 or Sephacryl S-200 HR 16/60 or Sephacryl S-300 HR 16/60. The NaCl was later removed by the Sephadex-G25 desalting column. The desalted fractions were quickly lyophilized and later dissolved in the required buffer when needed.

2.2.5. Biophysical characterization

Before understanding the self-assembly behavior of the conjugates, we characterized the samples using MALDI-TOF MS, CD, Fluorescence, and zeta potential measurements. The detailed procedures for characterization are given in the experimental section. Soon after characterization, we started self-assembly studies of these conjugates. The molecular weights of FAPs obtained in MALDI-TOF MS analysis of SEC fractions are in agreement with the calcd ones and the ones obtained in reaction mixtures. Desalting and then doing MALDI-TOF MS analysis also showed the same results (Fig. 2.32). In order to probe whether the secondary and tertiary structure of made FAPs is intact, detailed fluorescence studies were carried out.[36] Fluorescence studies reveal there was no shift in tryptophan emission (295 nm) or change intensity for Prok which indicates the 3D structure of a protein is preserved in the complexes as well. In the case of Try FAP, a slight increase of emission can be observed. We observed a significant decrease in intensities in emission for any Chy FAP with respect to the native protein. This could be because of the unfolding of the protein. The fluorescence study revealed that the sub conjugate is unstable, and we observed in redshift, denaturation could be the reason, and we saw visible precipitation of conjugate (Fig. 2.33).

Similar to fluorescence study, the CD spectroscopy also revealed that the Sub conjugate is unstable (Fig. 2.34) and completely lost the signature when compared with Sub native (4C3V) is 30% helical (9 helices; 83 residues) and 19% beta-sheet (14 strands; 54 residues).

We observed no perturbation to very small perturbation in the secondary structure of the conjugates / FAPs when compared with native proteins. ProK (5B1D) is 26% helical (9 helices; 74 residues) and 24% beta-sheet (19 strands; 67 residues), Try (4MTB) is 10% helical (4 helices; 23 residues) and 35% beta-sheet (21 strands; 80 residues), Chy (1GL0) is 11% helical (5 helices; 27 residues) and 35% beta-sheet (20 strands; 87 residues). However, decreased ellipticity can be seen for conjugates, which could be because of slight concentration differences. Zeta potential measurements were also carried out to investigate the surface charge of designed protein complexes. Both Try and Chy complexes are highly cationic like native proteins, but the magnitude decreased (+16.5 mV and +8.7 mV). Similarly, the ProK complex is slightly anionic (-3.2 mV), and Sub complex is close to neutral (-0.7 mV), both showed the decreased magnitude (Fig. 2.35).

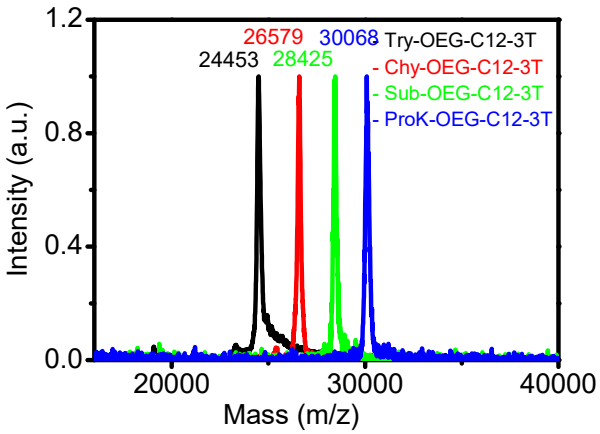


Figure 2.32. MALDI-TOF MS characterization of protein variant library after purification.

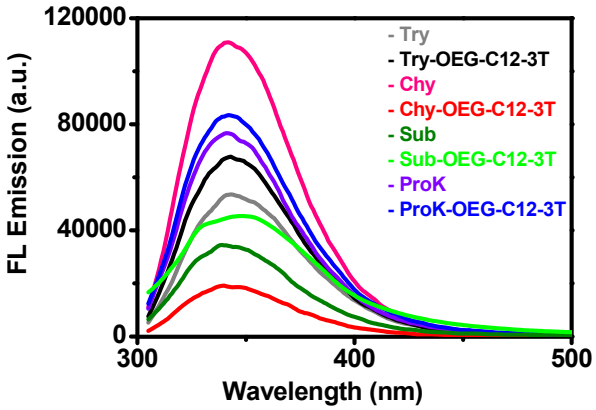


Figure 2.33. Fluorescence characterization of protein variant library after purification.

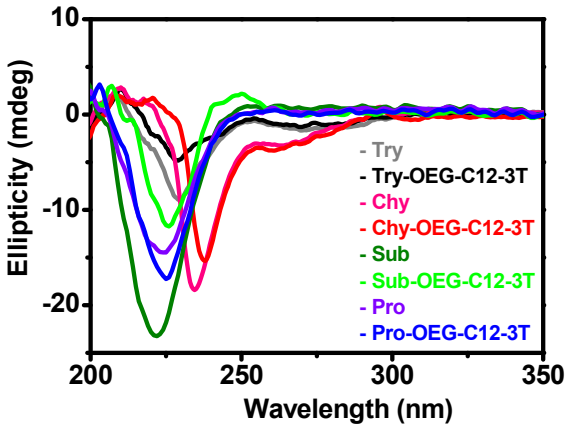


Figure 2.34. CD characterization of protein variant library after purification.

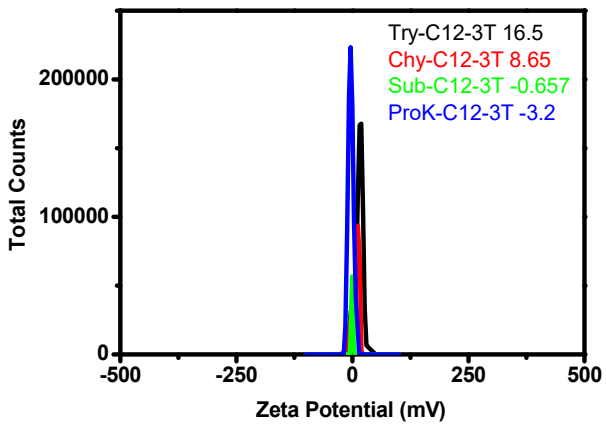


Figure 2.35. Zeta potential characterization of protein variant library after purification.

Table 2.4. Summary of characterization data of protein variants.

Mw of Native Protein (Da)	Mw of AABP-OEG-C12-3T (Da)	Mw of Conjugate Calcd (Da)	Mw of Conjugate Obtained (Da)	Zeta Potential (mV)
Try (23,272)	**1,188**	**24,460**	**24,453**	**+16.5**
Chy (25,425)	**1,188**	**26,613**	**26,579**	**+8.7**
ProK (28,909)	**1,188**	**30,097**	**30,068**	**-3.2**
Sub (27,294)	**1,188**	**28,482**	**28,425**	**-0.7**

2.2.6. Self-assembly studies

In order to test the self-assembling ability of designed protein variant FAPs, self-assembly studies were carried out using analytical SEC, DLS, SANS, and SEC-MALS. The detailed procedure for the same is given in the methods section at the end of this chapter. To determine the RMWs of soluble protein complexes, the protein variant FAPs were dissolved individually in Milli Q water, and a series of SEC runs were performed. Analytical SEC studies revealed that all four conjugates eluted at lower elution volumes (11.5 to 10.9 mL) indicative of large protein complexes, as shown below (Fig. 2.36). D_h obtained (10-16nm) indicated higher-order complex formation in all the conjugates (Fig. 2.37). Detailed SANS measurements were done to understand the oligomeric states, the radius of gyration (R_g), and shape (σ) of the complexes.[34] Both Try and Chy complexes exist as a 10 mer with the molecular mass of 245 kDa, whereas ProK conjugate self-assemble into a 22 mer with a molecular mass of 661 kDa. The R_g of protein complexes were obtained from the Guinier region of SANS data. The R_g of Try, Chy, and ProK complexes are 41 Å, 40 Å and 53 Å, respectively. R_g data showed an interesting trend similar to R_h data. We could not perform the SANS experiment for Sub complex as it precipitated during the 12 hrs period of SANS measurement. Nevertheless, it was evident from the SEC and DLS results that Sub conjugate is also capable of forming higher-order protein complex (Fig. 2.38). SEC-MALS analyses of the Try-OEG-C12-3T was performed on a Superdex-200 10/300 GL column (GE Healthcare)

connected to an Agilent HPLC system equipped with the 18-angle light scattering detector (Wyatt Dawn HELIOS II) and a refractive index detector (Wyatt OptilabTrEX) calibrated with BSA to determine the absolute molecular weights. The obtained AMW of 276 kDa is in agreement with the rest of the techniques, i.e., SEC, DLS, and SANS results (Fig. 2.39). The shape factor σ (R_g/R_h) was obtained from SANS and DLS data. The obtained σ values of all of the protein complexes are in the range of 0.60-0.79, indicating that the designed protein complexes are spherical, similar to both naturally occurring[26,35] and synthetic protein cages[7,14]. The data obtained in all the techniques are in agreement. The summary of the same is shown in the table below.

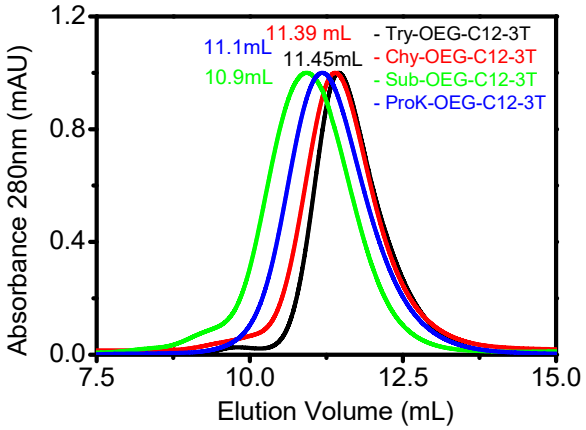


Figure 2.36. SEC data of protein variant library after purification.

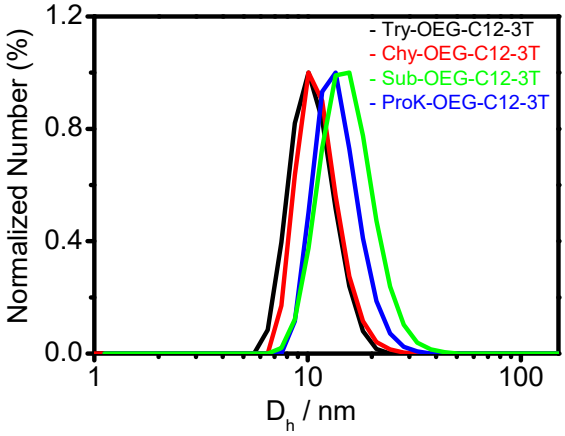


Figure 2.37. DLS data of protein variant library after purification.

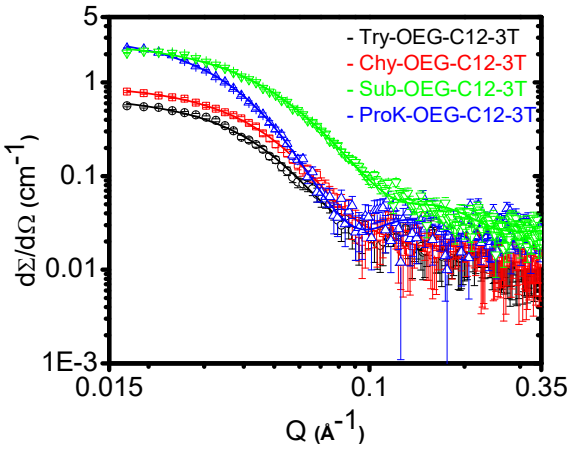


Figure 2.38. SANS data of protein variant library after purification.

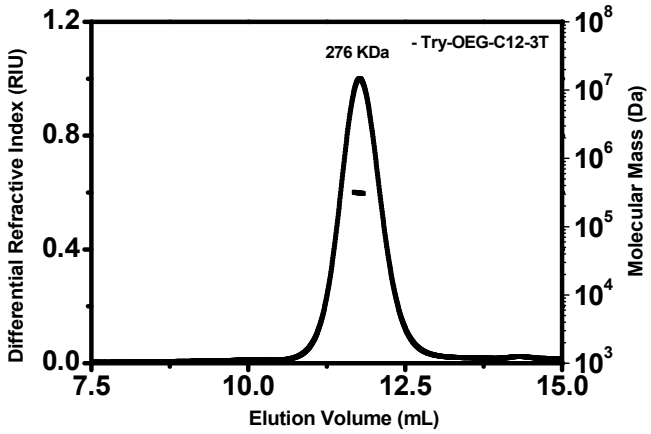


Figure 2.39. SEC-MALS data of Try-OEG-C12-3T of protein variant library.

Table 2.5. Summary of self-assembly data of protein variant library.

Protein Conjugate	Elution Vol SEC (mL)	Mol Wt -SEC (kDa)	Mol Wt SANS (kDa)	Oligomeric State (mer)	R_g from P(r) SANS (Å)	D_h DLS (nm)	Shape Factor ($\sigma = R_g/R_h$)
Try-OEG-C12-3T	**11.4**	**276**	**245**	**10**	**41**	**11**	**0.72**
Chy-OEG-C12-3T	**11.5**	**258**	**266**	**10**	**40**	**10**	**0.77**
Pro-OEG-C12-3T	**11.2**	**300**	**661**	**22**	**53**	**14**	**0.76**
Sub-OEG-C12-3T	**10.9**	**339**	**----**	**----**	**---**	**16**	**----**

Effect of pH and salt on self-assembly

In order to test whether the change in pH has any effect on self-assembly of designed FAPs,[14,35] SEC studies of Chy-OEG-C12-3T were carried out at different pH. The elution volume of the protein complex was shifted to the lower side by 0.4 mL at pH 10 (The particle seems to be larger at higher PHs). This result indicates the size of the protein complex enlarged with an increase in pH. Surprisingly, decreasing the solution pH to 5 resulted in the shrinkage of the size with respect to the dimension of a protein complex at pH 7.4. The effect of salt seems to be prominent in pH 5 (Fig. 2.41 and Fig. 2.42). At pH 10, Chy is anionic (predominantly negative); only negatively charged residues influence the oligomeric states and sizes. At pH 7.4, all the charged residues (both the positive and negative) influence the oligomeric states and sizes. At pH 5, it is cationic (predominantly positive); only positively charged residues influence the oligomeric states and sizes. The density and location of charges might influence the attractions and/or repulsions between protein subunits at any given pH. We have two possible hypotheses: **Hypothesis i:** It is possible that the net negative charges at 10 is significantly repulsive than net positive charges at pH 5, resulting in protein subunits staying away (loosely arranged) at pH 10 and might appear larger in SEC. **Hypothesis ii:** It is also possible that the net negative charges at pH 10 are significantly less

which might aid in the accommodation of more monomeric units in the complex, resulting in higher oligomeric states and so higher size. At pH 5, the significant repulsion between positively charged residues might result in complexes with lesser oligomeric states and so lesser sizes. The influence of salt and effect of pH on different protein complexes needs to studied in depth. Also, there could be time-dependent evolution of particles/ complexes, which, we have not studied /understood yet and needs to be investigated. The effect of the spacer is studied in the next subchapter 2.3.

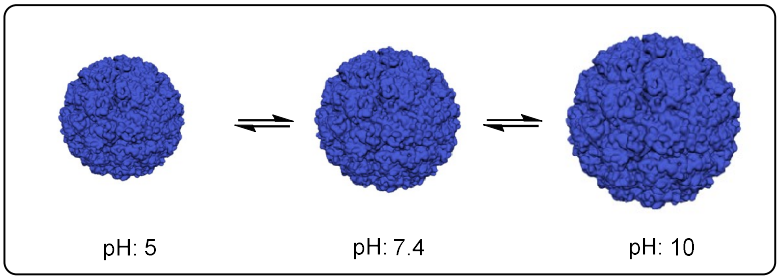


Figure 2.40. Cartoon representation of size evolution of particles with pH.

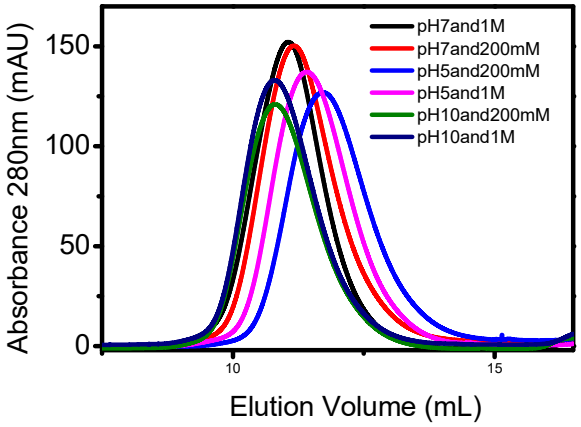


Figure 2.41. Size-evolution of particles with pH and salt as per SEC data.

2.2.7. Discussion

We have modified only 4 serine proteases using the MAPLab technology, but in principle, the method can be extended to all the serine proteases. All the protein conjugates in this library have an identical linker (OEG), and a hydrophobic tail (C12-3T) except the protein is systematically varied in the macromolecular design. We found that the effect of protein structure, size, and surface charge on self-assembly in this protein library is very significant.

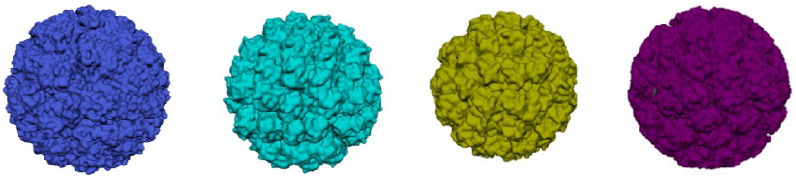


Figure 2.42. Cartoon representation of protein complexes in protein variant library.

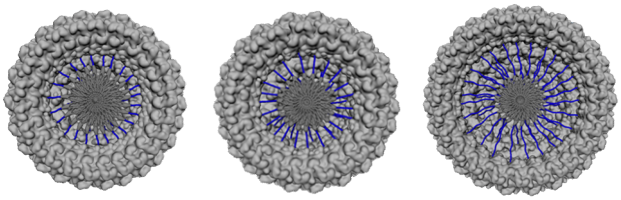
All the proteins formed higher-order complexes. Try, and Chy formed roughly 250 kDa complexes with an oligomeric state of 10. The latter two, Sub and proteins k formed 650 kDa particles with an oligomeric state of 22. The higher average oligomeric state of the latter two is quite remarkable when compared to earlier proteins (10 mer). The reason for this could be that Try and Chy are related and formed smaller, but similar complexes, Sub and ProK, are related and formed larger complexes with the higher-oligomeric state. This result can be explained based on the surface charge of proteins. ProK and Sub complexes have surface charges close to neutral, and hence the complexes can accommodate more subunits without much electrostatic repulsion. We believe that the complex formation and oligomeric state is a result of the complex interplay between opposing surface charge of proteins and attracting hydrophobicity of tails. Looking at the shape factor, the obtained complexes are spherical as expected (Theoretically: Rg/Rh = 0.66 for particles, 1.00 for vesicles, 1.77 for rods). Another interesting observation we have made here is that in all these complexes, the surface charge is retained, but the magnitude decreased almost by half. This chapter clearly showed that the protein heads have a significant effect on complex formed given the spacer and tail are the same.

The last conjugate, i.e., Sub-OEG-C12-3T, was found to be unstable. However, we were able to record SEC and DLS to confirm that it is forming a higher-order complex comparable to ProK. We believe that there is a delicate balance between microphase and macrophase separation. We are trying to stabilize this conjugate as well.

Table 2.6. Analysis of self-assembly data of protein variants.

Head Variant FAP Library	
Conjugate	**Oligomeric State (mer)**
TRY-OEG-C12-3T	**10mer, Smaller**
CHY-OEG-C12-3T	**10mer, Smaller**
SUB-OEG-C12-3T	**20mer, Larger**
PROK-OEG-C12-3T	**20mer, Larger**

2.3. Synthesis and Self-assembly Studies of Spacer Variant FAPs



2.3.1. Design

Encouraged by the results in the protein variant library, we investigated the effect of the spacer on self-assembly. The influence of hydrophobic domain location could be crucial when it comes to controlling the oligomeric state and molecular weights of protein complexes. The spacer can play a vital role in this regard; for example, if the hydrophobic domain is close to protein, we expect a different size and oligomeric state of the complex when compared with FAP, which has hydrophobic domain situated away from protein. We focused our efforts on the design and synthesis of 4 additional AABPs (Fig. 2.43) having different linker lengths, i.e., di, tetra, dodeca, and cetylethylene glycols (DEG, TEG, DDEG, CEG) apart from the OEG made in the previous subchapter.

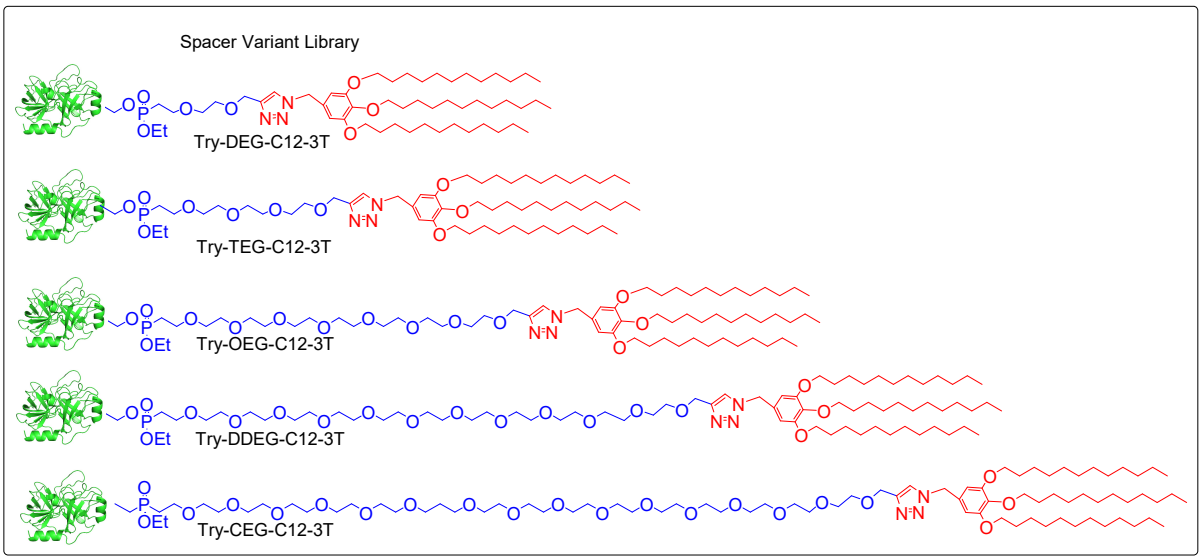
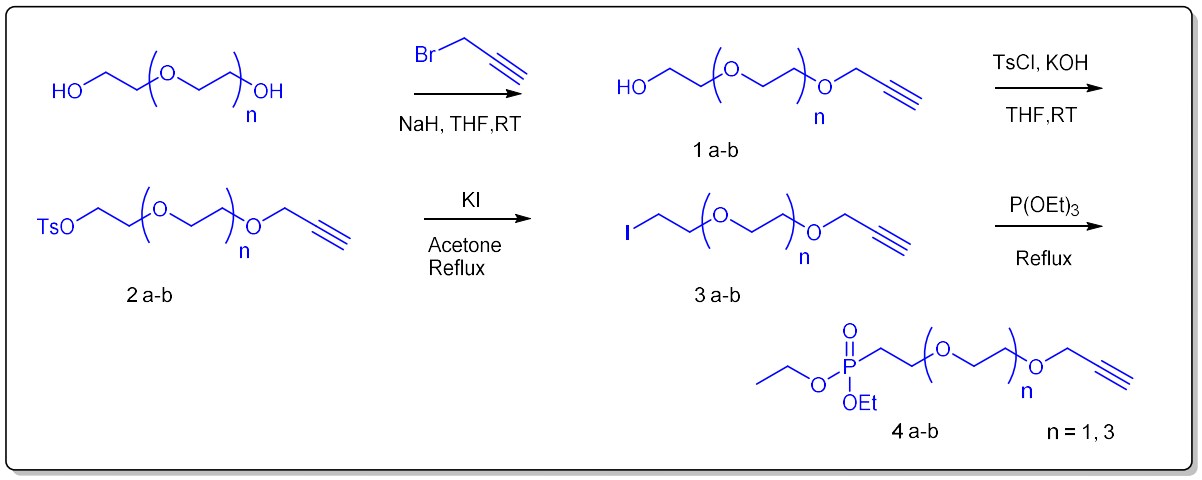


Figure 2.43. Structures of protein conjugates in the spacer variant library.

2.3.2. Synthesis

Synthesis of di and tetraethylene glycol spacers

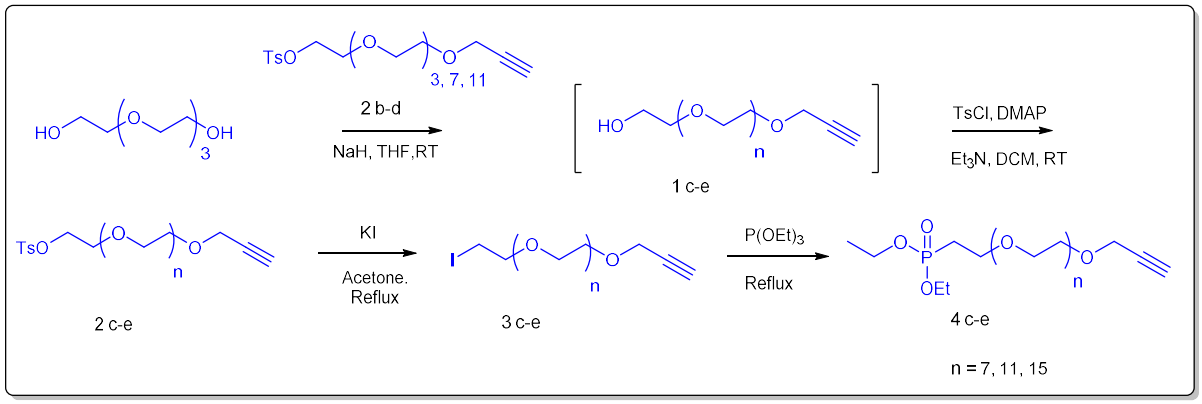


Scheme 2.13. General scheme for the synthesis of DEG or TEG spacers.

In an oven-dried RBF, DEG or TEG (1.0 eq) was dissolved with stirring in THF. NaH (1.0 eq) was added to the flask in small portions at 0 ºC. After 1 hr, ProBr (0.7 eq) was added dropwise, maintaining the reaction at the same temperature. Then, the reaction was stirred for 12 hrs at RT. Upon completion, excess NaH was quenched with dropwise addition of water. The resulting reaction mixture was extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified by NPC using MeOH / DCM as eluent. In an oven-dried RBF, monopropargylate (**1a-b**) (1.0 eq) was dissolved with stirring in THF at 0 ºC. Aqueous KOH (4 eq) was added to the flask in small portions immediately. After 10 mins, TsCl (1.2 eq) solution in THF was added dropwise and stirred for 12 hrs. Upon completion, the reaction was quenched with aqueous ammonium chloride and extracted with DCM thrice to get the crude product, which was purified by NPC using MeOH / DCM as eluent. A mixture of above-obtained tosylate (**2a-b**) (1.0 eq) and KI (4.0 eq) was refluxed in acetone for 18 hrs. Upon completion, excess KI was filtered and washed thrice with acetone. Collected acetone fraction was evaporated under vacuum to get residue, which was then washed with water and extracted with DCM. The combined organic layer was washed with aqueous Na_2CO_3 and then concentrated under vacuum to get the crude product, which was purified by NPC using MeOH / DCM as eluent. In an oven-dried RBF, iodide (**3a-b**) (1.0 eq) and $P(OEt)_3$ (4.0 eq) were taken and refluxed for 1 h. Upon completion of the reaction, the excess $P(OEt)_3$ was removed under vacuum, and

the reaction mixture was directly loaded onto a silica gel column, and the crude mixture was purified using MeOH / DCM as eluent (scheme 2.13)

Synthesis of octa, dodeca and cetylethylene glycol spacers



Scheme 2.14. General scheme for the synthesis of OEG, DDEG, and CEG spacers.

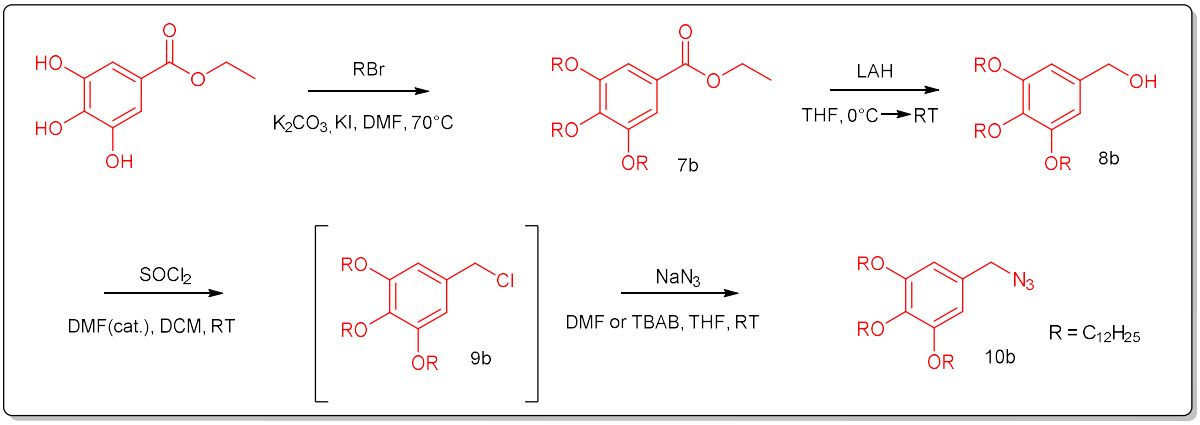
In an oven-dried RBF, a mixture of TEG (1.0 eq) and **TsO-(OEG)$_{n-4}$** (**2b-d**) (0.5 eq) was taken and dissolved in THF under stirring. The reaction mixture was cooled to 0 ºC and NaH (1.0 eq) was added in small portions. The resultant mixture was stirred at RT for 12 hrs. Upon completion of the reaction, excess of NaH was quenched by the dropwise addition of water. The resultant mixture was concentrated under reduced pressure. To the obtained residue, water was added and washed with EtOAc thoroughly. Now, the aqueous layer was concentrated under reduced pressure, and to the obtained residue, fresh EtOAc was added and filtered to remove salts. Finally, the collected EtOAc layer was dried over Na_2SO_4 and concentrated under reduced pressure to get product **HO-(OEG)$_{n+4}$** (**1c-e**) along with TEG impurity. This crude product was taken to the next step without purification. Then in an oven-dried RBF, the crude product HO-(OEG)$_{n+4}$ (**1c-e**) with TEG impurity (1.0 eq), 4-dimethylaminopyridine (DMAP) (0.5 eq) and TsCl (15 eq) were taken and dissolved in DCM under stirring. The mixture was cooled to 0 ºC, and triethylamine (Et_3N) (20 eq) was added dropwise. The resultant mixture was then allowed to react for 12 hrs at RT. Upon completion, the reaction was quenched by the dropwise addition of water and extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under reduced pressure to get a crude product, TsO-(OEG)$_{n+4}$ (**2c-e**), which was purified using NPC.

Note

1. TEG on treatment with TsO-$(OEG)_n$ and NaH in THF gives water-soluble HO-$(OEG)_{n+4}$. MALDI-TOF MS analysis of reaction mixture showed the presence of three components, i.e., excess of TEG, starting material TsO-$(OEG)_n$, and product HO-$(OEG)_{n+4}$.
2. The reaction mixture was partitioned between EtOAc and water. The product HO-$(OEG)_{n+4}$ and starting material TEG were obtained in the aqueous layer, whereas TsO-$(OEG)_n$ was present in the organic layer, which was discarded. The aqueous layer was concentrated under reduced pressure to get a mixture of HO-$(OEG)_{n+4}$ and starting material TEG.
3. Tosylation of the mixture of HO-$(OEG)_{n+4}$ and TEG gives corresponding tosylated products, i.e., TsO-TEG-OTs and TsO-$(OEG)_{n+4}$. The products were separated using NPC as they have sufficient R_f difference.

The TsO-$(OEG)_{n+4}$ (**2c-e**) on reflux in acetone in the presence of KI afforded corresponding iodide (**3c-e**) derivative. A mixture of above-obtained tosylate (**2c-e**) (1.0 eq) and KI (4.0 eq) was refluxed in acetone for 18 hrs. Upon completion, excess KI was filtered and washed thrice with acetone. Collected acetone fraction was evaporated under vacuum to get residue, which was then washed with water and extracted with DCM. The combined organic layer was washed with aqueous Na_2CO_3 and then concentrated under vacuum to get the crude product, which was purified by NPC using MeOH / DCM as eluent. Diphosphonate ester (**4c-e**) was synthesized later by refluxing iodide with $P(OEt)_3$. To brief, in an oven-dried RBF, iodide (**3c-e**) (1.0 eq) and $P(OEt)_3$ (4.0 eq) were taken and refluxed for 1 hr. Upon completion of the reaction, the excess $P(OEt)_3$ was removed under vacuum, and the crude reaction mixture was directly loaded onto an NPC and was purified using MeOH / DCM as eluent (scheme 2.14).

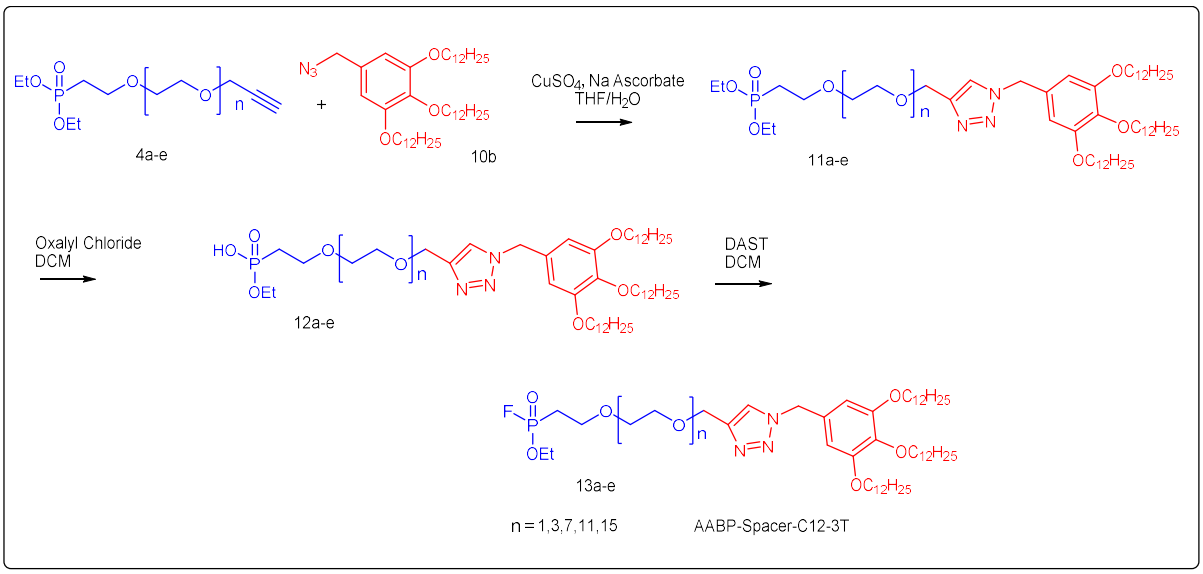
Synthesis of the hydrophobic tail



Scheme 2.15. Synthesis of C12-3T hydrophobic azide.

The synthesis of hydrophobic azide was discussed in chapter 2.1. To brief, Ethyl 3,4,5-trihydroxybenzoate was alkylated in the presence of RBr, K_2CO_3, KI to get alkylated ester (**7b**). Then in the presence of LAH, the ester was reduced to its corresponding alcohol (**8b**). This was then converted into chloride (**9b**) using $SOCl_2$ in the presence of DMF. Finally above obtained chloride (1.0 eq) in the presence of NaN_3 (1.5 eq) was converted into the corresponding azide (**10b**) (scheme 2.15)

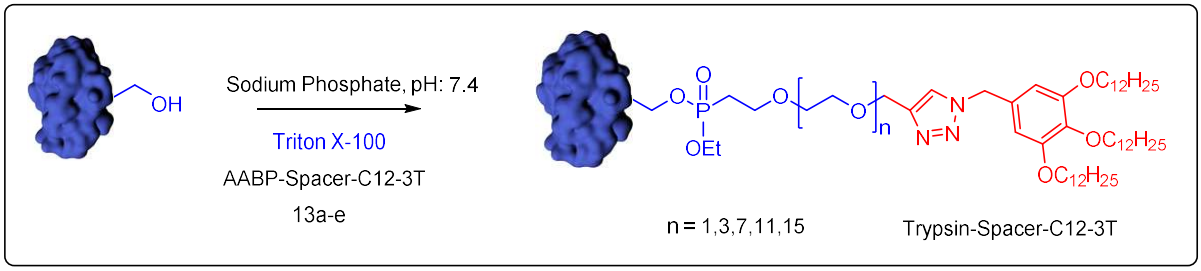
Synthesis of AABP



Scheme 2.16. Synthesis of the amphiphilic probe for making spacer variant library.

The synthesis of the AABPs in this sub-chapter was done as per the procedure mentioned in the previous sub-chapter 2.1. To brief, All the AABPs in this library were synthesized by [2 + 3] dipolar cycloaddition of the hydrophilic alkynes (**4a-e**) (diphosphonate esters) with hydrophobic azide (**10b**) in the presence of Na Asc. and $CuSO_4$ in 50 % THF / H_2O. Then, the resulting diphosphonate ester (**11a-e**) was treated with OxCl in DCM to get monophosphonate ester (**12a-e**) and finally fluorinated using DAST in DCM. These final AABPs (**13a-e**) were used for protein modification without further purification (scheme 2.16)

2.3.3. Conjugation and purification



Scheme 2.17. General conjugation scheme for making spacer variant library.

Try modification was carried out at the concentration of 100 µM, as in previous chapters. Both test scale and scale-up conjugations were done following the procedure mentioned in chapter 2.2. All the AABPs reacted except AABP with DEG (shortest linker) (scheme 2.13 and Fig. 2.44). We believe that the reason could be the unavailability of the active site of protein for this probe because of the sub-optimal linker length. To monitor the extent of modification, the samples were directly withdrawn from the reaction mixture using a pipette and analyzed using the protocol mentioned in chapter 2.1.

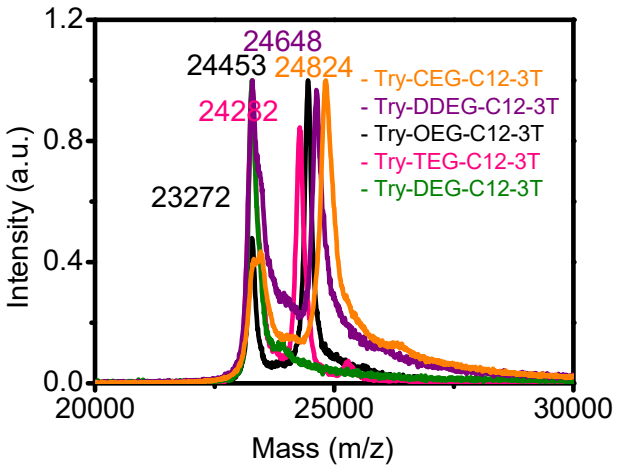


Figure 2.44. MALDI-TOF MS characterization of spacer variant library before purification (reaction mixture).

All the FAPs were purified using three-step purification, *i.e.,* IEX, SEC, and desalting, performed using FPLC as described earlier. As all the conjugates in this library are derived from Try (+30.4 mV), so these conjugates were purified by CEX in the first step using SP sepharose. Then SEC and desalting were performed. The complete purification protocol is mentioned in chapter 2.2. Monitoring of IEX, SEC, and desalted fractions was also carried out using the same procedure mentioned for native proteins and reaction mixtures except for the addition of triton X-100 in protein fractions.

2.3.4. Biophysical characterization

Before understanding the self-assembly behavior of the conjugates, we characterized the samples using MALDI-TOF MS, CD, fluorescence, and Zeta Potential measurements, as discussed in the previous chapter. The detailed procedures for characterization are given in the experimental section. The molecular weights of 4 FAPs, (we could not make conjugate with shortest linker length) obtained in MALDI-TOF MS analysis are in agreement with the calcd ones (Fig. 2.45). In order to probe if the secondary and tertiary structure of made FAPs is still intact, a detailed Fluorescence study was carried out.[36] This revealed that there was no shift in tryptophan emission maximum (295 nm), or change in intensities for the FAPs with respect to the native proteins. This indicates the 3D structure of a protein is preserved in the complex state, just like in the native state (Fig. 2.46). However, CD spectroscopy[37] (195-250 nm) indicates a very small to no perturbation in the secondary structure of some conjugates (Fig. 2.47). Finally, as part of characterization, zeta potential measurements were carried out to investigate the surface charge of designed protein complexes like in previous chapters. Since these are Try conjugates, we expected a positive zeta potential, and as expected, all the complexes are cationic (+16.5 mV to +4.6 mV) (Fig. 2.48).

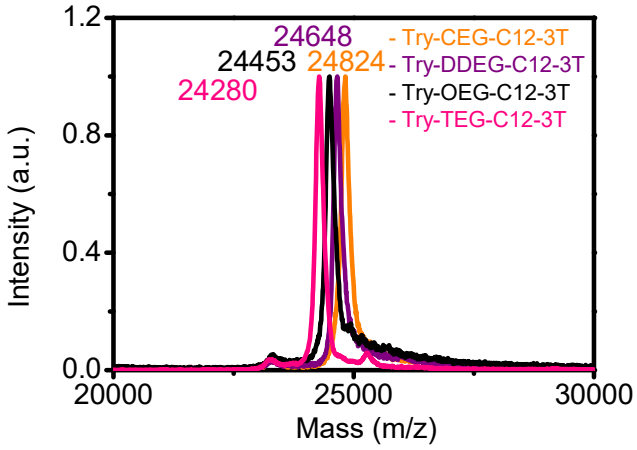


Figure 2.45. MALDI-TOF MS characterization of spacer variant library after purification.

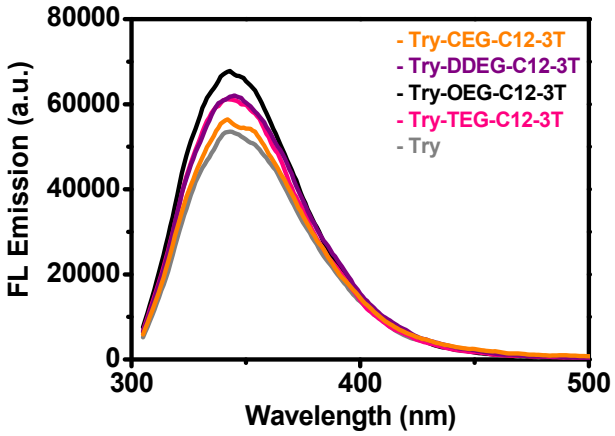


Figure 2.46. Fluorescence characterization of spacer variant library, after purification.

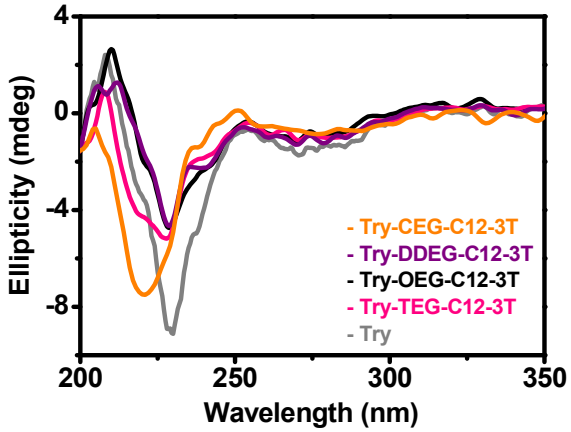


Figure 2.47. CD characterization of spacer variant library, after purification.

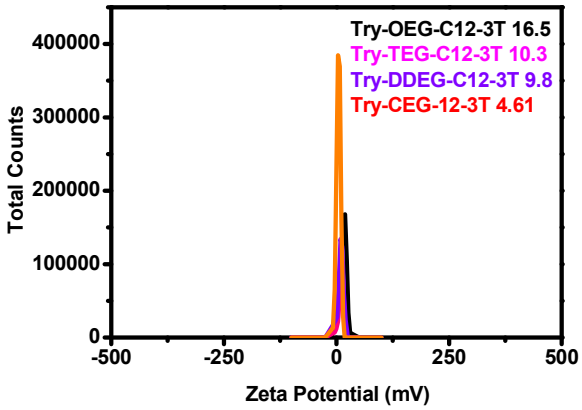


Figure 2.48. Zeta potential characterization of spacer variant library after purification.

Table 2.7. Summary of characterization data of spacer variants FAPs.

MW of AABPs (Da)	MW of Try (Da)	MW of Conjugates Calc. (Da)	MW of Conjugates Obtained (Da)	Zeta Potential (+mV)
AABP-DEG-C12-3T (923)	23,272	24,195	----------	---------
AABP-TEG-C12-3T (1,011)	23,272	24,283	24282	10.3
AABP-OEG-C12-3T (1,188)	23,272	24,460	24,445	16.5
AABP-DDEG-C12-3T (1,363)	23,272	24,635	24648	9.8
AABP-CEG-C12-3T (1,541)	23,272	24,813	24824	4.6

2.3.5. Self-assembly studies

After characterization, we started exploring the self-assembly behavior of these conjugates, in order to understand the effect of linker length. The detailed procedure for measuring SEC, SEC-MALS, SANS, and DLS is given in the Methods section at the end of this chapter. To determine the RMWs of soluble protein complexes, the protein conjugates were dissolved individually in Milli Q water, and a series of SEC runs were performed. All the spacer variant

FAPs formed larger protein complexes with elusion volumes 10.55 - 11.39 mL. Both CEG and DDEG (FAPs with larger linker length) formed higher-order complexes than OEG and TEG conjugates (FAPs with shorter linker length) (Fig. 2.49). The mean size of the protein complexes was measured using DLS using Zetasizer Nano 2590 (Malvern, UK). DLS results also showed a similar trend like SEC results, with D_h ranging from 11 - 14 nm (Fig. 2.50). Then, SANS measurements were done to understand the oligomeric states of the complexes. SANS studies revealed that the molecular mass of Try-TEG-C12-3T with the shortest linker (TEG) is 243 kDa with oligomerization state as 10 whereas Try-CEG-C12-3T with the longest linker (CEG) is 372 kDa with oligomerization state as 15. As predicted, the R_g of protein complexes also increased with an increase in spacer length. The size of Try-TEG-C12-3T is 32 Å, whereas Try-CEG-C12-3T is 49 Å (Fig. 2.51). SEC-MALS analyses of the FAP, Try-CEG-C12-3T was performed on a Superdex-200 10/300 GL column (GE Healthcare) connected to an Agilent HPLC system equipped with the 18-angle light scattering detector (Wyatt Dawn HELIOS II) and a refractive index detector (Wyatt OptilabTrEX) calibrated with BSA to determine the AMW. The molecular mass of Try-CEG-C12-3T was found to be 400 kDa (Fig. 2.52). All the protein complexes are narrow-disperse, as evident from SEC and SANS results. The obtained data from various studies are in agreement with each other.

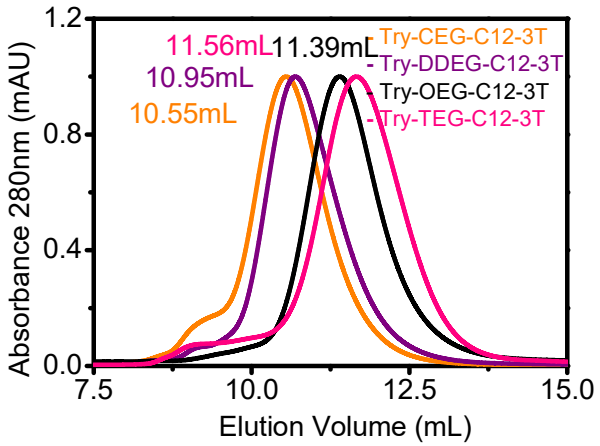


Figure 2.49. SEC data of spacer variant library after purification.

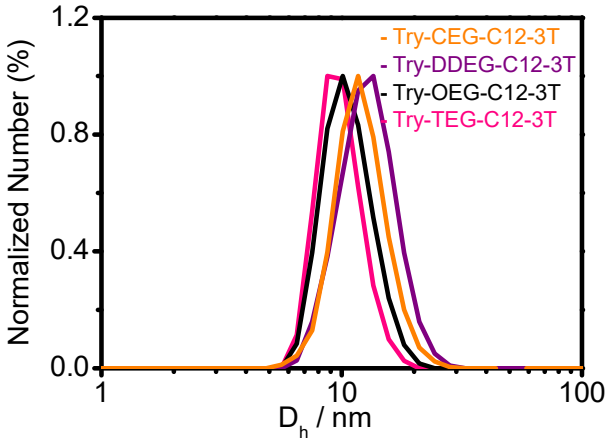


Figure 2.50. DLS data of the spacer variant library after purification.

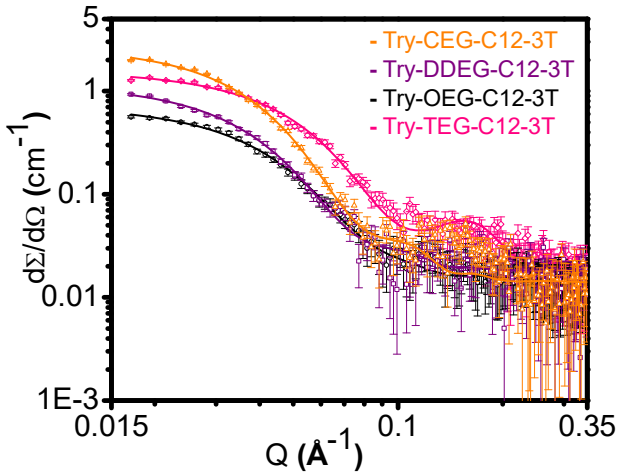


Figure 2.51. SANS data of spacer variant library after purification.

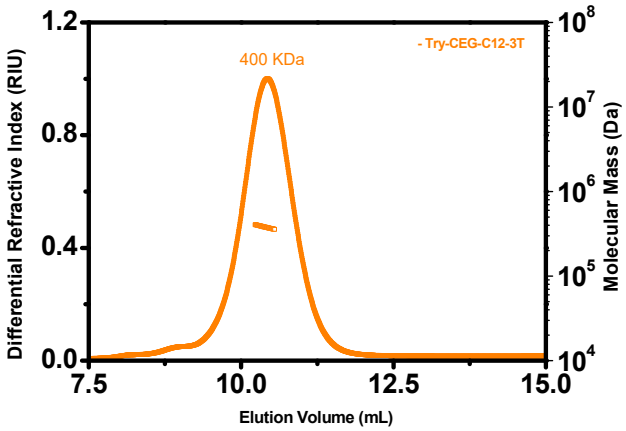


Figure 2.52. SEC-MALS data of spacer variant library after purification.

Table 2.8. Summary of self-assembly data of spacer variants.

Protein Conjugate	Elution Vol SEC (mL)	Mol Wt SEC (kDa)	Mol Wt SANS (kDa)	Oligo State (mer)	R_g from P(r) SANS (Å)	D_h DLS (nm)	Shape Factor ($\sigma = R_g/R_h$)
Try-TEG-C12-3T	**11.7**	**241**	**243**	**10**	**32**	**11**	**0.60**
***Try-OEG-C12-3T**	**11.4**	**276**	**245**	**10**	**41**	**11**	**0.72**
Try-DDEG-C12-3T	**10.7**	**375**	**320**	**13**	**49**	**14**	**0.67**
Try-CEG-C12-3T	**10.6**	**388**	**372**	**15**	**49**	**14**	**0.73**

2.3.6. Discussion

To summarize this library, FAPs in this library have an increasing number of tetraethylene glycol units, one at a time down the series. We could not make di because of the suboptimal linker. All the made FAPs formed higher-order complexes.

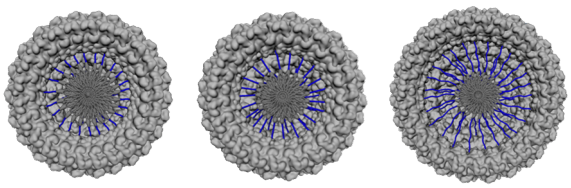


Figure 2.53. Cartoon representation of protein complexes in spacer variant library.

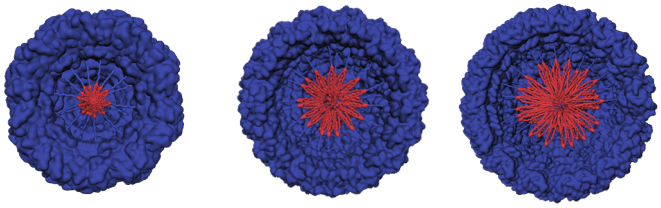
TEG and OEG formed 250 kDa particles with an oligomeric state of 10. Then the oligomeric state increased by 3 and 5 for DDEG and CEG complexes. Most importantly, the σ for these conjugates has a value of 0.7, emphasizing that these are also spherical assemblies. For the first time, we were able to engineer the protein complexes with such accuracy using chemical methods. It seems like, with increasing spacer length, the complex can accommodate more

monomers; consequently, the complex size, oligomeric state, and MW is increasing down the series. This data set reveals that the spacer does have a significant role to play. The next sub-chapter is the final library in this series, in which we made tail variant FAPs.

Table 2.9. Analysis of self-assembly data of spacer variants.

Spacer Variant FAP Library	
Conjugate	**Oligomeric State (mer)**
TRY-DEG-C12-3T	**Couldn't Make**
TRY-TEG-C12-3T	**10mer, Similar Attributes**
TRY-OEG-C12-3T	**10mer, Similar Attributes**
TRY-DDEG-C12-3T	**13mer, Increased Size and Oligomeric State**
TRY-CEG-C12-3T	**15mer, Further Increase in Size**

2.4. Synthesis and Self-assembly Studies of Tail Variant FAPs



2.4.1. Design

This sub-chapter studies the effect of the hydrophobic tail on self-assembly of FAPs upon systematic variation. We aimed to determine the minimum hydrophobicity needed for complex formation for a given protein and to determine the influence of branching in hydrophobicity. The influence of hydrophobicity is vital when it comes to controlling the oligomeric state and molecular weights of protein complexes. To demonstrate the systematic variation of the hydrophobic tails, 7 new AABPs (Fig. 2.54) were synthesized with varying alkyl chain length (hexyl, dodecyl, and octadecyl abbreviated as C6/C12/C18) and branching (1-tail/2-tails/3-tails abbreviated as 1T/2T/3T).

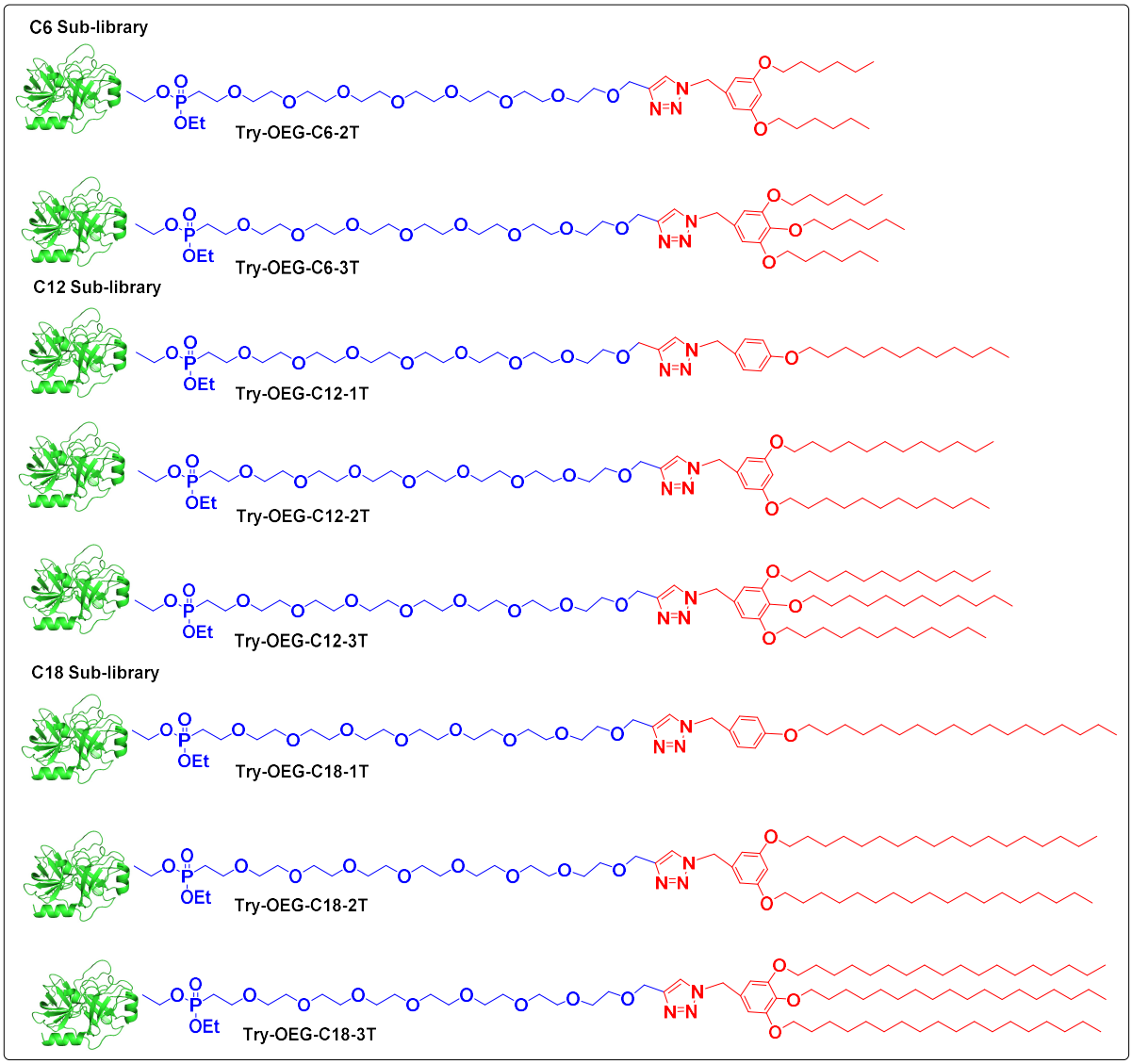
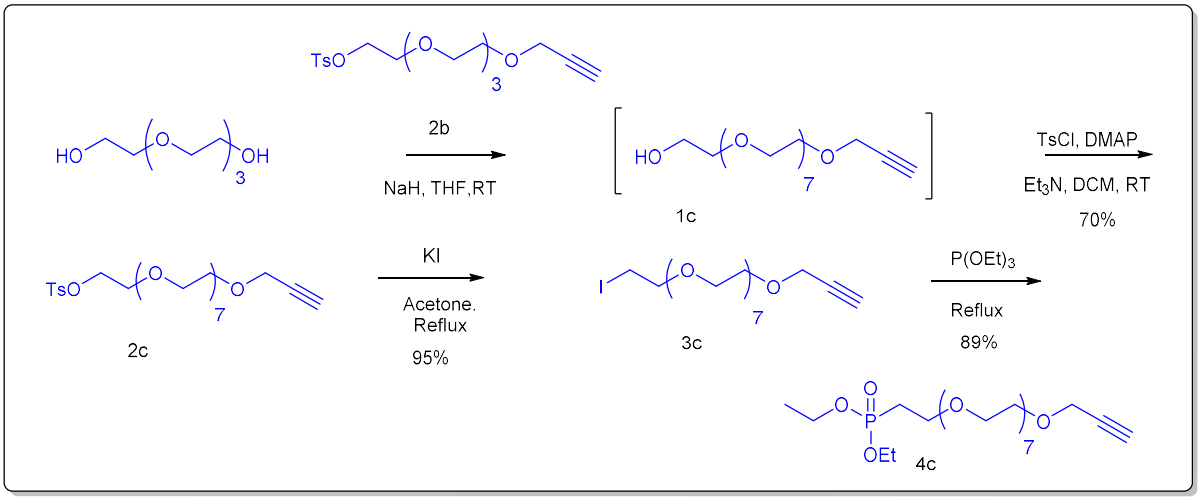


Figure 2.54. Structures of conjugates in the tail variant library.

2.4.2. Synthesis

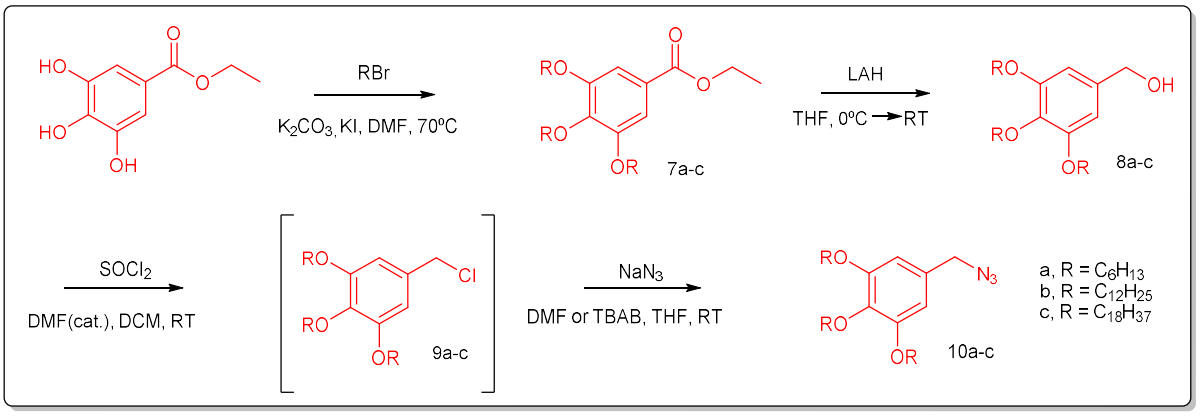
Synthesis of the hydrophilic spacer



Scheme 2.18. Synthesis of the OEG hydrophilic spacer

As discussed in chapter 2.1, the required diphosphonate ester (**4c**) was made in the following steps. Synthesis of the hydrophilic spacer (**1c**) was made from tetraethylene glycol (TEG) and the tosylate (**2b**) in the presence of NaH to get mono propargyl derivative (**1c**). This monopropargylate was tosylated in the presence of KOH. The obtained tosylate (**2c**) was refluxed in acetone for with KI to get the corresponding iodide. Finally, in the presence of $P(OEt)_3$, the iodide was converted into diphosphonate ester (**4c**) (scheme 2.18).

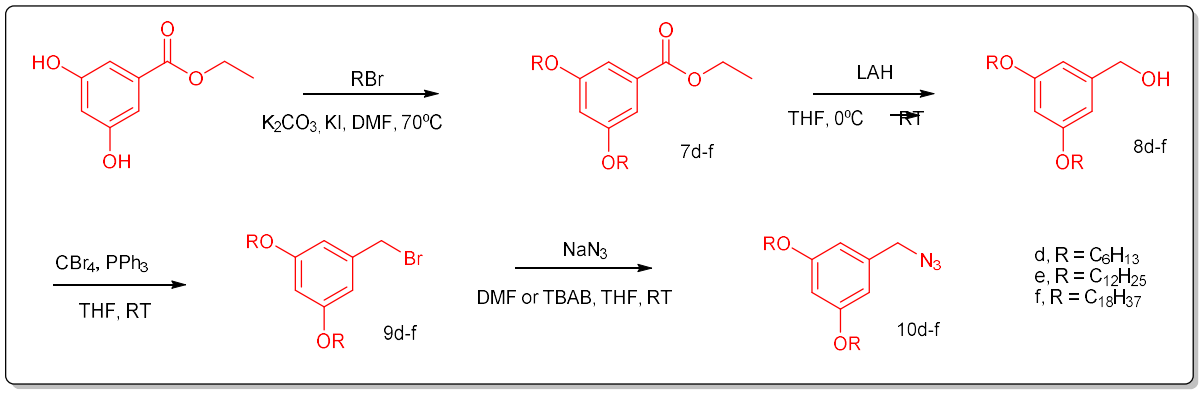
Synthesis of 3-Tail hydrophobic tails



Scheme 2.19. General scheme for the synthesis of the 3-Tail hydrophobic tail.

Ethyl 3,4,5-trihydroxybenzoate (1.0 eq), RBr (3.5 eq), K_2CO_3 (3.5 eq), KI (0.05 eq) were taken in an oven-dried RBF. The flask was then purged with nitrogen, and DMF was added under stirring to dissolve the mixture and heated at 70 ºC for 12 hrs. Upon completion, the reaction mixture was neutralized with acidic water and extracted thrice with EtOAc. The combined organic layer was dried over Na_2SO_4 and evaporated under vacuum to get the crude product, which was purified using NPC using EtOAc / PE. In an oven-dried RBF, above ester (**7a-c**) (1.0 eq) was taken and dissolved in THF with stirring at 0 ºC. Then LAH (3 eq) was added in portions, maintaining the reaction temperature at 0 ºC and after 30 mins, allowed to react at RT for 12 hrs. Upon completion, LAH was quenched by the dropwise addition of EtOAc at 0 ºC (transesterification was observed if the temperature was raised). Water was added to the reaction mixture and stirred until color changes to off-white. The resulting solution was filtered using Buchner funnel, and the residue was washed with EtOAc. The EtOAc layer was separated from water, and the residual water layer was again extracted thrice with EtOAc. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified by NPC using EtOAc / PE. In an oven-dried RBF, the above-obtained alcohol (**8a-c**) (1.0 eq) was dissolved in DCM, and a catalytic amount of DMF was added. Then, $SOCl_2$ (1.6 eq) was added dropwise and allowed to react under stirring for 30 mins at RT. Upon completion of the reaction, DCM and excess of $SOCl_2$ were evaporated under vacuum. The residue was dissolved in Et_2O and washed thrice with water. The Et_2O layer was then dried over Na_2SO_4 and concentrated under vacuum. The resulting crude product was used for next step without any purification. In an oven-dried RBF, a mixture of above-obtained chloride (**9a-c**) (1.0 eq), NaN_3 (1.5 eq) was taken, and DMF was added. The resulting mixture was stirred for 12 hrs at RT. Upon completion, water was added to the reaction mixture and extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified using NPC using EtOAc / PE (scheme 2.19).

Synthesis of 2-Tail hydrophobic tails

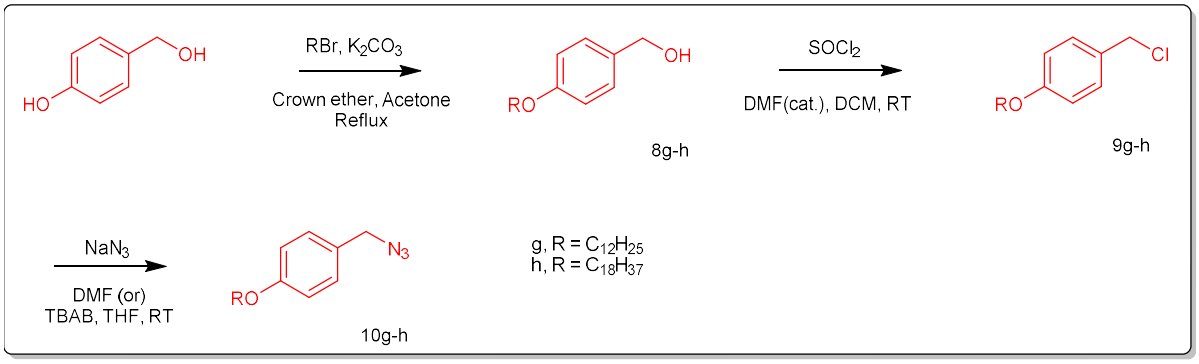


Scheme 2.20. General scheme for the synthesis of the 2-Tail hydrophobic tail.

Ethyl 3,5-dihydroxybenzoate (1.0 eq), RBr (2.5 eq), K_2CO_3 (3.5 eq), KI (0.05 eq) were taken in an oven-dried RBF. The flask was then purged with N_2, DMF was added under stirring to dissolve the mixture and heated at 70 ºC for 12 hrs. Upon completion, the reaction mixture was neutralized with acidic water and extracted thrice with EtOAc. The combined organic layer was dried over Na_2SO_4 and evaporated under vacuum to get the crude product, which was purified by NPC using EtOAc / PE. Then, in an oven-dried RBF, above ester (**7d-f**) (1.0 eq) was taken and dissolved in THF with stirring at 0 ºC. Then LAH (3 eq) was added in portions, maintaining the reaction temperature at 0 ºC. After 30 mins, allowed to react at RT for 12 hrs. Upon completion, LAH was quenched by the dropwise addition of EtOAc at 0 ºC (transesterification was observed if the temperature was raised). Water was added to the reaction mixture and stirred until color changes to off-white. The resulting solution was filtered using Buchner funnel, and the residue was washed with EtOAc. The EtOAc layer was separated from water, and the residual water layer was again extracted thrice with EtOAc. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified by NPC using EtOAc / PE. After this, in an oven-dried RBF, the above-obtained alcohol (**8d-f**) (1 eq) and PPh_3 (2.0 eq) were taken and dissolved with stirring in THF. Then, the solution of tetrabromomethane (CBr_4) (2.0 eq) in THF was added dropwise and stirred for 1 hr at RT. Upon completion, THF was evaporated under vacuum. Water was added to the residue and extracted thrice with EtOAc. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified using NPC using EtOAc / PE. In an oven-dried RBF, the mixture of above-obtained bromide (**9d-f**) (1.0 eq), NaN_3 (1.5 eq) was taken, and DMF was added.

The resulting mixture was stirred for 12 hrs at RT. Upon completion, water was added to the reaction mixture and extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified using NPC using EtOAc / PE (scheme 2.20).

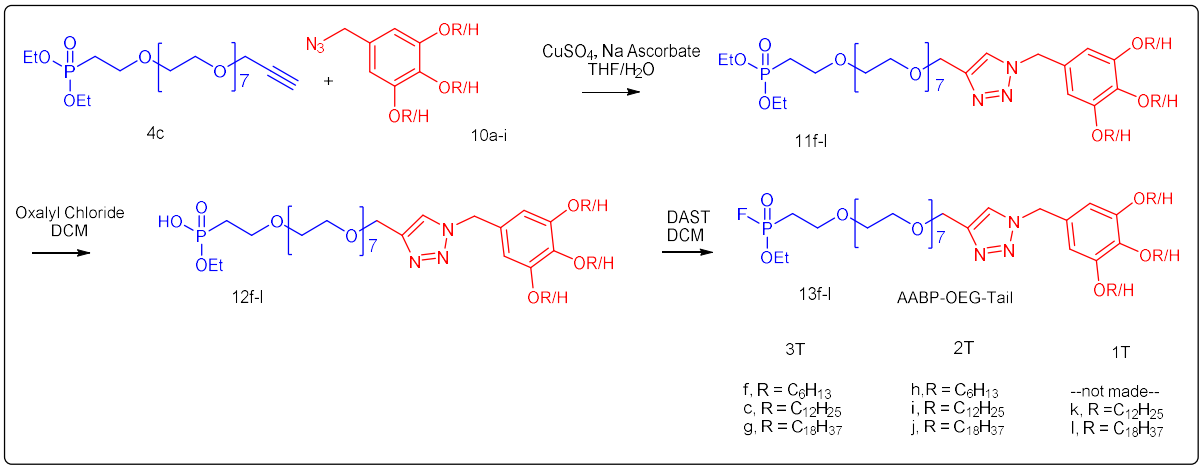
Synthesis of 1-Tail hydrophobic tails



Scheme 2.21. General scheme for the synthesis of the 1-Tail hydrophobic tail.

4-Hydroxybenzyl alcohol (1.0 eq), RBr (1.1 eq), K_2CO_3 (1.2 eq), crown ether (0.1 eq) were taken in an oven-dried RBF. The flask was then purged with N_2, and acetone was added under stirring to dissolve the mixture and refluxed for 24 hrs. Upon completion, the reaction mixture was concentrated and extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and evaporated under vacuum to get the crude product, which was purified using NPC using EtOAc / PE. Then, in an oven-dried RBF, the above-obtained alcohol (**8g-h**) (1.0 eq) was dissolved in DCM, and a catalytic amount of DMF was added. Then, $SOCl_2$ (1.6 eq) was added dropwise and allowed to react under stirring for 30 mins at RT. Upon completion of the reaction, DCM and excess of $SOCl_2$ were evaporated under vacuum. The residue was dissolved in Et_2O and washed thrice with water. The Et_2O layer was then dried over Na_2SO_4 and concentrated under vacuum. The resulting crude product was used for next step without any purification. In an oven-dried RBF, the mixture of above-obtained chloride (**9g-h**) (1.0 eq), NaN_3 (1.5 eq) was taken, and DMF was added. The resulting mixture was stirred for 12 hrs at RT. Upon completion, water was added to the reaction mixture and extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified using NPC using EtOAc / PE (scheme 2.21).

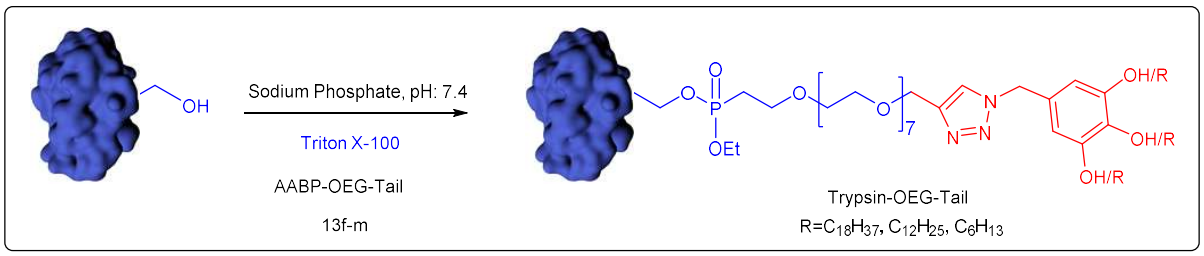
Synthesis of AABP



Scheme 2.22. Synthesis of the amphiphilic probe for making tail variant library.

The synthesis of the AABPs in this sub-chapter was done as per the procedure mentioned in the previous sub-chapter 2.1. To brief, All the AABPs in this library were synthesized by [2 + 3] dipolar cycloaddition of the hydrophilic alkyne (**4c**) (diphosphonate esters) with hydrophobic azides (**10a-l**) in the presence of Na Asc. and $CuSO_4$ in 50 % THF / H_2O. Then, the resulting diphosphonate ester (**11f-l**) was treated with OxCl in DCM to get monophosphonate ester (**12f-l**) and finally fluorinated using DAST in DCM (scheme 2.22). These final AABPs (**13f-l**) were used for protein modification without further purification.

2.4.3. Conjugation and purification



Scheme 2.23. General conjugation scheme for making tail variant library.

Both test scale and scale-up conjugations of Try were carried out as per the protocol mentioned in the previous chapter 2.1. The AABPs FP-OEG-C18-2T and FP-OEG-C18-3T, was used at 10 eq, which was just 2 eq for other AABPs. 4 % triton X-100 was added for these probes, which was just 2 % for other probes. Bath sonicator was used to solubilize these

two AABPs as were difficult to solubilize with just vortexing. All the probes reacted, including the most hydrophobic one, i.e., FP-OEG-C18-3T (scheme 2.18, Fig. 2.55, Fig. 2.56 and Fig. 2.57). To monitor the extent of Try modification, the samples were directly withdrawn from the reaction mixture using a pipette and analyzed, as mentioned in chapter 2.1.

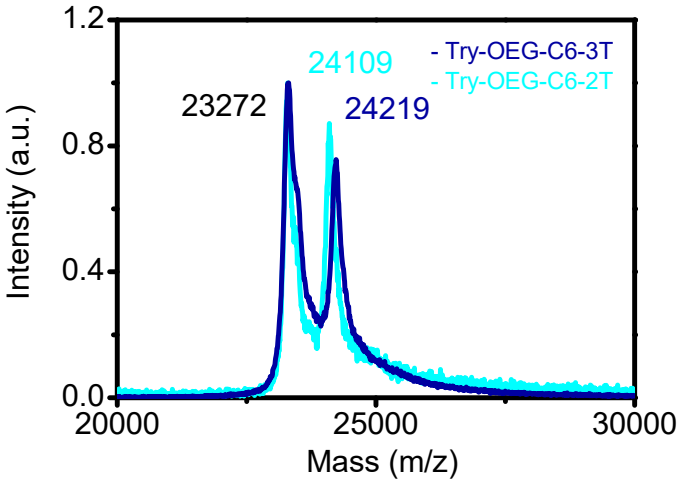


Figure 2.55. MALDI-TOF MS characterization of C6-tail variant library before purification (reaction mixtures).

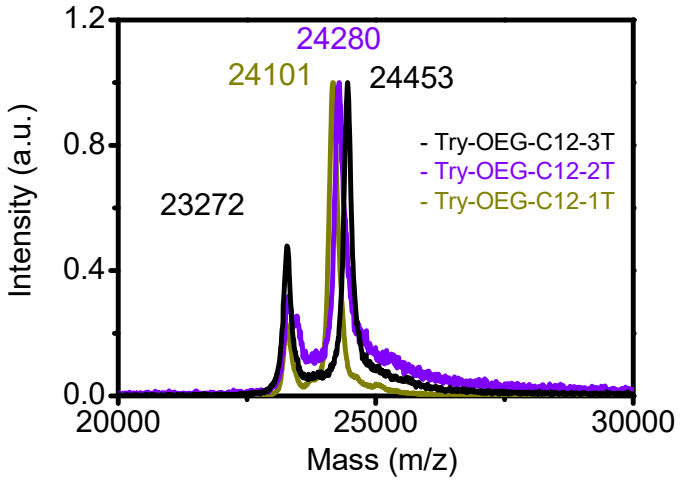


Figure 2.56. MALDI-TOF MS characterization of C12-tail variant library before purification (reaction mixtures).

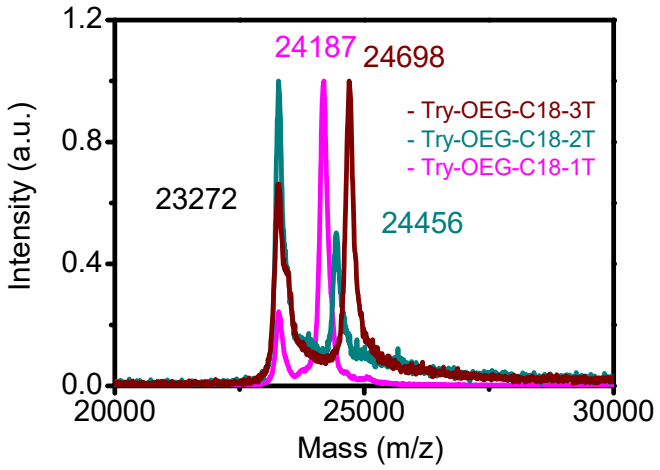


Figure 2.57. MALDI-TOF MS characterization of C18-tail variant library before purification (reaction mixtures).

All the FAPs were purified by three-step purification, *i.e.,* IEX, SEC, and desalting performed using FPLC, as described earlier. As all the conjugates in this library are of Try (+30.4 mV) and so these conjugates were purified by CEX in the first step using SP sepharose. Then SEC and desalting were performed. The complete purification protocol is mentioned in chapter 2.1. Monitoring of IEX, SEC, and desalted fractions were carried out using the same procedure mentioned for native proteins and reaction mixtures except for the addition of triton X-100 in protein fractions.

2.4.4. Biophysical characterization

Before understanding the self-assembly behavior of the conjugates, we characterized the samples using MALDI-TOF MS, CD, fluorescence, and zeta potential measurements like in precious sub-chapters. The detailed procedures for characterization are given in the experimental section. The molecular weights of conjugates obtained in MALDI-TOF MS analysis are in agreement with the calcd ones and ones obtained for reaction mixtures (Fig. 2.58 and Fig. 2.59). The secondary and tertiary structures of designed protein conjugates/complexes are intact when studied by fluorescence.[36] Denaturation might result in considerable redshift, and unfolding might result in decreased intensity. But we have observed any significant increase in the fluorescence emission for Try-OEG-C12-3T, Try-OEG-C18-1T, Try-OEG-C6-3T. This could be because of increased order, which needs further investigation (Fig. 2.60). We did not observe significant perturbation in the secondary

structure of the conjugates in CD spectroscopy (195-250 nm) studies (Fig. 2.61).[37] Zeta potential measurements were also carried out to investigate the changes in surface charges of made FAPs. Since all the conjugates are from Try, we observed that all the conjugates/complexes are cationic but varying (+16.5 mV and +5.5 mV). We couldn't understand the changes in zeta values at this point. A detailed investigation is needed in this regard (Fig. 2.62).

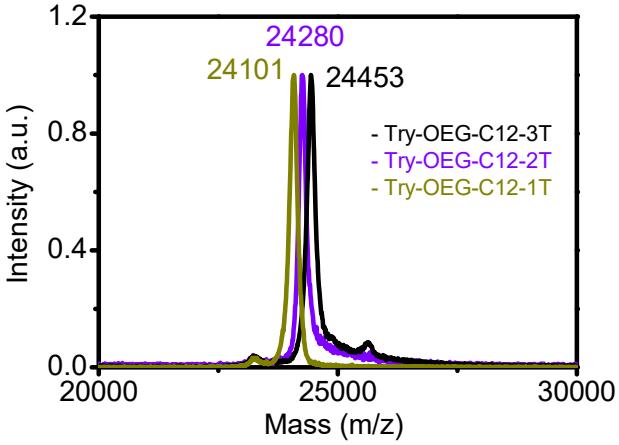


Figure 2.58. MALDI-TOF MS characterization of C12-tail variant library after purification.

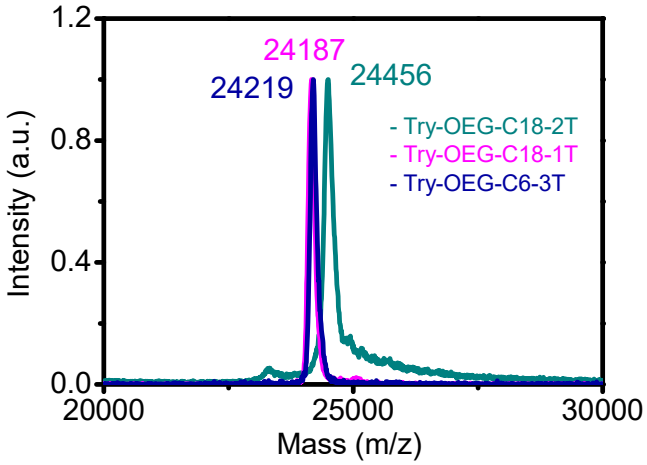


Figure 2.59. MALDI-TOF MS characterization of C6, C18-tail variant libraries after purification.

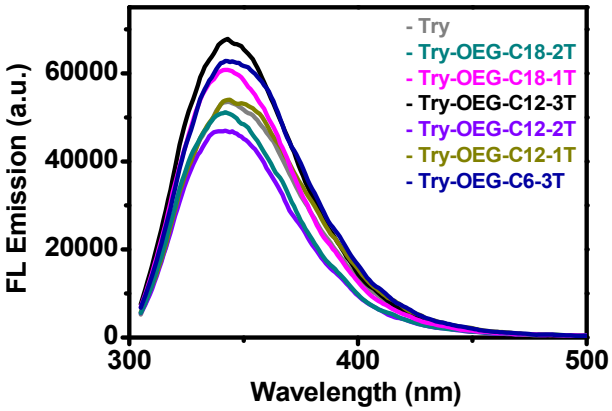


Figure 2.60. Fluorescence characterization of the tail variant library after purification.

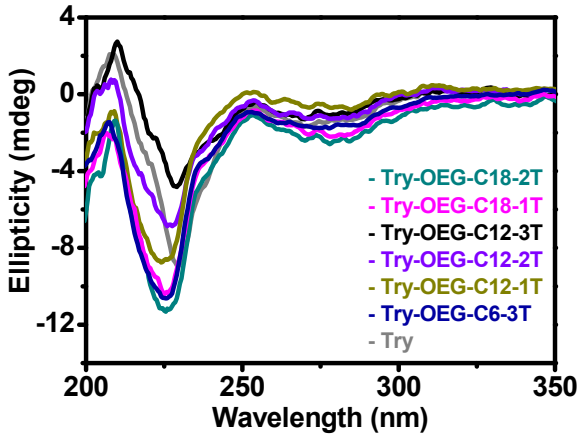


Figure 2.61. CD characterization of the tail variant library after purification.

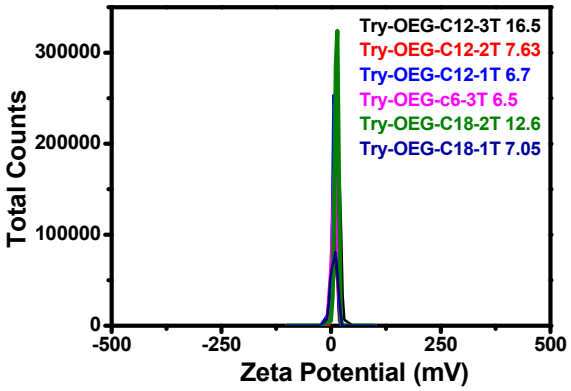


Figure 2.62. Zeta potential characterization of the spacer variant library after purification.

Table 2.10. Summary of characterization data of the tail variant FAPs.

Mw of AABPs (Da)	MW of Native Try (Da)	MW of Conjugate Calcd (Da)	MW of Conjugate Obtained (Da)	Zeta Potential (+mV)
AABP-OEG-C6-2T (835)	**23,272**	**24,107**	**24,146**	**--------**
AABP-OEG-C6-3T (936)	**23,272**	**24,208**	**24,219**	**6.5**
AABP-OEG-C12-1T (819)	**23,272**	**24091**	**24101**	**5.5**
AABP-OEG-C12-2T (1,004)	**23,272**	**24,276**	**24,280**	**7.6**
AABP-OEG-C12-3T (1,188)	**23,272**	**24,460**	**24,453**	**16.5**
AABP-OEG-C18-1T (904)	**23,272**	**24176**	**24187**	**7.1**
AABP-OEG-C18-2T (1,172)	**23,272**	**24,444**	**24,456**	**12.6**
AABP-OEG-C18-3T (1441)	**23,272**	**24,713**	**24,698**	**--------**

2.4.5. Self-assembly studies

The detailed procedure for measuring SEC, SEC-MALS, DLS, and SANS is given in the methods section at the end of this chapter. The results from the self-assembly studies of tail variant FAP's is given below. To determine the RMWs of soluble protein complexes, the protein conjugates were dissolved individually in Milli Q water, and a series of SEC runs were performed like in the previous chapter. Remarkably, the protein conjugate with just one hydrophobic tail (Try-OEG-C18-1T), self-assembled into higher-order protein complex in SEC studies. As expected, Try-OEG-C12-2T, Try-OEG-C12-3T, and Try-OEG-C18-2T self-assembled into higher-order protein complexes as expected (Fig. 2.63). On the other hand, control design Try-OEG-C6-3T with short hydrophobic tail predominantly existed as monomeric form at 800 μM concentration as predicted. Surprisingly Try-OEG-C12-1T also existed as a monomer in SEC studies. However, in both these conjugates, we observed a

higher-order complex like a peak. This peak could be mono labeled conjugate, which is in equilibrium with a monomer, which is predominant. We could not confirm it as conjugate using MALDI-TOF MS. It could also be a double-labeled conjugate. A detailed investigation is needed in this regard (Fig. 2.64). As expected, DLS studies showed that protein conjugates with sufficient tail length and hydrophobicity (Try-OEG-C12-2T/3T and Try-OEG-C18-1T/2T) self-assemble into large protein complexes. Surprisingly Try-OEG-C12-1T also behaved as a higher-order complex in DLS. However, Try-OEG-C6-3T self-assembled to form the only trimer at higher concentration (800 µM) (Fig. 2.65). SANS measurements were done to understand the oligomeric states and MWs of the complexes. Try-OEG-C6-3T self-assembled to form trimer at higher concentration (800 µM) in SANS also. The rest of the conjugates Try-OEG-C12-1T/2T/3T and Try-OEG-C18-1T/2T assembled to form the higher-order complexes with oligomeric states ranging from 10-18 (Fig. 2.66). SEC-MALS analyses of the FAP, Try-OEG-C12-2T was performed on a Superdex-200 10/300 GL column calibrated with BSA to determine the AMW as in the previous chapter. The molecular mass of Try-OEG-C12-2T obtained (261 kDa) matches with other results (Fig. 2.67). As expected, protein conjugates with sufficient tail length and hydrophobicity (Try-OEG-C12-2T Try-OEG-C12-3T, Try-OEG-C18-1T, and Try-OEG-C18-2T) self-assemble into higher-order protein complexes. This data obtained for these conjugates agree in all techniques. The control design Try-OEG-C6-3T (800 µM) and Try-OEG-C12-1T (800 µM) predominantly exists as monomer during SEC experiment, whereas the same conjugates behaved as trimer and higher-order complex when analyzed by SANS and DLS experiments. This is primarily attributed to the result of the dilution effect during SEC experiments.

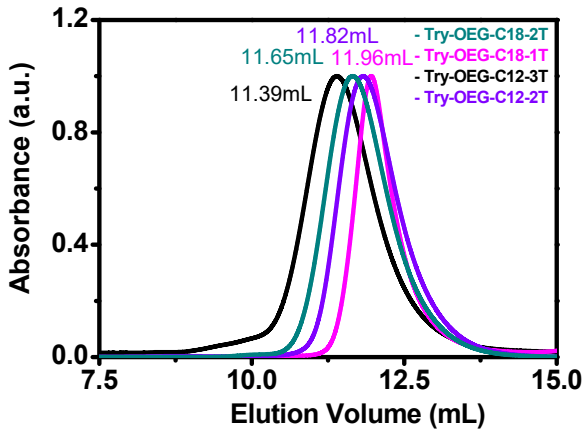


Figure 2.63. SEC data of the tail variant library after purification.

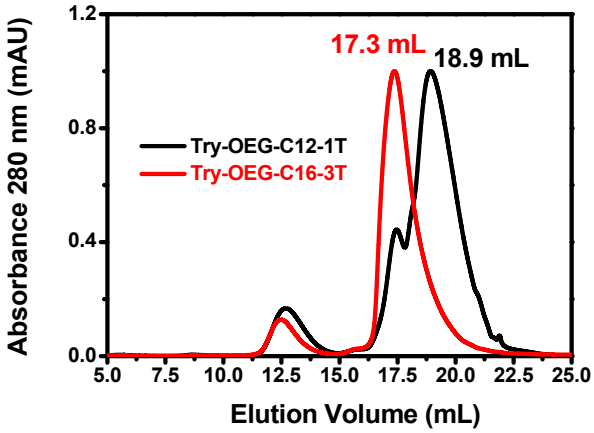


Figure 2.64. SEC data of FAPs that did not assemble (Try-OEG-C6-3T, Try-OEG-C12-1T) (Spacer variant library).

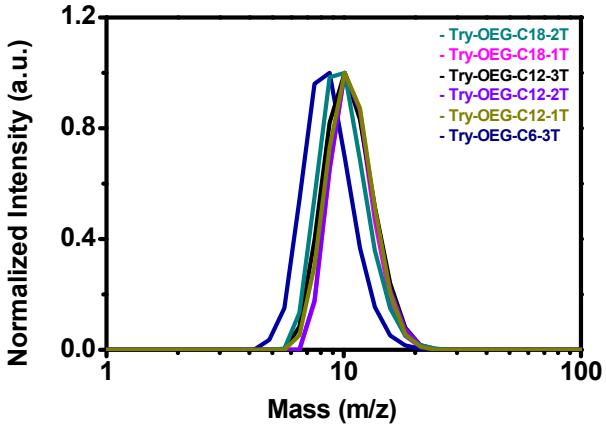


Figure 2.65. DLS data of the tail variant library after purification.

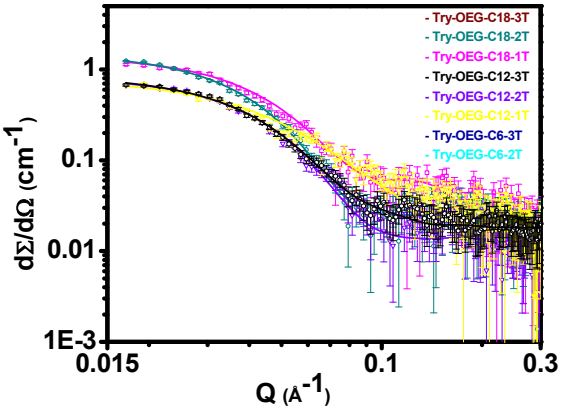


Figure 2.66. SANS data of the tail variant library after purification.

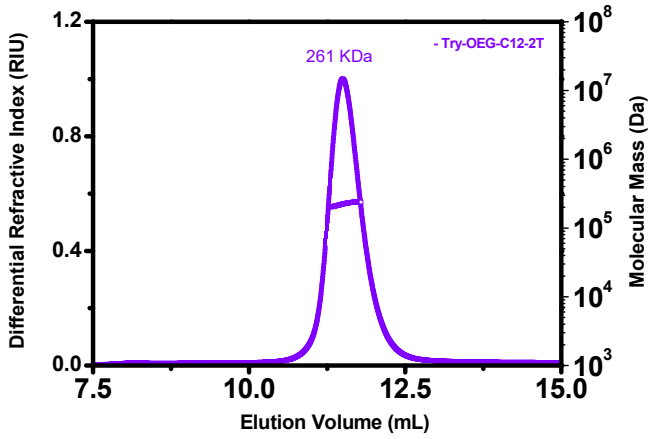


Figure 2.67. SEC-MALS data of the tail variant library after purification.

Table 2.11. Summary of self-assembly data of the tail variants.

Protein Conjugate	Elution Vol SEC (mL)	Mol Wt SEC (kDa)	Mol Wt SANS (kDa)	Oligo state (mer)	R_g from P(r) SANS (Å)	R_h DLS (nm)	Shape Factor ($\sigma = R_g/R_h$)
Try-OEG-C6-3T	17.3	23	73	3	35	4.2	-----
Try-OEG-C12-1T	17.3	23	289	12	37	5.0	0.74
Try-OEG-C12-2T	11.8	225	267	11	42	5.5	0.76
Try-OEG-C12-3T	11.4	276	245	10	41	5.7	0.72
Try-OEG-C18-1T	12.0	210	411	17	43	5.5	0.78
Try-OEG-C18-2T	11.7	241	440	18	47	5.9	0.79

2.4.6. Discussion

All the conjugates with enough hydrophobicity assembled to form higher-order complexes. The influence of hydrophobic tail in making complex is very simple that lack of enough hydrophobicity does not help in complex formation.

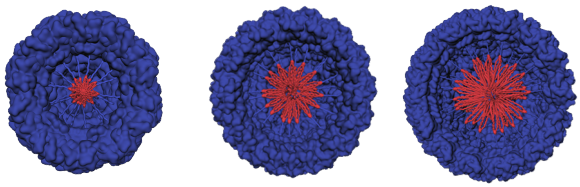


Figure 2.68. Cartoon representation of protein complexes (tail variant library).

In the C6 series, C6-1T was not made. C6-2T did not assemble as it does not have enough hydrophobicity. However, C63T formed trimer at very high concentration DLS and SANS only. In the C12 series, C12-1T formed a higher-complex in DLS and SANS but not in SEC, and the rest behaved as higher-order complexes in all the studies. In the C18 series, both 1T and 2T formed the higher-order complexes.

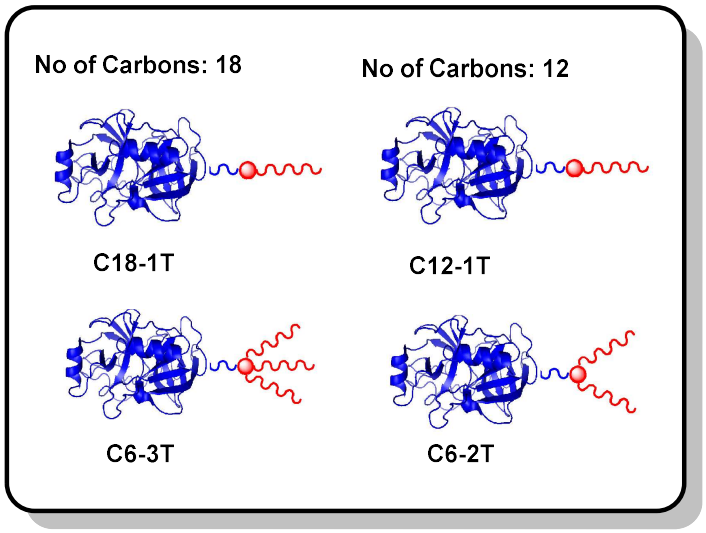


Figure 2.69. Tail variants library with the same number of carbons but different branching.

The most important observation from this library is that it is not just the minimum hydrophobicity that is needed for complex formation but also the arrangement of the hydrophobic groups that are crucial (Fig. 2.69). For example, 18C in linear fashion

assembled, whereas splitting the same 18C into 3Ts does not help. Similarly, 12C in linear fashion helped in complex formation to some extent than splitting those into 2Ts of 6C.

These results reveal that the hydrophobic tail is playing the most important role in complex formation. The protein self-assembly is mediated by strong favorable interaction between hydrophobic tails opposed by electrostatic repulsion between protein head group containing point charges (presence of acidic and basic amino acid at physiological pH and ionic strength)[39]. This mechanism is further supported by the fact that the protein conjugate with a short hydrophobic tail (Try-OEG-C6-2T) failed to self-assemble completely. This is because electrostatic repulsion between proteins would be much higher compared to weak hydrophobic interaction in the assembled state, which would be energetically unfavorable[39].

Table 2.12. Analysis of self-assembly data of tail variants.

No of Carbons: 18 varying branching pattern		**No of Carbons: 12 varying branching pattern**	
as 1 linear chain	**Split into 3 tails**	**as 1 linear chain**	**Split into 2 tails**
Higher-order complex in all studies	**Trimer in SANS and DLS**	**Higher-order complex in SANS and DLS**	**No assembly**

To conclude this chapter, we established the MAPLab technology and then we studied the influence of individual components (head, spacer, and tail) of FAP by systematically making a library of FAPs.[24]

2.5. Experimental section and methods

Selection of proteins used in this work

We chose 4 serine proteins for our study, *i.e.*, Try (23 kDa), Chy (25 kDa), Sub (27 kDa), and ProK (29 kDa). Proteins were purchased from commercial vendors, and molecular weights were determined prior to modification to track the molecular weight changes accurately.

Matrix preparation and molecular weight determination

Apart from the molecular weight determination of native proteins, all stages of protein modification and purification were also followed by MALDI-TOF MS. The samples were analyzed in Linear High Mass mode in AB Sciex 4800 plus MALDI-TOF/TOF analyzer with 4000 Series Explorer as software. Mass was scanned between 10,000 Da and 40,000 Da with a focus mass at 25,000-29,000 Da, depending on the protein analyzed.

4 mg of DHAP and 4.5 mg of DAHC were weighed separately in microcentrifuge tubes. 150 µL of EtOH and 200 µL of milli-Q water were added to each tube, respectively. Both the solutions were sonicated using bath sonicator for 1 min and vortexed for another min. Then 50 µL of DAHC aqueous solution was transferred to the DHAP solution, and the resulting solution was vortexed for one more min to yield matrix mixture. The sample preparation was carried out in a 500 µL microcentrifuge tube, 1:1:1 ratio of a protein sample, 2 % trifluoroacetic acid (TFA), and matrix mixture were mixed and vortexed for 1 min. The samples were then kept undisturbed. When the crystallization observed in centrifuge tubes, 0.5 µL of the samples were spotted on the MALDI-TOF MS plate and air-dried for 15 mins. The plate was then loaded and fired to get accurate molecular weight both in +1 and +2 states. 100 µM protein concentration was found to be optimum for MALDI-TOF MS analysis.

Protein modification, monitoring, and purification

Protein modification was carried out at the concentration of 100 µM, which was found to be optimum for MALDI-TOF MS monitoring. Triton X-100 was used to solubilize the AABPs at concentrations 100 times (20 mM) more than CMC or 2 % of the total volume of the reaction mixture. Typically, for test reactions, the final volume of the reaction mixture is 1 mL. Proteins were weighed (2.3 to 2.9 mg, depending on protein) in microcentrifuge tubes

and 500 µL 50 mM Sod. Phos. pH 7.4 was added and mixed gently with a pipette to make 200 µM solutions. Then AABPs (1 or 2 equivalents) were weighed in a different microcentrifuge tube, followed by the addition of 20 µL triton X-100 and 480 µL of 50 mM Sod. Phos. pH 7.4 and vortexed for 15 mins. When the AABP solution becomes clear, the protein solution was added into AABP solution to get 100 µM (1 mL) protein solutions and allowed to react for 24 h on rotospin at 20 rpm at 25 ºC. For probes FP-OEG-C18-2T and FP-OEG-C18-3T, 10 equivalents were weighed in a microcentrifuge tube and 4 % triton X-100 was added and sonicated till the mixture became homogenous (please note that we used bath sonicator to solubilize these two AABPs). Protein modifications were carried out in falcon tubes at 200 mg scale following the linear scale-up of the above-mentioned procedure for understanding the self-assembly behavior.

To monitor the extent of protein modification, the samples were directly withdrawn from the reaction mixture using a pipette and analyzed. In brief, 2 µL of the reaction mixture was mixed with 2 µL of 2 % TFA and 2 µL of matrix mixture as previously stated, vortexed and spotted on MALDI-TOF MS plate. Please note that the matrix preparation and sample preparation procedures remained the same here.

All the conjugates were purified by three-step purification, *i.e.,* IEX, SEC, and desalting performed using FPLC. IEX was performed to remove triton X-100 using either SP sepharose or Q sepharose resins depending on isoelectric point (pI) and surface charges of proteins (Try: 30.4, Chy: 19.5, Pro K: -8.4, Sub: -6.2). For example, to purify the reaction mixture of Try or Chy, we used SP sepharose, a cation-exchange resin at pH 7.4, and to purify Sub and ProK, we used Q sepharose, an anion-exchange resin at pH 10.

During CEX, the column was pre-equilibrated with the same buffer (50 mM Sod. Phos. pH 7.4) which was used for modification and then the sample was injected followed by post-injection equilibration for at least 2 CVs (CVs) or until the complete removal of triton X-100 for large scale reactions. The elution of native protein and its corresponding conjugate together as a mixture was later achieved using 50 mM Sod. Phos. pH 7.4, 1 M NaCl as elution buffer.

In the case of anion-exchange chromatography (AEX), the protein reaction mixtures were buffer-exchanged first to 50 mM tris base pH 10 (since the buffer used for modification was 50 mM Sod. Phos. pH 7.4). Then the anion-exchange column was pre-equilibrated using the

same buffer (50 mM tris base pH 10), which was used for buffer-exchange, and then the sample was injected, followed by post-injection equilibration for at least 2 CVs or until the complete removal of triton X-100 for large scale reactions. The elution of the native protein and its conjugate together as the mixture was later achieved using 50 mM tris base pH 10, 1 M NaCl as elution buffer.

The obtained IEX fractions were subjected to SEC immediately to remove the native proteins from the protein conjugates. For the separation of native protein from protein conjugate, 50 mM Sod. Phos. pH 7.4, 1 M NaCl was used as buffer using either Sephacryl S-100 HR 16/60 or Sephacryl S-200 HR 16/60 or Sephacryl S-300 HR 16/60. The NaCl was later removed by the Sephadex-G25 desalting column. The desalted fractions were quickly lyophilized and later dissolved in the required buffer when needed.

Monitoring of IEX, SEC, and desalted fractions were carried out using the same procedure mentioned for native proteins and reaction mixtures except for the addition of triton X-100 in protein fractions. In short, 98 µL of samples from each fraction were mixed with 2 µL of triton X-100 (2 % or 100 times the CMC) in a separate microcentrifuge tube and vortexed for 4 hrs. The samples were then analyzed using the same procedure as mentioned above. Please note that the matrix and sample preparation procedure remained the same.

Synthesis and purification of head variants

Synthesis and purification of Try-OEG-C12-3T

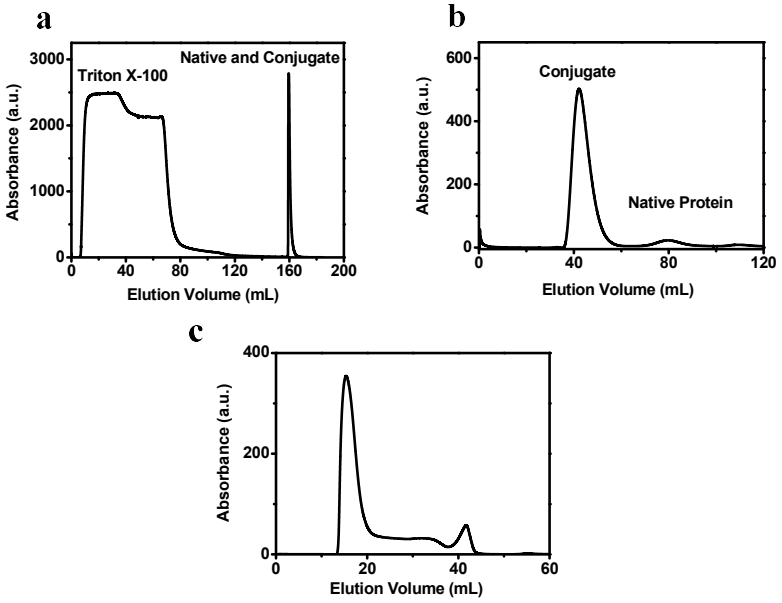


Figure 2A. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram.

Synthesis and purification of Chy-OEG-C12-3T

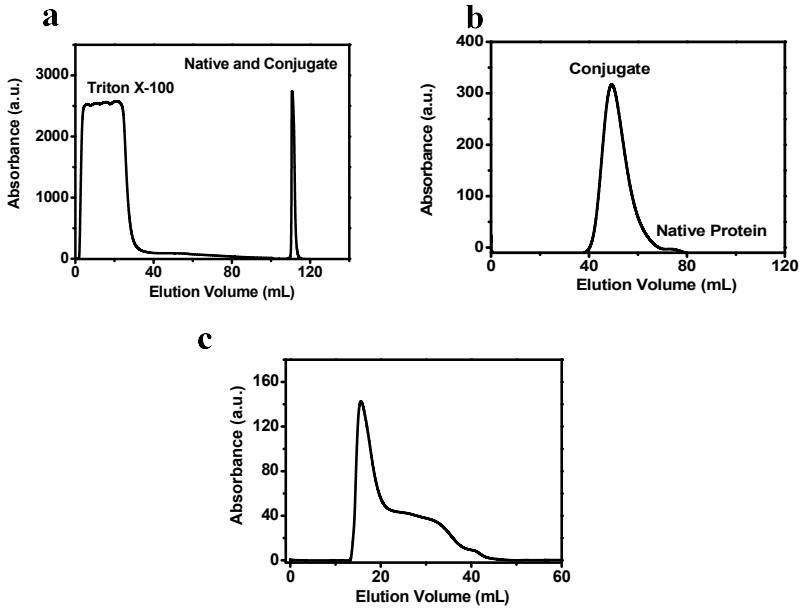


Figure 2B. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram**.**

Synthesis and purification of Sub-OEG-C12-3T

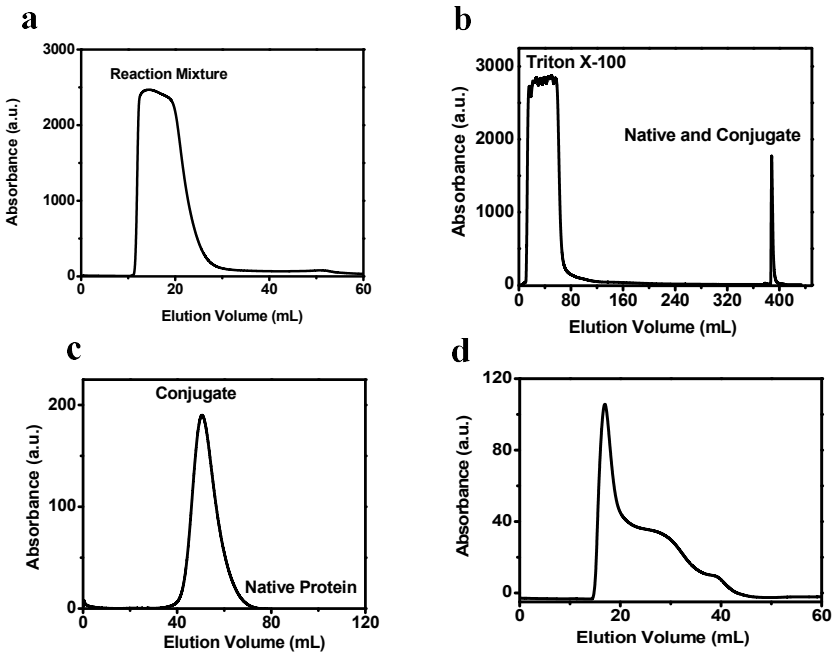


Figure 2C. a. Buffer exchange chromatogram, b. Ion exchange chromatogram, c. Size-exclusion chromatogram, d. Desalting chromatogram.

Synthesis and purification of Pro-OEG-C12-3T

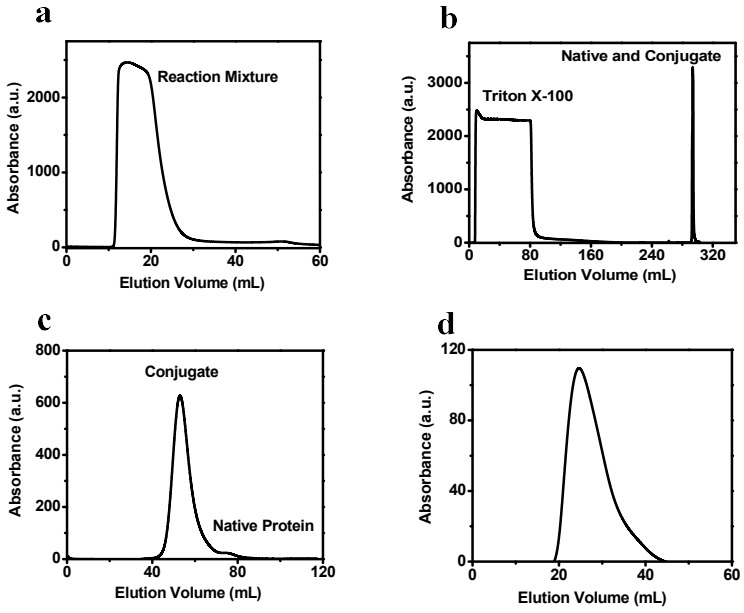


Figure 2D. a. Buffer exchange chromatogram, b. Ion exchange chromatogram, c. Size-exclusion chromatogram, d. Desalting chromatogram

Synthesis and purification of spacer variants conjugates

Synthesis and purification of Try-DEG-C12-3T.

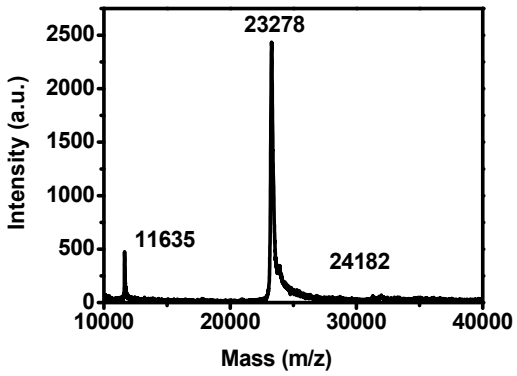


Figure 2E. MALDI-TOF MS spectrum of Try-DEG-C12-3T reaction mixture. **Note.** Reaction failed due to sub-optimal linker length, as evident from MALDI-TOF MS data.

Synthesis and purification of Try-TEG-C12-3T

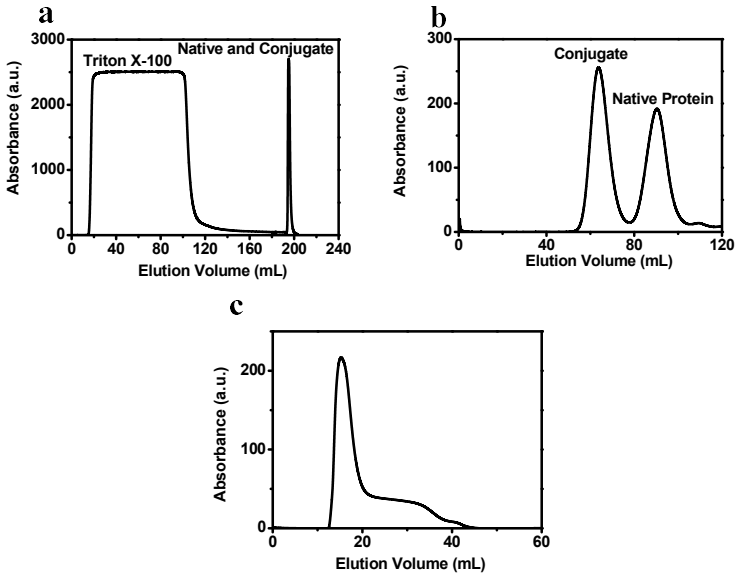


Figure 2F. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram.

Synthesis and purification of Try-OEG-C12-3T given in head variants

Synthesis and purification of Try-DDEG-C12-3T.

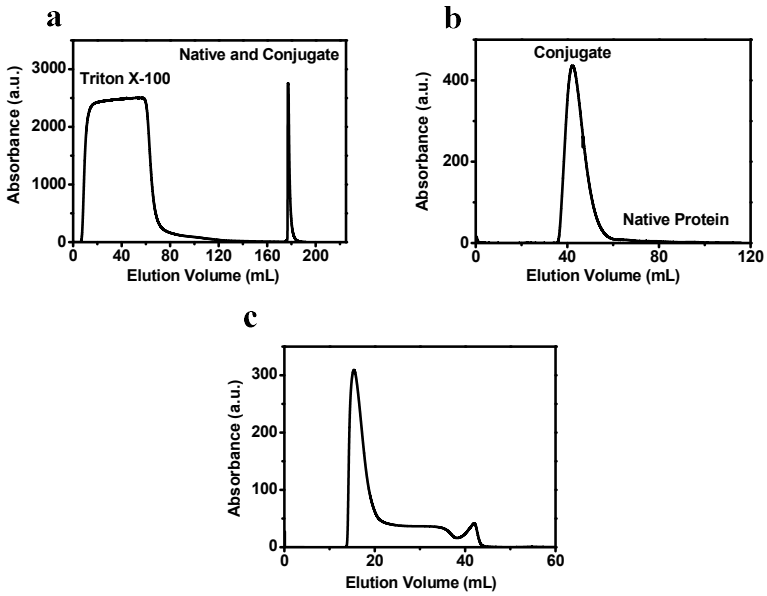


Figure 2G. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram.

Synthesis and purification of Try-CEG-C12-3T

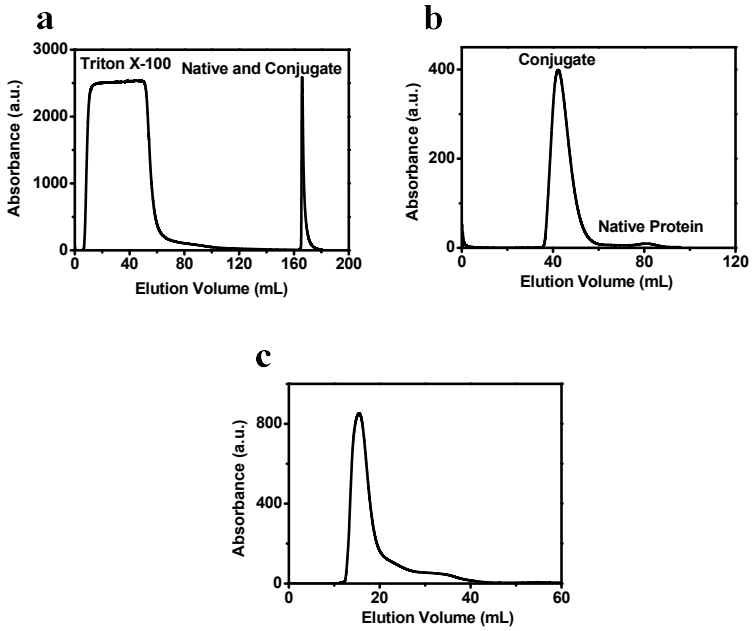


Figure 2H. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram

Synthesis and purification of tail variants

Synthesis and purification of Try-OEG-C6-3T

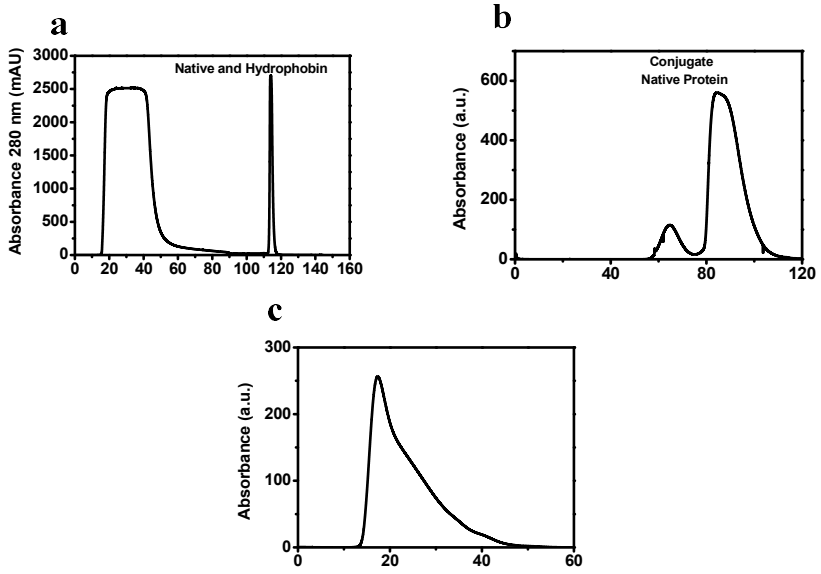


Figure 2I. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram.

Synthesis and purification of Try-OEG-C12-3T given in head variants

Synthesis and purification of Try-OEG-C18-3T

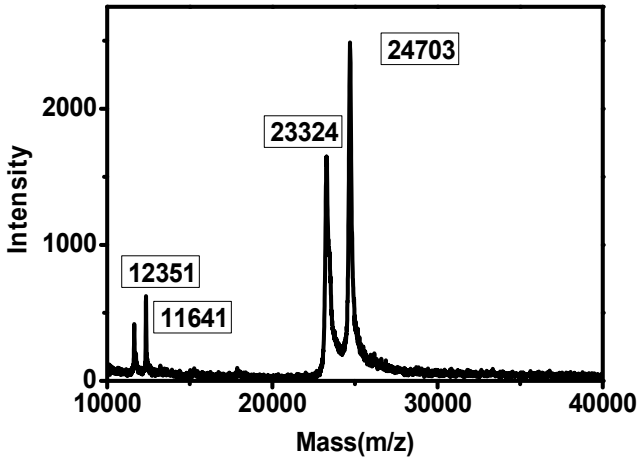


Figure 2J. MALDI-TOF MS spectrum of Try-OEG-C18-3T reaction mixture. **Note.** We were not able to purify this conjugate due to the extreme hydrophobicity.

Synthesis and purification of Try-OEG-C6-2T

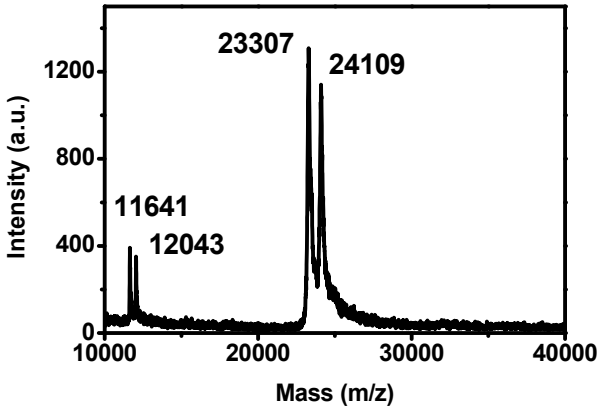


Figure 2K. MALDI-TOF MS spectrum of Try-OEG-C6-2T reaction mixture. **Note.** This conjugate did not self-assemble due to insufficient hydrophobicity.

Synthesis and purification of Try-OEG-C12-2T.

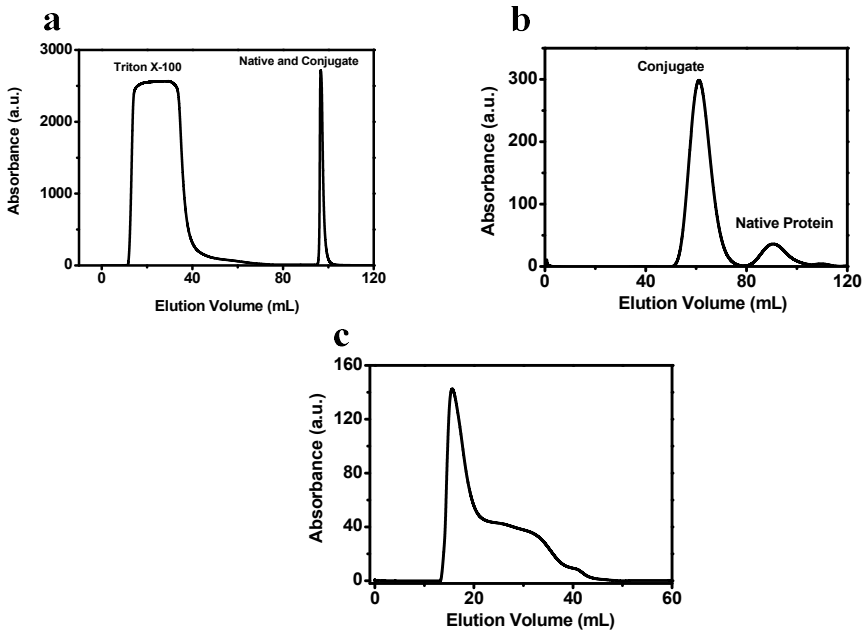


Figure 2L. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram.

Synthesis and purification of Try-OEG-C18-2T.

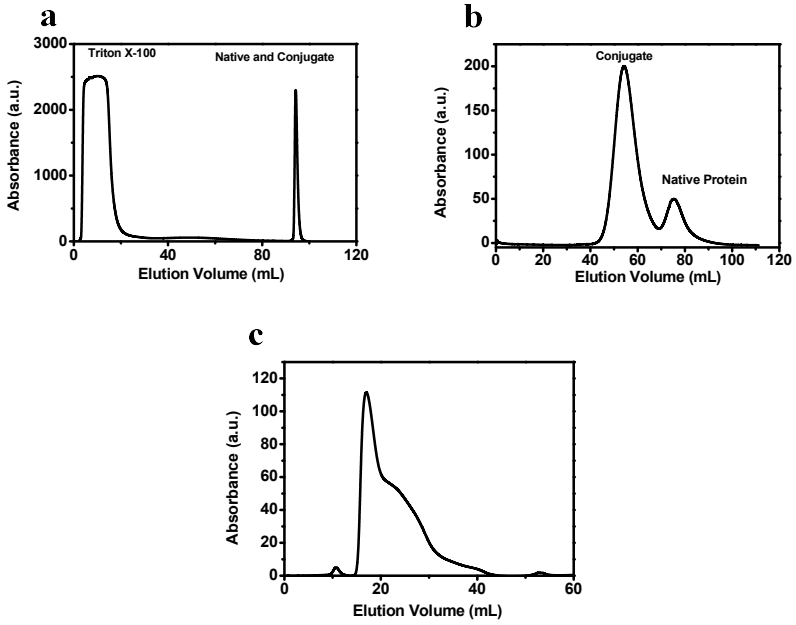


Figure 2M. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram. Try-OEG-C6-1T was not made.

Synthesis and purification of Try-OEG-C12-1T

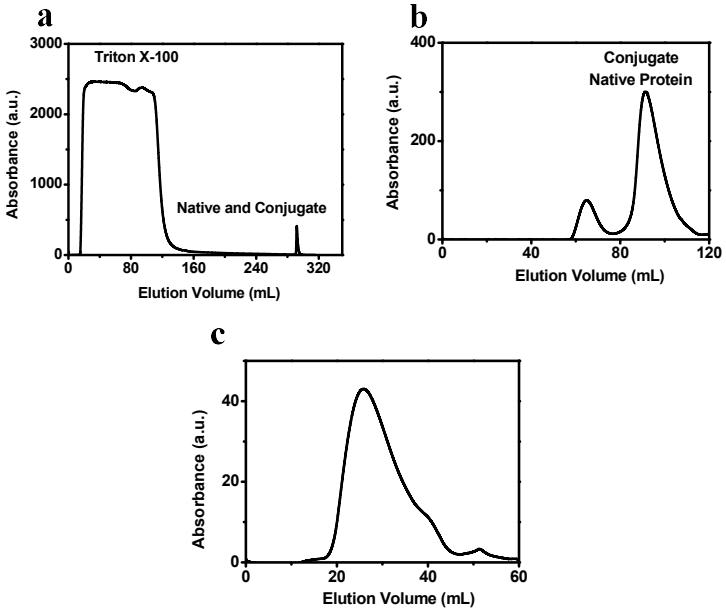


Figure 2N. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram.

Synthesis and purification of Try-OEG-C18-1T

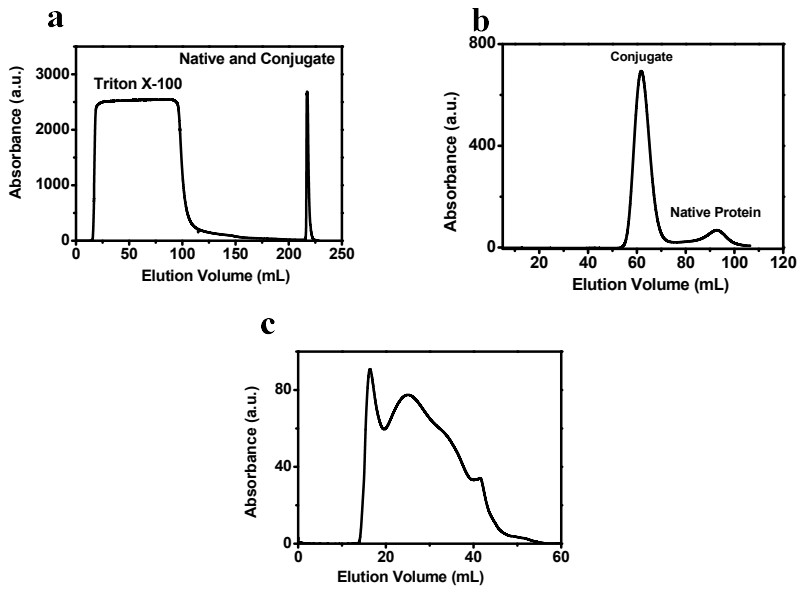


Figure 2O. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram.

Size exclusion-based molecular weight determination

To determine the molecular weights of soluble protein complexes, we performed SEC in Superdex-200 10/300 GL (GE healthcare). For this purpose, we chose standard proteins (GE Healthcare), *i.e.*, blue dextran (2000 kDa), to determine the void volume of this column, thyroglobulin (660 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa) and Try (23 kDa).

A series of size exclusion runs were performed with these proteins in 50 mM Sod. Phos. pH 7.4, 1 M NaCl with 0.25 mL/min as the flow rate. All the proteins were dissolved individually in Milli Q water at 3 mg/mL and 0.5 mL was injected. With the obtained elution volumes, the partition coefficients (K_{av}) for standard proteins were calcd according to standard protocol (GE Healthcare) using the formula $K_{av} = V_e - V_o / V_c - V_o$ where, V_o = column void volume (8.5 mL, from the elution volume of blue dextran), V_e = elution volume of the particular sample, and V_c = geometric CV (23.5 mL). The calibration curve was plotted for K_{av} of standards against RMWs of standard proteins. ($r^2 = 0.99$, please refer to the calibration curve). Now, to determine the molecular weights of soluble protein complexes, the protein conjugates were dissolved individually in Milli Q water at 5 mg/mL and 0.5 mL was injected.

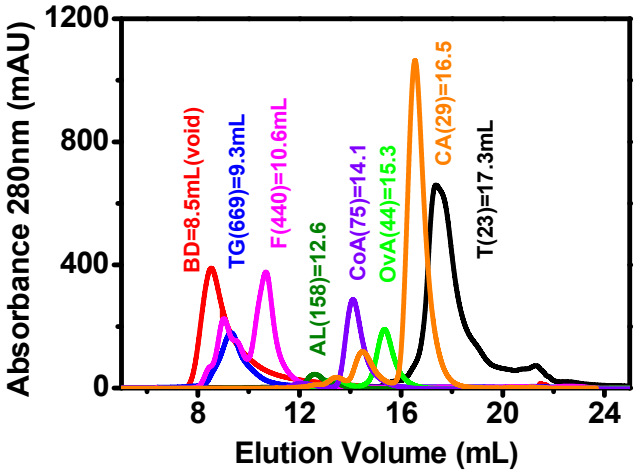


Figure 2P. SEC runs for standard proteins in Superdex 200.

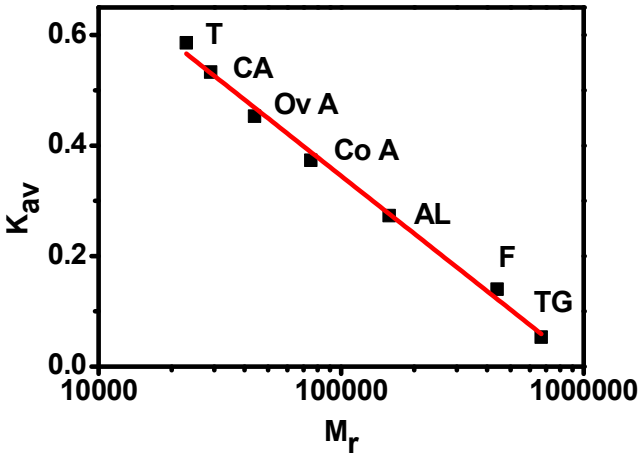


Figure 2Q. Calibration curve for Superdex 200.

Dynamic light scattering

Analysis of protein complexes was performed using DLS using Zetasizer Nano 2590 (Malvern, UK). Protein samples (5 mg/mL) were prepared in 50 mM Sod. Phos. pH 7.4. 1 mL of sample was taken in disposable polystyrene cells, and then the mean size of the complexes was measured at 90º scattering angle.

SANS measurements

SANS experiments were carried out at SANS spectrometer operating at Dhruva Reactor, Bhabha Atomic Research Centre (BARC), Mumbai, India[51]. The facility employs a neutron

velocity selector to provide a monochromatic neutron beam with a mean wavelength (λ) of 5.2 Å and wavelength spread (Δλ/λ) approximately 15 %. The scattered neutrons were detected using a one-dimensional ^{3}He position sensitive detector (PSD) in an angular range of 0.5-15° corresponding to wave vector transfer ($Q = 4\pi\sin(\theta/2)/\lambda$, where θ is scattering angle) range of 0.015-0.3 Å^{-1}. The measured SANS data were corrected for the background, the empty cell contribution, and the transmission. The corrected data were presented on an absolute scale using the standard protocols. For SANS measurements, the samples (5 mg/mL) were prepared in Sod. Phos. pH 7.4 buffer in D_2O in order to minimize the incoherent scattering and to increase contrast. The temperature was maintained at 30 ºC during the measurements.

SANS Analysis

SANS is a very promising technique to investigate self-assembled structures in nanometer length scales. In the SANS experiment, one usually measures the differential scattering cross-section ($d\Sigma/d\Omega$) per unit volume as a function of wave vector transfer (Q). For monodispersed particles in a medium, it can be expressed as[52]

$$\frac{d\Sigma}{d\Omega} = n\,P(Q)\,S(Q) + B \qquad (1)$$

Where n is the particle number density. $P(Q)$ is the intra-particle structure factor (square of the form factor), and S(Q) is the inter-particle structure factor. $P(Q)$ provides the specific size and shape of the scatterer, whereas $S(Q)$ decides the spatial arrangement of the particles and thereby gives the information about the inter-particle interaction. For a dilute system, S(Q) may be approximated to unity. B is a constant term representing the incoherent background arising mainly from the hydrogen atoms present in the sample.

$P(Q)$ for spherical particles of radius R is given by[53]

$$P(Q) = \frac{16\pi^2}{9}\left(\rho_p - \rho_s\right)^2 R^6 \left[3\frac{\sin(QR) - (QR)\cos(QR)}{(QR)^3}\right]^2 \qquad (2)$$

where ρ_p and ρ_s are the scattering length densities of particle and solvent, respectively.

For prolate ellipsoidal particles, P(Q) can be expressed as[53]

$$P(Q)=\frac{16\pi^2}{9}\left(\rho_p-\rho_s\right)^2\left(ab^2\right)^2\int_0^1\left[F\left(Q,\mu\right)\right]^2 d\mu \qquad (3)$$

where the functions are given by

$$F(Q,\mu)=\frac{3(\sin x-x\cos x)}{x^3} \text{ and } x=Q[a^2\mu^2+b^2(1-\mu^2)]^{1/2}$$

where a and $b = c$ are, respectively, semi-major and semi-minor axes of the prolate ellipsoid ($a > b = c$). The variable μ is the cosine of the angle between the directions of a and Q.

The following expression provides $P(Q)$ for a spherical core-shell structure[53]

$$P(Q)=\frac{16\pi^2}{9}[(R+dR)^3\left(\rho_{shell}-\rho_s\right)\frac{sinQ(R+dR)-Q(R+dR)cosQ(R+dR)}{Q^3(R+dR)^3}-R^3\left(\rho_{shell}-\rho_{core}\right)\frac{sinQR-QRcosQR}{Q^3R^3}]^2$$

Where R is core radius, and dR is shell thickness. $\rho_{core}$and ρ_{shell} represents the scattering length densities of the core and shell, respectively.

Throughout the data analysis corrections were made for instrumental smearing. The calcd scattering profiles were smeared by the appropriate resolution function to compare with the measured data. The parameters in the analysis were optimized by means of a nonlinear least-square fitting program[54].

SEC-MALS

SEC-MALS analyses of the protein complexes were performed on a Superdex-200 10/300 GL column (GE Healthcare) connected to an Agilent HPLC system equipped with the 18-angle light scattering detector (Wyatt Dawn HELIOS II) and a refractive index detector (Wyatt OptilabTrEX). The system was calibrated with BSA at a concentration of 2 mg/mL; 100 µL of the protein conjugates at concentrations 5 mg/mL were injected, and the molecular weights were calcd using ASTRA software (Wyatt Technologies).

Fluorescence Spectroscopy

The fluorescence of native proteins and protein conjugates was measured using the Xenon arc lamp on Fluorologspectrofluorimeter (HORIBA, Jobin-Yvon). Protein samples (1 mg/mL)

were prepared in 50 mM Sod. Phos. pH 7.4. Measurements were carried out at tryptophan specific excitation wavelength of 295 nm in a 10 mm path length cuvette. The samples remained clear during the measurement.

CD Spectroscopy

The CD spectra of native proteins and protein conjugates were recorded on the Jasco J-815 spectrometer. Protein samples (0.25 mg/mL) were prepared in 50 mM Sod. Phos. pH 7.4 and measurements were performed in a 1 mm path length cuvette. The spectra (background corrected) were obtained at a scanning speed of 50 nm/minute within range of 195-250 nm. Mean residue elasticity was converted from accumulated spectra of five measurements for each sample.

Zeta Potential Measurement

The zeta potential for native proteins and protein conjugates was measured using Zetasizer Nano 2590 (Malvern, UK). Protein samples (1 mg/mL) were prepared in milli-Q water.

Synthesis and purification of AABPs and their intermediates

General remarks

All compounds were characterized by Nuclear Magnetic Resonance Spectroscopy (NMR), *i.e.,* ^{1}H, ^{13}C, and ^{19}F in case of fluorinated compounds and also by Matrix-Assisted Laser Desorption/Ionisation-Time of Flight Mass Spectrometry (MALDI-TOF MS) or High-Resolution Mass Spectrometry (HRMS). All the NMR spectra were recorded either on a Bruker 400 MHz or on a Jeol 400 MHz instrument using Tetramethylsilane (TMS) as an internal standard (chemical shifts in δ). Unless otherwise stated, all ^{1}H units were mentioned in parts per million (ppm), and measured relative to the signals of the residual proton resonances of $CHCl_3$ (7.26 ppm) or methanol (CH_3OH) (4.87 ppm) in their respective deuterated solvents. All ^{13}C and ^{19}F NMR spectra were measured decoupled from ^{1}H nuclei and are reported in ppm.

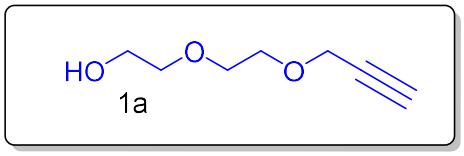
All the reactions were monitored by thin-layer chromatography over precoated silica gel plates, using ultraviolet (UV) lamp or phosphomolybdic acid (PMA) or ninhydrin (NH) staining. Chromatographic separations were performed on column chromatography using

normal-phase silica gel (100-200 mesh size) or reverse-phase silica gel (50 µm). The room temperature (RT) varied between 21 ºC - 35 ºC. To explain the multiplicities following abbreviations were used: s = singlet, d = doublet, dd = doublet of doublet, t = triplet, q = quartet, m = multiplet, quint = quintet.

Synthesis of hydrophilic spacers and their intermediates

Synthesis of diethylene glycol spacer and its intermediates

Synthesis of compound 1a



Mol. formula: $C_7H_{12}O_3$

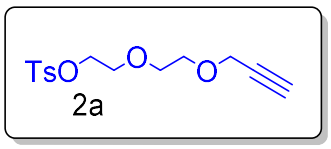
Mol. weight: 144.07 g/mol

Physical appearance: pale yellow liquid

Yield: 53 %

The **compound 1a** was prepared starting from DEG (1.2 g, 11 mmol), NaH (0.03 g, 1.4 mmol), ProBr (1.3 g, 10 mmol) in THF. The product was obtained as a pale yellow liquid (2.1 g, 15 mmol, 53 %) after purification by NPC using MeOH / DCM as eluent, R_f = 0.36 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.17 (d, J = 2.4 Hz, 2H), 3.70-3.63 (m, 7H), 3.57 (t, J = 4.4 Hz, 3H), 2.43 (t, J = 2.4 Hz, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 79.47, 77.16, 74.83, 72.58, 70.17, 69.12, 61.65, 58.41. **HRMS** (M+H): 145.12 g/mol.

Synthesis of compound 2a



Mol. formula: $C_{14}H_{18}O_5S$

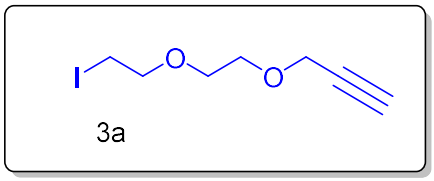
Mol. weight: 298.08 g/mol

Physical appearance: pale yellow liquid

Yield: 90 %

The **compound 2a** was prepared starting from **compound 1a** (1.0 g, 7 mmol), TsCl (1.5 g, 8 mmol), KOH (1.3 g, 24 mmol) in THF. The product was obtained as a pale yellow liquid (1.8 g, 6 mmol, 90 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.55 in 50 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.78 (d, J = 8 Hz, 2H), 7.32 (d, J = 8 Hz, 2H), 4.17-4.13 (m, 4H), 3.70-3.58 (m, 6H), 2.44-2.41(m, 4H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 144.92, 132.94, 129.91, 128.04, 79.57, 77.16, 74.73, 70.59, 69.30, 69.05, 68.73, 60.45, 58.48, 21.71, 21.12, 14.31. **HRMS** (M+Na): 321.07 g/mol.

Synthesis of compound 3a



3a

Mol. formula: $C_7H_{11}O_2I$

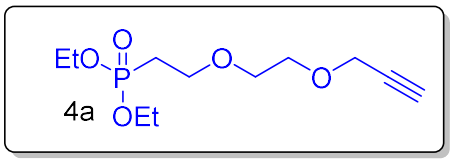
Mol. weight: 253.95 g/mol

Physical appearance: pale yellow liquid

Yield: 92 %

The **compound 3a** was prepared from **compound 2a** (1.5 g, 5 mmol), KI (3.3 g, 20 mmol) in acetone. The product was obtained as a pale yellow liquid (1.1 g, 4 mmol, 92 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.70 in 50 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.17 (d, J = 2.4 Hz, 2H), 3.70-3.63 (m, 7H), 3.57 (t, J = 4.4 Hz, 3H), 2.43 (t, J = 2.4 Hz, 1H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 79.37, 77.16, 74.93, 72.48, 70.17, 69.32, 61.65, 58.41 g/mol.

Synthesis of compound 4a



Mol. formula: $C_{11}H_{21}O_5P$

Mol. weight: 264.11 g/mol

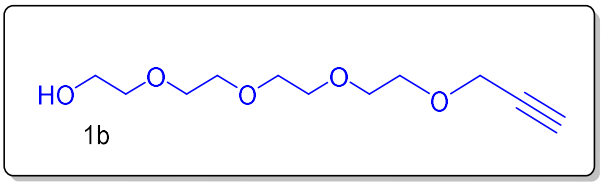
Physical appearance: pale yellow liquid

Yield: 93 %

The **compound 4a** was prepared from **compound 3a** (1.0 g, 4 mmol) and $P(OEt)_3$ (2.6 g, 15 mmol). The product was obtained as a pale yellow liquid (0.97 g, 3 mmol, 93 %) after purification by NPC using MeOH / DCM as eluent, R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.01-3.78 (m, 6H), 3.60-3.30 (m, 6H), 2.30 (m, 1H), 1.96-1.83 (m, 2H), 1.14-0.97 (m, 6H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 79.13, 77.16, 74.49, 69.55, 68.62, 64.66, 61.28, 57.86, 27.19, 25.80, 16.03. **HRMS** (M+Na): 287.12 g/mol.

Synthesis of tetraethylene glycol spacer and its intermediates

Synthesis of compound 1b



Mol. formula: $C_{11}H_{20}O_5$

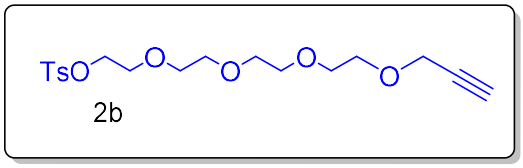
Mol. weight: 232.13 g/mol

Physical appearance: pale yellow liquid

Yield: 58 %

The **compound 1b** was prepared from TEG (15.0 g, 77 mmol), NaH (1.23 g, 51 mmol), ProBr (6.13 g, 51 mmol) in THF. The product was obtained as a pale yellow liquid (4.6 g, 30 mmol, 58 %), after purification by NPC using MeOH / DCM as eluent, R_f = 0.34 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 3.92 (d, J = 2.4 Hz, 2H), 3.44-3.36 (m, 15H), 3.31 (t, J = 4.4 Hz, 3H), 2.34 (t, J = 2.4 Hz, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 79.15, 74.46, 72.03, 69.95, 69.89, 69.86, 69.70, 69.68, 68.43, 60.82, 57.68, 53.30. **HRMS** (MW+Na): 255.12 g/mol.

Synthesis of compound 2b



2b

Mol. formula: $C_{18}H_{26}O_7S$

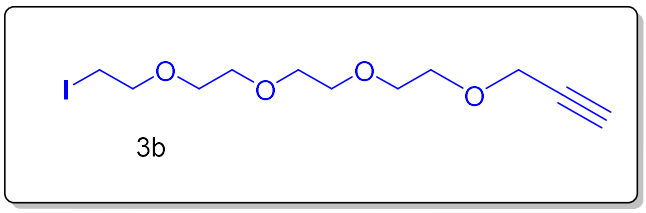
Mol. weight: 386.45 g/mol

Physical appearance: pale yellow liquid

Yield: 71 %

The **compound 2b** was prepared from **compound 1b** (7.0 g, 30 mmol), TsCl (4.1 g, 30 mmol), KOH(4.0 g, 70 mmol) in THF. The product was obtained as a pale yellow liquid (8.2 g, 21 mmol, 71 %) after purification by NPC using MeOH / DCM as eluent, R_f = 0.57 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.80 (d, J = 8 Hz, 2H), 7.34 (d, J = 8 Hz, 2H), 4.20-4.14 (m, 5H), 3.70-3.58 (m, 17H), 2.44-2.41 (m, 4H), 1.68 (s, 3H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 144.89, 129.90, 128.07, 79.71, 74.62, 70.81, 70.65, 70.60, 70.46, 69.33, 69.18, 68.75, 58.47, 21.73. **HRMS** (M+Na): 409.13 g/mol.

Synthesis of compound 3b



3b

Mol. formula: $C_{11}H_{19}O_4I$

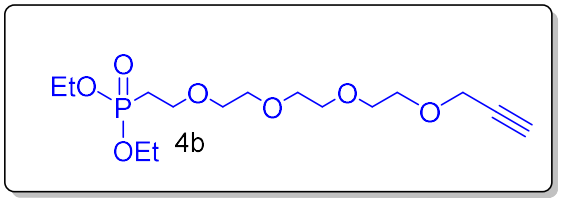
Mol. weight: 342.03 g/mol

Physical appearance: pale yellow liquid

Yield: 72 %

The **compound 3b** was made from **compound 2b** (26.0 g, 67 mmol), KI (40.0 g, 240 mmol) in acetone. The product was obtained as a pale yellow liquid (16.5 g, 48 mmol, 72 %) after purification by NPC using MeOH / DCM as eluent, R_f = 0.69 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.16 (d, J = 2.4 Hz, 2H), 3.73-3.62 (m, 15H), 3.22 (t, J = 7.8 Hz, 2H), 2.41 (t, J = 2.4 Hz, 1H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 79.46, 77.16, 74.61, 71.82, 70.51, 70.29, 70.08, 68.98, 58.30, 3.17. **HRMS** (M+Na): 365.02 g/mol.

Synthesis of compound 4b



Mol. formula: $C_{15}H_{29}O_7P$

Mol. weight: 352.17 g/mol

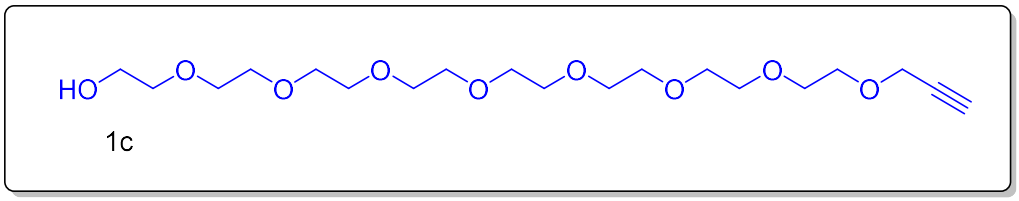
Physical appearance: pale yellow liquid

Yield: 75 %

The compound **4b** was prepared from **3b** (0.20 g, 0.5 mmol) and $P(OEt)_3$ (0.67 g, 4 mmol). The product was obtained as a pale yellow liquid (0.15 g, 0.4 mmol, 75 %) after purification by NPC using MeOH / DCM as eluent, R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.19 (s, 2H), 4.14-4.01 (m, 4H), 3.73-3.53 (m, 14H), 2.43 (t, J = 2.4 Hz, 1H), 2.18-1.97 (m, 2H), 1.31 (t, J = 7.2 Hz). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 79.82, 77.16, 74.70, 70.77, 70.69, 70.57, 70.35, 69.28, 65.29, 61.79, 61.73, 58.55, 27.87, 26.49, 16.59. **HRMS** (MW+Na): 375.15 g/mol.

Synthesis of octaethylene glycol spacer and its intermediates

Synthesis of compound 1c



Mol. formula: $C_{19}H_{36}O_9$

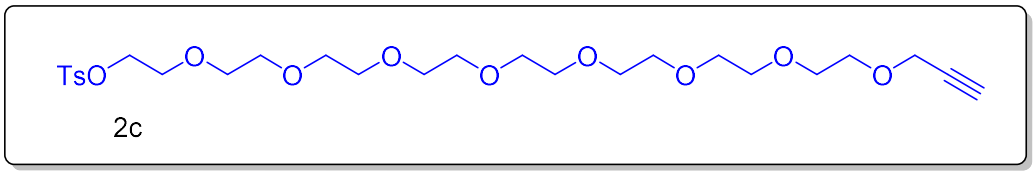
Mol. weight: 408.48 g/mol

Physical appearance: colorless liquid

Yield: not determined

The **compound 1c** was prepared from **compound 2b** (7.3 g, 18 mmol), TEG (9.0 g, 46 mmol), and NaH (0.9 g, 37 mmol) in THF. The product obtained was carried to the next step without purification, R_f = 0.34 in 5 % MeOH / DCM. **HRMS** (M+H): 409.24 g/mol.

Synthesis of compound 2c



Mol. formula: $C_{26}H_{42}O_{11}S$

Mol. weight: 562.67 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 2c** was prepared from **compound 1c** (9.9 g), TsCl (46.0 g, 422 mmol), Et_3N (57.0 g, 564 mmol) and DMAP (1.7 g, 13 mmol) in DCM. The product was obtained as pale yellow liquid (3.3 g, 5.0 mmol) after purification by NPC using MeOH / DCM as eluent, R_f =

0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.77 (d, J = 8 Hz, 2H), 7.32 (d, J = 8 Hz, 2H), 4.18 (d, J = 2.4 Hz, 2H), 4.14 (t, J = 3.2 Hz, 2H), 3.69-3.66 (m, 6H), 3.65-3.61 (m, 20H), 3.57 (s, 4H), 2.42 (m, 4H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 144.78, 132.79, 129.76, 127.87, 70.60, 70.42, 69.20, 68.56, 21.56. **MALDI-TOF MS** (M+K): 601.11 g/mol.

Synthesis of compound 3c



Mol. formula: $C_{19}H_{35}IO_8$

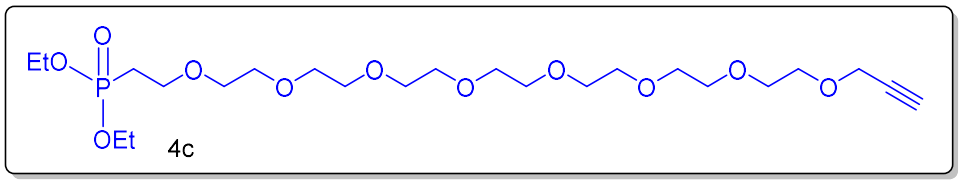
Mol. weight: 518.13 g/mol

Physical appearance: pale yellow liquid

Yield: 95 %

The **compound 3c** was prepared from **compound 2c** (4.6 g, 8.1 mmol), KI (5.4 g, 32 mmol) in acetone. The product was obtained as a pale yellow liquid (4.0 g, 7.7 mmol, 95 %), after purification by NPC using MeOH / DCM as eluent, R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.19 (d, J = 2.4 Hz, 2H), 3.75 (t, J = 7.2 Hz, 3H), 3.69-3.61 (m, 33H), 3.25 (t, J = 7.2 Hz, 2H), 2.43 (t, J = 2.4 Hz, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 79.77, 74.68, 72.09, 70.78, 70.70, 70.52, 70.33, 69.22, 58.53. **MALDI-TOF MS** (M+Na): 557.03 g/mol.

Synthesis of compound 4c



Mol. formula: $C_{23}H_{45}O_{11}P$

Mol. weight: 528.26 g/mol

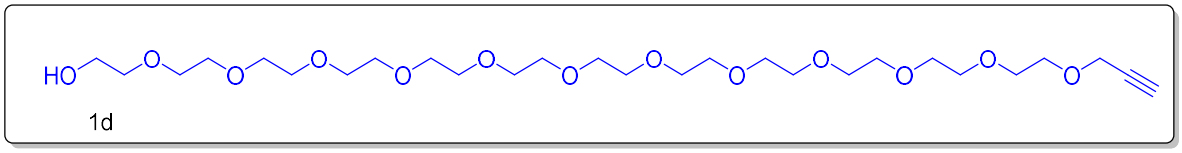
Physical appearance: pale yellow liquid

Yield: 89 %

The **compound 4c** was prepared from **compound 3c** (4.0 g, 7.7 mmol) and $P(OEt)_3$ (5.9 g, 32 mmol). The product was obtained as a pale yellow liquid (3.6 g, 6.8 mmol, 89 %) after purification by NPC using MeOH / DCM as eluent, R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.18 (m, 2H), 4.12-4.03 (m, 4H), 3.72-3.58 (m, 30H), 2.42 (t, *J* = 2.4 Hz, 1H) 2.16-2.07 (m, 2H), 1.30 (t, *J* = 7.2 Hz, 6H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 79.64, 74.65, 70.54, 70.43, 70.39, 70.17, 69.09, 65.12, 61.69, 61.63, 58.41. **MALDI-TOF MS** (M+K): 567.15 g/mol.

Synthesis of dodecaethylene glycol spacer and its intermediates

Synthesis of compound 1d



Mol. formula: $C_{27}H_{52}O_{13}$

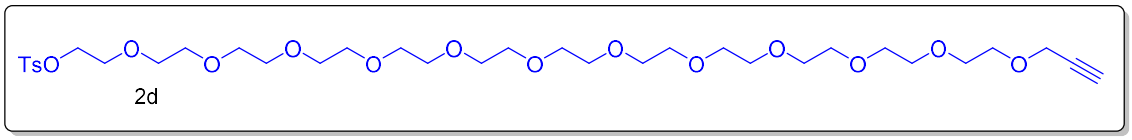
Mol. weight: 584.70 g/mol

Physical appearance: colorless liquid

Yield: not determined

The **compound 1d** was prepared from **compound 2c** (5.3 g, 9.0 mmol), TEG (4.5 g, 23 mmol) and NaH (0.45 g, 16 mmol) in THF. The product obtained was carried to the next step without purification, R_f = 0.34 in 5 % MeOH / DCM. **MALDI-TOF MS** (M+Na): 607.12 g/mol.

Synthesis of compound 2d



Mol. formula: $C_{34}H_{58}O_{15}S$

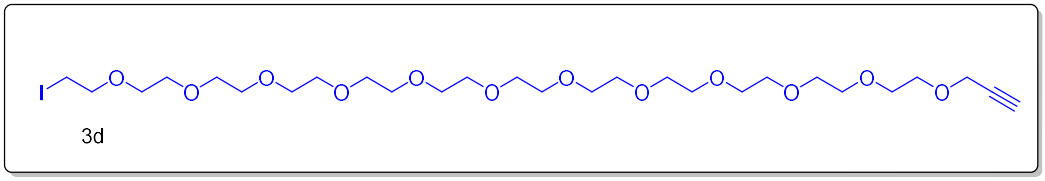
Mol. weight: 738.88 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 2d** was prepared from **compound 1d** (7.1 g), TsCl (33.0 g, 304 mmol), Et_3N (40 g, 405 mmol) and DMAP (1.2 g, 10 mmol) in DCM. The product was obtained as pale yellow liquid after purification by NPC using MeOH / DCM as eluent, R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.79 (d, J = 8 Hz, 2H), 7.34 (d, J = 8 Hz, 2H), 4.19-4.13 (m, 4H), 3.71-3.56 (m, 52H), 2.43-2.42 (m, 4H), 1.95 (s, 3H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 144.88, 133.03, 129.91, 12.07, 79.72, 74.65, 70.80, 70.66, 70.61, 70.57, 70.46, 69.32, 69.17, 68.74, 58.47, 21.73. **MALDI-TOF MS** (M+K): 777.05 g/mol.

Synthesis of compound 3d



Mol. formula: $C_{27}H_{51}IO_{12}$

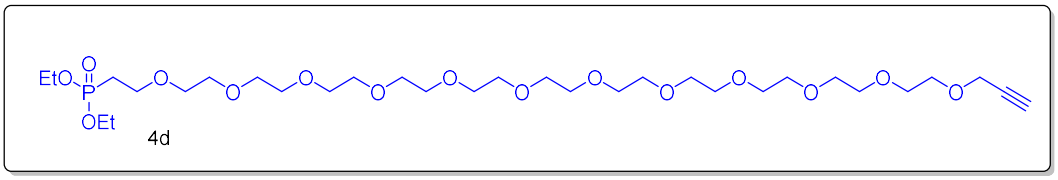
Mol. weight: 693.59 g/mol

Physical appearance: pale yellow liquid

Yield: 90 %

The **compound 3d** was prepared from **compound 2d** (0.40 g, 0.5 mmol), KI (0.25 g, 2.0 mmol) in acetone. The product was obtained as a pale yellow liquid (0.38 g, 0.5 mmol, 90 %) after purification by NPC using MeOH / DCM as eluent, R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.19 (d, J = 2.4 Hz, 2H), 3.75 (t, J = 7.2 Hz, 2H), 3.69-3.64 (m, 49H), 3.25 (t, J = 2.4 Hz, 1H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 79.79, 77.16, 74.67, 72.14, 70.78, 70.69, 70.67, 70.52, 70.34, 69.23, 58.53. **MALDI-TOF MS** (M+K): 732.97 g/mol.

Synthesis of compound 4d



Mol. formula: $C_{31}H_{61}O_{15}P$

Mol. weight: 704.78 g/mol

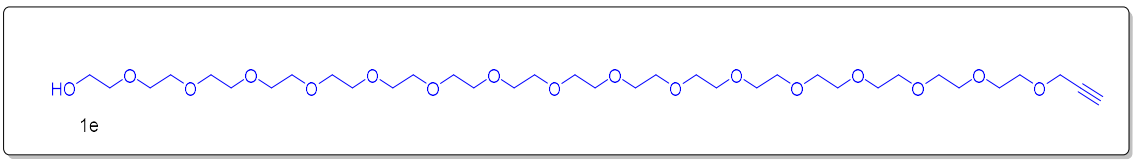
Physical appearance: pale yellow liquid

Yield: 87 %

The **compound 4d** was prepared from **compound 3d** (0.21 g, 0.30 mmol) and $P(OEt)_3$ (0.20 g, 1.2 mmol). The product was obtained as a pale yellow liquid (0.18 g, 0.25 mmol, 87 %) after purification by NPC using MeOH / DCM as eluent, R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.19 (m, 2H), 4.12-4.03 (m, 4H), 3.76-3.54 (m, 51H), 2.43 (t, J = 2.4 Hz, 1H), 2.17-2.06 (m, 2H), 1.30 (t, J = 7.8 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 79.76, 77.16, 74.68, 70.67, 70.28, 69.21, 65.24, 61.70, 58.51, 16.52. **MALDI-TOF MS** (M+K): 743.06 g/mol.

Synthesis of cetyl ethylene glycol spacer and its intermediates

Synthesis of compound 1e



Mol. formula: $C_{35}H_{68}O_{17}$

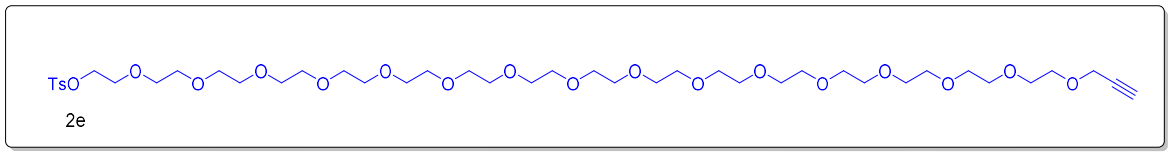
Mol. Weight: 760.91 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 1e** was prepared from **compound 2d** (0.77 g, 1 mmol), TEG (0.50 g, 2.5 mmol), and NaH (0.05 g, 2 mmol) in THF. The product obtained was carried to the next step without purification, R_f = 0.34 in 5 % MeOH / DCM. **MALDI-TOF MS** (M+Na): 783.20 g/mol.

Synthesis of compound 2e



Mol. formula: $C_{42}H_{74}O_{19}S$

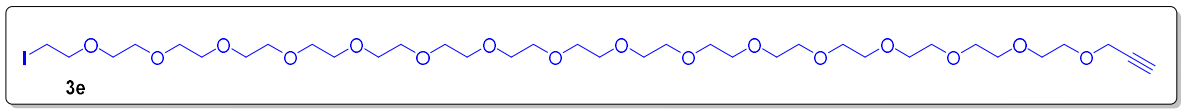
Mol. weight: 914.09 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 2e** was prepared from **compound 1e** (0.7 g), TsCl (2.3 g, 21 mmol) Et_3N (2.8 g, 27 mmol) and DMAP (0.08 g, 0.6 mmol) in DCM. The product was obtained as pale yellow liquid after purification by NPC using MeOH / DCM as eluent, R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, CD_3OD) δ_H 7.83 (d, J = 8 Hz, 2H), 7.48 (d, J = 8 Hz, 2H), 4.20-4.15 (m, 4H), 3.69-3.54 (m, 53H), 3.32 (s, 7H), 2.87 (t, J = 2.4 Hz, 1H), 2.48 (s, 3H). **^{13}C NMR** (100 MHz, CD_3OD) δ_C 146.44, 134.43, 130.83, 129.09, 76.09, 71.56, 71.37, 70.97, 70.10, 69.75, 59.03, 54.83, 49.00, 21.62. **MALDI-TOF MS** (M+K): 953.22 g/mol.

Synthesis of compound 3e



Mol. formula: $C_{35}H_{67}IO_{16}$

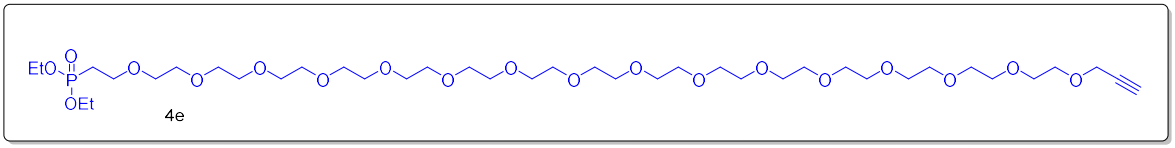
Mol. weight: 870.80 g/mol

Physical appearance: pale yellow liquid

Yield: 92 %

The **compound 3e** was prepared from **compound 2e** (0.21 g, 0.23 mmol), KI (0.15 g, 0.90 mmol) in acetone. The product was obtained as a pale yellow liquid (0.18 g, 0.21 mmol, 92 %) after purification by NPC using MeOH / DCM as eluent, R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, CD_3OD): δ_H 4.21 (d, J = 4 Hz, 2H), 3.78 (t, J = 8 Hz, 2H), 3.70-3.65 (m, 62H), 3.34-3.31 (m, 6H), 2.88 (t, J = 4 Hz, 1H). **^{13}C NMR (**100 MHz, CD_3OD): δ_C 80.65, 75.96, 73.10, 71.64, 71.56, 71.37, 71.17, 70.11, 59.05, 54.81, 49.00. **MALDI-TOF MS** (M+K): 909.26 g/mol.

Synthesis of compound 4e



Mol. formula: $C_{39}H_{77}O_{19}P$

Mol. weight: 880.99 g/mol

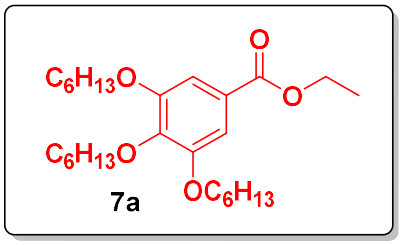
Physical appearance: pale yellow liquid

Yield: 93 %

The **compound 4e** was prepared from **compound 3e** (0.18 g, 0.21 mmol) and $P(OEt)_3$ (0.14 g, 0.84 mmol). The product was obtained as a pale yellow liquid (0.17 g, 0.2 mmol, 93 %) after purification by NPC using MeOH / DCM as eluent, R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, CD_3OD): δ_H 4.19 (d, J = 2 Hz, 2H), 4.16-4.03 (m, 4H), 3.76-3.58 (m, 64H), 2.43 (t, J = 2.4 Hz, 1H), 2.15-2.06 (m, 2H), 1.32 (t, J = 8 Hz, 6H). **^{13}C NMR (**100 MHz, CD_3OD): δ_C 109.07, 104.14, 99.85, 99.56, 98.20, 91.48, 87.18, 83.00, 77.16, 54.69, 44.89. **MALDI-TOF MS** (M+K): 919.18 g/mol.

Synthesis of 3-tail (3T) hydrophobic azides and their intermediates

Synthesis of compound 7a



Mol. formula: $C_{27}H_{46}O_5$

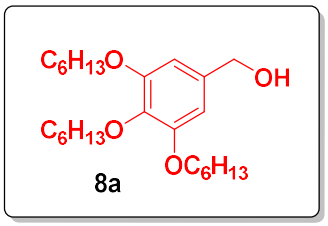
Mol. weight: 450.33 g/mol

Physical appearance: colorless liquid

Yield: 90 %

The **compound 7a** was prepared from ethyl 3,4,5-trihydroxybenzoate (5.2 g, 28 mmol), 1-hexyl bromide (20.0 g, 121 mmol), K_2CO_3 (11.0 g, 78 mmol), KI (0.25 g, 1.5 mmol) in DMF. The product was obtained as colourless liquid (12.0 g, 26 mmol, 90 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.46 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.25 (s, 2H), 4.34 (q, J = 7.2 Hz, 2H), 4.01 (t, J = 6.8 Hz, 6H), 1.84-1.70 (m, 6H), 1.50-1.20 (m, 21H), 0.88 (m, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 166.61, 152.89, 142.33, 125.15, 108.0, 73.59, 69.24, 61.10, 31.86, 31.68, 30.38, 29.38, 25.87, 22.75, 29.64, 14.54, 14.16. **MALDI-TOF MS** (M+K): 489.24 g/mol.

Synthesis of compound 8a



Mol. formula: $C_{25}H_{44}O_4$

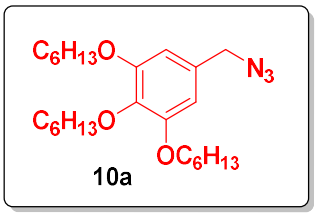
Mol. weight: 408.32 g/mol

Physical appearance: Colourless liquid

Yield: 82 %

The **compound 8a** was prepared from **compound 7a** (1.0 g, 2.2 mmol), LAH (0.25 g, 6.6 mmol) in THF. The product was obtained as colourless liquid (0.73 g, 1.8 mmol, 82 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.25 in 10 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.55 (s, 2H), 4.5 (s, 2H), 3.95 (m, 6H), 1.82-1.70 (m, 6H), 1.50-1.20 (m, 18H), 0.89 (m, 9H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 153.36, 137.56, 136.18, 105.35, 77.16, 73.55, 69.17, 65.78, 31.91, 31.70, 31.09, 30.39, 29.48, 25.90, 22.77, 14.24. **MALDI-TOF MS** (M+K): 447.28 g/mol.

Synthesis of compound 10a



10a

Mol. formula: $C_{25}H_{43}N_3O_3$

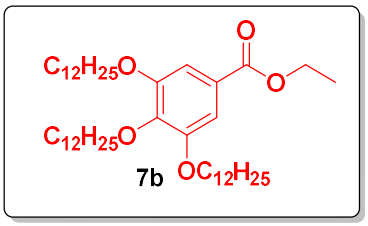
Mol. weight: 433.33 g/mol

Physical appearance: pale yellow liquid

Yield: 96 %

The **compound 9a** was prepared using **8a** (1.0 g, 2.4 mmol), $SOCl_2$ (0.45 g, 4 mmol) and catalytic amount of DMF in DCM. This **compound 9a** was directly used to make **compound 10a** using NaN_3 (0.23 g, 3.5 mmol) in DMF from chloride. The product was obtained as pale yellow liquid (1 g, 2 mmol, 96 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.44 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.48 (s, 2H), 4.24 (s, 2H), 3.96 (m, 6H), 1.83-1.70 (m, 6H), 1.50-1.20 (m, 18H), 0.90 (m, 9H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 153.47, 138.20, 130.48, 106.68, 73.55, 69.26, 55.21, 32.08, 30.46, 29.90, 29.85, 29.79, 29.55, 29.52, 26.23, 22.84, 14.27. **HRMS** (M+H): 434.33 g/mol.

Synthesis of compound 7b



Mol. formula: $C_{45}H_{82}H_5$

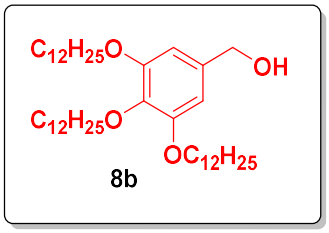
Mol. weight: 702.61 g/mol

Physical appearance: White solid

Yield: 93 %

The **compound 7b** was prepared from ethyl 3,4,5-trihydroxy ester (5.0 g, 25 mmol), 1-dodecyl bromide (25.0 g, 100 mmol), K_2CO_3 (17.0 g, 125 mmol), KI (0.2 g, 1.2 mmol) in DMF. The product was obtained as a white solid (18.0 g, 25 mmol, 93 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.46 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.25 (s, 2H), 4.34 (q, J = 7.2 Hz, 2H), 4.00 (t, J = 6.8 Hz, 6H), 1.86-1.70 (m, 6H), 1.50-1.25 (m, 56H), 0.88 (t, J = 6.8 Hz, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 166.45, 152.79, 142.31, 125.04, 107.98, 73.47, 69.16, 60.93, 31.94, 29.76, 29.74, 29.71, 29.68, 29.65, 29.59, 29.41, 29.38, 29.33, 26.08, 22.70. **MALDI-TOF MS** (M+K): 741.59 g/mol.

Synthesis of compound 8b



Mol. formula: $C_{43}H_{80}O_4$

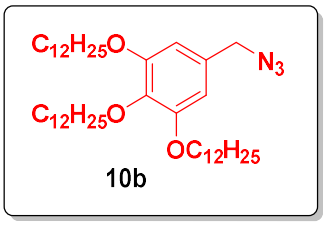
Mol. weight: 660.60 g/mol

Physical appearance: White solid

Yield: 75 %

The **compound 8b** was prepared by from **compound 7b** (0.70 g, 1.5 mmol), LAH (0.16 g, 4.1 mmol) in THF. The product was obtained as a white solid (0.45 g, 0.7 mmol, 75 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.25 in 10 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.55 (s, 2H), 4.53 (s, 2H), 3.97 (m, 6H), 1.83-1.67 (m, 6H), 1.48-1.26 (m, 55H), 0.87 (t, J = 7.2 Hz, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 153.34, 137.53, 136.22, 105.33, 73.56, 69.16, 65.73, 32.06, 30.44, 29.90, 29.84, 29.80, 29.54, 26.24, 22.83, 14.25. **MALDI-TOF MS** (M+K): 699.54 g/mol.

Synthesis of compound 10b



10b

Mol. formula: $C_{43}H_{79}O_3$

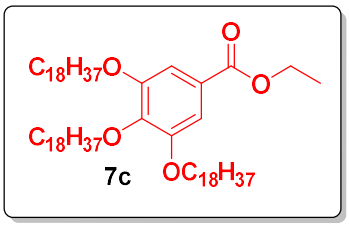
Mol. weight: 685.61 g/mol

Physical appearance: White solid

Yield: 86 %

The **compound 9b**, a chloride was achieved using **compound 8b** (1.8 g, 2.7 mmol), $SOCl_2$ (0.50 g, 4 mmol) and catalytic amount of DMF in DCM. Subsequently azide, **compound 10b** was synthesized using NaN_3 (0.20 g, 3 mmol) in DMF from **compound 9b** without purification. The product was obtained as a white solid (0.90 g, 1.3 mmol, 86 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.44 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.48 (s, 2H), 4.24 (s, 2H), 3.97 (m, 6H), 1.83-1.68 (m, 6H), 1.48-1.26 (m, 57H), 0.87 (t, J = 7.2 Hz, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 153.47, 138.20, 130.48, 106.68, 73.55, 69.26, 55.32, 32.08, 30.46, 29.90, 29.85, 29.79, 29.55, 29.52, 26.23, 22.84, 14.27. **MALDI-TOF MS** (M+Na): 708.60 g/mol.

Synthesis of compound 7c



Mol. formula: $C_{63}H_{118}O_5$

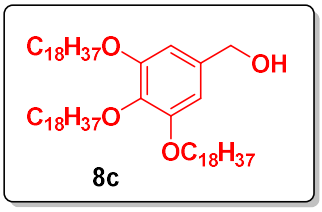
Mol. weight: 954.89 g/mol

Physical appearance: White solid

Yield: 95 %

The **compound 7c** was prepared from ethyl 3,4,5-trihydroxybenzoate (2.0 g, 10 mmol), 1-octadecyl bromide (18.4 g, 54 mmol), K_2CO_3 (12.1 g, 87 mmol), KI (0.2 g, 1.0 mmol) in DMF. The product was obtained as a white solid (9.2 g, 9.5 mmol, 95 %) after purification by NPC using DCM / PE as eluent, R_f = 0.46 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.25 (s, 2H), 4.34 (q, J = 7.2 Hz, 2H), 4.00 (t, J = 6.4 Hz, 6H), 1.84-1.70 (m, 6H), 1.50-1.19 (m, 109H), 0.876 (t, J = 6.8 Hz, 9H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 166.62, 161.35, 152.93, 142.42, 125.18, 108.09, 77.16, 73.62, 69.30, 61.10, 32.09, 29.88, 29.83, 29.80, 29.53, 29.47, 29.34, 26.24, 22.85, 14.55, 14.27. **MALDI-TOF MS** (M+Na): 977.88 g/mol.

Synthesis of compound 8c



Mol. formula: $C_{61}H_{116}O_4$

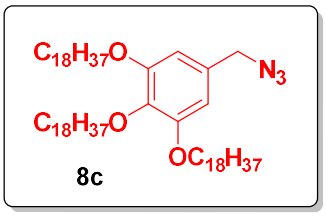
Mol. weight: 912.88 g/mol

Physical appearance: White solid

Yield: 86 %

The **compound 8c** was prepared from **compound 7c** (10.0 g, 10 mmol), LAH (2.0 g, 52 mmol) in THF. The product was obtained as a white solid (8.0 g, 9 mmol, 86 %) after purification by NPC using DCM / PE as eluent, R_f = 0.34 in 10 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.55 (s, 2H), 4.58 (s, 2H), 3.97-3.91 (m, 6H), 1.81-1.70 (m, 6H), 1.47-1.25 (m, 117H), 0.88 (t, J = 7.2 Hz, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 153.37, 137.58, 136.21, 105.37, 77.16, 73.63, 69.18, 65.76, 63.21, 32.07, 30.46, 29.87, 29.81, 29.57, 29.52, 26.28, 22.84, 14.27. **MALDI-TOF MS** (M+K): 951.82 g/mol.

Synthesis of compound 10c



8c

Mol. formula: $C_{61}H_{115}N_3O_3$

Mol. weight: 937.89 g/mol

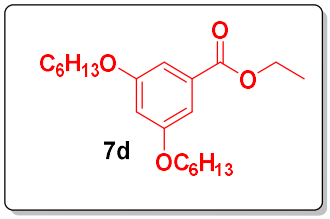
Physical appearance: White solid

Yield: 72 %

The **compound 9c**, a chloride was achieved using **compound 8c** (3.5 g, 3.8 mmol), $SOCl_2$ (1 g, 7.6 mmol) and catalytic amount of DMF in DCM. This was used for next step without purification. The **compound 10c**, azide was synthesized using NaN_3 (0.90 g, 14 mmol) and TBAB (1.2 g, 0.4 mmol) in THF. The product was obtained as white solid (1.74 g, 1.8 mmol, 72 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.44 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.48 (s, 2H), 4.24 (s, 2H), 3.95 (m, 6H), 1.84-1.70 (m, 6H), 1.50-1.19 (m, 97H), 0.87 (t, J = 6.8 Hz, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 153.48, 138.21, 130.48, 106.68, 73.55, 69.26, 55.32, 32.09, 30.47, 29.88, 29.83, 29.57, 29.53, 26.24, 22.85, 14.27. **MALDI-TOF MS** (M+K): 976.76 g/mol.

Synthesis of 2T hydrophobic azides and their intermediates

Synthesis of compound 7d



Mol. formula: $C_{21}H_{34}O_4$

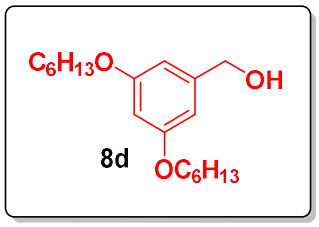
Mol. weight: 350.25 g/mol

Physical appearance: colorless liquid

Yield: 96 %

The **compound 7d** was prepared from ethyl 3,5-dihydroxybenzoate (5 g, 27 mmol), 1-hexyl bromide (11.5 g, 80 mmol), K_2CO_3 (11.5 g, 70 mmol), KI (0.22 g, 1.3 mmol) in DMF. The product was obtained as a colourless liquid (9 g, 25 mmol, 96 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.46 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.16 (d, J = 2.4 Hz, 2H), 6.63 (t, J = 2.4 Hz, 1H), 4.34 (q, J = 7.2 Hz, 2H), 3.97 (t, J = 6.8 Hz, 4H), 1.77 (quint, J = 6.8 Hz, 4H), 1.49-1.31 (m, 15H), 0.90 (t, J = 6.8 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 166.67, 160.25, 132.32, 107.74, 106.41, 68.43, 61.21, 31.70, 29.29, 25.83, 22.74, 14.47, 14.18. **MALDI-TOF MS** (M+K): 389.20 g/mol.

Synthesis of compound 8d



Mol. formula: $C_{19}H_{32}O_3$

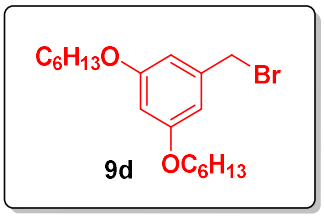
Mol. weight: 307.24 g/mol

Physical appearance: colorless liquid

Yield: 70 %

The **compound 8d** was synthesized from **compound 7d** (9.2 g, 25 mmol), LAH (2.9 g, 80 mmol) in THF. The product was obtained as a colourless liquid (6 g, 20 mmol, 70 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.32 in 10 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.49 (d, *J* = 2.4 Hz, 2H), 6.38 (t, *J* = 2.4 Hz, 1H), 4.61 (s, 2H), 3.93 (t, *J* = 6.8 Hz, 4H), 1.76 (quint, *J* = 6.4 Hz, 4H), 1.48-1.39 (m, 4H), 1.37-1.30 (m, 8H), 0.91 (t, *J* = 6.8 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 160.61, 143.34, 105.13, 100.68, 68.16, 65.52, 31.70, 29.33, 25.84, 22.74, 14.17. **HRMS** (M+H): 308.24 g/mol.

Synthesis of compound 9d



9d

Mol. formula: $C_{19}H_{31}BrO_2$

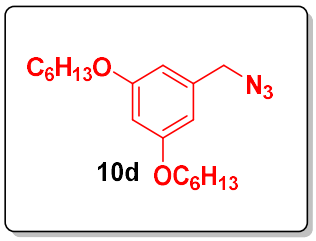
Mol. weight: 370.56 g/mol

Physical appearance: colorless liquid

Yield: 75 %

The **compound 9d** was synthesized from **compound 8d** (2.0 g, 6 mmol), PPh_3 (2.7 g, 10 mmol) and CBr_4 (3.4 g, 10 mmol) in THF. The product was obtained as a colourless liquid (1.8 g, 5 mmol, 75 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.49 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.52 (d, *J* = 2 Hz, 2H), 6.38 (t, *J* = 2 Hz, 1H), 4.41 (s, 2H), 3.92 (t, *J* = 6.8 Hz, 4H), 1.76 (quint, *J* = 6.4 Hz, 4H), 1.48-1.39 (m, 4H), 1.37-1.30 (m, 8H), 0.91 (t, *J* = 6.8 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 160.53, 139.66, 107.43, 101.50, 68.21, 33.95, 31.70, 29.31, 25.84, 22.74, 14.18. **MALDI-TOF MS** (M+Na): 393.11 g/mol.

Synthesis of compound 10d



Mol. formula: $C_{19}H_{31}N_3O_2$

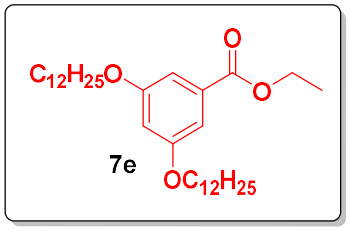
Mol. weight: 333.47 g/mol

Physical appearance: yellowish liquid

Yield: 75 %

The **compound 10d** was synthesized from **compound 9d** (1.5 g, 4 mmol), NaN_3 (0.46 g, 7 mmol) in DMF. The product was obtained as a yellowish liquid (1.0 g, 3 mmol, 75 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.48 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.44 (d, J = 2.4 Hz, 2H), 6.41 (t, J = 2.4 Hz, 1H), 4.24 (s, 2H), 3.93 (t, J = 6.8 Hz, 4H), 1.77 (quint, J = 6.8 Hz, 4H), 1.48-1.39 (m, 4H), 1.37-1.30 (m, 8H), 0.91 (t, J = 6.8 Hz, 6H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 160.70, 137.54, 106.46, 101.50, 68.21, 55.05, 31.70, 29.31, 25.84, 22.74, 14.18. **MALDI-TOF MS** (M+K): 372.25 g/mol.

Synthesis of compound 7e



Mol. formula: $C_{33}H_{58}O_4$

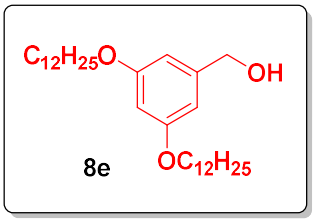
Mol. weight: 518.43 g/mol

Physical appearance: white solid

Yield: 78 %

The **compound 7e** was prepared from ethyl 3,5-dihydroxybenzoate (4.0 g, 20 mmol), 1-dodecyl bromide (13.5 g, 55 mmol), K_2CO_3 (12.1 g, 87 mmol), KI (0.2 g, 3 mmol) in acetonitrile (ACN) / THF (2:1). The product was obtained as a white solid (9.0 g, 17 mmol, 78 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.48 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.16 (d, J = 2 Hz, 2H), 6.63 (t, J = 2 Hz, 1H), 4.36 (q, J = 6.8 Hz 2H), 3.96 (t, J = 6.8 Hz, 4H), 1.77 (quint, J = 6.8 Hz, 4H), 1.48-1.20 (m, 41H), 0.88 (t, J = 7.2 Hz, 6H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 166.67, 160.24, 132.31, 107.73, 106.39, 68.43, 61.21, 32.07, 29.81, 29.79, 29.75, 29.73, 29.59, 29.51, 29.33, 26.16, 22.84, 14.47, 14.28. **HRMS** (M+H): 519.44 g/mol.

Synthesis of compound 8e



Mol. formula: $C_{31}H_{56}O_3$

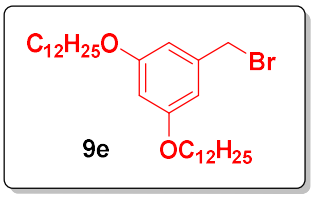
Mol. weight: 476.42 g/mol

Physical appearance: white solid

Yield: 86 %

The **compound 8e** was synthesized from **compound 7e** (5.0 g, 10 mmol), LAH (1.25 g, 32 mmol) in THF. The product was obtained as a white solid (3.7 g, 8 mmol, 86 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.32 in 10 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.49 (d, J = 2 Hz, 2H), 6.37 (t, J = 2 Hz, 1H), 4.61 (s, 2H), 3.93 (t, J = 6.8 Hz, 4H) 1.76 (quint, J = 6.4 Hz, 4H), 1.47-1.20 (m, 36H), 0.88 (t, J = 7.2 Hz, 6H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 160.65, 143.30, 105.14, 100.63, 68.18, 65.61, 32.06, 29.81, 29.78, 29.75, 29.73, 29.53, 29.503, 29.39, 26.18, 22.84, 14.28. **HRMS** (M+H): 477.43 g/mol.

Synthesis of compound 9e



Mol. formula: $C_{31}H_{55}BrO_2$

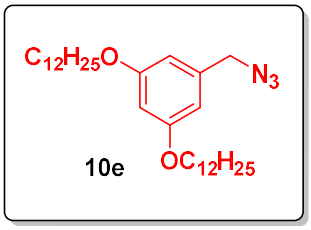
Mol. weight: 538.33 g/mol

Physical appearance: white solid

Yield: 85 %

The **compound 9e** was synthesized from **compound 8e** (3.5 g, 7 mmol), PPh_3 (3.3 g, 14 mmol) and CBr_4 (5.0 g, 14 mmol) in THF. The product was obtained as a white solid (3.3 g, 6 mmol, 85 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.49 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.51 (d, J = 2.4 Hz, 2H), 6.38 (t, J = 2.4 Hz, 1H), 4.40 (s, 2H), 3.92 (t, J = 6.8 Hz, 4H) 1.76 (quint, J = 6.8 Hz, 4H), 1.47-1.20 (m, 37H), 0.88 (t, J = 7.2 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 160.53, 139.64, 107.48, 101.51, 68.23, 33.96, 32.06, 29.81, 29.78, 29.75, 29.73, 29.53, 29.50, 29.39, 26.18, 22.84, 14.28. **HRMS** (M+H): 539.34 g/mol.

Synthesis of compound 10e



Mol. formula: $C_{31}H_{55}N_3O_2$

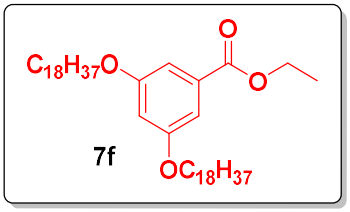
Mol. weight: 501.80 g/mol

Physical appearance: white solid

Yield: 86 %

The **compound 10e** was synthesized from bromide, **compound 9e** (1.2 g, 2 mmol), NaN_3 (0.21 g, 4 mmol) in DMF. The product was obtained as a white solid (0.95 g, 2 mmol, 86 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.48 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.43 (d, J = 2 Hz, 2H), 6.41 (t, J = 2 Hz, 1H), 4.25 (s, 2H), 3.93 (t, J = 6.8 Hz, 4H) 1.76 (quint, J = 7.2 Hz, 4H), 1.47-1.20 (m, 36H), 0.88 (t, J = 7.2 Hz, 6H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 160.73, 137.50, 106.56, 101.13, 68.24, 55.07, 32.06, 29.81, 29.78, 29.75, 29.73, 29.53, 29.50, 29.39, 26.18, 22.84, 14.28. **HRMS** (M+H): 502.43 g/mol.

Synthesis of compound 7f



Mol. formula: $C_{45}H_{82}O_4$

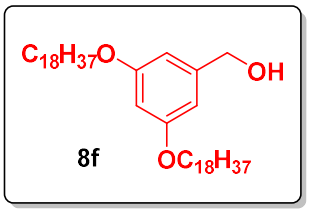
Mol. weight: 686.62 g/mol

Physical appearance: white solid

Yield: 91 %

The **compound 7f** was prepared from ethyl 3,5-dihydroxybenzoate (1.0 g, 5 mmol), 1-octadecyl bromide (6 g, 15 mmol), K_2CO_3 (4.0 g, 30 mmol), KI (0.045 g, 0.30 mmol), in DMF. The product was obtained as a white solid (3.4 g, 5 mmol, 91 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.46 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.37 (s, 2H), 6.66 (s, 1H), 4.32 (q, J = 6.8 Hz, 2H), 3.96 (t, J = 7.2 Hz, 4H), 1.77 (quint, J = 6.8 Hz, 4H), 1.48-1.20 (m, 67H), 0.88 (t, J = 7.2 Hz, 6H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 166.67, 160.28, 132.36, 107.81, 106.48, 68.48, 61.20, 32.08, 29.85, 29.81, 29.76, 29.73, 29.64, 29.53, 29.51, 29.35, 29.17, 22.84, 14.47, 14.26. **HRMS** (M+H): 687.62 g/mol.

Synthesis of compound 8f



Mol. formula: $C_{43}H_{80}O_3$

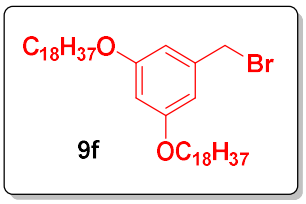
Mol. weight: 644.61 g/mol

Physical appearance: white solid

Yield: 96 %

The **compound 8f** was synthesized from **compound 7f** (7.0 g, 10 mmol), LAH (1.3 g, 33 mmol) in THF. The product was obtained as a white solid (7.0 g, 10 mmol, 96 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.32 in 10 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.50 (s, 2H), 6.37 (s, 1H), 4.61 (s, 2H), 3.94 (t, J = 6.4 Hz, 4H), 1.76 (quint, J = 6.4 Hz, 4H), 1.48-1.39 (m, 4H), 1.35-1.20 (m, 62H), 0.91 (t, J = 6.8 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 160.59, 143.36, 105.11, 100.59, 77.16, 68.14, 65.34, 32.07, 29.86, 29.85, 29.83, 29.82, 29.78, 29.75, 29.56, 29.52, 29.41, 26.20, 22.83, 14.23. **HRMS** (M+Na): 688.28 g/mol.

Synthesis of compound 9f



Mol. formula: $C_{43}H_{79}BrO_3$

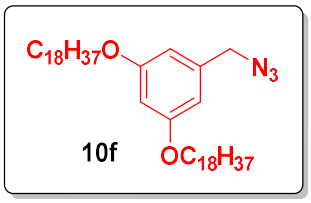
Mol. weight: 708.52 g/mol

Physical appearance: white solid

Yield: 82 %

The **compound 9f** was synthesized from **compound 8f** (3.0 g, 4.6 mmol), PPh_3 (2.4 g, 9.0 mmol) and CBr_4 (3.0 g, 9.0 mmol) in THF. The product was obtained as a white solid (2.6 g, 3.5 mmol, 82 %) after purification by NPC using EtOAc / PE eluent, R_f = 0.49 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.51 (s, 2H), 6.39 (s, 1H), 4.43 (s, 2H), 3.94 (t, J = 6.4 Hz, 4H), 1.76 (quint, J = 6.4 Hz, 4H), 1.48-1.39 (m, 4H), 1.35-1.20 (m, 60H), 0.91 (t, J = 6.8 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 160.58, 139.71, 107.81, 106.48, 68.48, 61.20, 32.08, 29.85, 29.81, 29.76, 29.73, 29.64, 29.53, 29.51, 29.35, 29.17, 22.84, 14.26. **HRMS** (M+H): 709.56 g/mol.

Synthesis of compound 10f



Mol. formula: $C_{43}H_{79}N_3O_3$

Mol. weight: 670.61 g/mol

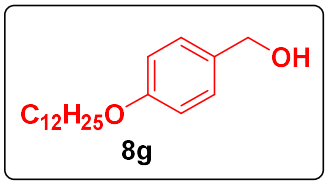
Physical appearance: white solid

Yield: 79 %

The **compound 10f** was synthesized starting from **compound 9f** (0.30 g, 0.4 mmol), NaN_3 (0.11 g, 1.6 mmol) and TBAB (0.10 g, 0.3 mmol) in THF. The product was obtained as a white solid (0.15 g, 0.3 mmol, 79 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.48 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 6.43 (s, 2H), 6.41 (s, 1H), 4.25 (s, 2H), 3.93 (t, J = 6.4 Hz, 4H), 1.76 (quint, J = 6.8 Hz, 4H), 1.48-1.39 (m, 4H), 1.35-1.20 (m, 59H), 0.87 (t, J = 7.2 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 160.79, 137.59, 107.81, 106.65, 101.26, 68.31, 55.12, 32.09, 29.85, 29.81, 29.76, 29.73, 29.64, 29.53, 29.51, 29.35, 29.17, 26.20, 22.84, 14.26. **HRMS** (M+Na): 693.28 g/mol.

Synthesis of 1T hydrophobic azides and their intermediates

Synthesis of compound 8g



Mol. formula: $C_{19}H_{32}O_2$

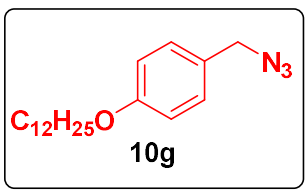
Mol. weight: 292.24 g/mol

Physical appearance: white solid

Yield: 73 %

The **compound 8g** was prepared from 4-hydroxybezyl alcohol (0.50 g, 4.2 mmol), 1-bromododecane (1.0 g, 4.4 mmol), K_2CO_3 (0.85 g, 5.1 mmol), 18-crwon-6 (0.10 g, 0.4 mmol) in acetone. The product was obtained as a white solid (0.90 g, 3.1 mmol, 73 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.25 in 75 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.23 (d, J = 8.8 Hz, 2H), 6.87 (d, J = 8.8 Hz, 2H), 4.56 (s, 2H), 3.93 (t, J = 6.8 Hz, 2H), 1.80-1.73 (m, 2H), 1.47-1.26 (m, 18H), 0.88 (t, J = 7.2 Hz, 3H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 158.93, 133.03, 128.75, 114.70, 77.16, 68.22, 65.23, 32.06, 29.80, 29.78, 29.74, 29.72, 29.54, 29.49, 29.40, 26.18, 22.83, 14.26. **MALDI-TOF MS** (M+K): 331.22 g/mol.

Synthesis of compound 10g



Mol. formula: $C_{19}H_{31}O$

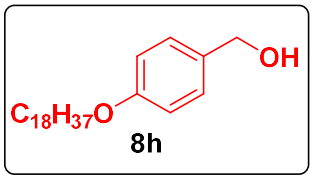
Mol. weight: 317.26 g/mol

Physical appearance: white solid

Yield: 60 %

The **compound 9g** a chloride was achieved starting from **compound 8g** (0.50 g, 1.7 mmol), $SOCl_2$ (0.32 g, 2.7 mmol) and catalytic amount of DMF in DCM. Subsequently azide, **compound 10g** was synthesized using NaN_3 (0.15 g, 2.5 mmol) in DMF using impure **compound 9g**. The product was obtained as a white solid (0.32 g, 1.0 mmol, 60 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.44 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.22 (d, J = 8.8 Hz, 2H), 6.88 (d, J = 8.8 Hz, 2H), 4.25 (s, 2H), 3.93 (t, J = 6.8 Hz, 2H), 1.80-1.73 (m, 2H), 1.51-1.30 (m, 18H), 0.90 (t, J = 7.2 Hz, 3H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 159.28, 129.70, 127.20, 114.75, 77.16, 68.04, 54.44, 32.02, 29.77, 29.74, 29.71, 29.68, 29.50, 29.45, 29.34, 26.13, 22.78, 14.17. **MALDI-TOF MS** (M+K): 356.22 g/mol.

Synthesis of compound 8h



8h

Mol. formula: $C_{25}H_{44}O_2$

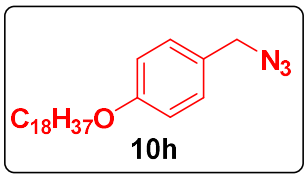
Mol. Weight: 376.33 g/mol

Physical appearance: White solid

Yield: 68 %

The **compound 8h** was prepared from 4-hydroxybezyl alcohol (0.52 g, 4.2 mmol), 1-bromooctadecane (1.5 g, 4.6 mmol), K_2CO_3 (0.85 g, 5.1 mmol), 18-crwon-6 (0.10 g, 0.4 mmol) in acetone. The product was obtained as a white solid (1.0 g, 2.8 mmol, 68 %) after purification by NPC using DCM / PE as eluent, R_f = 0.25 in 10 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.29 (d, J = 8.8 Hz, 2H), 6.90 (d, J = 8.8 Hz, 2H), 4.62(s, 2H), 3.97 (t, J = 6.8 Hz, 2H), 1.88-1.70 (m, 2H), 1.51-1.21 (m, 34H), 0.91 (t, J = 7.2 Hz, 3H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 158.95, 133.04, 128.76, 114.71, 77.16, 68.22, 65.25, 32.08, 29.85, 29.81, 29.76, 29.73, 29.56, 29.51, 29.42, 26.19, 22.84, 14.27. **MALDI-TOF MS** (M+K): 415.31 g/mol.

Synthesis of compound 10h



Mol. formula: $C_{25}H_{43}N_3O$

Mol. weight: 401.34 g/mol

Physical appearance: white solid

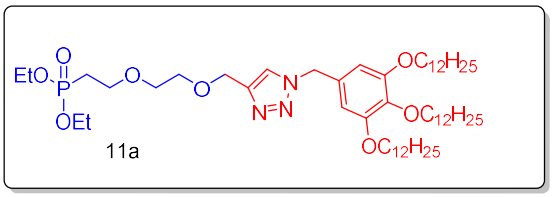
Yield: 72 %

The **compound 9h**, a chloride was achieved starting from **compound 8h** (0.30 g, 0.9 mmol), $SOCl_2$ (0.17 g, 1.5 mmol) and catalytic amount of DMF in DCM. Subsequently, using **compound 9h**, an azide, **compound 10h** was synthesized using NaN_3 (0.09 g, 1.4 mmol) and tetra-n-butylammonium bromide (TBAB) (0.3 g, 0.9 mmol) in tetrahydrofuran (THF). The product was obtained as white solid (0.27 g, 0.6 mmol, 72 %) after purification by NPC using EtOAc / PE as eluent, R_f = 0.44 in 5 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.25 (d, J = 8.8 Hz, 2H), 6.92 (d, J = 8.8 Hz, 2H), 4.29(s, 2H), 3.99 (t, J = 6.8 Hz, 2H), 1.87-1.74 (m, 2H), 1.53-1.25 (m, 32H), 0.92 (t, J = 7.2 Hz, 3H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 159.35, 129.81, 127.26, 114.86, 77.16, 68.18, 54.56, 32.07, 29.85, 29.81, 29.75, 29.73, 29.54, 29.51, 29.39, 26.18, 22.84, 14.25. **MALDI-TOF MS** (M+Na): 424.35 g/mol.

Synthesis of AABPs and their intermediates

Synthesis of intermediates for spacer variant AABPs

Synthesis of compound 11a



Mol. formula: $C_{54}H_{100}N_3O_{18}P$

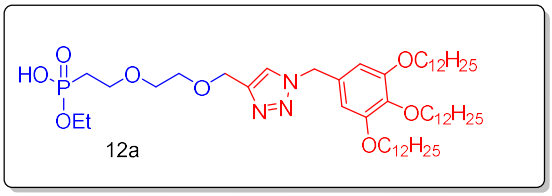
Mol. weight: 949.72 g/mol

Physical appearance: brownish liquid

Yield: 68 %

The **compound 11a** was synthesized from **compound 4a** (100 mg, 0.39 mmol) and **compound 10b** (azide) (271 mg, 0.39 mmol), $CuSO_4$ (3.1 mg, 0.02 mmol), Na Asc. (7.8 mg, 0.04 mmol) in 50 % THF / H_2O. The crude product obtained was purified using reverse phase column chromatography using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid (250 mg, 0.25 mmol, 68 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.49 (s, 1H), 6.46 (s, 2H), 5.38 (s, 2H), 4.65 (s, 2H), 4.17-4.00 (m, 4H), 3.95-3.90 (m, 6H), 3.75-3.50 (m, 5H), 2.15-2.03 (m, 2H), 184-1.68 (m, 13H), 1.54-1.19 (m, 62H), 0.87 (t, J = 6.8 Hz, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 153.79, 138.75, 129.55, 107.01, 77.16, 73.68, 70.39, 69.86, 69.51, 65.32, 64.95, 61.84, 54.71, 32.14, 30.53, 29.85, 29.63, 27.87, 26.33, 22.90, 16.53, 14.33. **MALDI-TOF MS** (M+K): 988.33 g/mol.

Synthesis of compound 12a



12a

Mol. formula: $C_{52}H_{96}N_3O_8P$

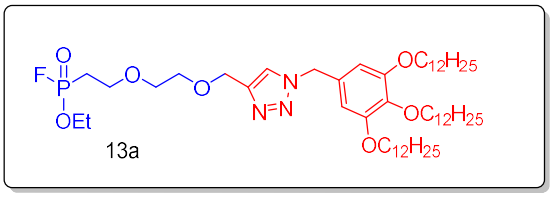
Mol. weight: 921.69 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12a** was synthesized from **compound 11a** (100 mg, 0.10 mmol), OxCl (49 mg, 0.42 mmol) in DCM. The product was carried to next step without purification. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.52 (s, 1H), 6.46 (s, 2H), 5.33 (s, 2H), 4.64 (s, 2H), 4.09-4.00 (m, 4H), 3.91-3.83 (m, 6H), 3.72-3.51 (m, 5H), 2.15-2.03 (m, 2H), 178-1.68 (m, 13H), 1.44-1.12 (m, 55H), 0.85 (t, J = 6.8 Hz, 9H). **MALDI-TOF MS** (M+K): 960.32 g/mol.

Synthesis of compound 13a



13a

Mol. formula: $C_{52}H_{95}N_3O_{12}PF$

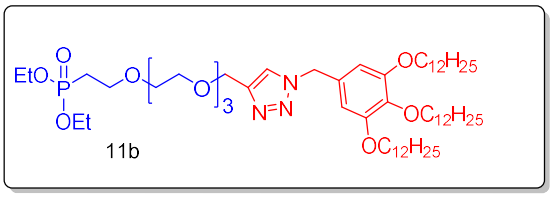
Mol. weight: 923.77 g/mol

Physical appearance: brownish liquid

Yield: not determined

The compound **22c** was synthesized from **22b** (80 mg, 0.08 mmol), DAST (50 mg, 0.34 mmol) in DCM. The product was used for modification without further purification. **^{19}F NMR** (400 MHz, $CDCl_3$) δ_F -59.91, -62.74. **MALDI-TOF MS** (M+Na): 946.78 g/mol.

Synthesis of compound 11b



11b

Mol. formula: $C_{58}H_{108}N_3O_{10}P$

Mol. weight: 1037.78 g/mol

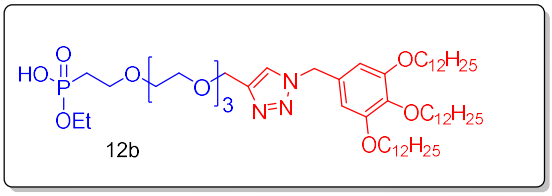
Physical appearance: brownish liquid

Yield: 62 %

The **compound 11b** was synthesized from **compound 4b** (250 mg, 0.70 mmol) and **compound 10b** (azide) (500 mg, 0.70 mmol), $CuSO_4$ (36 mg, 0.1 mmol), Na Asc. (14 mg, 0.070 mmol) in 50 % THF / H_2O. The crude product obtained was purified using reverse phase column chromatography using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid (330 mg, 0.30 mmol, 62 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H**

NMR (400 MHz, $CDCl_3$): δ_H 7.47 (s, 1H), 6.45 (s, 2H), 5.37 (s, 2H), 4.65 (s, 2H), 4.13-4.04 (m, 4H), 3.93-3.89 (q, *J* = 8 Hz, 6H), 3.74-3.56 (m, 15H), 2.15-2.07 (m, 2H), 1.80-1.70 (m, 10H), 1.45-1.25 (m, 68H), 0.87 (t, *J* = 8 Hz, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 153.74, 145.65, 138.86, 129.44, 122.56, 106.97, 77.16, 73.53, 70.73, 70.61, 70.32, 69.92, 69.43, 65.28, 64.80, 61.81, 61.74, 54.64, 32.08, 30.48, 29.90, 29.86, 29.79, 29.58, 29.44, 27.84, 26.46, 26.24, 22.84, 16.61, 16.55, 14.26. **MALDI-TOF MS** (M+K): 1076.79 g/mol.

Synthesis of compound12b



12b

Mol. formula: $C_{56}H_{104}N_3O_{10}P$

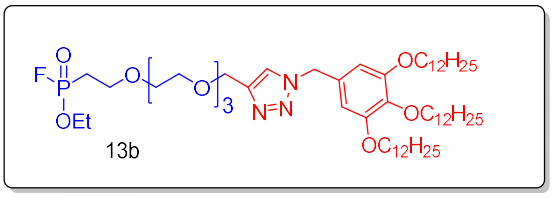
Mol. weight: 1009.75 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12b** was synthesized from **compound 11b** (296 mg, 0.28 mmol), OxCl (170 mg, 1.1 mmol) in DCM. The product was carried to next step without purification. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.50 (s, 1H), 6.49 (s, 2H), 5.40 (s, 2H), 4.60 (s, 2H), 4.12-4.03 (m, 2H), 3.94-3.91 (q, *J* = 8 Hz, 6H), 3.81-3.62 (m, 15H), 2.15-2.05 (m, 2H), 1.82-1.69 (m, 10H), 1.47-1.27 (m, 68H), 0.89 (t, *J* = 8 Hz, 9H). **MALDI-TOF MS** (M+K): 1048.84 g/mol.

Synthesis of compound 13b



13b

Mol. formula: $C_{56}H_{103}N_3O_9PF$

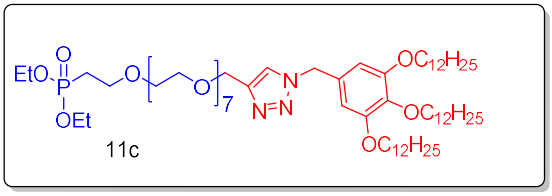
Mol. weight: 1011.74 g/mol

Physical appearance: brownish liquid

Yield: not determined

The **compound 13b** was synthesized from **compound 12b** (170 mg, 0.10 mmol), DAST (11 mg, 0.6 mmol) in DCM. The product was used for modification without further purification. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (M+Na): 1034.88 g/mol.

Synthesis of compound 11c



11c

Mol. formula: $C_{66}H_{124}N_3O_{14}P$

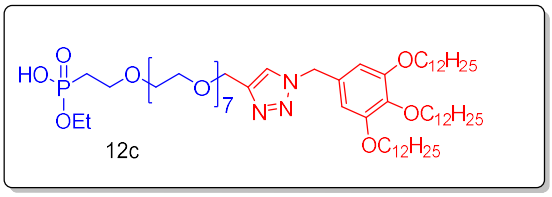
Mol. weight: 1213.88 g/mol

Physical appearance: brownish liquid

Yield: 65 %

The **compound 11c** was synthesized from **compound 4c** (200 mg, 0.37 mmol) and **compound 10b** (azide) (259 mg, 38 mmol), $CuSO_4$ (4.71 mg, 0.018 mmol), Na Asc. (7.5 mg, 0.037 mmol) in 50% THF / H_2O. The product obtained was purified using RPC using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid (0.325 g, 0.26 mmol, 65 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.49 (s, 1H), 6.45 (s, 2H), 5.37 (s, 2H), 4.65 (s, 2H), 4.14-4.04 (m, 4H), 3.98-3.88 (m, 6H), 3.74-3.56 (m, 31H), 2.16-2.06 (m, 2H), 1.80-1.68 (m, 6H), 1.48-1.38 (m, 6H), 1.33-1.20 (m, 57H), 0.87 (t, J = 6.8 Hz, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 153.51, 138.52, 129.47, 106.81, 73.50, 70.64, 70.60, 70.54, 70.50, 70.23, 69.81, 69.29, 65.18, 64.79, 61.70, 61.64, 31.99, 30.38, 29.81, 29.79, 29.72, 29.69, 29.66, 29.47, 29.42, 27.73, 26.35, 26.17, 26.14, 22.75, 16.52, 16.45, 14.17. **MALDI-TOF MS** (M+K): 1,252.84 g/mol.

Synthesis of compound 12c



12c

Mol. formula: $C_{64}H_{120}N_3O_{14}P$

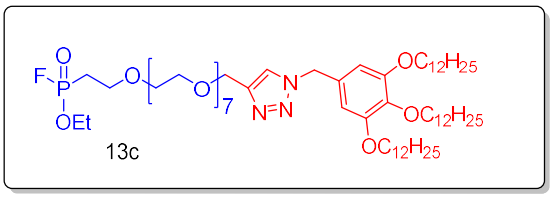
Mol. weight: 1185.85 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12c** was synthesized from **compound 11c** (257 mg, 0.21 mmol), OxCl (136 mg, 0.84 mmol) in DCM. The product obtained was carried to next step without purification. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.49 (s, 1H), 6.46 (s, 2H), 5.38 (s, 2H), 4.66 (s, 2H), 4.30-4.25 (m, 2H), 3.94-3.68 (m, 38H), 2.31-2.21 (m, 2H), 1.79-1.69 (m, 6H), 1.39-1.26 (m, 68H), 0.88 (t, J = 6.8 Hz, 9H). **MALDI-TOF MS** (M+Na): 1,208.90 g/mol.

Synthesis of compound 13c



13c

Mol. formula: $C_{64}H_{119}N_3O_{14}PF$

Mol. weight: 1187.84 g/mol

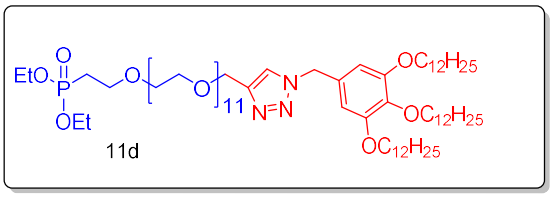
Physical appearance: brownish liquid

Yield: not determined

The **compound 13c** was synthesized from **compound 12c** (250 mg, 0.21 mmol), DAST (129 mg, 0.80 mmol) in DCM. The product obtained was used for modification without

purification. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (M+K): 1,226.77 g/mol.

Synthesis of compound 11d



11d

Mol. formula: $C_{74}H_{140}N_3O_{18}P$

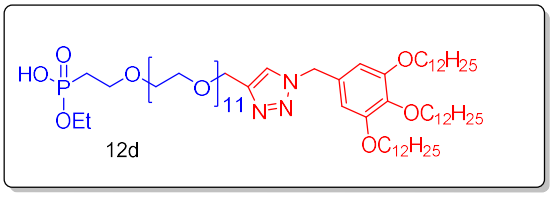
Mol. weight: 1390.99 g/mol

Physical appearance: brownish liquid

Yield: 57 %

The **compound 11d** was synthesized from **compound 4d** (150 mg, 0.21 mmol) and **compound 10b** (azide) (219 mg, 0.31 mmol), $CuSO_4$ (10 mg, 0.040 mmol), Na Asc. (4 mg, 0.020 mmol) in 50 % THF / H_2O. The crude product obtained was purified using reverse phase column chromatography using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid (168 mg, 0.12 mmol, 57 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.48 (s, 1H), 6.44 (s, 2H), 5.36 (s, 2H), 4.64 (s, 2H), 4.10-4.03 (m, 4H), 3.88 (q, J = 8 Hz, 6H), 3.73-3.57 (m, 55H), 2.15-2.06 (m, 2H), 1.80-1.66 (m, 13H), 1.48-1.23 (m, 74H), 0.85 (t, J = 8 Hz, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 153.65, 138.49, 129.49, 122.61, 106.79, 77.16, 73.56, 72.79, 70.30, 69.84, 69.31, 64.81, 61.75, 54.57, 32.03, 30.41, 29.85, 29.81, 29.75, 29.49, 29.47, 26.19, 22.80, 14.23. **MALDI-TOF MS** (M+K): 1,429.61 g/mol.

Synthesis of compound 12d



12d

Mol. formula: $C_{72}H_{136}N_3O_{18}P$

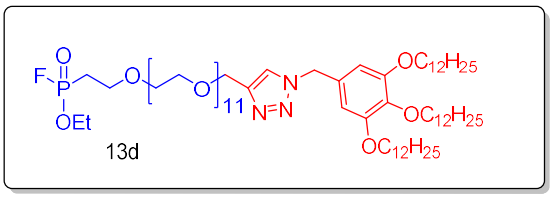
Mol. weight: 1361.66 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12d** was synthesized from **compound 11d** (45 mg, 0.043 mmol), OxCl (22 mg, 0.17 mmol) in DCM. The product was carried to next step without purification; **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.50 (s, 1H), 6.46 (s, 2H), 5.38 (s, 2H), 4.66 (s, 2H), 4.14-4.06 (m, 2H), 3.90 (q, J = 8 Hz, 6H), 3.78-3.63 (m, 55H), 2.16-2.08 (m, 2H), 1.80-1.68 (m, 13H), 1.44-1.26 (m, 74H), 0.87 (t, J = 8 Hz, 9H). **MALDI-TOF MS** (M+K): 1,400.60 g/mol.

Synthesis of compound 13d



13d

Mol. formula: $C_{72}H_{135}N_3O_{17}PF$

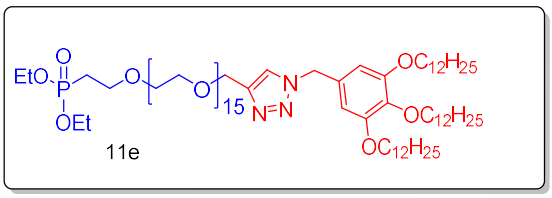
Mol. weight: 1363.66 g/mol

Physical appearance: brownish liquid

Yield: not determined

The **compound 13d** was synthesized from **compound 12d** (44 mg, 0.043 mmol), DAST (257 mg, 0.16 mmol) in DCM. The product was used for modification without further purification. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F - 59.91, -62.74. **MALDI-TOF MS** (M+K): 1402.68 g/mol.

Synthesis of compound 11e



11e

Mol. formula: $C_{82}H_{156}N_3O_{22}P$

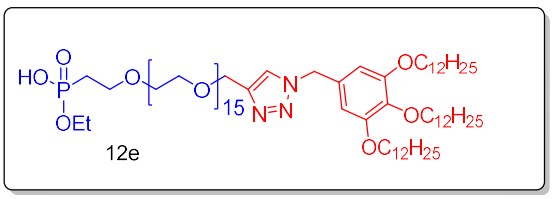
Mol. weight: 1565.69 g/mol

Physical appearance: brownish liquid

Yield: 48 %

The **compound 11e** was synthesized from **compound 4e** (170 mg, 0.19 mmol), **compound 10b** (azide) (160 mg, 0.23 mmol), $CuSO_4$ (9 mg, 0.03 mmol), Na Asc. (3 mg, 0.01 mmol) in 50 % THF / H_2O. The crude product obtained was purified using reverse phase column chromatography using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid (40 mg, 0.09 mmol, 48%), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, CD_3OD): δ_H 7.96 (s, 1H), 6.60 (m, 3H), 5.47 (s, 2H), 4.60 (s, 2H), 4.12-4.05 (m, 4H), 3.94-3.86 (m, 6H), 3.72-3.56 (m, 36H), 3.29-3.27 (m, 13H), 2.16-2.08 (m, 2H), 1.78-1.65 (m, 6H), 1.51-1.27 (m, 68H), 0.88 (t, J = 8 Hz, 9H). **^{13}C NMR (**100 MHz, CD_3OD): δ_C 154.67, 125.14, 124.54, 107.55, 74.44, 71.69, 71.63, 71.57, 71.54, 71.28, 70.78, 70.01, 49.00, 33.15, 30.99, 30.91, 30.87, 30.81, 30.59, 30.57, 27.39, 23.80, 14.52.. **MALDI-TOF MS** (M+K): 1,604.78 g/mol.

Synthesis of compound 12e



12e

Mol. formula: $C_{80}H_{152}N_3O_{22}P$

Mol. weight: 1538.67 g/mol

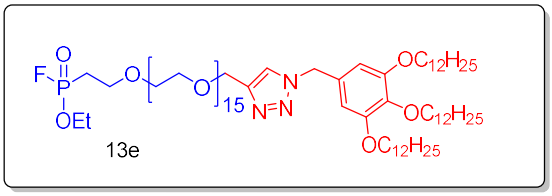
Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12e** was synthesized from **compound 11e** (145 mg, 0.09 mmol), OxCl(71 mg, 0.50 mmol) in DCM. The product was carried to next step without purification. **^{1}H NMR** (400 MHz, CD_3OD): δ_H 7.96 (s, 1H), 6.59 (m, 3H), 5.46 (s, 2H), 4.60 (s, 2H), 3.93-3.86 (m,

2H), 3.78-3.69 (m, 6H), 3.61-3.56 (m, 36H), 3.29-3.27 (m, 13H), 2.37-2.30 (m, 2H), 1.77-1.65 (m, 6H), 1.50-1.11 (m, 68H), 0.87 (t, J = 8 Hz, 9H). **MALDI-TOF MS** (M+K): 1,578.69 g/mol.

Synthesis of compound 13e



13e

Mol. formula: $C_{80}H_{151}N_3O_{21}PF$

Mol. weight: 1541.10 g/mol

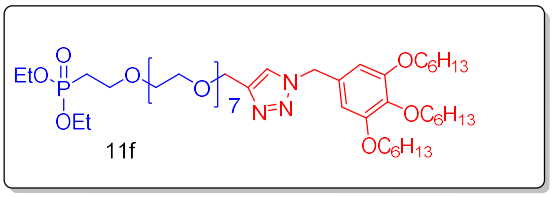
Physical appearance: brownish liquid

Yield: not determined

The **compound 13e** was synthesized from **compound 12e** (44 mg, 0.043 mmol), DAST (18 mg, 0.11 mmol) in DCM. The product was used for modification without further purification. **^{19}F NMR** (400 MHz, CD_3OD): δ_F -59.91, -62.74; **MALDI-TOF MS** (M+K): 1,580.56 g/mol.

Synthesis of 3-tail variant AABPs and their intermediates

Synthesis of compound 11f



11f

Mol. formula: $C_{48}H_{88}N_3O_{14}P$

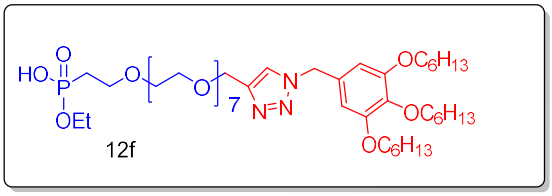
Mol. weight: 961.60 g/mol

Physical appearance: brownish liquid

Yield: 60 %

The **compound 11f** was synthesized from **compound 4c** (alkyne) (182 mg, 334 mmol), **10a** (115 mg, 334 mmol), $CuSO_4$ (2.3 mg, 0.094 mmol) Na Asc. (3.7 mg, 0.0189 mmol) in 50% THF / H_2O. The crude product obtained was purified using RPC using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid (215 mg, 229 mmol, 60 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.49 (s, 1H), 6.45 (s, 2H), 5.38 (s, 2H), 4.65 (s, 2H), 4.11-4.05 (m, 4H), 3.94-3.86 (m, 6H), 3.74-3.56 (m, 34H), 2.16-2.08 (m, 2H), 1.77-1.67 (6H), 1.49-1.27 (m, 27H), 0.89 (m, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 153.28, 129.39, 106.64, 77.16, 73.30, 70.07, 69.65, 69.09, 65.01, 64.62, 61.54, 54.32, 53.43, 31.63, 31.44, 29.21, 25.62, 22.49, 16.29, 13.91. **MALDI-TOF MS** (M+K): 1,000.62 g/mol.

Synthesis of compound 12f



12f

Mol. formula: $C_{46}H_{84}N_3O_{14}P$

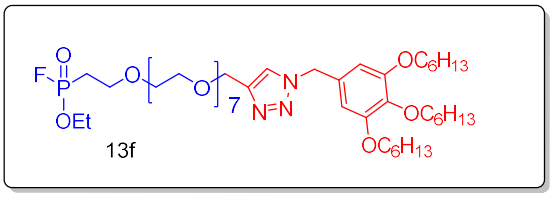
Mol. weight: 933.56 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12f** was synthesized from **compound 11f** (53 mg, 51 mmol), OxCl (28 mg, 0.114 mmol) in DCM. The product obtained was carried to next step without purification. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.48 (s, 1H), 6.42 (s, 2H), 5.34 (s, 2H), 4.61 (s, 2H), 3.90-3.81 (m, 8H), 3.62-3.53 (m, 32H), 1.87-1.64 (m, 8H), 1.29- 10.85 (m, 22H), 0.83 (m, 9H). **MALDI-TOF MS** (M+K): 972.63 g/mol.

Synthesis of compound 13f



13f

Mol. formula: $C_{46}H_{83}N_3O_{14}PF$

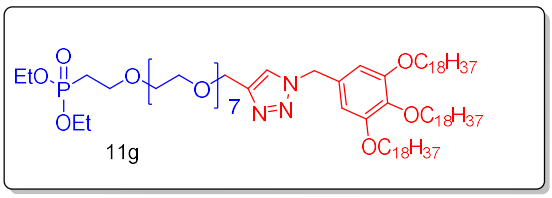
Mol. weight: 935.56 g/mol

Physical appearance: brownish liquid

Yield: not determined

The **compound 13f** was synthesized from **compound 12f** (45 mg, 0.042 mmol), DAST (31 mg, 0.19 mmol) in DCM. The product obtained was used for modification without further purification. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (M+K): 974.62 g/mol.

Synthesis of compound 11g



11g

Mol. formula: $C_{84}H_{160}N_3O_{14}P$

Mol. weight: 1466.16 g/mol

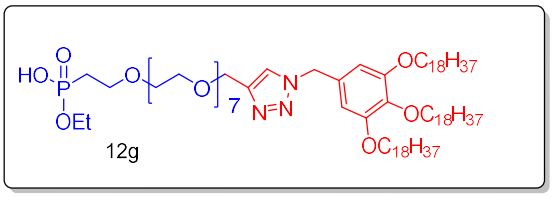
Physical appearance: brownish liquid

Yield: 82 %

The **compound 11g** was synthesized from **compound 4c** (alkyne) (86 mg, 0.16 mmol), **compound 10c** (150 mg, 0.16 mmol), $CuSO_4$ (8.1 mg, 0.032 mmol), Na Asc. (3.2 mg, 0.016 mmol) in 50:25:25 THF / H_2O / DCM. The crude product obtained was purified using RPC using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid

(195 mg, 0.13 mmol, 82 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.48 (s, 1H), 6.44 (s, 2H), 5.37 (s, 2H), 4.64 (s, 2H), 4.12-4.03 (m, 4H), 3.95-3.85 (m, 6H), 3.73-3.59 (m, 32H), 2.17-1.99 (m, 2H), 1.79-1.69 (m, 6H), 1.44-1.24 (m, 100H), 0.86 (t, J = 7.2 Hz, 9H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 153.66, 138.54, 129.50, 106.83, 77.16, 73.48, 70.64, 70.58, 70.54, 70.27, 69.84, 69.33, 65.22, 64.81, 61.76, 54.56, 53.54, 32.03, 30.43, 29.86, 29.83, 29.76, 29.72, 29.47, 27.76, 26.38, 26.20, 22.79, 16.56, 16.50, 14.22. **MALDI-TOF MS** (M+K): 1,505.23 g/mol.

Synthesis of compound 12g



12g

Mol. formula: $C_{82}H_{156}N_3O_{14}P$

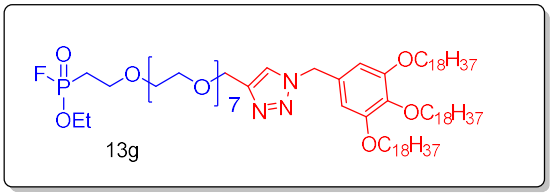
Mol. weight: 1438.13 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12g** was synthesized from **compound 11g** (195 mg, 0.13 mmol), OxCl (85 mg, 0.52 mmol) in DCM. The product obtained was carried to next step without purification. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.49 (s, 1H), 6.6 (s, 2H), 5.37 (s, 2H), 4.65 (s, 2H), 4.11-4.04 (m, 2H), 3.92-3.88 (m, 6H), 3.77-3.62 (m, 31H), 2.16-2.05 (m, 2H), 1.79-1.68 (m, 6H), 1.46-1.20 (m, 113H), 0.87 (t, J = 7.2 Hz, 9H). **MALDI-TOF MS** (M+K): 1,477.21 g/mol.

Synthesis of compound 13g



13g

Mol. formula: $C_{82}H_{155}N_3O_{14}PF$

Mol. weight: 1440.12 g/mol

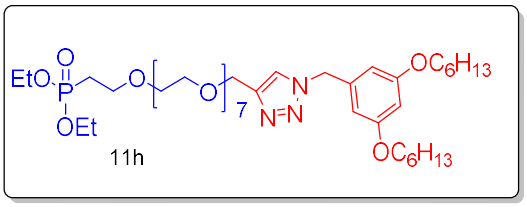
Physical appearance: brownish liquid

Yield: not determined

The **compound 13g** was synthesized from **compound 12g** (17 mg, 0.12 mmol), DAST (78 mg, 0.46 mmol) in DCM. The product obtained was used for modification without purification. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (M+K): 1,479.10 g/mol.

Synthesis of 2-tail variant AABPs and their intermediates

Synthesis of compound 11h



11h

Mol. formula: $C_{42}H_{76}N_3O_{13}P$

Mol. weight: 861.51 g/mol

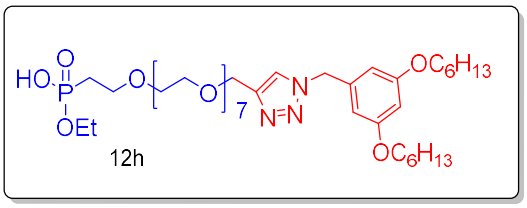
Physical appearance: brownish liquid

Yield: 65 %

The **compound 11h** was synthesized from **compound 4c** (alkyne) (470 mg, 0.90 mmol), **compound 10d** (300 mg, 0.90 mmol), $CuSO_4$ (11 mg, 0.044 mmol), Na Asc. (17 mg, 0.085 mmol) in 50% THF / H_2O. The crude product obtained was purified using reverse phase column chromatography using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid (500 mg, 0.82 mmol, 65 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.49 (s, 1H), 6.41-6.35 (m, 3H), 5.38 (S, 2H), 4.63 (s, 2H), 4.11-4.00 (m, 4H), 3.85 (t, J = 6.4 Hz, 4H), 3.72-3.52 (m, 32H), 2.12-2.03 (m, 2H), 1.75-1.70 (m, 4H), 1.46-1.27 (m, 19H), 0.872 (m, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 160.91, 145.62, 136.58, 122.63, 106.70, 101.38, 70.68, 70.64, 70.57, 70.54, 70.27, 69.83, 68.27, 65.21, 64.81, 61.74,

54.34, 31.63, 29.28, 27.88, 26.40, 25.76, 22.65, 16.47, 14.09. **MALDI-TOF MS** (M+K): 900.56 g/mol.

Synthesis of compound 12h



Mol. formula: $C_{40}H_{72}N_3O_{13}P$

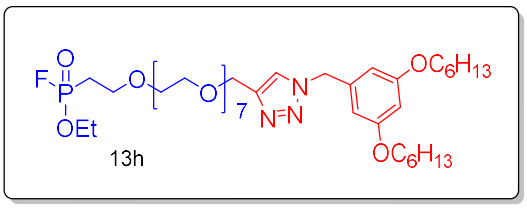
Mol. weight: 833.48 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12h** was synthesized from **compound 11h** (490 mg, 0.569 mmol), OxCl(290 mg, 2.2 mmol) in DCM. The product obtained was carried to next step without purification. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.49 (s, 1H), 6.38-6.36 (m, 3H), 5.38 (S, 2H), 4.64 (s, 2H), 4.10-4.03 (m, 2H), 3.88 (t, J = 6.4 Hz, 4H), 3.72-3.50 (m, 32H), 2.11-2.04 (m, 2H), 1.70-1.66 (m, 4H), 1.48-1.23 (m, 16H), 0.88 (m, 6H). **MALDI-TOF MS** (M+K): 872.37 g/mol.

Synthesis of compound 13h



Mol. formula: $C_{40}H_{71}N_3O_{12}PF$

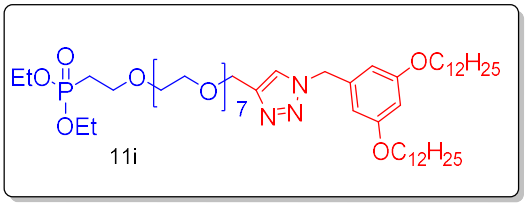
Mol. weight: 835.47 g/mol

Physical appearance: brownish liquid

Yield: not determined

The **compound 13h** was synthesized from **compound 12h** (425 mg, 0.510 mmol), DAST (374 mg, 2.3 mmol) in DCM. The product was used for modification without further purification. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (M+Na): 858.45 g/mol.

Synthesis of compound 11i



11i

Mol. formula: $C_{54}H_{100}N_3O_{13}P$

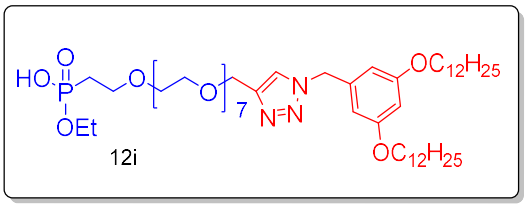
Mol. weight: 1029.69 g/mol

Physical appearance: brownish liquid

Yield: 68 %

The **compound 11i** was synthesized from **compound 4c** (alkyne) (100 mg, 0.19 mmol), **compound 10e** (0.94 g, 0.19 mmol), $CuSO_4$ (2.48 mg, 0.010 mmol), Na Asc. (3.94 mg, 0.0198 mmol) in 50 % THF / H_2O. The crude product obtained was purified using reverse phase column chromatography using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid (133 mg, 2.3 mmol, 68 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.50 (s, 1H), 6.40-6.37 (m, 3H), 5.39 (s, 2H), 4.65 (s, 2H), 4.14-4.03 (m, 4H), 3.88 (t, J = 6.4 Hz, 4H), 3.73-3.53 (m, 30H), 2.16-2.08 (m, 2H), 1.47-1.19 (m, 60H), 0.87 (t, J = 6.8 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 160.91, 145.62, 136.58, 122.63, 106.70, 101.38, 70.68, 70.64, 70.57, 70.54, 70.27, 69.83, 68.27, 65.21, 64.81, 61.74, 54.34, 31.63, 29.81, 29.79, 29.75, 29.73, 29.54, 29.51, 29.36, 26.18, 22.77, 16.47, 14.09. **MALDI-TOF MS** (M+K): 1,068.71 g/mol.

Synthesis of compound 12i



12i

Mol. formula: $C_{52}H_{96}N_3O_{13}P$

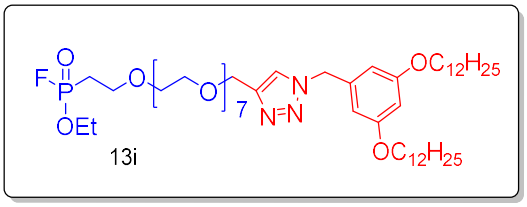
Mol. weight: 1001.66 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12i** was synthesized from **compound 11i** (45 mg, 0.043 mmol), OxCl(22 mg, 0.17 mmol) in DCM. The product obtained was carried to next step without purification. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.48 (s, 1H), 6.33-6.32 (m, 3H), 5.34 (s, 2H), 4.59 (s, 2H), 3.86-3.80 (m, 6H), 3.60-3.56 (m, 34H), 3.53-3.35 (m, 7H), 1.89-1.64 (m, 2H), 1.36–1.32 (m, 4H), 1.20–0.89 (m, 41H), 0.79 (t, J = 6.8 Hz, 6H). **MALDI-TOF MS** (M+K): 1,040.63 g/mol.

Synthesis of compound 13i



13i

Mol. formula: $C_{52}H_{95}N_3O_{12}PF$

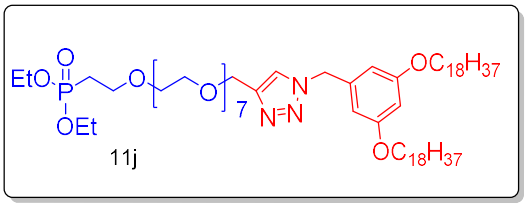
Mol. weight: 1003.66 g/mol

Physical appearance: brownish liquid

Yield: not determined

The **compound 13i** was synthesized from **compound 12i** (44 mg, 0.043 mmol), DAST (29 mg, 0.16 mmol) in DCM. The product was used for modification without further purification. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (M+K): 1,042.96 g/mol.

Synthesis of compound 11j



11j

Mol. formula: $C_{66}H_{124}N_3O_{13}P$

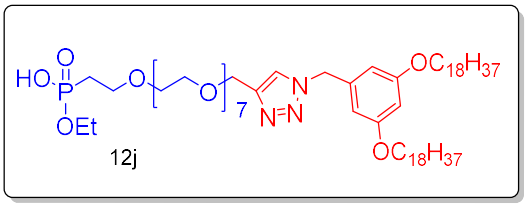
Mol. weight: 1197.88 g/mol

Physical appearance: brownish liquid

Yield: 68 %

The **compound 11j** was synthesized from **compound 4c** (alkyne) (100 mg, 189 mmol), **compound 10f** (142 mg, 212 mmol), $CuSO_4$ (2.0 mg, 0.008 mmol) Na Asc. (3.37 mg, 0.00189 mmol) in 50 % THF / H_2O. The crude product obtained was purified using reverse phase column chromatography using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid (159 mg, 135 mmol, 68 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.56 (s, 1H), 6.42-6.39 (m, 3H), 5.41 (s 2H), 4.67 (s, 2H), 4.15-4.05 (m, 4H), 3.90 (t, J = 6.4 Hz, 4H), 3.75-3.57 (m, 33H), 2.18-2.08 (m, 2H), 1.77-1.70 (m, 4H), 1.45-1.27 (m, 81H), 0.894 (t, J = 6.8 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 160.79, 136.49, 106.55, 101.27, 70.65, 70.47, 70.17, 69.64, 68.13, 65.11, 64.70, 61.62, 54.33, 31.91, 29.69, 29.60, 29.56, 29.38, 29.35, 29.18, 27.67, 26.28, 26.12, 22.57, 16.45, 14.11. **MALDI-TOF MS** (M+K): 1,236.84 g/mol.

Synthesis of compound 12j



12j

Mol. formula: $C_{64}H_{120}N_3O_{13}P$

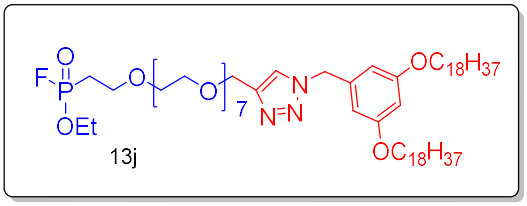
Mol. weight: 1169.85 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12j** was synthesized from **compound 11j** (100 mg, 0.0834 mmol), OxCl (53 mg, 0.329 mmol) in DCM. The product obtained was carried to next step without purification. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.50 (s, 1H), 6.36-6.35 (m, 3H), 5.37 (s, 2H), 4.62 (s, 2H), 3.85-3.81 (m, 2H), 3.63-3.40 (m, 36H), 1.89-1.72 (m, 2H), 1.71-1.69 (m, 4H), 1.39-0.88 (m, 63H), 0.83 (t, J = 6.8 Hz, 6H). **MALDI-TOF MS** (M+K): 1,208.89 g/mol.

Synthesis of compound 13j



13j

Mol. formula: $C_{64}H_{119}N_3O_{13}PF$

Mol. weight: 1171.85 g/mol

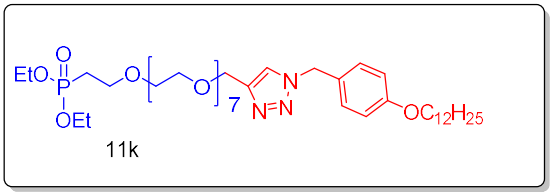
Physical appearance: brownish liquid

Yield: not determined

The **compound 13j** was synthesized from **compound 12j** (70 mg, 0.0597 mmol), DAST (22 mg, 0.136 mmol) in DCM. The product was used for modification without further purification. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (M+K): 1,210.73 g/mol.

Synthesis of 1-tail variant AABPs and their intermediates

Synthesis of compound 11k



11k

Mol. formula: $C_{42}H_{76}N_3O_{12}P$

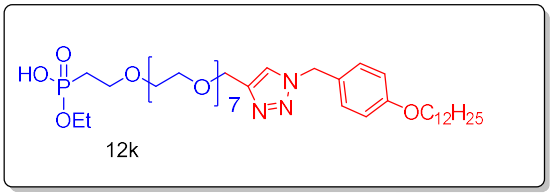
Mol. weight: 845.51 g/mol

Physical appearance: brownish liquid

Yield: 65 %

The **compound 11k** was synthesized from **compound 4c** (alkyne) (200 mg, 0.37 mmol), **compound 10g** (120 mg, 0.37 mmol), $CuSO_4$ (4.71 mg, 0.0189 mmol), Na Asc. (7.5 mg, 0.0378 mmol) in 50 % THF / H_2O. The product obtained was purified using RPC using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid (210 mg, 0.24 mmol, 65 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.45 (s, 1H), 7.19 (d, J = 8.8 Hz, 2H), 6.86 (d, J = 8.8 Hz, 2H), 5.47 (s, 2H), 4.63 (s, 2H), 4.19-4.00 (m, 4H), 3.93 (t, J = 6.4 Hz, 2H), 3.74-3.57 (m, 31H), 2.18-2.03 (m, 2H), 1.81-1.71 (m, 2H), 1.47-1.19 (m, 27H), 0.87 (t, J = 6.8 Hz, 3H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 159.38, 129.57, 126.26, 114.87, 77.16, 70.45, 70.43, 70.35, 70.08, 69.60, 67.99, 65.01, 64.59, 61.54, 53.59, 31.80, 29.54, 29.52, 29.48, 29.46, 29.27, 29.23, 29.09, 27.57, 25.92, 22.57, 14.03. **MALDI-TOF MS** (M+K): 884.53 g/mol.

Synthesis of compound 12k



12k

Mol. formula: $C_{40}H_{72}N_3O_{12}P$

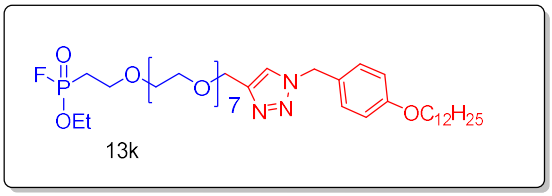
Mol. weight: 817.48 g/mol

Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12k** was synthesized from **compound 11k** (120 mg, 0.14 mmol), OxCl (94 mg, 0.56 mmol) in DCM. The product obtained was carried to next step without purification. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.30 (s, 1H), 7.15 (d, J = 8.8 Hz, 2H), 6.81 (d, J = 8.8 Hz, 2H), 5.38 (s, 2H), 4.59 (s, 2H), 4.28-4.14 (m, 4H), 3.88 (t, J = 6.4 Hz, 2H), 3.76-3.52 (m, 31H), 2.25-2.17 (m, 2H), 1.78-1.63 (m, 2H), 1.43-1.16 (m, 27H), 0.82 (t, J = 6.8 Hz, 3H).

Synthesis of compound 13k



13k

Mol. formula: $C_{40}H_{71}N_3O_{11}PF$

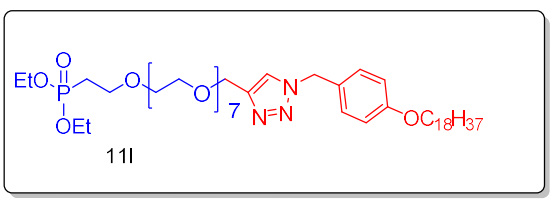
Mol. weight: 819.48 g/mol

Physical appearance: brownish liquid

Yield: not determined

The **compound 13k** was synthesized from **compound 12k** (100 mg, 0.12 mmol), DAST (99 mg, 0.59 mmol) in DCM. The product obtained was used for modification without purification. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (M+K): 858.50 g/mol.

Synthesis of compound 11l



11l

Mol. formula: $C_{48}H_{88}N_3O_{12}P$

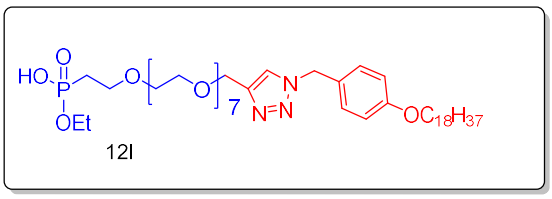
Mol. weight: 929.61 g/mol

Physical appearance: brownish liquid

Yield: 76 %

The **compound 11l** was synthesized from **compound 4c** (alkyne) (100 mg, 0.19 mmol), **compound 10h** (70 mg, 0.19 mmol), $CuSO_4$ (2.3 mg, 0.009 mmol), Na Asc. (3.7 mg, 0.0162 mmol) in 50:25:25 in THF / H_2O / DCM. The crude product obtained was purified using RPC using ACN / H_2O followed by $CHCl_3$ and NPC using MeOH / DCM to get brownish liquid (130 mg, 0.14 mmol, 76 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.45 (s, 1H), 7.38 (d, J = 8.8 Hz, 2H), 6.85 (d, J = 8.8 Hz, 2H), 5.42 (s, 2H), 4.63 (s, 2H), 4.12-3.99 (m, 4H), 3.92 (t, J = 6.4 Hz, 2H), 3.72-3.57 (m, 31H), 2.16-2.03 (m, 2H), 1.83-1.69 (m, 2H), 1.46-1.16 (m, 38H), 0.86 (t, J = 6.8 Hz, 3H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C. 159.58, 145.51, 129.80, 126.36, 122.47, 115.09, 77.16, 70.61, 70.58, 70.55, 70.51, 70.47, 70.12, 69.81, 68.23, 65.27, 64.73, 61.26, 53.86, 53.55, 32.02, 29.79, 29.76, 29.71, 29.68, 29.50, 29.46, 29.30, 28.05, 26.66, 26.13, 22.79, 14.23. **MALDI-TOF MS** (M+K): 968.68 g/mol.

Synthesis of compound 12l



12l

Mol. formula: $C_{46}H_{84}N_3O_{12}P$

Mol. weight: 901.57 g/mol

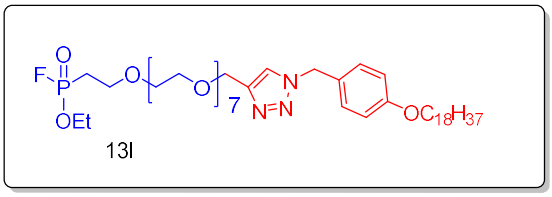
Physical appearance: pale yellow liquid

Yield: not determined

The **compound 12l** was synthesized from **compound 11l** (100 mg, 0.13 mmol), OxCl (70 mg, 0.43 mmol) in DCM. The product obtained was carried to next step without purification.

^{1}H NMR (400 MHz, $CDCl_3$): δ_H 7.46 (s, 1H), 7.19 (d, *J* = 8.8 Hz, 2H), 6.85 (d, *J* = 8.8 Hz, 2H), 5.42 (s, 2H), 4.63 (s, 2H), 4.14-3.97 (m, 2H), 3.92 (t, *J* = 6.4 Hz, 2H), 3.77-3.51 (m, 30H), 2.16-2.03 (m, 2H), 1.77-1.68 (m, 2H), 1.46-1.16 (m, 35H), 0.86 (t, *J* = 6.8 Hz, 3H). **MALDI-TOF MS** (M+Na): 924.64 g/mol.

Synthesis of compound 13l



13l

Mol. formula: $C_{46}H_{83}N_3O_{11}PF$

Mol. weight: 903.57 g/mol

Physical appearance: brownish liquid

Yield: not determined

The **compound 13l** was synthesized from **compound 12l** (80 mg, 0.08 mmol), DAST (60 mg, 0.39 mmol) in DCM. The product obtained was used for modification without purification. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (M+Na): 926.24 g/mol.

References

1 Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **2016**, 537, 320-327.

2 Yeates, T. O. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu. Rev. Biophys.* **2017**, 46, 23-42.

3 Salgado, E. N., Radford, R. J. & Tezcan, F. A. Metal-directed protein self-assembly. *Acc. Chem. Res.* **2010**, 43, 661-672.

4 Reynhout, I. C., Cornelissen, J. J. & Nolte, R. J. Synthesis of polymer− biohybrids: from small to giant surfactants. *Acc. Chem. Res.* **2009**, 42, 681-692.

5 Luo, Q., Hou, C., Bai, Y., Wang, R. & Liu, J. Protein Assembly: Versatile Approaches to Construct Highly Ordered Nanostructures. *Chem. Rev.* **2016**, 116, 13571-13632.

6 King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **2014**, 510, 103-108.

7 Wörsdörfer, B., Woycechowsky, K. J. & Hilvert, D. Directed evolution of a protein container. *Science* **2011**, 331, 589-592.

8 Lai, Y.-T., Tsai, K.-L., Sawaya, M. R., Asturias, F. J. & Yeates, T. O. Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. *J. Am. Chem. Soc.* **2013**, 135, 7738-7743.

9 Velonia, K., Rowan, A. E. & Nolte, R. J. Lipase polystyrene giant amphiphiles. *J. Am. Chem. Soc.* **2002**, 124, 4224-4225.

10 Heredia, K. L. *et al.* In situ preparation of protein−"smart" polymer conjugates with retention of bioactivity. *J. Am. Chem. Soc.* **2005,** 127, 16955-16960.

11 Biswas, S. *et al.* A tubular biocontainer: metal ion-induced 1D assembly of a molecularly engineered chaperonin. *J. Am. Chem. Soc.* **2009**, 131, 7556-7557.

12 Sim, S., Niwa, T., Taguchi, H. & Aida, T. Supramolecular Nanotube of Chaperonin GroEL: Length Control for Cellular Uptake Using Single-Ring GroEL Mutant as End-Capper. *J. Am. Chem. Soc.* **2016**, 138, 11152-11155.

13 Ren, Q., Kwan, A. H. & Sunde, M. Two forms and two faces, multiple states and multiple uses: Properties and applications of the self-assembling fungal hydrophobins. *Biopolymers* **2013**, 100, 601-612.

14 Hobley, L. *et al.* BslA is a self-assembling bacterial hydrophobin that coats the Bacillus subtilis biofilm. *Proc. Natl Acad. Sci. USA* **2013**, 110, 13600-13605.

15 Theil, E. C. Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. *Ann. Rev. Biochem.* **1987**, 56, 289-315.

16 Kobayashi, K. & Iwano, M. BslA (YuaB) forms a hydrophobic layer on the surface of Bacillus subtilis biofilms. *Mol. Microbiol.* **2012**, 85, 51-66.

17 Dykes, G. W., Crepeau, R. H. & Edelstein, S. J. Three-dimensional reconstruction of the 14-filament fibers of hemoglobin S. *J. Mol. Biol.* **1979**, 130, 451-472.

18 Garcia-Seisdedos, H., Empereur-Mot, C., Elad, N. & Levy, E. D. Proteins evolve on the edge of supramolecular self-assembly. *Nature* **2017**,

19 Israelachvili, J. N. *Intermolecular and surface forces*. San Diego: Academic press, 2015.

20 Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **2002**, 102, 4501-4524.

21 Long, J. Z. & Cravatt, B. F. The metabolic serine hydrolases and their functions in mammalian physiology and disease. *Chem. Rev.* **2011**, 111, 6022-6063.

22 Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise huisgen cycloaddition process: copper (I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Edn. Engl.* **2002**, 114, 2708-2711.

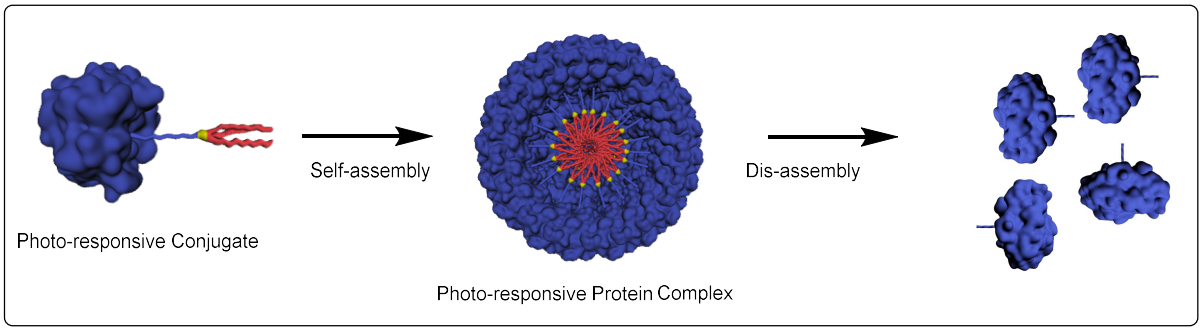
23 Liu, Y., Patricelli, M. P. & Cravatt, B. F. Activity-based protein profiling: the serine hydrolases. *Proc. Natl Acad. Sci. USA* **1999**, 96, 14694-14699.

24 Sandanaraj, B. S., Reddy, M. M., Bhandari, P. J., Kumar, S. & Aswal, V. K. Rational design of supramolecular dynamic protein assemblies by using a micelle-assisted activity-based protein-labeling technology. *Chem. - Eur. J.* **2018,** 24, 16085-16096.

Chapter 3

FAPs for the Accurate Design of Stimuli-responsive Protein Complexes

3.1. Synthesis, Self-assembly and Dis-assembly Studies of Photo-responsive FAP



3.1.1. Design

This sub-chapter deals with a photo-responsive protein complex. Natural protein complexes evolved to perform advanced functions *in vivo* through several reversible/irreversible post-translational modifications, thus converting the static structure into a dynamic one and vice versa.[1] Genetic methods often yield high-resolution structures but fail when it comes to the design of stimuli-responsive protein complexes. To construct stimuli-responsive protein complexes with dynamic/adaptive behavior, we need more building blocks/chemical entities. So, genetic methods should take advantage of the incorporation of Unnatural amino acids (demonstrated long back[2]). Very few reports incorporated unnatural amino acids for *de* novo protein complex design.[3]

On the other hand, protein engineering using chemical strategy provides an opportunity for the design of stimuli-responsive protein complexes because of access to innumerable building blocks,[4,5], but there are few reports in this direction[5-9]. Here, in this chapter, we report the rational design of monodisperse protein complexes with stimuli-responsive function/adaptive behavior. The other reason why we are interested in creating these stimuli-responsive systems is our own curiosity, i.e., So far, we have achieved the complex formation from monomeric proteins by adding hydrophobicity. If this is true, if we selectively clip off this hydrophobic tail, we should get back to monomeric proteins (Fig. 3.1).

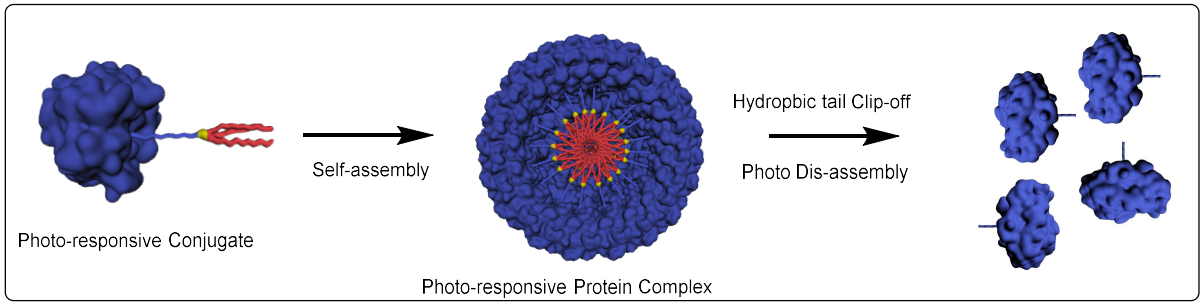


Figure 3.1. Cartoon representation of photo-responsive complex assembly and dis-assembly.

To investigate this, we have created sophisticated photo-responsive FAP in this sub-chapter. For creating photo-responsive complex, we made a similar conjugate using Try, but between the spacer and hydrophobic tail, we have introduced a photo-responsive 2-nitrobenzyl (2-NB) group (Fig. 3.2).

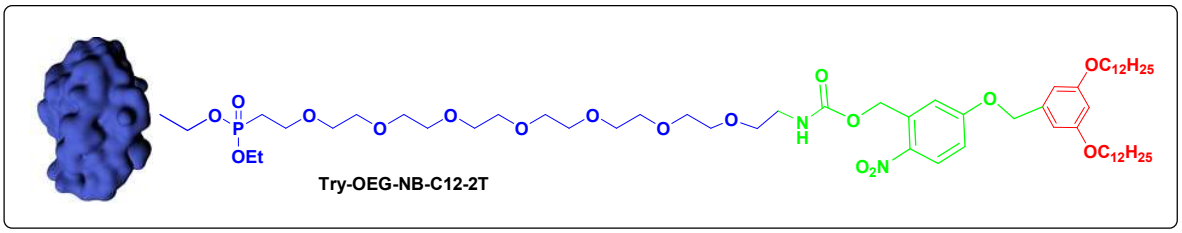
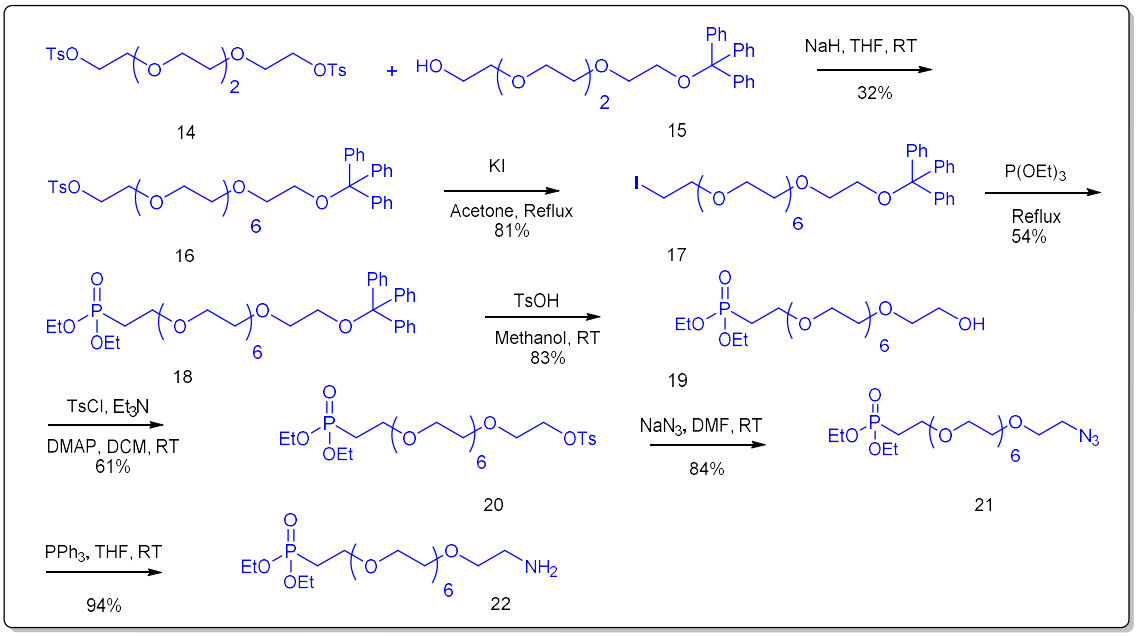


Figure 3.2. Structure of photo-responsive protein conjugate.

3.1.2. Synthesis

Synthesis of the amine-terminated spacer

To install a photo-sensitive 2-nitrobenzyl group between the spacer and tail, we chose octaethylene glycol as we know this spacer is optimum for making FAPs from chapter 2. Using similar synthesis, we made an amine-terminated spacer rather than alkyne terminated one used in the previous chapter (Scheme 3.1).

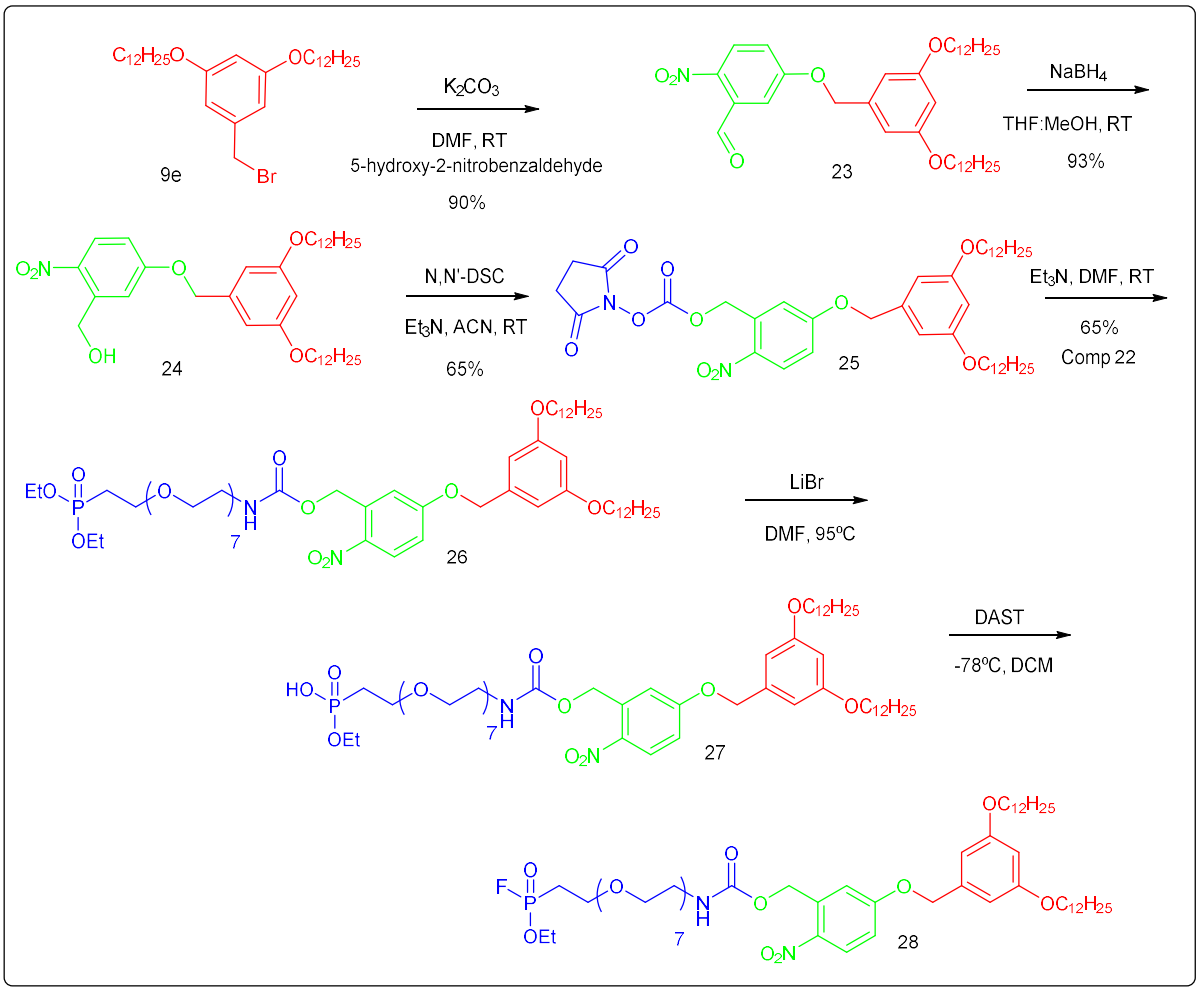


Scheme 3.1. Synthesis of amine-terminated OEG spacer.

In an oven-dried RBF, a mixture of ditosyl TEG (**14**) and monotrityl TEG (**15**) was taken and dissolved in THF under stirring at 0 ºC, and NaH was added in small portions. The resultant mixture was stirred at RT for 12 hrs. Upon completion of the reaction, excess of NaH was quenched by the dropwise addition of water. The resultant mixture was concentrated under reduced pressure. To the obtained residue, water was added and extracted thrice with DCM.

Then, the combined organic layer was dried over Na_2SO_4 and concentrated under reduced pressure to get the crude product, which was purified by NPC using EtOAc / PE as eluent to get brownish liquid (**3c**). A mixture of above-obtained tosylate (**16**) and KI was refluxed in acetone for 18 hrs. Upon completion, excess KI was filtered and washed thrice with acetone. Collected acetone fraction was evaporated under vacuum to get residue, which was then washed with water and extracted with DCM. The combined organic layer was washed with aqueous Na_2CO_3 and then concentrated under vacuum to get the crude product, which was purified by NPC using MeOH / DCM as eluent. In an oven-dried RBF, iodide (**17**) and $P(OEt)_3$ were taken and refluxed for 1 h. Upon completion of the reaction, the excess $P(OEt)_3$ was removed under vacuum, and the crude reaction mixture was directly loaded onto a silica gel column and was purified using MeOH / DCM as eluent. In an oven-dried RBF, a mixture of diphosphonate ester with monotrityl terminus (**18**) and TsOH was dissolved in MeOH and stirred for 12 hrs at RT. Upon completion of the reaction, the solvent was evaporated under reduced pressure. Then, the obtained residue purified using NPC using MeOH / DCM as eluent. In an oven-dried RBF, above obtained alcohol (**19**), TsCl and DMAP were taken and were dissolved in DCM under 0 ºC. Then, Et_3N was added to the above mixture and allowed to stir for 12 hrs at RT. Upon completion, the reaction was quenched with water and extracted with DCM for thrice. The combined organic layer was dried over Na_2SO_4 and concentrated under reduced pressure to get the crude product, which was purified using NPC using MeOH / DCM as eluent. In an oven-dried RBF, above obtained tosylate (**20**) and NaN_3 was taken and dissolved in DMF under stirring. The resultant mixture was allowed to react for 12 hrs. Upon completion of the reaction, DMF was evaporated under reduced pressure. Then, the obtained residue was purified using NPC using MeOH / DCM as eluent. In an oven-dried RBF, azide (**21**) was dissolved in THF under stirring. Then, PPh_3 was added to the above mixture at 0 ºC and allowed to react for 18 hrs at RT. Upon completion of the reaction, a few drops of water were added and again stirred for 1 hr. Then, THF was evaporated under reduced pressure, and the obtained water layer was washed with toluene thoroughly. The water layer was then concentrated under reduced pressure to get the product (**22**), which was used without further purification (Scheme 3.1).

Synthesis of AABP

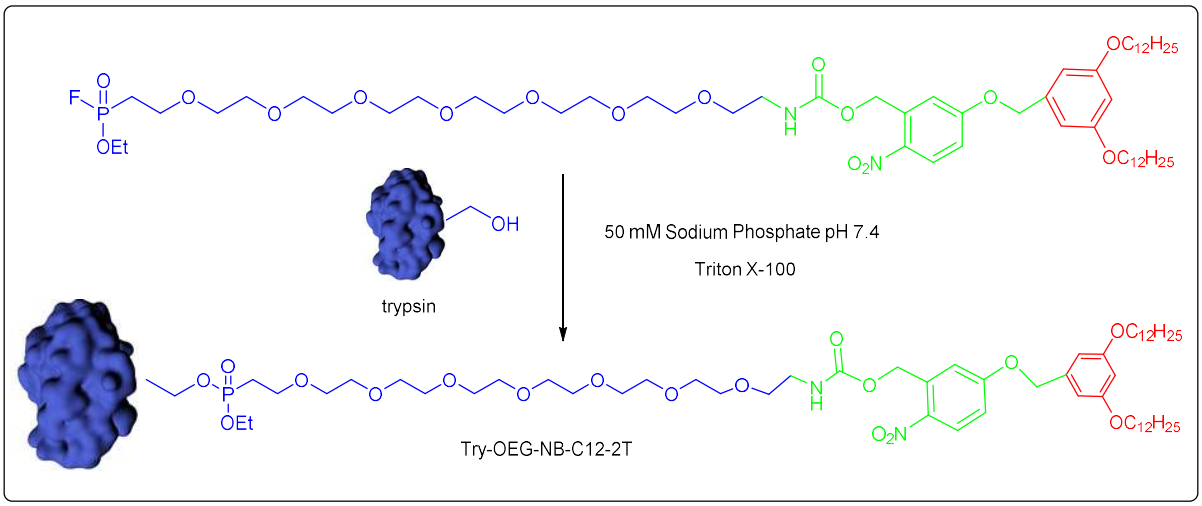


Scheme 3.2. Synthesis of photo-responsive AABP.

In an oven-dried RBF, 5-hydroxy-2-nitrobenzaldehyde (1.1 eq), K_2CO_3 (1.1 eq), and bromide (**9e**) (1 eq) (synthesis explained in chapter 2) was dissolved in DMF under stirring at RT and allowed to react for 12 hrs. Upon completion of the reaction, water was added and extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and evaporated under vacuum to get crude product which was then purified using NPC using EtOAc / PE as eluent. In an oven-dried RBF, the above-obtained aldehyde (**23**) (1 eq) was dissolved in THF : MeOH (2:1) under stirring at 0 ºC and Sodium Borohydride ($NaBH_4$) (1.5 eq) was added in small portions and allowed to react for 1 hr. Upon completion, the reaction was quenched with Na_2CO_3 and extracted thrice with EtOAc. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified using NPC using EtOAc / PE as eluent. In an oven-dried RBF, above obtained alcohol (**24**) (1 eq)

and *N,N'*-DSC (5 eq) were dissolved in ACN under stirring. Then, Et_3N (5 eq) was then added slowly at RT and allowed to react for 12 hrs. Upon completion of the reaction, ACN and Et_3N were evaporated under vacuum. The obtained residue was directly purified using NPC using EtOAc / PE as eluent. In an oven-dried RBF, the above obtained activated ester (**25**) (1.1 eq) and previously made spacer (**3i**) (1 eq) were dissolved in DMF under stirring. Then, Et_3N (1.1 eq) was added slowly to the reaction mixture and stirred at RT for 12 hrs. Upon completion of the reaction, DMF and Et_3N were evaporated under vacuum. To the obtained residue, water was added and extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was purified by NPC using MeOH / DCM as eluent. In an oven-dried RBF, the above-obtained diphosphonate ester (**26**) (1 eq) was taken, and LiBr (20 eq) was added. To the mixture, DMF was added and heated at 95 ºC for 20 hrs. Upon completion, water was added and extracted thrice with EtOAc. The water layer was collected, and 2N HCl was added and stirred for another 30 mins. The mixture was then extracted thrice with EtOAc. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get crude product which was used without purification. OxCl was used for this step in the previous chapter. To the stirring solution of the above-obtained monophosphate ester (**27**) (1 eq) in DCM, DAST (4 eq) was added dropwise at -78 ºC and allowed to react for 15 mins. Upon completion of the reaction, excess of DAST and DCM was evaporated under vacuum. To the obtained residue, water was added and stirred for 2 more mins to quench any residual DAST. The reaction mixture was then extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product (**28**), which was used without purification (Scheme 3.2).

3.1.3. Conjugation and purification



Scheme 3.3. Synthesis of photo-responsive FAP (Try-OEG-NB-C12-2T).

After the synthesis and characterization of photo-responsive AABP, we conjugated the probe to Try using MAPLab (Scheme 3.3). The protocol for both the test scale and scale-up conjugation is mentioned in the methods section of this chapter. To monitor the extent of protein modification, the samples were directly withdrawn from the reaction mixture and analyzed, as mentioned in the methods section. The representative MALDI-TOF MS spectrum of the reaction mixture is given below (Fig. 3.3).

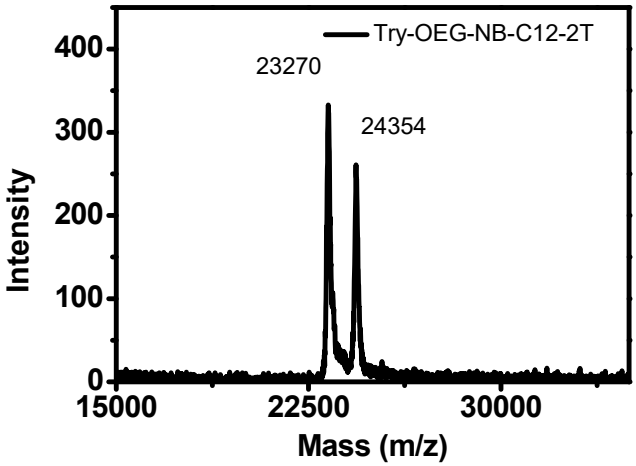


Figure 3.3. MALDI-TOF MS analysis of reaction mixture.

The conjugate was purified by three-step purification; as explained earlier, CEX was performed to remove triton X-100 using SP sepharose at pH 7.4. The obtained IEX fractions

were subjected to SEC immediately to remove the native proteins from the protein conjugates in 50 mM Sod. Phos. pH 7.4, 1 M NaCl. The NaCl was later removed by the Sephadex-G25 desalting column. The desalted fractions were quickly lyophilized and later dissolved in the required buffer when needed. The detailed procedure for purification is given in the experimental section.

3.1.4. Characterization

After conjugation and during purification, the purity of IEX, SEC, and desalted fractions were analyzed by MALDI-TOF MS using the same procedure mentioned for native proteins and reaction mixtures except for the addition of triton X-100 in protein fractions. The detailed procedure is given in the experimental section. The obtained MW of FAP is the same as expected / calcd one without any ambiguity (Fig. 3.4).

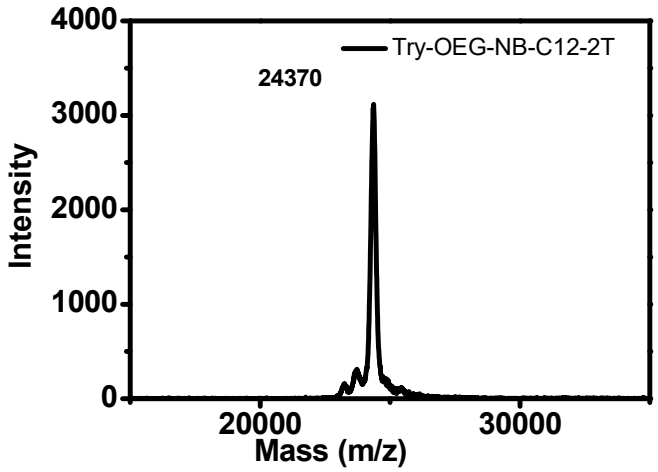


Figure 3.4. MALDI-TOF MS spectrum of purified photo-responsive FAP.

3.1.5. Self-assembly studies

Before dis-assembly, we wanted to study the self-assembly behavior of this conjugate. After the purification, we did SEC, SEC-MALS, and DLS to confirm that the photo-responsive FAP can assemble to form a higher-order complex (Fig. 3.5).

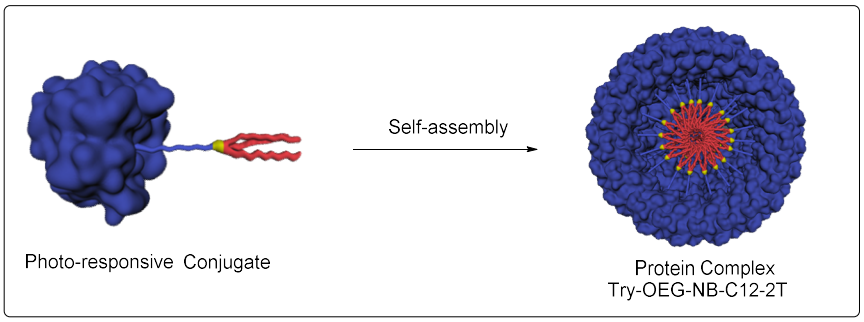


Figure 3.5. Cartoon representation of self-assembly of photo-responsive FAP.

The mean size of the photo-responsive protein complex was measured using DLS, which revealed that the complex formed has D_h of about 10 nm size (Fig. 3.6). To determine the RMW of the soluble protein complex, the FAP was dissolved individually in Milli Q water, and the SEC run was performed. The complex eluted at 11.49 mL, while native protein eluted at 17 mL suggesting the formation of higher-order complex (Fig. 3.7). SEC-MALS analysis of the protein complex was performed on a Superose 6, GL column calibrated with BSA to determine the AMWs. The results showed that this FAP could form a higher-order complex 348 kDa (Fig. 3.8). The elution volume from SEC is in agreement with the DLS studies. We this minimum study, we performed disassembly studies.

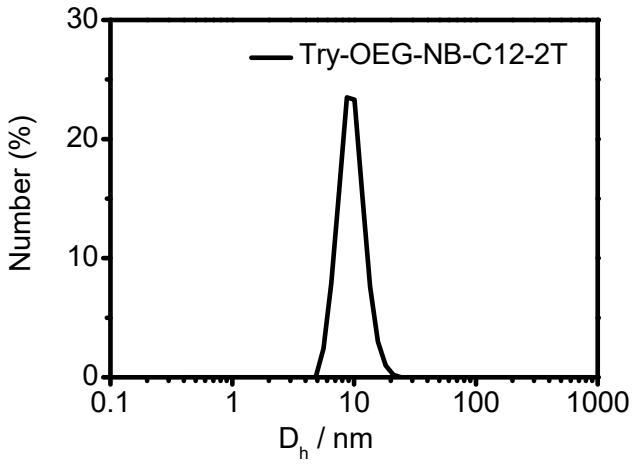


Figure 3.6. DLS data of photo-responsive FAP after purification.

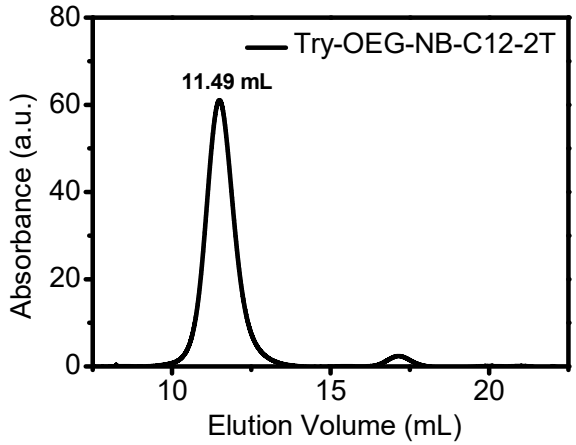


Figure 3.7. SEC data of photo-responsive FAP after purification (Superdex-200).

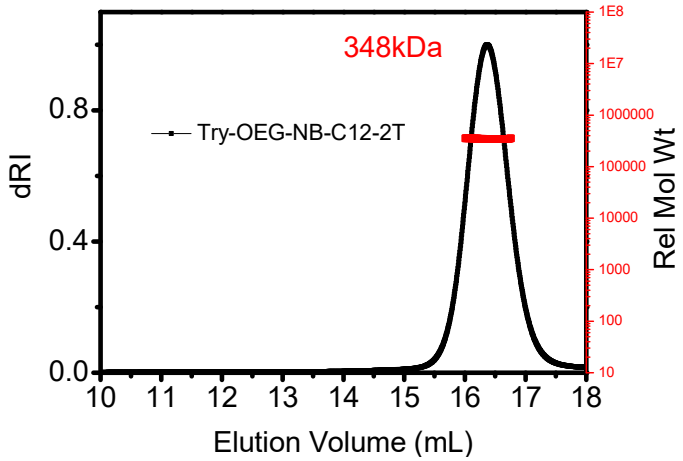


Figure 3.8. SEC-MALS data of photo-responsive FAP after purification (Superose 6).

Table 3.1. Summary of self-assembly data of the photo-responsive FAP.

Protein Conjugate	Elution Vol SEC (mL)	Mol Wt MALS (kDa)	Oligomeric State MALS (mer)	D_h DLS (nm)
Try-OEG-NB-C12-3T	11.4	348	14	11

3.1.6. photo-disassembly studies

The perform photo-disassembly studies, photo-responsive FAP was dissolved in 50 mM Sod. Phos. pH 7.4 in a 50 mL beaker and temperature was maintained at 0 ºC using an ice bath over the entire reaction (Fig. 3.9). The sample was exposed to the Sankyo Denki G8T5E UVB linear lamp (8 W output), which emits ultraviolet rays between 280 to 360 nm (peak at 305-315 nm). Samples were withdrawn at different time points using a pipette, and analysis of the same was carried out using SEC, MALDI-TOF MS, and DLS.

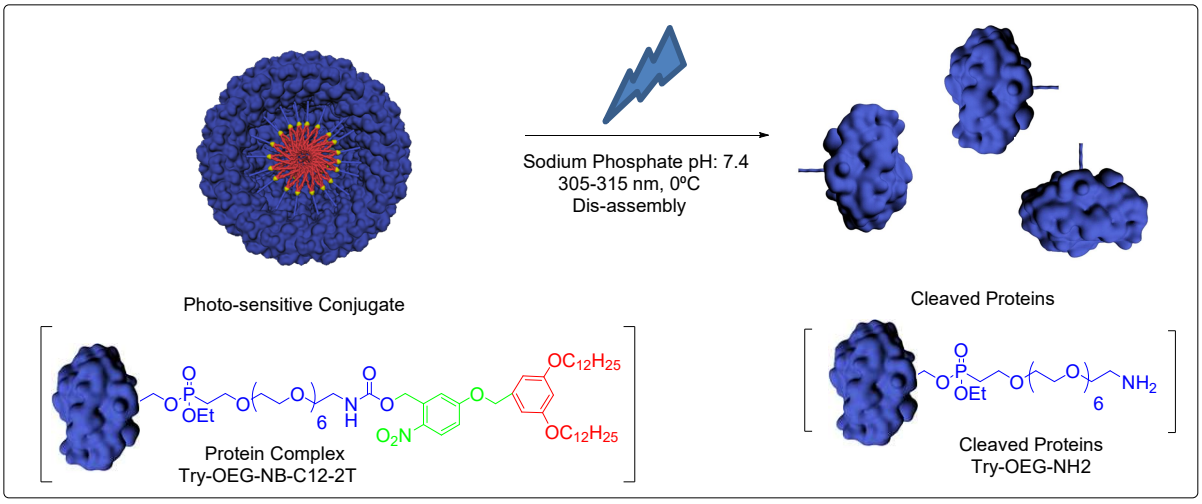


Figure 3.9. Cartoon representation of the dis-assembly process of the photo-responsive protein complex.

Upon photo-irradiation, cleavage of a 2-NB group should lead to a separation of the hydrophobic domain from the rest of the hydrophilic domain, which results in dis-assembly of a protein complex, as the assembly is kept intact by strong hydrophobic interaction. Time-dependent photochemical reactions were carried out on the photo-responsive FAP (50 µM), and the samples were analyzed by analytical SEC. As expected, at time zero, the protein conjugate eluted at 11.5 mL indicative of higher-order complex. However, at intermediate time intervals (2 mins to 45 mins), two peaks were observed, the first peak at 11.5 mL and the second peak at 17.0 mL corresponding to a protein complex and a cleaved, monomeric protein conjugate, respectively. The concentration of protein complex gradually decreased with a concomitant increase in the concentration of a monomeric protein as a function of sample exposure time to UV light. However, the peak at 10.5 mL did not vanish completely even after prolonged irradiation time of 90 mins (Fig. 3.10).

In parallel, detailed MALDI-TOF MS analyses were carried out for reaction mixtures at different time points. At time zero, a single species at 24,362 Da corresponding to the full conjugate Try-OEG-NB-C12-2T was observed. At 2 mins, two species were seen, one major peak for the full conjugate and a very small peak for a cleaved conjugate, Try-OEG-NH_2 at 23,723 Da. With an increase in UV light exposure, an appreciable decrease in peak height for the full conjugate and concomitant increase for a cleaved conjugate was observed. At 90 mins, complete conversion of full conjugate into a cleaved protein was achieved (Fig. 3.11, Fig. 3.12 and Fig. 3.13). DLS data also revealed that there are two species (smaller and larger) upon exposure (90 mins) when compared with unexposed FAP. This is in agreement with SEC, a study that showed a remnant peak even after 90 mins exposure (Fig. 3.14). The photo-responsive complex, upon irradiation at 0 ºC got cleaved and disassembled. We have followed this process using SEC, MALDI-TOF MS, and DLS with respect to time. All most 100 % cleavage happened as per MALDI-TOF MS data, whereas the SEC and DLS showed some remnant higher complex.

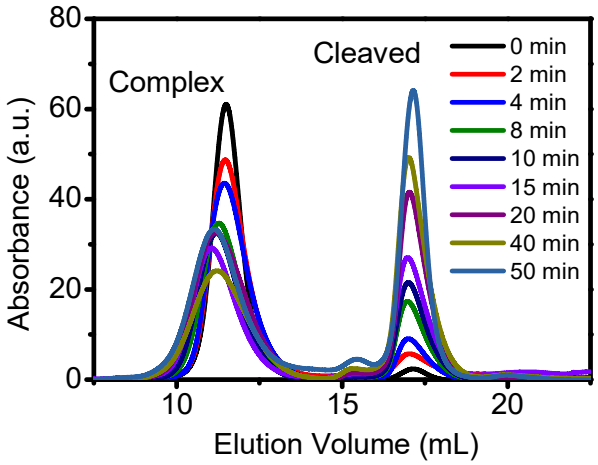


Figure 3.10. Programmed time-dependent irreversible dis-assembly of photo-responsive protein complex followed using SEC (Superdex-200).

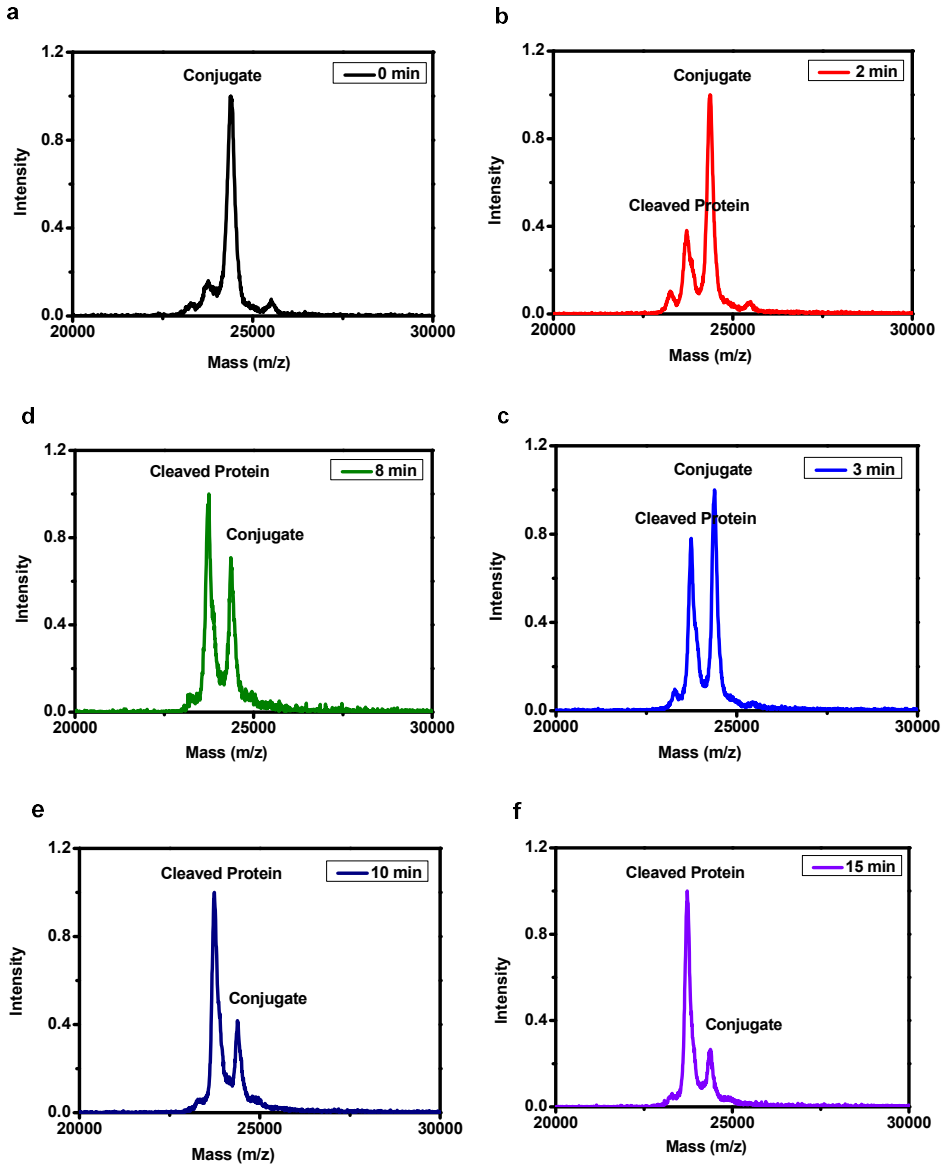


Figure 3.11. Time-dependent MALDI-TOF MS studies of programmed dis-assembly of the photo-responsive protein complex, showing a systematic decrease in the intensity of photo-responsive FAP and an increase in the cleaved conjugate.

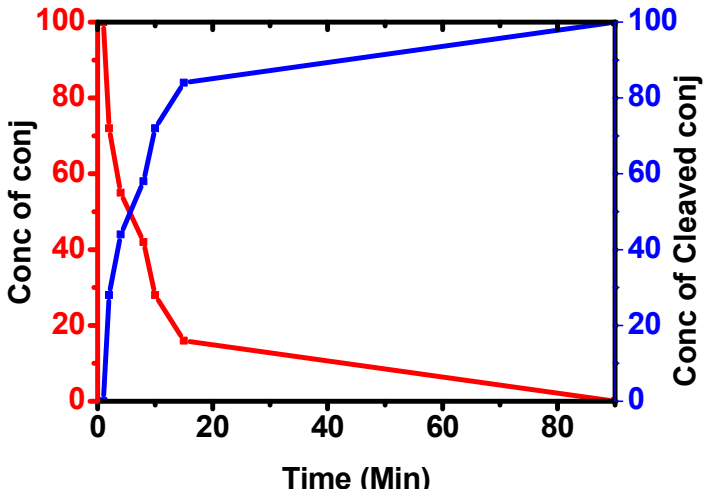


Figure 3.12. Graphic representation of Time-dependent molecular weight changes of the conjugate and the cleaved conjugate as per MALDI-TOF MS studies.

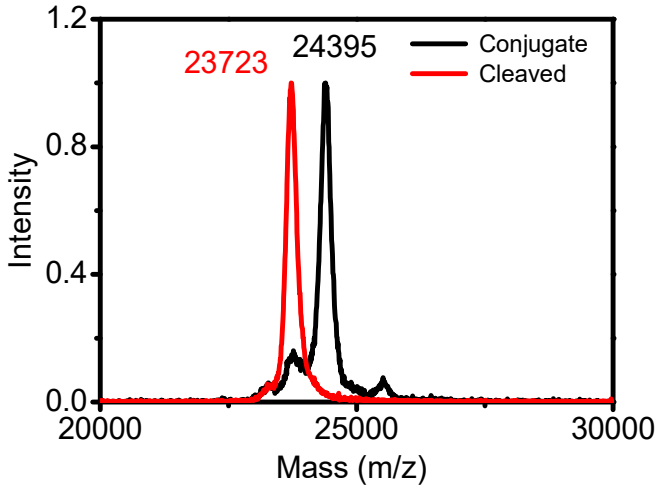


Figure 3.13. MALDI-TOF MS data of protein conjugate before and after 90 mins of UV exposure.

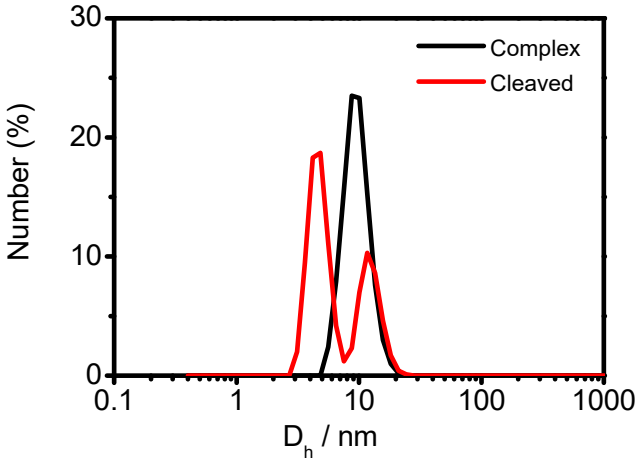


Figure 3.14. DLS data for programmed dis-assembly of photo-responsive protein complex before and after cleavage.

3.1.7. Control studies

The control experiments were performed to prove that the disassembly of the Try-OEG-NB-C12-2T protein complex is only because of selective cleavage of 2-NB in response to UV light.

Control study I

First, we verified the role of UV light by exposing the Try-OEG-NB-C12-2T conjugate to visible light for 90 mins. The SEC profile for the exposed samples showed that the protein conjugate didn't undergo any cleavage, in spite of a 2-NB group. The SEC profile is nearly identical to the unexposed Try-OEG-NB-C12-2T conjugate (Fig. 3.15).

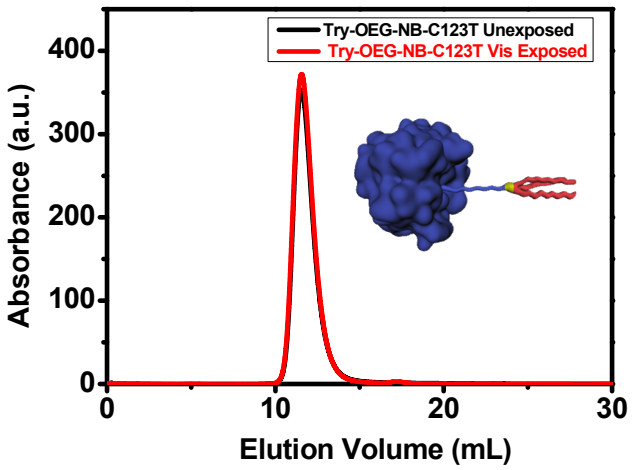


Figure 3.15. SEC data obtained for Photo-responsive Try-OEG-NB-C12-2T in response to the visible light. (i) Black trace: not exposed to visible light (ii) red trace: exposed to light for 90 min (Superdex-200).

Control study II

Similarly, to prove that protein conjugate without the nitro-benzyl group, i.e., Try-OEG-C12-3T is stable upon exposure to UV light, we exposed the conjugate to UV light for 90 mins. The SEC profile for the exposed samples showed that the protein conjugate, in spite of exposure to UV light, for 90 mins, maintained profile, which is nearly the same as the unexposed Try-OEG-C12-2T conjugate (Fig. 3.16).

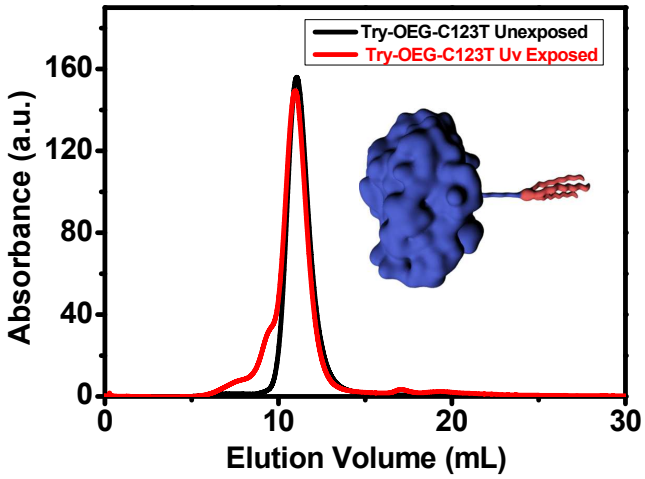


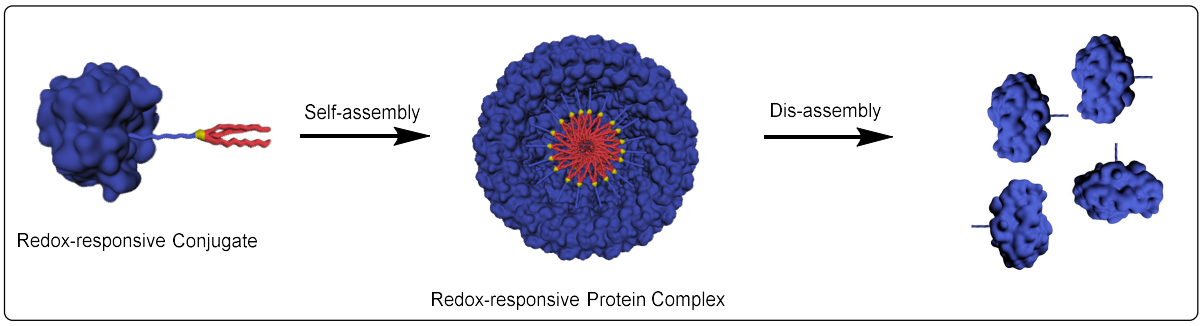
Figure 3.16. SEC data obtained for protein conjugate without 2-NB group, Try-OEG-C12-3T (control) in response to the UV light. (i) Black trace: not exposed to UV light (ii) Red trace: exposed to UV light for 90 min (Superdex-200).

3.1.8. Discussion

We have successfully demonstrated the construction of a sophisticated photo-responsive protein complex. For creating photo-responsive complex, Try was used as a head group, and between the spacer and hydrophobic tail, we have introduced a photo-responsive 2-nitrobenzyl (2-NB) group. 2-NB group dint interfere in conjugation and self-assembly. The photo-responsive protein complex, upon photo exposure, got disassembled because of the cleavage of the 2-NB group. The process of cleavage of conjugate and dis-assembly of the complex was followed through SEC and DLS studies. However, we have observed

incomplete disassembly. When verified using MALDI-TOF MS analysis, we have observed the complete cleavage of the conjugate. Careful observation of SEC profile revealed that the new assembly is not our original protein complex but altogether a new species resulting from the cleaved tail group. We are investigating this right now in our lab, and we will reveal these results soon.

3.2. Synthesis, Self-assembly and Dis-assembly Studies of Redox-responsive FAP



3.2.1. Design

This sub-chapter deals with redox-responsive FAP. After the demonstration of the intelligent complex with photo-sensitive features in our previous subchapter, we thought of making redox-sensitive complex for a couple of reasons. A redox-responsive complex can be used for targeted drug delivery in response to the specific intracellular and extracellular environment of tumor tissues. Disulfide linkage will be cleaved at a high concentration of GSH in the intracellular reducing environment of tumor cells (2-10 mM) while remaining stable in the extracellular environment (2-20 μM).[10] The other reason for making the redox-sensitive complex is to verify the complete cleavage of complex upon redox-stimuli as we have experienced incomplete cleavage with photo-sensitive complex (Fig. 3.17).

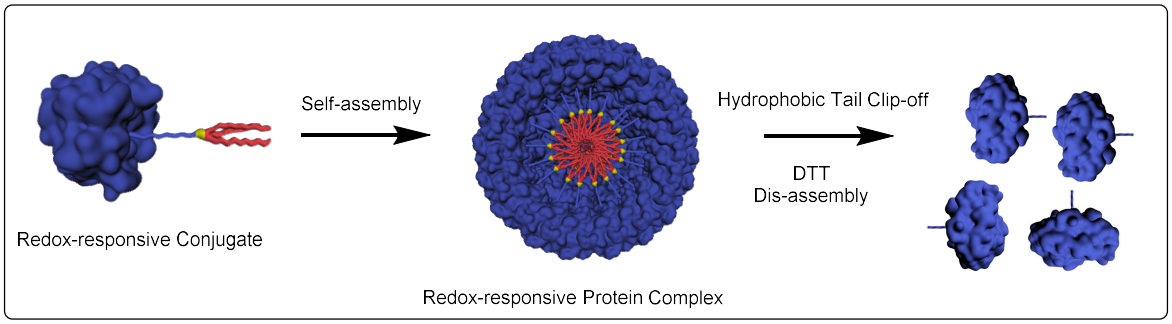


Figure 3.17. Cartoon representation of assembly and disassembly of the redox-responsive conjugate.

We chose the OEG spacer and C12-2T tail, just like the previous chapter. We don't want to invoke too many variables in this subchapter except the stimulus. The synthesis of this spacer was just discussed. For creating a redox-responsive complex, we made a similar conjugate even here, but between the space and hydrophobic tail, we have introduced a disulfide bridge (Fig. 3.17).

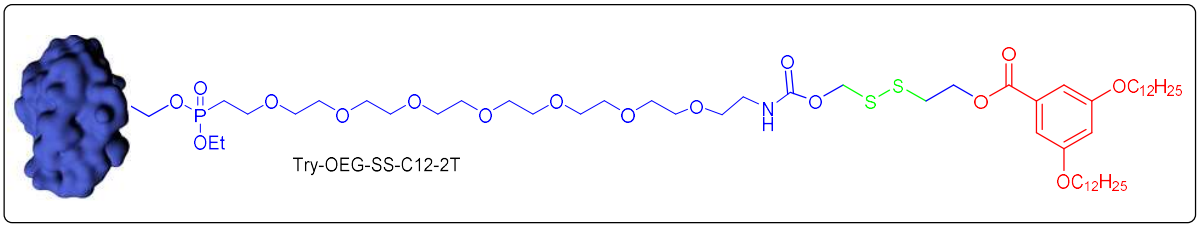
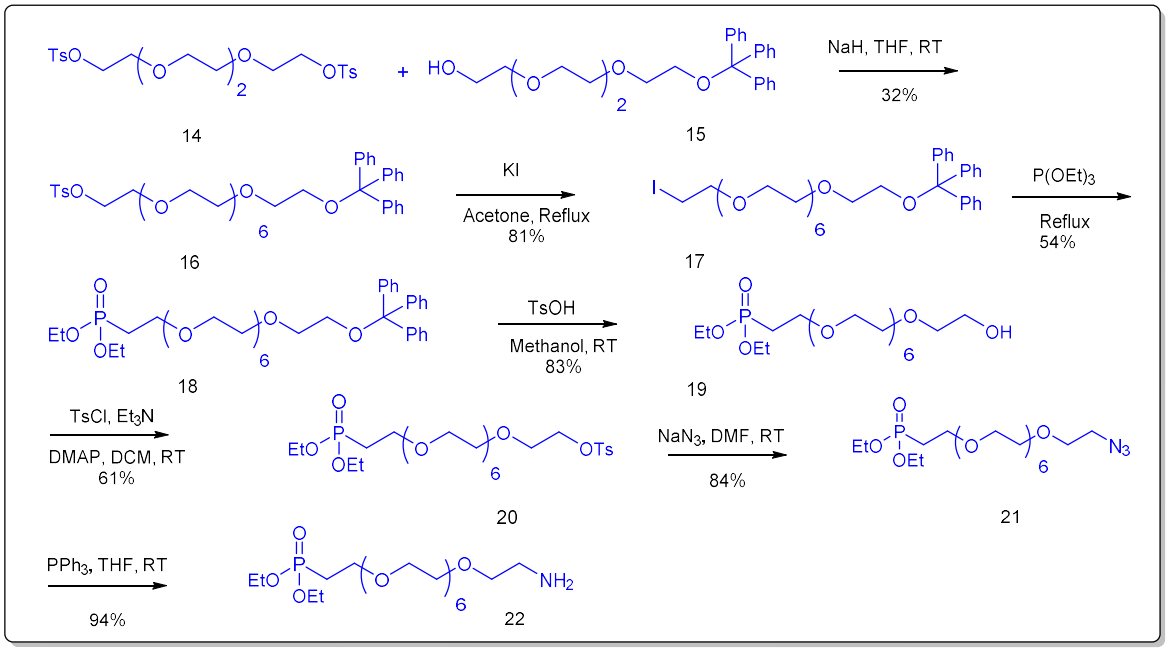


Figure 3.18. Chemical Structure of redox-responsive protein conjugate.

3.2.2. Synthesis

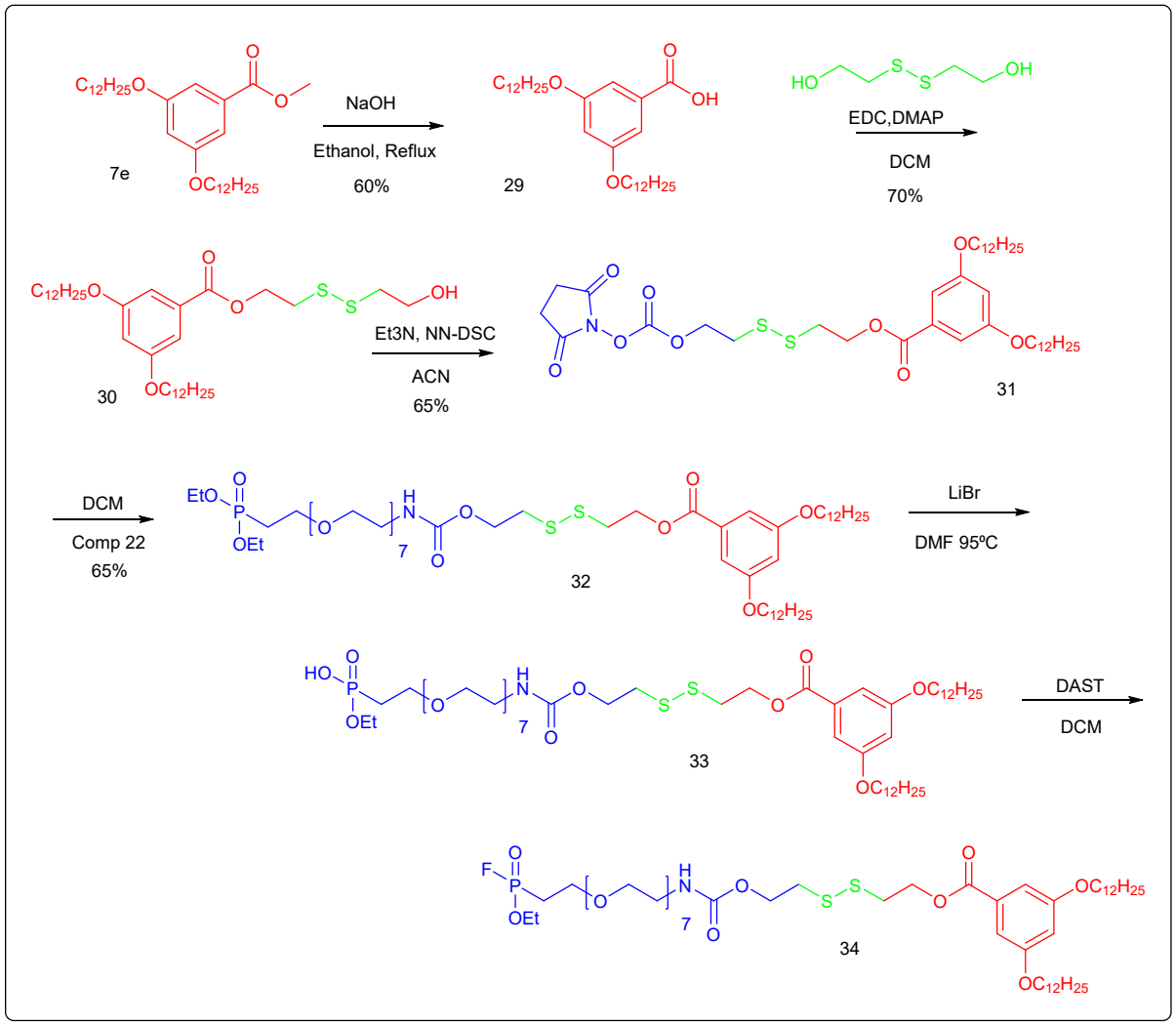
Synthesis of the amine-terminated spacer



Scheme 3.4. Synthesis of amine-terminated OEG spacer.

The synthesis of this is discussed before in chapter 3.1. To brief, the reaction of monotrityl TEG (**15**) and ditosyl TEG (**14**) in the presence of NaH in THF afforded tosylate (**16**). The obtained tosylate (**16**) on reflux with KI in acetone yielded corresponding iodo derivative (**3d**). The iodide (**17**) on reflux with $P(OEt)_3$ provided compound **19**. Then, the treatment of diphosphonate ester (**18**) with TsOH in MeOH yielded corresponding hydroxyl (**19**), which upon tosylation in DCM gave tosylated diphosphonate ester (**20**). This tosylate (**20**) upon treatment with NaN_3 yielded azide (**21**). Finally, treating azide (**21**) with PPh_3 gave amine (**22**) (Scheme 3.4).

Synthesis of AABP



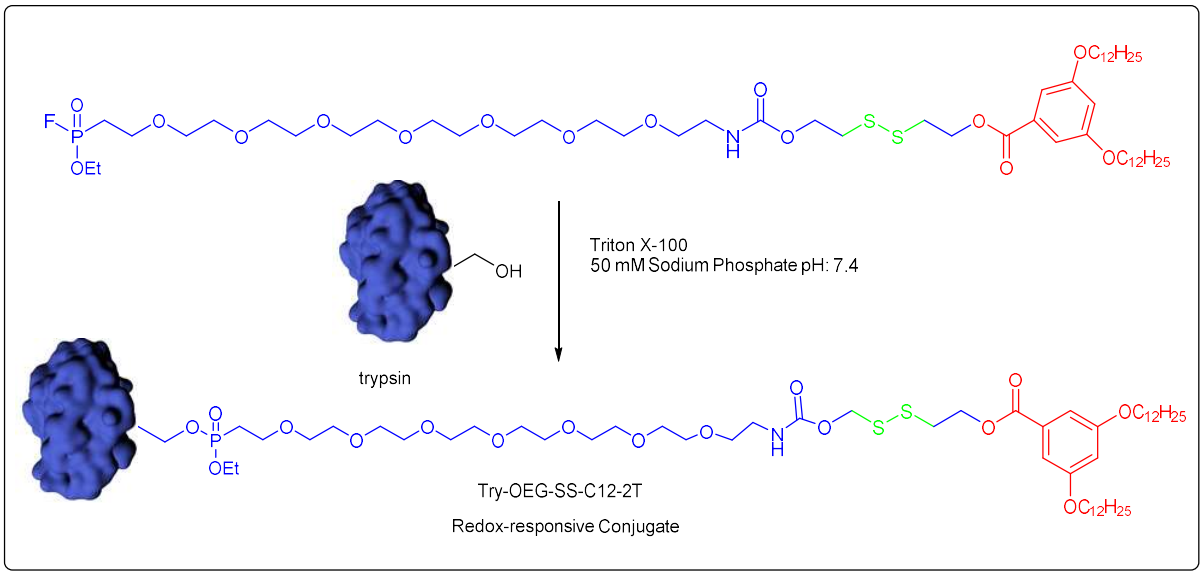
Scheme 3.5. Synthesis of redox-responsive AABP.

The ester and NaOH were dissolved in EtOH in an RBF. The mixture was then refluxed for 4 hrs. Upon completion, the reaction was quenched with dropwise addition of water and acidified with conc. HCl. The obtained precipitate was filtered and washed with EtOH thrice. The combined organic layer dried over Na_2SO_4 and concentrated under reduced pressure to get the crude product, which was utilized for the next reaction without further purification. In an oven-dried RBF above obtained acid (**29**), 2,2'-disulfanediylbis(ethan-1-ol) and Et_3N were taken and dissolved in DCM under stirring. EDC and DMAP were added slowly to the above mixture and allowed to stir for 12 hrs. Upon completion, the reaction was quenched with water, and the resulting content was extracted in DCM thrice. Combined organic layers were dried over Na_2SO_4 and concentrated under reduced pressure to get crude product which was

purified using silica gel column chromatography. In an oven-dried RBF, above alcohol (**30**) (1 eq) and *N,N'*-DSC (5 eq) were dissolved in ACN under stirring. Then, Et_3N (5 eq) was then added slowly at RT and allowed to react for 12 hrs. Upon completion of the reaction, ACN and Et_3N were evaporated under vacuum. The obtained residue was directly purified using NPC using EtOAc / PE as eluent. In an oven-dried RBF, the above obtained activates ester (**31**) (1.1 eq) and amine (**22**) (1 eq) were dissolved in DMF under stirring. Then, Et_3N (1.1 eq) was added slowly to the reaction mixture and stirred at RT for 12 hrs. Upon completion of the reaction, DMF and Et_3N were evaporated under vacuum. To the obtained residue, water was added and extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get crude product which was purified by NPC using MeOH / DCM as eluent

In an oven dried RBF, above obtained diphosphonate ester (**32**) (1 eq) was taken and LiBr (20 eq) was added. To the mixture, DMF was added and heated at 95 ºC for 20 hrs. Upon completion, water was added and extracted thrice with EtOAc. The water layer was collected, and 2 N HCl was added and stirred for another 30 mins. The mixture was then extracted thrice with EtOAc. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get crude product which was used for next step without purification. To the stirring solution of the above obtained monophosphonate ester (**33**) (1 eq) in DCM, DAST (4 eq) was added dropwise at -78 ºC and allowed to react for 15 mins. Upon completion of the reaction, excess of DAST and DCM were evaporated under vacuum. To the obtained residue, water was added and stirred for 2 more mins to quench any residual DAST. The reaction mixture was then extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product (**34**), which was used for conjugation without purification (Scheme 3.5).

3.2.3. Conjugation and purification



Scheme 3.6. Synthesis of redox-responsive FAP (Try-OEG-SS-C12-2T).

After the synthesis and characterization of redox-responsive AABP, we conjugated the probe to Try using MAPLab (Scheme 3.6). Both the test scale and scale-up conjugations were done as per protocol mentioned in the methods section. To monitor the extent of protein modification, the samples were directly withdrawn from the reaction mixture and analyzed. The representative MALDI-TOF MS spectrum of the reaction mixture is given below (Fig. 3.19).

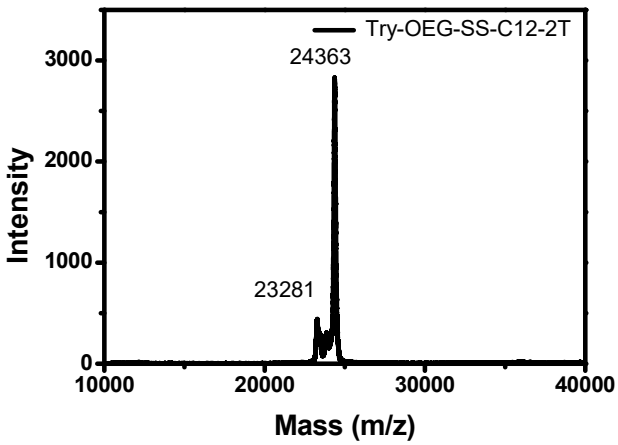


Figure 3.19. MALDI-TOF MS analysis of reaction mixture.

The conjugate was purified by three-step purification, as explained earlier; IEX was performed to remove triton X-100 using SP sepharose, a cation-exchange resin at pH 7.4. The obtained IEX fractions were subjected to SEC immediately to remove the native proteins from the redox-responsive FAP in 50 mM Sod. Phos. pH 7.4, 1 M NaCl. The NaCl was later removed by the Sephadex-G25 desalting column. The desalted fractions were quickly lyophilized and later dissolved in the required buffer when needed. The detailed procedure is given in the experimental section.

3.2.4. Characterization

The purity of IEX, SEC, and desalted fractions were analyzed by MALDI-TOF MS using the same procedure mentioned for native proteins and reaction mixtures except for the addition of triton X-100 in protein fractions. The detailed procedure is given in the experimental section. The Obtained MW is in agreement with the calcd ones (Fig. 3.20).

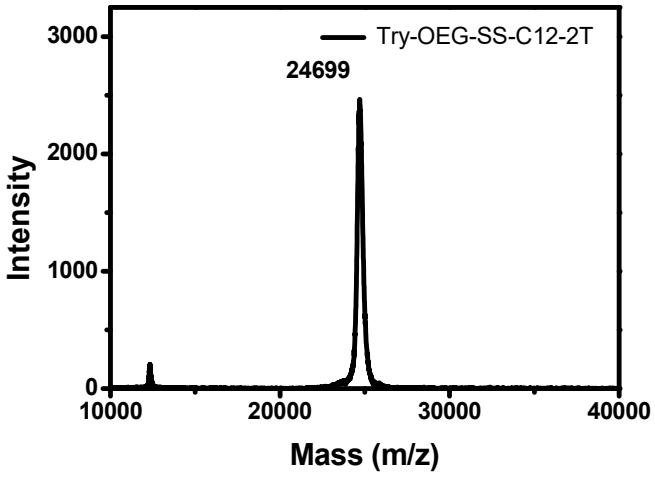


Figure 3.20. MALDI-TOF MS spectrum of the purified conjugate.

3.2.5. Self-assembly studies

After making and purifying the conjugate, we investigated the self-assembly of this FAP (Fig. 3.21). We have characterized this complex using complementary techniques like DLS, SEC, and SEC-MALS.

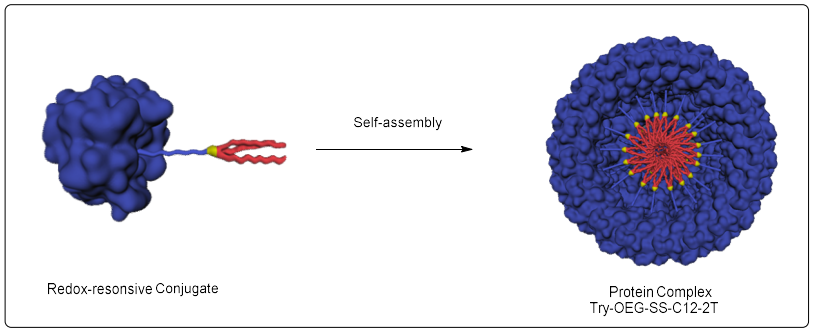


Figure 3.21. Cartoon representation of self-assembly of photo-responsive protein conjugate into a complex.

The mean size of the protein complexes was measured using DLS, which revealed that the complex formed has D_h of about 10 nm in size (Fig. 3.22). To determine the RMWs of the protein complex, this FAP was individually in Milli Q water, and the SEC run was performed. The complex eluted at 16.6 mL, while native protein eluted later, suggesting the formation of higher-order complex (Fig. 3.23). SEC-MALS analyses of the protein complex were performed on a Superose 6 column calibrated with BSA to determine the AMW. This study confirmed that the conjugate could assemble to form a higher-order complex of about 425 kDa. This is surprisingly larger when compared with the sizes of FAPs obtained in chapter 2. Provided the spacer and tail constant, this can be explained on the AABP structure. To include the disulfide bridge in the AABP design, we have eliminated the triazole ring or 2-NB ring from the structure. We believe that this might result in a flexible structure with better packing in the complex state (Fig. 3.24). Thus the data from SEC studies agree with the DLS and SEC-MALS studies.

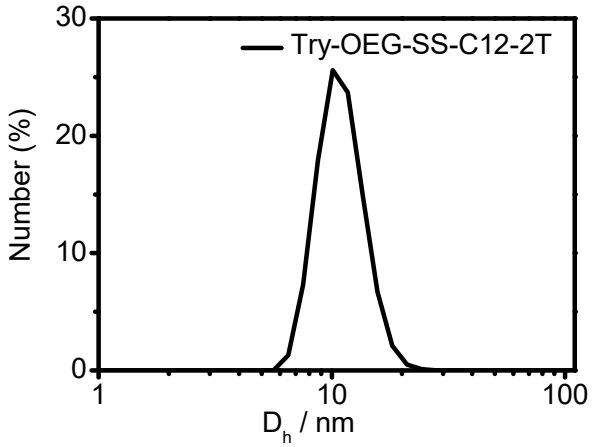


Fig 3.22. DLS data of redox-responsive FAP after purification.

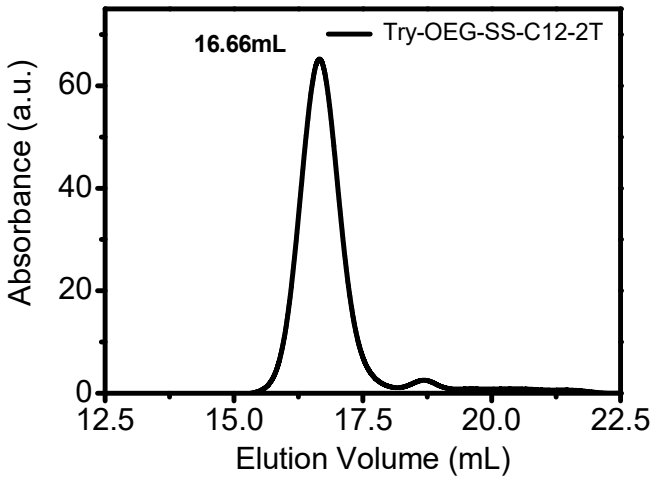


Figure 3.23. SEC data of redox-responsive FAP after purification (Superose 6).

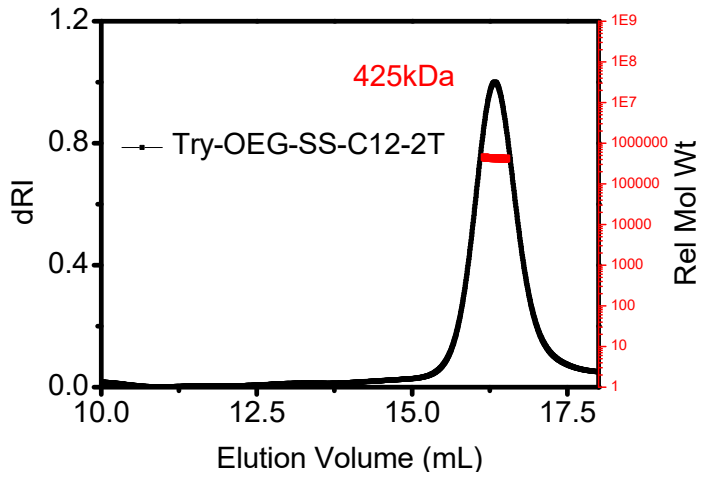


Figure 3.24. SEC-MALS data of redox-responsive FAP after purification (Superose 6).

Table 3.2. Summary of self-assembly data of the redox-responsive protein conjugate.

Protein Conjugate	Elution Vol SEC (mL)	Mol Wt MALS (kDa)	Oligomeric State MALS (mer)	D_h DLS (nm)
Try-OEG-SS-C12-3T	16.66	425	17	10

3.2.6. Redox-disassembly studies

To perform the redox-disassembly studies, the redox-responsive FAP was dissolved in 50 mM Sod. Phos. pH 7.4 in a 2 mL centrifuge tube, and the temperature was maintained at RT over the entire reaction. Different eq of DTT (aq. solution) was added, and then the sample was incubated for varying time points after gentle mixing with a pipette. After that, the sample was analyzed using SEC and MALDI-TOF MS. We did two different studies here: equivalent and time-based disassembly.

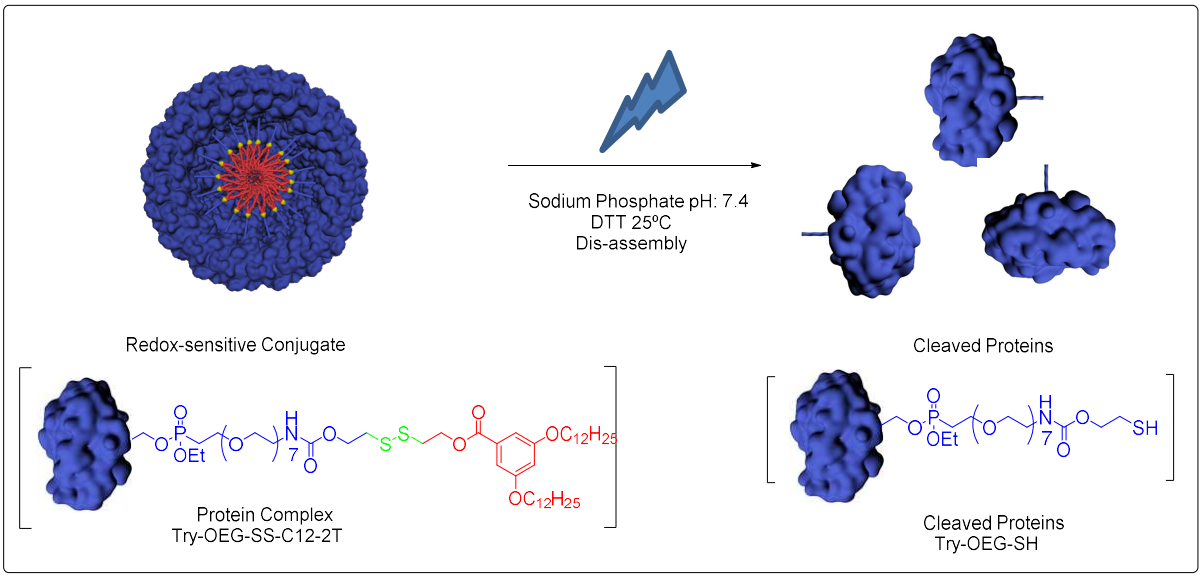


Figure 3.25. Cartoon representation of the disassembly process of the redox-responsive protein complex.

This complex, upon treatment with DTT for 1 hr, got cleaved and disassembled to different extents with respect to equivalents of DTT. We found that just 10 eq of DTT can cleave the conjugate to almost 90 %, but 30 eq is needed for complete cleavage. We have followed this process using SEC (Fig. 3.26). The above study revealed that 30 eq of DTT is sufficient to dis-assemble the complex completely in 1 hr. Similarly, we did time-dependent study. The complex was incubated with 30 eq of DTT for different time periods. The SEC studies revealed that redox-responsive FAP dis-assembles to 90% within 15 mins, while the complete cleavage is observed only at 60 mins. This study concluded that upon treatment with 30 equivalents of DTT for 15 mins, the complex can be cleaved to a significant extent but disassembled completely in 1 hr (Fig. 3.27). All most 100 % cleavage happened when we monitored this in MALDI-TOF MS, unlike the previous photo responsive complex. The

redox-responsive FAP was treated with 30 eq of DTT for 1 hr and analysis as carried out. The obtained molecular weight corresponds to the cleaved conjugate (Fig. 3.28). All the studies concluded that the 100 % cleavage is happening for redox-responsive FAP upon treatment with DTT.

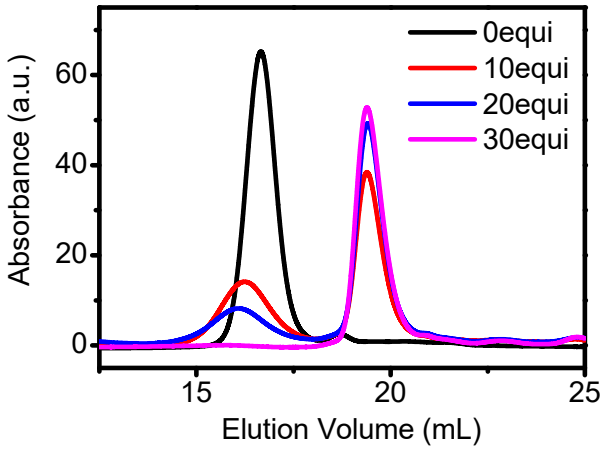


Figure 3.26. Equivalent-dependent dis-assembly study for Try-OEG-SS-C12-2T complex followed using SEC (Superose 6)..

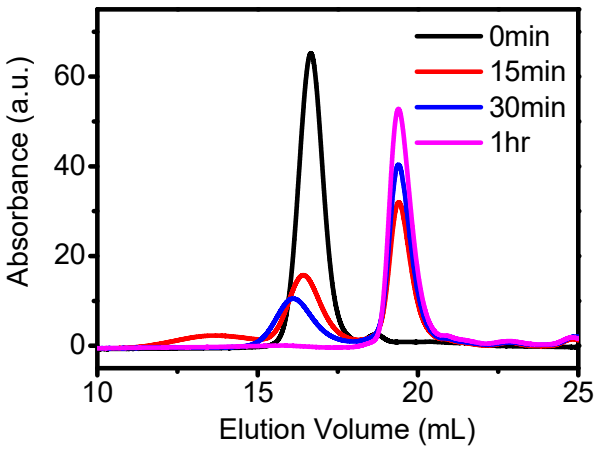


Figure 3.27. Time-dependent dis-assembly study for Try-OEG-SS-C12-2T complex followed using SEC (Superose 6).

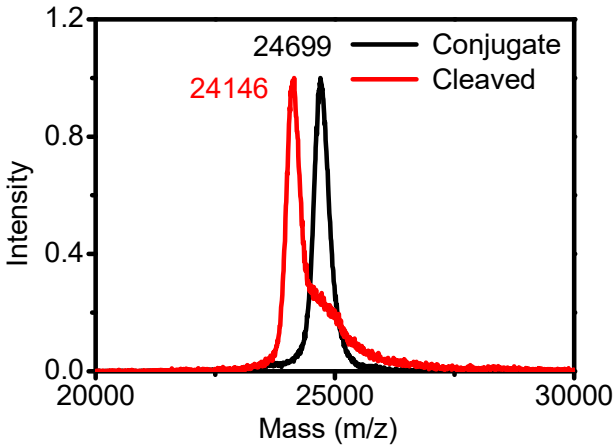


Figure 3.28. MALDI-TOF MS characterization of dis-assembly, molecular weight changes as seen for conjugate before and after cleavage.

3.1.7. Control studies

Control study I

In order to confirm that the disassembly is happening only because of the SS bridge in FAP and DTT, we have done three control experiments. In the first experiment, we have incubated the native Try (Fig. 3.29) with DTT for 12 hrs. The SEC profile shows that the protein undergoes degradation due to the presence of DTT and autolysis (Fig. 3.20).

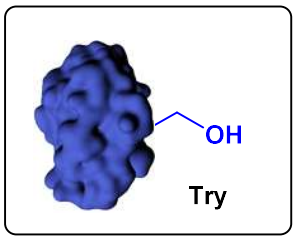


Figure 3.29. Structure of native Try.

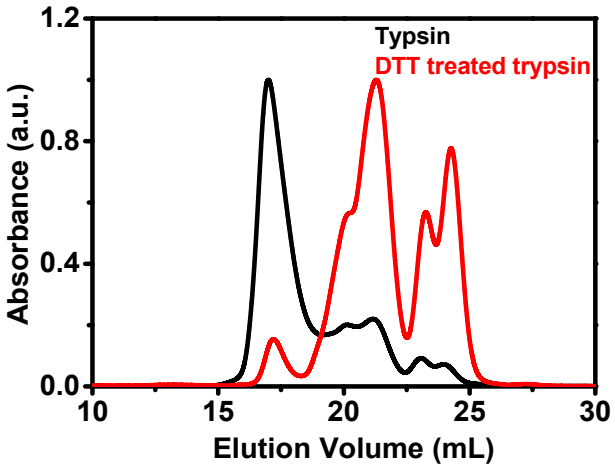


Figure 3.30. SEC chromatogram of native Try (black) DTT treated Try (red) in SEC (Superose 6).

Control study II

In the second experiment, we blocked the active site of protein with the flurophosphonate to prevent the autolysis (Fig. 3.31), and then we treated the conjugate with DTT for 12hrs. The SEC experiment revealed that the DTT alone has no effect on the elution volume when autolysis is prevented by blocking the active site (Fig. 3.32).

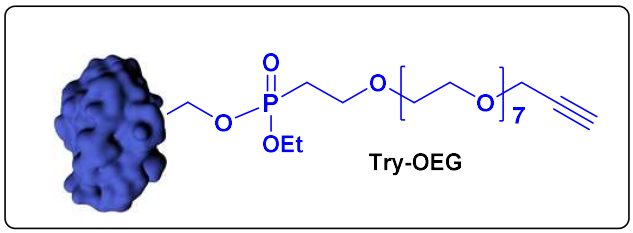


Figure 3.31. Structure of inhibited Try with FP-half probe.

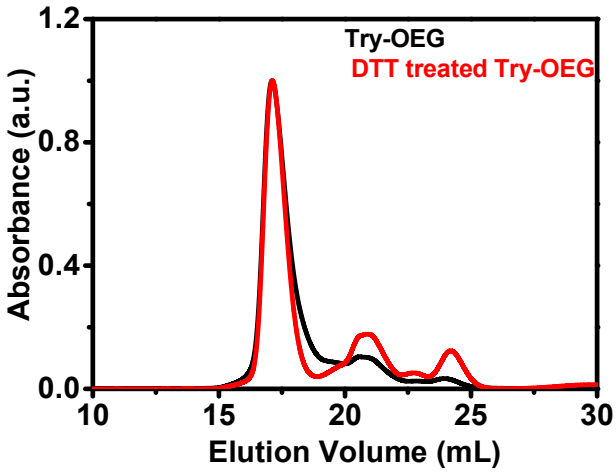


Figure 3.32. Try-OEG and (black) DTT treated try-OEG in SEC (Superose 6).

Control study III

Further, we have incubated the Try-OEG-C12-2T (a non-stimuli-responsive complex) with DTT to check the effect of the same (Fig. 3.33). This study also showed no effect on elution volume, indicating that complex is stable in the presence of DTT if it lacks the SS bridge in FAP design (Fig. 3.34).

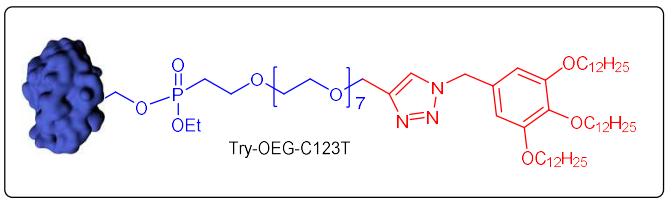


Figure 3.33. Structure of Try conjugate with FP- full probe.

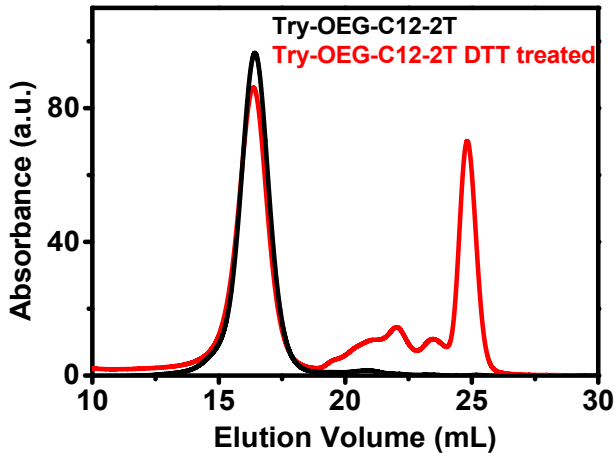
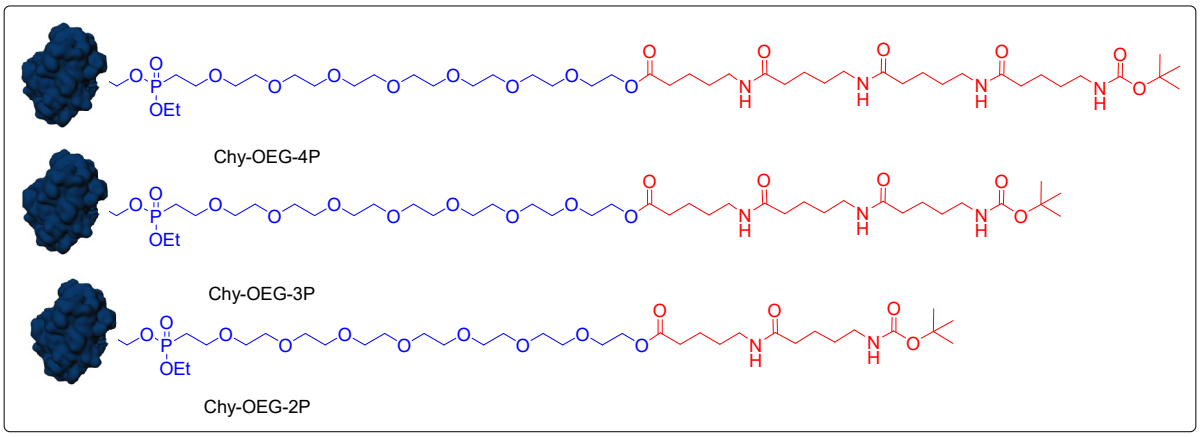


Figure 3.34. Try-OEG-C12-2T (black) and DTT treated Try-OEG-C12-2T, respectively in SEC (Superose 6).

3.2.8. Discussion

We have successfully demonstrated a sophisticated redox-responsive protein complex like never before in this chapter. By incorporating the SS bridge between tail and spacer, we have designed a new AABP for the construction of a redox-responsive protein complex. Try was converted into redox-responsive FAP using MAPLab technology. This FAP, self-assembled into higher order-complex with the unexpected high oligomeric state (17) obtained for any FAP with the OEG and C12-2T. We believe that this is a direct result of this FAP's flexibility and efficient packing since there is no triazole ring or 2-NB groups in the macromolecular design. Upon DTT treatment, the hydrophobic domain was lost, forcing the complex to disassemble. Unlike the photo-responsive complex, we have observed complete disassembly here with 30 eq of DTT within 60 mins. The process of cleavage of conjugate and dis-assembly of the complex was followed using time and eq dependent DTT exposure studies. The complete cleavage of the conjugate was also verified using SEC, MALDI-TOF MS analysis.

3.3. Synthesis and Self-assembly Studies of Peptide-based FAPs



3.3.1. Design

This chapter is about peptide-based FAPs. In the previous subchapters, we exploited the hydrophobic interaction as the only driving force for the construction of protein complexes. Natural assemblies are often driven by several non-covalent interactions like electrostatic force, hydrogen bonding, and hydrophobic interaction to drive the protein self-assembly.[11] As a first step in mimicking nature, in this subchapter, we disclose a new molecular design, in which both the hydrophobic and possible hydrogen-bonding forces drive the self-assembly process (Fig. 3.35). Peptide-based FAPs were made by changing hydrophobic tail groups into hydrophobic peptides, which provide the chance of hydrogen-bonding.

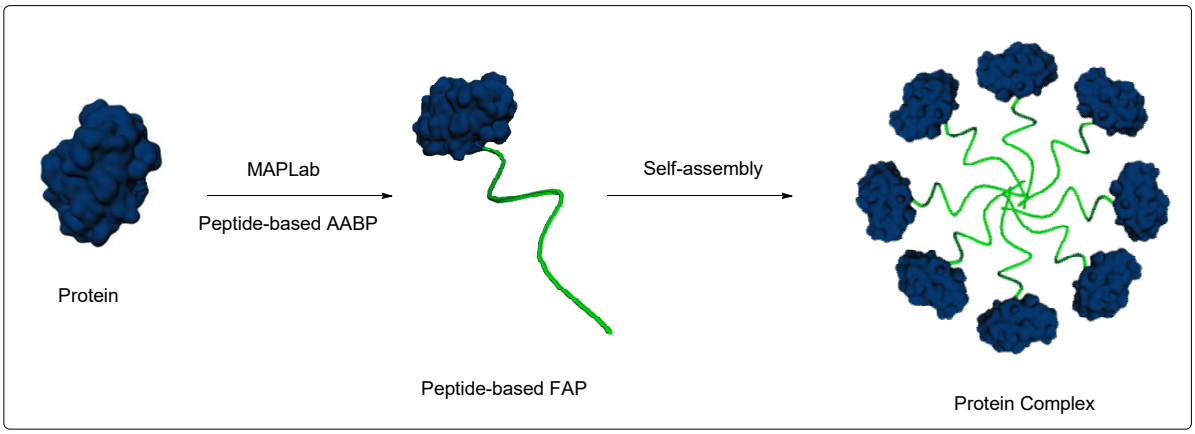


Figure 3.35. Cartoon representation of design, synthesis, and self-assembly of peptide-based FAPs.

We designed, made, and investigated two libraries here, protein library and peptide library. We conjugated the hydrophobic tetrapeptide AABP to 3 serine proteases using MAPLab to investigate whether the protein group has a similar effect, as observed in chapter 2 (Fig. 3.36). Then we have studied the tail library to know the influence of hydrophobic peptide tail group on self-assembly (Fig. 3.37).

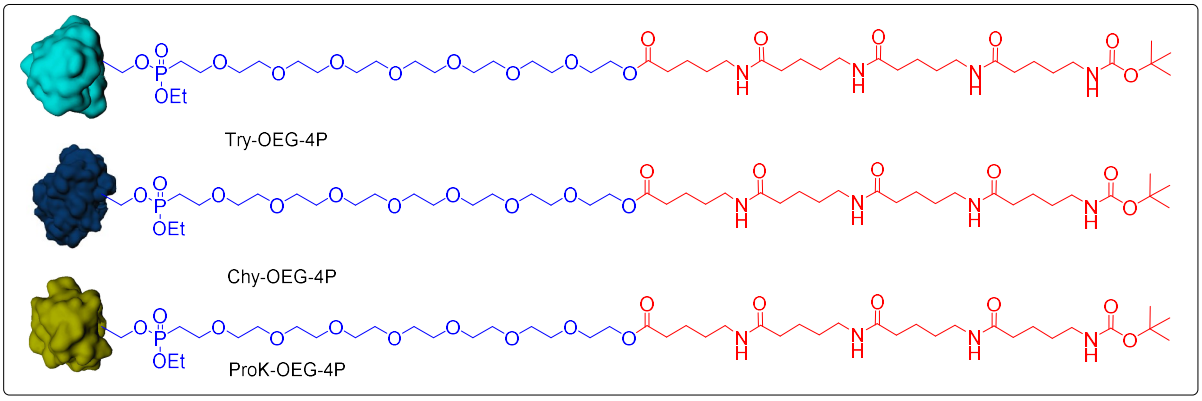


Figure 3.36. Structures of peptide-based FAPs in protein variant library.

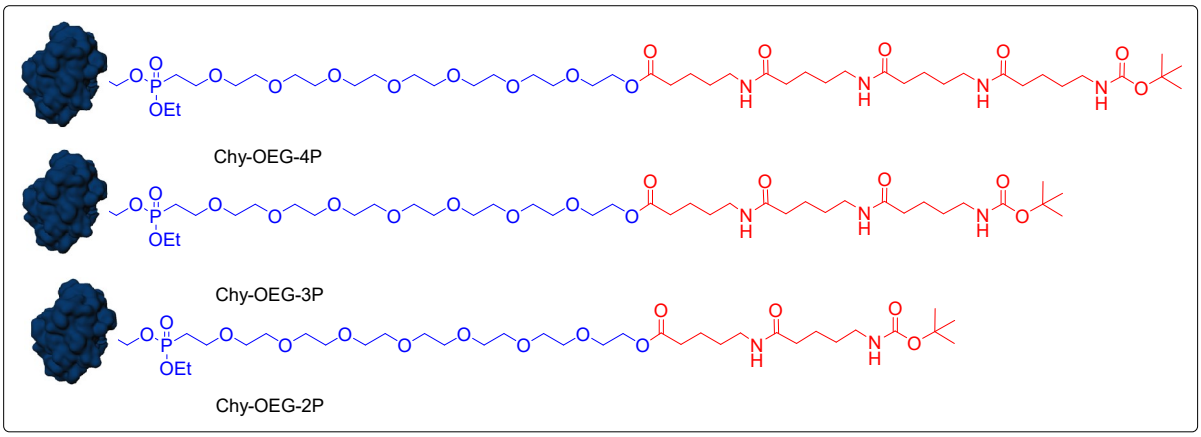
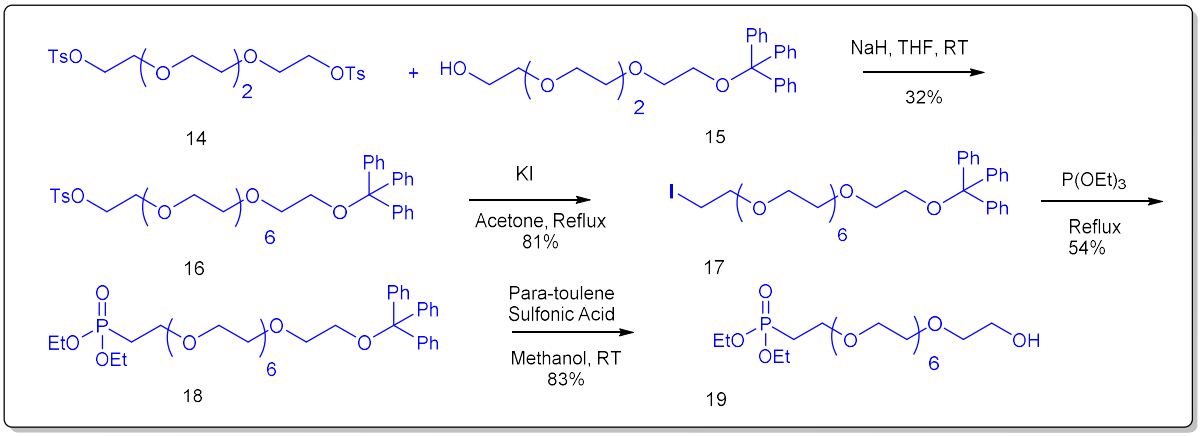


Figure 3.37. Structures of peptide-based FAPs in peptide variant library.

3.3.2. Synthesis

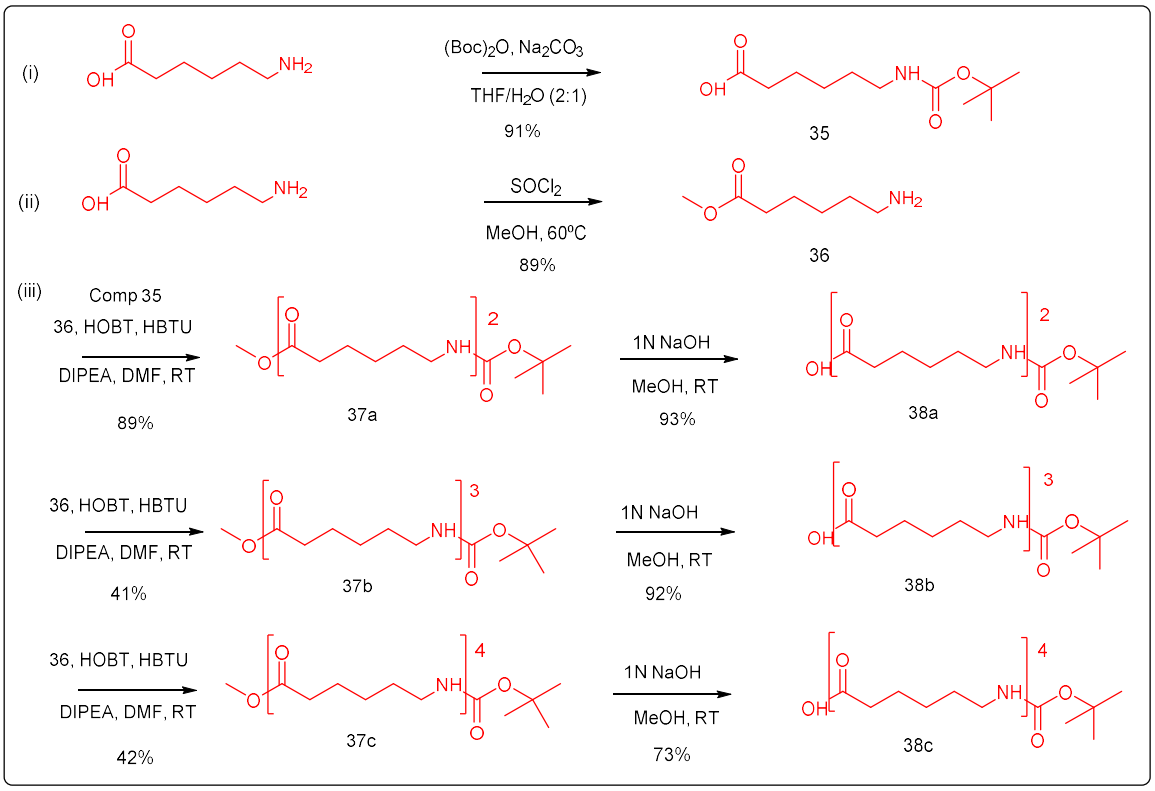
Synthesis of alcohol terminated



Scheme 3.7. Synthesis of alcohol-terminated OEG spacer.

The synthesis of this is discussed before in chapter 3.1. To brief, the reaction of monotrityl TEG (**15**) and ditosyl TEG (**14**) in the presence of NaH in THF afforded tosylate (**16**). The obtained tosylate (**16**) on reflux with KI in acetone yielded corresponding iodo derivative (**3d**). The iodide (**17**) on reflux with $P(OEt)_3$ provided compound **18**. Then, the treatment of diphosphonate ester (**18**) with TsOH in MeOH yielded hydroxyl derivative (**19**) (Scheme 3.7).

Synthesis hydrophobic peptides

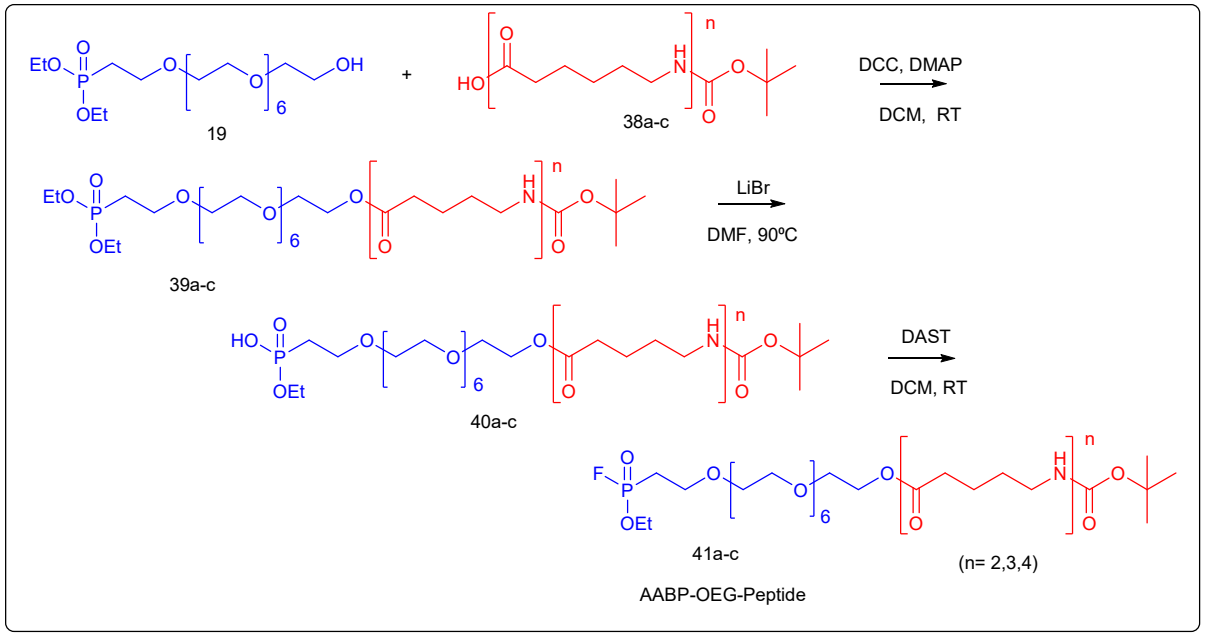


Scheme 3.8. Synthesis of Boc-protected hydrophobic peptides with free acid functionality.

For making peptide-based FAPs, we have made three different peptides, i.e., tetra, tri, and dipeptide, as shown here. 6-aminocaproic acid was amine protected in the first step. In an oven-dried RBF, a mixture of 6-aminocaproic acid and Boc anhydride ($(BOC)_2O$) was taken and dissolved in THF : H_2O (2:1) at 0 ºC under stirring. Subsequently, Na_2CO_3 solution was added to the above reaction mixture in small portions, and then the reaction mixture was stirred for 12 hrs at RT. Upon completion, excess of Na_2CO_3 was neutralized with 1N HCl, and the resultant mixture was extracted with EtOAc for thrice. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product (**35**), which was utilized in the next step without further purification. 6-aminocaproic acid was acid protected in the next step. To the suspension of 6-aminocaproic acid in MeOH, $SOCl_2$ was added dropwise at 0 ºC under a nitrogen atmosphere. Then, the reaction mixture was refluxed for 12 hrs. Upon completion, the solvent and excess $SOCl_2$ was removed under vacuum. Further, the solid residue was washed with dry Et_2O. The obtained solid residue (**36**) was used in the next step without any further purification. Then, the resultant acid (**35**) and amine (**36**) were coupled in this step. Mixture of acid (**35**) (1 eq), hydroxybenzotriazole (HOBT) (1

eq), and (2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) (1 eq) was dissolved in DMF. To this mixture, N,N-diisopropylethylamine (DIPEA) (3.5 eq) was added dropwise at 0 ºC under stirring. Further, the mixture was allowed to attend a room temperature. Then, the solution of amine (**36**) (1 eq) in DMF was added to the above reaction mixture stirred for 12 hrs RT. Upon completion, DMF was evaporated under reduced pressure. To the obtained residue, excess of EtOAc was added. The organic layer was then sequentially washed with 1 N HCl, saturated $NaHCO_3$ solution, and brine solution. Then, the washed organic layer was dried over Na_2SO_4 concentrated under reduced pressure to get the crude product, which was purified using column chromatography to get pale yellow liquid (**37a**). Keeping the amine protected, we hydrolyzed the acid functionality to get hydrophobic dipeptide (**38a)** in the next step. To brief, Boc protected polyamide (**37a**) (1 eq) was dissolved in MeOH at 0 ºC and 1 N NaOH (3 eq) was added dropwise. The reaction mixture was allowed to warm to room temperature and stirred for 8 hrs. After completion, the reaction mixture was concentrated under vacuum, and the obtained residue was acidified with 1 N HCl. The acidified solution was then extracted with EtOAc thrice. The combined organic layer was dried over Na_2SO_4 and concentrated under reduced pressure to get the crude product. Using this compound, we have similarly made tri and tetrapeptides with free acid terminus (**38b-c)** (Scheme 3.8).

Synthesis amphiphilic peptide-based AABPs

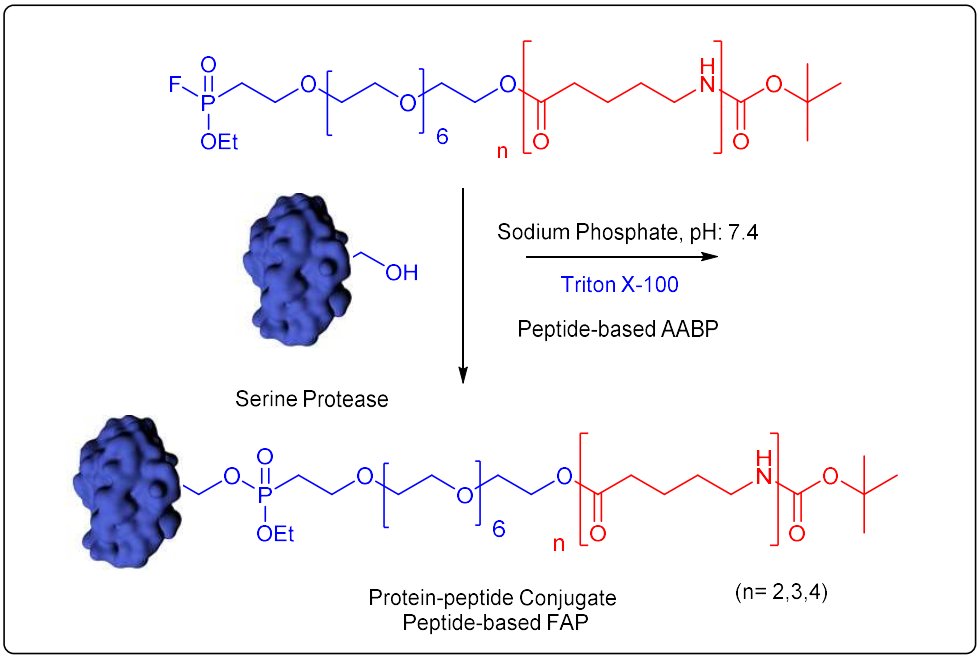


Scheme 3.9. General scheme for the synthesis of peptide-based AABPs.

After making tetra, tri, and dipeptides, as shown above, we have made 3 different AABPs with alcohol terminated OEG spacer (**19**). In an oven-dried RBF, acid (**38a-c**), alcohol (**19**), DCC (1.2 eq), and DMAP (1 eq) were taken and dissolved in DCM at 0 ºC under stirring. The reaction mixture was allowed to warm to room temperature and stirred for 12 hrs. Upon completion, EtOAc (excess) was added, and the resultant mixture was washed sequentially with 1 N HCl, saturated $NaHCO_3$, and brine solution, respectively. Then, the obtained organic layer was concentrated under reduced pressure to get crude product which was purified using column chromatography. In an oven-dried RBF, diphosphonate ester (**39a-c**) (1 eq) was taken, and LiBr (20 eq) was added. To the mixture, DMF was added and heated at 95 ºC for 20 hrs. Upon completion, water was added and extracted thrice with EtOAc. The water layer was collected, and 2 N HCl was added and stirred for another 30 mins. The mixture was then extracted thrice with EtOAc. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get crude product which was used without purification. To the stirring solution of monophosphonate ester (**40a-c)** (1 eq) in DCM, DAST (4 eq) was added dropwise at RT and allowed to react for 4 hrs. Excess of DAST and DCM were evaporated under reduced pressure. To the obtained residue, water was added and stirred for 2 more mins to quench any residual DAST. The reaction mixture was then extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product (**41a-c)** (Scheme 3.9).

3.3.3. Conjugation and purification

After the synthesis and characterization of peptide-based AABPs, we conjugated the probes to serine proteases using MAPLab (Scheme 3.10). Protein modification was carried out at a concentration of 100 µM. The test scale was done in a microcentrifuge tube at a 2.3 mg scale, and scale-up conjugation was done falcon tubes at 200 mg scale. The detailed protocol is given in the methods section of this chapter.



Scheme 3.10. General scheme for conjugation of peptide-based AABPs.

The conjugates were purified by IEX only, unlike previous chapters. IEX was performed to remove triton X-100 using SP sepharose (a cation-exchange resin at pH 7.4) for Try (30.4), Chy (19.5), and Q sepharose (an anion-exchange resin at pH 10) for Pro K (-8.4). The obtained IEX fractions were stored at -80 ºC without SEC and des immediately. To perform biophysical characterization and self-assembly studies, SEC was performed, and these fractions were used for further studies. The detailed procedure is given in the experimental section.

3.3.4. Biophysical characterization

Before investigating the self-assembly, we characterized these FAPs using MALDI-TOF MS, CD, and fluorescence studies, which did not show any significant structural changes in conjugates. To perform these biophysical characterization SEC purified fractions were used.

MALDI-TOF MS showed expected molecular weight changes for all the conjugates, as shown below (Fig. 3.32). In order to check if the secondary and tertiary structure of made protein FAPs is intact, detailed fluorescence studies were done.[36] We did not observe any significant increase or decrease in intensities in emission for any conjugate with respect to the native protein. Unfolding might result in decreased intensity. These studies revealed that, for Chy and ProK, there was no shift in tryptophan emission maximum (295 nm), which indicates the 3D structure of proteins is preserved in the complex state. Denaturation might result in redshift. But we have observed a blueshift in the case of TryOEG-4P, which needs further investigation (Fig. 3.39). CD spectroscopy[37] (195-250 nm) also indicated no perturbation in the secondary structure of the Chy-OEG-4P and Prok-OEG-4P FAPs but similar to fluorescence studies Try-OEG-4P showed perturbation in the structure when compared with native (Fig. 3.40).

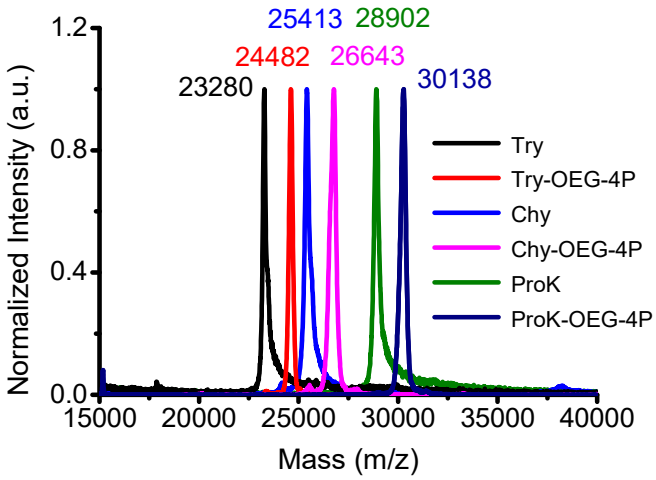


Figure 3.38. MALDI-TOF MS analysis of purified FAPs in protein variant library after purification.

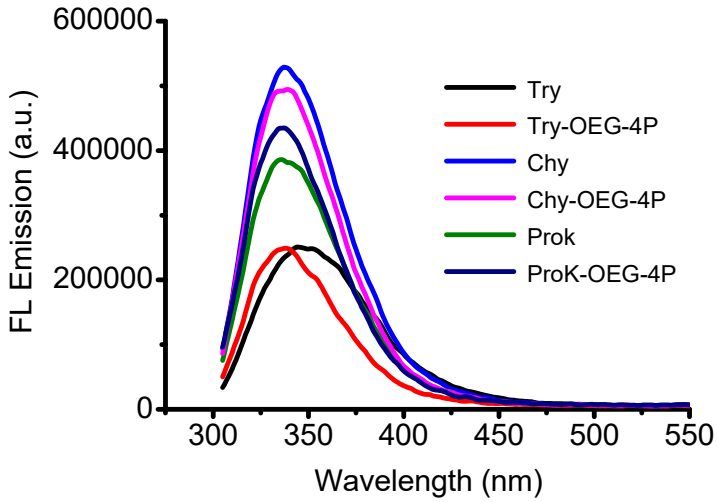


Figure 3.39. Fluorescence studies for purified peptide-based FAPs in protein variant library after purification.

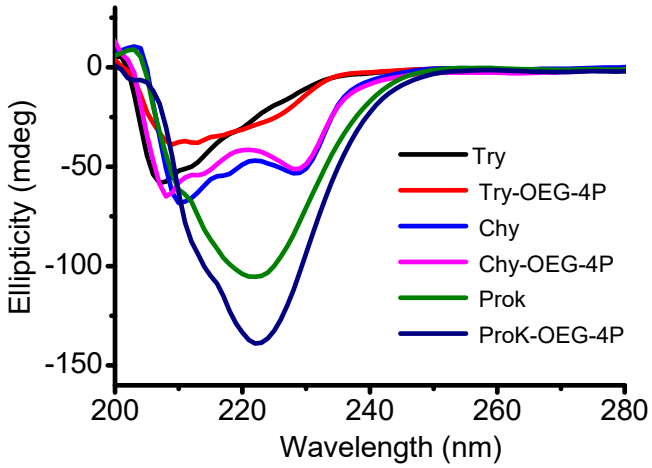


Figure 3.40. CD studies for purified peptide-based FAPs in protein variant library after purification.

3.3.5. Self-assembly studies

Self-assembly studies of protein variant library

After characterization, we focused our attention on self-assembly studies of peptide-based FAPs in the protein variant using complementary techniques like analytical size-exclusion chromatography (SEC), dynamic light scattering (DLS), SEC-coupled with multi-angle light scattering (SEC-MALS). To determine the relative molecular weights (RMWs) of soluble protein complexes, the FAP fractions stored at -80 ºC were subjected series of large scale SEC runs to remove native proteins. Then SEC studies were done on the analytical scale on Superose 6, which revealed that all the peptide-based FAPS in this protein variant library self-assembled into protein complexes of bigger size as evident from their lower elution volumes (15-17 mL). This result indicates that the tetrapeptide-AABP has sufficient hydrophobicity to drive the self-assembly of highly charged large globular proteins. SEC studies revealed that the protein complex formed by ProK-OEG-4P is slightly bigger than that of Try-OEG-4P and Chy-OEG-4P (Fig. 3.41). The mean size of the protein complexes was measured using DLS, and the D_h of protein complexes were obtained in the range of 10-13 nm, indicating the formation of higher-order protein complexes (Fig. 3.42). To get information such as AMWs and oligomeric states of the protein complexes, SEC-MALS studies were done on a Superose 6 column. For detailed procedures, refer to the methods section of this chapter. SEC-MALS results revealed that the molecular mass of protein complexes formed by Prok is about 550 kDa, whereas Try and Chy is about 322 and 246 kDa, respectively. The oligomeric state of ProK-OEG-4P is about 18 mer, whereas Chy-OEG-4P and Try-OEG-4P is about 9 mer and 13 mer, respectively (Fig. 3.41). Both SEC, DLS results are in agreement with SEC-MALS results. All studies showed that proK-OEG-4P is forming larger complex than the rest and the tetrapeptide have sufficient hydrophobicity to drive the self-assembly

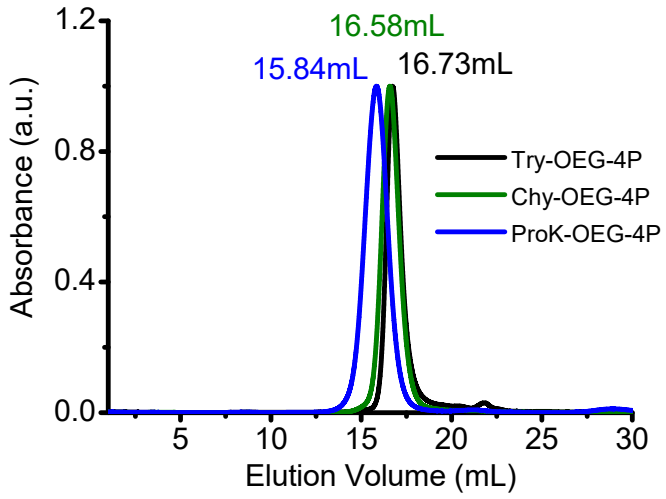


Figure 3.41. SEC data of peptide-based FAPs in protein variant library after purification.

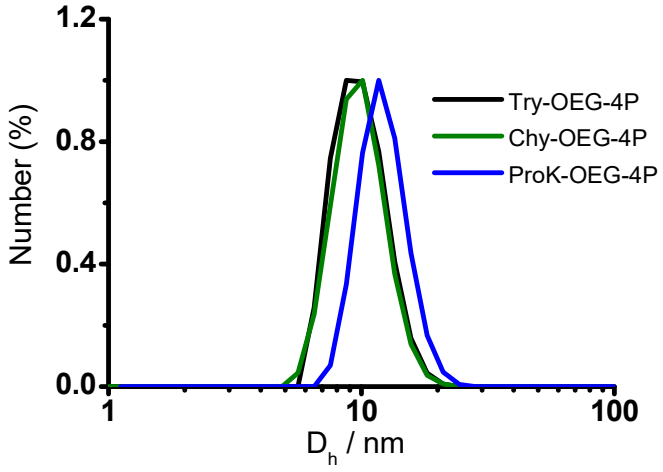


Figure 3.42. DLS data of peptide-based FAPs in protein variant library after purification.

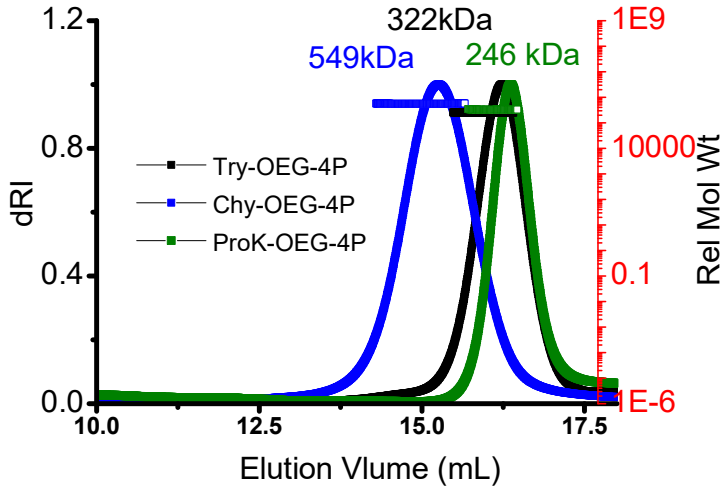


Figure 3.43. SEC-MALS data of peptide-based FAPs in protein variant library after purification.

Table 3.3. Summary of self-assembly data of the FAPs in protein variant library.

Protein-Peptide Conjugates	Elu Vol SEC (mL)	M_w SEC (kDa)	M_w MALS (kDa)	Oligomeric State MALS (mer)	D_h DLS (nm)
Try-OEG-4P	**16.7**	**150**	**322**	**13**	**10**
Chy-OEG-4P	**16.6**	**175**	**246**	**9**	**9**
ProK-OEG-4P	**15.8**	**300**	**549**	**18**	**13**

Self-assembly studies of peptide variant library

Next, we focused our attention on the self-assembly studies of peptide-based FAPs having different hydrophobic peptide length (peptide variant library). We choose the Superdex-200 column to analyze size differences in peptide variant FAPs (Chy-OEG-4P, Chy-OEG-3P, Chy-OEG-2P). FAP with tetrapeptide, showed higher-order complex with the similar oligomeric state (SEC calibration curve derived RMW were also similar). Further, these SEC studies reveal that the size of Chy-OEG-3P with tripeptide is smaller (2 mer) than Chy-OEG-4P with tetrapeptide (9 mer). Interestingly, Chy-OEG-2P with dipeptide having a short hydrophobic domain failed to oligomerize and hence eluted at the same volume as native monomeric native protein (Fig. 3.44). As expected, in this chapter, we saw an increase in the size of the complex, with an increase in peptide length. These results are in agreement with

our previous chapter 2 in which the conjugates in the C6 series, 3-Tail variant (Try-OEG-C6-3T), formed a trimer at much higher concentrations. Whereas Try-OEG-C6-2T dint assemble as it does not have enough hydrophobicity.[10]

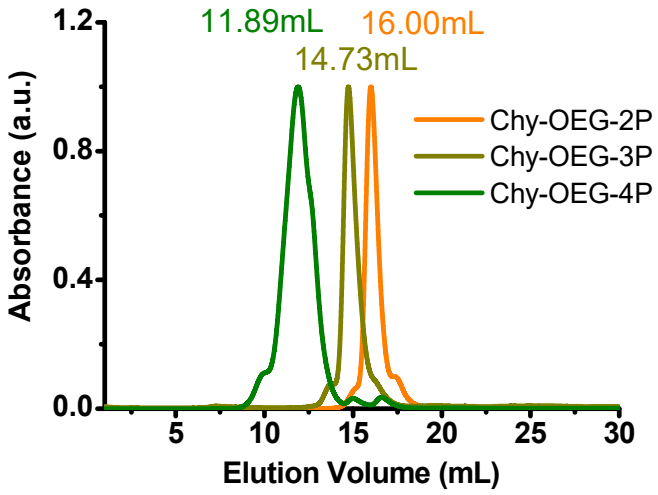


Figure 3.44. SEC data of peptide-based FAPs in peptide variant library after purification.

Table 3.4. Summary of self-assembly data of the FAPs in peptide variant library.

Protein-Peptide Conjugates	Elu Vol SEC (mL)	Mw SEC (kDa)	Oligo State SEC (mer)	D_h DLS (nm)
Chy-OEG-4P	11.9	190	9	9
Chy-OEG-3P	14.7	60	2	--
Chy-OEG-2P	16.0	30	1	--

3.3.6. Discussion

Protein variant library

In this study, not surprisingly, we saw the similar influence of the head group in this sub-library, like the Try and Chy forming smaller complexes and Prok forming a much higher order complex. The above results are similar to previous results, which suggest that the protein having low surface charge (close to neutral) has the tendency to oligomerize better compare to proteins with high surface charges. In our previous study also, we have shown

that the size of the protein complex is dictated by the size of the native protein (R_g) provided the linker, and the hydrophobic tail is the same.[12] The results from the present study also indicate the same[13].

Peptide variant library

When we investigated the effect of peptide tail keeping head group the same as Chy, as expected, dipeptide and tripeptide FAPs did not form higher-order complexes. These results manifest that the peptide domain should have minimum hydrophobicity to drive the complex formation of proteins.[13] Our results are in agreement with other reports as well.[12] From the above results, we propose the following mechanism; the main driving force for the self-assembly of peptide-based FAPs is the attractive interaction of hydrophobic peptide domains however this is opposed by repulsive electrostatic repulsion between the hydrophilic globular proteins (because of the presence of charged amino acids at physiological pH i.e. Asp and Glu, Lys and Arg). This concludes this stimuli-responsive chapter 3, which brings us to my last chapter I am currently working on.

Table 3.5. Analysis of self-assembly data of peptide-based FAPs.

Protein-Peptide Conjugates	M_w MALS (kDa)	Oligo State MALS (mer)	Observation
ProK-OEG-4P	**549**	**18**	**similar to chapter 2.2**
Try-OEG-4P	**322**	**13**	**similar to chapter 2.2**
Chy-OEG-4P	**246**	**9**	**similar to chapter 2.2 & 2.4**
Chy-OEG-3P	**---**	**2**	**similar to chapter 2.4**
Chy-OEG-2P	**---**	**1**	**similar to chapter 2.4**

3.4. Experimental Section and Methods

Selection of proteins used in this work

We chose 4 serine proteins for our study, *i.e.*, Try (23 kDa), Chy (25 kDa), and ProK (29 kDa). Proteins were purchased from commercial vendors, and molecular weights were determined prior to modification to track the molecular weight changes accurately. Matrix preparation and molecular weight determination were done in the same instruments, using the same protocol mentioned in chapter 2. Protein modification, monitoring, and purification were also done in the same instruments, using the same protocol mentioned in chapter 2.

Synthesis and purification of photo-sensitive protein conjugate (Try-OEG-NB-C12-2T)

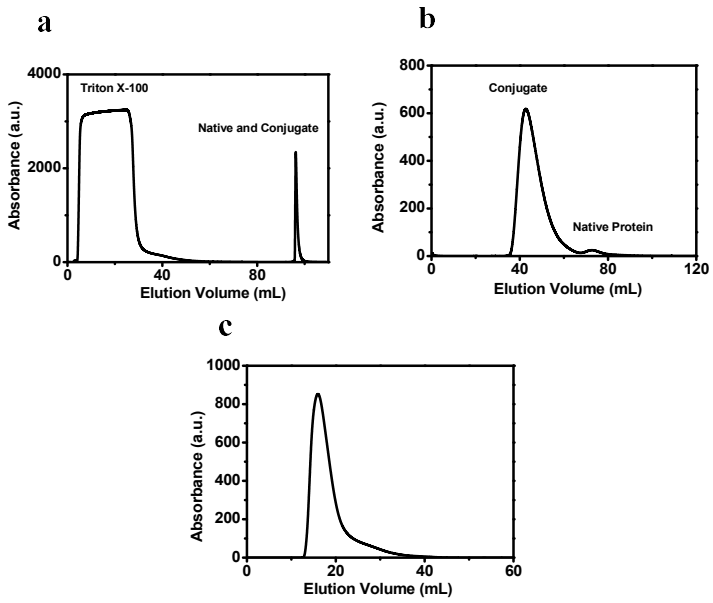


Figure 3A. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram

Synthesis and purification of redox-sensitive protein conjugate (Try-OEG-SS-C12-2T)

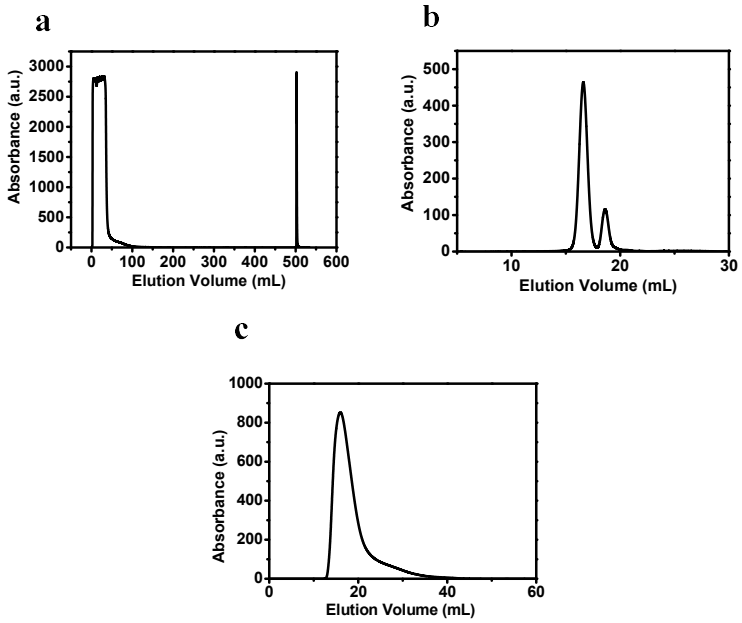


Figure 3B. a. MALDI-TOF MS spectrum of the reaction mixture, b. Ion exchange chromatogram, c. Size-exclusion chromatogram, d. Desalting chromatogram.

Synthesis and purification of Try-OEG-4P conjugate

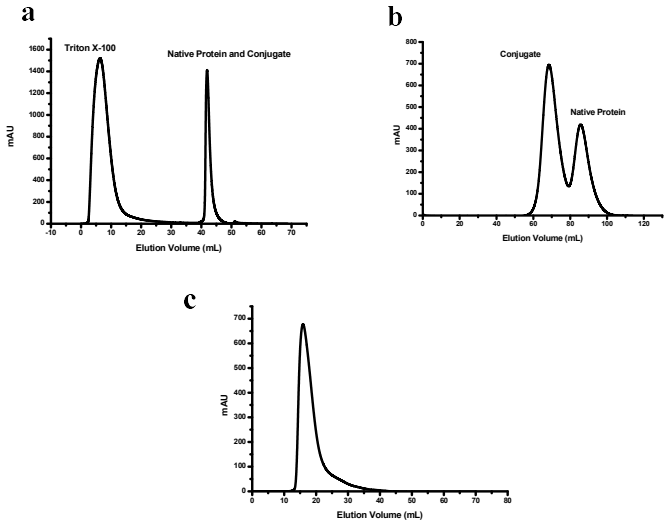


Figure 3C. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram.

Synthesis and purification of Chy-OEG-4P conjugate

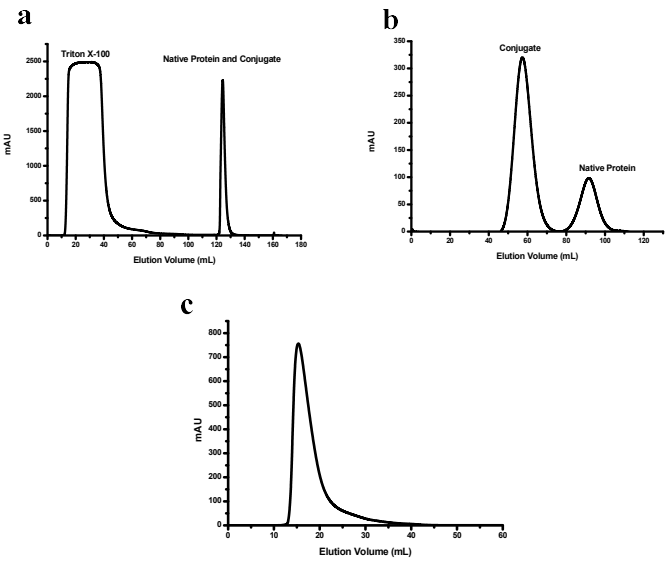


Figure 3D. a. Ion exchange chromatogram, b. Size-exclusion chromatogram, c. Desalting chromatogram.

Synthesis and purification of ProK-OEG-4P

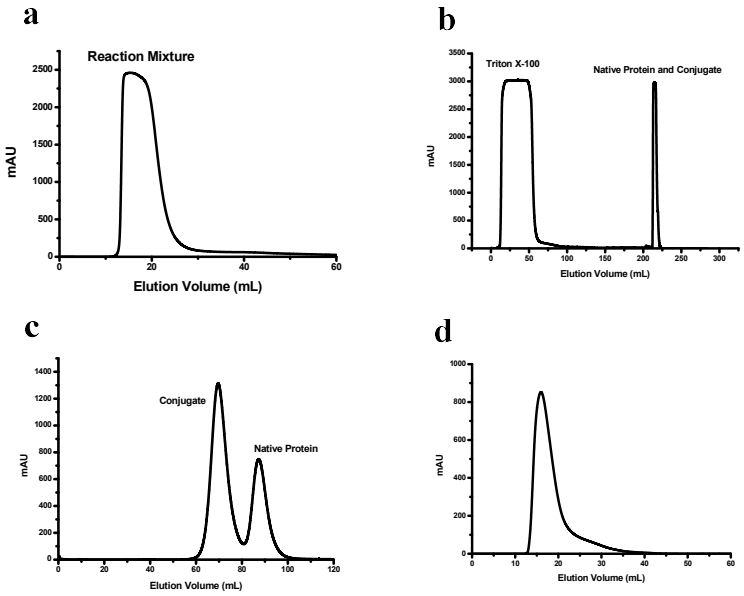


Figure 3E. a. Buffer exclusion chromatogram, b. Ion exchange chromatogram c. Size-exclusion chromatogram c. Desalting chromatogram.

Synthesis and purification of Chy-OEG-3P conjugate

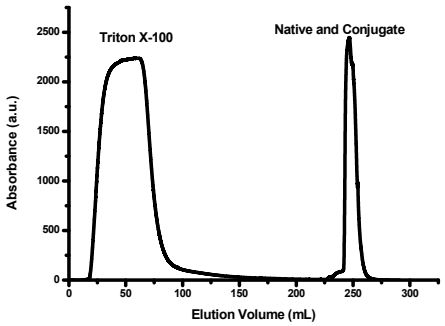


Figure 3F. Ion exchange chromatogram

Synthesis and purification of Chy-OEG-2P

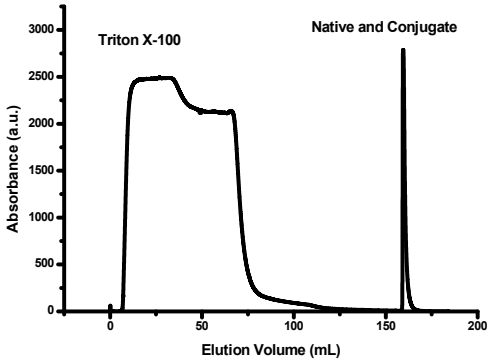


Figure 3G. Ion exchange chromatogram

Size exclusion-based molecular weight determination

To determine the molecular weights of soluble protein complexes, we performed SEC runs for standard proteins in Superdex-200 10/300 GL Superose-6 Increase 10/300 GL (GE healthcare). For this purpose, we chose standard proteins the same as chapter 2.

SEC based MW determination in Superdex-200

A series of size exclusion runs were performed with these proteins in 50 mM Sod. Phos. pH 7.4, 1 M NaCl with 0.25 mL/min as the flow rate. All the proteins were dissolved individually in Milli Q water at 3 mg/mL and 0.5 mL was injected. With the obtained elution volumes and the partition coefficients (K_{av}), the calibration curve was plotted, and the RMWs of soluble protein complexes was determind.

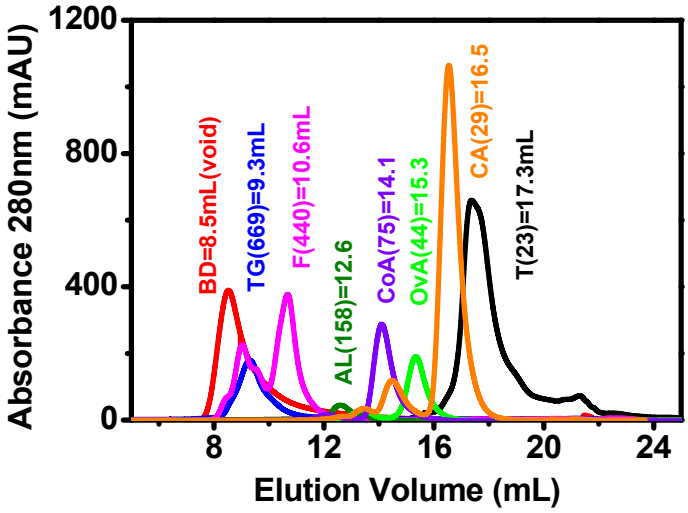


Figure 3H. SEC runs for standard proteins in Superdex-200

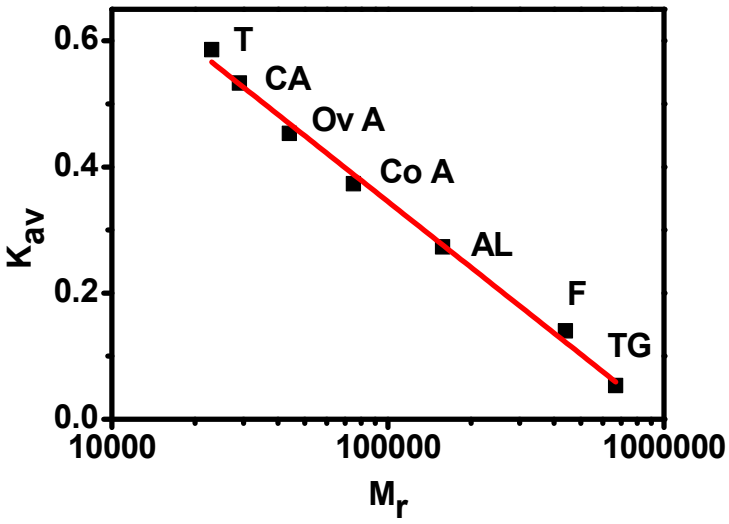


Figure 3I. Calibration curve for Superdex-200

SEC based MW determination in Superose-6

A series of size exclusion runs were performed with these proteins in 50 mM Sod. Phos. pH 7.4, 200 M NaCl with 0.25 mL/min as the flow rate. All the proteins were dissolved individually in Milli Q water at 3 mg/mL and 0.5 mL was injected. With the obtained elution volumes and the partition coefficients (K_{av}), the calibration curve was plotted, and the MW of soluble protein complexes was determined.

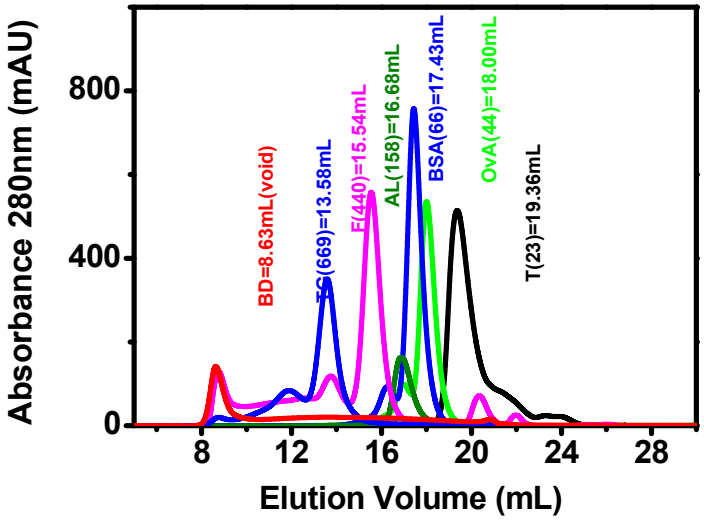


Figure 3J. SEC runs for standard proteins in Superose-6

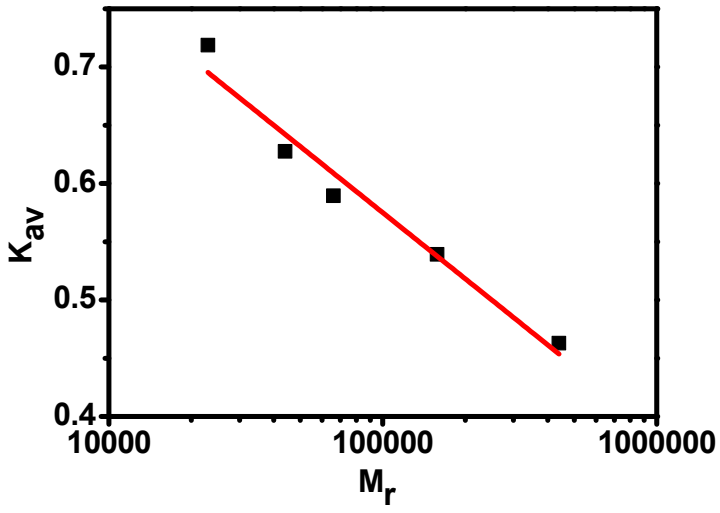


Figure 3K. Calibration curve for Superose-6

Dynamic light scattering, fluorescence spectroscopy, CD spectroscopy, and zeta potential measurement were done in the same instruments, using the same protocol mentioned in chapter 2

Programmed Dis-assembly of the photo-sensitive supramolecular protein complex

The photo-sensitive protein conjugate was dissolved in 50 mM Sod. Phos. pH 7.4 in a 50 mL beaker and temperature was maintained at 0 ºC using an ice bath over the entire reaction. The sample was exposed to the Sankyo Denki G8T5E UVB linear lamp (8 W output), which

emits ultraviolet rays between 280 to 360 nm (peak at 305-315 nm). Samples were withdrawn at different time points using a pipette, and analysis of the same was carried out using SEC and MALDI-TOF MS.

Programmed Dis-assembly of the redox-sensitive supramolecular protein complex

The redox-sensitive protein conjugate was dissolved in 50 mM Sod. Phos. pH 7.4 in a 2 mL centrifuge tube, and the room temperature was maintained over the entire reaction. The different equivalent of DTT (aq. solution) was added, and then the sample was incubated for 12 hrs after gentle mixing with a pipette. After that, the sample was analyzed using SEC and MALDI-TOF MS. This study revealed that the 30 eq of DTT is sufficient to dis-assemble the complex completely due to the clipping of the hydrophobic tail by DTT. Further, the complex was incubated with 30 eq of DTT for different time points. The SEC studies revealed that in 15 mins, 80 % of complex dis-assembles, whereas in 30 mins, it disassembles to 90 %, while the complete cleavage is observed in 60 mins.

Synthesis and purification of AABPs and their intermediates

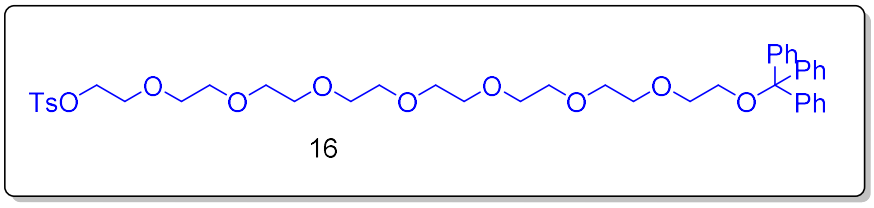
General remarks

All compounds were made, characterized using similar conditions and instruments, as mentioned in chapter 2.

AABPs in chapter 3.1 and their intermediates

Synthesis of amine-terminated octaethylene glycol spacer and its intermediates

Synthesis of compound 16



16

Mol. formula: $C_{42}H_{54}O_{11}S$

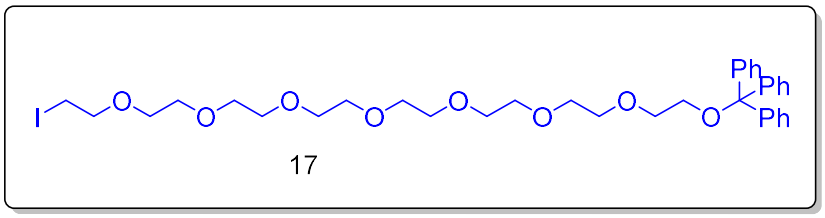
Mol. weight: 766.34 g/mol

Physical appearance: brownish liquid

Yield: 32 %

In an oven-dried RBF, a mixture of ditosyl TEG (**14**) (4.5 g, 10.4 mmol) and monotrityl TEG (**15**) (12.6 g, 25.0 mmol) was taken and dissolved in THF under stirring. The reaction mixture was cooled to 0 ºC and NaH (1.25 g, 52.0 mmol) was added in small portions. The resultant mixture was stirred at RT for 12 hrs. Upon completion of the reaction, excess of NaH was quenched by the dropwise addition of water. The resultant mixture was concentrated under reduced pressure. To the obtained residue, water was added and extracted thrice with DCM. Then, the combined organic layer was dried over Na_2SO_4 and concentrated under reduced pressure to get the crude product, which was purified by NPC using EtOAc / hexane as eluent to get brownish liquid (2.6 g, 3.39 mmol, 32 %), R_f = 0.40 in 100 % EtOAc. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.78 (d, J = 7.6 Hz, 2H), 7.51-7.41 (m, 5H), 7.37-7.16 (m, 10H), 4.14 (t, J = 4.2 Hz, 2H), 3.72-3.50 (m, 26H), 3.23 (t, J = 5.2 Hz, 2H), 2.41 (s, 3H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 144.84, 144.15, 133.00, 129.87, 128.74, 128.02, 127.97, 127.93, 127.80, 126.96, 86.55, 70.81, 70.75, 70.69, 70.62, 70.58, 70.54, 69.30, 68.69, 63.35, 21.69. **MALDI-TOF MS** (MW+K): 805.13 g/mol.

Synthesis of compound 17



17

Mol. formula: $C_{35}H_{47}IO_8$

Mol. weight: 722.23 g/mol

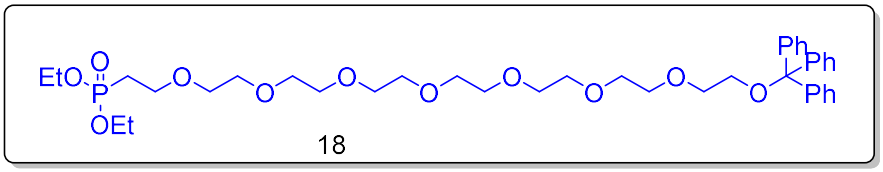
Physical appearance: brown-yellow liquid

Yield: 81 %

The **compound 17** was prepared from **16** (2.6 g, 3.4 mmol), KI (1.7 g, 10.1 mmol) in acetone. The product was obtained as brownish yellow liquid (2.0 g, 2.7 mmol, 81%), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.54-7.40(m, 6H), 7.34-7.19(m, 9H), 3.74 (t, J = 6.8 Hz, 2H), 3.70-3.57 (m, 25H), 3.30-3.16 (m, 4H). **^{13}C NMR (**100 MHz,

$CDCl_3$): δ_C 144.15, 128.75, 127.80, 126.96, 86.57, 72.00, 70.80, 70.72, 70.70, 70.68, 70.62, 70.59, 70.54, 70.23, 63.36. **MALDI-TOF MS** (MW+K): 761.13 g/mol.

Synthesis of compound 18



18

Mol. formula: $C_{39}H_{57}O_{11}P$

Mol. weight: 732.36 g/mol

Physical appearance: brownish liquid

Yield: 54 %

The **compound 18** was prepared from **compound 17** (2.0 g, 2.7 mmol) and $P(OEt)_3$ (2.2 g, 13.8 mmol). The product was obtained as brownish liquid (1.1 g, 1.5 mmol, 54 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.52-7.41 (m, 6H), 7.34-7.16 (m, 9H), 4.20-4.04 (m, 4H), 3.82-3.53 (m, 28H), 3.23 (t, J = 6.8 Hz, 2H) 2.18-2.03 (m, 2H), 1.31 (t, J = 7.2 Hz, 6H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 144.10, 128.68, 127.73, 126.89, 86.50, 70.76, 70.68, 70.64, 70.58, 70.53, 70.42, 70.16, 65.11, 63.31, 61.56, 27.66, 26.28, 16.40. **MALDI-TOF MS** (MW+K): 771.25 g/mol.

Synthesis of compound 19



19

Mol. formula: $C_{20}H_{43}O_{11}P$

Mol. weight: 490.52 g/mol

Physical appearance: pale yellow liquid

Yield: 83 %

In an oven dried RBF, a mixture of **compound 18** (9.3 g, 12.7 mmol) and TsOH (3.6 g, 19.1 mmol) was dissolved in MeOH and stirred for 12 hrs at RT. Upon completion of reaction, solvent was evaporated under reduced pressure. Then, the obtained residue purified using NPC using MeOH / DCM as eluent to get brownish liquid (**compound 19)** (5.9 g, 12.0 mmol, 83 %), R_f = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.11-3.98 (m, 4H), 3.70-3.49 (m, 30H), 2.14-2.03 (m, 2H), 1.27 (t, J = 7.2 Hz, 6H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 72.63, 70.60, 70.56, 70.55, 70.52, 70.45, 70.29, 70.20, 65.13, 27.68, 26.30, 16.43. **MALDI-TOF MS** (MW+K): 529.25 g/mol.

Synthesis of compound 20



20

Mol. formula: $C_{27}H_{49}O_{13}PS$

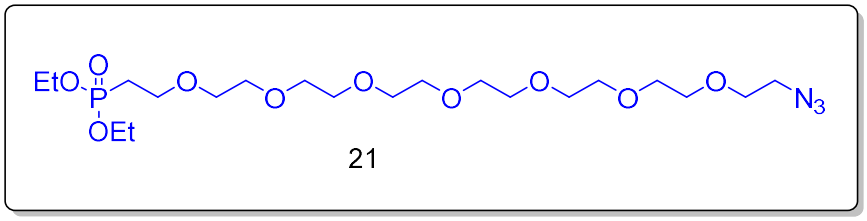
Mol. weight: 644.26 g/mol

Physical appearance: pale yellow liquid

Yield: 61 %

In an oven dried RBF, **compound 19** (5.2 g, 11.8 mmol), TsCl (6.2 g, 33 mmol), DMAP (0.2 g, 1.6 mmol) were taken and was dissolved in DCM under at 0 ºC. Then, Et_3N (4.7 g, 47.3 mmol) was added to the above mixture and allowed to stir for 12 hrs at RT. Upon completion, the reaction was quenched with water and extracted in DCM for thrice. The combined organic layer was dried over Na_2SO_4 and concentrated under reduced pressure to get crude product which was purified using NPC using MeOH / DCM as eluent to get pale yellow liquid (**compound 20)** (4.2 g, 6.5 mmol, 61%), R_f = 0.40 in 5% MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.75 (d, J = 8 Hz, 2H), 7.30(d, J = 8 Hz, 2H), 4.17-4.00 (m, 6H), 3.74-3.50 (m, 30H), 2.42 (s, 3H), 2.13-2.05 (m, 2H), 1.29 (t, J = 7.2 Hz, 6H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 144.85, 133.05, 129.89, 128.03, 77.16, 70.78, 70.65, 70.60, 70.49, 70.23, 69.31, 68.73, 65.17, 61.73, 61.66, 27.71, 26.33, 21.70, 16.52, 16.46. **MALDI-TOF MS** (MW+K): 683.25 g/mol.

Synthesis of compound 21



21

Mol. formula: $C_{20}H_{42}N_3O_{10}P$

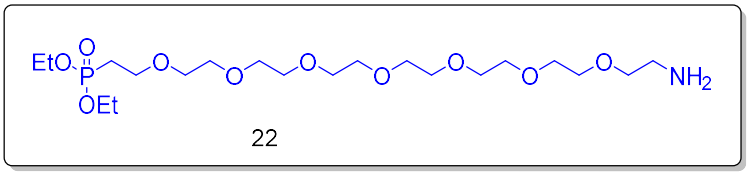
Mol. weight: 515.54 g/mol

Physical appearance: pale yellow liquid

Yield: 84 %

In an oven-dried RBF, a mixture of **compound 20** (4.1 g, 6.0 mmol) and NaN_3 (4.1 g, 63.0 mmol) was taken and dissolved in DMF under stirring. The resultant mixture was allowed to react for 12 hrs. Upon completion of reaction, DMF was evaporated under reduced pressure. Then, the obtained residue was purified using NPC using MeOH / DCM as eluent to get pale yellow liquid (**compound 21)** (2.7 g, 5.2 mmol, 84 %), R_f = 0.40 in 5% MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.13-3.99 (m, 4H), 3.74-3.54 (m, 29H), 3.36 (t, J = 6.8 Hz, 2H), 2.21-2.01 (m, 2H), 1.29 (t, J = 7.2 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 70.75, 70.72, 70.68, 70.62, 70.50, 70.24, 70.09, 65.18, 50.73, 27.71, 26.33, 16.47. **MALDI-TOF MS** (MW+K): 554.35 g/mol.

Synthesis of compound 22



22

Mol. formula: $C_{20}H_{44}O_{10}P$

Mol. weight: 489.27 g/mol

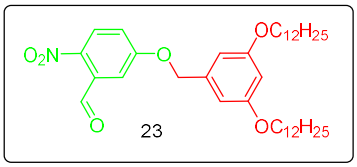
Physical appearance: pale yellow liquid

Yield: 96 %

In an oven-dried RBF, **compound 21** (2.7 g, 5.2 mmol) was taken and dissolved in THF under stirring. Then, PPh_3 (2.0 g, 7.0 mmol) was added to the above mixture at 0 ºC and allowed to react for 18 hrs at RT. Upon completion of the reaction, a few drops of water were added and again stirred for 1 hr. Then, THF was evaporated under reduced pressure, and the obtained water layer was washed with toluene for at least four times. The water layer was then concentrated under reduced pressure to get the product that was used without further purification (**compound 22)** (2.4 g, 4.9 mmol 96 %), R_f = 0.59 in 2 % Et_3N in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.19- 4.02 (m, 4H), 3.78-3.55 (m, 31H), 2.17-2.08 (m, 4H), 1.31 (t, J = 7.2 Hz, 6H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 70.65, 70.62, 70.42, 70.62, 70.50, 70.24, 70.09, 65.17, 50.74, 27.76, 26.37, 16.37. **MALDI-TOF MS** (MW+K): 528.45 g/mol.

Synthesis of photo-sensitive AABP and its intermediates

Synthesis of compound 23



Mol. formula: $C_{38}H_{59}NO_6$

Mol. weight: 625.43 g/mol

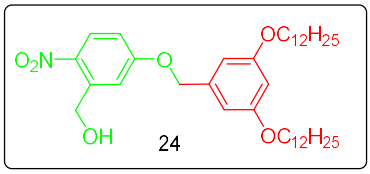
Physical appearance: pale yellow solid

Yield: 90 %

In an oven-dried RBF, 5-hydroxy-2-nitrobenzaldehyde (0.11 g, 0.65 mmol), K_2CO_3 (0.11 g, 0.79 mmol) and compound **9e** (0.40 g, 0.74 mmol) were dissolved in DMF under stirring at RT and allowed to react for 12 hrs. Upon completion of reaction, water was added and extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and evaporated under vacuum to get crude product which was purified using NPC using EtOAc / PE as eluent to get pale solid (0.36 g, 0.57 mmol, 90 %), R_f = 0.52 in 20 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 10.47 (s, 1H), 8.15 (d, J = 8.9 Hz, 1H), 7.40 (d, J = 8.9 Hz, 2.8 Hz, 1H), 7.18 (d, J = 2.8 Hz, 1H), 6.52 (d, J = 2.4 Hz, 2H), 6.42 (t, J = 2.4 Hz, 1H), 5.12 (s, 2H), 3.93 (t, J = 6.4 Hz, 4H), 1.85-1.68 (m, 4H), 1.48-1.26 (m, 44H), 0.88 (t, J = 7.2 Hz, 6H).

^{13}C NMR (100 MHz, $CDCl_3$): δ_C 188.57, 163.17, 160.82, 142.45, 137.13, 134.40, 127.40, 119.29, 114.50, 105.78, 101.22, 77.16, 71.14, 68.25, 32.04, 29.79, 29.77, 29.73, 29.71, 29.52, 29.48, 29.34, 26.16, 22.82, 14.26. **MALDI-TOF MS** (MW+K): 663.73 g/mol.

Synthesis of compound 24



Mol. formula: $C_{38}H_{61}NO_6$

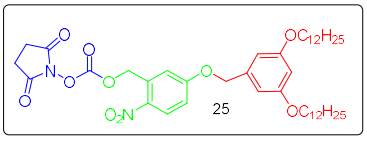
Mol. weight: 627.43 g/mol

Physical appearance: pale yellow solid

Yield: 93 %

In an oven dried RBF, compound **23** (0.40 g, 0.64 mmol) was dissolved in THF (2 mL) and MeOH (0.8 mL) under stirring. Then, the mixture was cooled to 0 ºC and $NaBH_4$ (0.036 g, 0.94 mmol) was added in small portions and allowed to react for 1 hr. Upon completion, reaction was quenched with Na_2CO_3 and extracted thrice with EtOAc . The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get crude product which was purified using NPC using EtOAc / PE as eluent to get pale yellow solid (0.37 g, 0.59 mmol, 93 %), R_f = 0.20 in 20 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 8.15 (d, J = 8.9 Hz, 1H), 7.40 (dd, J = 8.9, 2.8 Hz, 1H), 7.18 (d, J = 2.8 Hz, 1H), 6.52 (d, J = 2.4 Hz, 2H), 6.42 (t, J = 2.4 Hz, 1H), 5.12 (s, 2H), 4.80 (s, 2H), 3.93 (t, J = 6.4 Hz, 4H), 1.85-1.68 (m, 4H), 1.48-1.26 (m, 44H), 0.88 (t, J = 7.2 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 163.17, 160.82, 142.45, 137.13, 134.40, 127.40, 119.29, 114.50, 105.78, 101.22, 77.16, 71.14, 68.25, 32.04, 29.79, 29.77, 29.73, 29.71, 29.52, 29.48, 29.34, 26.16, 22.82, 14.26. **MALDI-TOF MS** (MW+K): 666.73 g/mol.

Synthesis of compound 25



Mol. formula: $C_{43}H_{64}N_2O_{10}$

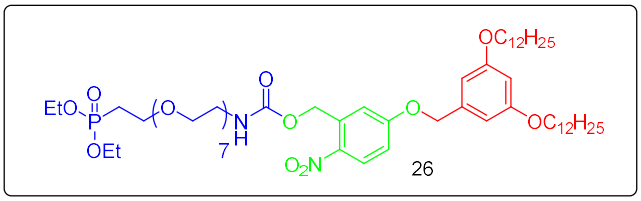
Mol. weight: 768.46 g/mol

Physical appearance: pale yellow solid

Yield: 65 %

In an oven dried RBF, compound **24** (0.31 g, 0.49 mmol) and *N,N'*-DSC (0.63 g, 2.4 mmol) were dissolved in ACN under stirring. Then, Et_3N (0.23 g, 2.4 mmol) was then added slowly at RT and allowed to react for 12 hrs. Upon completion of reaction, ACN and Et_3N were evaporated under vacuum. The obtained residue was directly purified using NPC using EtOAc / PE as eluent to get pale yellow solid (0.25 g, 0.32 mmol, 65 %), R_f = 0.18 in 25 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 8.15 (d, *J* = 8.9 Hz, 1H), 7.40 (dd, *J* = 8.9 Hz, 2.8 Hz, 1H), 7.18 (d, *J* = 2.8 Hz, 1H), 6.52 (d, *J* = 2.4 Hz, 2H), 6.42 (t, *J* = 2.4 Hz, 1H), 5.20 (s, 2H), 5.12 (s, 2H), 3.93 (t, *J* = 6.4 Hz, 4H), 2.40 (s, 4H), 1.85-1.68 (m, 4H), 1.48-1.26 (m, 44H), 0.88 (t, *J* = 7.2 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 168.55, 163.64, 160.78, 151.47, 139.58, 137.57, 133.63, 128.31, 115.07, 112.90, 105.93, 101.29, 70.98, 69.32, 68.25, 32.05, 29.80, 29.77, 29.74, 29.72, 29.54, 29.48, 29.37, 26.18, 25.60, 22.82, 14.26. **MALDI-TOF MS** (MW+K): 807.73 g/mol.

Synthesis of compound 26



26

Mol. formula: $C_{59}H_{102}NO_{18}P$

Mol. weight: 1143.68 g/mol

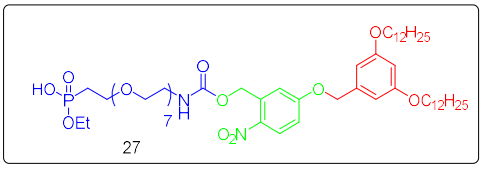
Physical appearance: pale yellow liquid

Yield: 65 %

In an oven dried RBF, compound **22** (0.10 g, 0.13 mmol) and compound **25** (0.06 g, 0.11 mmol) were dissolved in DMF under stirring. Then, Et_3N (0.014 g, 13 mmol) was added slowly to the reaction mixture and stirred at RT for 12 hrs. Upon completion of reaction,

DMF and Et_3N were evaporated under vacuum. To the obtained residue water was added and extracted thrice with DCM. Combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get crude product which was purified by NPC using MeOH / DCM as eluent to get pale yellow liquid (0.09 g, 0.07 mmol, 65 %), R_f = 0.40 in 5% MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 8.07 (d, J = 8.9 Hz, 1H), 7.10 (dd, J = 8.9, 2.8 Hz, 1H), 6.84 (d, J = 2.8 Hz, 1H), 6.47 (d, J = 2.4 Hz, 2H), 6.37 (t, J = 2.4 Hz, 1H), 5.47 (s, 2H), 4.99 (s, 2H), 4.07-3.98 (m, 4H), 3.87 (t, J = 6.4 Hz, 4H), 3.69-3.47 (m, 28H), 3.39-3.29 (m, 2H), 2.10-2.00 (m, 2H), 1.78-1.63 (m, 4H), 1.45-1.12 (m, 47H), 0.81 (t, J = 7.2 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 162.98, 160.56, 155.86, 139.98, 137.52, 136.89, 127.80, 114.27, 112.98, 105.81, 100.85, 77.16, 68.01, 65.01, 63.31, 61.57, 61.51, 40.93, 31.83, 29.58, 29.56, 29.52, 29.51, 29.32, 29.27, 29.16, 27.55, 26.17, 25.98, 22.61, 16.32, 14.06. **MALDI-TOF MS** (MW+K): 1182.63 g/mol.

Synthesis of compound 27



27

Mol. formula: $C_{57}H_{98}NO_{18}P$

Mol. Weight: 1115.65 g/mol

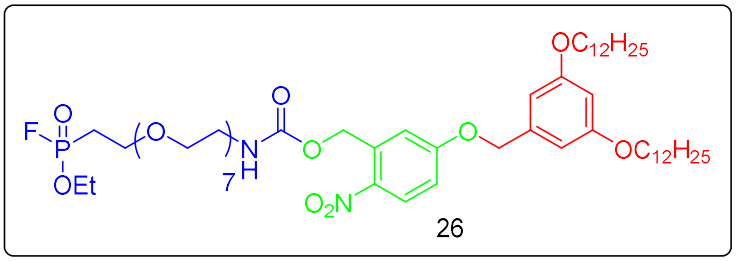
Physical appearance: pale yellow liquid

Yield: Not obtained

In an oven dried RBF, compound **26** (0.070 g, 0.06 mmol) was taken and LiBr (0.050 g, 0.6 mmol) was added. To the mixture, DMF was added and heated at 95 ºC for 20 hrs. Upon completion, water was added and extracted thrice with EtOAc . The water layer was collected and 2 N HCl was added and stirred for another 30 mins. The mixture was then extracted thrice with EtOAc. Combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get crude product which was used without purification. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 8.14 (d, J = 8.9 Hz, 1H), 7.14 (s, 1H), 6.87 (dd, J = 8.9, 2.8 Hz, 1H), 6.51 (d, J = 2.8 Hz, 1H), 6.39 (d, J = 2.4 Hz, 2H), 5.51 (s, 2H), 5.03 (s, 2H), 4.11-3.99 (m, 2H), 3.90 (t, J

= 6.4 Hz, 4H), 3.77-3.49 (m, 28H), 3.43-3.33 (m, 2H), 2.16-2.05 (m, 2H), 1.80-1.67 (m, 4H), 1.47-1.14 (m, 47H), 0.85 (t, *J* = 7.2 Hz, 6H).

Synthesis of compound 26



26

Mol. formula: $C_{57}H_{97}FNO_{17}P$

Mol. weight: 1117.65 g/mol

Physical appearance: pale yellow liquid

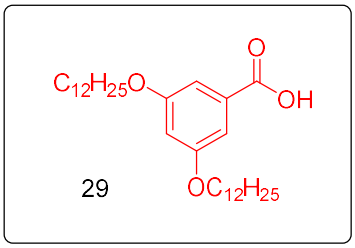
Yield: Not obtained

To the stirring solution of compound **27** (0.04 g, 0.03 mmol) in DCM, DAST (0.024 g, 0.14 mmol) was added dropwise at -78 ºC and allowed to react for 15 mins. Upon completion of the reaction, excess of DAST and DCM were evaporated under vacuum. To the obtained residue, water was added and stirred for 2 more mins to quench any residual DAST. The reaction mixture was then extracted thrice with DCM. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (MW+Na): 1140.63 g/mol.

AABPs in chapter 3.2 and their intermediates

Synthesis of redox-responsive AABP and its intermediates

Synthesis of compound 29



29

Mol. formula: $C_{31}H_{54}O_4$

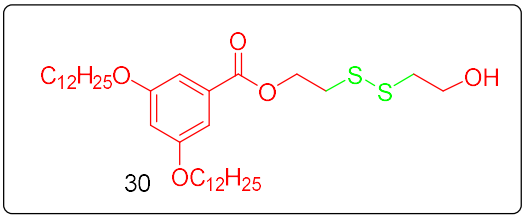
Mol. Weight: 490.76 g/mol

Physical appearance: White solid

Yield: 60 %

The compound (**7e**) (3 g, 5 mmol), sodium hydroxide (NaOH) (0.5 g, 10 mmol) were dissolved in EtOH. The mixture was then refluxed for 4 hrs. Upon completion, the reaction was quenched with dropwise addition of water and acidified with conc. HCl. The obtained precipitate was filtered and washed with EtOH for two times. The combined organic layer dried over Na_2SO_4 and concentrated under reduced pressure to get the crude product, which was utilized for the next reaction without further purification. The product was obtained as white solid (1.7 g, 3 mmol, 60 %), Rf = 0.10 in 10 % EtOAc /PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.21 (d, *J* = 2.4 Hz, 2H), 6.68 (t, *J* = 2.4 Hz, 1H), 3.98 (t, *J* = 6.4 Hz, 2H), 1.85-1.73 (m, 4H), 1.52-1.17 (m, 38H), 0.88 (t, *J* = 6.8 Hz ,6H).

Synthesis of compound 30



30

Mol. formula: $C_{35}H_{62}O_5S_2$

Mol. Weight: 626.99 g/mol

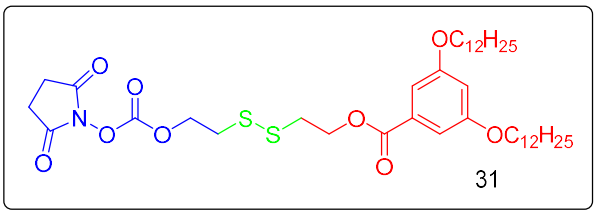
Physical appearance: White solid

Yield: 70 %

In an oven dried RBF **compound 29** (2.5 g, 5 mmol), 2,2'-disulfanediylbis(ethan-1-ol) (1.9 g, 10 mmol) and DMAP (0.3 g, 2 mmol) were taken and dissolved in DCM under stirring. Then, solution of EDC in DCM (2 g, 10 mmol) was added slowly to above mixture and allowed to stir for 12 hrs. Upon completion, reaction was quenched with water and resulting content was extracted in DCM thrice. Combined organic layers was dried over Na_2SO_4 and concentrated under reduced pressure to get crude product which was purified using silica gel column

chromatography. The product (**30**) obtained was white solid (2.2 g, 3 mmol, 70 %), Rf = 0.20 in 10 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.16 (d, *J* = 2.4 Hz, 2H), 6.64 (t, *J* = 2.4 Hz, 1H), 4.57 (t, *J* = 6.4 Hz, 2H), 3.96 (t, *J* = 6.4 Hz, 2H), 3.91-9.87 (m, 2H),3.05 (t, *J* = 6.8 Hz, 2H), 2.90 (t, *J* = 6.8 Hz, 2H), 1.84-1.73 (m, 4H), 1.52-1.26 (m, 38H), 0.88 (t, *J* = 6.8 Hz, 6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 166.48, 160.30, 131.58, 107.88, 106.76, 77.16, 68.47, 63.12,60.33, 41.81, 37.12, 32.05, 29.79, 29.77, 29.73, 29.71, 29.49, 29.48, 29.30, 26.14, 22.82, 14.25**. MALDI-TOF MS** (M+Na): 665.47 g/mol.

Synthesis of compound 31



Mol. formula: $C_{40}H_{65}NO_9S_2$

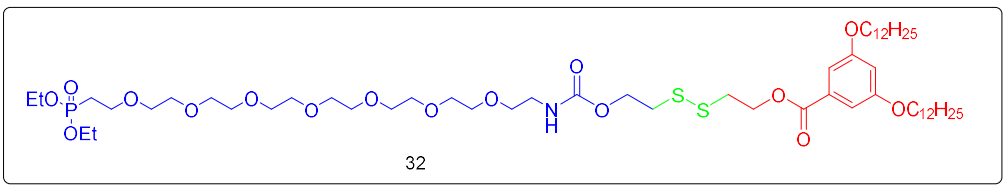
Mol. Weight: 768.07 g/mol

Physical appearance: White solid

Yield: 65 %

The **compound 31** was prepared from **compound 30** (0.5 g, 0.8 mmol) and N,N′-DSC (1.6 g, 6 mmol) and Et_3N (0.6 g, 5 mmol). The product obtained was pale yellow solid (0.4 g, 0.5 mmol, 65 %), Rf = 0.18 in 25 % EtOAc / PE. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.14 (d, *J* = 2.4 Hz, 2H), 6.63 (t, *J* = 2.4 Hz, 1H), 4.59-4.54 (m, 4H), 3.96 (t, *J* = 6.4 Hz, 2H), 3.08-3.01 (m, 4H), 2.82(s, 4H), 1.80-1.73 (m, 4H), 1.47-1.26 (m, 38H), 0.87 (t, *J* = 6.8 Hz ,6H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 168.60, 166.33, 160.28, 151.54, 131.60, 107.85, 106.75, 68.84, 68.45, 62.96, 37.32, 36.39, 32.03, 29.78, 29.75, 29.72, 29.69, 29.49, 29.47, 29.29, 26.13, 25.56, 22.80, 14.24.

Synthesis of compound 32



Mol. formula: $C_{56}H_{104}NO_{10}PS_2$

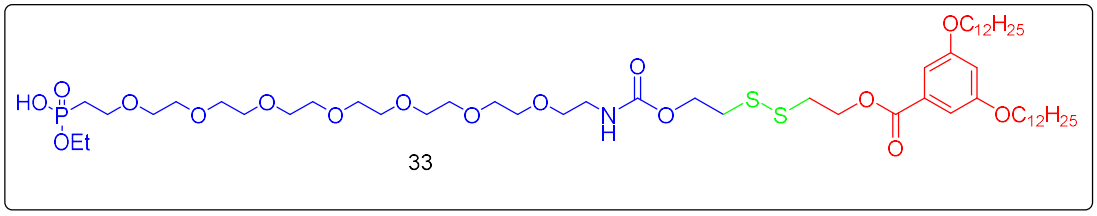
Mol. Weight: 1142.53 g/mol

Physical appearance: White solid

Yield: 65 %

The **compound 32** was prepared from **compound 31** (0.10 g, 0.13 mmol) and **compound 22** (0.06 g, 0.11 mmol). The product obtained was pale yellow liquid (0.09 g, 0.07 mmol, 65 %), Rf = 0.40 in 5 % MeOH / DCM. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.14 (d, *J* = 2.4 Hz, 2H), 6.62 (t, *J* = 2.4 Hz, 1H), 4.54 (t, *J* = 6.8 Hz, 2H), 4.30 (m, 2H), 4.15-4.03 (m, 4H), 3.95 (t, *J* = 6.4 Hz, 2H), 3.77-3.47(m, 30H), 3.03 (t, *J* = 6.4 Hz, 2H), 2.93 (t, *J* = 6.8 Hz, 2H), 2.20-2.02 (m, 2H), 1.79-1.72 (m, 4H), 1.51-1.18(m, 46H), 0.86 (t, *J* = 6.8 Hz ,6H).

Synthesis of compound 33



33

Mol. formula: $C_{54}H_{100}NO_{16}PS_2$

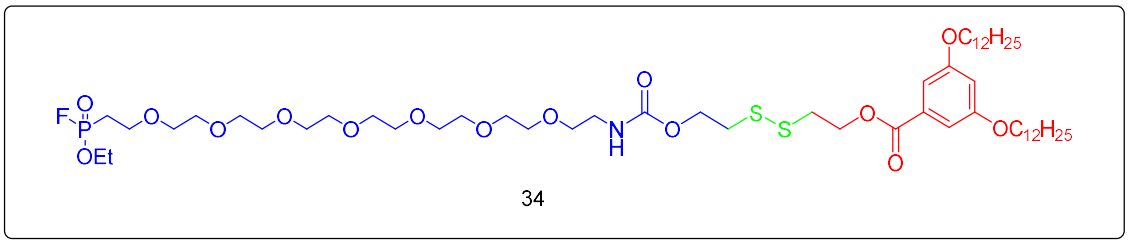
Mol. Weight: 1114.47 g/mol

Physical appearance: White solid

Yield: Not determined

The **compound 33** was prepared from compound **32** (0.070 g, 0.06 mmol) and LiBr (0.050 g, 0.6 mmol). The obtained get crude product, which was used without purification.

Synthesis of compound 34



34

Mol. formula: $C_{54}H_{99}FNO_{15}PS_2$

Mol. Weight: 1116.47 g/mol

Physical appearance: White solid

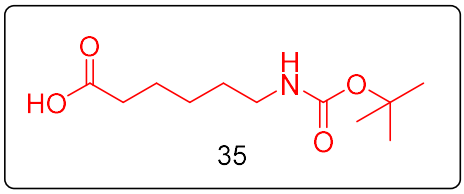
Yield: Not determined

The compound **34** was prepared from compound **33** (0.04 g, 0.03 mmol), DAST (0.024 g, 0.14 mmol. The obtained get crude product, which was used without purification. **^{19}F NMR** (400 MHz, $CDCl_3$): $^{\delta}$F -59.91, -62.74.

AABPs in chapter 3.3 and their intermediates

Synthesis of hydrophobic peptides (polyamides) and their intermediates

Synthesis of compound 35



35

Mol. formula: $C_{11}H_{21}NO_4$

Mol. Weight: 231.20 g/mol

Physical appearance: White solid

Yield: 91 %

In an oven-dried RBF, a mixture of 6-aminocaproic acid (5 gm, 38 mmol) and Boc anhydride (10g, 45mmol) was taken and dissolved in THF : H_2O (2:1) at 0 ºC under stirring. Subsequently, Na_2CO_3 (6 g, 57 mmol) solution was added to the above reaction mixture in small portions, and then the reaction mixture was stirred for 12 hrs at RT. Upon completion, excess of Na_2CO_3 was neutralized with 1 N HCl, and the resultant mixture was extracted with EtOAc for thrice. The combined organic layer was dried over Na_2SO_4 and concentrated under vacuum to get the crude product, which was utilized in the next step without further purification (8 g, 34 mmol).

Synthesis of compound 36



Mol. formula: $C_7H_{15}NO_2$

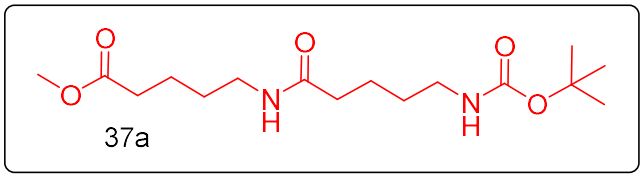
Mol. Weight: 145.20 g/mol

Physical appearance: White solid

Yield: 89 %

To the suspension of 6-aminocaproic acid (5 gm, 38 mmol) in MeOH, $SOCl_2$ (18 g, 114 mmol) was added dropwise at 0 ºC under a nitrogen atmosphere. Then, the reaction mixture was refluxed for 12 hrs. Upon completion, the solvent and excess $SOCl_2$ was removed under vacuum. Further, the solid residue was washed with dry diethyl ether. The obtained solid residue was used in the next step without any further purification (4.9 g, 34 mmol). **^{1}H NMR** (400 MHz, $CDCl_3$): $\delta_{H\,5}$ 3.73(s, 3H), 3.04(t, *J* = 8 Hz, 2H), 2.47(t, *J* = 8 Hz, 2H), 1.76-1.65 (m, 4H), 1.50-1.40(m, 2H).

Synthesis of compound 37a



Mol. formula: $C_{18}H_{34}N_2O_5$

Mol. Weight: 358.47 g/mol

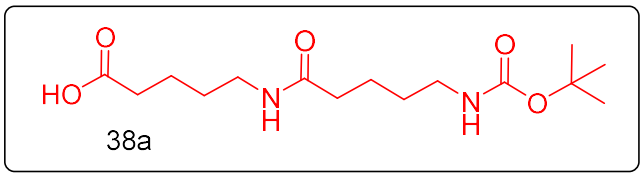
Physical appearance: White solid

Yield: 89 %

The **compound 37a** was prepared from **compound 35** (4.7g, 21 mmol), HOBT (3.3 g, 21 mmol), HBTU (8.3 g, 21 mmol), DIPEA (9.5 g, 73 mmol) and **compound 36** (3.8 g, 25 mmol) in DMF. The product was obtained as a white solid (5.4 g, 15 mmol). **^{1}H NMR** (400

MHz, $CDCl_3$): δ_H 3.64 (s, 3H), 3.24-3.19 (q, *J* = 8 Hz, 2H), 3.10-3.05(q, *J* = 4 Hz, 2H), 2.31-2.27 (t, *J* = 8 Hz, 2H), 2.15-2.12 (t, *J* = 8 Hz, 2H), 2.65-2.58 (qt, *J* = 8 Hz, 4H), 2.52-2.45 (qt, *J* = 8 Hz, 4H), 1.41(s, 9H), 1.36-1.23 (m, 4H). **MALDI-TOF MS** (MW+K):397.20 g/mol.

Synthesis of compound 38a



38a

Mol. formula: $C_{17}H_{32}N_2O_5$

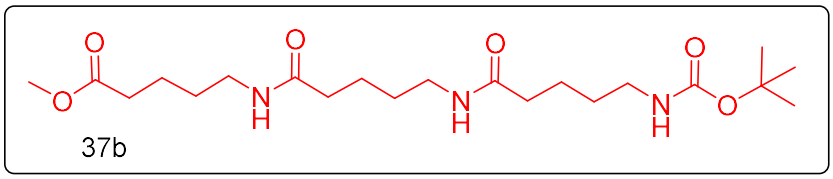
Mol. Weight: 344.45 g/mol

Physical appearance: White solid

Yield: 93 %

The **compound 38a** was prepared starting from **37a** (1.6 g, 4.7 mmol), NaOH (0.5 g, 15 mmol) in MeOH. The product was obtained as a white solid (1.5 g, 4.5 mmol). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 3.24-3.19 (q, *J* = 8 Hz, 2H), 3.10-3.05 (q, *J* = 4 Hz, 2H), 2.31-2.27 (t, *J* = 8 Hz, 2H), 2.15-2.12(t, *J* = 8 Hz, 2H), 2.65-2.58 (qt, *J* = 8 Hz, 4H), 2.52-2.45 (qt, *J* = 8 Hz, 4H), 1.41(s, 9H), 1.36-1.23 (m,4H). **MALDI-TOF MS** (MW+K):382.42 g/mol.

Synthesis of compound 37b



37b

Mol. formula: $C_{24}H_{25}N_3O_6$

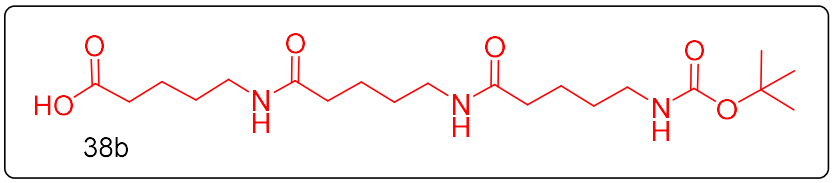
Mol. Weight: 471.63 g/mol

Physical appearance: White solid

Yield: 41 %

The compound **37b** was prepared from **36** (0.434 g, 1.21 mmol), HOBT (0.19 g, 1.21 mmol), HBTU (0.47g, 1.21mmol), DIPEA (0.56 g, 4.2 mmol), and **38a** (0.21 g, 1.34 mmol) in DMF. The product was obtained as a white solid (0.24 g, 0.50 mmol). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 3.66 (s, 3H), 3.27-3.22 (q, *J* = 8 Hz, 4H), 3.12-3.07 (q, *J* = 4 Hz, 2H), 2.33-2.29 (t, *J* = 8 Hz, 4H), 2.18-2.14 (t, *J* = 8 Hz, 2H), 1.68.-1.60 (qt, *J* = 8 Hz, 13H), 1.54-1.47 (qt, *J* = 8 Hz, 6H), 1.41(s, 9H), 1.36-1.23 (m, 6H). **MALDI-TOF MS** (MW+Na): 510.23 g/mol.

Synthesis of compound 38b



38b

Mol. formula: $C_{23}H_{43}N_3O_6$

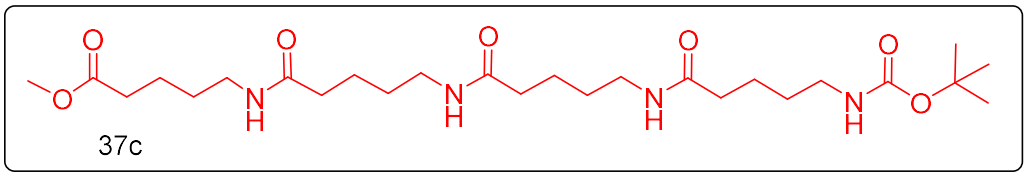
Mol. Weight: 457.61 g/mol

Physical appearance: White solid

Yield: 92 %

The **compound 38b** was from **compound 37b** (0.19 g, 0.42 mmol), NaOH (0.05 g, 1.27 mmol) in MeOH. The product was obtained as a white solid (0.17 g, 0.37 mmol). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 3.27-3.22 (q, *J* = 8 Hz, 4H), 3.12-3.07(q, *J* = 4 Hz, 2H), 2.33-2.29 (t, *J* = 8 Hz, 4H), 2.18-2.14 (t, *J* = 8 Hz, 2H), 1.68.-1.60 (qt, *J* = 8 Hz, 13H), 1.54-1.47 (qt, *J* = 8 Hz, 6H), 1.41(s, 9H), 1.36-1.23(m,6H). **MALDI-TOF MS** (MW+K): 496.24 g/mol.

Synthesis of compound 37c



37c

Mol. formula: $C_{30}H_{56}N_4O_7$

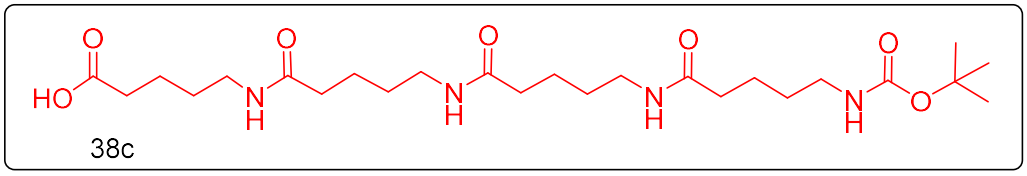
Mol. Weight: 584.79 g/mol

Physical appearance: Pale yellow liquid

Yield: 42 %

The compound **37c** was prepared from **36** (0.461 g, 1 mmol), HOBT (0.13 g, 1 mmol), HBTU (0.33 g, 1 mmol), DIPEA (0.4 g, 4 mmol), and **38b** (0.15 g, 0.35 mmol) in DMF. The product was obtained as a pale yellow liquid (0.21 g, 15 mmol). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 3.66 (s, 3H), 3.24-3.19 (q, *J* = 8 Hz, 6H), 3.10-3.05 (q, *J* = 4 Hz, 2H), 2.31-2.27 (t, *J* = 8 Hz, 6H), 2.15-2.12 (t, *J* = 8 Hz, 2H), 2.65-2.58(qt, *J* = 8 Hz, 13H), 2.52-2.45 (qt, *J* = 8 Hz, 8H), 1.41(s, 9H), 1.36-1.23 (m, 9H). **MALDI-TOF MS** (MW+K):622.25 g/mol.

Synthesis of compound 38c



Mol. formula: $C_{29}H_{54}N_4O_7$

Mol. Weight: 570.77 g/mol

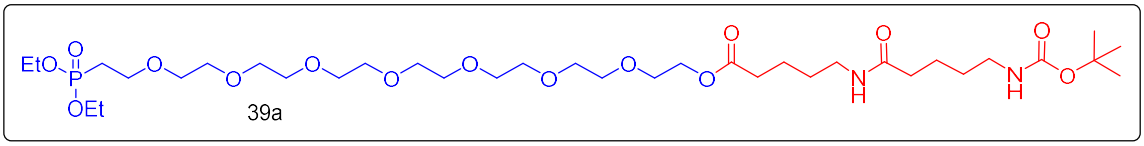
Physical appearance: White solid

Yield: 73 %

The compound **38c** was prepared from **37c** (0.103 g, 0.17 mmol), NaOH (0.05 g, 0.51 mmol) in MeOH. The product was obtained as a white solid (1.5g, 4.5 mmol).

Synthesis of peptide-based AABPs and their intermediates

Synthesis of compound 39a



Mol. formula: $C_{37}H_{73}N_2O_{15}P$

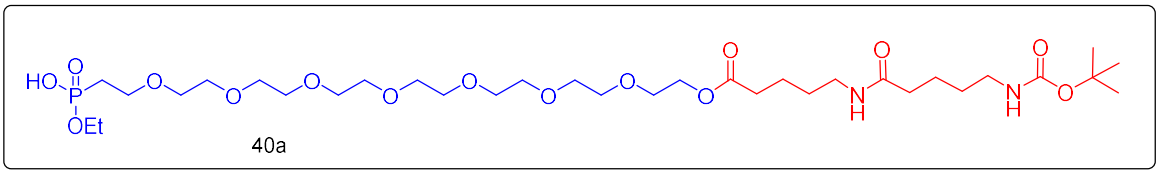
Mol. Weight: 816.96 g/mol

Physical appearance: Pale yellow liquid

Yield: 84 %

The **compound 39a** was prepared from **38a** (0.07 g, 0.20 mmol), alcohol (**19**) (0.1 g, 2.4 mmol), DCC (0.06 g, 0.3 mmol), DMAP (0.02 g, 0.2 mmol), in DCM. The product was obtained as a pale yellow liquid (0.14g, 0.18 mmol). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.22 (t, *J* = 8 Hz, 2H), 4.12-4.03(m, 4H), 3.75-3.59 (m, 28H), 3.25-3.20 (m, 2H), 3.09 (t, *J* = 8 Hz, 2H), 2.33(t, *J* = 4 Hz, 2H0, 2.16-2.03(m, 4H), 1.67-1.58 (qt, *J* = 8 Hz, 4H), 1.52-1.44 (qt, *J* = 8 Hz, 4H), 1.43 (s, 9H), 1.36-1.27 (m, 14H). **MALDI-TOF MS** (MW+K):855.25 g/mol.

Synthesis of compound 40a



40a

Mol. formula: $C_{35}H_{69}N_2O_{15}P$

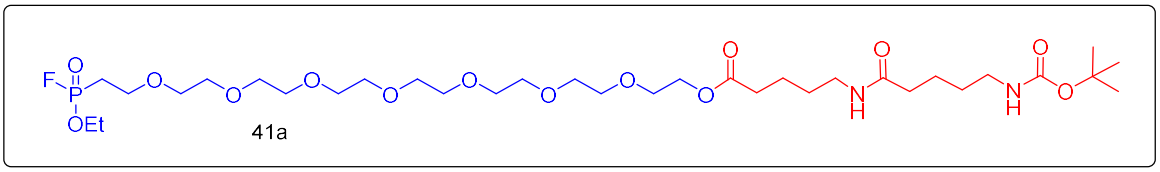
Mol. Weight: 788.90 g/mol

Physical appearance: Pale yellow liquid

Yield: Not obtained

The compound **40a** was prepared from **39a** (0.04 g, 0.05 mmol), LiBr (0.23 g, 2.4 mmol), in DMF. The product was obtained as a pale yellow liquid. **MALDI-TOF MS** (MW+K): 817.36 g/mol.

Synthesis of compound 41a



41a

Mol. formula: $C_{35}H_{69}FN_2O_{14}P$

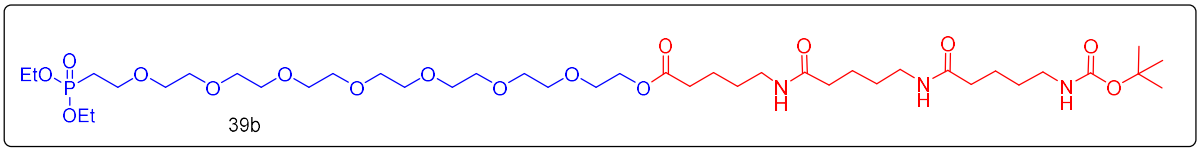
Mol. Weight: 790.90 g/mol

Physical appearance: Pale Yellow Liquid

Yield: Not obtained

The **compound 41a** was prepared from **40a** (0.01 g, 0.012 mmol), and DAST (0.03 g, 0.2 mmol), in DMF. The product was obtained as a pale yellow liquid. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (MW+K): 829.62 g/mol.

Synthesis of compound 39b



Mol. formula: $C_{37}H_{73}N_2O_{15}P$

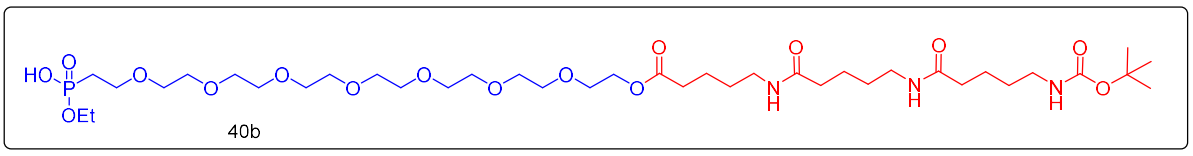
Mol. Weight: 929.29 g/mol

Physical appearance: Pale Yellow Liquid

Yield: 62 %

The **compound 39b** was prepared from **38b** (0.14 g, 0.32 mmol), alcohol **(compound 19)** (0.15 g, 0.4mmol), DCC (0.08 g, 0.3 mmol), DMAP (0.04 g, 0.2 mmol) in DCM. The product was obtained as a pale yellow liquid (0.15 g, 0.12 mmol). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 4.22 (t, J = 8 Hz, 2H), 4.12-4.03 (m, 4H), 3.75-3.59(m, 28H), 3.27-3.22(m, 4H), 3.09 (t, J = 8 Hz, 2H), 2.33 (t, J = 4 Hz, 2H), 2.16-2.03 (m, 6H), 1.67-1.58 (qt, J = 8 Hz, 4H), 1.52-1.44 (qt, J = 8 Hz, 4H), 1.43(s, 9H), 1.36-1.27(m, 14H). **MALDI-TOF** MS (MW+K): 968.52 g/mol.

Synthesis of compound 40b



Mol. formula: $C_{41}H_{80}N_3O_{16}P$

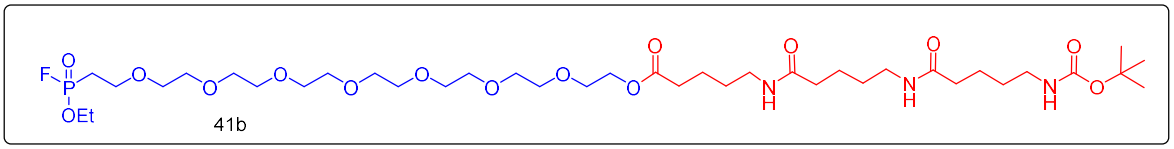
Mol. Weight: 902.62 g/mol

Physical appearance: Pale yellow liquid

Yield: Not obtained

The **compound 40b** was prepared from **compound 39b** (0.06 g, 0.06 mmol), LiBr (0.13 g, 1.6 mmol), in DMF. The product was obtained as a pale yellow liquid. **MALDI-TOF MS** (M+K): 940.23 g/mol.

Synthesis of compound 41b



Mol. formula: $C41H_{80}FN_2O_{14}P$

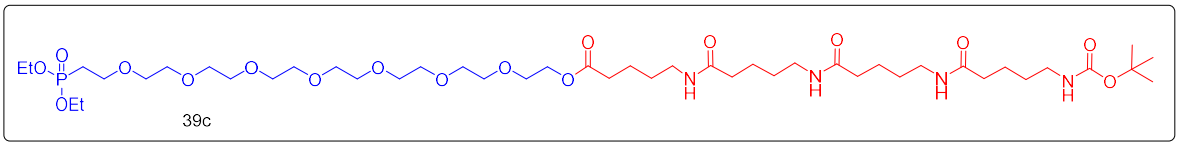
Mol. Weight: 904.06 g/mol

Physical appearance: Pale yellow liquid

Yield: Not obtained

The **compound 41b** was prepared from **40b** (0.03 g, 0.03 mmol), DAST (0.02 g, 0.12 mmol), in DMF. The product was obtained as a pale yellow liquid. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (MW+K): 943.15 g/mol.

Synthesis of compound 39c



Mol. formula: $C_{50}H_{97}N_4O_{17}P$

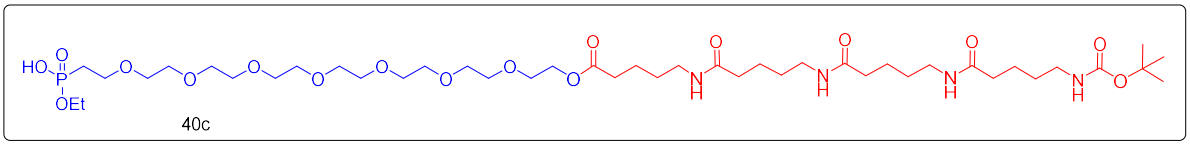
Mol. Weight: 1108.61 g/mol

Physical appearance: Pale yellow liquid

Yield: not obtained

The **compound 39c** was prepared from compound **38c** (0.7 g, 1 mmol), alcohol **(19)** (0.7 g, 1 mmol), DCC (0.02 g, 1.5 mmol), DMAP (0.01 g, 1 mmol), in DCM. The product was obtained as a pale yellow liquid (0.7 g, 0.6 mmol). **MALDI-TOF MS** (MW+K):1131.25 g/mol.

Synthesis of compound 40c



40c

Mol. formula: $C_{48}H_{93}N_4O_{17}P$

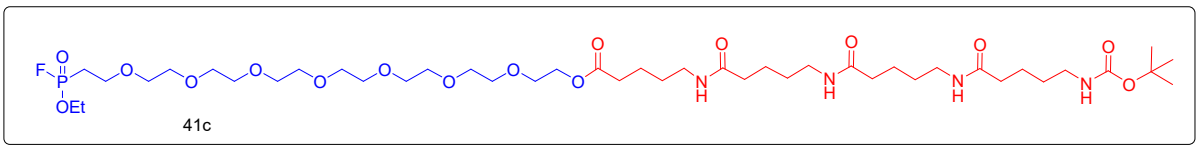
Mol. Weight: 1079.25 g/mol

Physical appearance: Pale yellow liquid

Yield: Not obtained

The **compound 40c** was prepared from **39c** (0.1 g, 0.08 mmol), LiBr (0.07 g, 0.84 mmol), in DMF. The product was obtained as a pale yellow liquid. **MALDI-TOF MS** (MW+K):1118.41 g/mol.

Synthesis of compound 41c



41c

Mol. formula: $C_{48}H_{92}FN_4O_{17}P$

Mol. Weight: 1081.90 g/mol

Physical appearance: Pale yellow liquid

Yield: Not obtained

The compound **41c** was prepared from **40c** (0.06 g, 0.050 mol), and DAST (0.04 g, 0.24 mmol), in DCM. The product was obtained as a pale yellow liquid. **^{19}F NMR** (400 MHz, $CDCl_3$): δ_F -59.91, -62.74. **MALDI-TOF MS** (MW+K): 1120.32 g/mol.

References

1 Bowman, G. D. & Poirier, M. G. Post-translational modifications of histones that influence nucleosome dynamics. *Chem. Rev.* **2014**, 115, 2274-2295.

2 Cornish, V. W., Mendel, D. & Schultz, P. G. Probing protein structure and function with an expanded genetic code. *Angew. Chem. Int. Edn. Engl.* **1995**, 34, 621-633.

3 Mills, J. H. *et al.* Computational design of a homotrimeric metalloprotein with a trisbipyridyl core. *Proc. Natl Acad. Sci. USA* **2016**, 201600188.

4 Lavigueur, C. *et al.* Thermoresponsive giant biohybrid amphiphiles. *Polymer chemistry* **2011**, 2, 333-340.

5 Xia, Y., Tang, S. & Olsen, B. D. Site-specific conjugation of RAFT polymers to proteins via expressed protein ligation. *Chem. Commun* **2013**, 49, 2566-2568.

6 Wan, X. & Liu, S. Fabrication of a thermoresponsive biohybrid double hydrophilic block copolymer by a cofactor reconstitution approach. *Macromol. Rapid Commun.* **2010**, 31, 2070-2076.

7 Moatsou, D. *et al.* Self-assembly of temperature-responsive protein–polymer bioconjugates. *Bioconjugate Chem.* **2015**, 26, 1890-1899.

8 Averick, S. *et al.* Cooperative, Reversible Self-Assembly of Covalently Pre-Linked Proteins into Giant Fibrous Structures. *Angew. Chem. Int. Edn. Engl.* **2014**, 53, 8050-8055.

9 Suma, T. *et al.* Generalizable Strategy for Engineering Protein Particles with pH-Triggered Disassembly and Recoverable Protein Functionality. *ACS Macro Lett.* **2015**, 4, 160-164.

10 Lee, M. H. *et al.* Disulfide-cleavage-triggered chemosensors and their biological applications. *Chem. Rev.* **2013**, 113, 5071-5109.

11 Levy, E. D., Erba, E. B., Robinson, C. V. & Teichmann, S. A. Assembly reflects evolution of protein complexes. *Nature* **2008**, 453, 1262-1265.

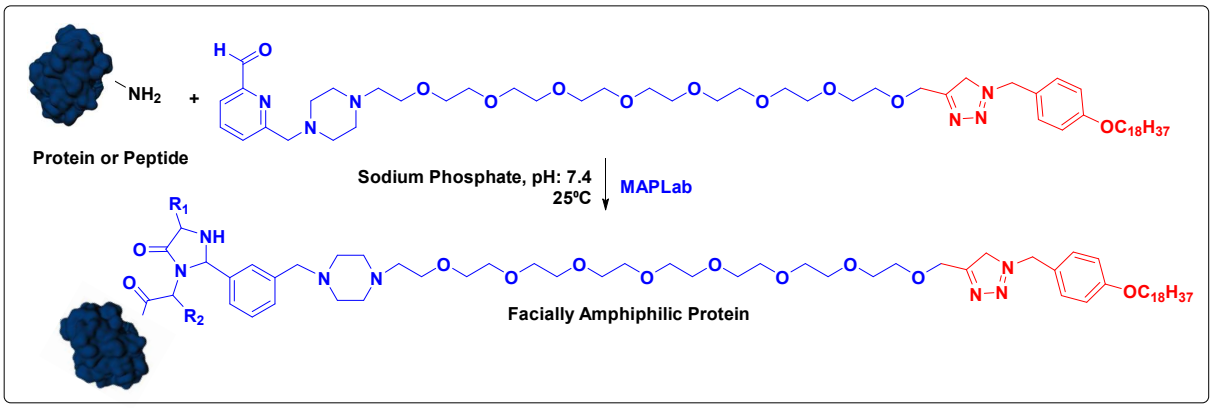
12 Sandanaraj, B. S., Reddy, M. M., Bhandari, P. J., Kumar, S. & Aswal, V. K. Rational design of supramolecular dynamic protein assemblies by using a micelle-assisted activity-based protein-labeling technology. *Chem. - Eur. J.* **2018,** 24, 16085-16096.

13 Sandanaraj, B. S., Reddy, M. M., Rao, K. J. & Bhandari, P. J. Rational design of semi-synthetic protein c with the defined oligomeric state. *ChemistrySelect* **2019,** 4, 6397-6402.

Chapter 4

Extending MAPLab Technology to N-terminus and Thiol Labelling

4.1. Targeting N-terminus for Universal Methodology



4.1.1. Design

The Cravatt group pioneered, activity-based protein profiling (ABPP) provides a powerful approach for the monitoring of enzyme functional states in complex biological systems. Activity-based probes (ABPs) with specific warheads exclusively react with active enzymes and have been developed for most of the enzyme classes like hydrolases, kinases, phosphatases, glycosidases, histone deacetylases, and oxidoreductases[1]. Most of the warheads have been developed originally as irreversible inhibitors. We have utilized such a warhead FP group in the previous two chapters for protein engineering applications for the first time.[2,3] Work in the previous chapters demonstrates the creation of a protein complex from monomeric protein by converting it into FAP, which then self-assembles into a protein complex. This was achieved using AABPs and MAPLab technology. But the methodology was restricted only to serine proteases (about 200 proteins in humans) because of the chemistry we choose. This chapter is about developing a new methodology that can be used for converting any protein/peptide of interest in general.

To be used extensively in protein engineering, the protein labeling reaction needs to be compatible with proteins and should be efficient enough to proceed at ambient temperatures and aqueous conditions. The reaction kinetics should be sufficient enough to proceed in uM or less concentrations of proteins. A wide variety of chemical reactions have been developed to specifically target the side chains of amino acids. Among the 20 canonical amino acids, Lys, Glu, Asp, and Cys are the most common targets for direct chemical modification. But, modifying a single site on the protein surface by targeting amino acids like Lys, Asp, and Glu is extremely difficult because of the high abundance and the high solvent accessibility of these amino acids. Lysine is one of the most common targets for protein bioconjugation reactions. Owing to high natural abundance on protein surfaces, a major challenge in targeting lysine is the control of the modification number. Heterogeneous products with multiple modifications can be controlled to some extent by limiting the equivalents of reagent reaction, but this may lead to a drastic decrease in modification yields. Alternatively, cysteine is a great target for site-selective modification because of low its low abundance, which is utilized in the next subchapter.

On the other hand, Noncanonical / Artificial amino acids in protein engineering are becoming increasingly common. Unlike the natural amino acids, these unnatural amino acids carry a bioorthogonal functional group, and groups are not typically present in biological systems and do not react with natural functionalities. Incorporating these exogenous functional groups into a target protein enables unpresidential site-selective modification, but the only issue is that these need to be introduced on the proteins in the very first step. Enzyme-mediated protein modifications are also being explored equally to the chemical protein modifications. Under mild conditions, enzyme-mediated bioconjugation has the ability to modify specific molecular tags. Some of the most important enzymes are sortase A, N-Myristoyltransferase, Subtiligase, Butelase 1, microbial transglutaminase.

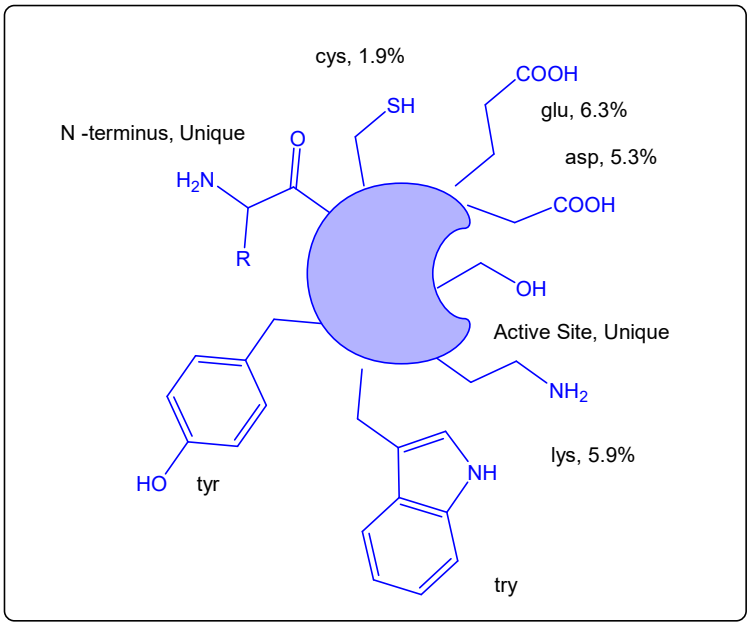


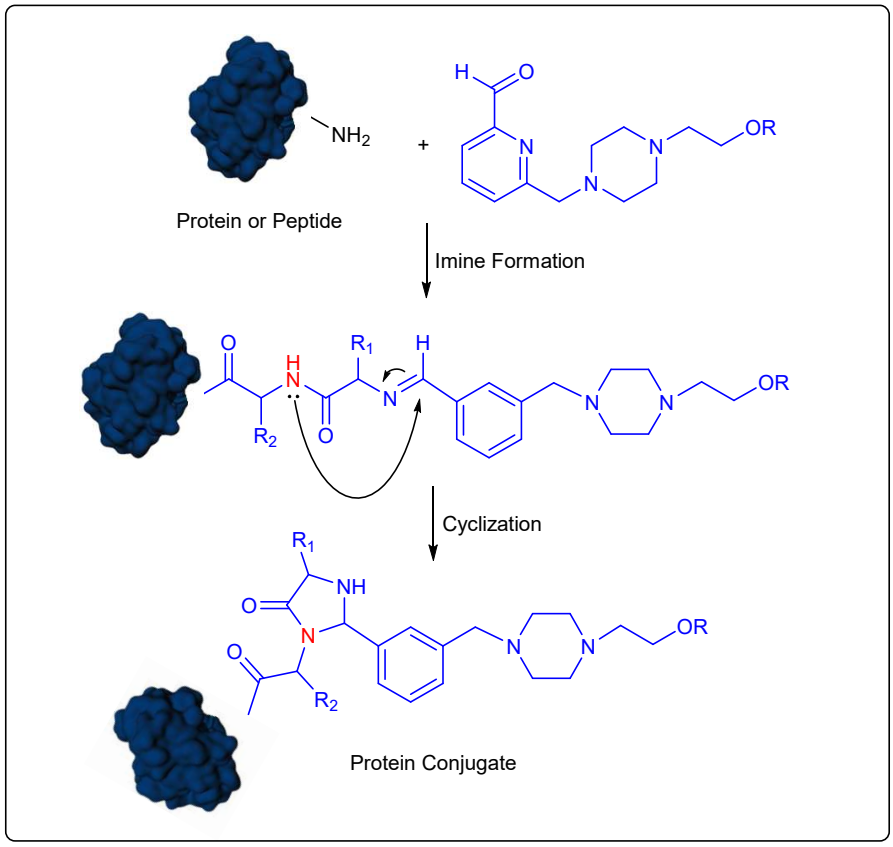
Figure 4.1. Relative abundance and uniqueness of amino acids.

Apart from natural and artificial amino acids, C-terminus presents a very good opportunity for site-specific modification. Very few strategies are present for the modification of C terminus. But in contrast to few examples for C-terminal modification, chemical strategies for targetting N-terminus are plenty. Just as C-terminus, N-terminus stands out as a unique site as single-chain proteins have only one N-terminus. While amine groups are abundant in proteins, the "α-amine of the N terminus" stands out as a uniquely reactive site as single-chain proteins have only one N

terminus. As a result, an increasing number of site-specific modification strategies are now targeting this position for applications in chemical and synthetic biology.[4]

The present sub-chapter discloses a novel method for the synthesis of protein complexes by site-specific modification of N-terminal amino acid (Fig. 4.1). This protein modification is, in principle, universal, and hence any protein/peptide can be converted into amphiphilic proteins/peptides. Until now, several chemical methods are used to modify N-terminal amino acids directly or convert them into unique functional groups for further ligations. N-terminus-specific enzymatic ligation strategies are also emerging as powerful strategies. One-step site-selective modification of protein N-termini that does not require any genetic engineering of the protein was reported recently[5]. This reaction is demonstrated for 12 different proteins, including the soluble domain of the human estrogen receptor.

Mechanism



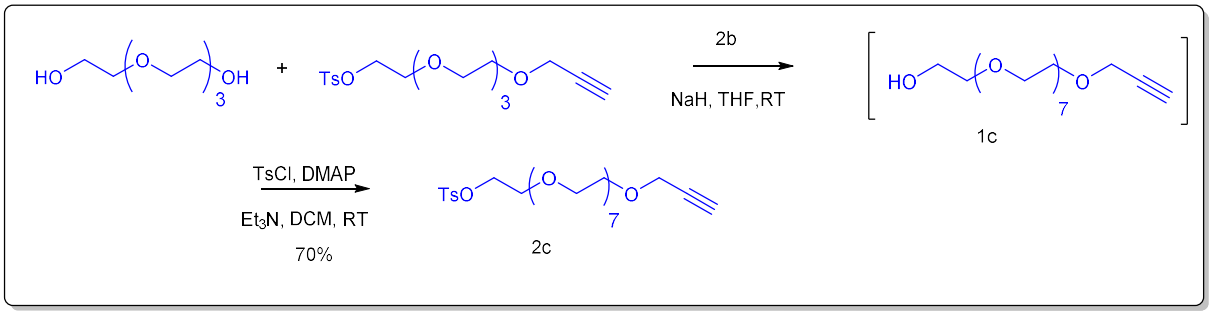
Scheme 4.4. Conjugation mechanism at the N-terminus with 2-PCA.

The method is specific to the N terminus in the first step of protein modification, because of the following mechanism. The reactive 'α-amine at the N-terminus' of the protein /peptide has pKa value ranging from 6-8, and because of this increased availability of deprotonated α-amine groups, imine formation happens between the α-amino group at N-terminus and probe/AABP, and then the nucleophilic attack of the neighboring amide nitrogen happens, on the initially formed imine leading to the stable conjugate as shown below[2].

4.1.2. Synthesis

We have slightly changed the design of AABP in this sub-chapter. In the place of the FP group, we have used of 2-pyridine carboxaldehyde (2-PCA) for targeting N-terminus. The synthesis is discussed below. Based on our previous chapter, we choose the OEG spacer and C18-1T for this N-terminus AABP.

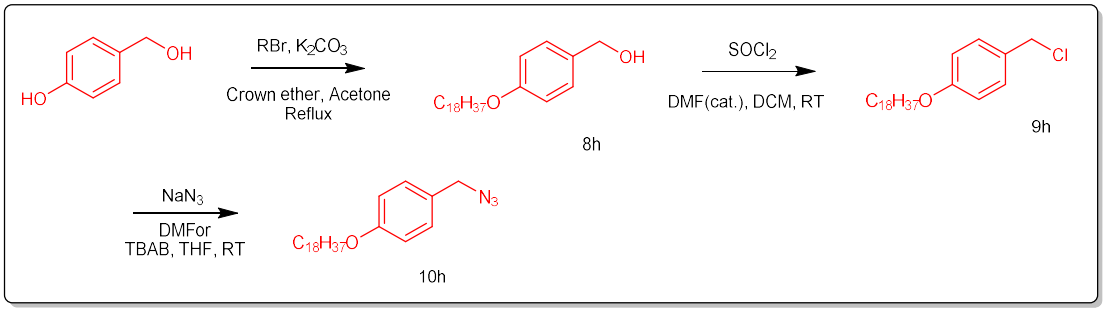
Synthesis of hydrophilic spacer



Scheme 4.1. Synthesis of the OEG hydrophilic spacer.

As discussed in chapter 2.1, the required tosylate (**2c**) was made in the following steps. Synthesis of the hydrophilic spacer (**1c**) was done from tetra ethylene glycol (TEG) and the tosylate (**2b**) in the presence of NaH to get mono propargyl derivative (**1c**). This monopropargylate was tosylated in the presence of KOH to get tosylate (**2c**).

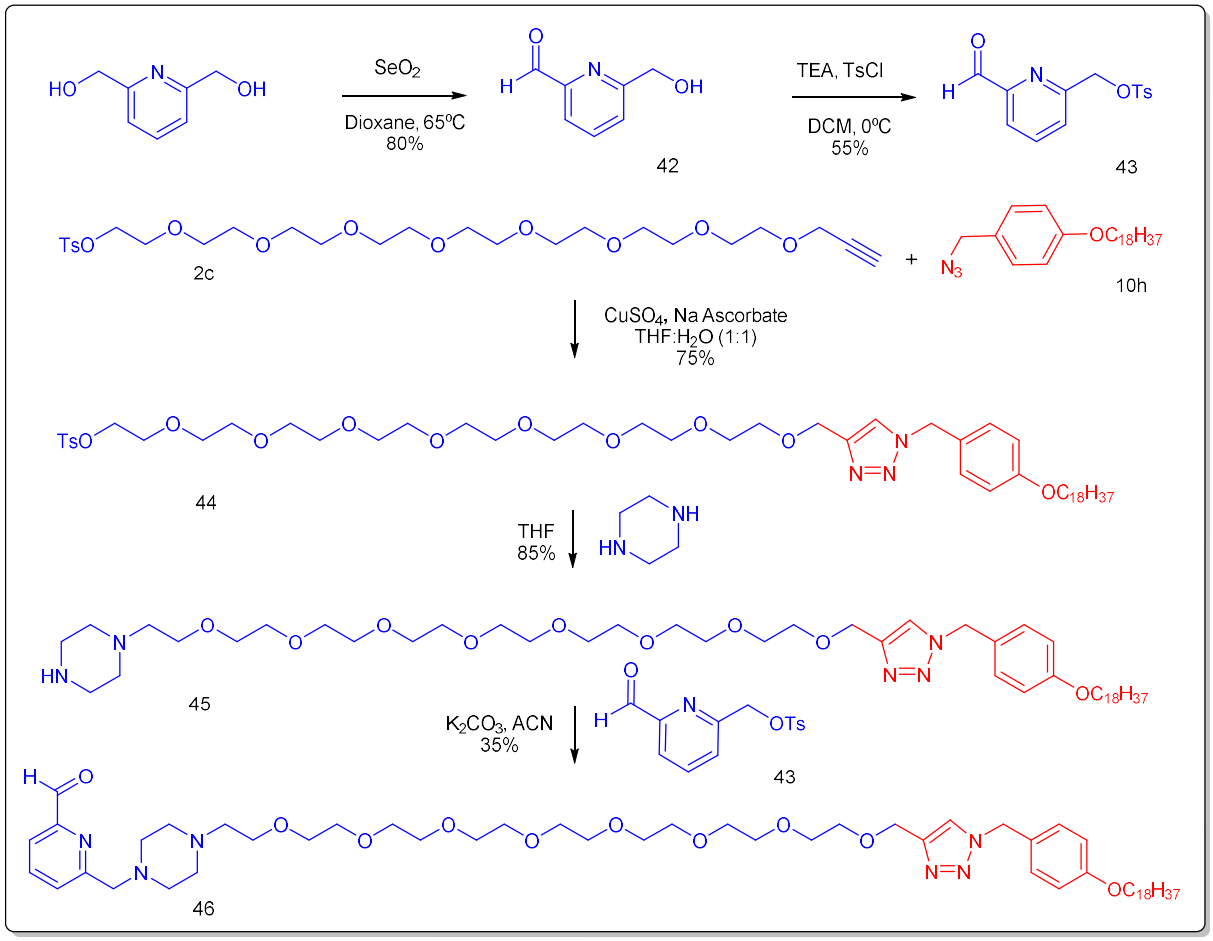
Synthesis of hydrophobic tail



Scheme 4.2. Synthesis of C181T hydrophobic tail.

The synthesis of hydrophobic azide (**10h**) was discussed in chapter 2.4. To brief, 4-(hydroxymethyl)phenol was alkylated in the presence of RBr, K_2CO_3, to get the alkylated product (**8h**). This was then converted into a chloride (**9h**) using $SOCl_2$ in the presence of DMF. Finally above obtained chloride (**9h**) (1.0 eq) in the presence of NaN_3 (1.5 eq) was converted into azide (**10h**).

Synthesis of AABP

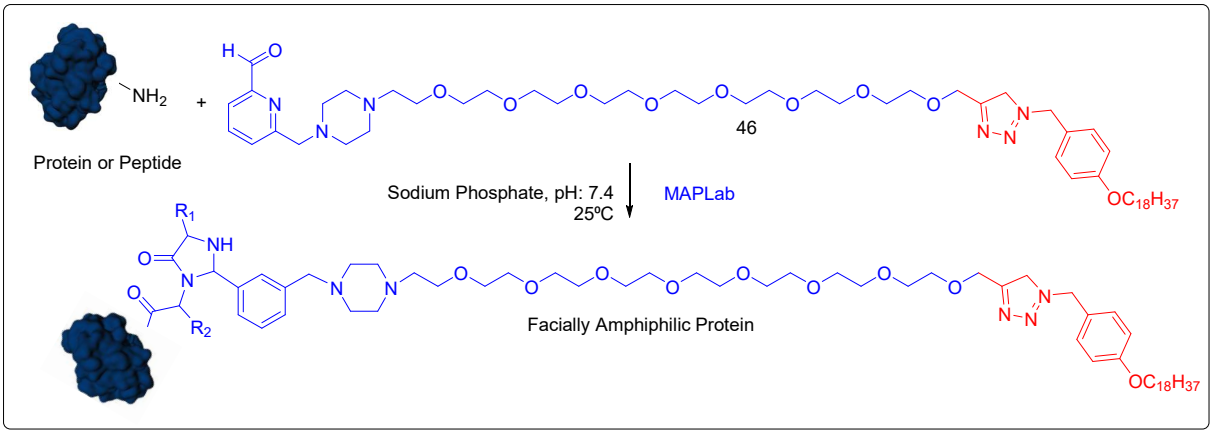


Scheme 4.3. Synthesis of AABP for targeting N-terminus.

To a solution of 2,6-pyridinedimethanol in 1,4-dioxane, selenium dioxide (SeO_2) was added. The resulting mixture was sonicated for 5 mins and then stirred at 65 ºC for 24 hrs. Then the reaction was cooled to room temperature and diluted with DCM. The mixture was filtered through celite, and the filtrate was concentrated under reduced pressure. The resulting crude material was purified by NPC using PE / EtOAc to get a liquid that became off-white solid later (**42**). TsCl was added to the solution of alcohol (**42**) in DCM at 0 ºC. Et_3N was then added to the above mixture, maintaining the temperature at 0 ºC. After 1 hr at RT, the reaction mixture was concentrated and purified using NPC using PE / EtOAc to get tosylate product (**43**). Hydrophilic alkyne (**2c**) (1 eq) and hydrophobic azide (**10h**) (1 eq) were dissolved in degassed THF and

stirred until a clear solution was obtained, then degassed H_2O was added and stirred vigorously for 10 more mins. Freshly prepared 1M Na Asc. (0.05 eq) and 1M $CuSO_4$ (0.1 eq) was added to the reaction mixture thrice in intervals of 45 mins and allowed to react for 16 hrs at RT. Upon completion, reaction mixture was extracted with DCM and combined organic layer was dried over Na_2SO_4 and concentrated under reduced pressure to get crude product, which was purified by NPC using MeOH / DCM solvent system to get clicked product (**44**). Tosylate (**44**) (1 eq) and piperazine (10 eq) were dissolved in THF in an RBF. Then the mixture was refluxed at 65 ºC. After 12 hrs, the reaction mixture was concentrated and purified by NPC using PE / DCM. Compound **45** (1.5 eq)**,** tosylate (**43**) (1 eq) and K_2CO_3 (1.6 eq) were weighed in RBF. Then the mixture was dissolved in ACN and refluxed at 65 ºC. After 16 hrs, the reaction mixture was concentrated and purified by NPC using PE / DCM to get the final compound (**46**).

4.1.3. Conjugation and purification



Scheme 4.4. Conjugation of the amphiphilic probe to the protein of interest at the N-terminus.

Proteins were weighed (1.4 to 6.6 mg, depending on protein) in microcentrifuge tubes, and 500 µL 50 mM sodium phosphate pH 7.4 was added and mixed gently with a pipette to make 200 µM solutions. Then AABP (50 or 100 eq) were weighed in a different microcentrifuge tube, followed by the addition of 2 µL triton X-100 and 480 µL of 50 mM sodium phosphate pH 7.4 and vortexed for 15 mins. When the AABP solution becomes clear, the protein solution was added into AABP solution to get 100 µM (1 mL) protein solutions and allowed to react for 24 hrs on rotospin at 20 rpm at 25 ºC. We saw more than 60-90 % conversions with 50-100 equivalents with a test scale (Fig. 4.2). We followed a different MALDI-TOF Ms monitoring in

this chapter (refer to the methods section). The AABP was used at 50 or 100 eq, which was just 2 eq in previous chapters. Triton X-100, which was used at concentrations 100 times (20 mM) more than CMC or 2% of the total volume of the reaction mixture to solubilize the AABPs in the previous chapter, gave noisy MALDI-TOF Spectra. So we decreased the Triton X-100 concentration to 10 times more than CMC. Heating the reaction mixture at 37ºC triggered precipitation of protein conjugates in the reaction mixtures, so we did the conjugation at RT even with decreased yields. Protein modifications were carried out in falcon tubes at 200 mg scale following the linear scale-up of the above-mentioned procedure for understanding the self-assembly behavior. We have modified only 4 proteins using this technology, but in principle, the method can be extended to a vast number of proteins.

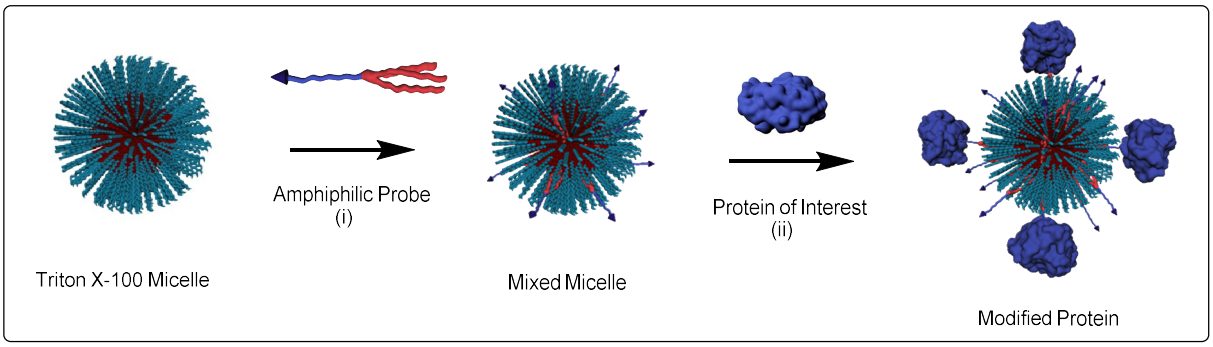


Figure 4.2. MAPLab technology: Conjugation steps and mechanisms.

The molecular weight changes can be seen in the reaction mixtures MALDI-TOF MS shown here (Fig. 4.3, Fig. 4.4, Fig. 4.5 and Fig. 4.6). To monitor the extent of protein modification, the samples were directly withdrawn from the reaction mixture using a pipette and analyzed. In brief, the sample and matrix mixture was mixed at a 1:10 ratio. When the crystallization observed in centrifuge tubes, 1-2µL of the samples was spotted on the MALDI-ToF MS plate and air-dried for 15 minutes. Then the plate was loaded and fired to get accurate molecular weights. Please refer to the methods section for the matrix preparation and sample preparation procedures.

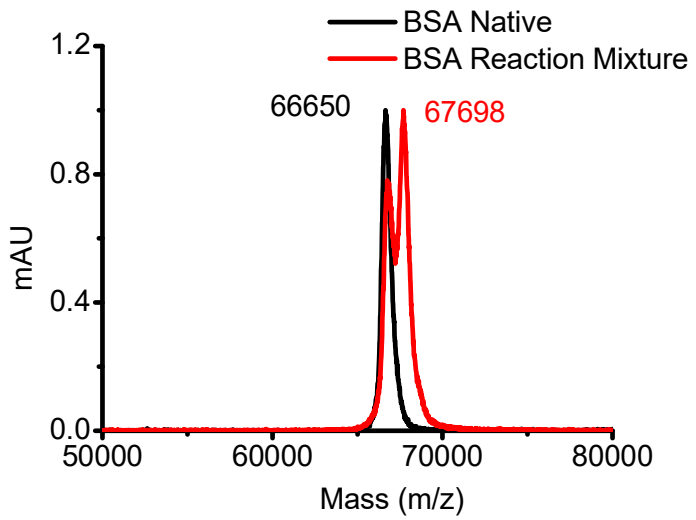


Figure 4.3. MALDI-TOF MS data of native and reaction mixtures of BSA.

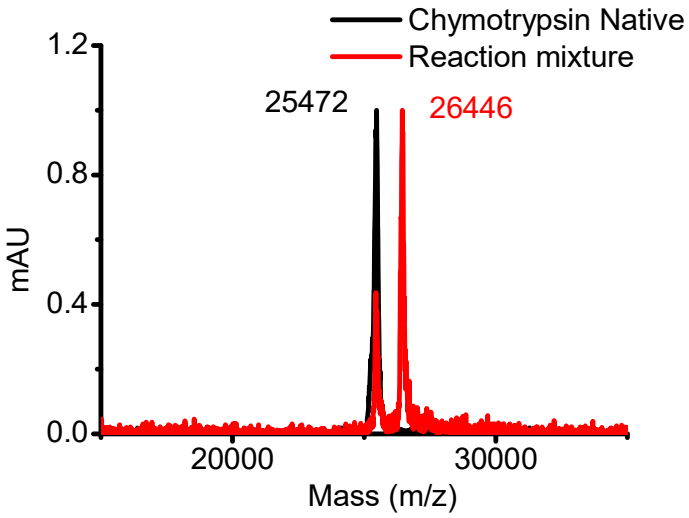


Figure 4.4. MALDI-TOF MS data of native and reaction mixtures of Chy.

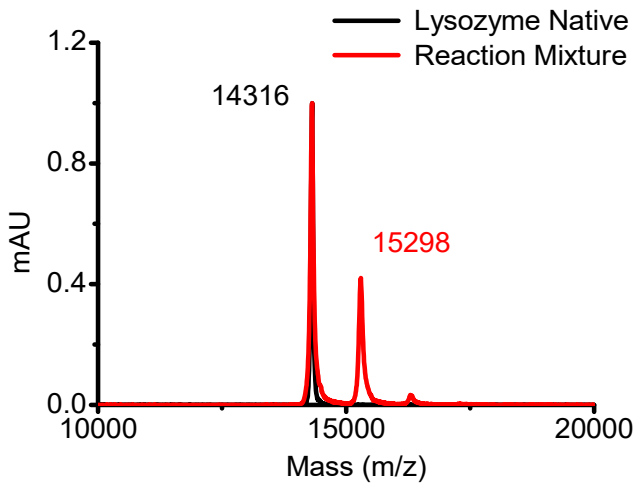


Figure 4.5. MALDI-TOF MS data of native and reaction mixtures of LZ.

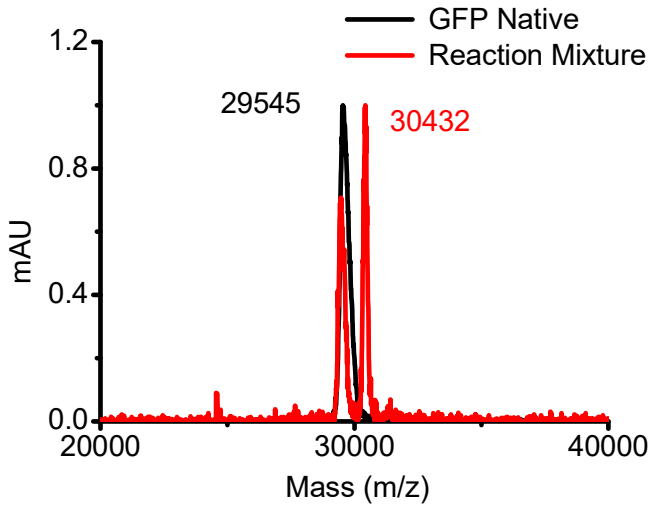


Figure 4.6. MALDI-TOF MS data of native and reaction mixtures of GFP.

All the conjugates were purified by two-step purification, *i.e.,* IEX and SEC, performed using FPLC. IEX was performed to remove triton X-100 using either SP sepharose or Q sepharose resins (GE) depending on isoelectric point (pI) and surface charges of proteins. For example, to purify the reaction mixture of LZ (pI: 11.35) and Chy (pI: 8.75) we used SP sepharose, a cation-exchange resin at pH 7.4 and to purify BSA (pI: 4.7) and GFP is (pI: 5.80) we used Q sepharose, an anion-exchange resin at pH 7.4. The obtained IEX fractions were subjected to SEC immediately to remove the native proteins from the protein conjugates in 50 mM sodium

phosphate pH 7.4, 1 M NaCl using either Sephacryl S-100 HR 16/60 or Sephacryl S-200 HR 16/60 or Sephacryl S-300 HR 16/60. After SEC the samples were stored at -80 ºC

4.1.4. Characterization

Before understanding the self-assembly behavior of the conjugates, we characterized the samples using MALDI-TOF MS. The detailed procedures for characterization are given in the experimental section. Soon after characterization, we started self-assembly studies of these conjugates. The MW obtained for Chy-OEG-C18-1T is in agreement with the calcd value (Fig. 4.7). BSA, on the other hand, showed ambiguity (double labeling like peaks) (Fig. 4.8), which could be because of know non-specific association of hydrophobic binding pockets on BSA with hydrophobic molecules and in this case with AABP-OEG-C18-1T.

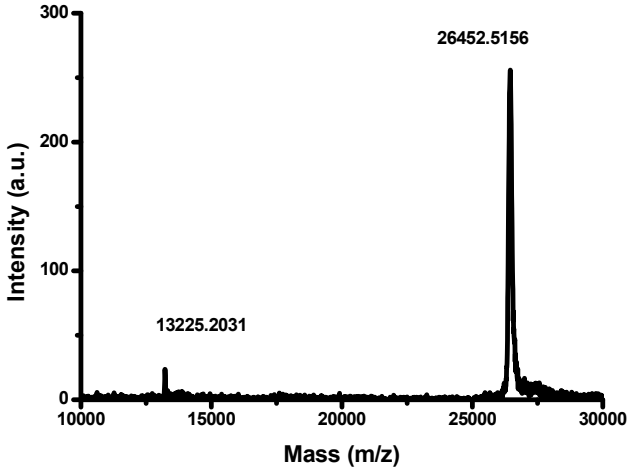


Figure 4.7. MALDI-TOF MS spectrum of purified Chy conjugate.

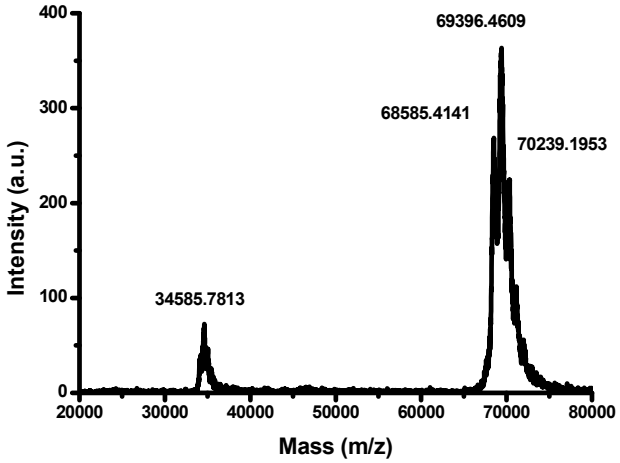
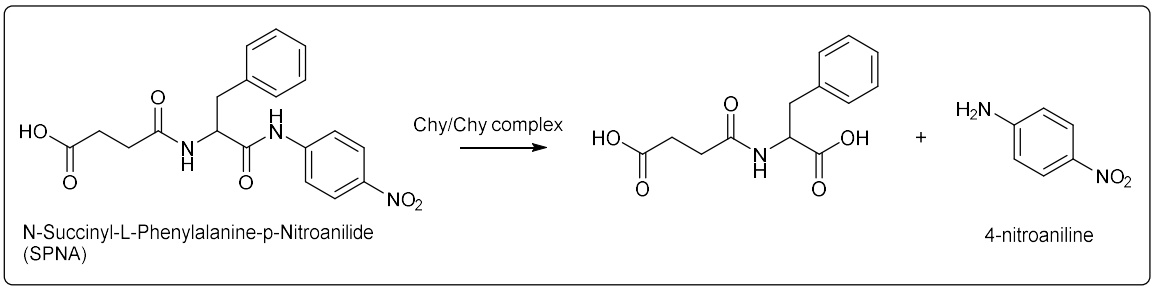


Figure 4.8. MALDI-TOF MS spectrum of purified BSA conjugate.

Chy-OEG-C181T activity assay

Unlike previous chapters, 2 and 3, N-terminus conjugation retains other functionalities/amino acids. So, Chy-OEG-C181T conjugate should retain its enzymatic activity. To compare the enzymatic activity of this conjugate with respect to the native Chy, we have performed hydrolysis of N-succinyl-Lphenylalanine-p-nitroanilide (SPNA)-based activity assay.



Scheme 4.5. Schematic representation of hydrolysis of SPNA.

The hydrolysis of SPNA results in the release of 4-nitroaniline, which absorption maximum is 410nm. Both native Chy and Chy-OEG-C181T were taken at 80 µM concentration in Sod. Phos. pH: 7.4 and SPNA was added to the solution at a final concentration of 4mM. Then the absorbance at 410 nm is measured for a period of 14 mins with a time interval of 2 mins. We found that the activity Chy-OEG-C181T is comparable to Chy native (Fig. 4.11). This is direct

evidence of active site intactness, and moreover, the active site location is accessible in the complex state as well (Fig. 4.9).

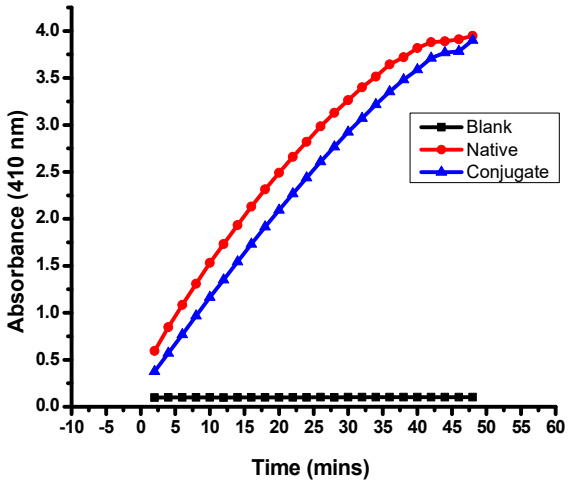


Figure 4.9. Activity of Chy Vs Chy-OEG-C181T conjugate.

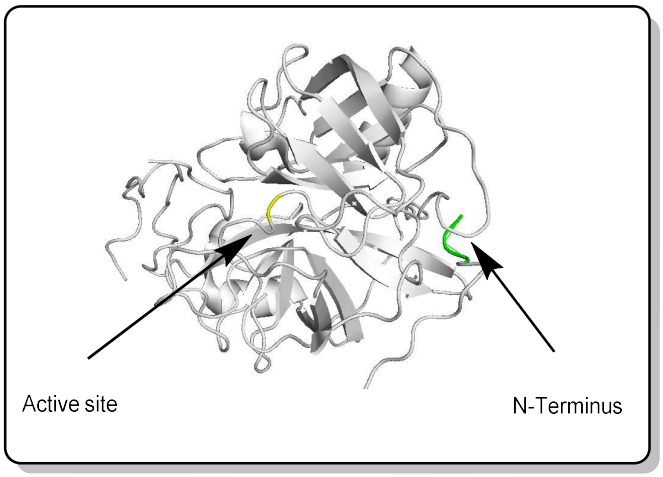


Figure 4.10. Chy (PDB accession: 1YPH) structure showing Active site Ser residue (yellow) with respect to N-terminus location (green).

4.1.5. Self-assembly studies

In order to test the self-assembling ability of the made FAPs in this chapter, detailed self-assembly studies were carried out using analytical SEC, dynamic light scattering (DLS). The detailed procedure for the same is given in the methods section at the end of this chapter. The SEC revealed that all the IEX purified conjugates could self-assemble to form higher-order complexes. BSA formed 700 kDa complex, and Chy formed a 400 kDa complex corresponding to the elution volumes of 13 mL and 14.6 mL, respectively. On the other hand, GFP formed 330 kDa complex, and surprisingly LZ showed 50 kDa complex. The RMWs were derived from the SEC calibration plot. Please refer to the methods section of this chapter (Fig. 4.12). The DLS studies were done for BSA and Chy, which is in agreement with SEC studies with D_h ranging from 11 - 14 nm (Fig. 4.12).

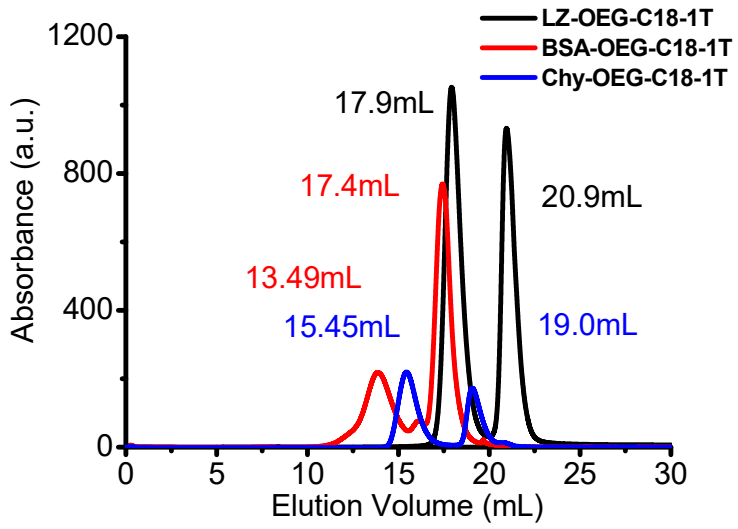


Figure 4.11. SEC data of IEX purified FAPs native and conjugate.

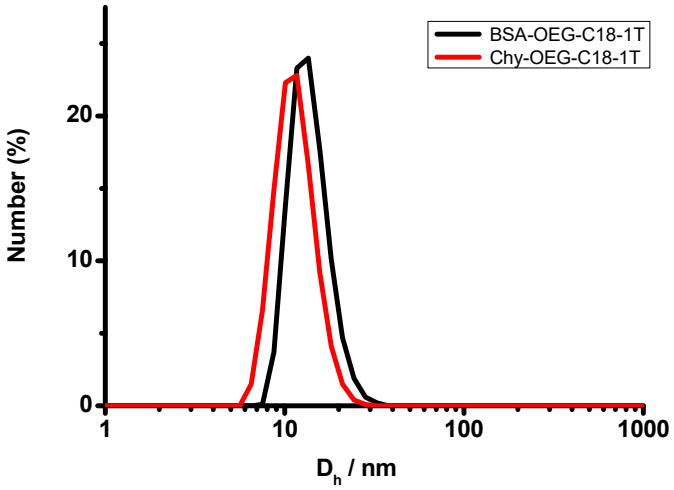


Figure 4.12. DLS data of SEC purified FAPs.

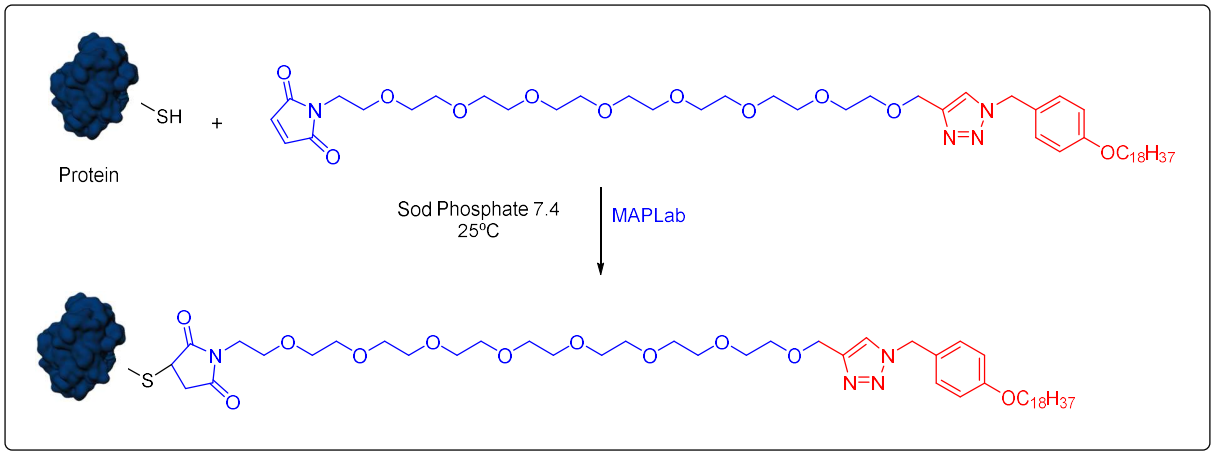
Table 4.1. Summary of self-assembly data of N-terminus FAPs (conjugates).

S.No	Protein Conjugate	Elution Vol SEC (mL)	Mol Wt SEC (kDa)	Oligomeric State (mer)	D_h DLS (nm)
1	BSA Nat	17.4	66	1	--
2	BSA-OEG-C18-1T	13	700	10	13.24
3	Chy Nat	19.0	25	1	--
4	Chy-OEG-C18-1T	14.6	400	16	11.73
5	GFP Nat	18.7	30	1	--
6	GFP-OEG-C18-1T	14	330	10	--
7	Lz Nat	20.9	14	1	--
8	Lz-OEG-C18-1T	17.9	50	3	--

4.1.6. Discussion

After exploring serine protease and FP chemistry, a new, universal method for converting any protein of interest into a bigger protein complex was explored in this chapter. The method is based on N-terminus labeling using the 2-PCA group reported by Francis and coworkers. In the AABP design, 2-PCA was incorporated in the place of the FP group, and the MAPLab was tried on four different proteins. MAPLab technology worked with this chemistry as well. We have successfully converted four proteins into FAPs using this method, and all the conjugates formed higher-order complexes when analyzed in the SEC. One interesting observation is that this chemistry is slow and requires higher equivalents. Heating the conjugation reaction mixtures increased the conversions as evident from MALDI-TOF MS analysis but resulted in precipitation of conjugates. So performing the conjugation at RT is the best condition in order to purify the conjugates without loss. Unlike previous chapters, N-terminus conjugation retains other functionalities/amino acids. When the activity was compared, Chy-OEG-C181T conjugate retained its enzymatic activity comparable to the activity of this native Chy.

4.2. Targeting Thiol for Universal Methodology



4.2.1. Design

The universal method presented in the previous chapter based on 2-PCA labeling is compatible with the majority of proteins but cannot be applied to proteins that are N-terminally acylated. It is also not applicable for proteins with Pro in the second position or those with N-termini that is not exposed to bulk solution[5]. To further increase the repertoire of chemistries that can be used for the creation of synthetic protein complexes, the present sub-chapter discloses a method for synthesis of protein amphiphiles/complexes by site-specific modification Cys amino acid (native or introduced). As mentioned in the previous sub-chapter, the Cys is a great target for site-selective modification because of low its low natural abundance. However, cysteine residues are often present as disulfide bonds that are crucial to the structural integrity of proteins, and so bioconjugation reactions may disrupt these bonds. The nucleophile side chain of cysteine can be a great target for site-selective protein modification. The relatively low abundance of cysteine facilitates the creation of well-defined bioconjugates. But cysteine residues are often part of disulfide bonds that are crucial to the folding and integrity of proteins, and so bioconjugation reactions, after reducing these disulfide bonds, may disrupt the structure of proteins. This means that the proteins with free Cys need to be targeted, and the proteins with free cysteine even rarer.

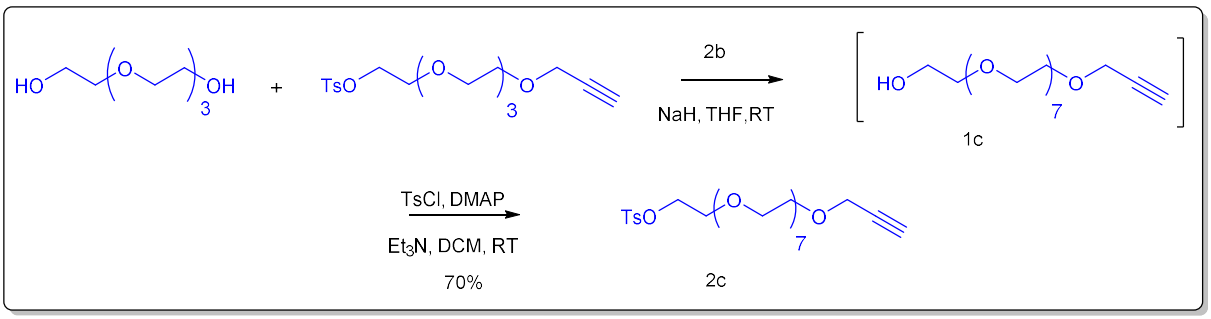
Nevertheless, most of the chemical protein engineering methods are based on Cys modification of BSA, and this particular combination played a key role in the development of chemical protein engineering strategies, for example, BSA polystyrene amphiphiles,[6] BSA-NIPAAm thermosensitive conjugates,[7,8] BSA-PNIPAAm proteinosomes[9] are some of the key reports in the literature. On the other hand, protein engineering reports based on Cys modification of other proteins are also present in the literature. For example, the pioneering work by Nolte (lipase-polystyrene amphiphile) was created using thiol-maleimide chemistry.[10] Olsen et al. reported myoglobin (Mb)-PNIPAM bioconjugates,[11] mCherry- PNIPAM conjugates,[12] thoroughly studied effect of small molecule osmolytes,[13] effects of polymer chemistry[14], and nature of protein interactions.[15] Thiol-maleimide based bioconjugation strategy was also used in conjunction with Expressed Protein Ligation (EPL) in protein engineering.[16]

Therefore this Cys modification is as good as a universal method, as genetic engineering and introduction of Cys are very common, and hence any protein/peptide can be converted into

amphiphilic proteins/peptides using this method as well. We have slightly changed the design of AABP in this sub-chapter, similar to the previous one, but in the place of the FP group or 2-PCA group, we have introduced the maleimide (MA) group for thiol targeting. Even though we know that the MA conjugate is prone to hydrolysis and better methods are known[17], we went with chemistry as proof of concept study. The synthesis is discussed below.

4.2.2. Synthesis

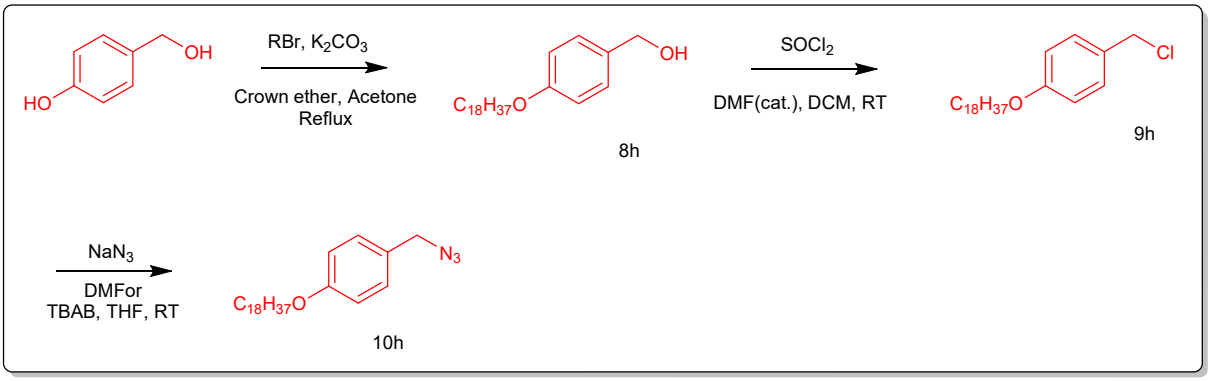
Synthesis of hydrophilic spacer



Scheme 4.6. Synthesis of the OEG hydrophilic spacer

As discussed in chapter 2.1, the required tosylate (**2c**) was made in the following steps. Synthesis of the hydrophilic spacer (**1c**) was done from tetra ethylene glycol (TEG) and the tosylate (**2b**) in the presence of NaH to get mono propargyl derivative (**1c**). This monopropargylate was tosylated in the presence of KOH to get tosylate (**2c**)

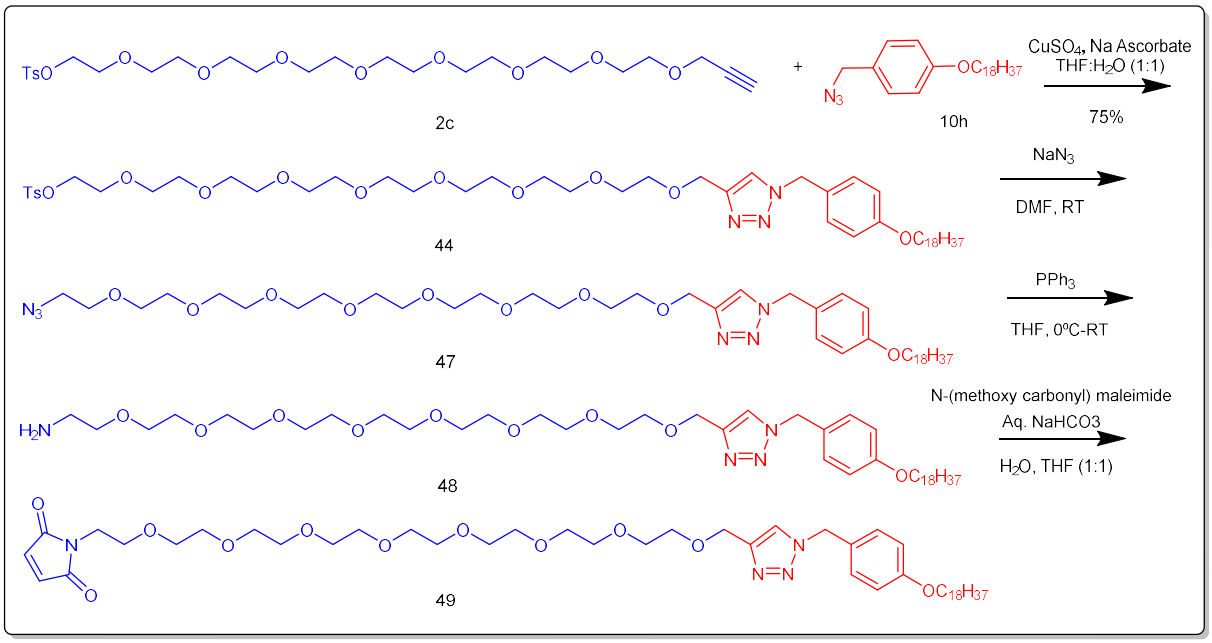
Synthesis of hydrophobic tail



Scheme 4.7. Synthesis of C181T hydrophobic tail

Synthesis of hydrophobic azide (**10h**) was discussed in chapter 2.4. To brief, 4-(hydroxymethyl)phenol was alkylated in the presence of RBr, K_2CO_3, to get the alkylated product (**8h**). This was then converted into a chloride (**9h**) using $SOCl_2$ in the presence of DMF. Finally above obtained chloride (**9h**) (1.0 eq) in the presence of NaN_3 (1.5 eq) was converted into azide (**10h**).

Synthesis of AABP

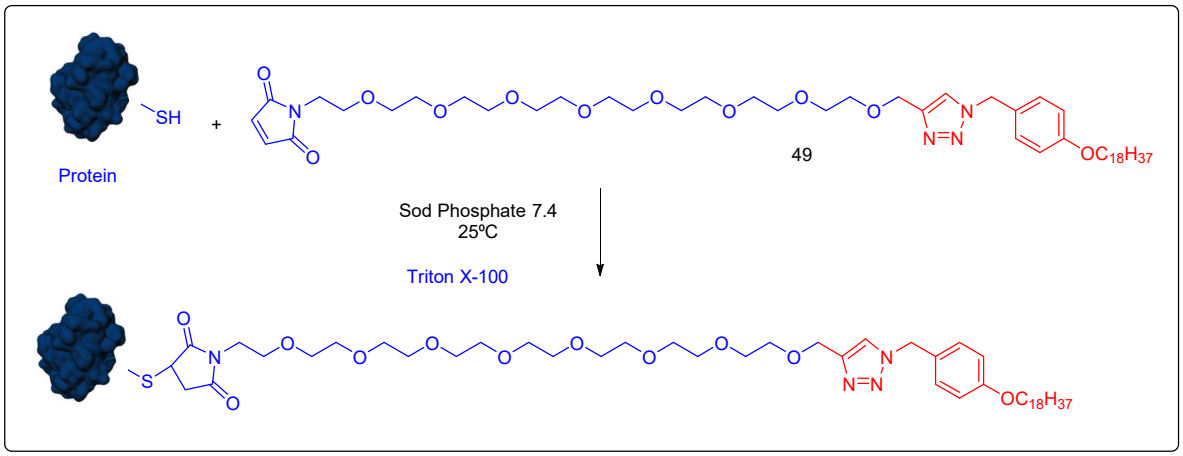


Scheme 4.8. Synthesis of the AABP for targeting thiol residue.

Hydrophilic alkyne (**2c**) (1 eq) and hydrophobic azide (**10h**) (1 eq) were dissolved in degassed THF and stirred until a clear solution was obtained, then degassed H_2O was added and stirred vigorously for 10 more mins. Freshly prepared 1M Na Asc. (0.05 eq) and 1M $CuSO_4$ (0.1 eq) was added to the reaction mixture thrice in intervals of 45 mins and allowed to react for 16 hrs at RT. Upon completion, the reaction mixture was extracted with DCM, and the combined organic layer was dried over Na_2SO_4 and concentrated under reduced pressure to get crude product which was purified by NPC using MeOH / DCM solvent system to get clicked product (**44**). The amphiphilic tosylate (**44**) was dissolved in DMF at RT. Then NaN_3 was added to the reaction

mixture and allowed to react at RT for 12 hrs. After this, the reaction mixture was concentrated and purified by NPC without any workup. Above obtained azide (**47**) was dissolved in THF at 0 ºC and THF dissolved PPh3 and was added at 0 ºC. Then the reaction mixture was allowed to react at RT. Water was then added to the reaction mixture after 12 hrs and allowed to stir for another 1 hour. Then the reaction mixture was extracted with DCM and purified using NPC. Amine (**48**) was dissolved in a sat. Aq solution of $NaHCO_3$ and cooled on an ice bath. Then N-(methoxycarbonyl) maleimide was added in portions under stirring. The mixture was stirred for 1 hr at 0 ºC, followed by 1 hr at RT. After extraction with DCM, the organic phase was dried over anhydrous Na_2SO_4, concentrated and purified by NPC with (MeOH/DCM) yielded the product (**49**).

4.2.3. Conjugation and purification



Scheme 4.9. Conjugation of the amphiphilic probe to the protein of interest (BSA) at the thiol residue.

The modification of BSA was carried out at the concentration of 100 µM following the procedure mentioned in chapter 4.1. Triton X-100 was used to solubilize the AABP only at concentrations 10 times (2 mM) more than CMC or 0.2% of the total volume of the reaction mixture. The AABP was used in 25 - 50 eq. We saw more than 70 % conversion with just 25 eq in test scale, unlike 2-PCA chemistry, which required 100 eq. The molecular weight changes can be seen in the reaction mixtures MALDI-TOF MS shown here (Fig. 4.15). To monitor the extent of protein modification, the samples were directly withdrawn from the reaction mixture using a

pipette and analyzed. In brief, the sample and matrix mixture was mixed at a 1:10 ratio. When the crystallization observed in centrifuge tubes, 1-2µL of the samples was spotted on the MALDI-ToF MS plate and air-dried for 15 minutes. Then the plate was loaded and fired to get accurate molecular weights. Please refer to the methods section for the matrix preparation and sample preparation procedures.

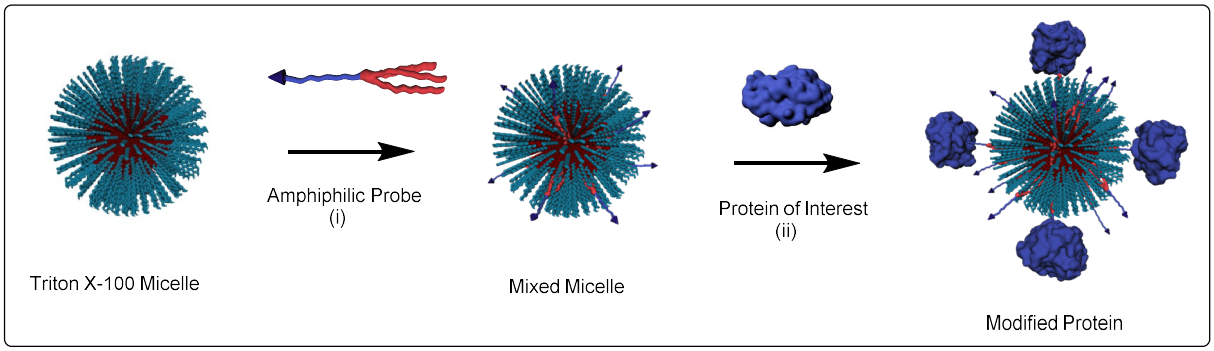


Figure 4.13. MAPLab technology: Conjugation steps and mechanism.

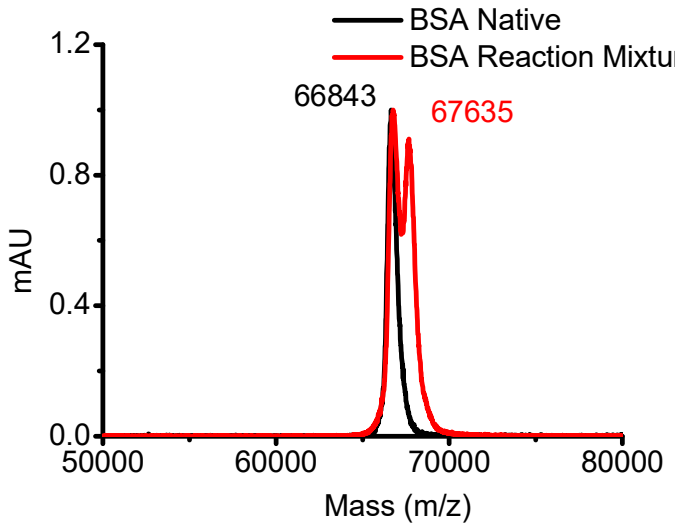


Figure 4.14. MALDI-TOF MS data of native and reaction mixtures of BSA.

BSA conjugate was purified by 2-step purification, *i.e.,* IEX and SEC only like mentioned in the previous subchapter 4.1. Then the fractions with conjugate were stored at -80 ºC without further desalting.

4.2.4. Characterization

Before understanding the self-assembly behavior of the conjugates, we characterized the samples using MALDI-TOF MS (Fig. 4.16). The detailed procedures for characterization are given in the experimental section. Soon after characterization, we started to study the self-assembly of this conjugate.

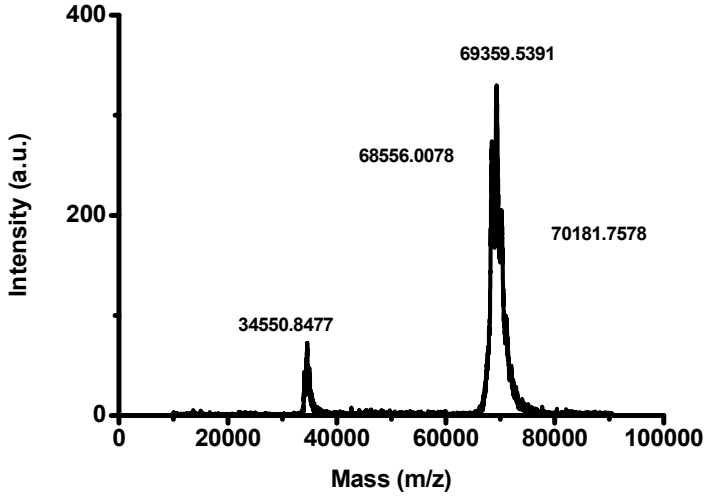


Figure 4.15. MALDI-TOF MS data of reaction mixture of BSA-Thiol Conjugate.

4.2.5. Self-assembly studies

In order to test the self-assembling ability of designed BSA conjugate, self-assembly studies were carried out using analytical SEC. The detailed procedure for the same is given in the methods section at the end of this chapter. When compared with the BSA conjugate of N-terminus, the thiol conjugate is relatively smaller in size (Fig. 4.17). This is because of the fact that the change in the site of conjugation might result in different exposure of charged residues (Fig. 4.18).

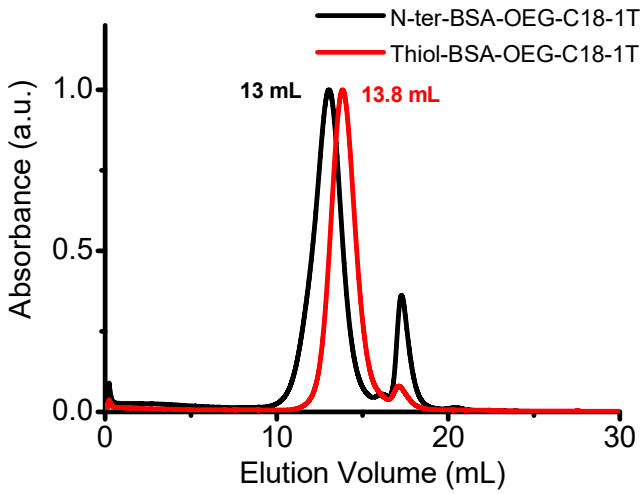


Figure 4.16. SEC data of IEX purified BSA thiol conjugate with respect to N-terminus conjugate.

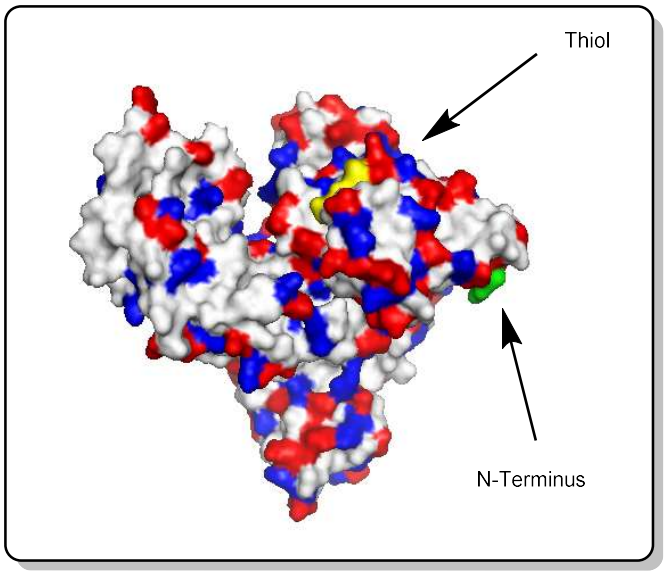


Figure 4.17. BSA (PDB accession: 3V03) structure showing thiol residue (yellow) with respect to N-terminus location (green). Negatively charged residues like Asp and Glu are colored red. Positively charged residues like Arg and Lys are colored blue.

Table 4.2. Comparison of SEC data of N-terminus vs. thiol conjugate of BSA.

S.No	Protein Conjugate	Elution Vol SEC (mL)	RMW SEC (kDa)	Oligomeric state SEC (mer)	D_h DLS (nm)
1	BSA Nat	17.4	66	1	--
2	BSA-OEG-C18-1T	13	700	10	13.24
2	BSA-OEG-C18-1T	13.8	630	9	--

4.2.6. Discussion

To address the drawback of the proposed N-terminus based universal method in the previous chapter, site-specific modification of thiol was explored in this chapter to make protein complex by selecting BSA as a model protein. So, we have used maleimide (MA) chemistry for thiol targeting and converted BSA into a FAP using MALab technology. When compared with BSA FAP of N-terminus, the thiol FAP is relatively smaller in size because the different sites of conjugation might expose the different surface to bulk of the solvent and to each BSA monomer in the complex, resulting in changes in oligomeric state and molecular weight. This protein modification is as good as a universal method as genetic engineering, and the introduction of thiol is very common.

4.3. Experimental section and methods

Selection of proteins used in this work

We chose 4 proteins for our study, *i.e.*, Lysozyme (14.3 kDa), chymotrypsin (25.4 kDa), GFP (29.5 kDa), and BSA (66.6 kDa). Proteins were purchased from commercial vendors (Except GFP whose plasmid was generously donated by Dr. Thomas Pucadyil and expressed by Ms. Puneeta), and molecular weights were determined prior to modification to track the molecular weight changes accurately.

Matrix preparation and molecular weight determination

Apart from the molecular weight determination of native proteins, all stages of protein modification and purification were also followed by MALDI-TOF MS. The samples were analyzed in Linear High Mass mode in AB Sciex 4800 plus MALDI-TOF/TOF analyzer with 4000 Series Explorer as software. Mass was scanned between 10,000 Da and 90,000 Da with a focus mass, selected depending on the protein analyzed.

We followed a different MALDI-TOF Ms Procedure in this chapter. To brief, 15mg of sinapinic acid was weighed in microcentrifuge tube then add 1.0 ml of matrix solution (70:30 water/acetonitrile with 0.1 % TFA final concentration) was added and vortexed to get the matrix mixture. The sample and matrix mixture were mixed at a 1:10 ratio. When the crystallization observed in centrifuge tubes, 1-2µL of the samples were spotted on the plate and air-dried for 15 minutes. Then the plate was loaded and fired to get accurate molecular weights.

Protein modification, monitoring, and purification

Protein modification, monitoring, and purification were done using a similar protocol given in chapter 2. However, some changes were made in the equivalents we used and triton X-100 concentrations we used. They are mentioned here. The AABP was used varied between 25 - 100 eq, which was just 2 eq in previous chapters. Triton X-100, which was used concentrations 100 times (20 mM) more than CMC or 2% of the total volume of the reaction mixture to solubilize the AABPs in the previous chapter, gave noisy MALDI-TOF Spectra. So we decreased the Triton X-100 concentration to 10 times more than CMC. Heating the reaction mixture at 37ºC

triggered precipitation of protein conjugates in the reaction mixtures. So even though we see increased conversions with heating, we did the conjugation at RT.

All the conjugates were purified by two-step purification, *i.e.,* IEX and SEC, performed using FPLC. IEX was performed to remove triton X-100 using either SP sepharose or Q sepharose resins (GE) depending on isoelectric point (pI) and surface charges of proteins. For example, to purify the reaction mixture of LZ (pI: 11.35) and Chy (pI: 8.75) we used SP sepharose, a cation-exchange resin at pH 7.4 and to purify BSA (pI: 4.7) and GFP is (pI: 5.80) we used Q sepharose, an anion-exchange resin at pH 7.4. The obtained IEX fractions were subjected to SEC immediately to remove the native proteins from the protein conjugates in 50 mM sodium phosphate pH 7.4, 1 M NaCl using either Sephacryl S-100 HR 16/60 or Sephacryl S-200 HR 16/60 or Sephacryl S-300 HR 16/60. After SEC the samples were stored at -80 ºC

Synthesis and purification of N-terminus BSA conjugate

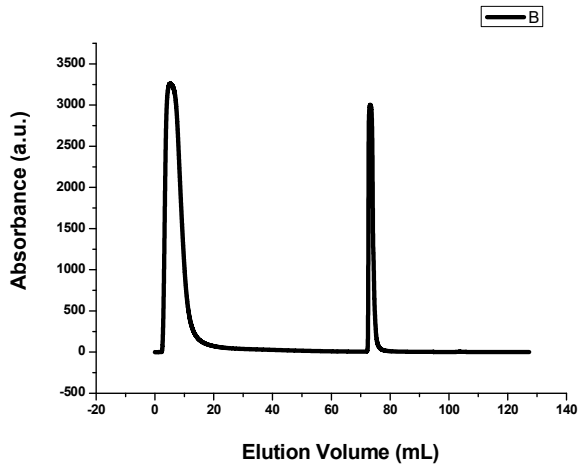


Figure 4A. IEX chromatogram of BSA reaction mixture.

Synthesis and purification of N-terminus chymotrypsin conjugate

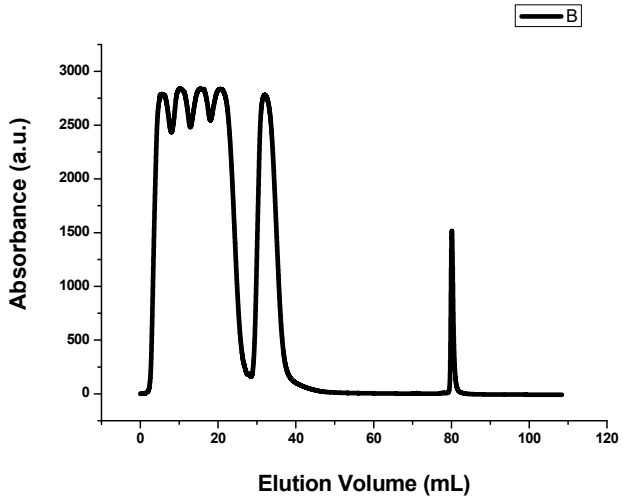


Figure 4B. IEX chromatogram of the chymotrypsin reaction mixture.

Synthesis and purification of N-terminus GFP conjugate

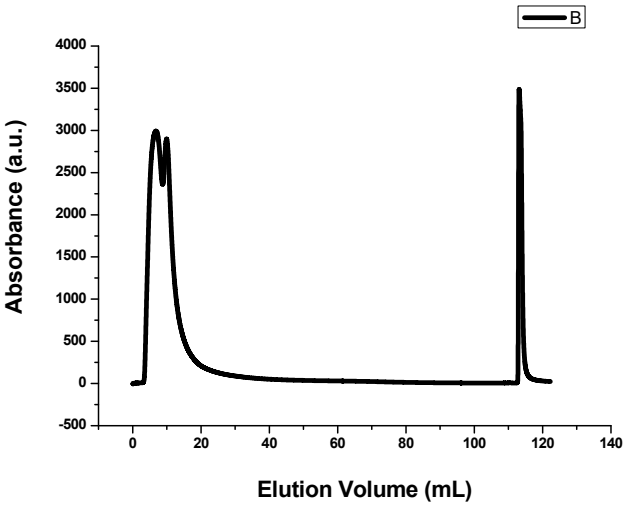


Figure 4C. IEX chromatogram of lysozyme reaction mixture.

Synthesis and purification of N-terminus GFP conjugate

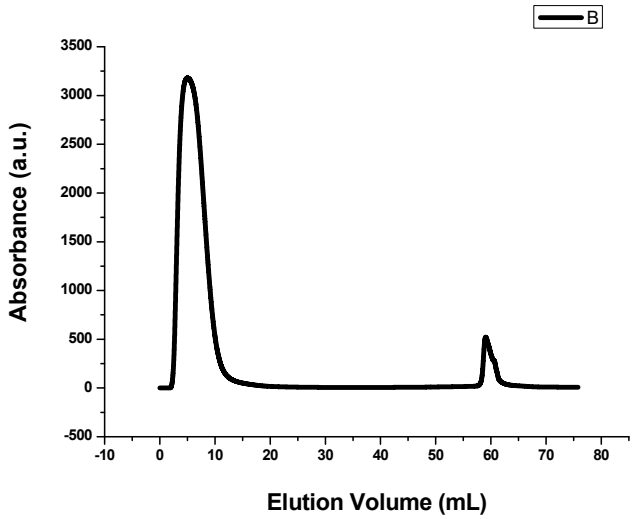


Figure 4D. IEX chromatogram of GFP reaction mixture.

Synthesis and purification of BSA-Thiol conjugate

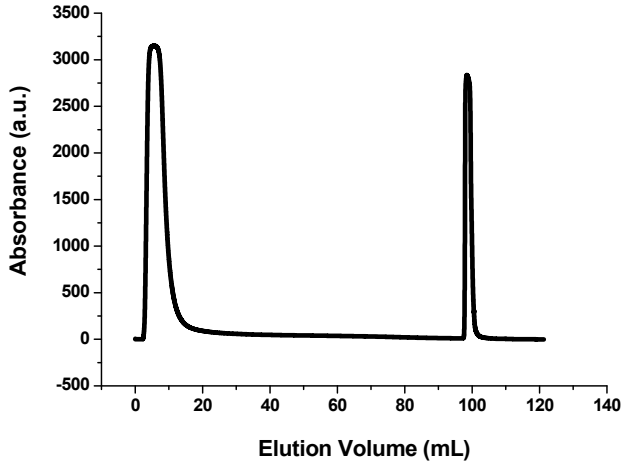


Figure 4E. IEX chromatogram of BSA-Thiol conjugation.

Size exclusion-based molecular weight determination

To determine the molecular weights of soluble protein complexes, we performed SEC in Superose 6 (GE healthcare). For this purpose, we chose standard proteins (GE Healthcare), *i.e.*, blue dextran (2000 kDa), to determine the void volume of this column, thyroglobulin (660 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa) and trypsin (23 kDa).

A series of size exclusion runs were performed with these proteins in 50 mM sodium phosphate pH 7.4, 200 mM NaCl with 0.25 mL/min as the flow rate. All the proteins were dissolved individually in Milli Q water at 3 mg/mL and 0.5 mL was injected. With the obtained elution volumes, the partition coefficients (K_{av}) for standard proteins were calculated according to standard protocol (GE Healthcare) using the formula $K_{av} = V_e - V_o / V_c - V_o$ where, V_o = column void volume (8.5 mL, from the elution volume of blue dextran), V_e = elution volume of the particular sample, and V_c = geometric CV (23.5 mL). The calibration curve was plotted for K_{av} of standards against relative molecular weights (RMW) of standard proteins. ($r^2 = 0.99$, please refer to the calibration curve). Now, to determine the molecular weights of soluble protein complexes, the protein conjugates were dissolved individually in Milli Q water at 5 mg/mL and 0.5 mL was injected.

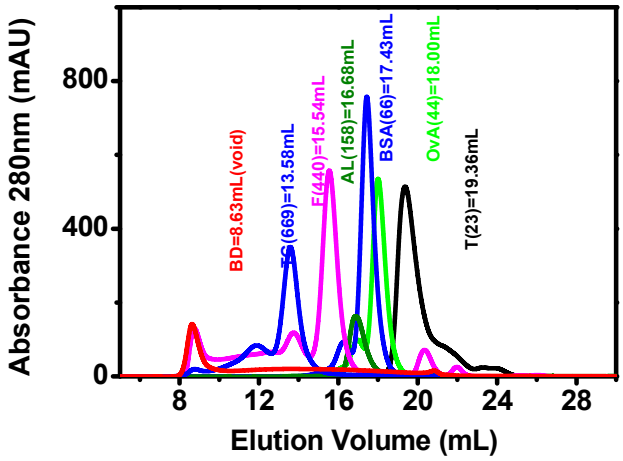


Figure 4F. SEC runs for standard proteins in Superose 6.

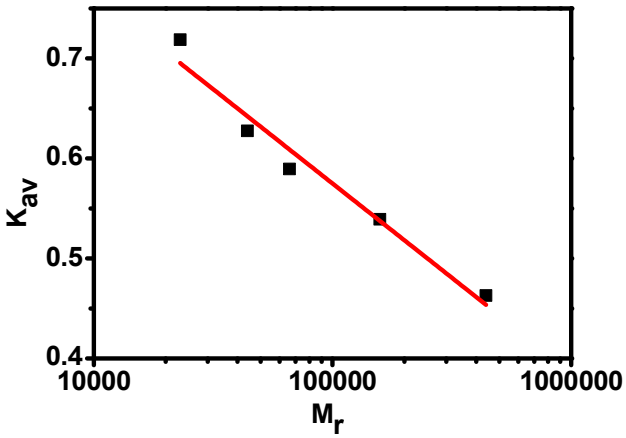


Figure 4G. Calibration curve for Superose 6.

Dynamic light scattering, Fluorescence Spectroscopy, Circular Dichroism (CD) Spectroscopy, Zeta potential measurements were done in the same fashion using the same instruments given in chapter 2.

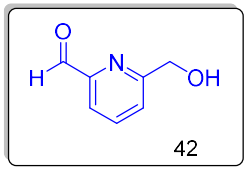
Synthesis and purification of AABPs and their intermediates

General remarks

All compounds were made, characterized using similar conditions and instruments as mentioned in chapter 2.

Synthesis of individual compounds in chapter 4.1

Compound 42



Mol. formula: $C_7H_7NO_2$

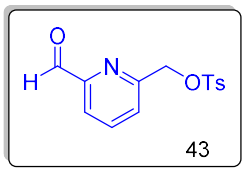
Mol. Weight: 137.13 g/mol

Physical appearance: white solid

Yield: 60 %

To a solution of 2,6-pyridinedimethanol (500 mg, 3.62 mmol) in 1,4-dioxane (10 mL), SeO_2 (200 mg, 1.81 mmol) was added. The resulting mixture was sonicated for 5 mins and then stirred at 65 ºC for 24 hrs. Then the reaction was cooled to room temperature and diluted with dichloromethane. The mixture was filtered through celite, and the filtrate was concentrated under reduced pressure. The resulting crude material was purified by chromatography using PE/EtOAc to get a liquid that became off-white solid later (300mg, 60%). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 10.02 (s, 1H), 7.84 (m, 2H), 6.52 (m, 1H), 4.48 (s, 1H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 193.12, 160.43, 151.66, 137.86, 125.00, 120.65, 77.16, 64.24.

Compound 43



Mol. formula: $C_{14}H_{13}NO_4S$

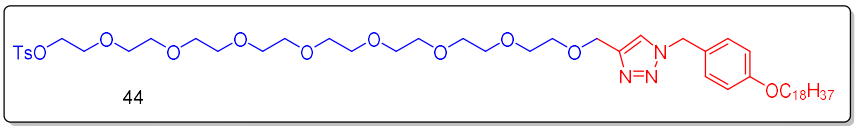
Mol. weight: 291.32 g/mol

Physical appearance: white solid

Yield: 55 %

Tosyl chloride (542 mg, 2.64 mmol) was added to the solution of alcohol (**compound 42**) (300 mg, 2.2 mmol) in DCM (10mL) at 0 ºC. Et_3N (661 mg, 906 µl, 6.54 mmol) was then added to the above mixture, maintaining the temperature at 0 ºC. After 1hour at RT, the reaction mixture was concentrated and purified using column chromatography using PE / EtOAc (350mg, 55%). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 9.93 (s, 1H), 7.85 (m, 4H), 7.66 (d, 1H), 7.35 (d, J = 8.4 Hz, 2H), 5.22 (s, 2H), 2.43 (s, 3H). **^{13}C NMR (**100 MHz, $CDCl_3$): δ_C 154.80, 152.24, 145.41, 138.22, 132.67, 130.09, 128.18, 126.10, 121.35, 77.16, 71.19, 21.75. **HRMS** (MW+H): 292.06 g/mol.

Compound 44



Mol. formula: $C_{51}H_{85}N_3O_{12}S$

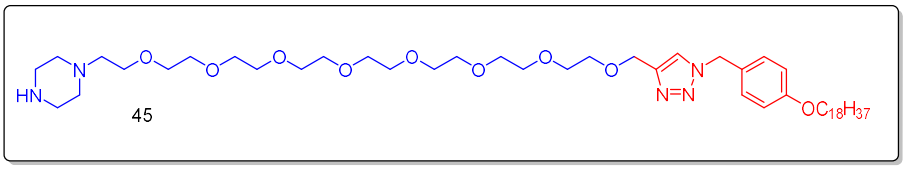
Mol. weight: 964.31 g/mol

Physical appearance: yellow waxy solid

Yield: 75 %

Synthesis of alkyne (**compound 2c**) and azide (**compound 10h**) was done as per our previous report.[18] Alkyne (5 g, 8.89 mmol) and azide (2.82 g, 8.89 mmol) were weighed in an oven dried RBF and THF was added, followed by water with vigorous stirring. Then, Na Asc. (6.6 mg, 0.033 mmol) was added followed by $CUSO_4$ (2.6 mg, 0.016 mmol) and allowed to react overnight. The reaction was extracted with DCM and then concentrated under reduced pressure. The resulting crude material was purified by NPC using PE/EtOAc (6.4 g, 75 %). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.78 (d, J = 8.4, 2H),7.47 (s, 1H), 7.33 (d, J = 8, 2H), 7.20 (d, J = 6.8 Hz, 2H), 6.86 (d, J = 8 Hz, 2H), 5.41 (s, 2H), 4.62 (s, 2H), 4.1 (m, J = 6.8Hz, 2H), 3.92 (t, J = 6.4 Hz, 2H), 3.63 (m, 30H), 3.5 (m, 5H), 1.75 (t, J = 7.2 Hz, 2H), 1.42 (m, 2H), 1.24 (m, 34H), 0.86 (t, J = 6.4 Hz, 3H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 159.57, 144.89, 133.11, 129.79, 128.07, 126.37, 115.05, 77.16, 70.62, 69.78, 69.49, 68.75, 68.19, 32.01, 29.78, 29.69, 29.67, 29.48, 29.45, 29.28, 26.11, 22.78, 21.74, 14.23. **MALDI-TOF MS** (MW+ Na): 986.55 g/mol.

Compound 45



Mol. formula: $C_{48}H_{87}N_5O_9$

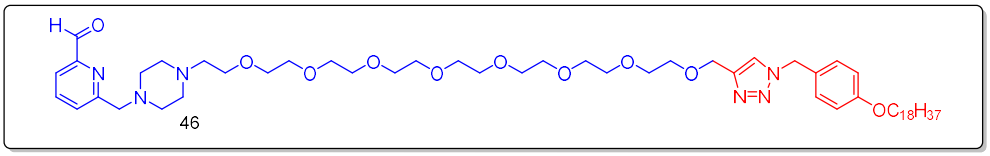
Mol. weight: 878.25 g/mol

Physical appearance: yellow waxy solid

Yield: 82 %

Tosylate **(compound 44)** (2 g, 2.07 mmol) and piperzine (1.9 g, 22.09 mmol) were dissolved in THF (30 mL). Then the mixture was refluxed at 65 ºC. After 12 hrs, reaction mixture was concentrated and purified by NPC (1.5 g, 82 %). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.48 (s, 1H), 7.12 (d, *J* = 8.8 Hz, 2H), 6.77 (d, *J* = 8.8 Hz, 2H), 5.43 (s, 2H), 4.53 (s, 2H), 3.82 (t, *J* = 6.4 Hz, 2H), 3.51 (m, 31H), 3.04 (t, *J* = 4.4 Hz, 4H), 2.51 (m, 5H), 1.66 (m, 2H), 1.34 (m, 3H), 1.15 (m, 31H), 0.769 (t, *J* = 6.8 Hz, 3H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 159.28, 145.12, 129.51, 126.20, 122.23, 114.77, 77.16, 70.36, 70.17, 69.50, 68.58, 68.34, 67.90, 64.47, 57.48, 57.17, 53.43, 50.12, 45.72, 43.55. **MALDI-TOF MS** (MW+ Na): 900.72 g/mol.

Compound 46



46

Mol. formula: $C_{55}H_{92}N_6O_{10}$

Mol. weight: 997.37 g/mol

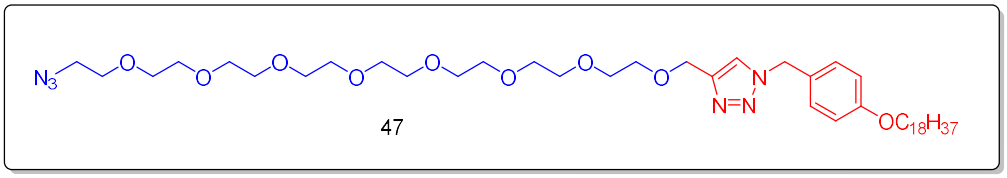
Physical appearance: yellow waxy solid

Yield: 33 %

Compound 45 (1.6 g, 1.82 mmol), K_2CO_3 (0.500 g, 3.62 mmol) and tosylate (**compound 43**) (0.530g, 1.82 mmol) were weighed in RBF. Then the mixture was dissolved in ACN (10 mL) and refluxed at 65 ºC. After 16 hrs, reaction mixture was concentrated and purified by NPC (0.6 g, 33 %). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 10.06 (s, 1H), 7.83 (m, 2H), 7.8 (m, 1H), 7.43 (s, 1H), 7.21 (d, *J* = 8.4 Hz, 2H), 7.67 (d, *J* = 8.4, 2H), 5.42 (s, 2H), 4.63 (s, 2H), 3.92 (t, *J* = 6.4 Hz, 2H), 3.75 (s, 2H), 3.61 (m, 32H), 2.59 (m, 9H), 1.75 (m, 2H), 1.42 (m, 2H), 1.24 (m, 32H), 0.86 (t, *J* = 6.8 Hz, 3H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ_C 193.77, 137.51, 129.81, 120.29, 115.08, 77.16, 70.67, 70.48, 68.23, 64.91, 64.17, 57.80, 53.83, 53.68, 53.34, 32.04, 29.81, 29.61, 29.31, 26.14, 22.81, 14.25. **MALDI-TOF MS** (MW+ Na):1019.60 g/mol.

Synthesis of individual compounds in Chapter 4.1

Compound 47



47

Mol. formula: $C_{44}H_{78}N_6O_9$

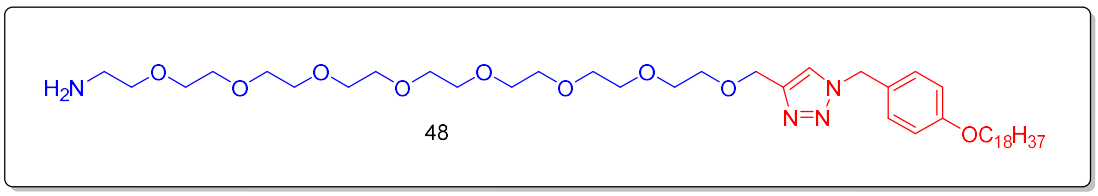
Mol. weight: 835.14 g/mol

Physical appearance: yellow waxy solid

Yield: 70 %).

Tosylate (**compound 46**) (2 g, 2.0 mmol) was dissolved in DMF (10 mL) at RT. Then NaN_3 (0.674 g, 10.3 mmol) was added to the reaction mixture and allowed to react at RT for 12 hrs. After this, the reaction mixture was concentrated and purified by NPC without any workup (1.2 g, 70 %). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.40 (s, 1H), 7.20 (d, J = 8 Hz, 2H), 6.86 (d, J = 8 Hz, 2H), 5.41 (s, 2H), 4.62 (s, 2H), 3.91 (t, J = 6.8 Hz, 2H), 3.63 (m, 32H), 3.35 (t, J = 4.8 Hz, 2H), 1.74 (t, J = 6.8 Hz, 2H), 1.42 (m, 2H), 1.2 (m, 32H), 0.85 (t, J = 6.4, 3H). **MALDI-TOF MS** (MW+ Na): 874.58 g/mol.

Compound 48



48

Mol. formula: $C_{44}H_{80}N_4O_9$

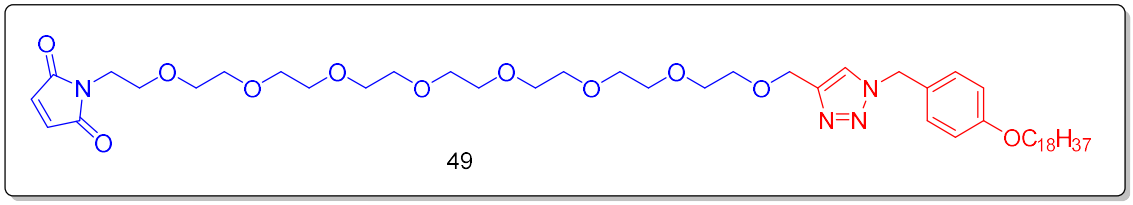
Mol. weight: 808.14 g/mol

Physical appearance: yellow waxy solid

Yield: 51 %

Azide (**compound 47**) (400 mg, 0.479 mmol) was dissolved in THF (7 mL) at 0 ºC and PPh_3 (250 mg, 0.954 mmol) dissolved in THF (3 mL) and was added at 0 ºC. Then the reaction mixture was allowed to react at RT. Water was then added to the reaction mixture after 12 hrs and allowed to stir for another 1 hr. Then the reaction mixture was extracted with DCM and purified using NPC (200 mg, 51 %). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.40 (s, 1H), 7.20 (d, J = 8.8 Hz, 2H), 6.86 (d, J = 8.8 Hz, 2H), 5.43 (s, 2H), 4.64 (s, 2H), 3.93 (t, J = 6.8 Hz, 2H), 3.68 (m, 32H), 3.38 (t, J = 4.8 Hz, 2H), 1.76 (m, 2H), 1.42 (m, 2H), 1.33 (m, 30H), 0.87 (t, J = 7.2 Hz, 3H). **MALDI-TOF MS** (MW+ Na): 831.53 g/mol.

Compound 49



49

Mol. formula: $C_{48}H_{80}N_4O_{11}$

Mol. weight: 889.58 g/mol

Physical appearance: yellow waxy solid

Yield: 50 %

Amine (**compound 48)** (100 mg, 0.123 mmol) was dissolved in a mixture of saturated aqueous solution of $NaHCO_3$ (5 ml) and THF (5 ml) cooled on an ice bath. Then N-(methoxy carbonyl) maleimide (23 mg, 0.148 mmol) was added in portions over 5 mins under vigorous stirring. The mixture was stirred for 1 hr at 0 ºC, followed by 1 hr at room temperature. After extraction with DCM, the organic phase was dried over anhydrous Na_2SO_4, filtered, and concentrated. Purification by silica gel column chromatography (MeOH/DCM) yielded the product as colorless oil after evaporation with DCM (55 mg, 50 %). **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H 7.43 (s, 1H), 7.20 (d, J = 8.8 Hz, 2H), 6.86 (d, J = 8.8 Hz, 2H), 6.68 (s, 2H), 5.46 (s, 2H), 4.62 (s, 2H), 3.91 (t, J = 6.1 Hz, 2H), 3.70 (t, J = 5.6 Hz, 2H), 3.63 (m, 32H), 1.74 (m, 2H), 1.41 (m, 2H), 1.23 (m, 32H), 0.85 (t, J = 7.2 Hz, 3H). **MALDI-TOF MS** (MW+ Na): 911.58 g/mol.

References

1 Cravatt, B. F., Wright, A. T. & Kozarich, J. W. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* **2008**, 77, 383-414.

2 Sandanaraj, B. S., Reddy, M. M., Rao, K. J. & Bhandari, P. J. Rational Design of Semi-Synthetic Protein Complexes with the Defined Oligomeric State. *ChemistrySelect* **2019**, 4, 6397-6402.

3 Sandanaraj, B. S., Bhandari, P. J., Reddy, M. M., Lohote, A. B. & Sahoo, B. Design, Synthesis, and Self-Assembly Studies of a Suite of Monodisperse, Facially Amphiphilic, Protein–Dendron Conjugates. *ChemBioChem* **2019**,

4 Rosen, C. B. & Francis, M. B. Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol.* **2017**, 13, 697.

5 MacDonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **2015**, 11, 326-331.

6 Dirks, A. T. *et al.* Preparation of biohybrid amphiphiles via the copper catalysed Huisgen [3+ 2] dipolar cycloaddition reaction. *Chem. Commun.* **2005**, 4172-4174.

7 Heredia, K. L. *et al.* In situ preparation of protein−"smart" polymer conjugates with retention of bioactivity. *J. Am. Chem. Soc.* **2005**, 127, 16955-16960.

8 Han, G., Wang, J.-T., Ji, X., Liu, L. & Zhao, H. Nanoscale Proteinosomes Fabricated by Self-Assembly of a Supra-molecular Protein-Polymer Conjugate. *Bioconjugate Chem.* **2016**,

9 Huang, X. *et al.* Interfacial assembly of protein-polymer nano-conjugates into stimulus-responsive biomimetic protocells. *Nat. Commun.* **2013**, 4, 2239.

10 Velonia, K., Rowan, A. E. & Nolte, R. J. Lipase polystyrene giant amphiphiles. *J. Am. Chem. Soc.* **2002**, 124, 4224-4225.

11 Huang, A., Qin, G. & Olsen, B. D. Highly Active Biocatalytic Coatings from Protein–Polymer Diblock Copolymers. *ACS Appl. Mater. Interfaces* **2015**, 7, 14660-14669.

12 Thomas, C. S., Xu, L. & Olsen, B. D. Kinetically Controlled nanostructure formation in self-assembled globular protein–polymer diblock copolymers. *Biomacromolecules* **2012**, 13, 2781-2792.

13 Thomas, C. S., Xu, L. & Olsen, B. D. Effect of Small Molecule Osmolytes on the Self-Assembly and Functionality of Globular Protein–Polymer Diblock Copolymers. *Biomacromolecules* **2013**, 14, 3064-3072.

14 Chang, D., Lam, C. N., Tang, S. & Olsen, B. D. Effect of polymer chemistry on globular protein–polymer block copolymer self-assembly. *Polym. Chem.* **2014**, 5, 4884-4895.

15 Lam, C. N. *et al.* The Nature of Protein Interactions Governing Globular Protein–Polymer Block Copolymer Self-Assembly. *Biomacromolecules* **2014**, 15, 1248-1258.

16 Xia, Y., Tang, S. & Olsen, B. D. Site-specific conjugation of RAFT polymers to proteins via expressed protein ligation. *Chem. Commun.* **2013**, 49, 2566-2568.

17 Kalia, D., Malekar, P. V. & Parthasarathy, M. Exocyclic olefinic maleimides: Synthesis and application for stable and thiol-selective bioconjugation. *Angew. Chem. Int. Edn. Engl.* **2016**, 55, 1432-1435.

18 Sandanaraj, B., Reddy, M., Bhandari, P., Kumar, S. & Aswal, V. Rational Design of Supramolecular Dynamic Protein Assemblies Using a Micelle-Assisted Activity-based Protein Labeling Technology. *Chem. - Eur. J.* **2018**,

Appendix-I: Synopsis and Outlook

Rational Design, Synthesis and Self-assembly Studies of Facially Amphiphilic Proteins (FAPs)

The research presented here attempts to provide a modular methodology to assemble proteins into higher order complexes. The methodology should address drawbacks in chemical strategies (synthesis of proteins-polymer conjugates) like non ubiquitous, non specific modification, polydispersity and poor characterization. The research presented here also attempts to provide a modular methodology to assemble proteins into stimuli responsive complexes. The research presents a universal method to assembly any protein of interest into a complex.

Chapter 1: Introduction to Protein Engineering

Some of the important protein engineering literature is covered in this chapter in the following subheadings

- Engineering protein cages
- Engineering protein oligomers
- Engineering monomeric proteins

Chapter 2: Facially Amphiphilic Proteins (FAPs) and Micelle Assisted Protein Labelling (MAPLab) Technology Development

In this chapter, we have developed a new chemical strategy to convert water soluble monomeric proteins into higher order complexes. Some model serine proteases were converted into higher order complexes using fluorophosphonate chemistry. This method is based on the generation of simple scaffold which we call "Facially Amphiphilic Proteins (FAPs)" from native proteins. This design is based on naturally occurring Facially Amphiphilic proteins called Hydrophobins. The generation of FAP was achieved using a Micelle-Assisted protein labeling (MAPLab) technology. The FAP, driven by strong hydrophobicity assembled to form a higher order protein complex.

The proposed methodology is amenable for the synthesis of protein complex libraries with molecular weights and dimensions comparable to naturally occurring protein cages. The simple design allowed us to study the behaviour of the scaffold with respect to individual components (protein head, hydrophilic spacer, and hydrophobic tail). We have then made and explored library of FAPs in a systematic way in order to understand the ways of controlling the attributes (the molecular weights, oligomeric states and dimensions of the synthesized protein complexes) of assemblies precisely. The designed protein complexes display very rich structural diversity, oligomeric states, sizes and surface charges that can be engineered through macromolecular design. The most striking features of our method are (i) simple design, (ii) high success rate, (iii) rapid generation of library of supramolecular protein complexes with unique structural diversity, and (iv) the size, oligomeric state and surface charge of protein complex can be precisely engineered through macromolecular design. This chapter is dealt in the following sub chapters

2.1. Synthesis and Self-assembly Studies of Prototype FAP

- In this chapter, we have developed the actual chemical methodology called micelle-assisted activity-based protein labelling (MAPLab) technology to make protein complexes from monomeric water soluble proteins.
- We have made a prototype FAP from trypsin using FP chemistry
- We have established the purification protocol for the FAP, characterized it and studied its self-assembly behaviour.

2.2. Synthesis and Self-assembly Studies of Protein Variant FAPs

- In this sub-chapter, MAPLab technology has been extended to make protein variant FAPs from serine proteases using FP chemistry.
- We have studied influence of protein head in the self-assembly of FAPs.
- We have found that the influence of protein is significant and proteins with low overall surface charge, resulted in protein complex with high aggregation number and molecular weight. Thus the oligomeric state of complex could be tuned by careful choice of Protein group.

2.3. Synthesis and Self-assembly Studies of Spacer Variant FAPs

- In this sub-chapter, MAPLab technology has been extended to make spacer variant FAPs from trypsin using FP chemistry.
- We have studied influence of spacer group in the self-assembly of FAPs.
- We have found that the length of spacer is also crucial factor in controlling the aggregation number. Longer spacer resulted in increased aggregation numbers

2.4. Synthesis and Self-assembly Studies of Tail Variant FAPs

- In this sub-chapter, MAPLab technology has been extended to make tail variant FAPs from trypsin using FP chemistry.
- We have studied influence of tail group in the self-assembly of FAPs.
- We have found that tail group plays a dominant role in controlling the self-assembly.
- We found the minimum hydrophobicity needed for self-assembly of FAPs.
- We have also investigated the influence of branching in hydrophobicity in the self-assembly of FAPs.

Chapter 3: Design of FAPs for the Accurate Design of Stimuli-responsive Protein Complexes

In this chapter, we have developed a new way of creating stimuli-responsive protein-complexes. We demonstrated sophisticated photo-responsive and redox- responsive protein-complexes that can be programmed to assemble/disassemble in a reversible/irreversible fashion using light and ph or as a trigger respectively. Our results demonstrate that accurate design of most sophisticated photo-responsive molecular machine can be achieved through chemical method. We firmly believe this work would inspire genetic methods (both computational and rational approaches) for construction of stimuli-responsive protein complexes using unnatural amino acids as building blocks through bottom-up approach. This chapter also deals with peptide-based FAPs which can drive the self-assembly through possible H-bonding apart from hydrophobicity. This chapter is discussed as following sub-chapters:

3.1. Synthesis, Self-assembly and Dis-assembly Studies of Photo-responsive FAPs

- In this sub-chapter, MAPLab technology has been extended to make Photo-responsive FAPs from Serine proteases using FP chemistry.
- To make this Photo-responsive complex, we have installed a stimuli-responsive 2-NB group between spacer and tail in the design of FAP.
- Upon photo-irradiation, the hydrophobic tail got cleaved leading to destabilization of complex and dis-assembly because of increased electrostatic repulsion between proteins.

3.2. Synthesis, Self-assembly and Dis-assembly Studies of Redox-responsive FAPs

- In this sub-chapter, MAPLab technology has been extended to make Redox-responsive FAPs from Serine proteases using FP chemistry.
- To make this Redox-responsive complex, we have installed a stimuli-responsive S-S group between spacer and tail in the design of FAP.
- Upon DTT treatment, the hydrophobic tail got cleaved and complex dis-assembled emphasizing the fact that favourable interaction between the hydrophobic tails is the major driving force for protein conjugate to self-assemble.

3.3. Synthesis and Self-assembly Studies of Peptide-based FAPs

- In this sub-chapter, MAPLab technology has been extended to make peptide-based FAPs from Serine proteases using FP chemistry.
- We have made two libraries in this study, i. Protein variant library and Peptide variant library.
- We have found that the influence of Protein is similar to that of chapter 2. We have also found that the Peptide hydrophobicity is also influencing the self-assembly

Chapter 4: Targeting N-terminus and Thiol for Universal Methodology

N-terminus and Thiol resides were explored to develop a Universal method to make complexes in this chapter. We believe that the methods developed in this chapter have immense potential in the area of vaccine development. Recent advances in the

identification of antigens capable of generating protective immunity have driven the development of subunit vaccines, where instead of using whole pathogens; specific portions are displayed and delivered using virus-like particles (VLPs). This requires the presentation of minimal, monomeric antigenic proteins into repetitive or multivalent/polyvalent fashion. However, using our methodologies discussed in this chapter, antigens can be converted into particles directly without the need for a viral scaffold to display.

This chapter is discussed as following sub-chapters:

4.1. Targeting N-terminus for Universal Methodology

- In this sub-chapter, MAPLab technology has been extended to convert any protein of interest into protein complexes with few exceptions by targeting n-terminus.
- Some model proteins and 2-PCA chemistry are chosen for this purpose.

4.2. Targeting Thiol for Universal Methodology

- In this sub-chapter, MAPLab technology has been extended to convert any protein of interest into protein complexes with few exceptions by targeting thiol residue.
- BSA and maleimide chemistry are chosen for this purpose.

Outlook

New viral outbreaks are always a threat to humanity, deadly and are capable of wiping out half of humanity. SARS-CoV, MERS-CoV, H1N1, Ebola, Zika, and recent 2019-nCoV outbreaks are such examples. Antiviral drugs can't be discovered effectively in many cases, so there is a never-ending quest for vaccine design/discovery for the viruses. Even with a better understanding of the genome and proteome of viruses, even today, we are at least 24-36 months away from creating a vaccine for a new outbreak.

We believe that this thesis work has immense potential in the area of vaccine development. Traditional vaccines like killed or attenuated organisms or their purified components have proven to be the most effective methods but with some safety issues including the risk of

reversion to a virulent organism and the possibility of disease in immune-compromised individuals. However recent advances in the identification of antigens capable of generating protective immunity have driven the development of subunit vaccines, where instead of using whole pathogens; specific portions are displayed and delivered as a Virus-Like Particles (VLPs). This requires the presentation of minimal/monomeric antigenic proteins into repetitive / multivalent/ polyvalent fashion.

Significant protein engineering efforts are directed towards this area especially the computational methods. Antigenic proteins were decorated on synthetic protein cages. The chemical approaches have also been used to functionalize natural viruses with foreign polypeptides or antigenic proteins. However, these methods can be used universally.

Our approach however is very simple, yet powerful way of creating protein complexes. We were able to control the molecular weights, oligomeric states and dimensions synthesized protein complexes precisely. We have also developed a new way of creating smart stimuli-responsive protein-complexes. Further, most importantly we extend this method to a universal chemistry to convert any protein of interest into proteins complexes with few exceptions.

Appendix-II: List of Figures

List of figures in chapter 1

Figure 1.1. a. The EGFP-capsid is mixed with wild-type capsid at pH 7.5 and then dialyzed against pH 5.0 to induce capsid formation resulting in EGFP encapsulation. b. Cascade reaction performed by co-encapsulated GOx and GCK at the interior of the CCMV capsids (the conversion of glucose to D-gluconate-6-P). c. The encapsulation of PC stacks into CCMV VLPs. Adapted with permission from ref [16], [17], and [18].

Figure 1.2. ELP-CP and it's Self- assembly: pH-induced assembly and ELP-induced assembly. Adapted with permission from ref [20].

Figure 1.3. a. conjugation of the DOX to CPMV. b. CPMV, with its SS RNA molecule (in red), is exposed to a solution containing the cargo (in blue), washing, and dialysis to get intact CPMV with infused cargo. Adapted with permission from ref [28] and [29].

Figure 1.4. a. Qβ plasmids used for particle production and the EGF-display. b. Qβ, plasmids used for chimeric VLP display of b122a. Adapted with permission from ref [33] and [35].

Figure 1.5. Synthesis of Qβ-Polymer conjugates to evaluate the immune recognition of polymer conjugated particles. Adapted with permission from ref [38].

Figure 1.6. a. The representative figure to show introduced pyrenes on the formation and stability of the TMV. b. *In situ* assembly of the TMV nanotube onto a rigid DNA origami structure. c. The RR-TMV disks functionalized with DOX) and PEG for improved solubility. Adapted with permission from ref [39], [40] and [41].

Figure 1.7. Fluorescently engineered HBV capsids. a) Encapsulation of FPs surface-display of cancer-targeting peptides. b) Linker-length dependent localization of fluorophores. c) Construction of double-layered fluorescent protein nanoparticle with cancer-targeting capability. Adapted with permission from ref [50].

Figure 1.8. a. Encapsulation of toxic protein (HIV protease) in an engineered variant of lumazine synthase. b. Encapsulating active enzymes in a protein cage. GFP (+36) (green) serves as a directing tag to load appended cargo proteins (blue) into empty AaLS-13 capsids. c. Electrostatic interactions-based encapsulation of cargo proteins by engineered AaLS cages. d. Design and structure of a nucleic-acid packaging protein capsule cpAaLS-RX possessing an oligoarginine tag (RX). Adapted with permission from ref [52], [53], [54], and [55].

Figure 1.9. a. Representative figure of AP-1 peptide inserted ferritin subunit and assembled AP1-PBNCs. b. the formation of the condensed structure of the poly-siRNA-proteinticle complex. Adapted with permission from ref [69] and [70].

Figure 1.10. Formation of ferritin-MOFs. Adapted with permission from ref [71].

Figure 1.11. Preparation of vault-pNIPAAm conjugates and thermo-responsive behavior. Adapted with permission from ref [80]

Figure 1.12. a. Overview of the computational methodology for the construction of the tetrahedral cage from two distinct trimeric building blocks. b. computational methodology for the construction of the Icosahedral cage from the trimeric building block. c. Overview of the construction of icosahedral cages in three different ways using pentamers+trimer, pentamers+dimers, and trimer+dimmers. Adapted with permission from ref [90], [91], and [92].

Figure 1.13. Overview of the two-component protein nanoparticle vaccine for the respiratory syncytial virus. Adapted with permission from ref [93] ,[94]

Figure 1.14. a. A general strategy for creating protein assemblies (symmetric nanostructures) by fusing dimeric protein and a trimeric protein. b. Atomic structure of a twelve subunit-tetrahedral protein cage by fusing two naturally oligomeric protein domains. Adapted with permission from ref [87] and [97].

Figure 1.15. a. Schematic construction of protein assemblies construction by fusing WA20 dimer and foldon trimer. b. Schematic construction of protein nanoparticle by fusing LSm (PDB 3BY7) and MyoX-Coil (PDB 2LW9). Adapted with permission from ref [88] and [100].

Figure 1.16. a. Design of a self-assembling octahedral protein cage from of the trimeric esterase (PDB 1ZOI) and the tetrameric coiled-coil (PDB 3R4A). b. Design of protein cages of different geometries by fusing the trimeric protein through a flexible linker to coiled coils with different oligomerization states. Adapted with permission from ref [89] and [103].

Figure 1.17. Cartoon illustrating the metal-induced trimerization of Tet8-M fusion protein. Adapted with permission from ref [104].

Figure 1.18. a. Design of the self-assembling polypeptide tetrahedron (Arrows denote the orientation of the coiled-coil elements in the assembly). b. Design of two tetrahedral structures based on protein origami. Adapted with permission from ref [105] and [106].

Figure 1.19. a. The self-assembly of recombinant protein amphiphiles (ZR-ELP, mCherry-ZE, and EGFP-ZE) into vesicles. b. Schematic representation of a growth factor-tethered protein nanoparticle via coiled-coil formation. c. Temperature responsive 2D networks of an elastomeric polymer created using helical zippers. Adapted with permission from ref [107,108,109] and [110].

Figure 1.20. a. CB8 induced a 2D sheet assembly of a hexameric hemoprotein. b. *N*-(1-pyrenyl) maleimide induced supramolecular dimerization of the hexameric hemoprotein. c. PNIPAAm induced thermoresponsive micellar assembly formation of the hexameric hemoprotein. Adapted with permission from ref [112,113] and [114].

Figure 1.21. a. Three state behavior of protein-ELPBC fusion, unimer, micelle, and a polydisperse micrometer-sized aggregate. b. designed mCherry-ELP fusion block copolymers using two different ELP sequences (ELP0 and ELP1), showing different SAXS profiles. Adapted with permission from ref [117] and ref [118].

Figure 1.22. a. Schematic representation of the supramolecular hemoprotein linear polymerization. b. Schematic representation of 2D network of hemoprotein. Adapted with permission from ref [119] and ref.[120]

Figure 1.23. a. Schematic representation of the supramolecular myoglobin linear polymerization. b. Schematic representation of heterotropic protein-protein complexation

between homotetrameric streptavidin and dimeric apomyoglobin. Adapted with permission from ref[121] and ref[122].

Figure 1.24. a. Cartoon representation of maleimide-tethered heme showing self-assembled states: linear oligomer and ring-shaped trimer. Adapted with permission from ref [123].

Figure 1.25. a. cyclic protein assembly of ecDHFR2 and MTX2-C9. b. synthetic scheme of linear and cyclic forms of the CS dimer, CS-SC. Adapted with permission from ref [125] and ref[126].

Figure 1.26. a. GFP (nano) polygons were achieved by connecting the 11th β strand of GFP to the N terminus of the rest of GFP (1–10 β strands) using a short peptide linker. b. Cartoon representation of genetically incorporation of a hydrophobic domain at the C-terminus of the IDP, and self-assembly around an artificial core. Adapted with permission from ref [127] and [128].

Figure 1.27. a. Synthesis of maleimide-functionalized pNIPAAm and its bioconjugation to mCherry S131C to create a protein-polymer diblock copolymer. b. hMb-PNIPAM bioconjugates synthesized using a thiol-maleimide coupling. Adapted with permission from ref [135] and [140].

Figure 1.28. a. Streptavidin-initiated polymerization of NIPAAm. b. Schematic representation of in situ growth of HSA-PHPMA nanostructures. Adapted with permission from ref [143] and [145].

Figure 1.29. a. Synthesis of linear (top) and Y-shaped (bottom) GFP–block copolymer conjugates and their proposed micellar assembly. b. synthetic strategy for sfGFP-PEGMA bioconjugates and its temperature-dependent aggregation. Adapted with permission from ref [148] and [149].

Figure 1.30. a. protein-polymer nano-conjugates (BSA-NH2/PNIPAAm) base proteinosomes generation protocol. b. Proteinosomes based on self-assembly of a

supramolecular bioconjugate of βCD modified BSA and Ad-PNIPAM. Adapted with permission from ref [152] and [153].

List of figures in chapter 2

List of figures in chapter 2.1.

Figure 2.1. Process of converting monomeric protein into a complex.

Figure 2.2. Crystal structure of fungal and bacterial hydrophobins (PDB accession: 2B97 and 4BHU) and designed FAP consists of globular water-soluble protein and hydrophobic tail connected by the hydrophilic linker (hydrophilic and hydrophobic faces are shown in blue and red, respectively).

Figure 2.3. The catalytic triad of a serine protease and its reacting mechanism with the FP probe.

Figure 2.4. Process of converting monomeric protein into a complex.

Figure 2.5. Molecular weight of native trypsin with optimized protocol.

Figure 2.6. Structure comparison of tritons and tweens with respect to AABP.

Figure 2.7. Proposed scheme of MAPLab technology (approach iii). The first step involves solubilization or mixed micellization of the probe in triton X-100. In this state, the FP group of the probe will available for labeling. In the next step, we are adding protein solution to this micellar solution to do the conjugation. After surfactant screening, we have optimised MAPLab technology, Triton X-100 was found to be best, as it satisfies all the requirements for protein labelling reaction and we did the test scale reaction.

Figure 2.8. MALDI-TOF MS analysis of reaction mixture (MAPLab, approach iii).

Figure 2.9. Cloud point based separation of native and conjugate.

Figure 2.10. First step of three-step purification for separating protein conjugate from reaction mixtures. Representative IEX chromatogram showing the separation of triton X-100 form protein mixture (native and modified protein).

Figure 2.11. Representative SEC chromatogram showing for separation of protein conjugate from native. The modified protein appears larger in size as evident from elution volume, because of complex formation.

Figure 2.12. Third step of three-step purification for separating salt from protein conjugate. Representative desalting chromatogram showing the removal of salt.

Figure 2.13. MALDI-TOF MS characterization of Try-OEG-C123T conjugate after purification.

Figure 2.14. Flouroscence characterization of Try-OEG-C123T conjugate after purification.

Figure 2.15. CD characterization of Try-OEG-C123T conjugate after purification.

Figure 2.16. Zeta potential characterization of Try-OEG-C123T conjugate after purification.

Figure 2.17. SEC data of Try-OEG-C123T conjugate.

Figure 2.18. DLS data of Try-OEG-C123T conjugate.

Figure 2.19. SANS data of Try-OEG-C123T conjugate.

Figure 2.20. SEC-MALS data of Try-OEG-C123T conjugate.

Figure 2.21. SEC based CMC estimation of Try-OEG-C123T.

Figure 2.22. Summary of chemical methodology to make FAPs. i. The AABP was solubilised in the presence of triton X-100, ii. MAPLab. iii. Three-step purification to remove triton X-100 and the unreacted protein using IEX, SEC and desalting. iv. Self-assembly of FAP**.**

Figure 2.23. Crystal structures of serine proteases used in this study (Try, Chy, Sub, ProKP(DB: 4MTB, 1GL0, 4C3V, 5B1D respectively) and Chemical structures of hydrophilic linker and hydrophobic tails.

List of figures in chapter 2.2.

Figure 2.24. Overlaid crystal structures of serine proteases used in this study a. Try, and Chy (4MTB, 1GLO respectively), b. Sub and ProK (4C3V, 5B1D respectively)

Figure 2.25. Structures of protein conjugates (head variant Library).

Figure 2.26. MALDI-TOF MS characterization of native proteins (Try, Chy, ProK and Sub).

Figure 2.27. SANS for native proteins (Try and Chy).

Figure 2.28. SANS for native proteins (ProK and Sub).

Figure 2.29. Zeta potential measurement of native proteins (Try, Chy, ProK and Sub).

Figure 2.30. Electrostatic surface potential of native proteins. Negatively charged residues like Asp and Glu are colored red. Positively charged residues like Arg and Lys are colored blue. His residues are colored green and active site Ser alone is represented in magenta (site of conjugation).

Figure 2.31. MALDI-TOF MS characterization of protein variant library before purification (reaction mixture).

Figure 2.32. MALDI-TOF MS characterization of protein variant library after purification.

Figure 2.33. Fluorescence characterization of protein variant library after purification.

Figure 2.34. CD characterization of protein variant library after purification.

Figure 2.35. Zeta potential characterization of protein variant library after purification.

Figure 2.36. SEC data of protein variant library after purification.

Figure 2.37. DLS data of protein variant library after purification.

Figure 2.38. SANS data of protein variant library after purification.

Figure 2.39. SEC-MALS data of Try-OEG-C12-3T of protein variant library.

Figure 2.40. Cartoon representation of size evolution of particles with pH.

Figure 2.41. Size evolution of particles with pH and salt as per SEC data.

Figure 2.42. Cartoon representation of protein complexes in protein variant library.

List of figures in chapter 2.3.

Figure 2.43. Structures of protein conjugates in spacer variant library.

Figure 2.44. MALDI-TOF MS characterization of spacer variant library before purification (reaction mixture).

Figure 2.45. MALDI-TOF MS characterization of spacer variant library after purification.

Figure 2.46. Fluorescence characterization of spacer variant library, after purification.

Figure 2.47. CD characterization of spacer variant library, after purification.

Figure 2.48. Zeta potential characterization of spacer variant library after purification.

Figure 2.49. SEC data of spacer variant library after purification.

Figure 2.50. DLS data of spacer variant library after purification.

Figure 2.51. SANS data of spacer variant library after purification.

Figure 2.52. SEC-MALS data of spacer variant library after purification.

Figure 2.53. Cartoon representation of protein complexes in spacer variant library.

List of figures in chapter 2.4.

Figure 2.54. Structures of conjugates in tail variant library.

Figure 2.55. MALDI-TOF MS characterization of C6-tail variant library before purification (reaction mixtures).

Figure 2.56. MALDI-TOF MS characterization of C12-tail variant library before purification (reaction mixtures).

Figure 2.57. MALDI-TOF MS characterization of C18-tail variant library before purification (reaction mixtures).

Figure 2.58. MALDI-TOF MS characterization of C12-tail variant library after purification.

Figure 2.59. MALDI-TOF MS characterization of C6, C18-tail variant libraries after purification.

Figure 2.60. Fluorescence characterization of tail variant library after purification.

Figure 2.61. CD characterization of tail variant library after purification.

Figure 2.62. Zeta potential characterization of spacer variant library after purification.

Figure 2.63. SEC data of tail variant library.

Figure 2.64. SEC data of FAPs that did not assemble (Try-OEG-C6-3T, Try-OEG-C12-1T) (Spacer variant library).

Figure 2.65. DLS data of tail variant library after purification.

Figure 2.66. SANS data of tail variant library after purification.

Figure 2.67. SEC-MALS data of tail variant library after purification.

Figure 2.68. Cartoon representation of protein complexes (tail variant library).

Figure 2.69. Tail variants library with same number of carbons but different branching.

List of figures in chapter 3

List of figures in chapter 3.1.

Figure 3.1. Cartoon representation of photo-responsive complex assembly and dis-assembly.

Figure 3.2. Structure of photo-responsive protein conjugate.

Figure 3.3. MALDI-TOF MS analysis of reaction mixture.

Figure 3.4. MALDI-TOF MS spectrum of purified photo-responsive FAP.

Figure 3.5. Cartoon representation of self-assembly of photo-responsive FAP.

Figure 3.6. DLS data of photo-responsive FAP after purification.

Figure 3.7. SEC data of photo-responsive FAP after purification (Superdex-200).

Figure 3.8. SEC-MALS data of photo-responsive FAP after purification (Superose 6).

Figure 3.9. Cartoon representation of dis-assembly process of photo-responsive protein complex.

Figure 3.10. Programmed time-dependent irreversible dis-assembly of photo-responsive protein complex followed using SEC (Superdex-200).

Figure 3.11. Time-dependent MALDI-TOF MS studies of programmed dis-assembly of photo-responsive protein complex showing systematic decrease in the intensity of conjugate and increase in cleaved conjugate.

Figure 3.12. Graphic representation of Time-dependent molecular weight changes of conjugate and cleaved conjugate as per MALDI-TOF MS studies.

Figure 3.13. MALDI-TOF MS data of protein conjugate before and after 90 mins of UV exposure.

Figure 3.14. DLS data for programmed dis-assembly of photo-responsive protein complex before and after cleavage.

Figure 3.15. SEC data obtained for Photo-responsive Try-OEG-NB-C12-2T in response to the visible light. (i) Black trace - not exposed to visible light (ii) red trace - exposed to light for 90 min (Superdex-200).

Figure 3.16. SEC data obtained for protein conjugate without 2-NB group, Try-OEG-C12-3T (control) in response to the UV light. (i) Black trace - not exposed to UV light (ii) Red trace - exposed to UV light for 90 min (Superdex-200).

List of figures in chapter 3.2.

Figure 3.17. Cartoon representation of assembly and dis-assembly of redox-responsive conjugate.

Figure 3.18. Chemical Structure of redox-responsive protein conjugate.

Figure 3.19. MALDI-TOF MS analysis of reaction mixture.

Figure 3.20. MALDI-TOF MS spectrum of purified conjugate.

Figure 3.21. Cartoon representation of self-assembly of photo-responsive protein conjugate into a complex.

Fig 3.22. DLS data of redox-responsive FAP after purification.

Figure 3.23. SEC data of redox-responsive FAP after purification (Superose 6).

Figure 3.24. SEC-MALS data of redox-responsive FAP after purification (Superose 6).

Figure 3.25. Cartoon representation of dis-assembly process of redox-responsive protein complex.

Figure 3.26. Equivalent-dependent dis-assembly study for Try-OEG-SS-C12-2T complex followed using SEC (Superose 6).

Figure 3.27. Time-dependent dis-assembly study for Try-OEG-SS-C12-2T complex followed using SEC (Superose 6).

Figure 3.28. MALDI-TOF MS characterization of dis-assembly, molecular weight changes as seen for conjugate before and after cleavage.

Figure 3.29. Structure of native Try.

Figure 3.30. SEC chromatogram of native Try (black) DTT treated Try (red) in SEC (Superose 6).

Figure 3.31. Structure of inhibited Try with FP-half probe.

Figure 3.32. Try-OEG and (black) DTT treated try-OEG in SEC (Superose 6).

Figure 3.33. Structure of Try conjugate with FP- full probe.

Figure 3.34. Try-OEG-C12-2T (black) and DTT treated Try-OEG-C12-2T respectively in SEC (Superose 6).

List of figures in chapter 3.3.

Figure 3.35. Cartoon representation of design, synthesis and self-assembly of peptide-based FAPs.

Figure 3.36. Structures of peptide-based FAPs in protein variant library.

Figure 3.37. Structures of peptide-based FAPs in peptide variant library.

Figure 3.38. MALDI-TOF MS analysis of purified FAPs in protein variant library after purification.

Figure 3.39. Fluorescence studies for purified peptide-based FAPs in protein variant library after purification.

Figure 3.40. CD studies for purified peptide-based FAPs in protein variant library after purification.

Figure 3.41. SEC data of peptide-based FAPs in protein variant library after purification.

Figure 3.42. DLS data of peptide-based FAPs in protein variant library after purification.

Figure 3.43. SEC-MALS data of peptide-based FAPs in protein variant library after purification.

Figure 3.44. SEC data of peptide-based FAPs in peptide variant library after purification.

List of figures in chapter 4

List of figures in chapter 4.1.

Figure 4.1. Relative abundance and uniqueness of amino acids.

Figure 4.2. MAPLab Technology: Conjugation steps and mechanism.

Figure 4.3. MALDI-TOF MS data of native and reaction mixtures of BSA.

Figure 4.4. MALDI-TOF MS data of native and reaction mixtures of Chy.

Figure 4.5. MALDI-TOF MS data of native and reaction mixtures of LZ.

Figure 4.6. MALDI-TOF MS data of native and reaction mixtures of GFP.

Figure 4.7. MALDI-TOF MS spectrum of purified Chy conjugate.

Figure 4.8. MALDI-TOF MS spectrum of purified BSA conjugate.

Figure 4.9. Activity of Chy Vs Chy-OEG-C181T conjugate.

Figure 4.10. Chy (PDB accession: 1YPH) structure showing Active site Ser residue (yellow) with respect to N-terminus location (green).

Figure 4.11. SEC data of IEX purified FAPs native and conjugate.

Figure 4.12. DLS data of SEC purified FAPs.

List of figures in chapter 4.2.

Figure 4.13. MAPLab technology: Conjugation steps and mechanism.

Figure 4.14. MALDI-TOF MS data of native and reaction mixtures of BSA.

Figure 4.15. MALDI-TOF MS data of reaction mixture of BSA-Thiol Conjugate.

Figure 4.16. SEC data of IEX purified BSA thiol conjugate with respect to N-terminus conjugate.

Figure 4.17. BSA (PDB accession: 3V03) structure showing thiol residue (yellow) with respect to N-terminus location (green). Negatively charged residues like Asp and Glu are colored red. Positively charged residues like Arg and Lys are colored blue.

Appendix-III: List of Schemes

List of schemes in chapter 2

List of schemes in chapter 2.1

Scheme 2.1. Proposed step by step conjugation to make FAP as per approach i.

Scheme 2.2. Synthesis of the octaethylene glycol hydrophilic spacer

Scheme 2.3. Synthesis of the C12-3T hydrophobic tail.

Scheme 2.4. Attempted step by step conjugation to make FAP as per approach i.

Scheme 2.5. Proposed conjugation to make FAP as per approach ii.

Scheme 2.6. Synthesis of the amphiphilic probe.

Scheme 2.7. Attempted conjugation of amphiphilic probe with co-solvents.

Scheme 2.8. Proposed conjugation strategy to make FAP (MAPLab Technology, approach iii).

List of schemes in chapter 2.2

Scheme 2.9. Synthesis of the OEG spacer.

Scheme 2.10. Synthesis of C12-3T hydrophobic azide.

Scheme 2.11. Synthesis of amphiphilic probe for protein variant library.

Scheme 2.12. General conjugation scheme for making protein variant library.

List of schemes in chapter 2.3

Scheme 2.13. General scheme for the synthesis of DEG or TEG spacers

Scheme 2.14. General scheme for the synthesis of OEG, DDEG and CEG spacers.

Scheme 2.15. Synthesis of C123T hydrophobic azide.

Scheme 2.16. Synthesis of amphiphilic probe for making spacer variant library.

Scheme 2.17. General conjugation scheme for making spacer variant library.

List of schemes in chapter 2.4

Scheme 2.18. Synthesis of the OEG hydrophilic spacer

Scheme 2.19. General scheme for synthesis of 3-Tail hydrophobic tail.

Scheme 2.20. General scheme for synthesis of 2-Tail hydrophobic tail

Scheme 2.21. General scheme for synthesis of 1-Tail hydrophobic tail

Scheme 2.22. Synthesis of amphiphilic probe for making tail variant library.

Scheme 2.23. General conjugation scheme for making tail variant library.

List of schemes in chapter 3

List of schemes in chapter 3.1

Scheme 3.1. Synthesis of amine-terminated OEG spacer.

Scheme 3.2. Synthesis of photo-responsive AABP.

Scheme 3.3. Synthesis of photo-responsive FAP (Try-OEG-NB-C12-2T).

List of schemes in chapter 3.2

Scheme 3.4. Synthesis of amine-terminated OEG spacer.

Scheme 3.5. Synthesis of redox-responsive AABP.

Scheme 3.6. Synthesis of redox-responsive FAP (Try-OEG-SS-C12-2T).

List of schemes in chapter 3.3

Scheme 3.7. Synthesis of alcohol-terminated OEG spacer.

Scheme 3.8. Synthesis of Boc-protected hydrophobic peptides with free acid functionality.

Scheme 3.9. General scheme for synthesis of peptide-based AABPs.

Scheme 3.10. General scheme for conjugation of peptide-based AABPs.

List of schemes n chapter 4

List of schemes n chapter 4.1

Scheme 4.1. Synthesis of the OEG hydrophilic spacer

Scheme 4.2. Synthesis of C181T hydrophobic tail.

Scheme 4.3. Synthesis of AABP for targeting N-terminus.

Scheme 4.4. Conjugation of amphiphilic probe to protein of interest at the N-terminus.

Scheme 4.5. Schematic representation of hydrolysis of SPNA

List of schemes n chapter 4.2

Scheme 4.6. Synthesis of the OEG hydrophilic spacer

Scheme 4.7. Synthesis of C181T hydrophobic tail

Scheme 4.8. Synthesis of AABP for targeting thiol residue.

Scheme 4.9. Conjugation of amphiphilic probe to protein of interest (BSA) at the thiol residue.

Appendix-IV: List of Tables

List of tables in chapter 2

Table 2.1. Summary of characterization data of prototype FAP.

Table 2.2. Summary of self-assembly data of prototype FAP.

Table 2.3. Summary of characterization data of native proteins.

Table 2.4. Summary of characterization data of protein variants

Table 2.5. Summary of self-assembly data of protein variants.

Table 2.6. Analysis of self-assembly data of protein variants.

Table 2.7. Summary of characterization data of spacer variants

Table 2.8. Summary of self-assembly data of spacer variants.

Table 2.9. Analysis of Self-assembly data of spacer variants.

Table 2.10. Summary of characterization data of tail variants.

Table 2.11. Summary of self-assembly data of tail variants.

Table 2.12. Analysis of self-assembly data of tail variants.

List of tables in chapter 3

Table 3.1. Summary of self-assembly data of the photo-responsive protein conjugate.

Table 3.2. Summary of self-assembly data of the redox-responsive protein conjugate.

Table 3.3. Summary of self-assembly data of the FAPs in protein variant library.

Table 3.4. Summary of self-assembly data of the FAPs in peptide variant library.

Table 3.5. Analysis of self-assembly data of peptide-based FAPs.

List of tables in chapter 4

Table 4.1. Summary of Self-assembly data of N-terminus FAPs (conjugates).

Table 4.2. Comparison of SEC data of N-terminus vs. thiol conjugate of BSA.

Appendix-V: List of Publications and Patents

List of Publications

1. Sandanaraj, B. S.*; **Reddy, M.M.;** Bhandari, P.; Kumar, S.; Aswal. V. Rational Design of Supramolecular Dynamic Protein Assemblies Using a Micelle-Assisted Activity-based Protein Labeling Technology. Chem. Eur. J., **2018,** 24, 16085-16096.

2. Sandanaraj, B. S.*; **Reddy, M. M.;** Jagan, k. ; Bhandari, P. J. Rational Design of Semi-Synthetic Protein Complexes with the Defined Oligomeric State. Chem. Select. **2019,** 4, 6397-6402.

3. Sandanaraj, B. S.*; **Reddy, M. M.;** Jagan, k. ; Bhandari, P. J; Lohote, A. B; Sahoo, B. Design, Synthesis and Self-assembly Studies of Suite of Monodisperse Facially Amphiphilic Protein-Dendron Conjugates. Chem Bio Chem. **2019,** Accepted Article.

4. Sandanaraj, B. S.*; **Reddy, M. M.;** Universal method for assembling any protein of interest. **2020,** Manuscript under preparation.

List of Patents

5. Sandanaraj, B. S.; **Reddy, M. M.;** Bhandari, P. J.; Jagan, k. Hydrophobin Mimics: Process for Preparation thereof. **US Patent, 2017, granted, 10188136.**

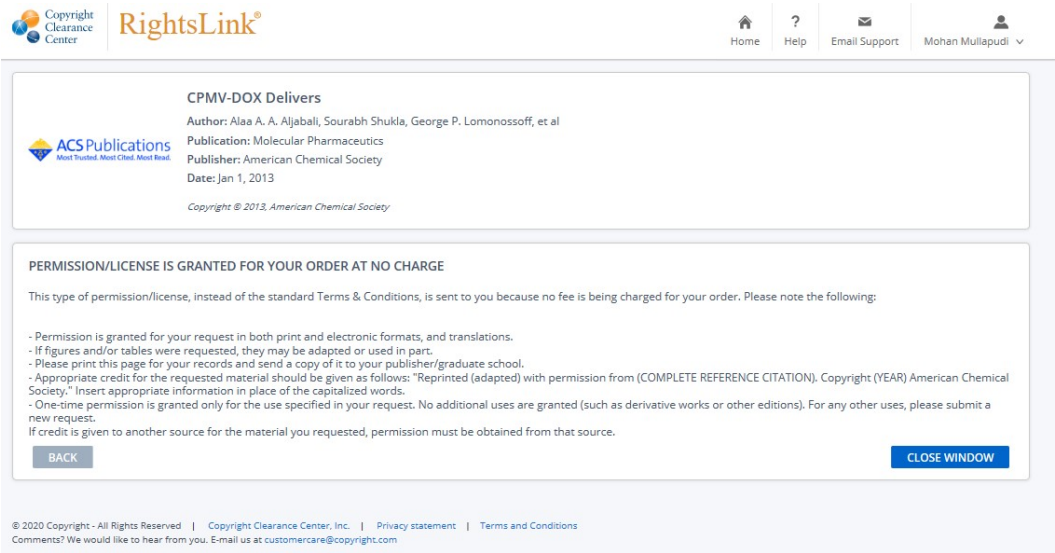
6. Sandanaraj, B. S.; Bhandari, P. J.; **Reddy, M. M.** Supramolecular Protein Assemblies with Advanced Functions and Synthesis thereof. **US Patent, 2018, published, 16146891**.

7. Sandanaraj, B. S.; **Reddy, M. M.** Chemical Method for Synthesis of Protein Amphiphiles. **US Patent, filed, 2019.**

8. Sandanaraj, B. S.; **Reddy, M. M.;** Bhandari. Generation-dependent Supramolecular Assemblies of Protein-dendron Conjugates. **US Patent, filed, 2020.**

Appendix-Vi: Acquired Copyrights from the Publishers

ACS articles:



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

CPMV-DOX Delivers

Author: Alaa A. A. Aljabali, Sourabh Shukla, George P. Lomonossoff, et al
Publication: Molecular Pharmaceutics
Publisher: American Chemical Society
Date: Jan 1, 2013

Copyright © 2013, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

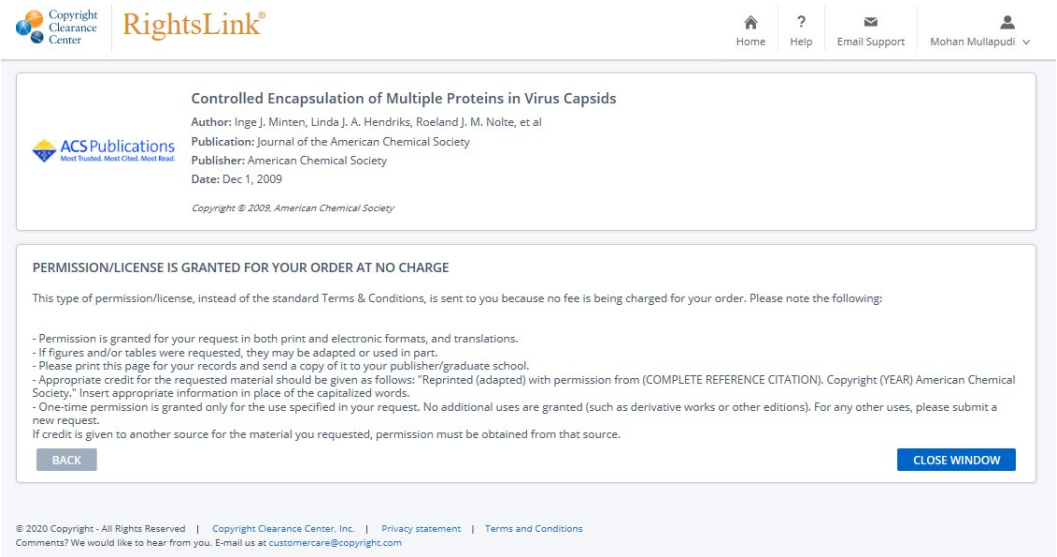
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

Controlled Encapsulation of Multiple Proteins in Virus Capsids

Author: Inge J. Minten, Linda J. A. Hendriks, Roeland J. M. Nolte, et al
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: Dec 1, 2009

Copyright © 2009, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

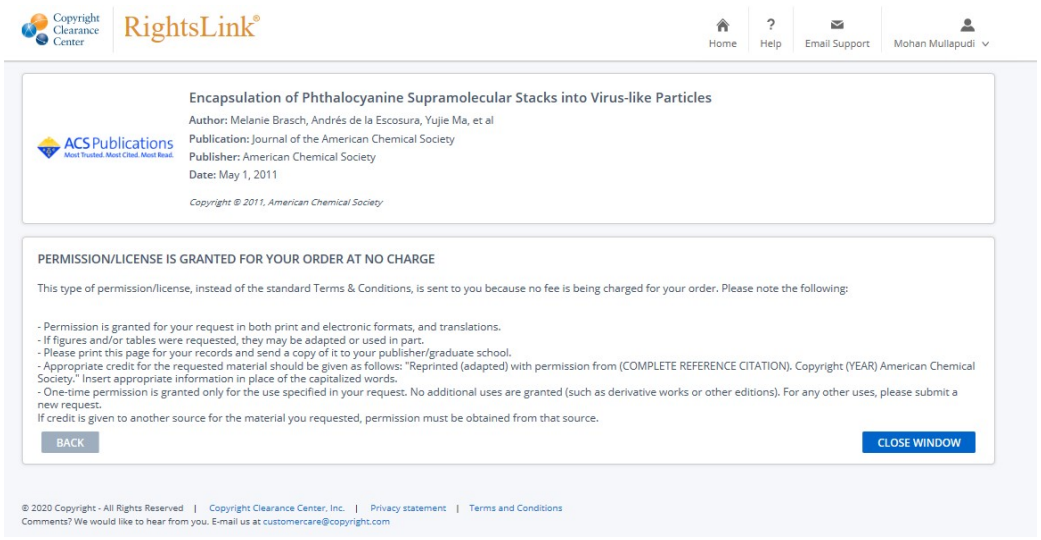
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications Most Trusted. Most Cited. Most Read.

Encapsulation of Phthalocyanine Supramolecular Stacks into Virus-like Particles

Author: Melanie Brasch, Andrés de la Escosura, Yujie Ma, et al
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: May 1, 2011

Copyright © 2011, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

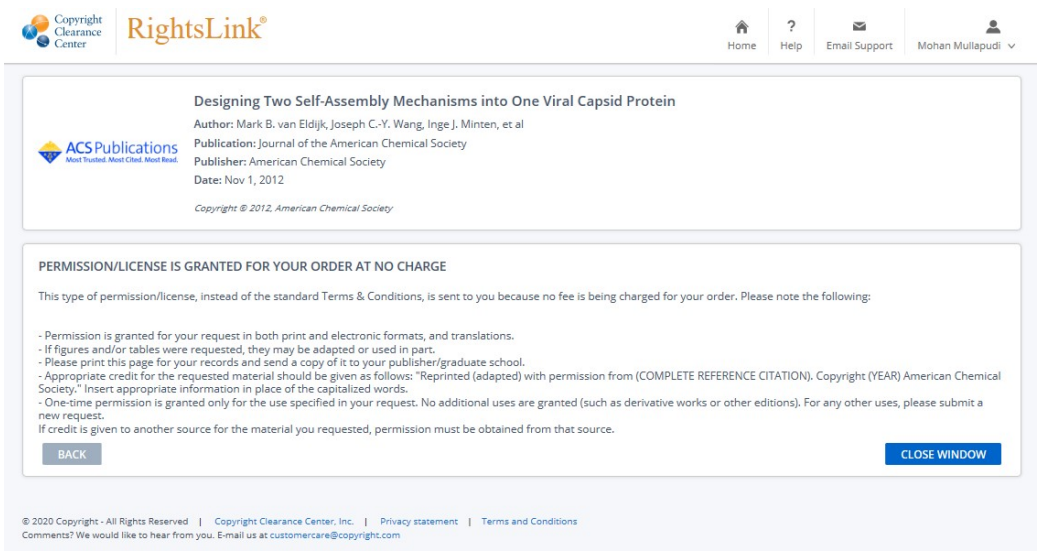
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications Most Trusted. Most Cited. Most Read.

Designing Two Self-Assembly Mechanisms into One Viral Capsid Protein

Author: Mark B. van Eldijk, Joseph C.-Y. Wang, Inge J. Minten, et al
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: Nov 1, 2012

Copyright © 2012, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

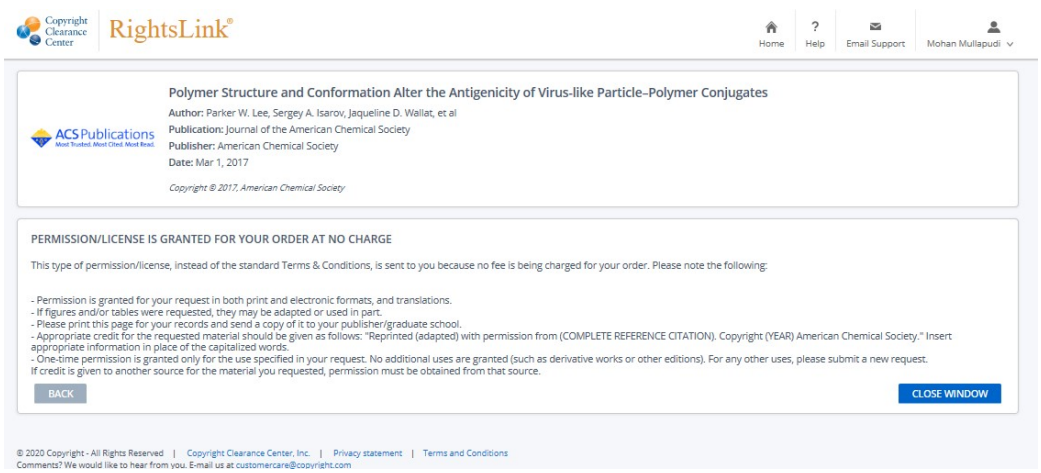
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications Most Trusted. Most Cited. Most Read.

Polymer Structure and Conformation Alter the Antigenicity of Virus-like Particle–Polymer Conjugates

Author: Parker W. Lee, Sergey A. Isarov, Jaqueline D. Wallat, et al
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: Mar 1, 2017

Copyright © 2017, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

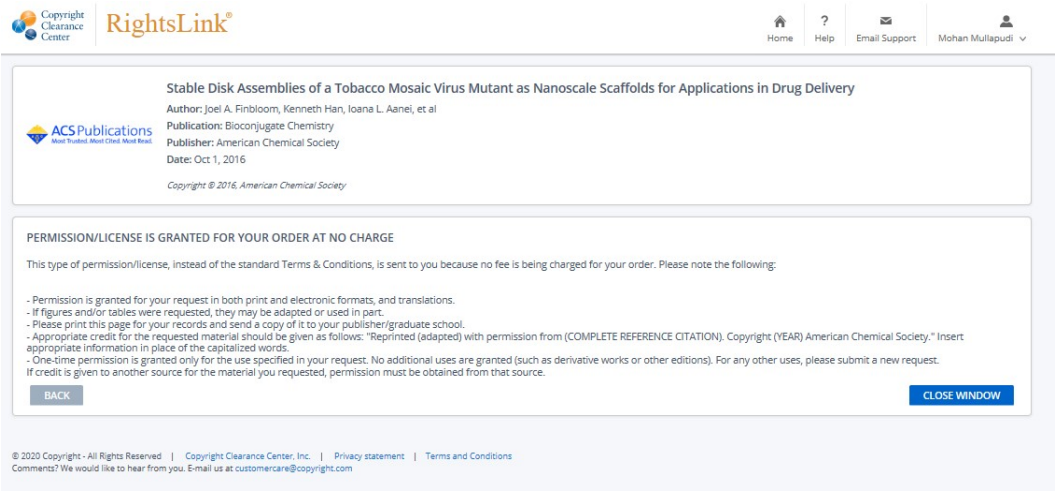
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications

Stable Disk Assemblies of a Tobacco Mosaic Virus Mutant as Nanoscale Scaffolds for Applications in Drug Delivery

Author: Joel A. Finbloom, Kenneth Han, Ioana L. Aanei, et al
Publication: Bioconjugate Chemistry
Publisher: American Chemical Society
Date: Oct 1, 2016

Copyright © 2016, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

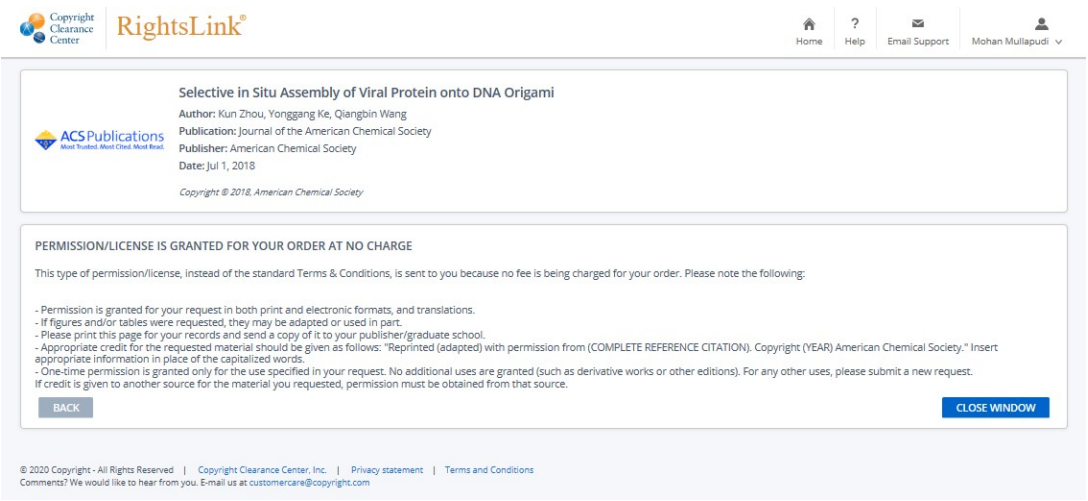
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications

Selective in Situ Assembly of Viral Protein onto DNA Origami

Author: Kun Zhou, Yonggang Ke, Qiangbin Wang
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: Jul 1, 2018

Copyright © 2018, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

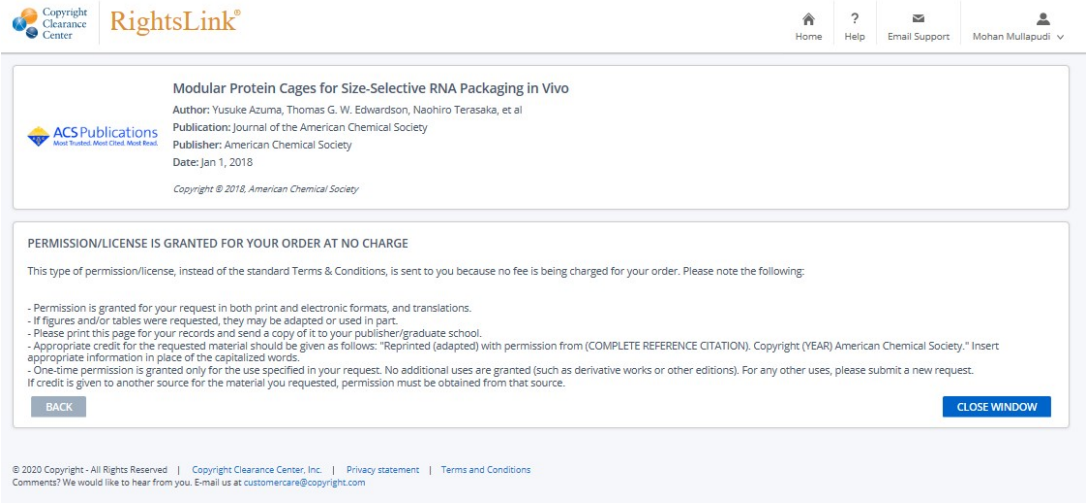
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications

Modular Protein Cages for Size-Selective RNA Packaging in Vivo

Author: Yusuke Azuma, Thomas G. W. Edwardson, Naohiro Terasaka, et al
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: Jan 1, 2018

Copyright © 2018, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

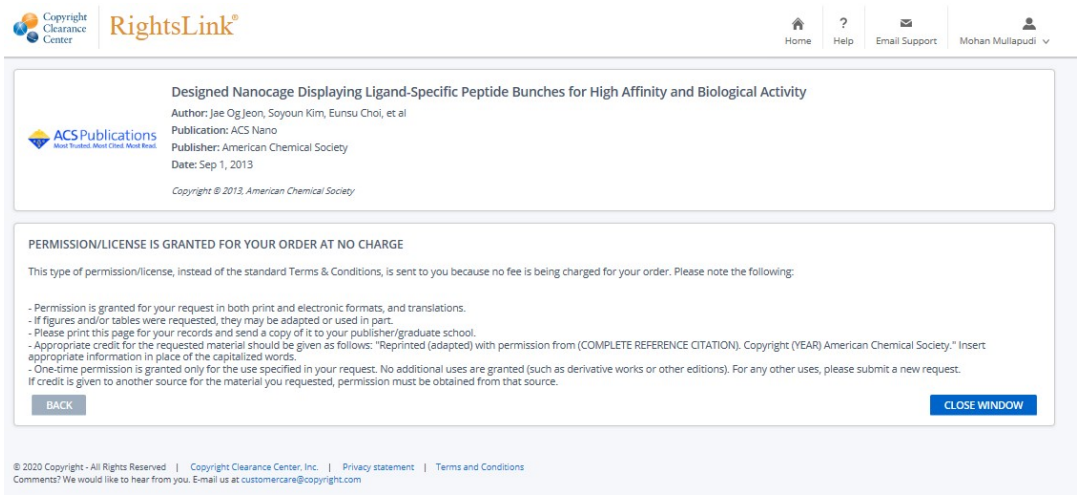
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications Most Trusted. Most Cited. Most Read.

Designed Nanocage Displaying Ligand-Specific Peptide Bunches for High Affinity and Biological Activity

Author: Jae Og Jeon, Soyoun Kim, Eunsu Choi, et al
Publication: ACS Nano
Publisher: American Chemical Society
Date: Sep 1, 2013

Copyright © 2013, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

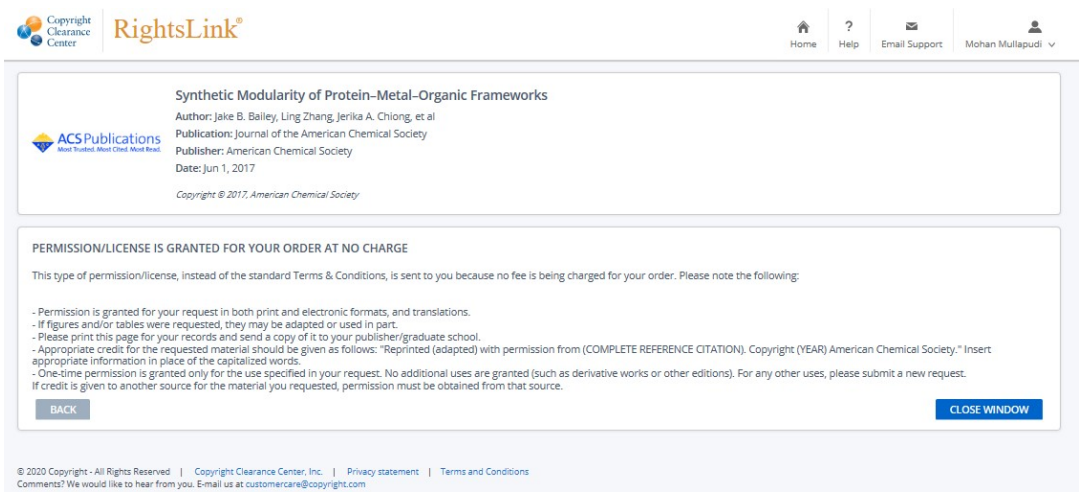
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications Most Trusted. Most Cited. Most Read.

Synthetic Modularity of Protein–Metal–Organic Frameworks

Author: Jake B. Bailey, Ling Zhang, Jerika A. Chiong, et al
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: Jun 1, 2017

Copyright © 2017, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

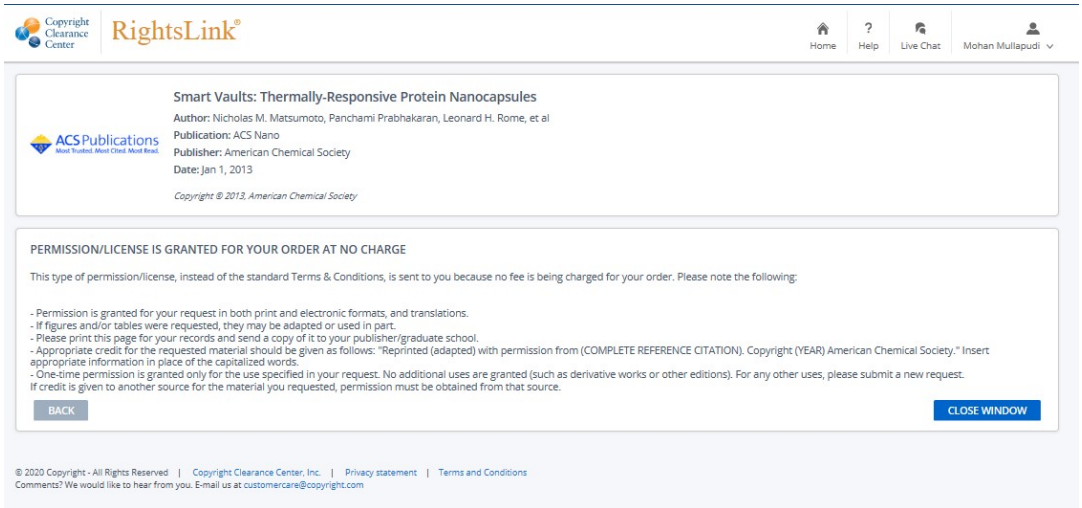
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center RightsLink®

Home | Help | Live Chat | Mohan Mullapudi

ACS Publications Most Trusted. Most Cited. Most Read.

Smart Vaults: Thermally-Responsive Protein Nanocapsules

Author: Nicholas M. Matsumoto, Panchami Prabhakaran, Leonard H. Rome, et al
Publication: ACS Nano
Publisher: American Chemical Society
Date: Jan 1, 2013

Copyright © 2013, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

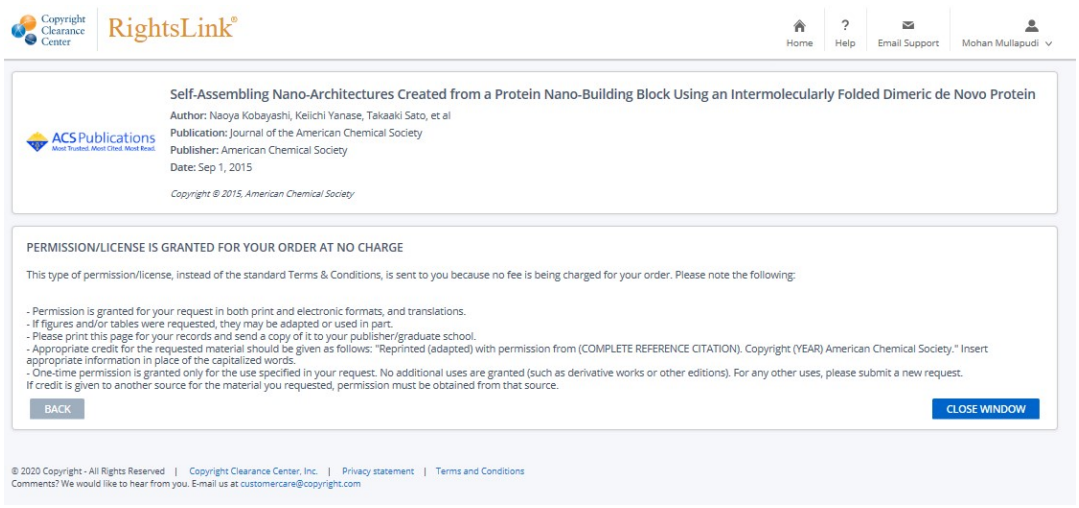
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric de Novo Protein

Author: Naoya Kobayashi, Keiichi Yanase, Takaaki Sato, et al
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: Sep 1, 2015

Copyright © 2015, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

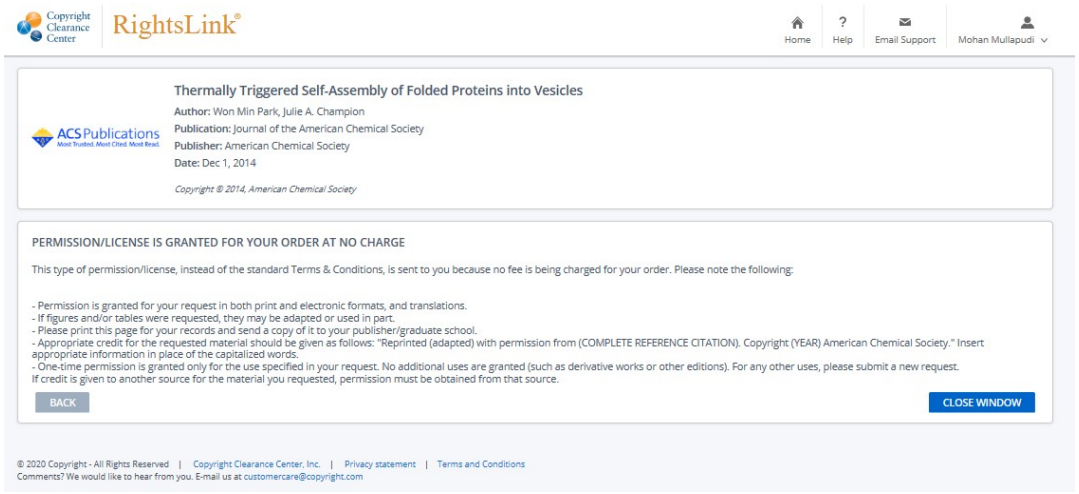
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

Thermally Triggered Self-Assembly of Folded Proteins into Vesicles

Author: Won Min Park, Julie A. Champion
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: Dec 1, 2014

Copyright © 2014, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

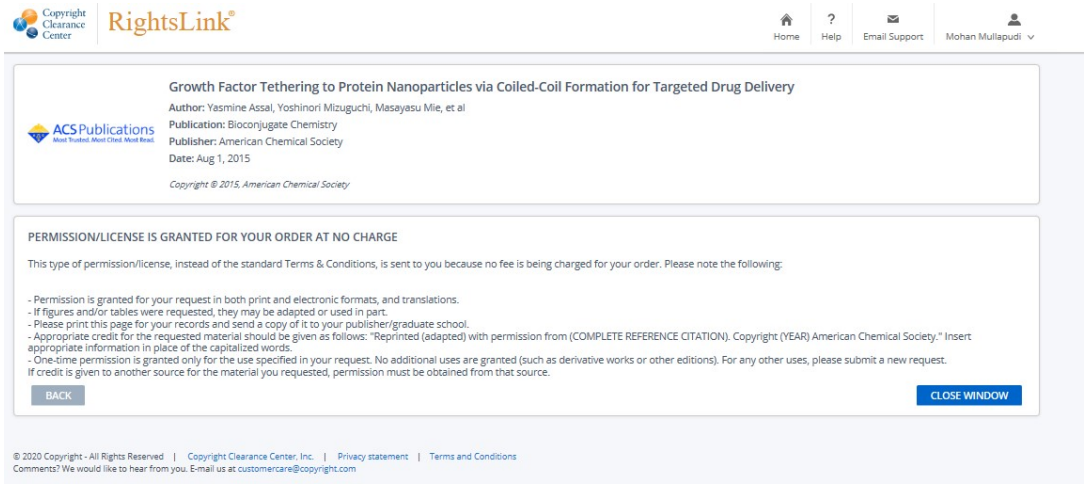
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

Growth Factor Tethering to Protein Nanoparticles via Coiled-Coil Formation for Targeted Drug Delivery

Author: Yasmine Assal, Yoshinori Mizuguchi, Masayasu Mie, et al
Publication: Bioconjugate Chemistry
Publisher: American Chemical Society
Date: Aug 1, 2015

Copyright © 2015, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

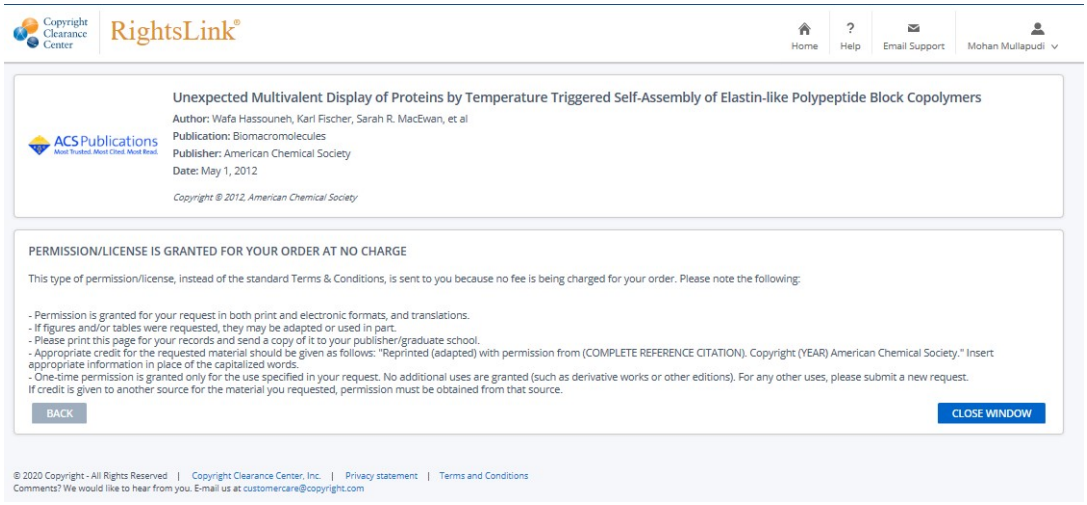
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

Unexpected Multivalent Display of Proteins by Temperature Triggered Self-Assembly of Elastin-like Polypeptide Block Copolymers

Author: Wafa Hassouneh, Karl Fischer, Sarah R. MacEwan, et al
Publication: Biomacromolecules
Publisher: American Chemical Society
Date: May 1, 2012

Copyright © 2012, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

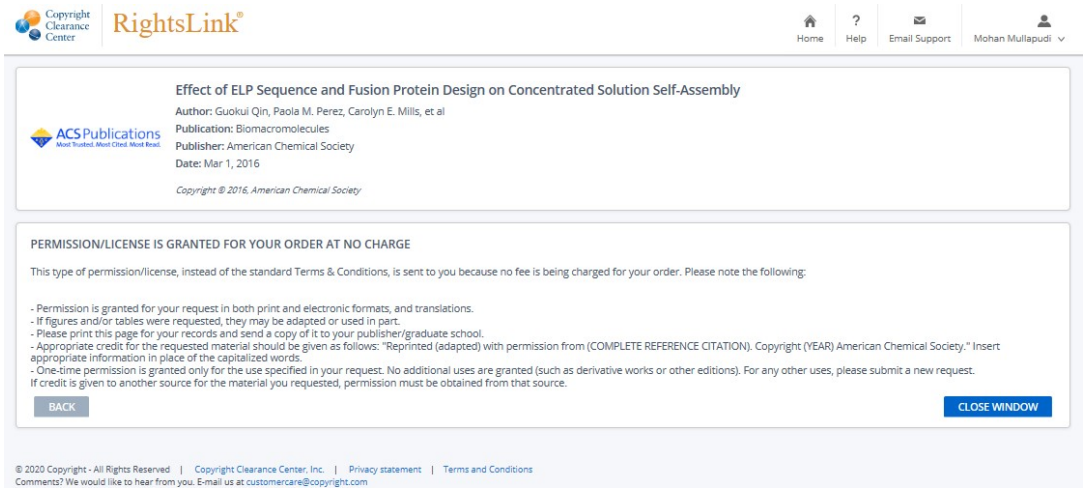
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

Effect of ELP Sequence and Fusion Protein Design on Concentrated Solution Self-Assembly

Author: Guokui Qin, Paola M. Perez, Carolyn E. Mills, et al
Publication: Biomacromolecules
Publisher: American Chemical Society
Date: Mar 1, 2016

Copyright © 2016, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

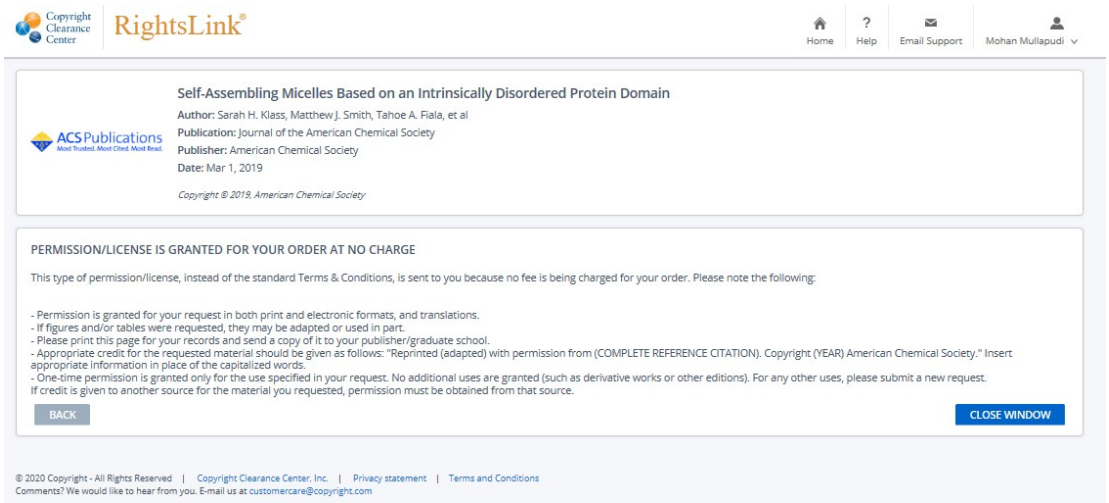
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

Self-Assembling Micelles Based on an Intrinsically Disordered Protein Domain

Author: Sarah H. Klass, Matthew J. Smith, Tahoe A. Fiala, et al
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: Mar 1, 2019

Copyright © 2019, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

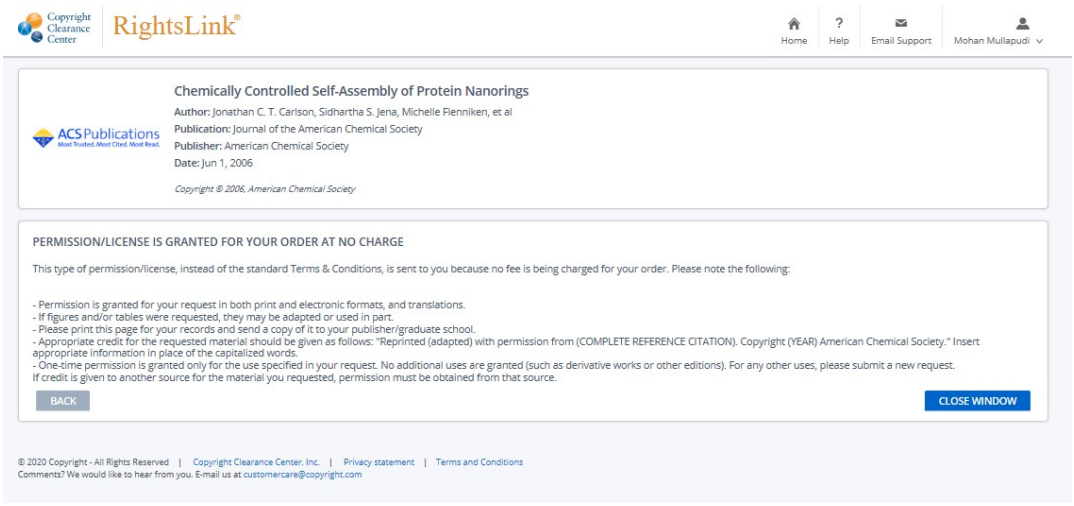
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

Chemically Controlled Self-Assembly of Protein Nanorings

Author: Jonathan C. T. Carlson, Sidhartha S. Jena, Michelle Flenniken, et al
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: Jun 1, 2006

Copyright © 2006, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

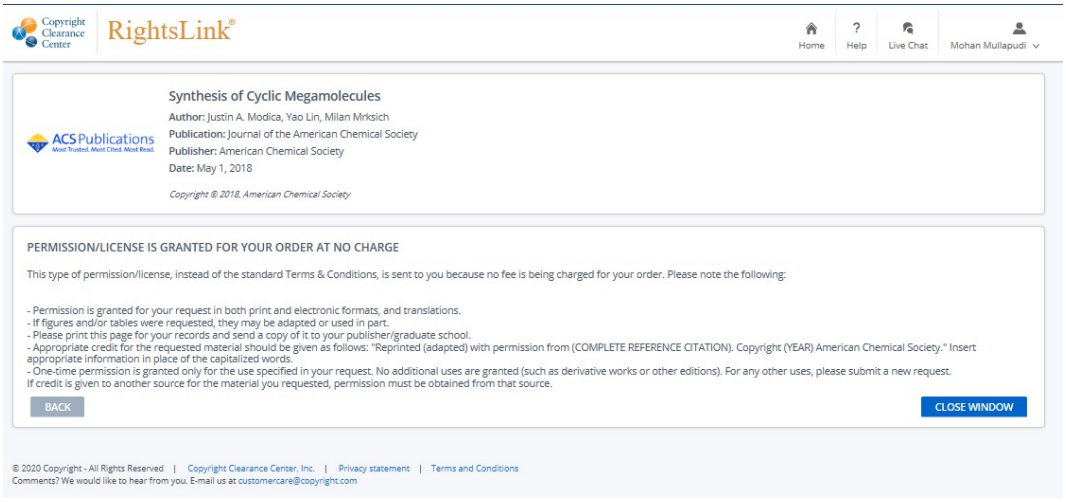
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Live Chat | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

Synthesis of Cyclic Megamolecules

Author: Justin A. Modica, Yao Lin, Milan Mrksich
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: May 1, 2018

Copyright © 2018, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

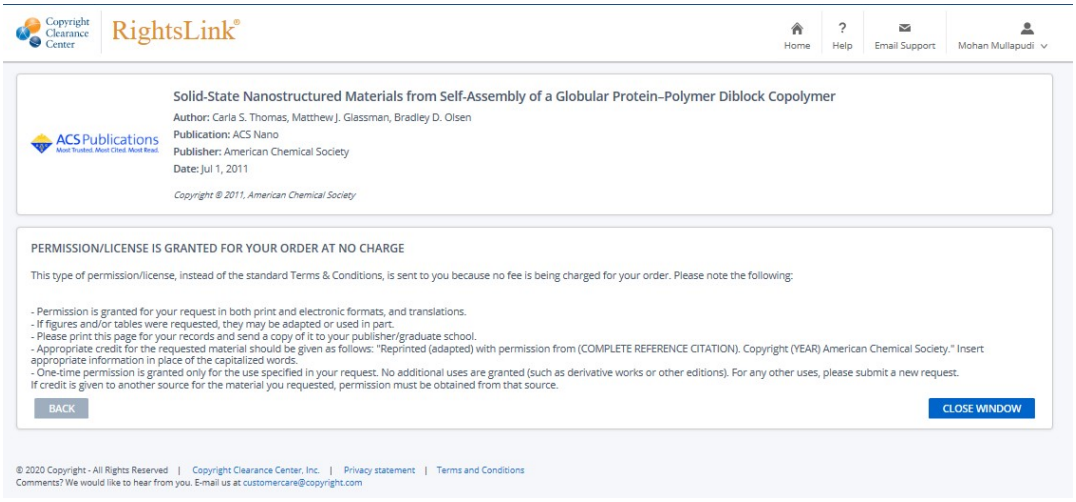
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

Solid-State Nanostructured Materials from Self-Assembly of a Globular Protein–Polymer Diblock Copolymer

Author: Carla S. Thomas, Matthew J. Glassman, Bradley D. Olsen
Publication: ACS Nano
Publisher: American Chemical Society
Date: Jul 1, 2011

Copyright © 2011, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

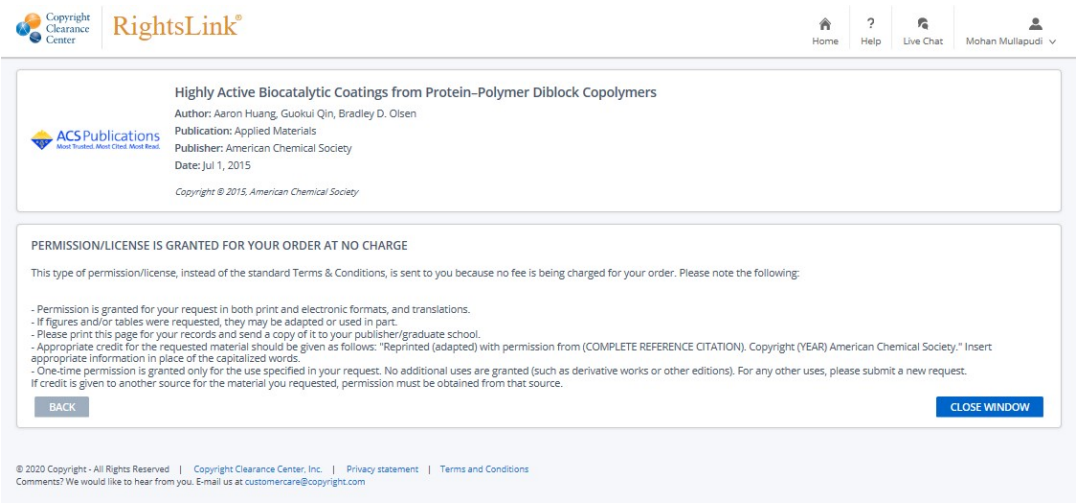
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Highly Active Biocatalytic Coatings from Protein–Polymer Diblock Copolymers

Author: Aaron Huang, Guokui Qin, Bradley D. Olsen
Publication: Applied Materials
Publisher: American Chemical Society
Date: Jul 1, 2015

Copyright © 2015, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

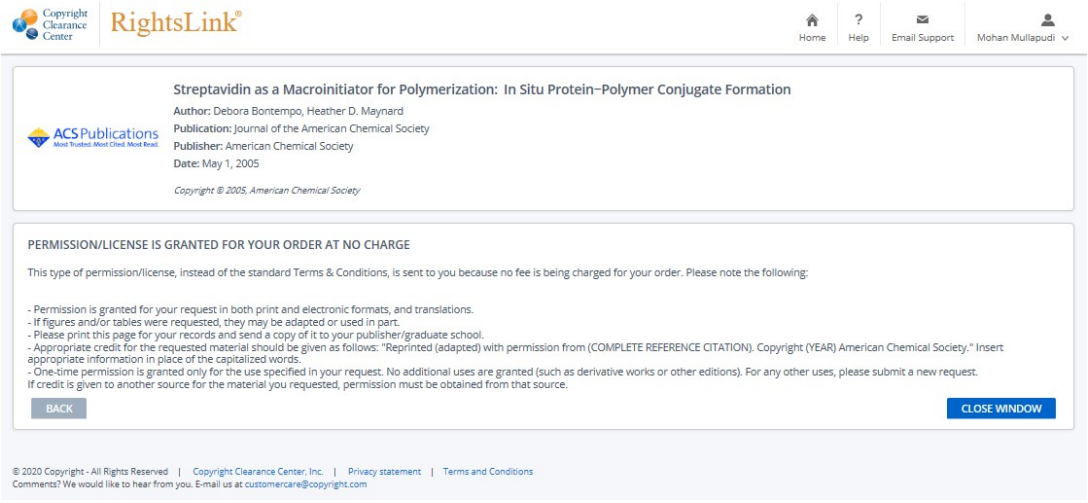
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Streptavidin as a Macroinitiator for Polymerization: In Situ Protein−Polymer Conjugate Formation

Author: Debora Bontempo, Heather D. Maynard
Publication: Journal of the American Chemical Society
Publisher: American Chemical Society
Date: May 1, 2005

Copyright © 2005, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

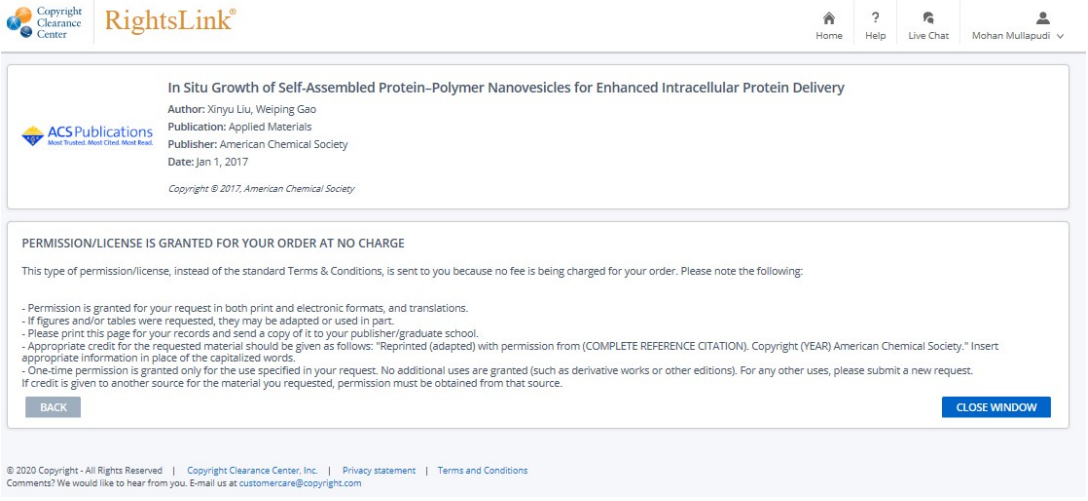
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



In Situ Growth of Self-Assembled Protein–Polymer Nanovesicles for Enhanced Intracellular Protein Delivery

Author: Xinyu Liu, Weiping Gao
Publication: Applied Materials
Publisher: American Chemical Society
Date: Jan 1, 2017

Copyright © 2017, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

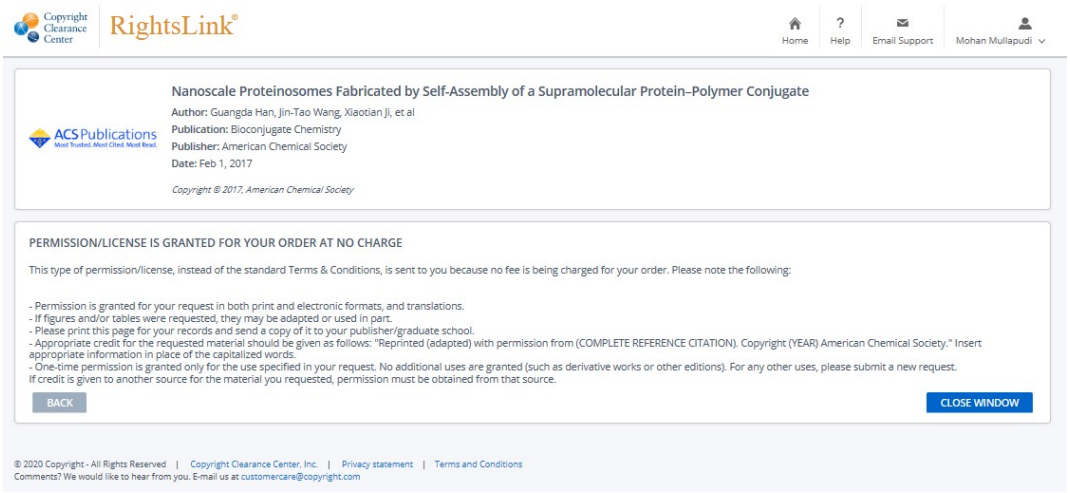
This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

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

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK CLOSE WINDOW

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

ACS Publications
Most Trusted. Most Cited. Most Read.

Nanoscale Proteinosomes Fabricated by Self-Assembly of a Supramolecular Protein–Polymer Conjugate

Author: Guangda Han, Jin-Tao Wang, Xiaotian Ji, et al
Publication: Bioconjugate Chemistry
Publisher: American Chemical Society
Date: Feb 1, 2017

Copyright © 2017, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

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

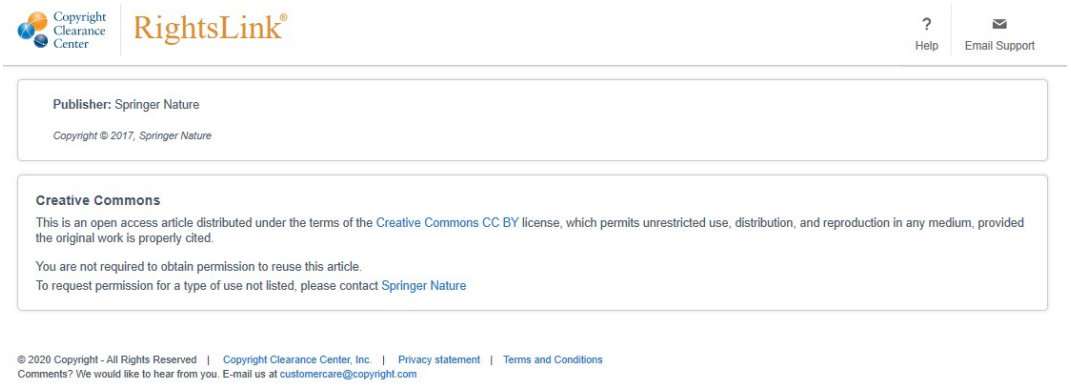
- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for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

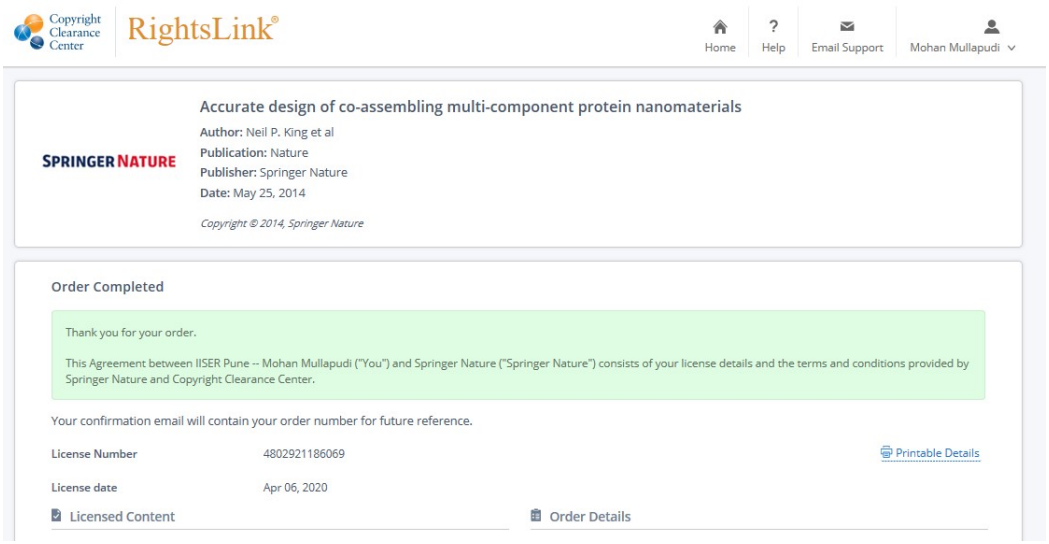
If credit is given to another source for the material you requested, permission must be obtained from that source.

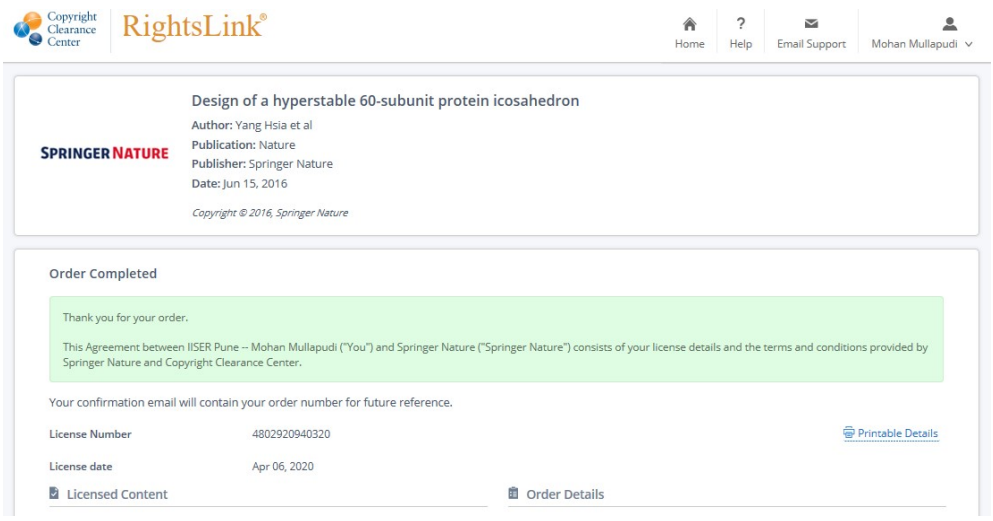
BACK | CLOSE WINDOW

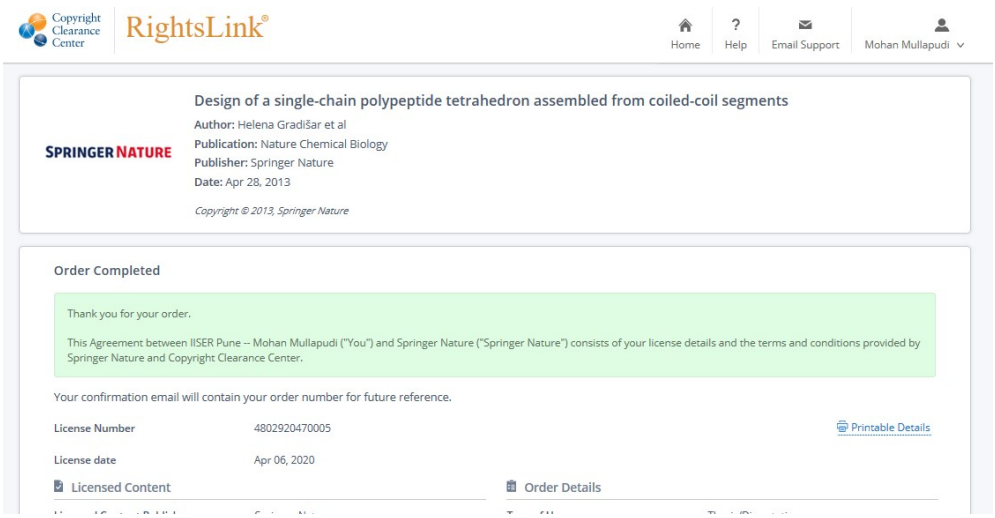
© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Nature articles:









Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

SPRINGER NATURE

Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments

Author: Helena Gradišar et al
Publication: Nature Chemical Biology
Publisher: Springer Nature
Date: Apr 28, 2013

Copyright © 2013, Springer Nature

Order Completed

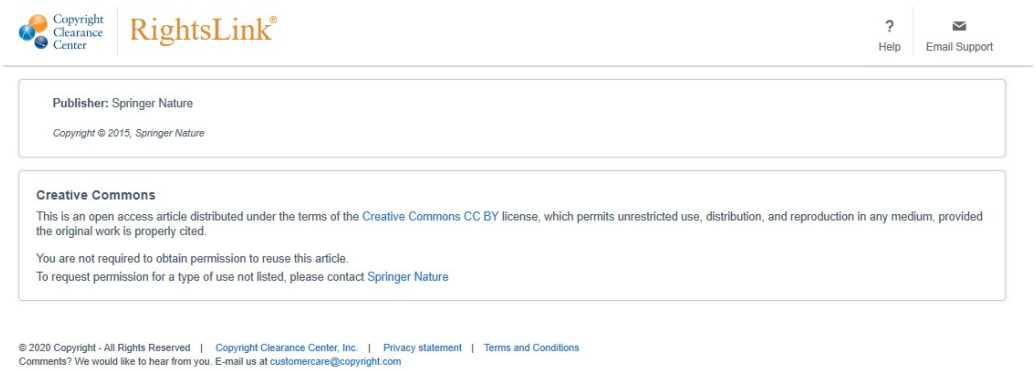
Thank you for your order.

This Agreement between IISER Pune -- Mohan Mullapudi ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802920470005	Printable Details
License date	Apr 06, 2020	

Licensed Content | **Order Details**



Copyright Clearance Center | RightsLink®

Help | Email Support

Publisher: Springer Nature

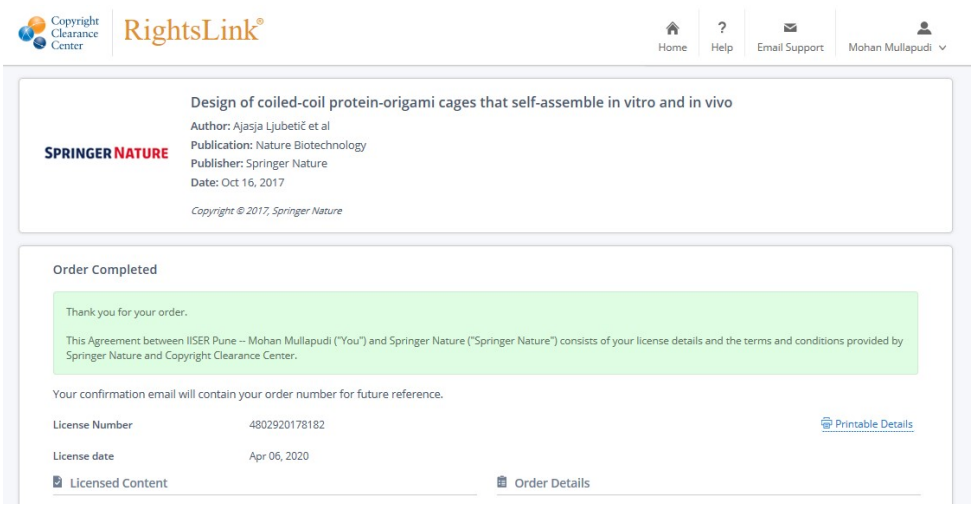
Copyright © 2015, Springer Nature

Creative Commons

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.
To request permission for a type of use not listed, please contact Springer Nature

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

SPRINGER NATURE

Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo

Author: Ajasja Ljubetič et al
Publication: Nature Biotechnology
Publisher: Springer Nature
Date: Oct 16, 2017

Copyright © 2017, Springer Nature

Order Completed

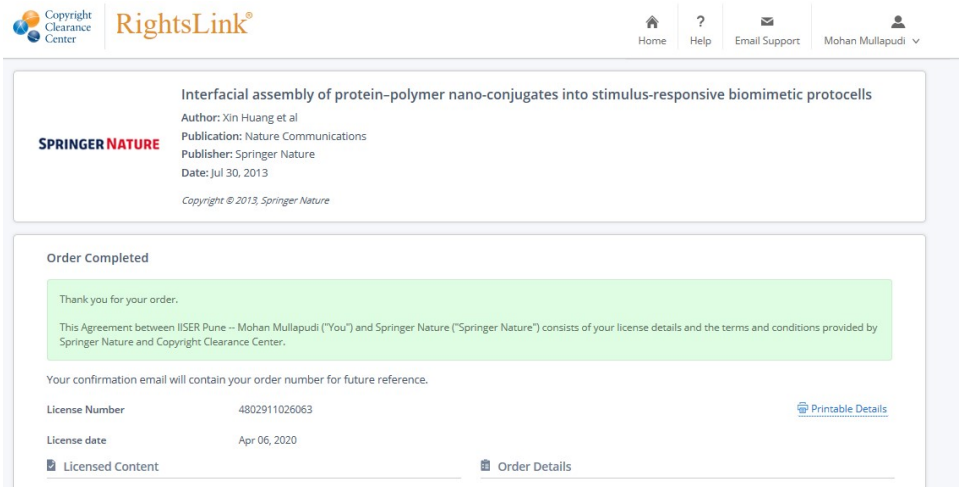
Thank you for your order.

This Agreement between IISER Pune -- Mohan Mullapudi ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802920178182	Printable Details
License date	Apr 06, 2020	

Licensed Content | Order Details



Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Mohan Mullapudi

SPRINGER NATURE

Interfacial assembly of protein–polymer nano-conjugates into stimulus-responsive biomimetic protocells

Author: Xin Huang et al
Publication: Nature Communications
Publisher: Springer Nature
Date: Jul 30, 2013

Copyright © 2013, Springer Nature

Order Completed

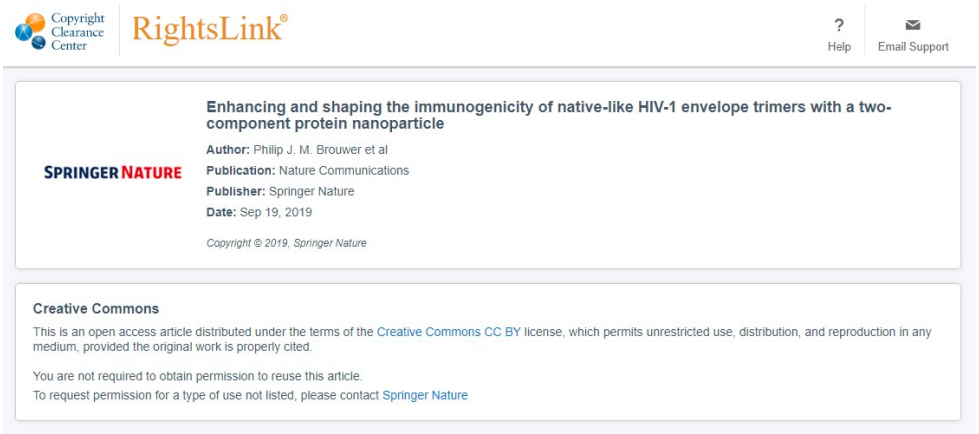
Thank you for your order.

This Agreement between IISER Pune -- Mohan Mullapudi ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802911026063	Printable Details
License date	Apr 06, 2020	

Licensed Content | Order Details



Copyright Clearance Center | RightsLink®

Help | Email Support

SPRINGER NATURE

Enhancing and shaping the immunogenicity of native-like HIV-1 envelope trimers with a two-component protein nanoparticle

Author: Philip J. M. Brouwer et al
Publication: Nature Communications
Publisher: Springer Nature
Date: Sep 19, 2019

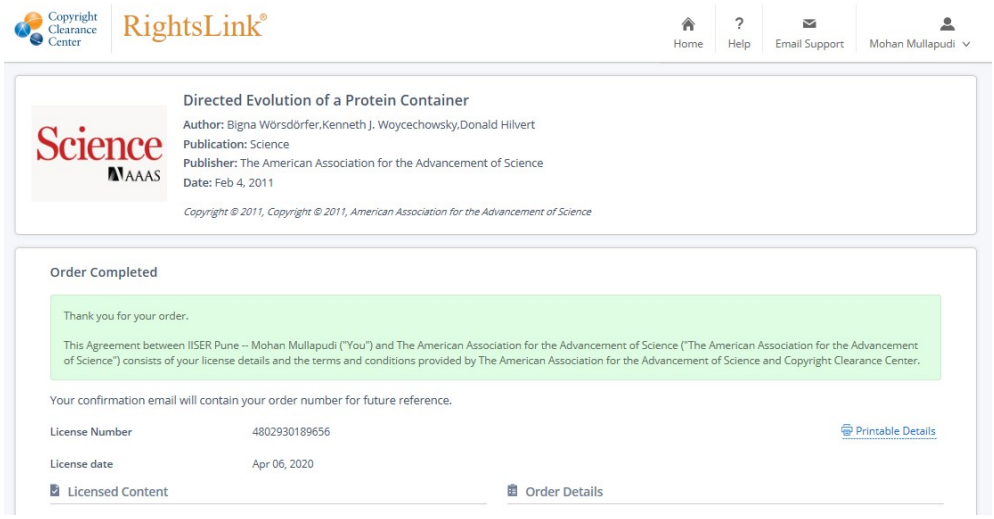
Copyright © 2019, Springer Nature

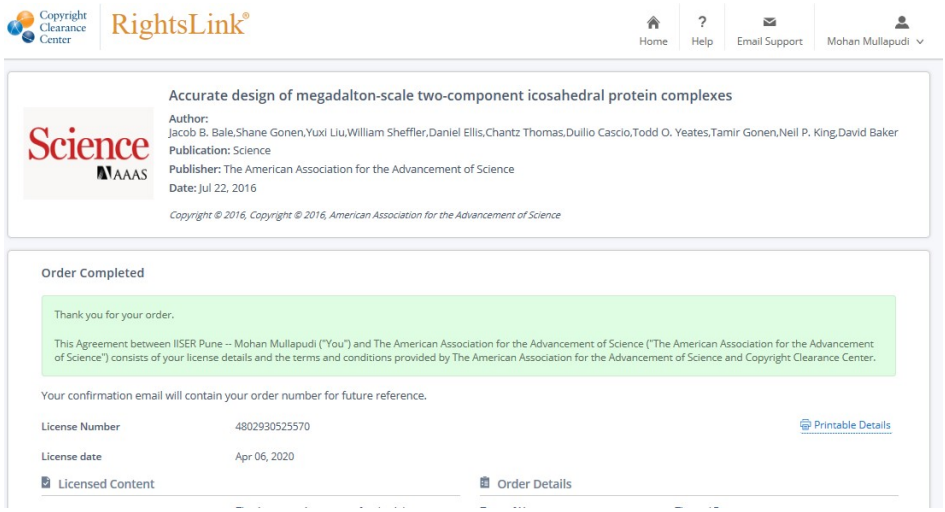
Creative Commons

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.
To request permission for a type of use not listed, please contact Springer Nature

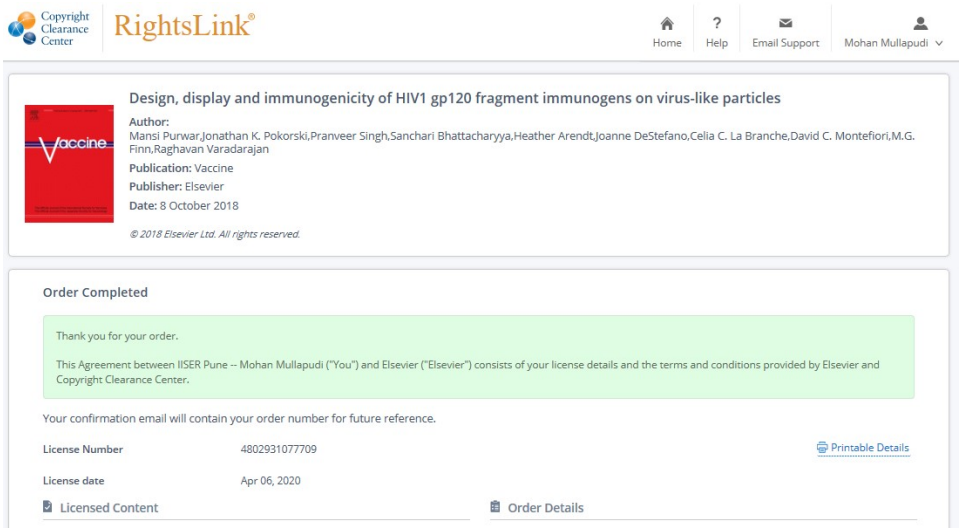
Science articles:

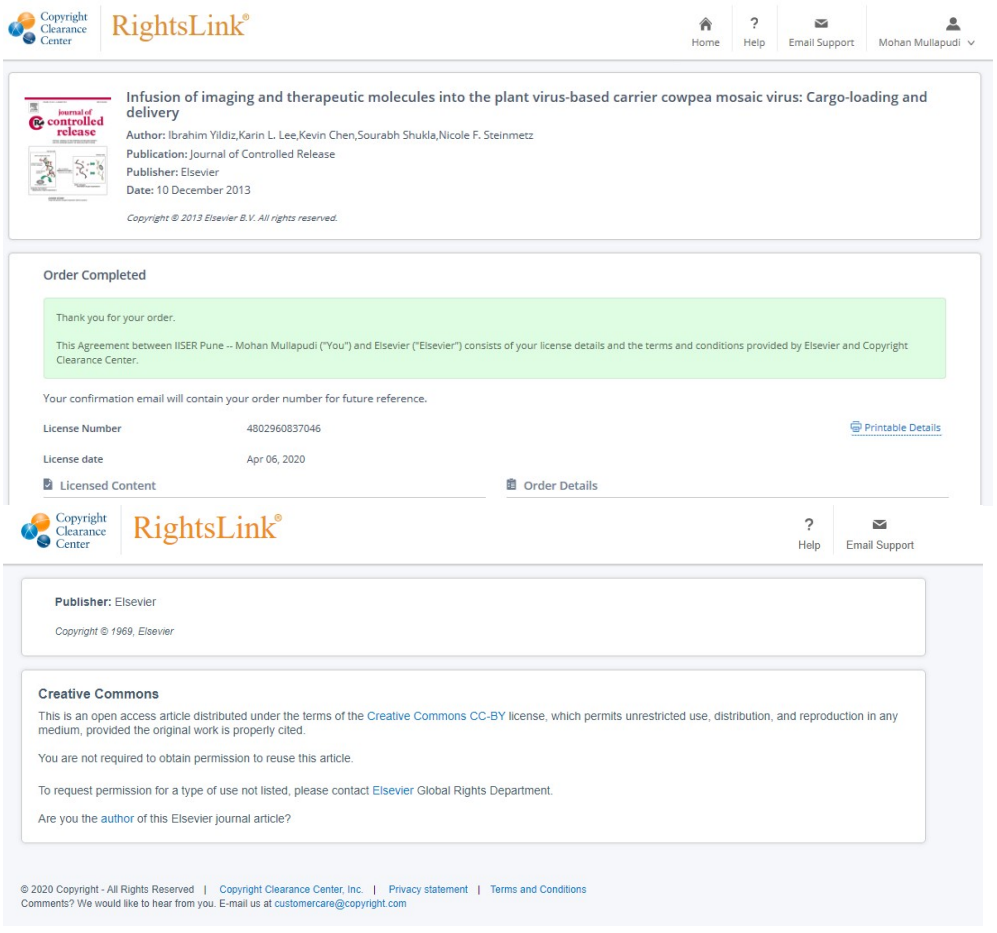




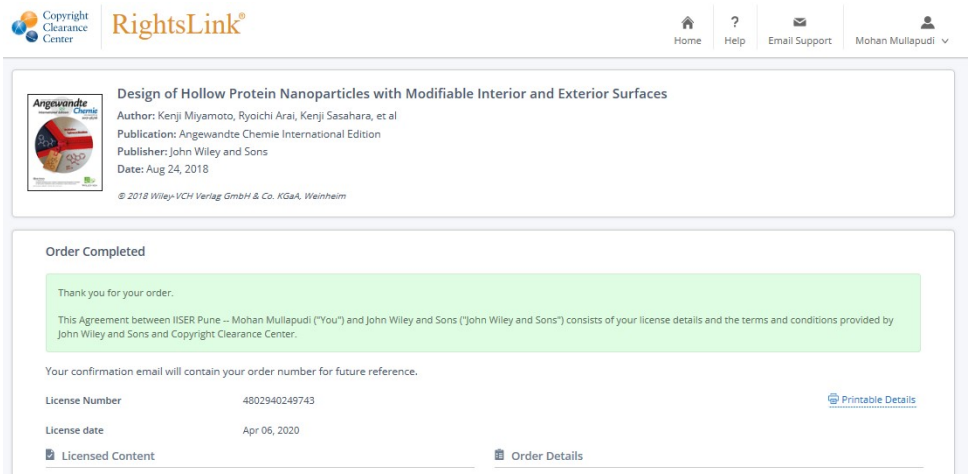


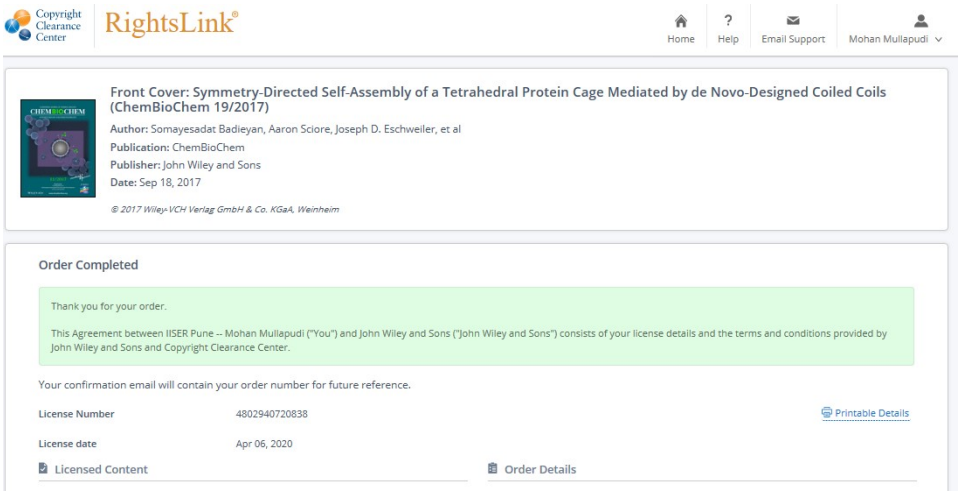
Elsevier articles:

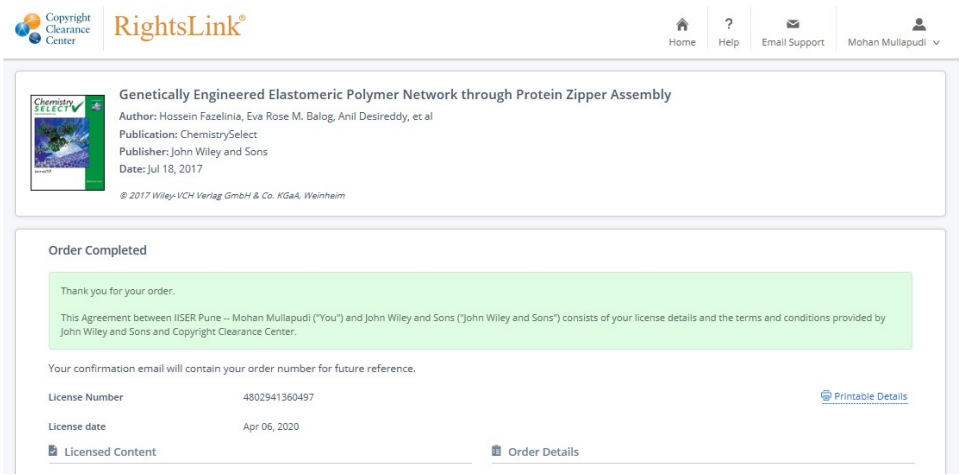


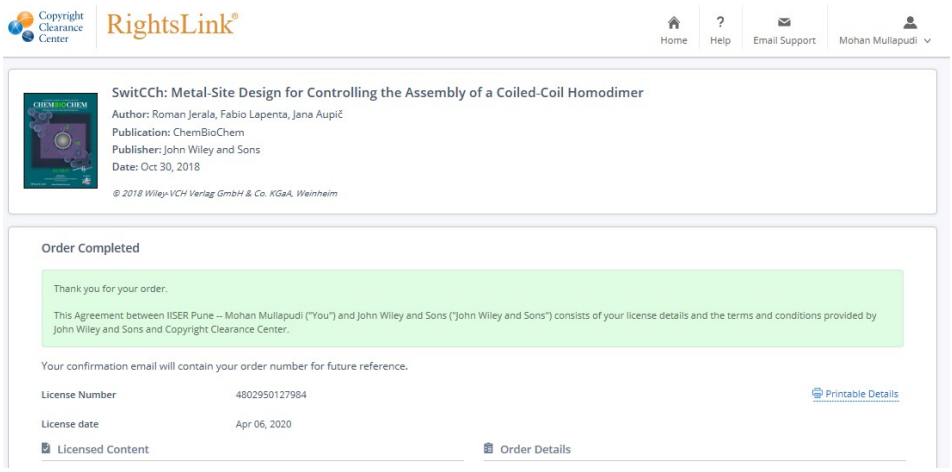


John Wiley and Sons articles:









Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

SwitCCh: Metal-Site Design for Controlling the Assembly of a Coiled-Coil Homodimer

Author: Roman Jerala, Fabio Lapenta, Jana Aupič
Publication: ChemBioChem
Publisher: John Wiley and Sons
Date: Oct 30, 2018

© 2018 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Order Completed

Thank you for your order.

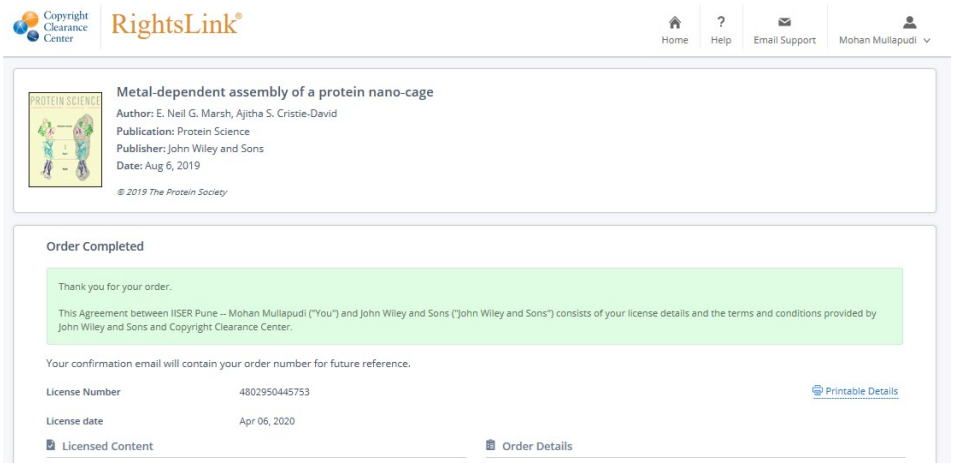
This Agreement between IISER Pune -- Mohan Mullapudi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802950127984
License date	Apr 06, 2020

Printable Details

Licensed Content | Order Details



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

Metal-dependent assembly of a protein nano-cage

Author: E. Neil G. Marsh, Ajitha S. Cristie-David
Publication: Protein Science
Publisher: John Wiley and Sons
Date: Aug 6, 2019

© 2019 The Protein Society

Order Completed

Thank you for your order.

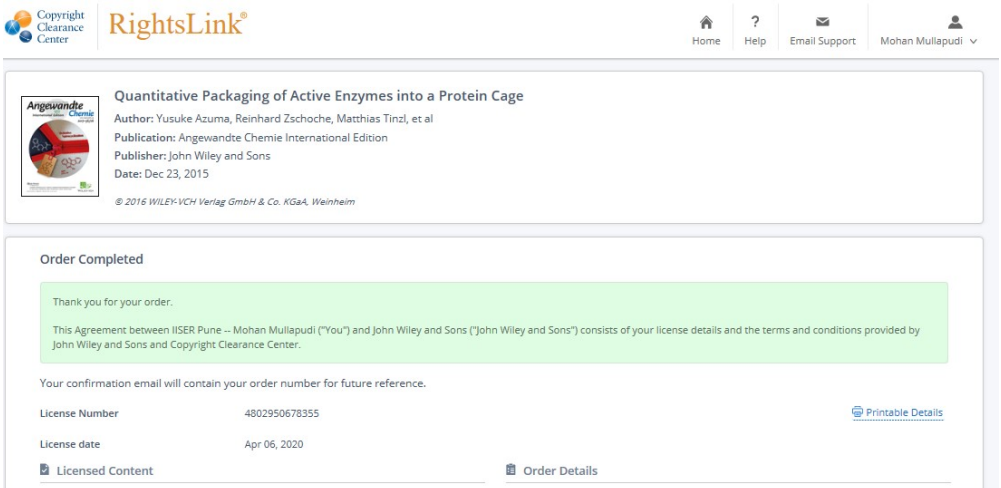
This Agreement between IISER Pune -- Mohan Mullapudi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802950445753
License date	Apr 06, 2020

Printable Details

Licensed Content | Order Details



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

Quantitative Packaging of Active Enzymes into a Protein Cage

Author: Yusuke Azuma, Reinhard Zschoche, Matthias Tinzl, et al
Publication: Angewandte Chemie International Edition
Publisher: John Wiley and Sons
Date: Dec 23, 2015

© 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Order Completed

Thank you for your order.

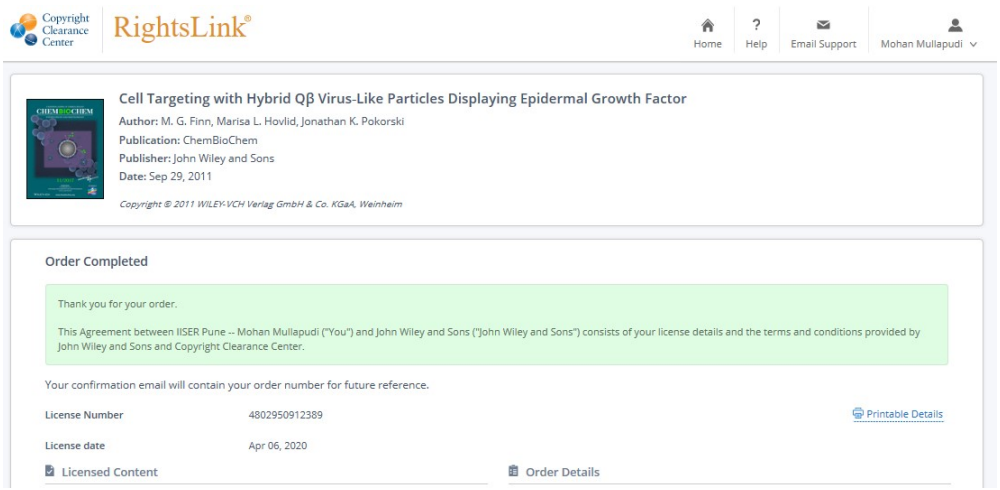
This Agreement between IISER Pune -- Mohan Mullapudi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802950678355
License date	Apr 06, 2020

Printable Details

Licensed Content | Order Details



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

Cell Targeting with Hybrid Qβ Virus-Like Particles Displaying Epidermal Growth Factor

Author: M. G. Finn, Marisa L. Hovlid, Jonathan K. Pokorski
Publication: ChemBioChem
Publisher: John Wiley and Sons
Date: Sep 29, 2011

Copyright © 2011 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Order Completed

Thank you for your order.

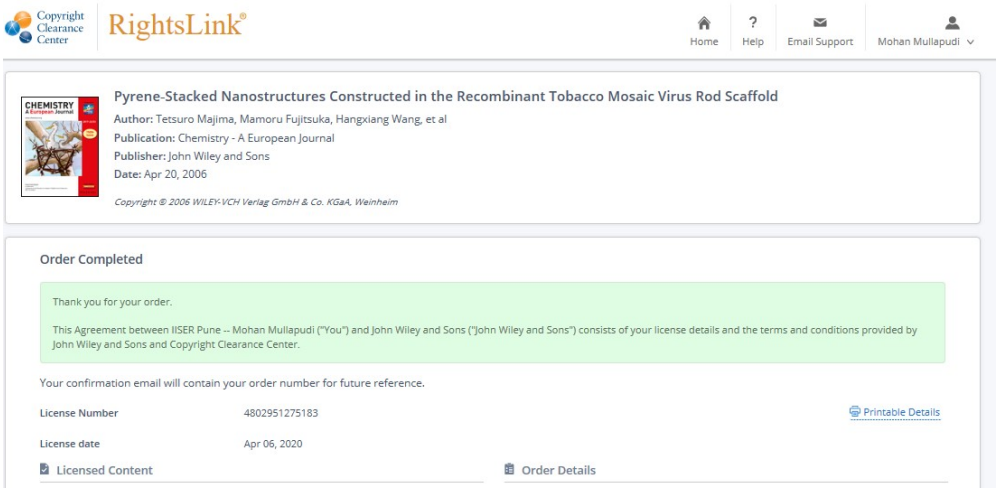
This Agreement between IISER Pune -- Mohan Mullapudi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802950912389
License date	Apr 06, 2020

Printable Details

Licensed Content | Order Details



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

Pyrene-Stacked Nanostructures Constructed in the Recombinant Tobacco Mosaic Virus Rod Scaffold

Author: Tetsuro Majima, Mamoru Fujitsuka, Hangxiang Wang, et al
Publication: Chemistry - A European Journal
Publisher: John Wiley and Sons
Date: Apr 20, 2006

Copyright © 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Order Completed

Thank you for your order.

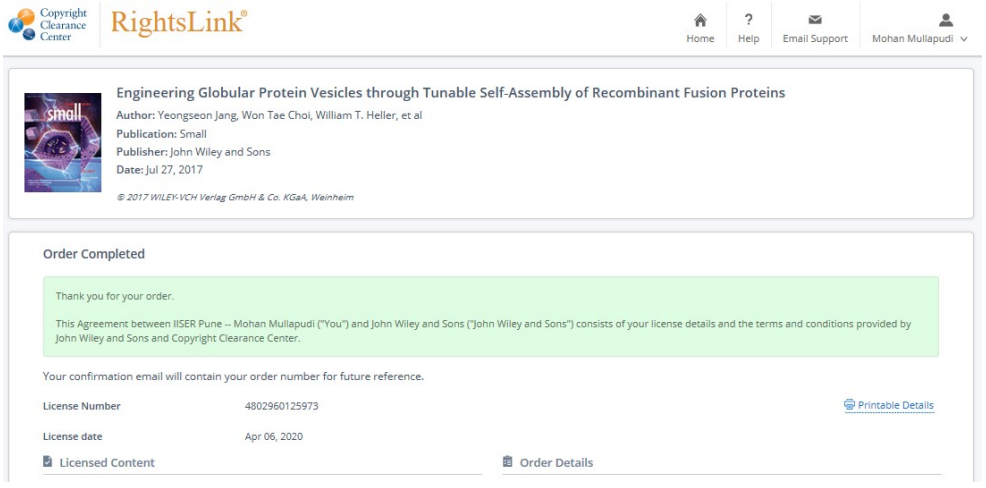
This Agreement between IISER Pune -- Mohan Mullapudi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802951275183
License date	Apr 06, 2020

Printable Details

Licensed Content | Order Details



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

Engineering Globular Protein Vesicles through Tunable Self-Assembly of Recombinant Fusion Proteins

Author: Yeongseon Jang, Won Tae Choi, William T. Heller, et al
Publication: Small
Publisher: John Wiley and Sons
Date: Jul 27, 2017

© 2017 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Order Completed

Thank you for your order.

This Agreement between IISER Pune -- Mohan Mullapudi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802960125973
License date	Apr 06, 2020

Printable Details

Licensed Content | Order Details

Copyright Clearance Center RightsLink®

Help | Email Support

Publisher: John Wiley and Sons

© 2017 The Authors. Published by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Open Access Article

This article is available under the terms of the Creative Commons Attribution License (CC BY) (which may be updated from time to time) and permits use, distribution and reproduction in any medium, provided that the Contribution is properly cited.

For an understanding of what is meant by the terms of the Creative Commons License, please refer to Wiley's Open Access Terms and Conditions.

Permission is not required for this type of reuse.

Wiley offers a professional reprint service for high quality reproduction of articles from over 1400 scientific and medical journals. Wiley's reprint service offers:

- Peer reviewed research or reviews
- Tailored collections of articles
- A professional high quality finish
- Glossy journal style color covers
- Company or brand customisation
- Language translations
- Prompt turnaround times and delivery directly to your office, warehouse or congress.

Please contact our Reprints department for a quotation. Email corporatesaleseurope@wiley.com or corporatesalesusa@wiley.com or corporatesalesDE@wiley.com.

© 2020 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi



Engineered Proteinticles for Targeted Delivery of siRNA to Cancer Cells

Author: Jeewon Lee, Kwangmeyung Kim, Ick Chan Kwon, et al
Publication: Advanced Functional Materials
Publisher: John Wiley and Sons
Date: Jan 16, 2015

© 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Order Completed

Thank you for your order.

This Agreement between IISER Pune -- Mohan Mullapudi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802960567155	Printable Details
License date	Apr 06, 2020	

Licensed Content | Order Details

John Wiley and Sons articles: My own articles.



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

Rational Design of Supramolecular Dynamic Protein Assemblies by Using a Micelle-Assisted Activity-Based Protein-Labeling Technology

Author: Vinod K. Aswal, Sugam Kumar, Pavankumar Janardhan Bhandari, et al
Publication: Chemistry - A European Journal
Publisher: John Wiley and Sons
Date: Oct 4, 2018

© 2018 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Order Completed

Thank you for your order.

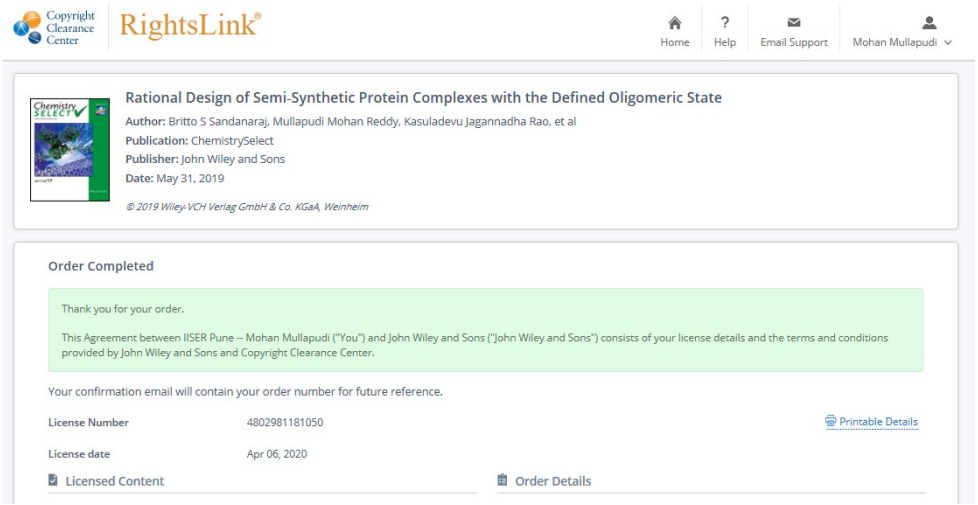
This Agreement between IISER Pune -- Mohan Mullapudi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802980490919
License date	Apr 06, 2020

Printable Details

Licensed Content | Order Details



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Mohan Mullapudi

Rational Design of Semi-Synthetic Protein Complexes with the Defined Oligomeric State

Author: Britto S Sandanaraj, Mullapudi Mohan Reddy, Kasuladevu Jagannadha Rao, et al
Publication: ChemistrySelect
Publisher: John Wiley and Sons
Date: May 31, 2019

© 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Order Completed

Thank you for your order.

This Agreement between IISER Pune -- Mohan Mullapudi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4802981181050
License date	Apr 06, 2020

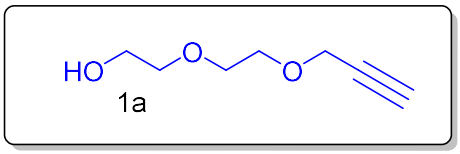
Printable Details

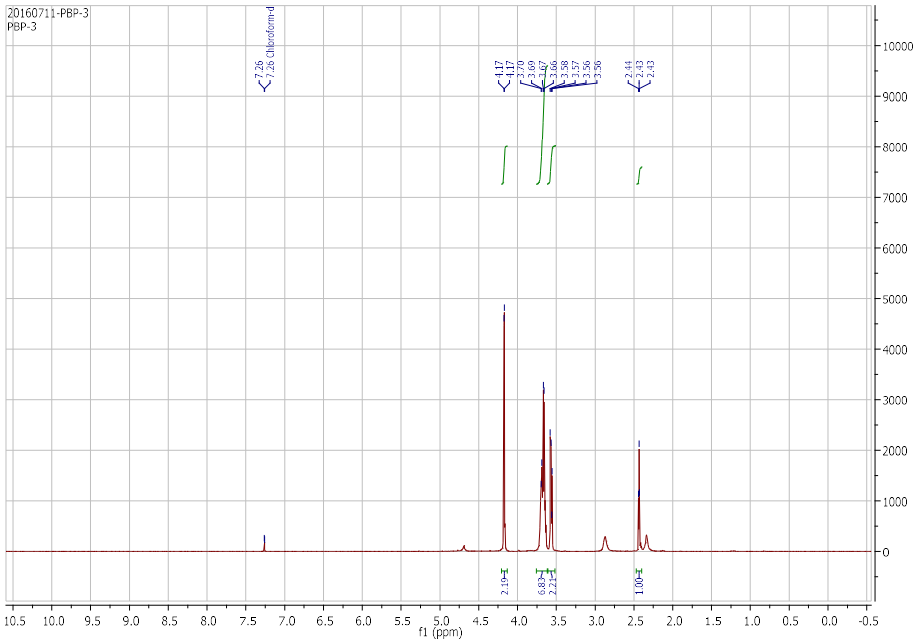
Licensed Content | Order Details

Appendix-VII: Spectral Data

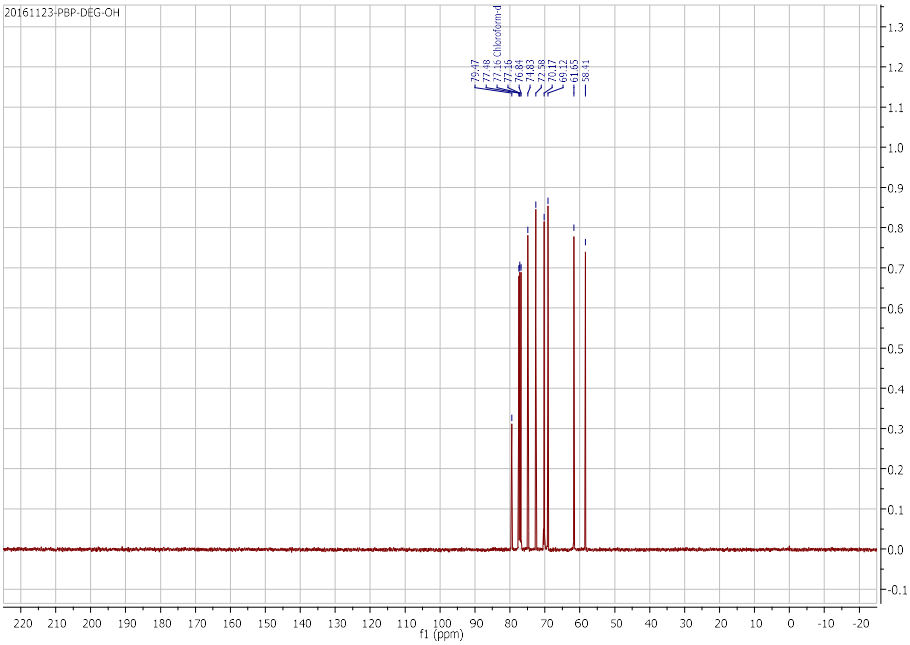
Chapter 2 Spectral Data

Compound 1a



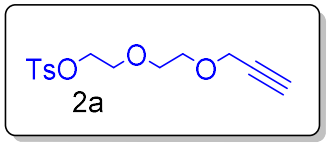


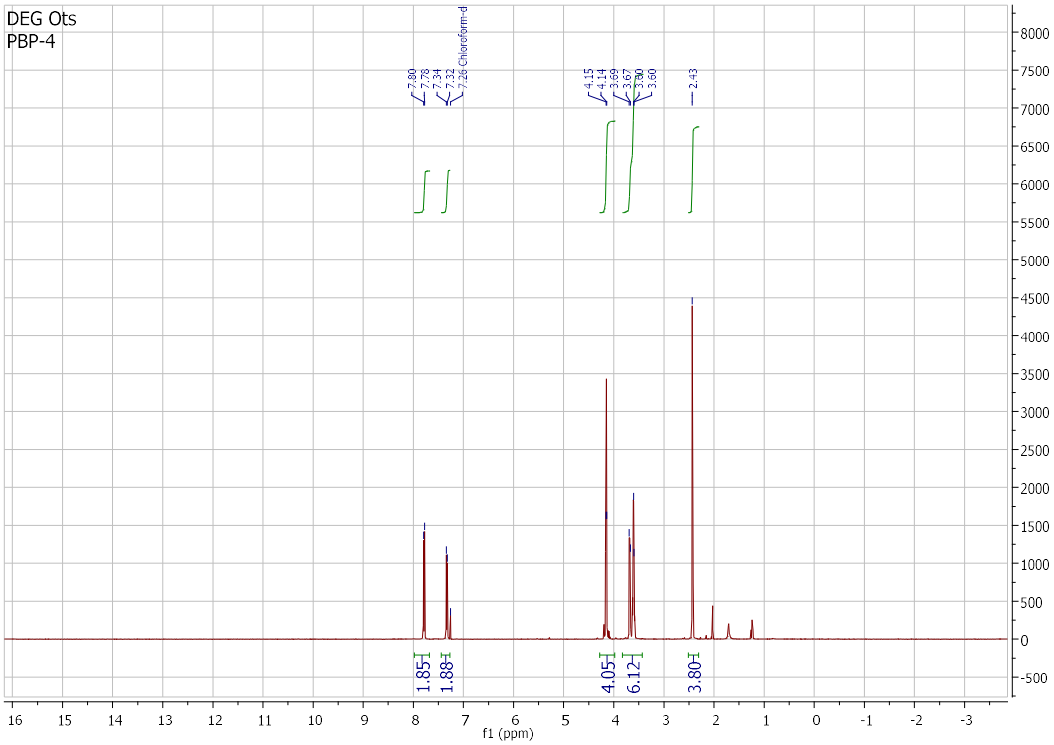
^{1}H NMR spectrum



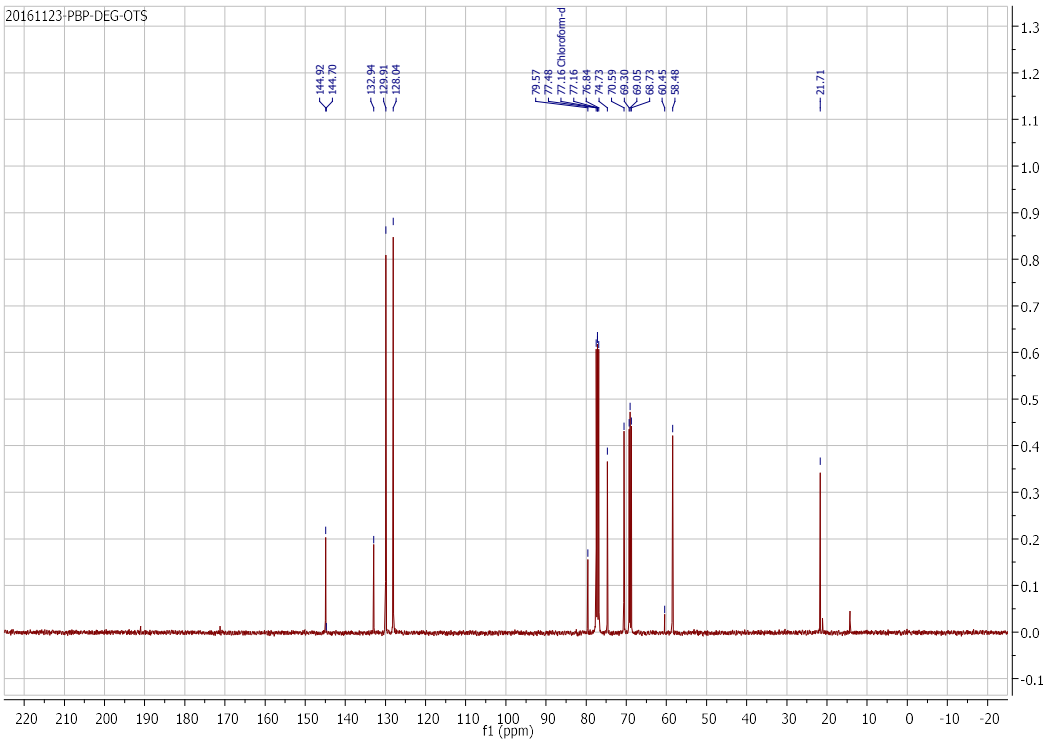
^{13}C NMR spectrum

Compound 2a



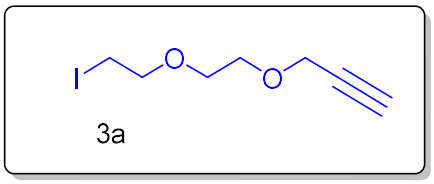


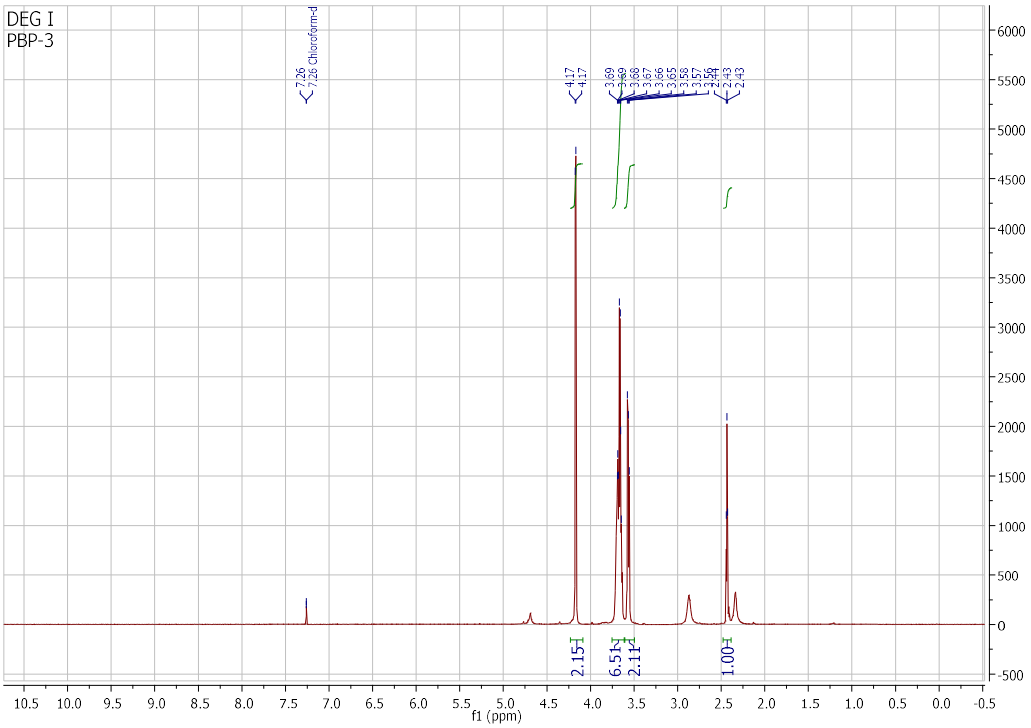
^{1}H NMR spectrum



^{13}C NMR spectrum

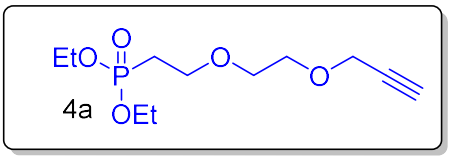
Compound 3a

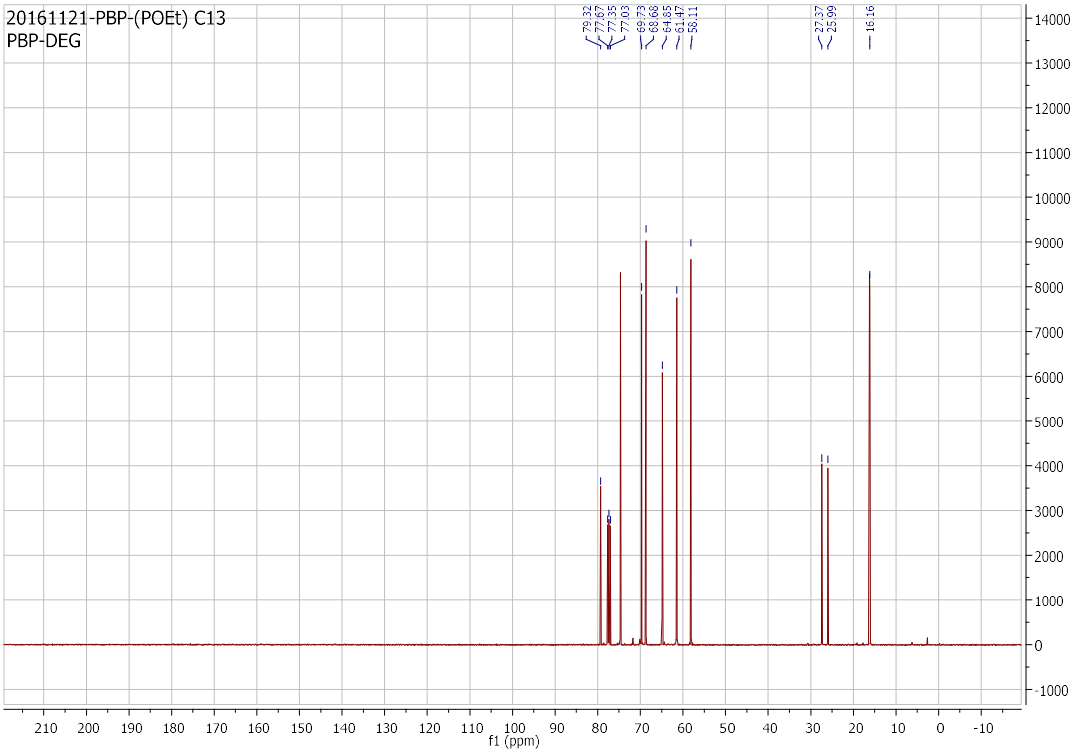




^{1}H NMR spectrum

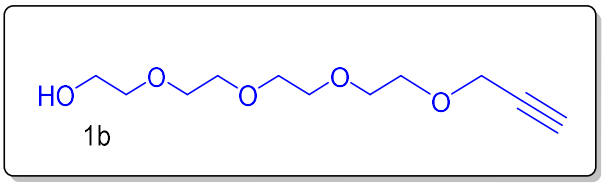
Compound 4a

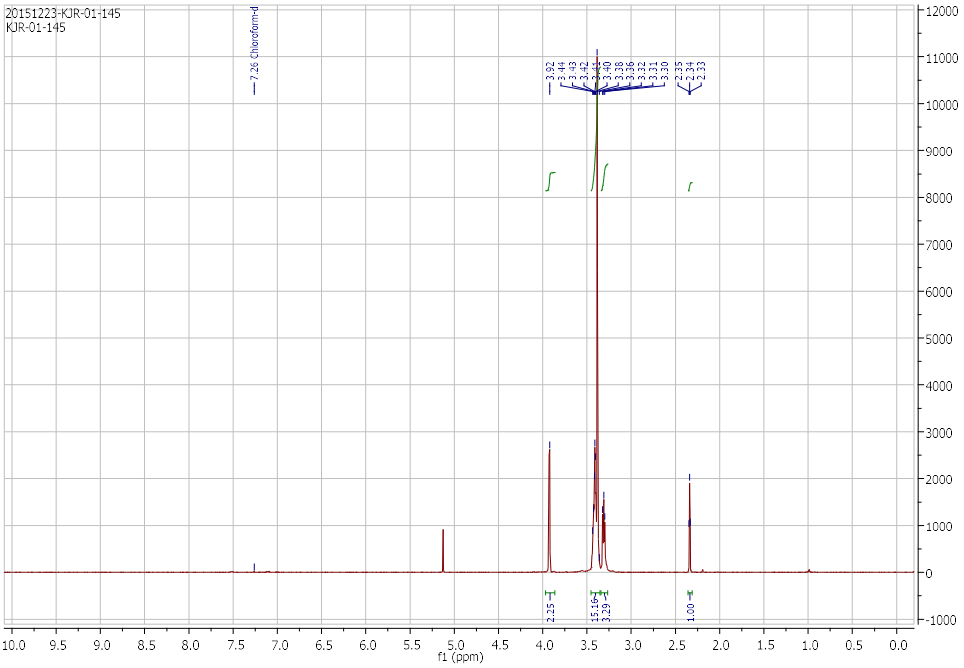




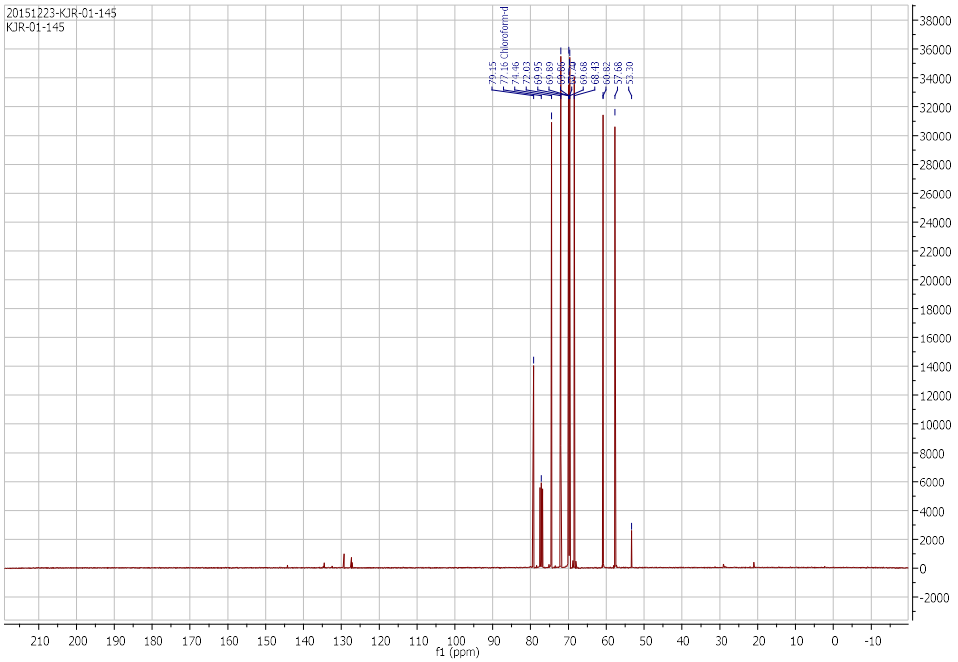
^{13}C NMR spectrum

Compound 1b



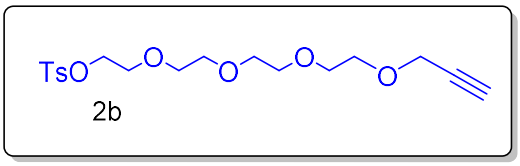


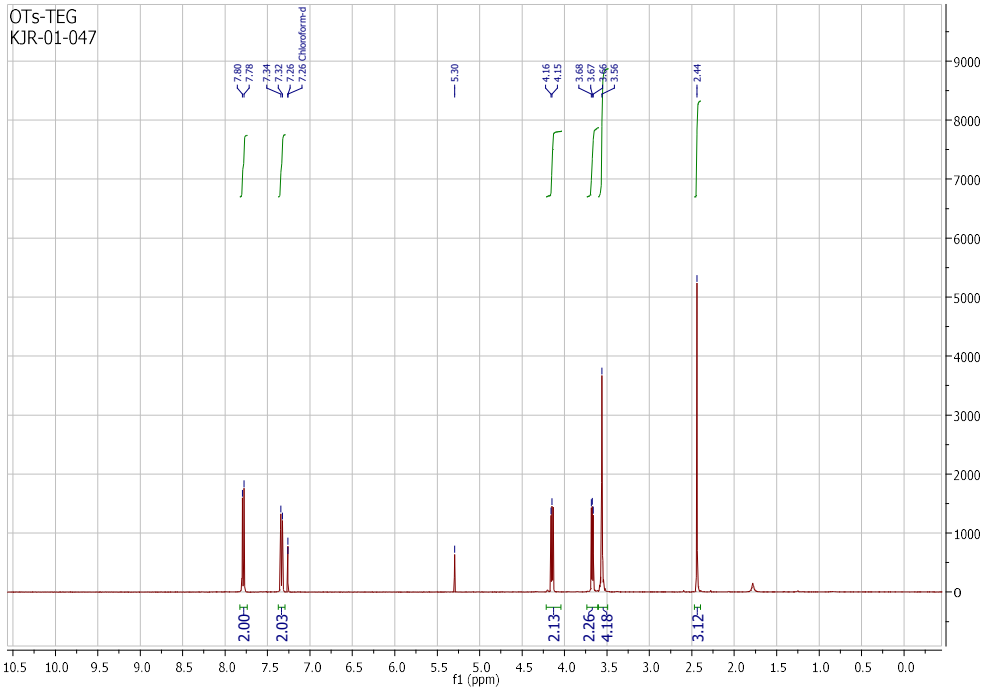
^{1}H NMR spectrum



^{13}C NMR spectrum

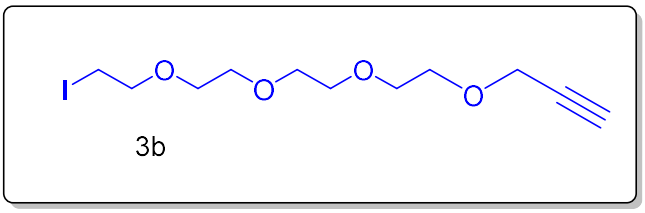
Compound 2b

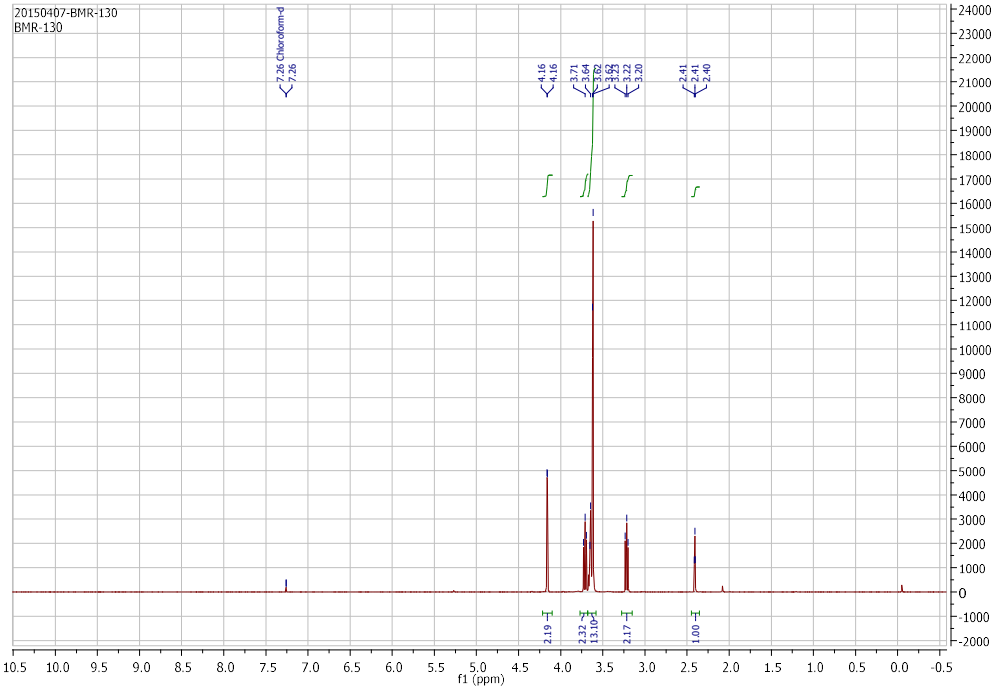




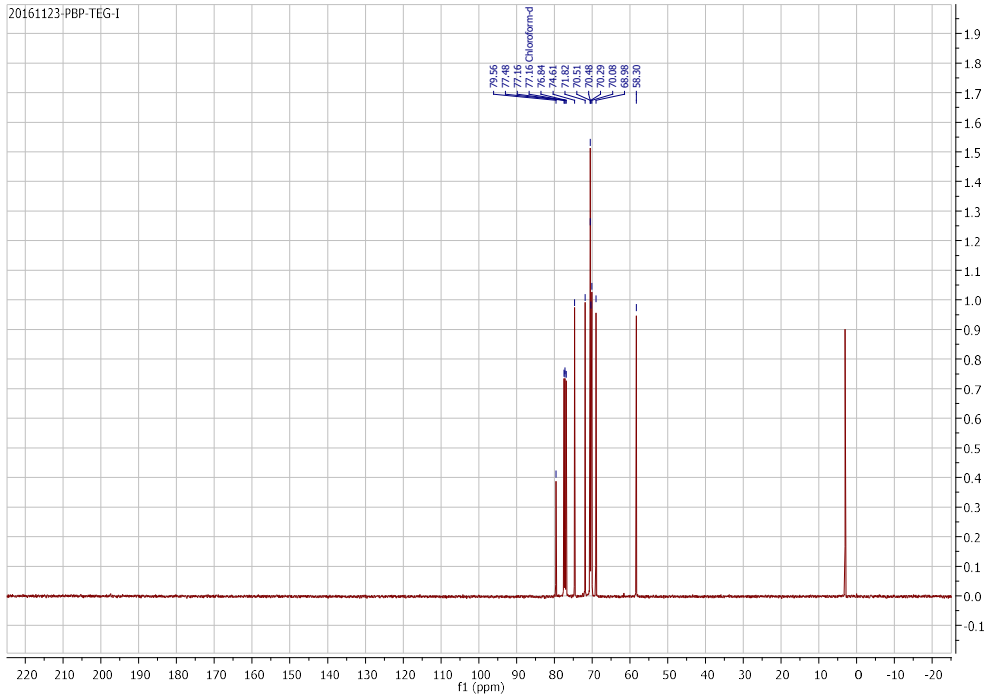
^{1}H NMR spectrum

Compound 3b



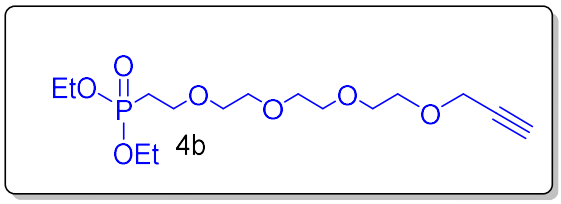


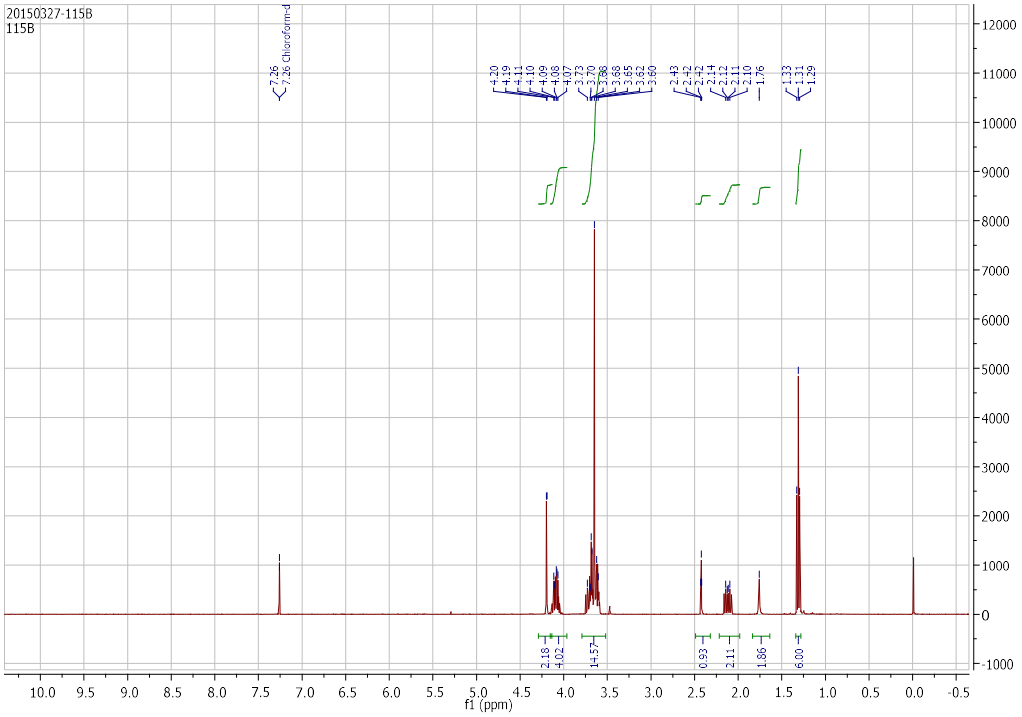
^{1}H NMR spectrum



^{13}C NMR spectrum

Compound 4b

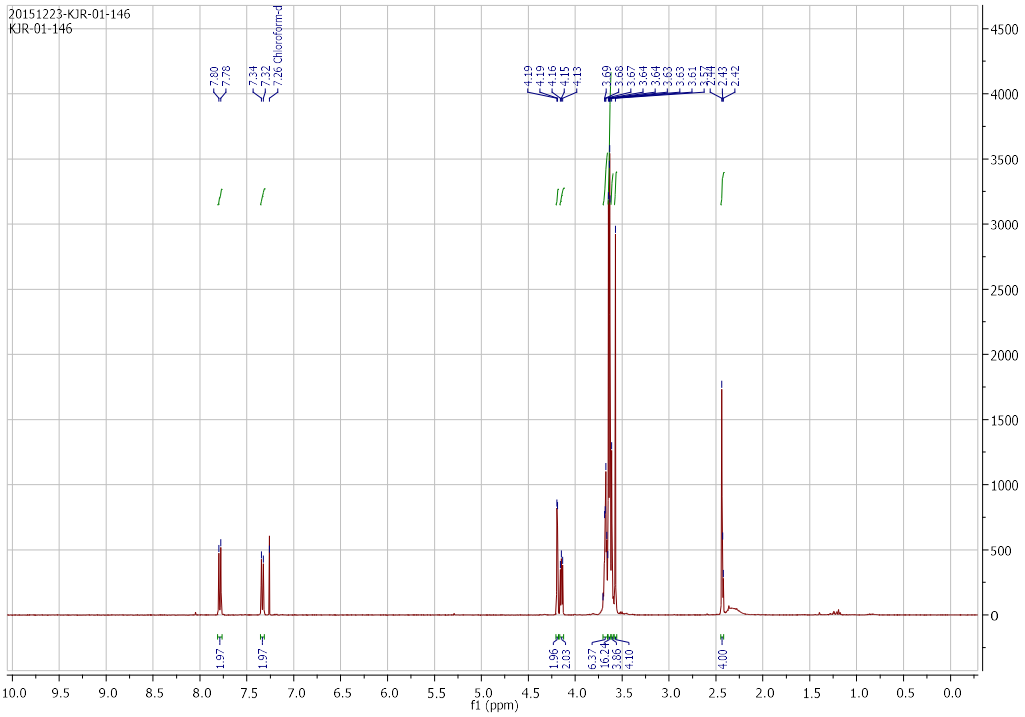




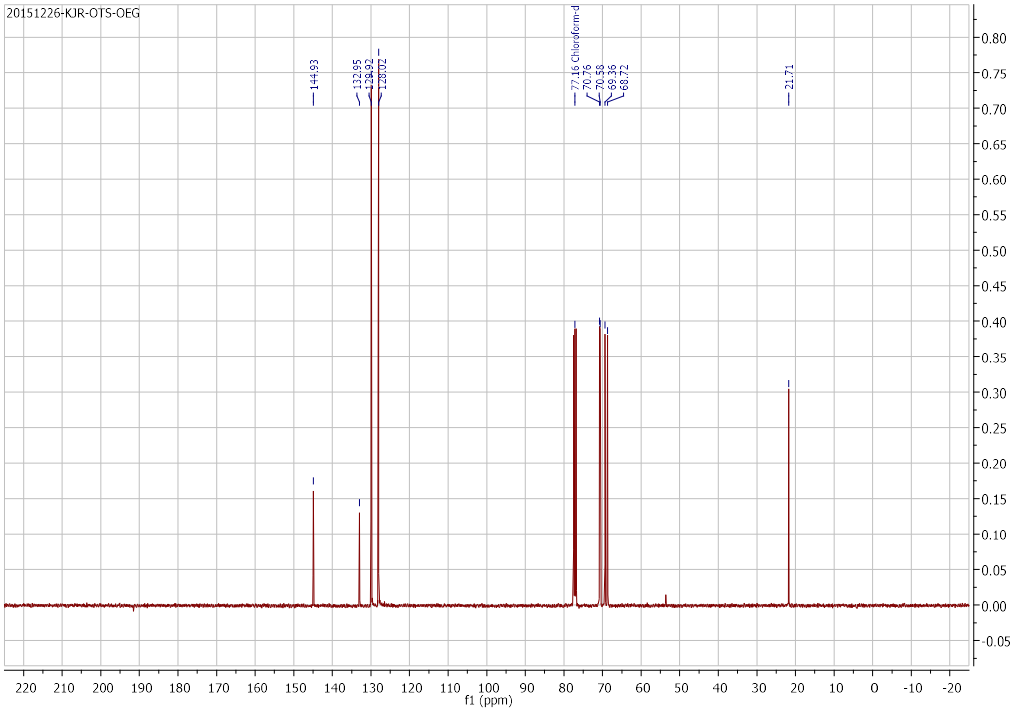
^{1}H NMR spectrum

Compound 2c



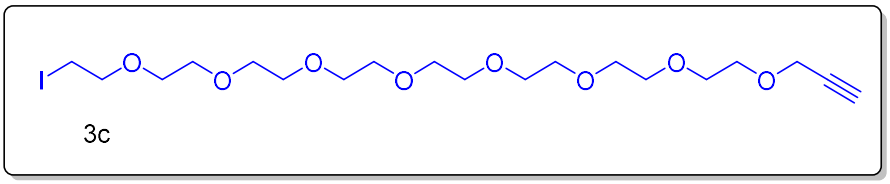


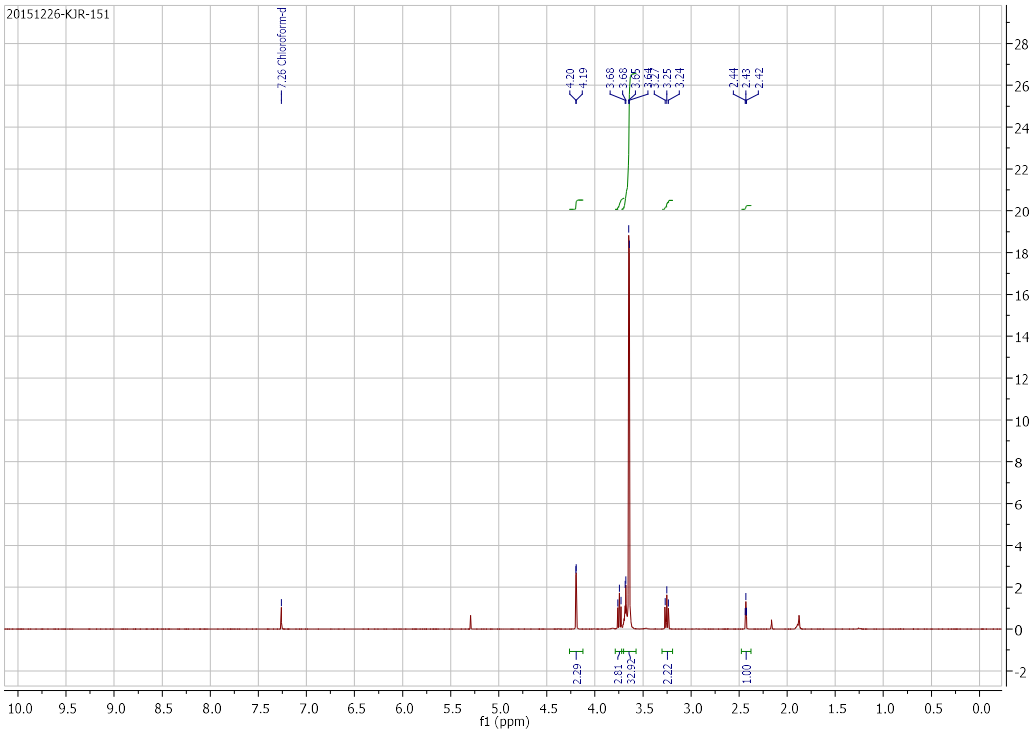
^{1}H NMR spectrum



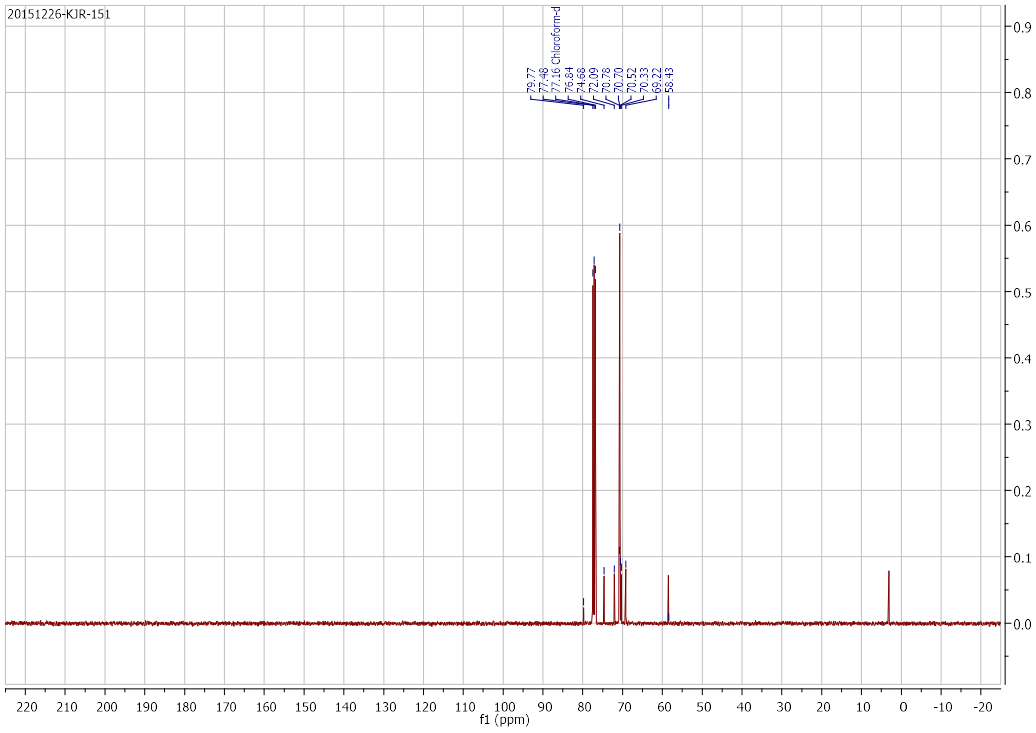
^{13}C NMR spectrum

Compound 3c



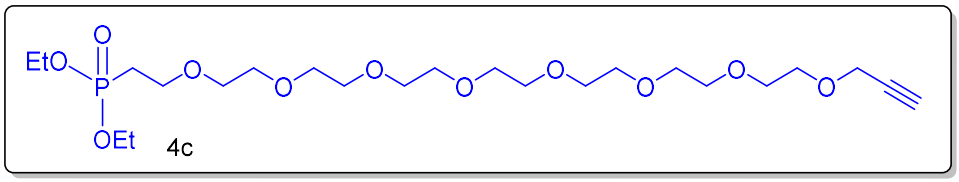


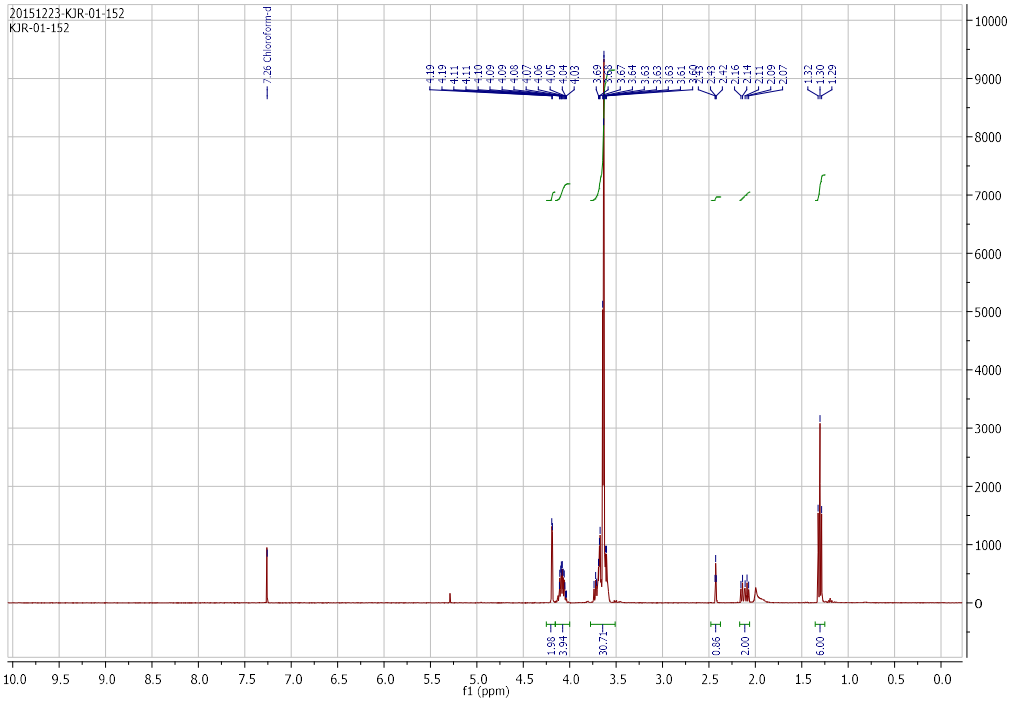
^{1}H NMR spectrum



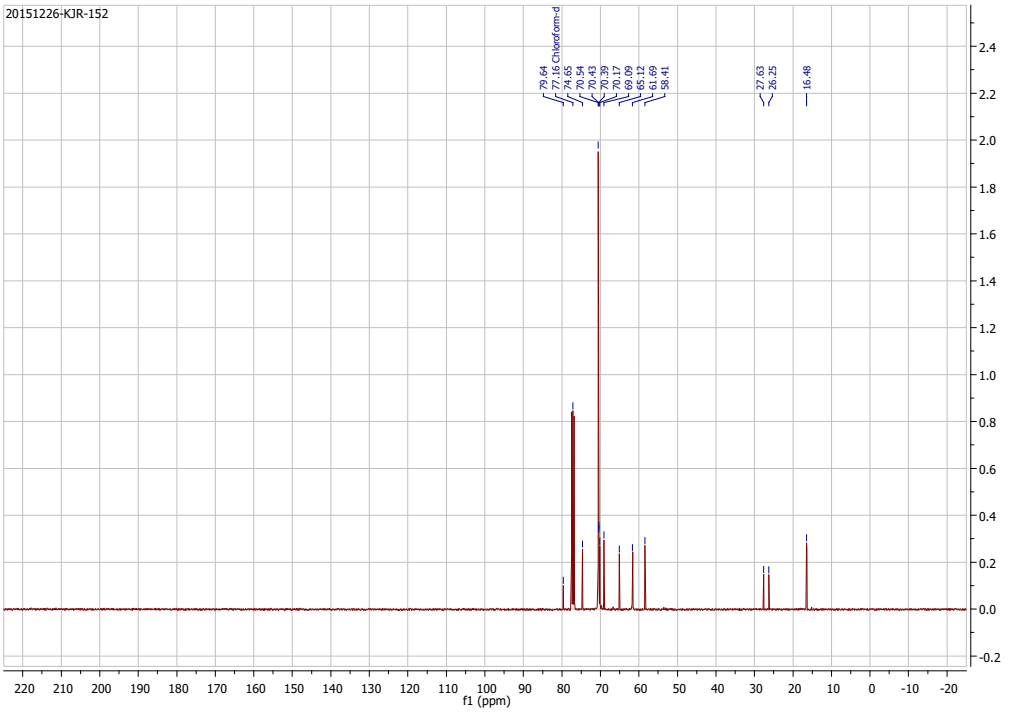
^{13}C NMR spectrum

Compound 4c



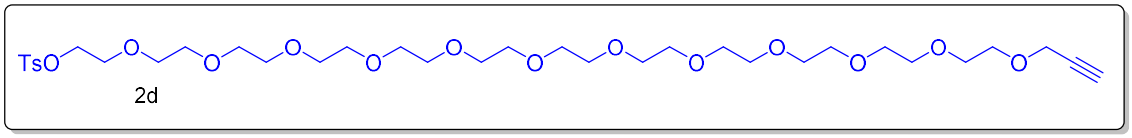


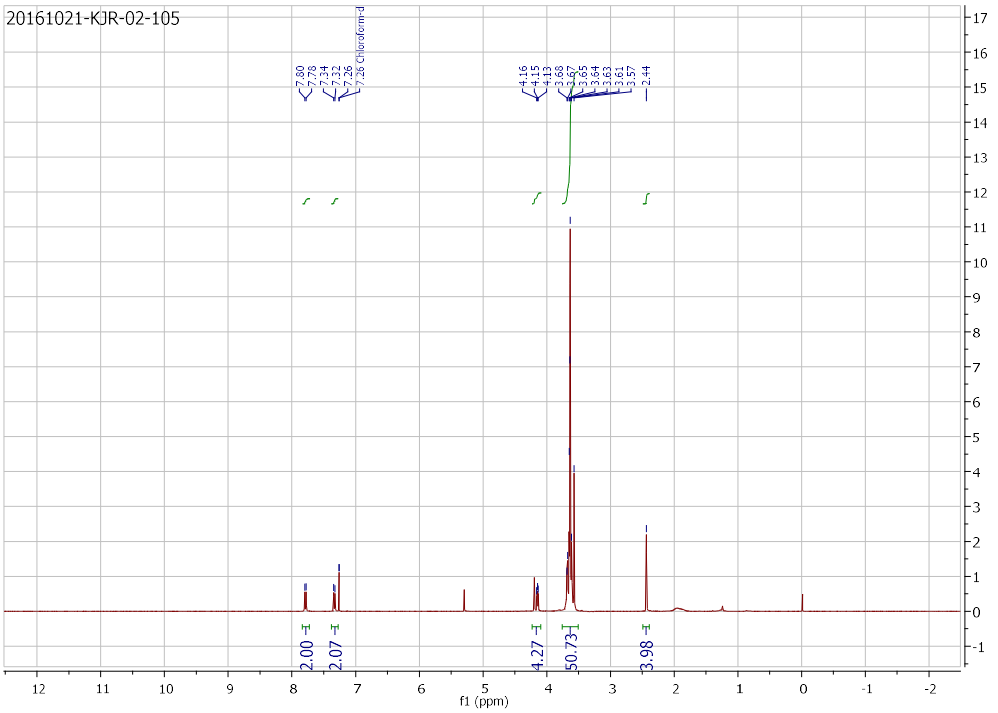
^{1}H NMR spectrum



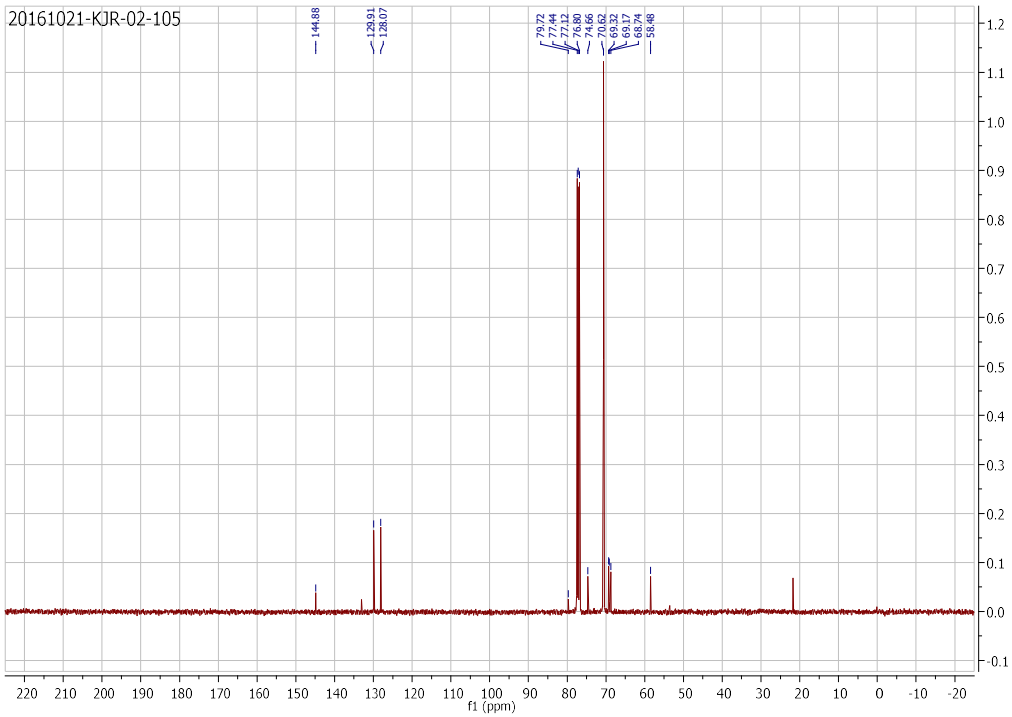
^{13}C NMR spectrum

Compound 2d



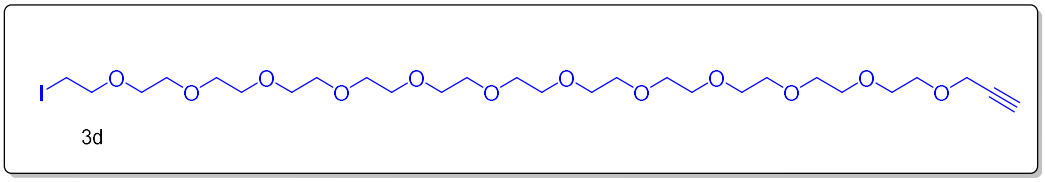


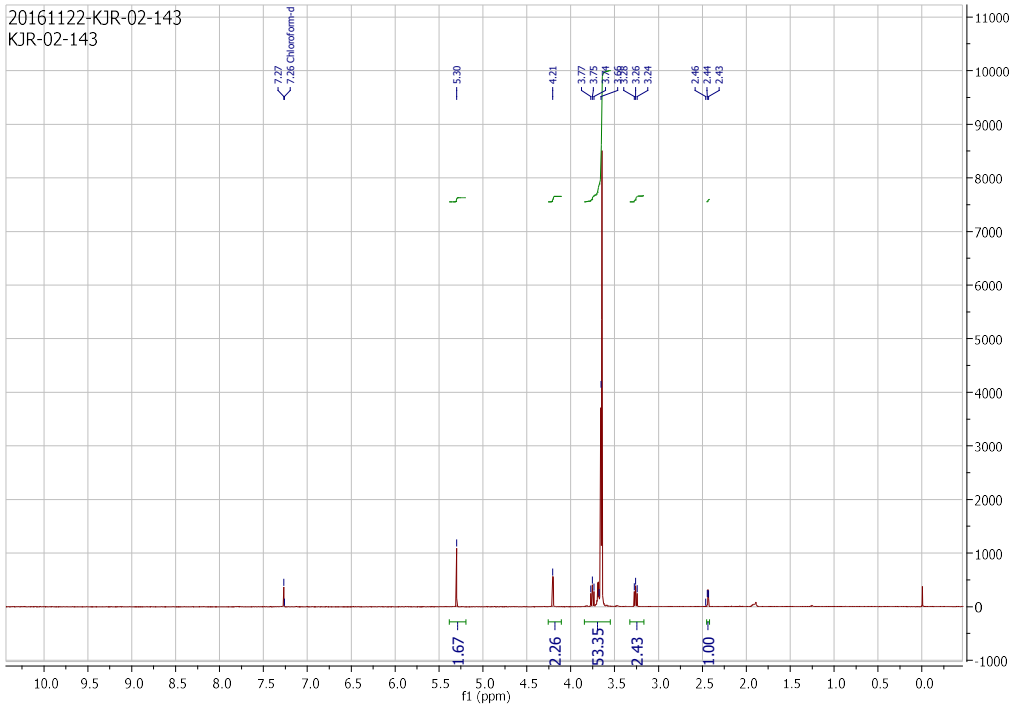
^{1}H NMR spectrum



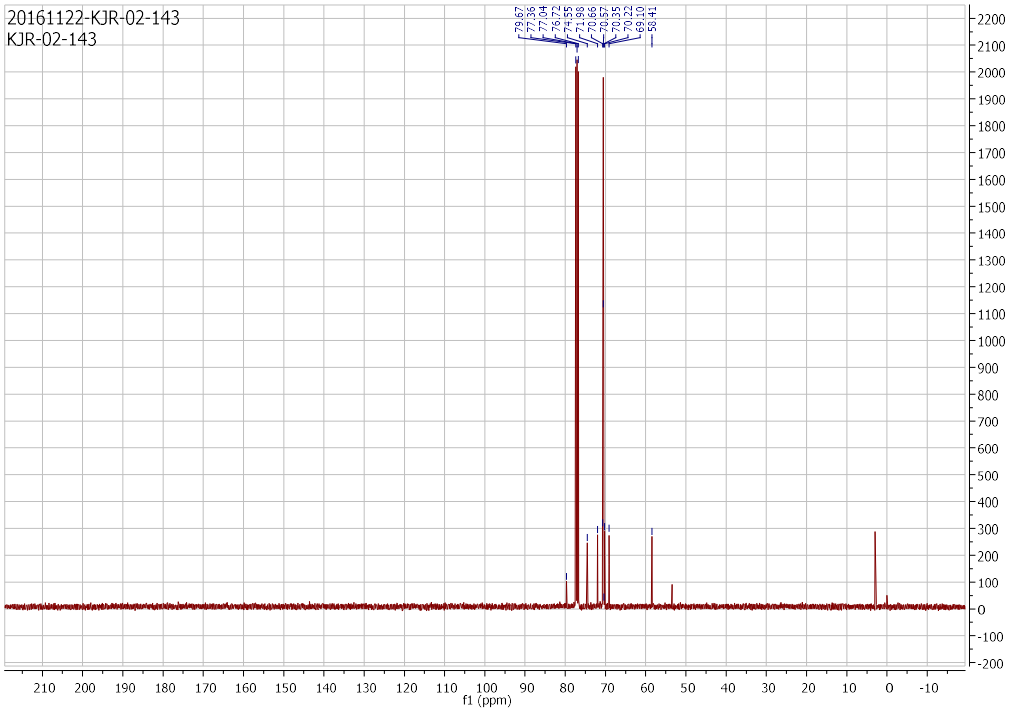
^{13}C NMR spectrum

Compound 3d



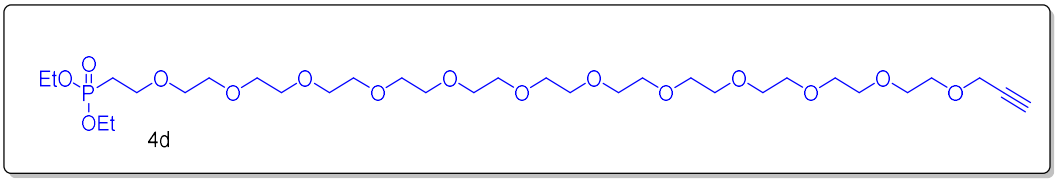


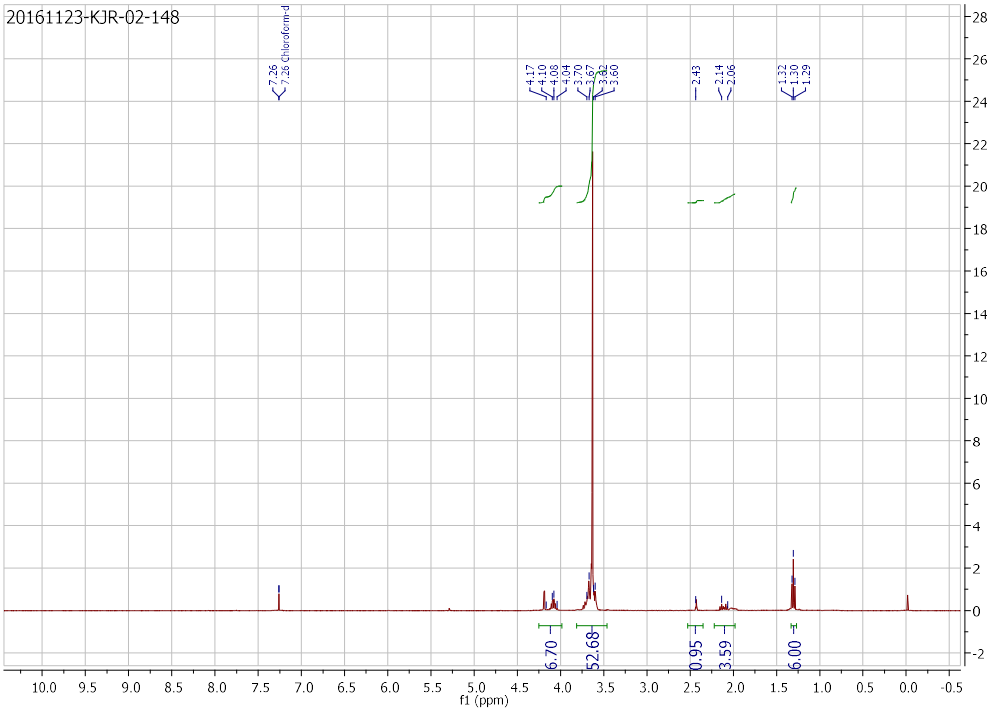
^{1}H NMR spectrum



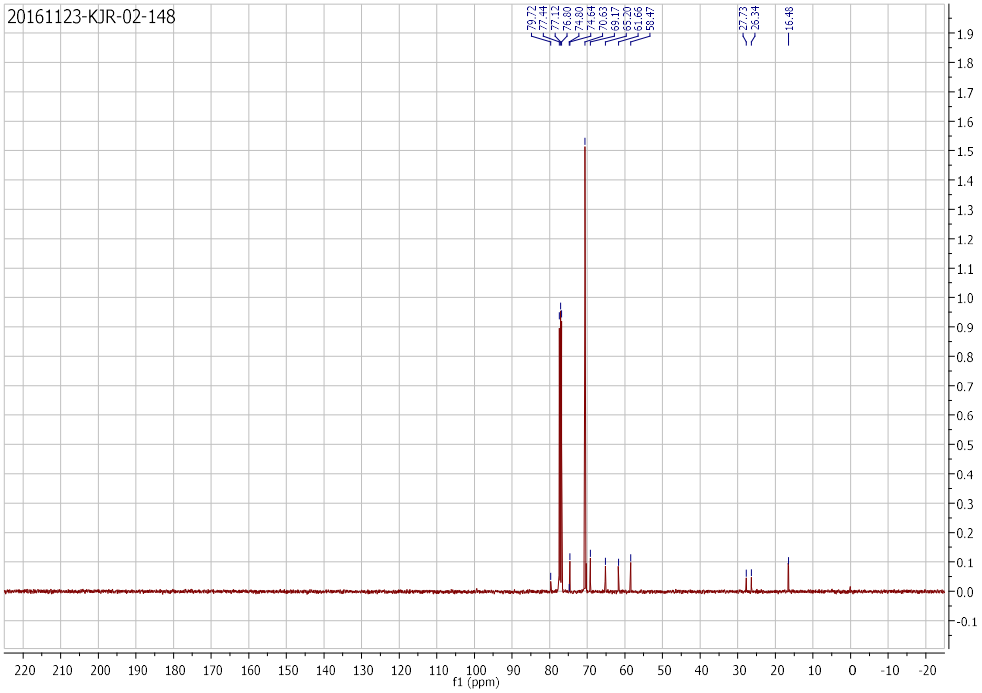
^{13}C NMR spectrum

Compound 4d



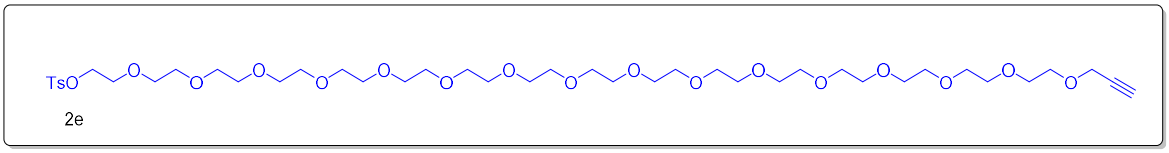


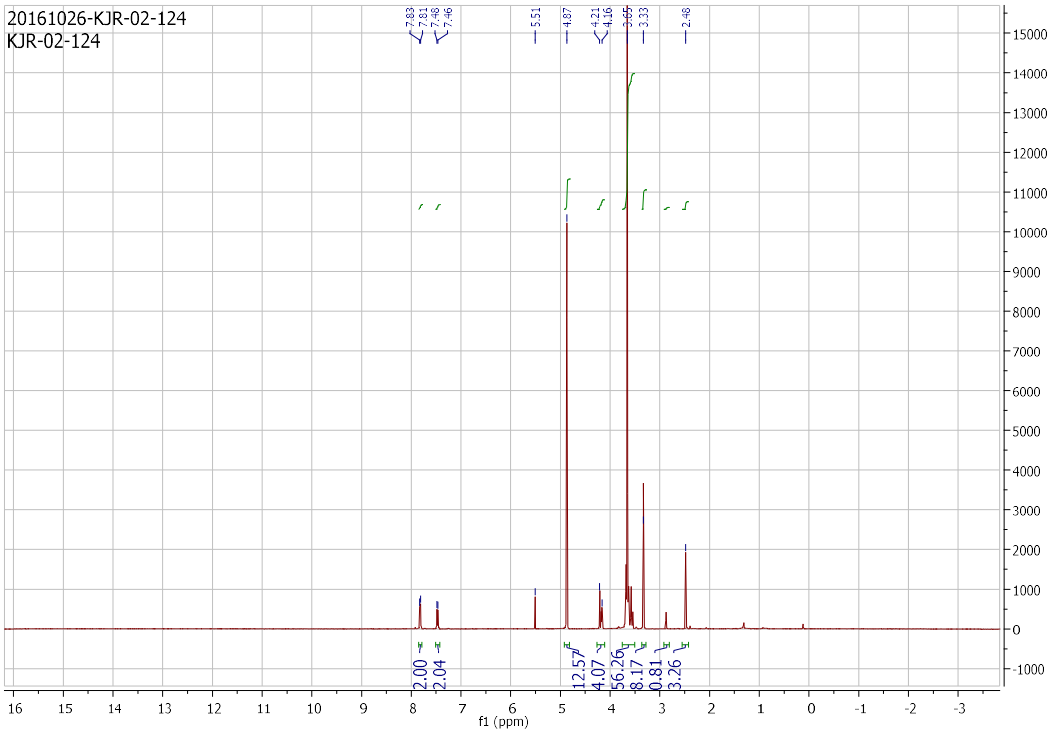
^{1}H NMR spectrum



^{13}C NMR spectrum

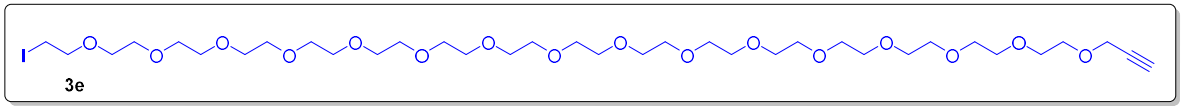
Compound 2e

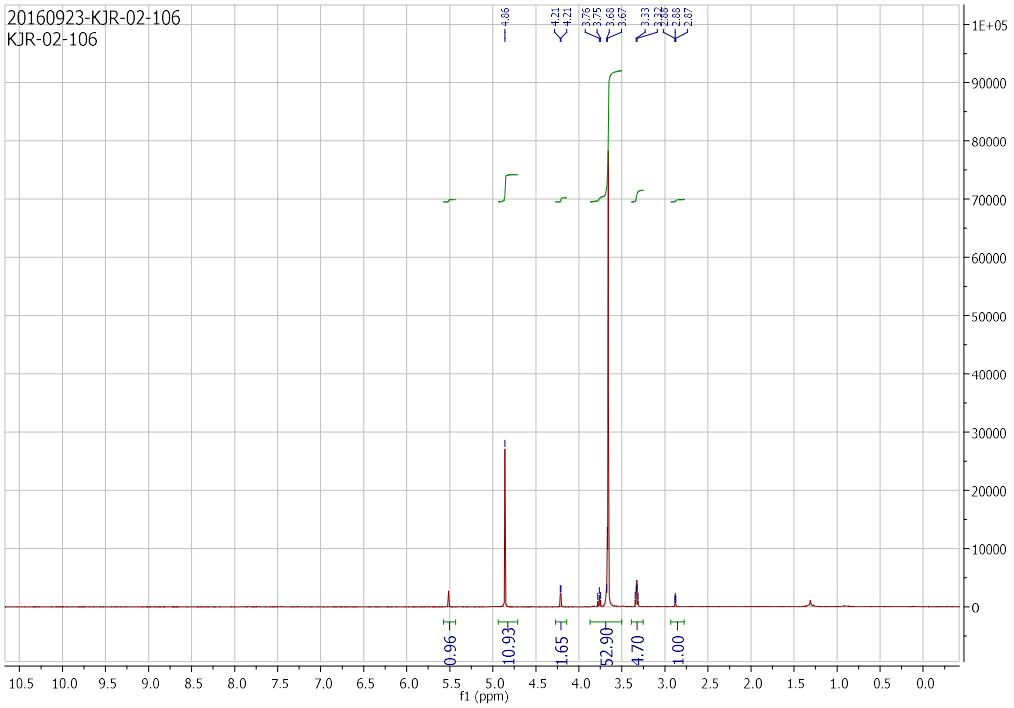




^{1}H NMR spectrum

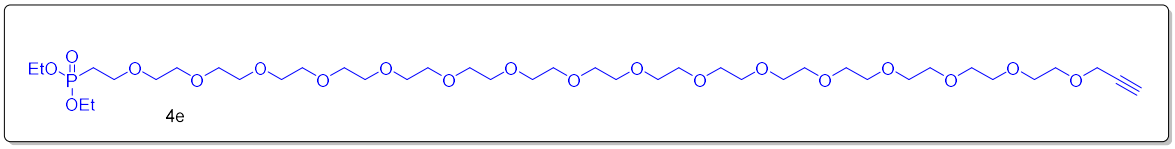
Compound 3e

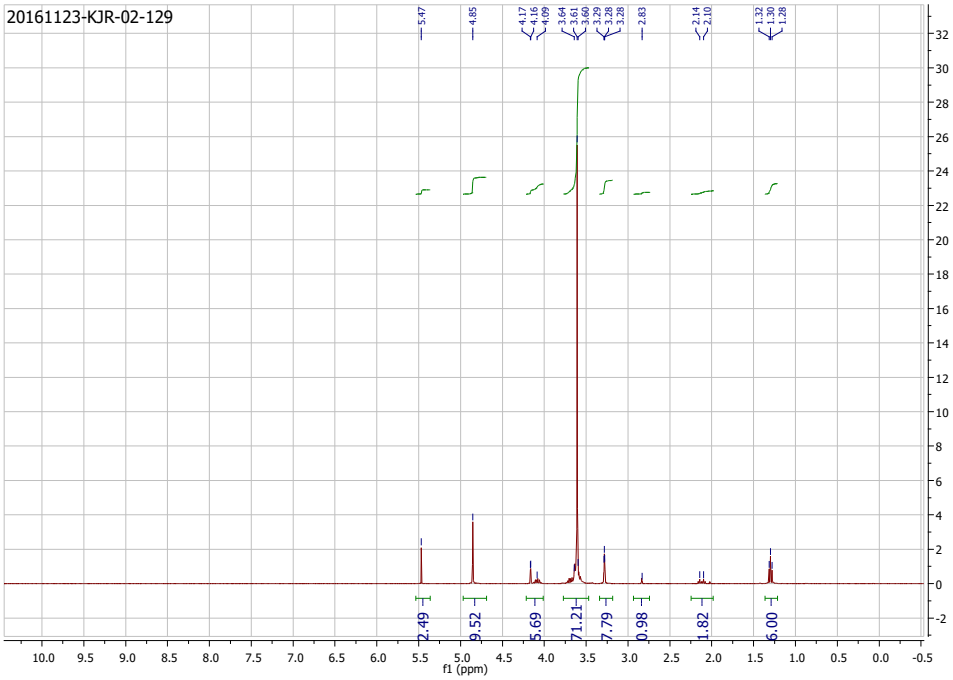




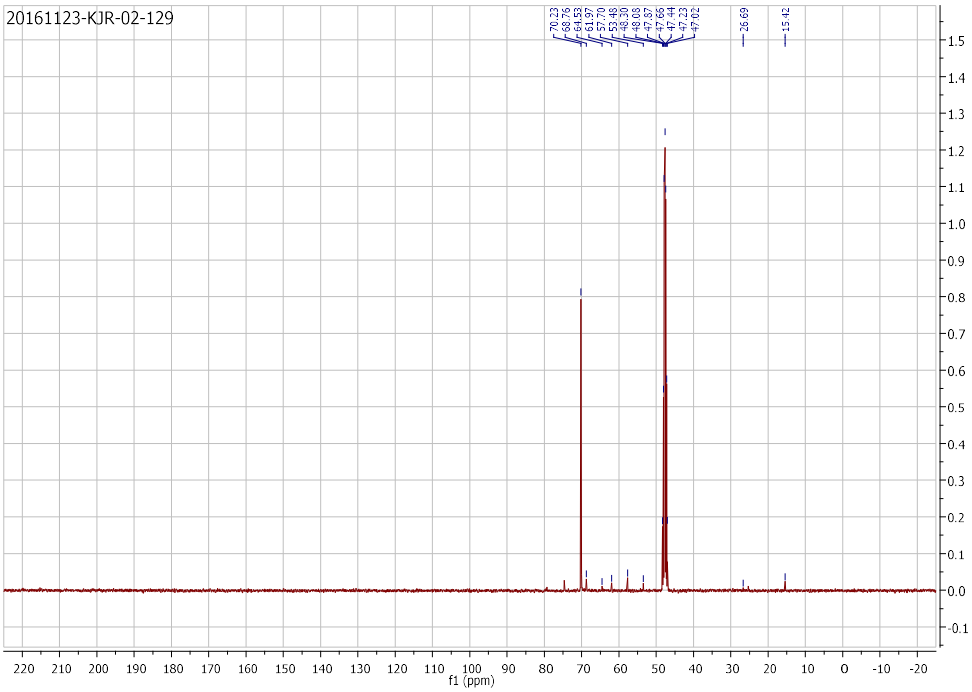
^{1}H NMR spectrum

Compound 4e



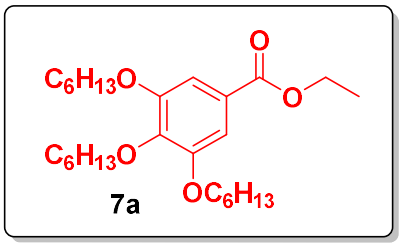


^{1}H NMR spectrum

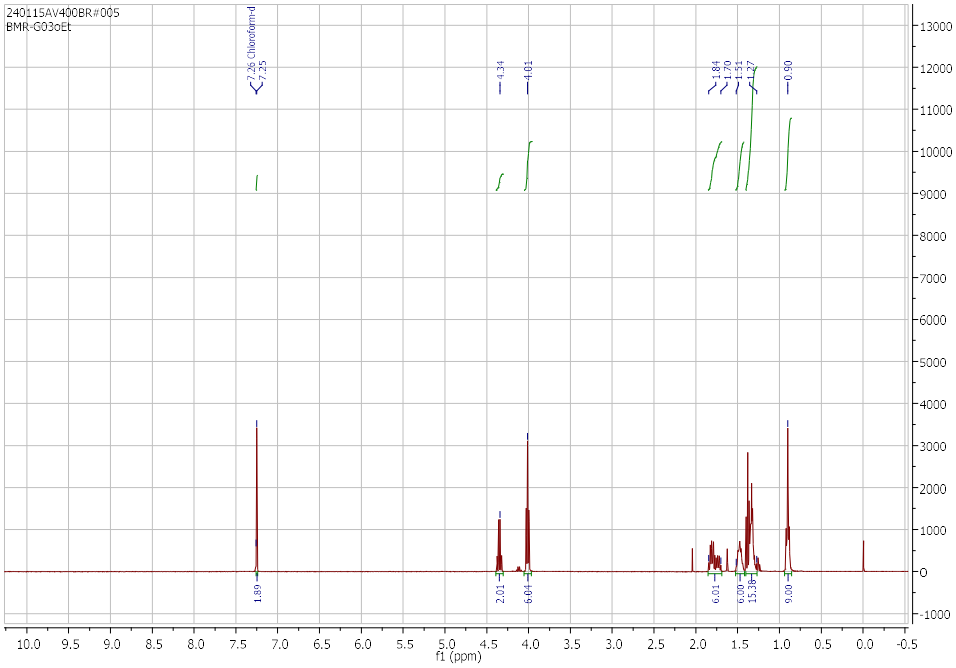


^{13}C NMR spectrum

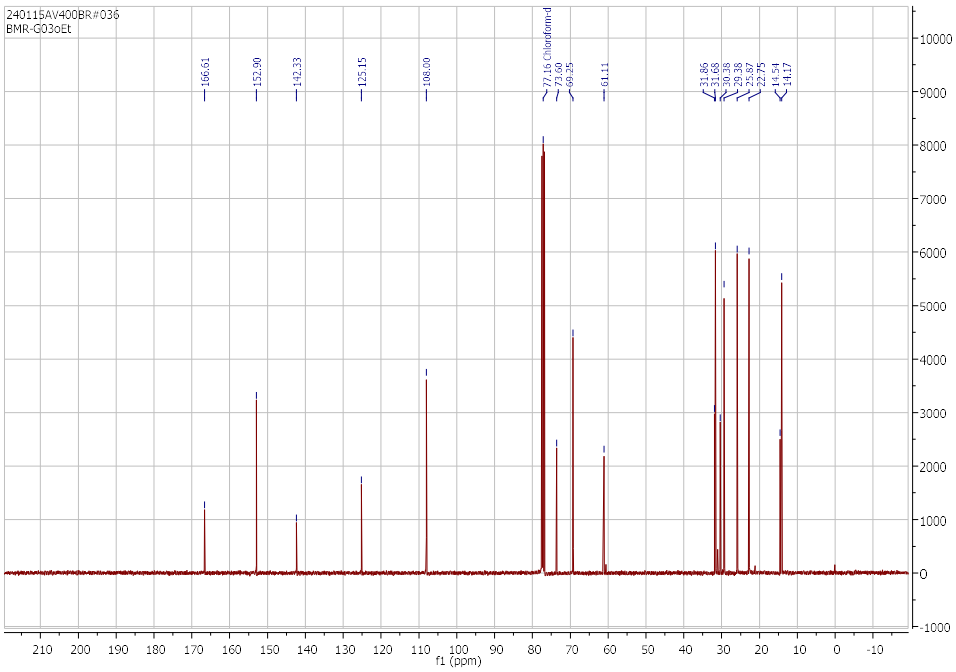
Compound 7a



7a

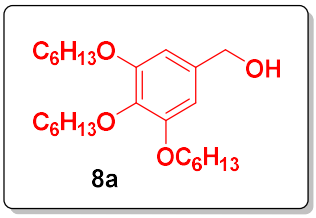


^{1}H NMR spectrum

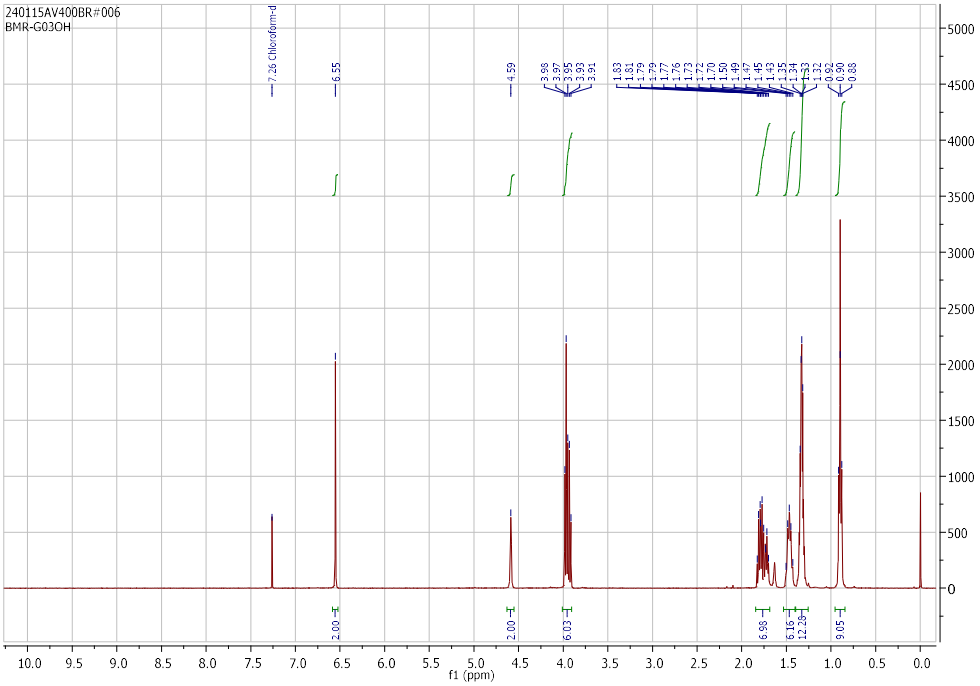


^{13}C NMR spectrum

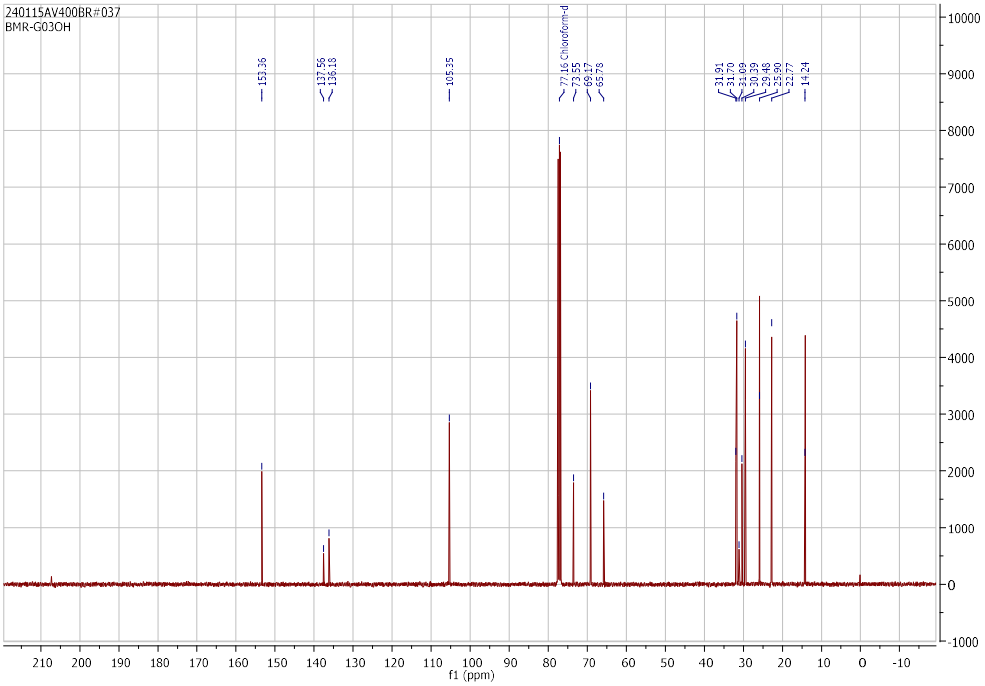
Compound 8a



8a

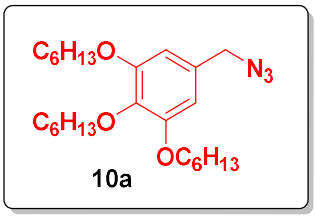


^{1}H NMR spectrum

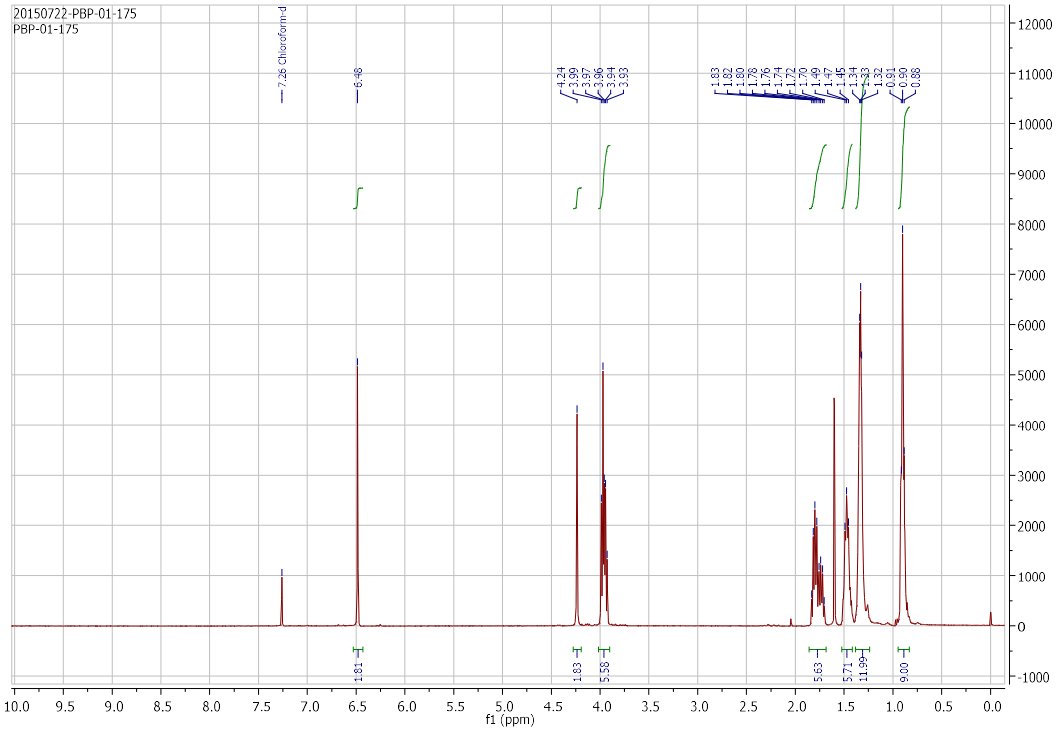


^{13}C NMR spectrum

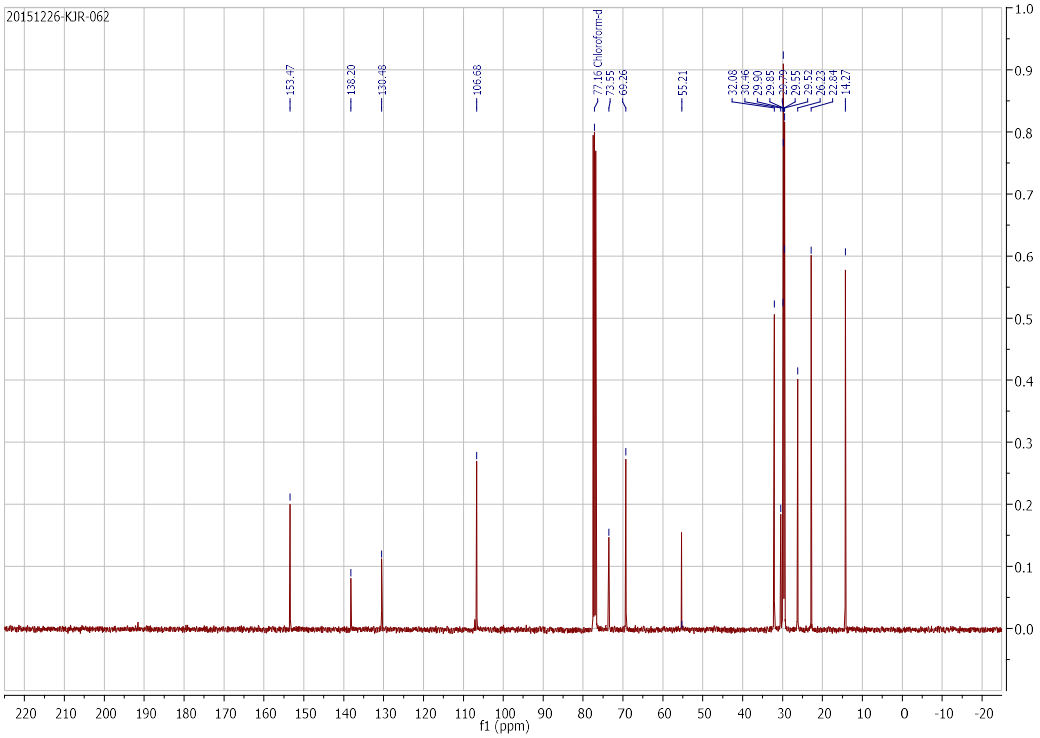
Compound 10a



10a

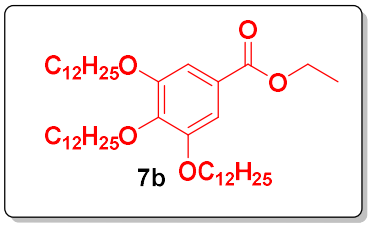


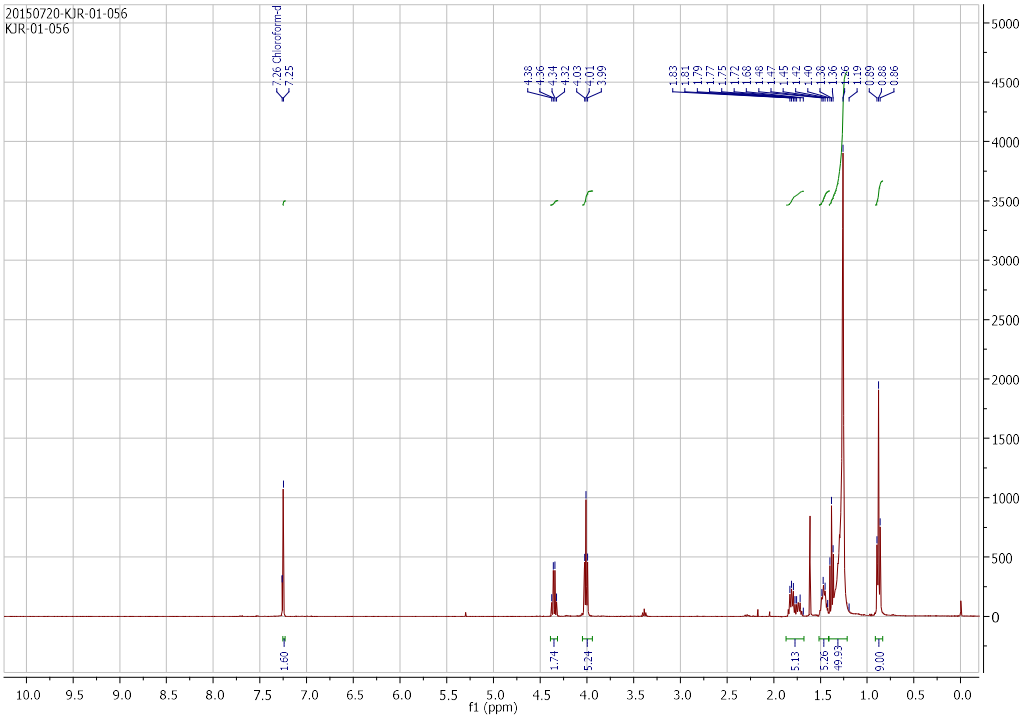
^{1}H NMR spectrum



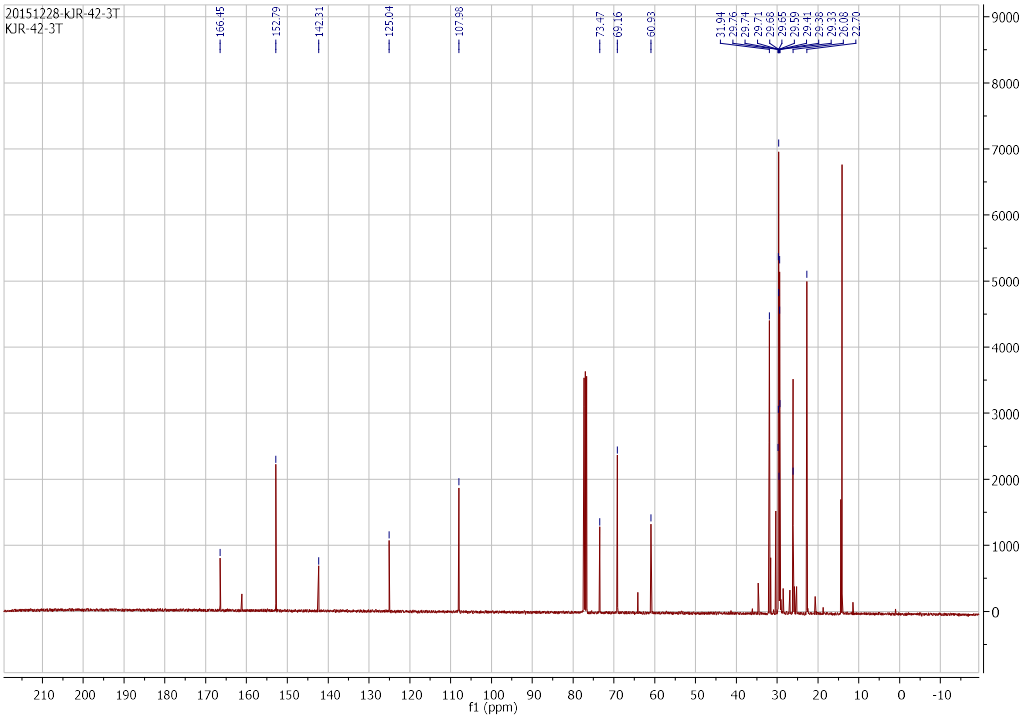
^{13}C NMR spectrum

Compound 7b



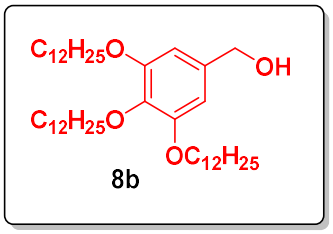


^{1}H NMR spectrum

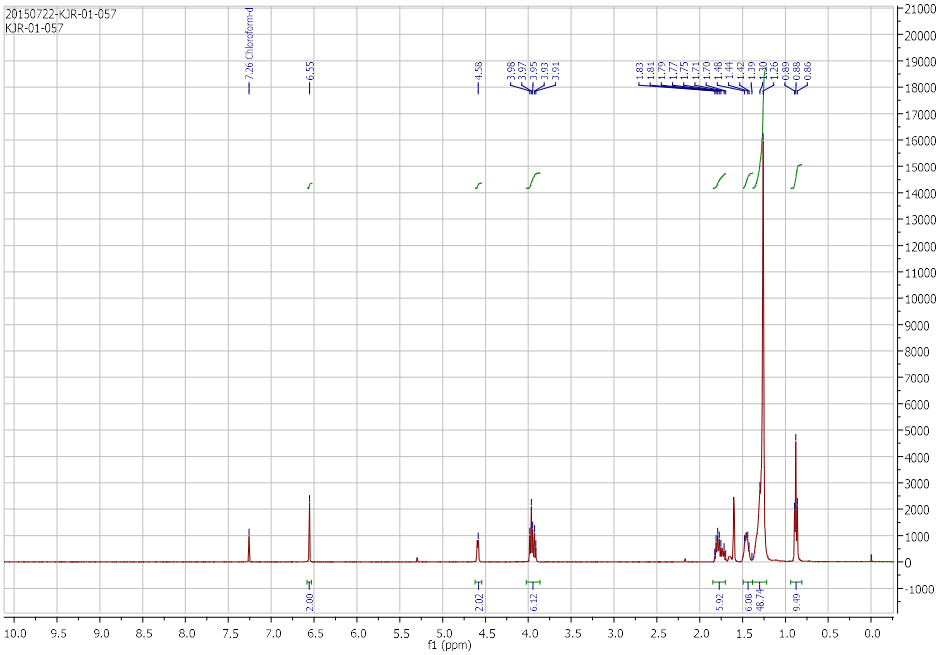


^{13}C NMR spectrum

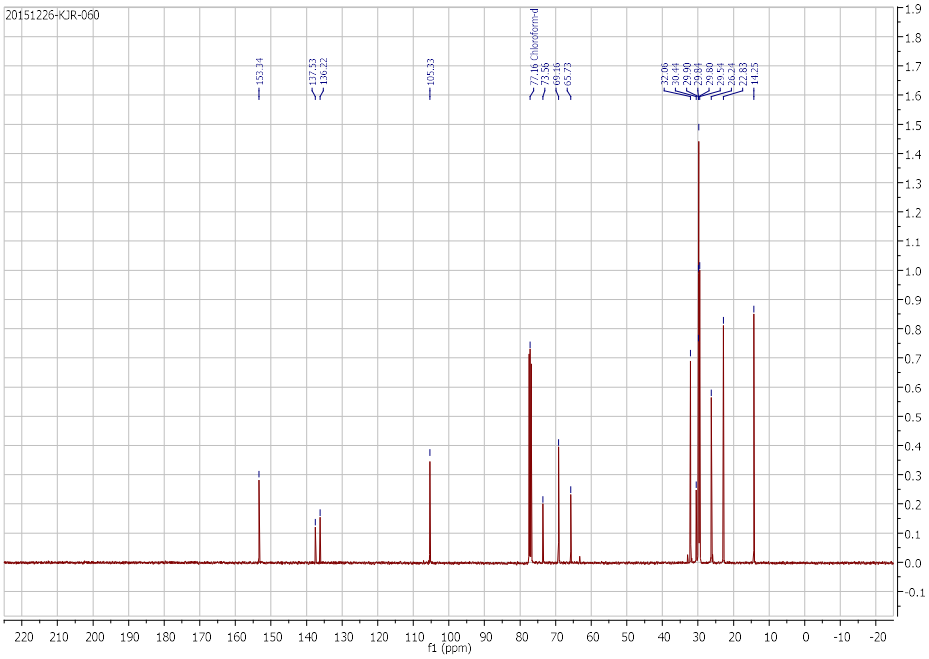
Compound 8b



8b

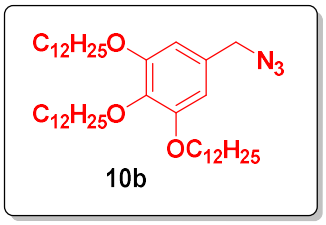


^{1}H NMR spectrum

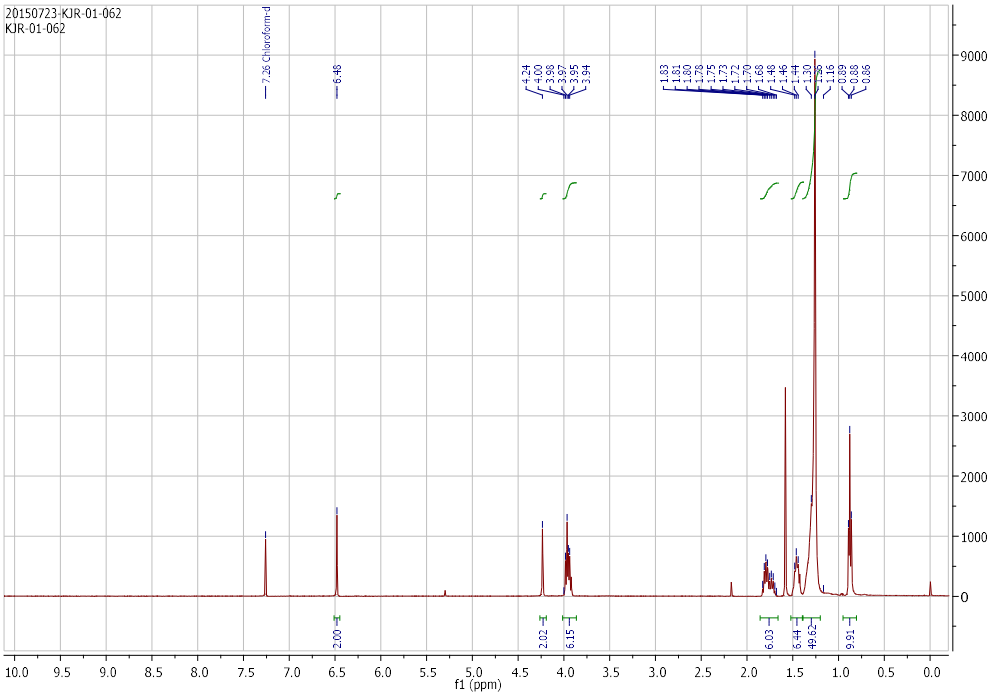


^{13}C NMR spectrum

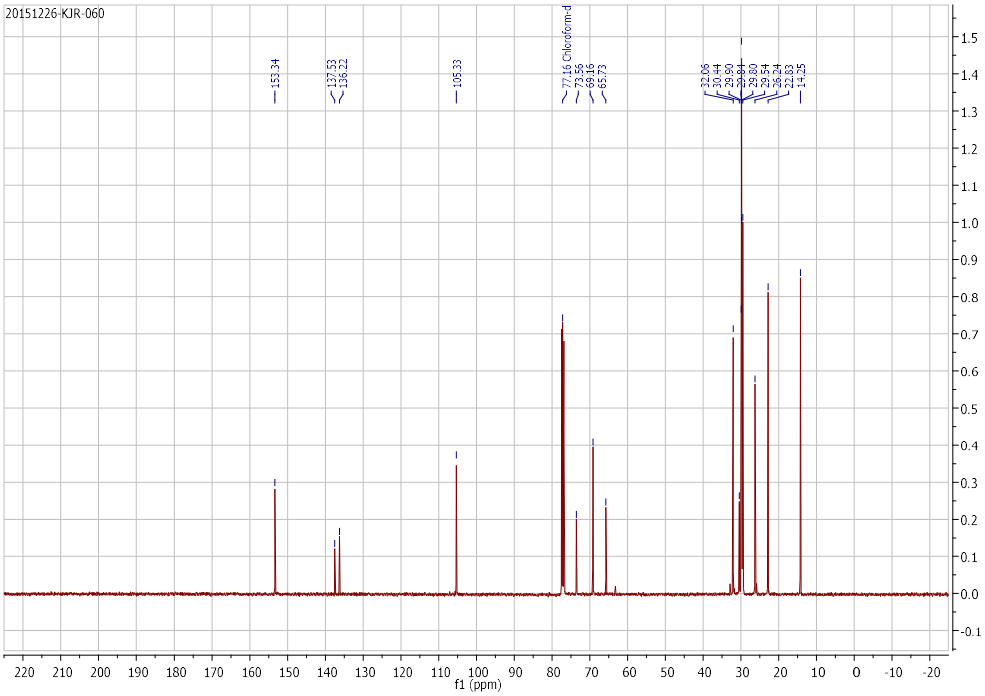
Compound 10b



10b

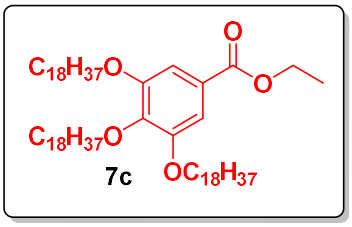


^{1}H NMR spectrum

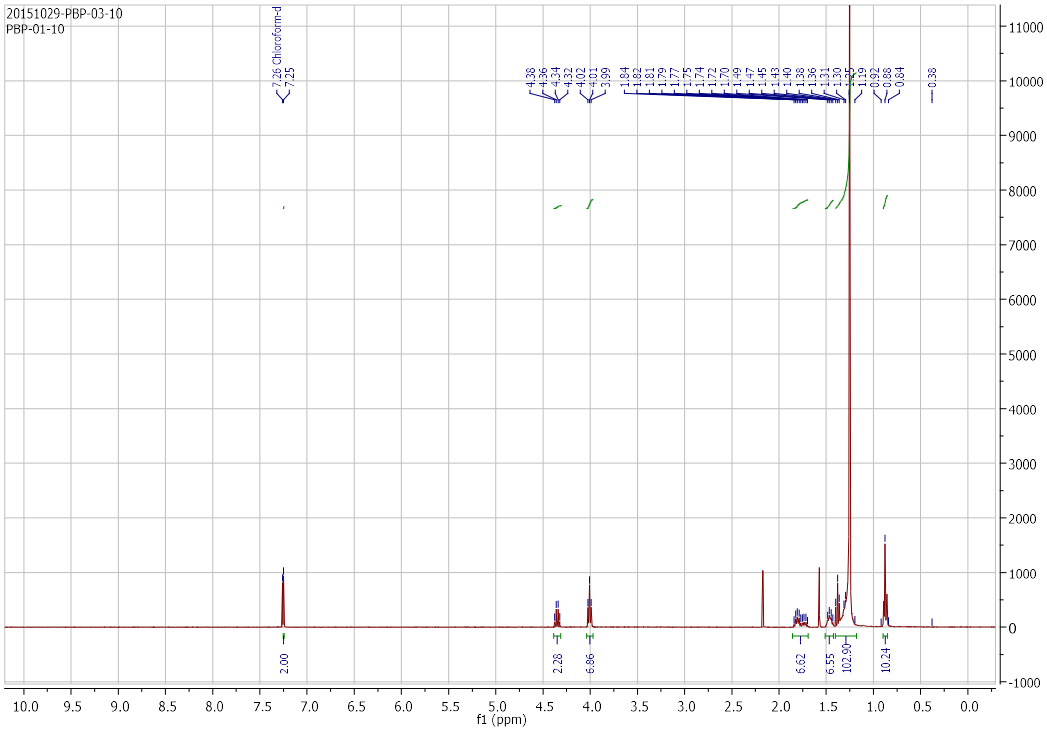


^{13}C NMR spectrum

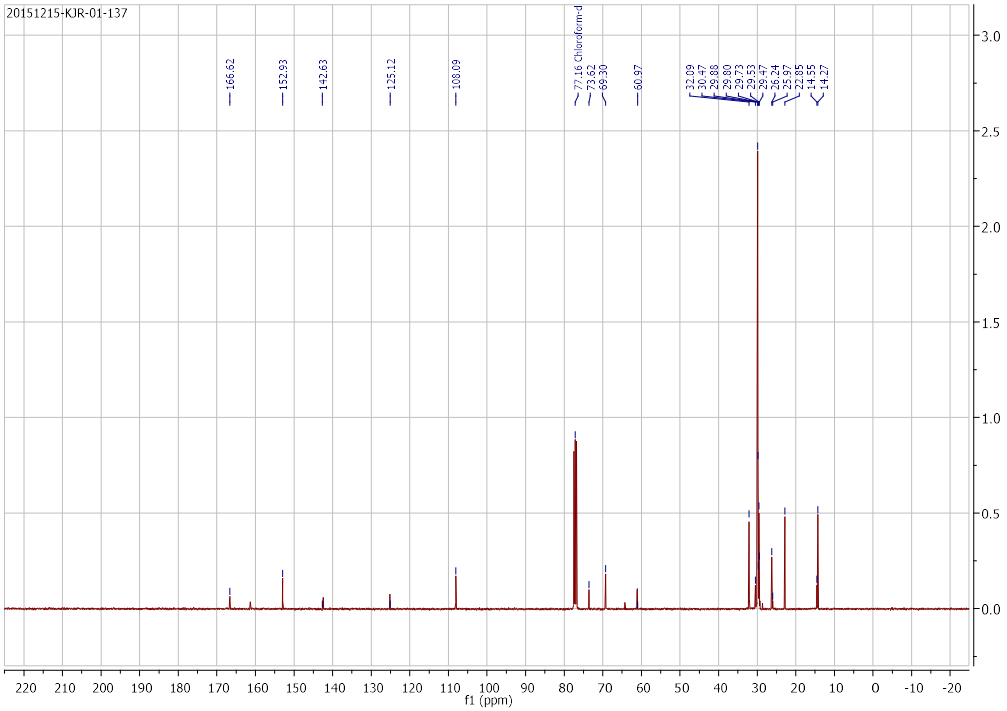
Compound 7c



7c

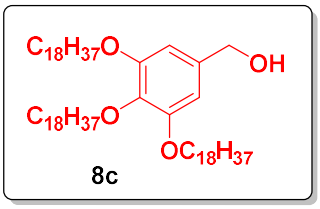


^{1}H NMR spectrum

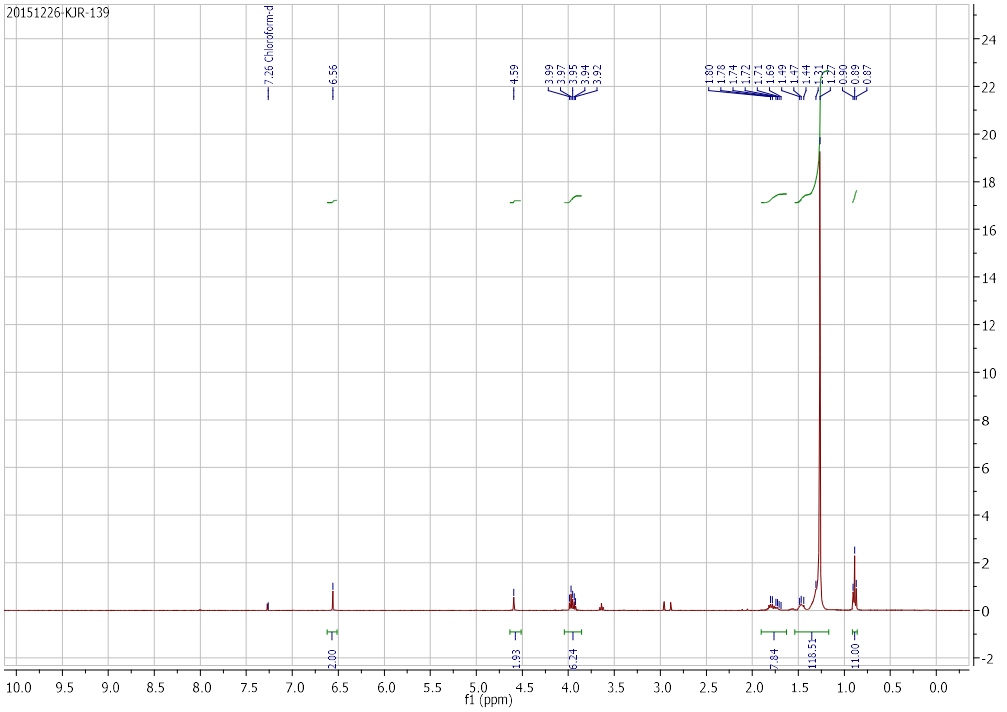


^{13}C NMR spectrum

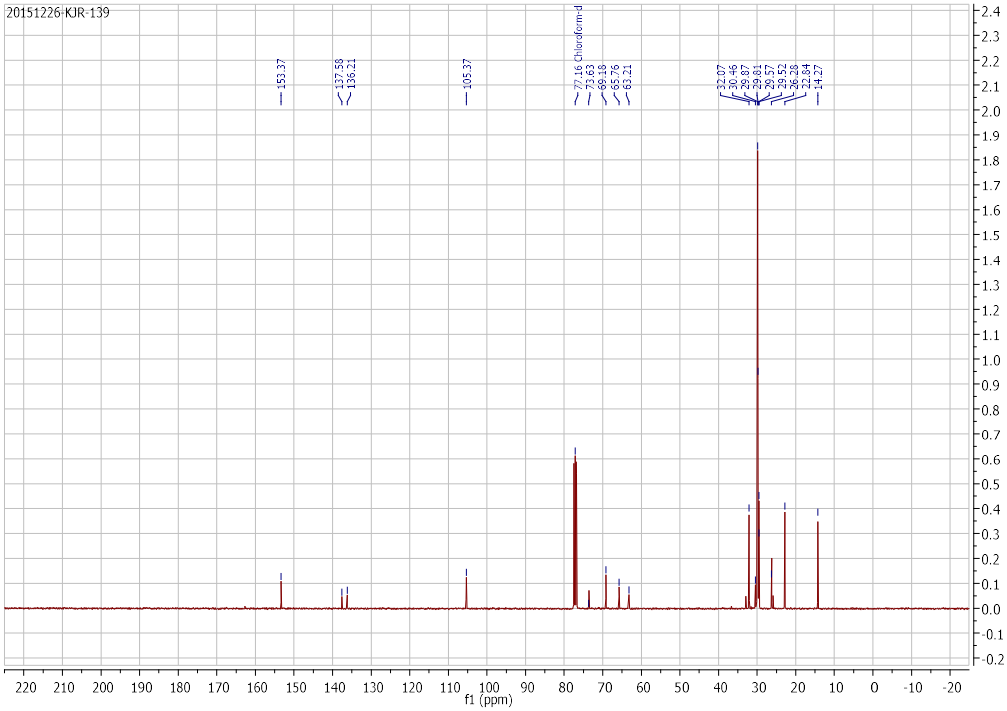
Compound 8c



8c

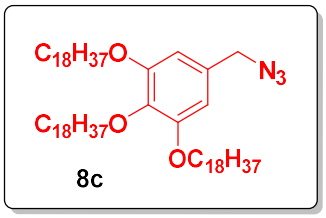


^{1}H NMR spectrum

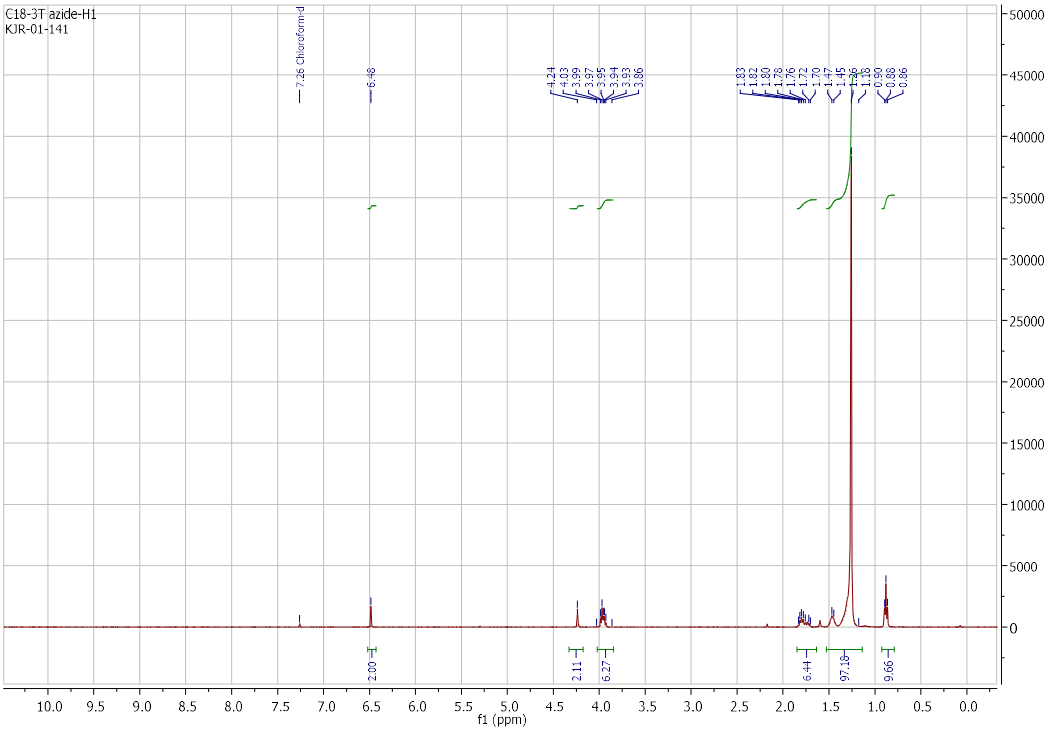


^{13}C NMR spectrum

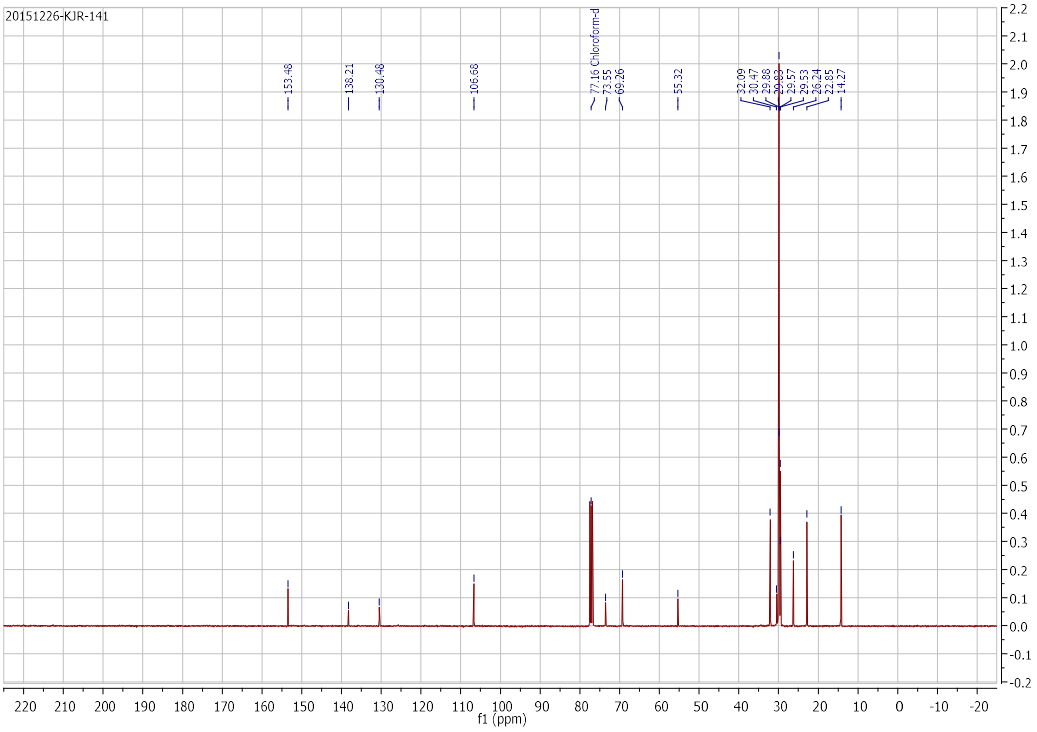
Compound 10c



8c

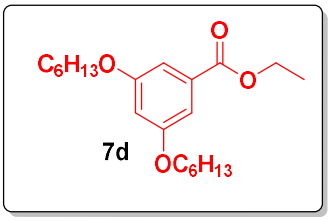


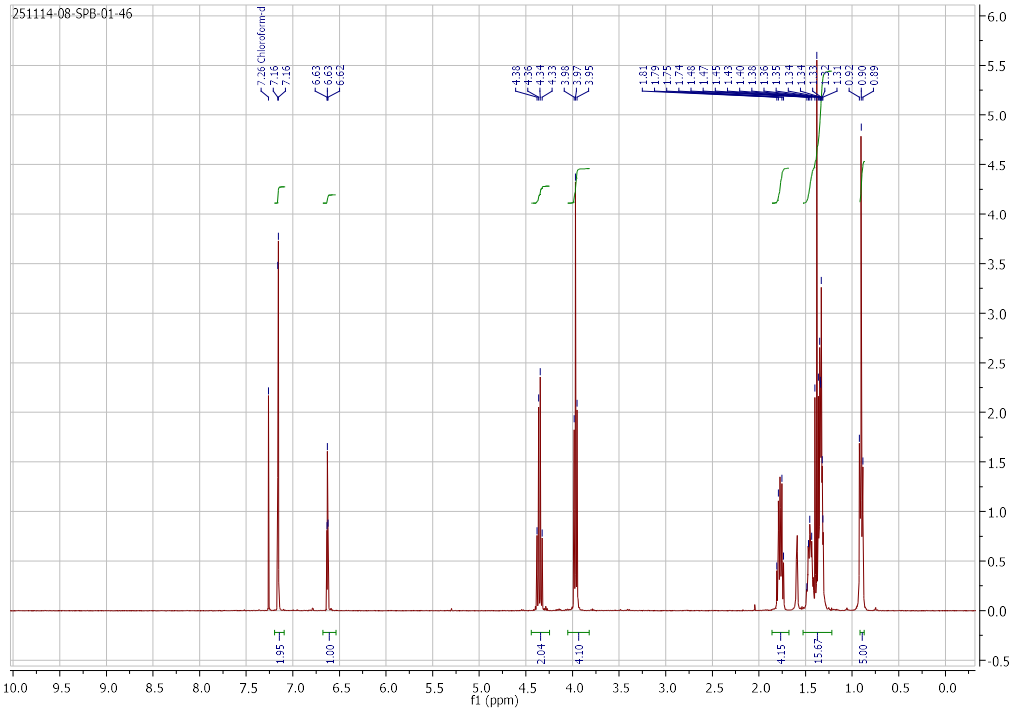
^{1}H NMR spectrum



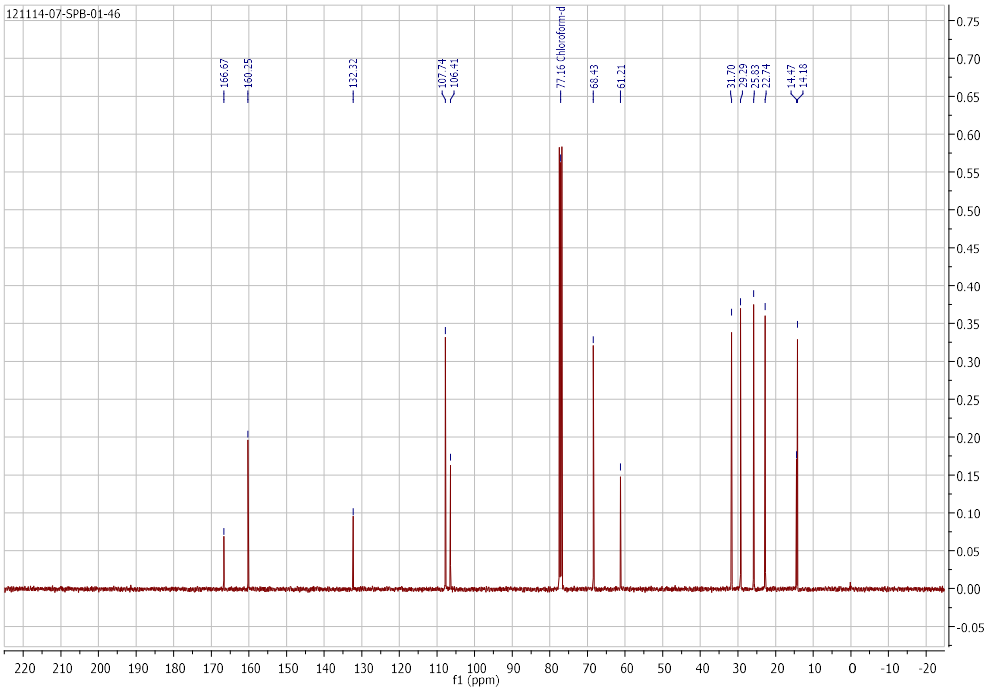
^{13}C NMR spectrum

Compound 7d



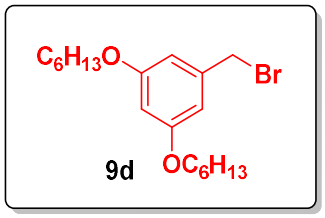


^{1}H NMR spectrum

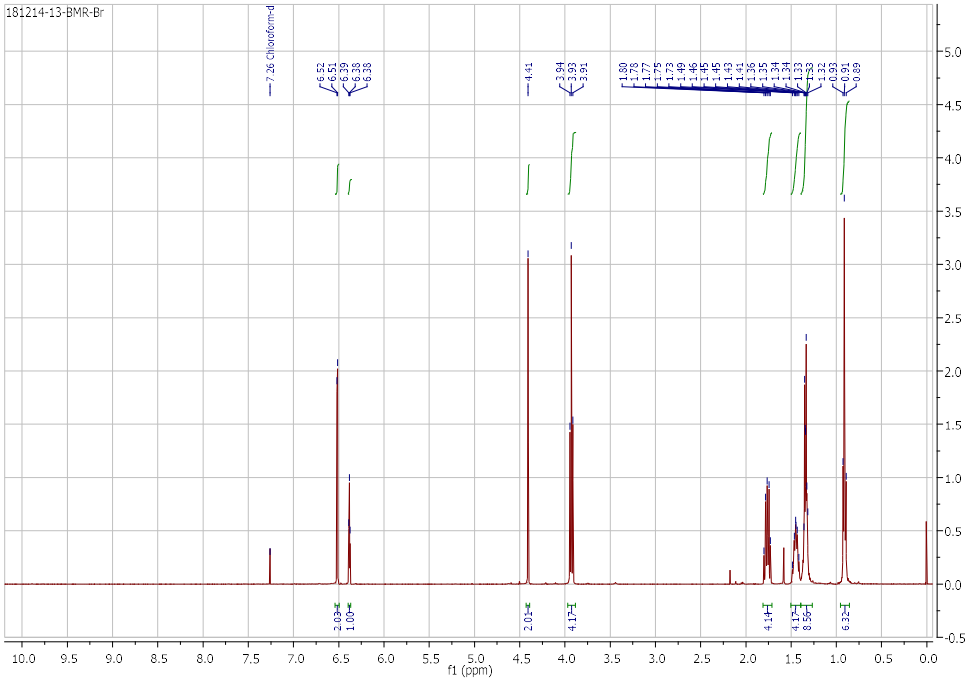


^{13}C NMR spectrum

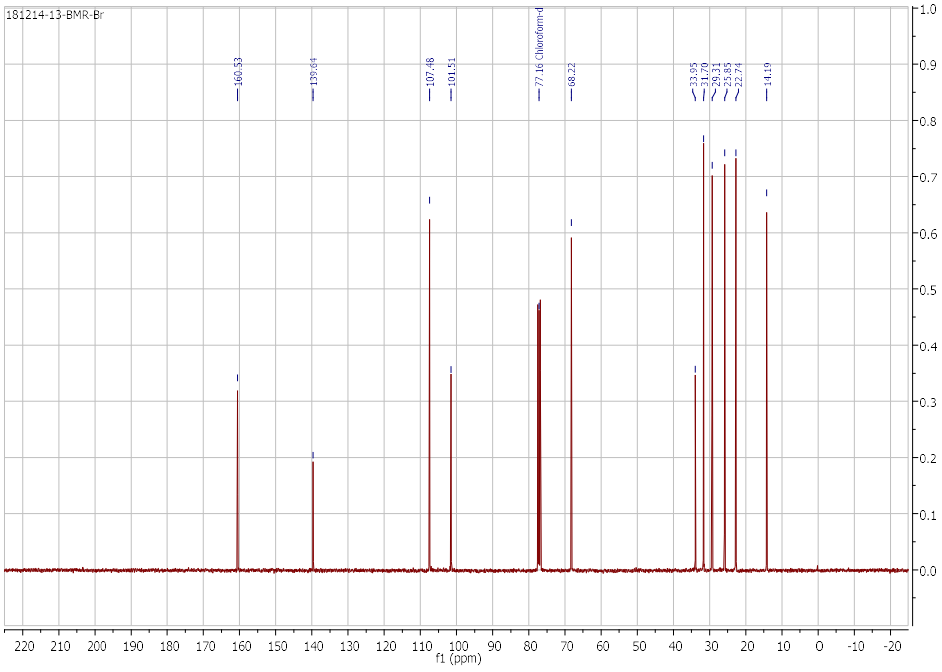
Compound 9d



9d

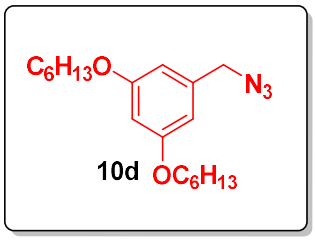


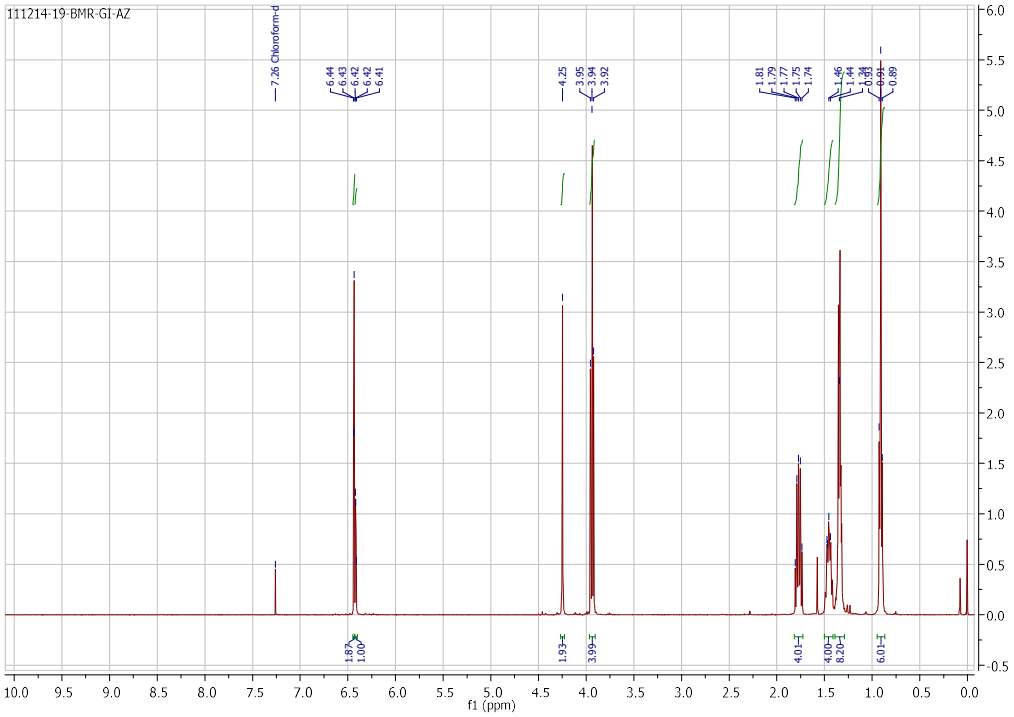
^{1}H NMR spectrum



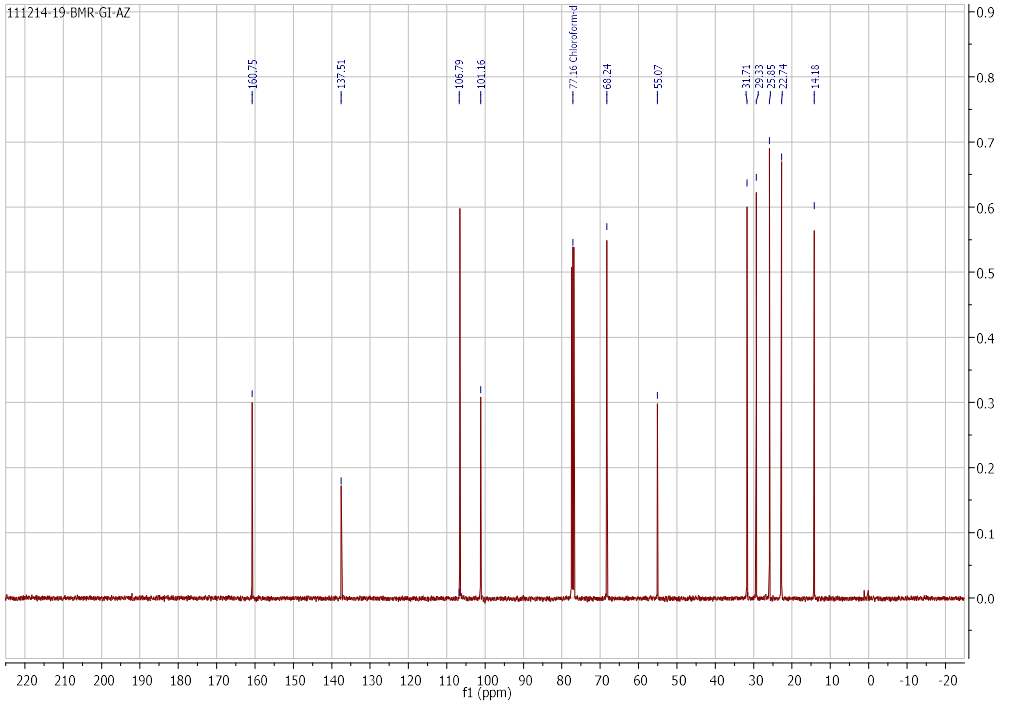
^{13}C NMR spectrum

Compound 10d



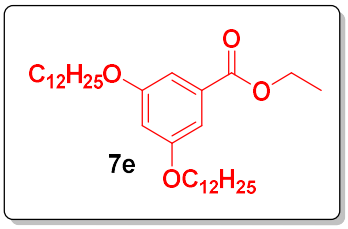


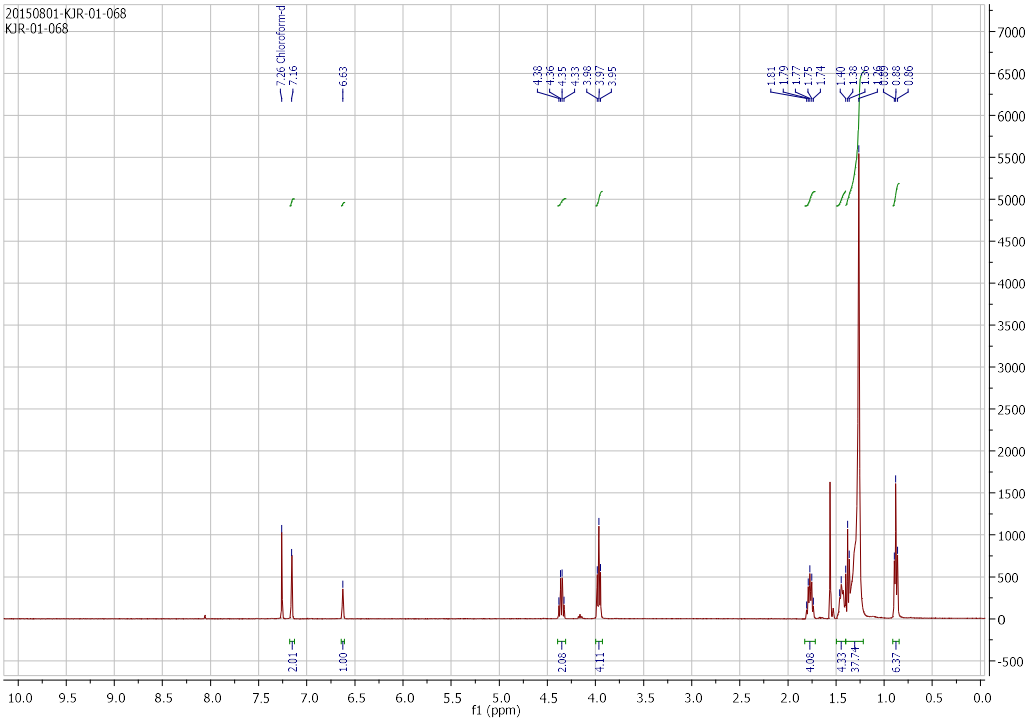
^{1}H NMR spectrum



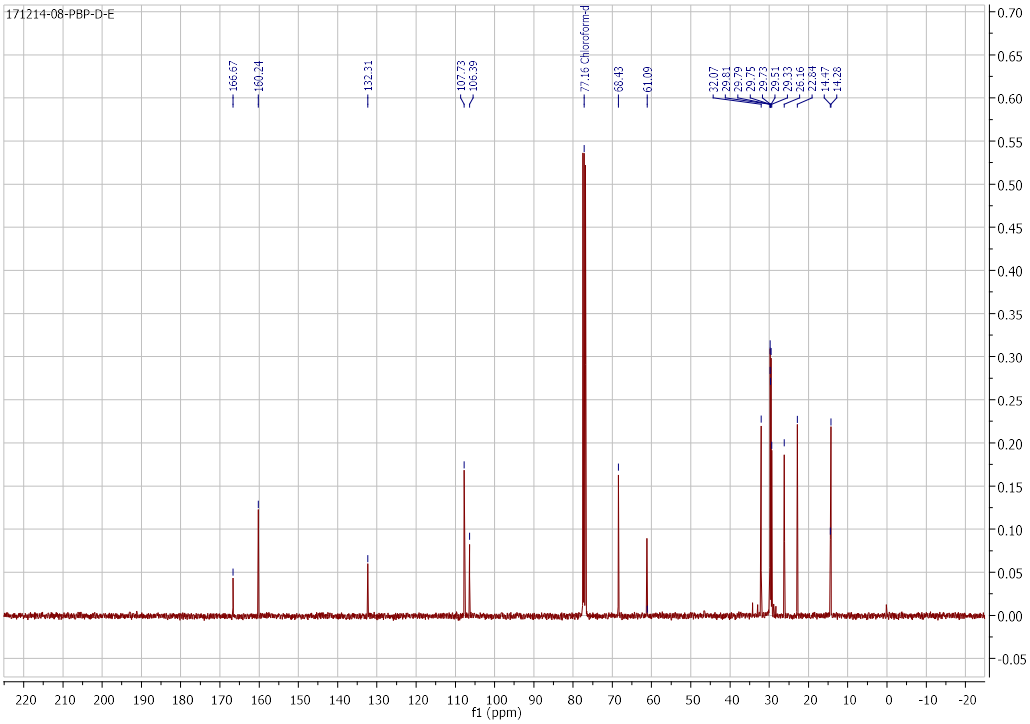
^{13}C spectrum

Compound 7e



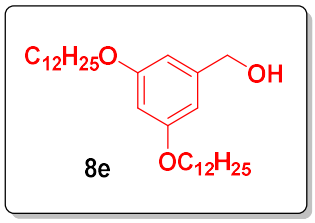


^{1}H NMR spectrum

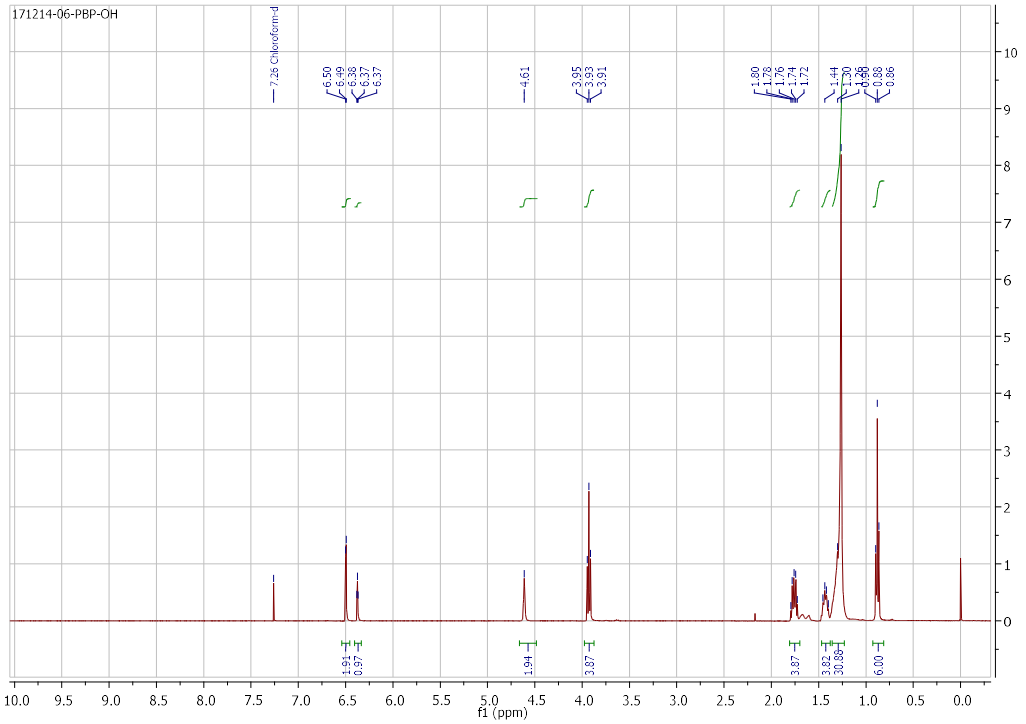


^{13}C NMR spectrum

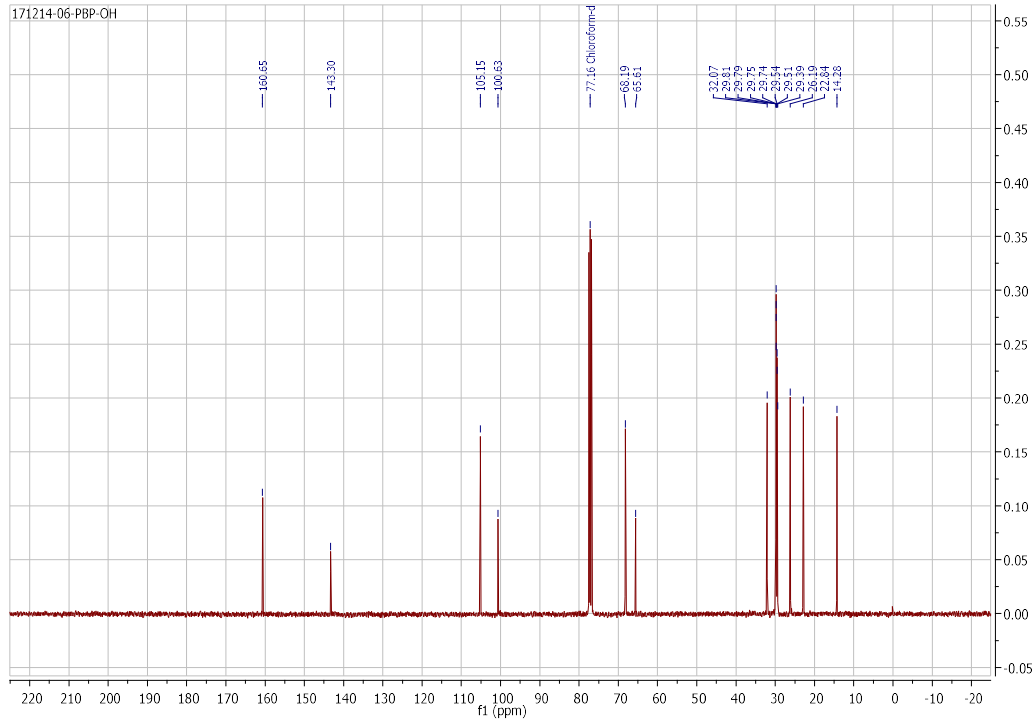
Compound 8e



8e

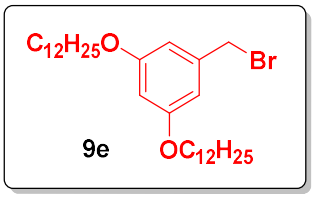


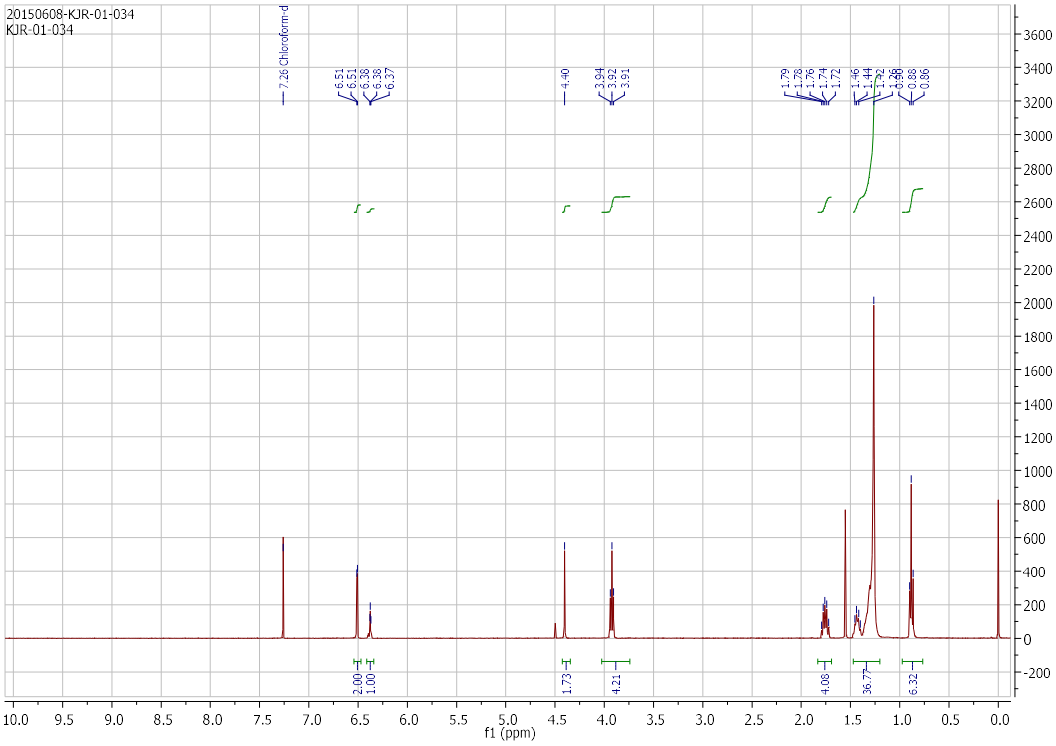
^{1}H NMR spectrum



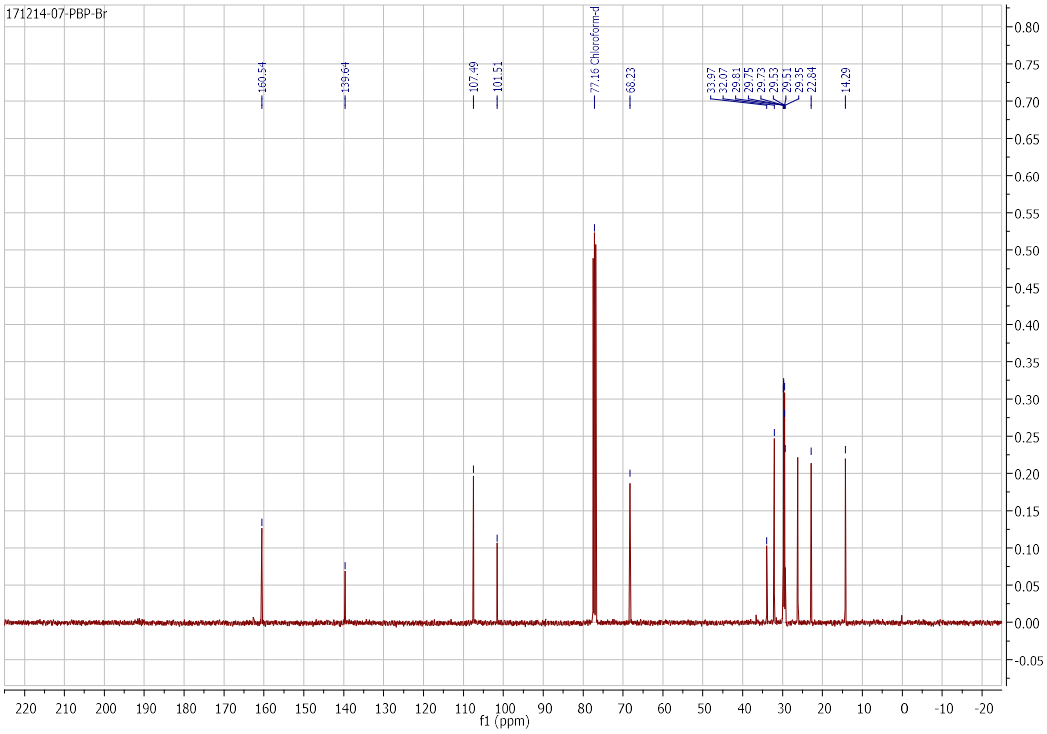
^{13}C NMR spectrum

Compound 9e



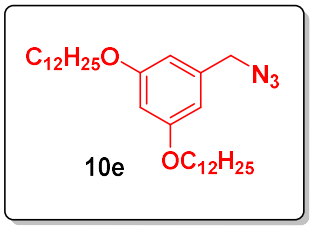


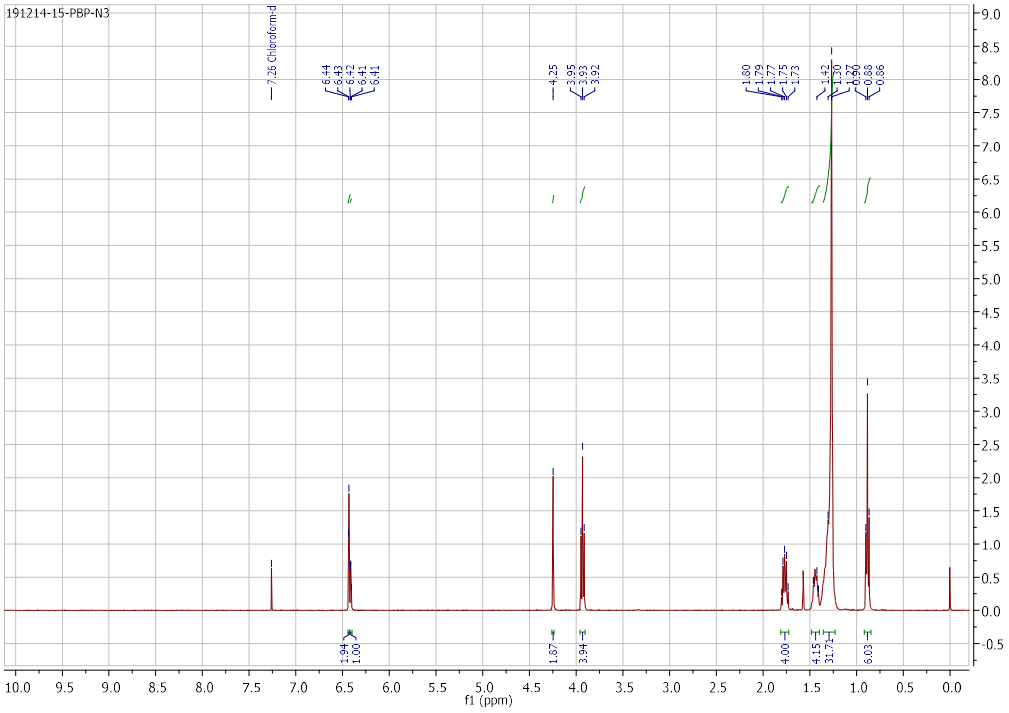
^{1}H NMR spectrum



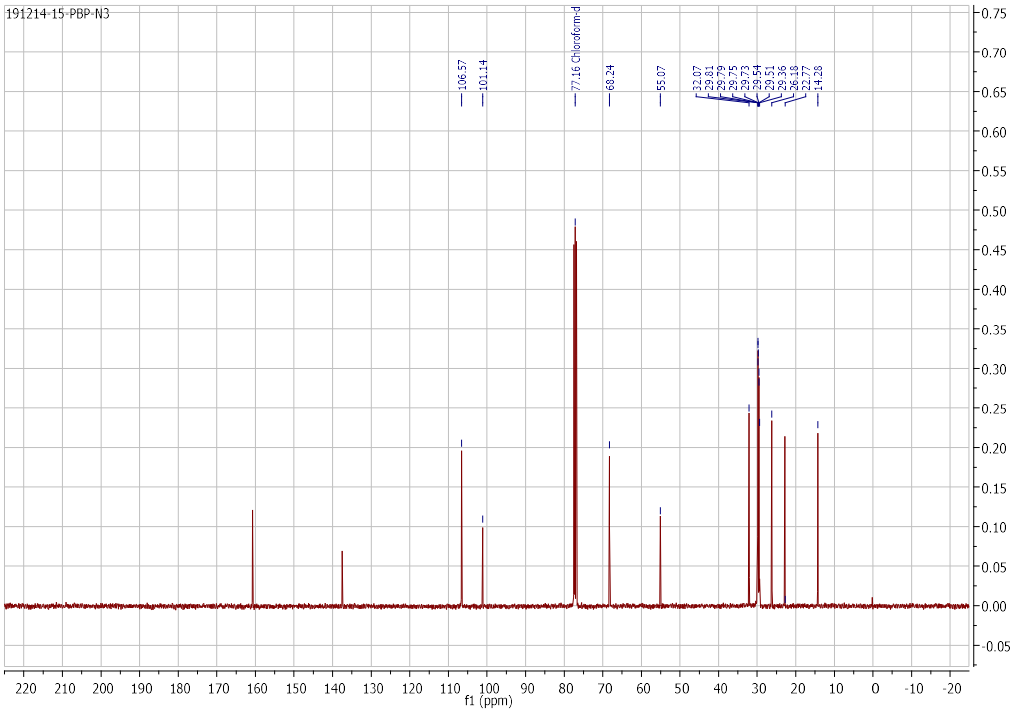
^{13}C NMR spectrum

Compound 10e



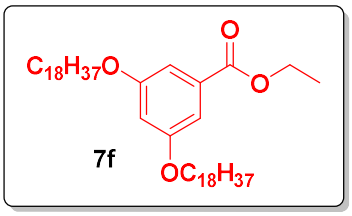


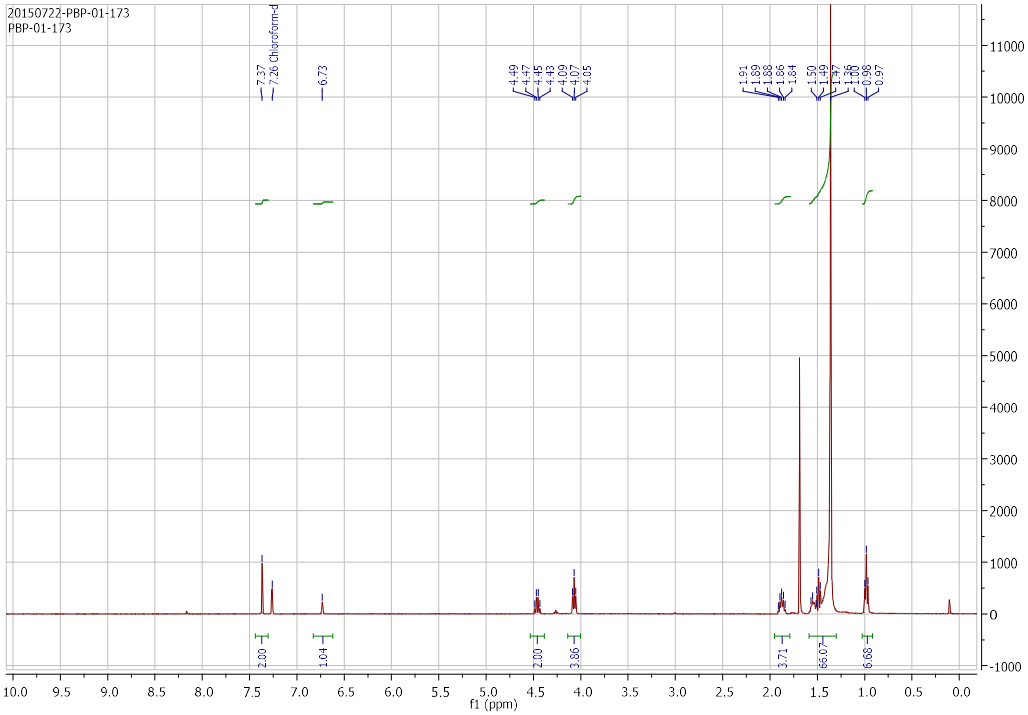
^{1}H NMR spectrum



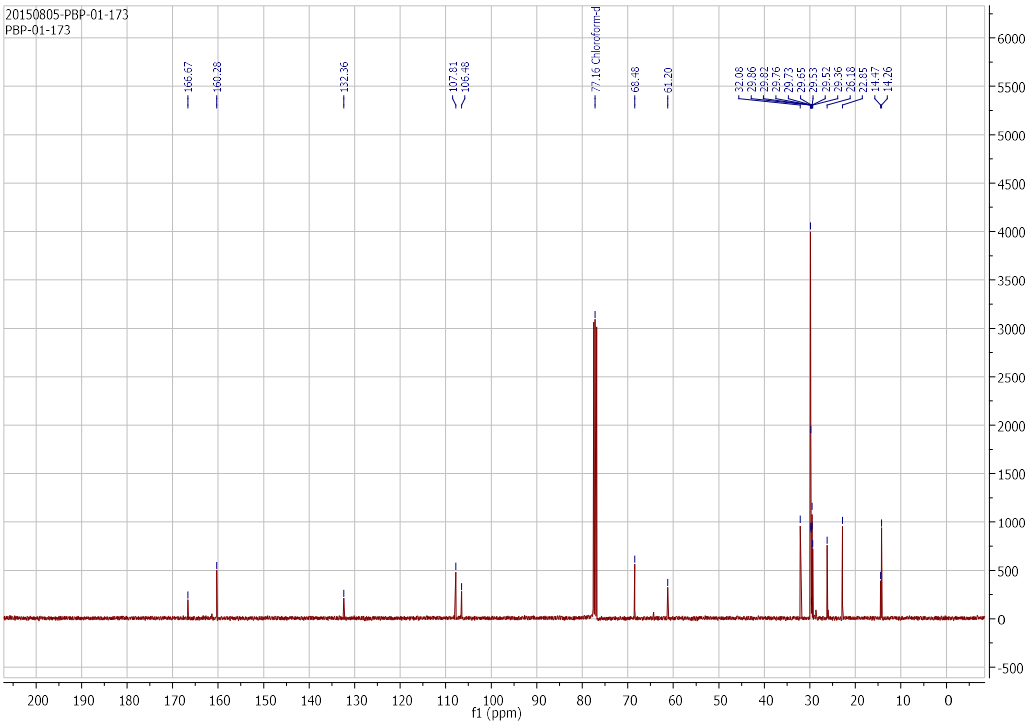
^{13}C NMR spectrum

Compound 7f



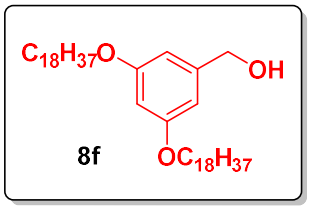


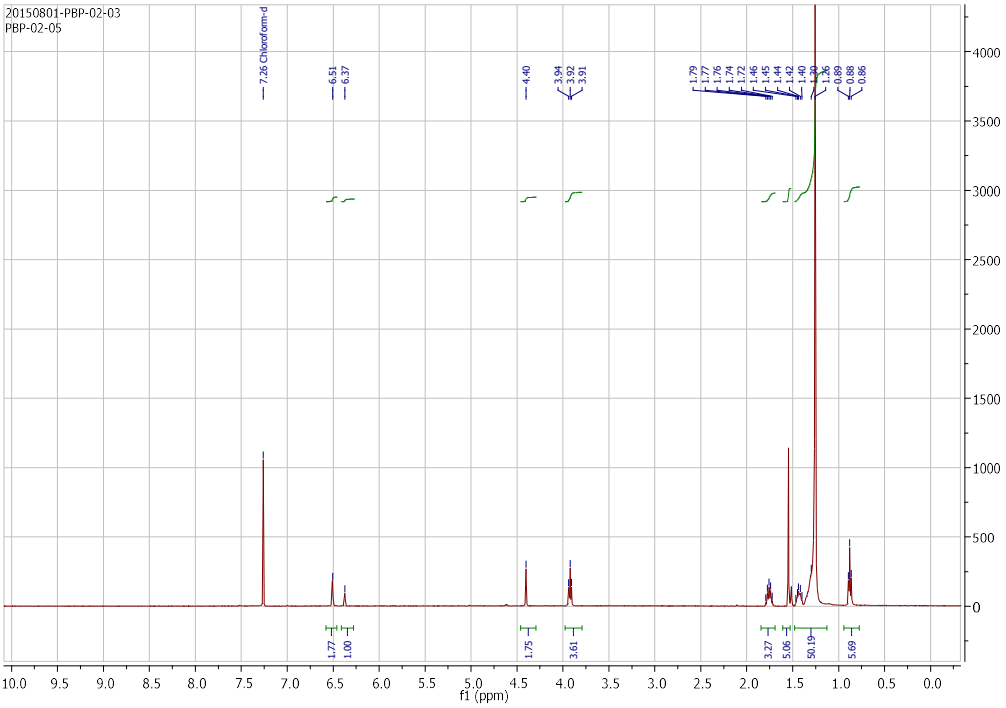
^{1}H NMR spectrum



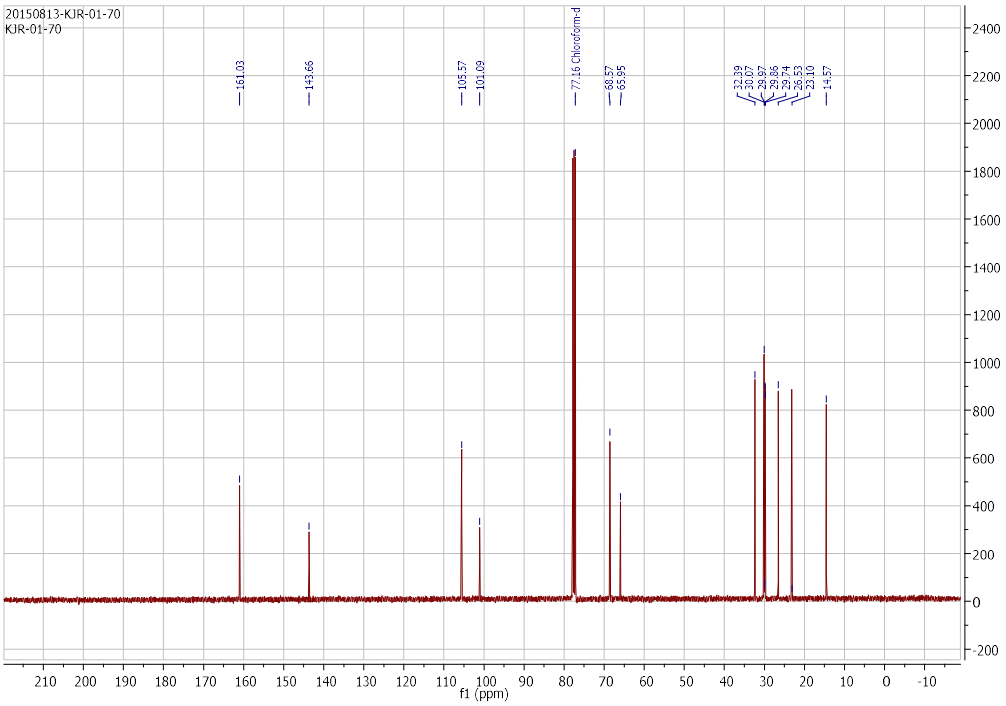
^{13}C NMR spectrum

Compound 8f



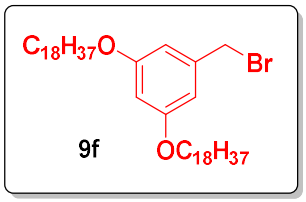


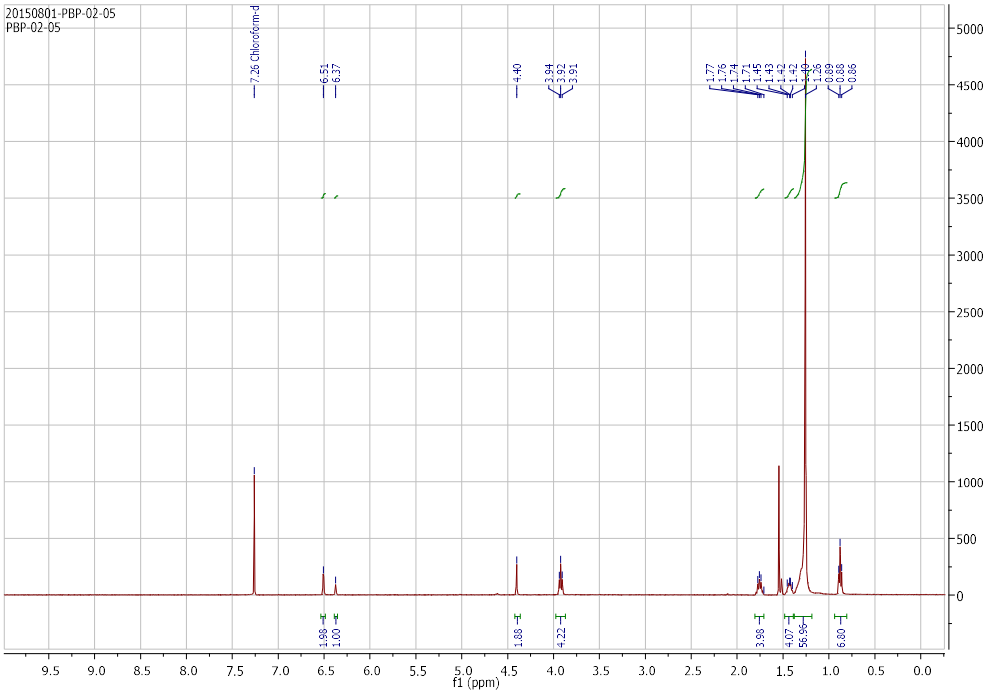
^{1}H NMR spectrum



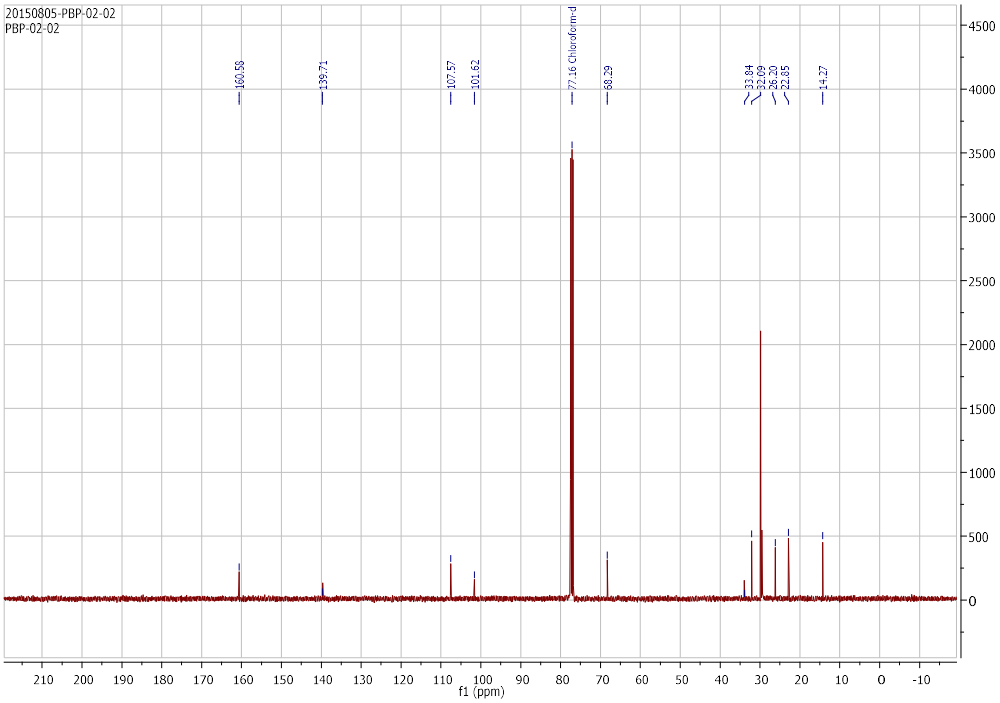
^{13}C NMR spectrum

Compound 9f



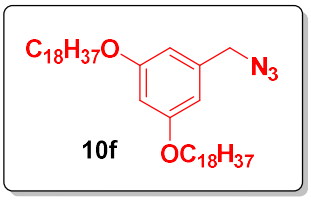


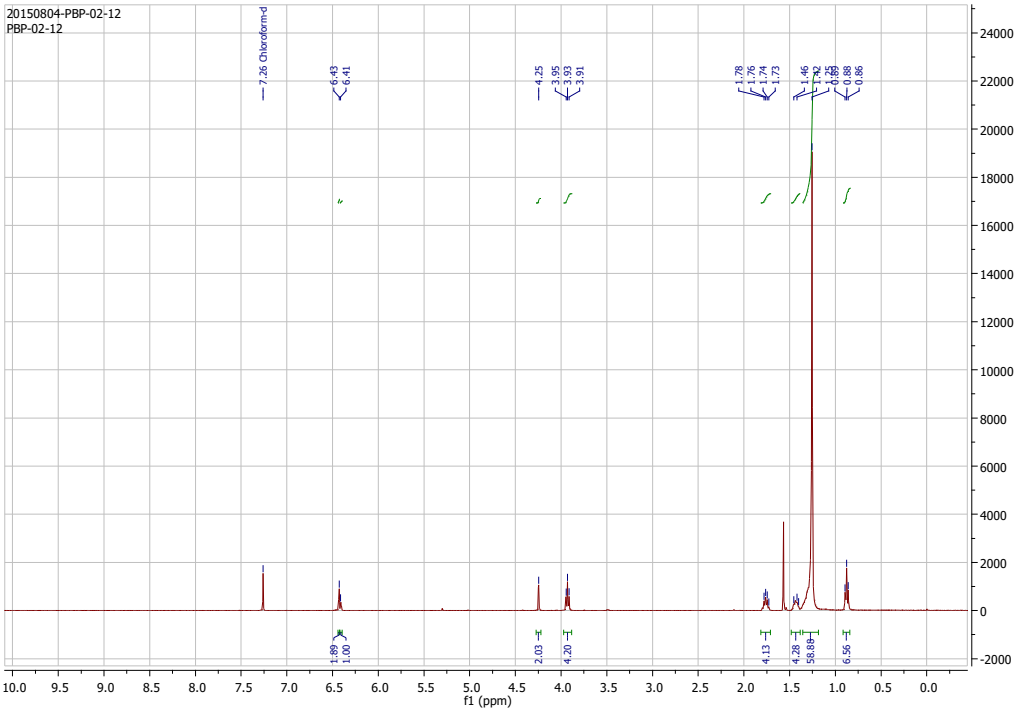
^{1}H NMR spectrum



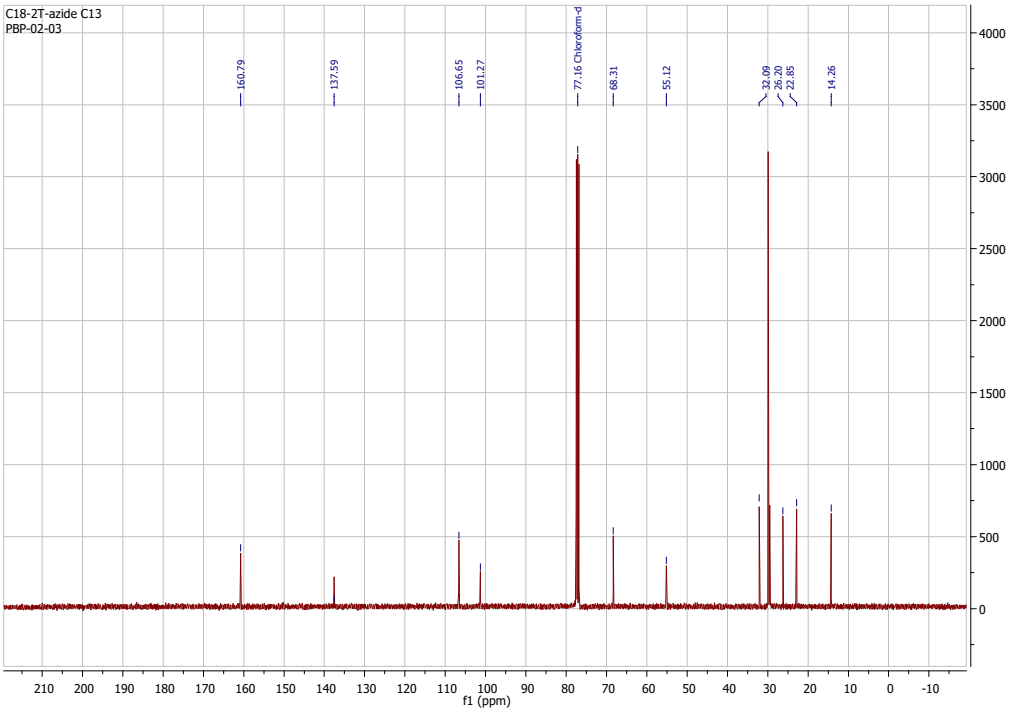
^{13}C NMR spectrum

Compound 10f



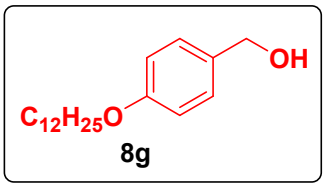


^{1}H NMR spectrum

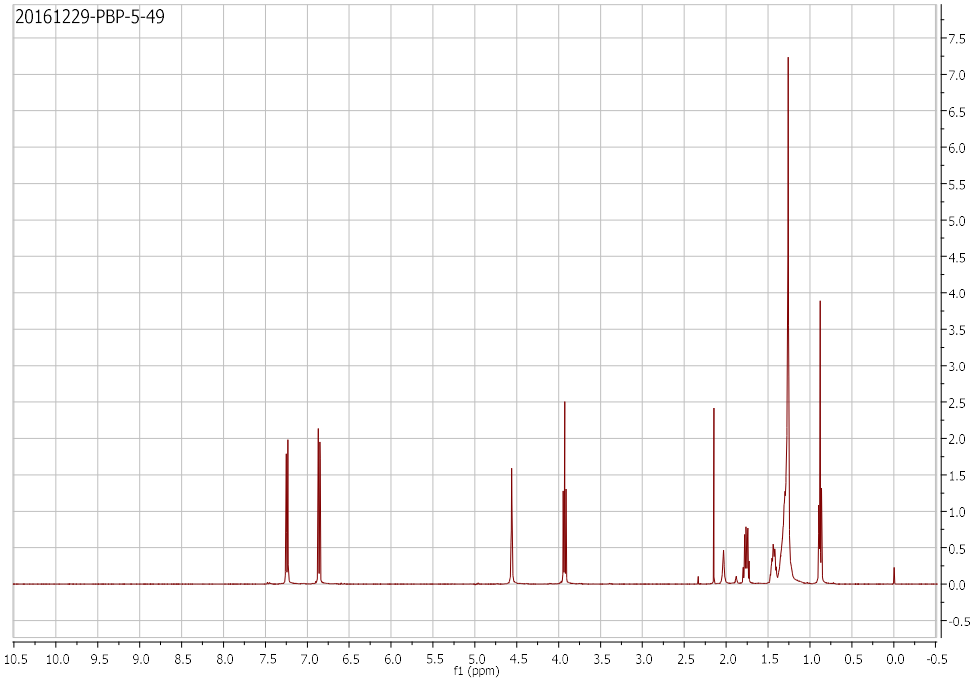


^{13}C NMR spectrum

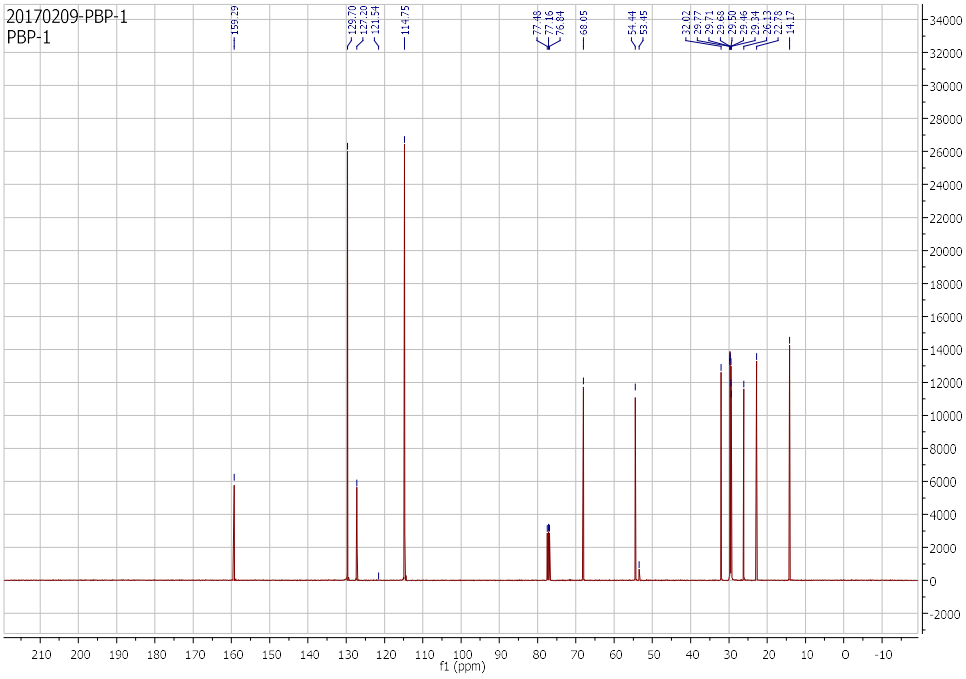
Compound 8g



8g

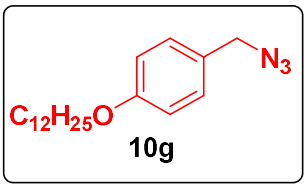


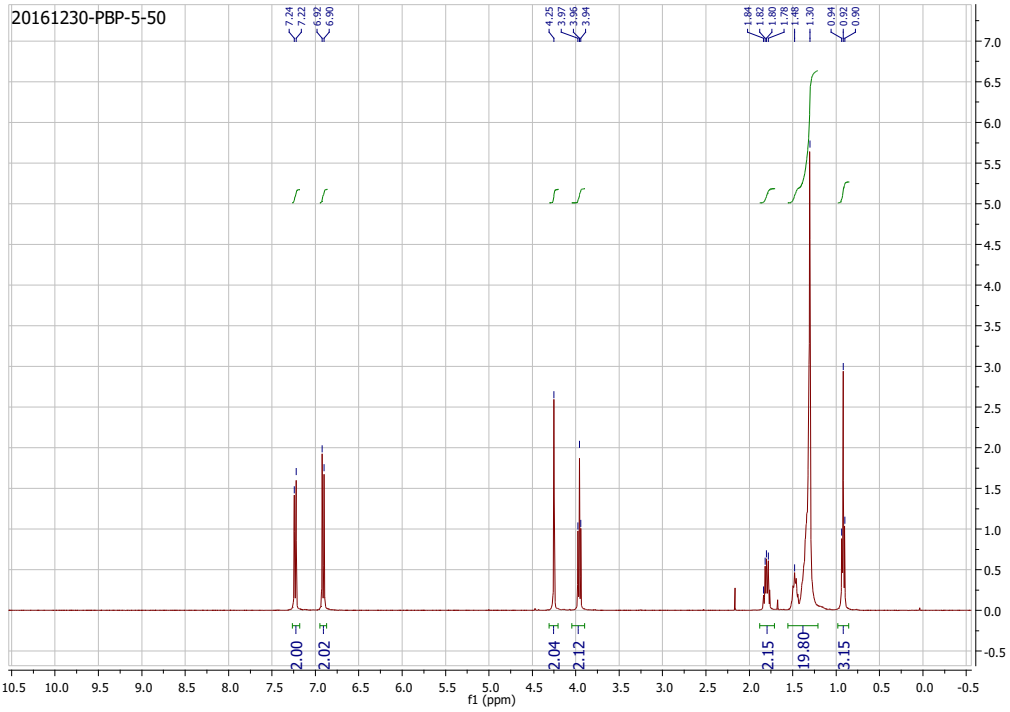
^{1}H NMR spectrum



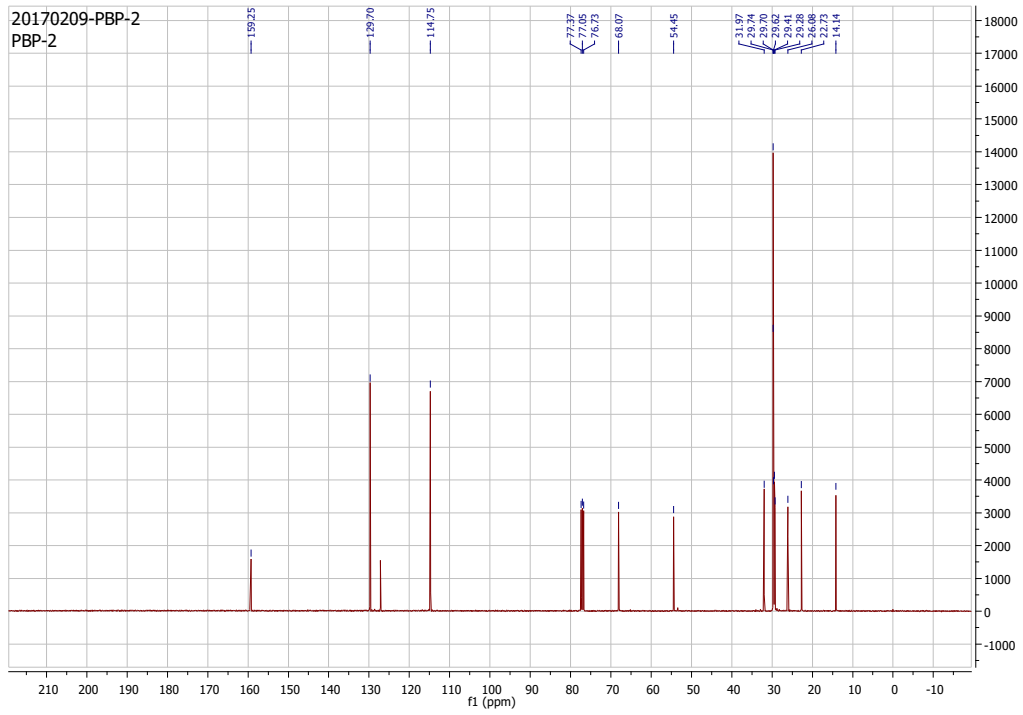
^{13}C NMR spectrum

Compound 10g



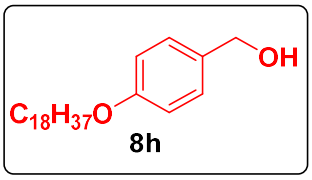


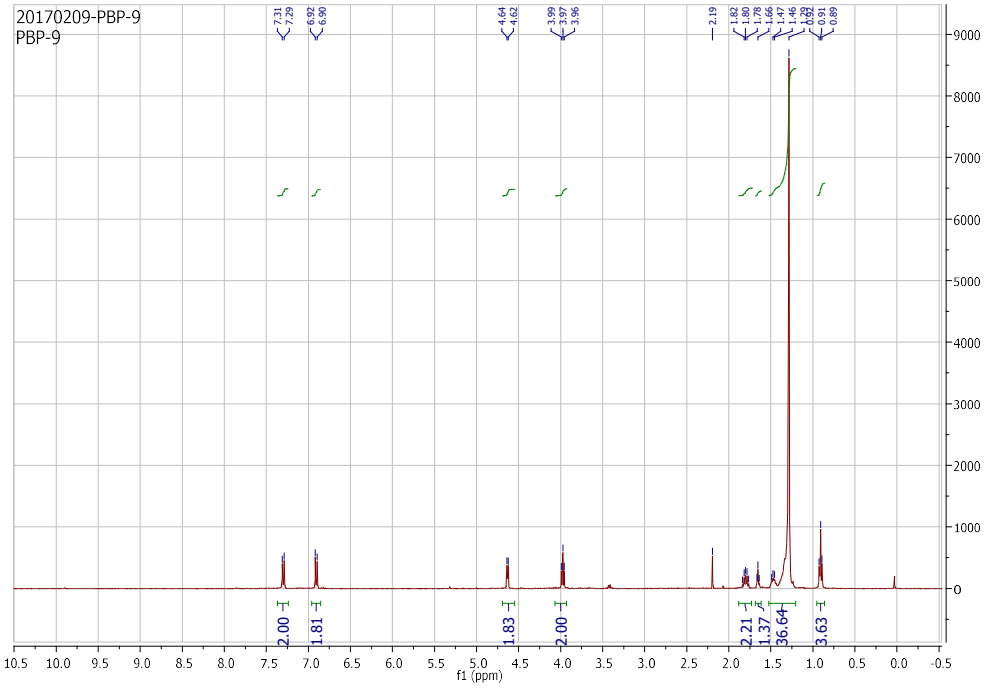
^{1}H NMR spectrum



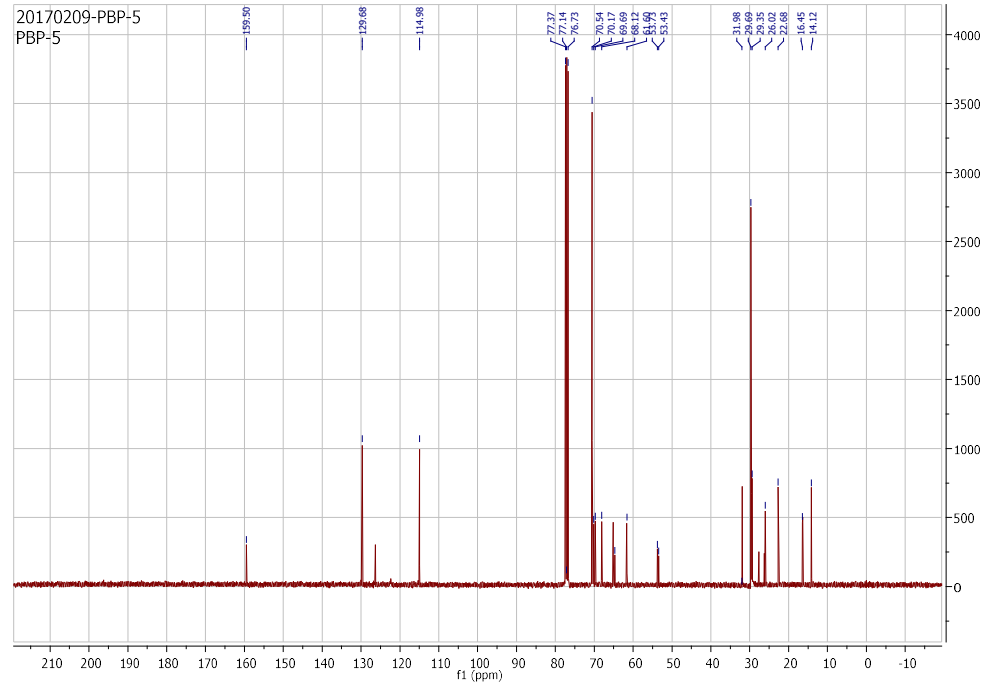
^{13}C NMR spectrum

Compound 8h



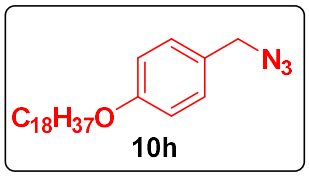


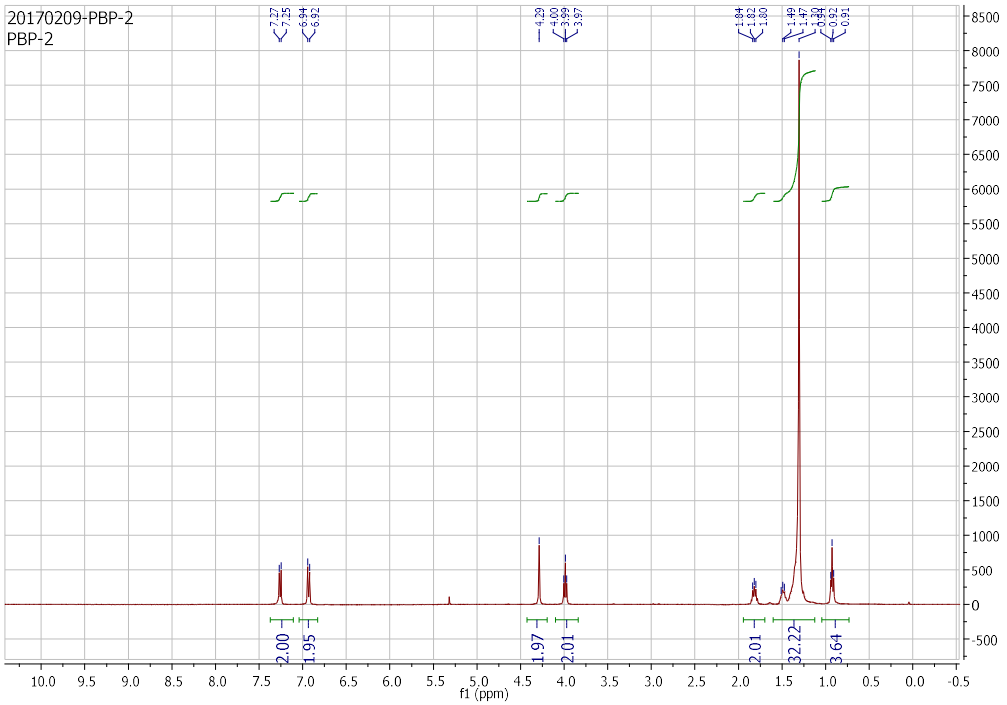
^{1}H NMR spectrum



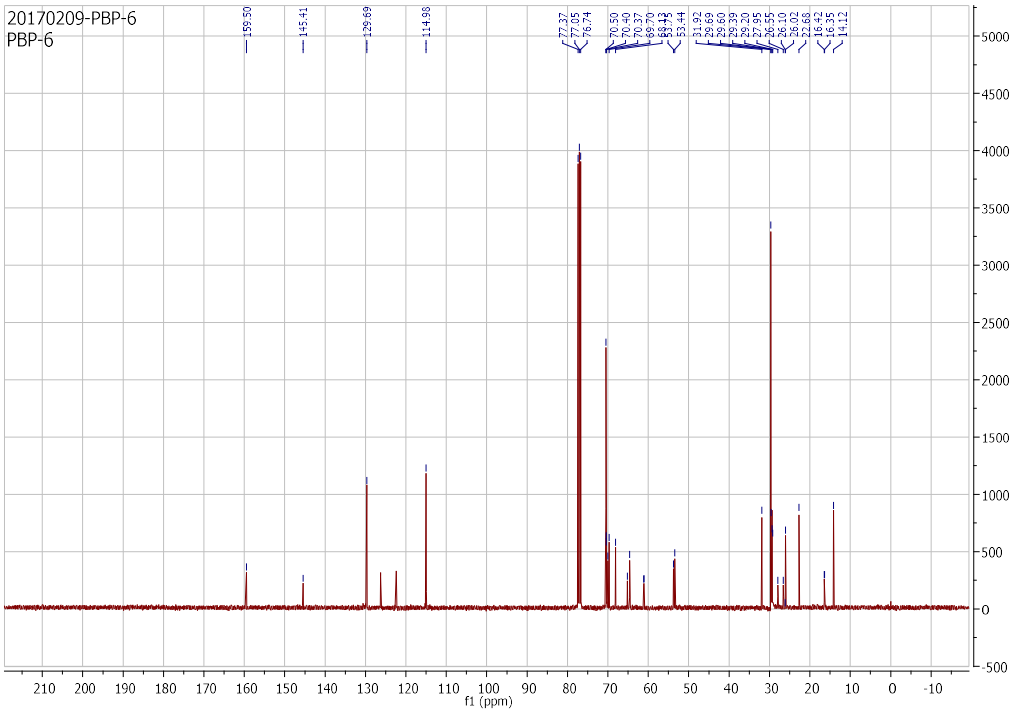
^{13}C NMR spectrum

Compound 10h



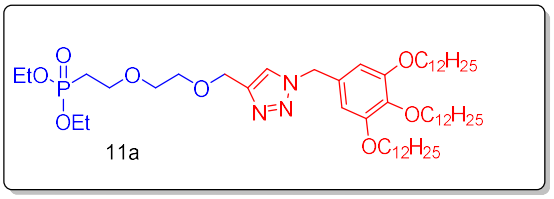


^{1}H NMR spectrum

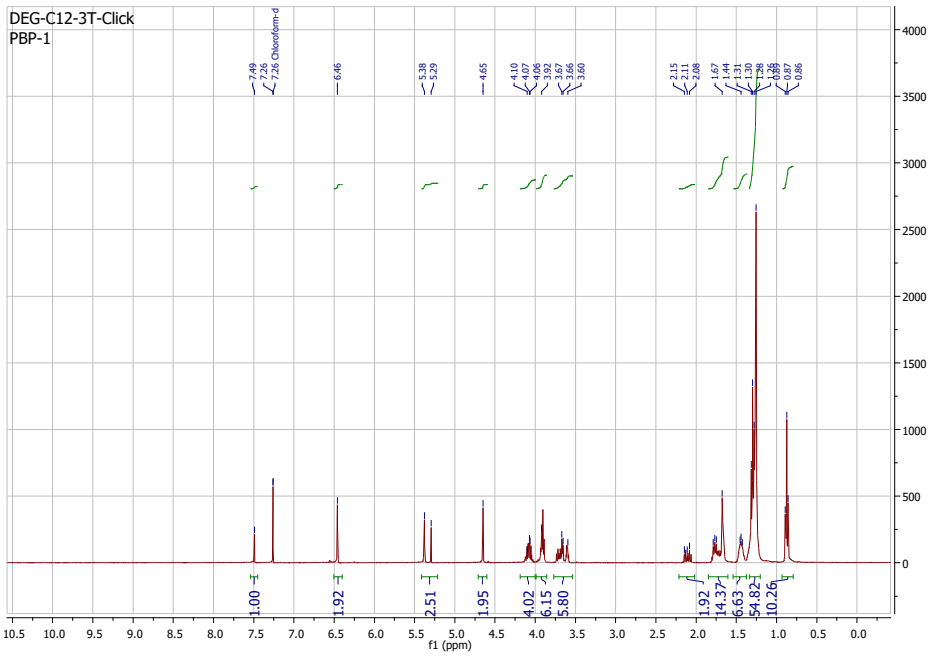


^{13}C NMR spectrum

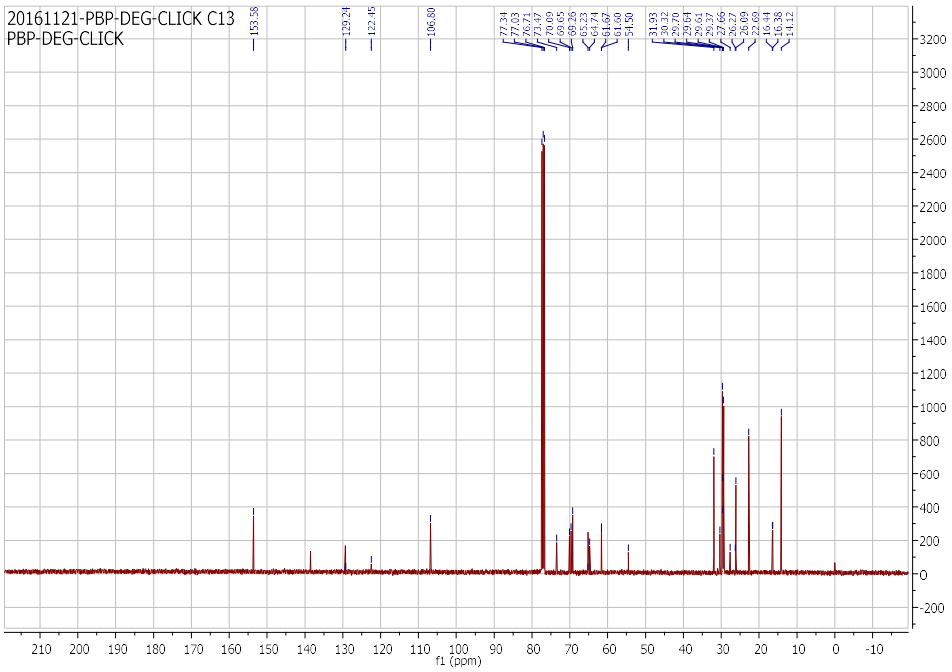
Compound 11a



11a

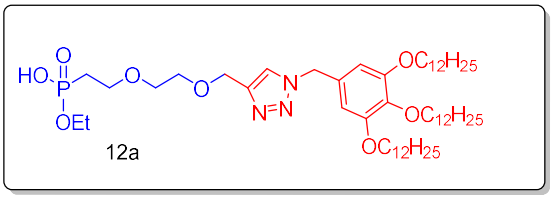


^{1}H NMR spectrum

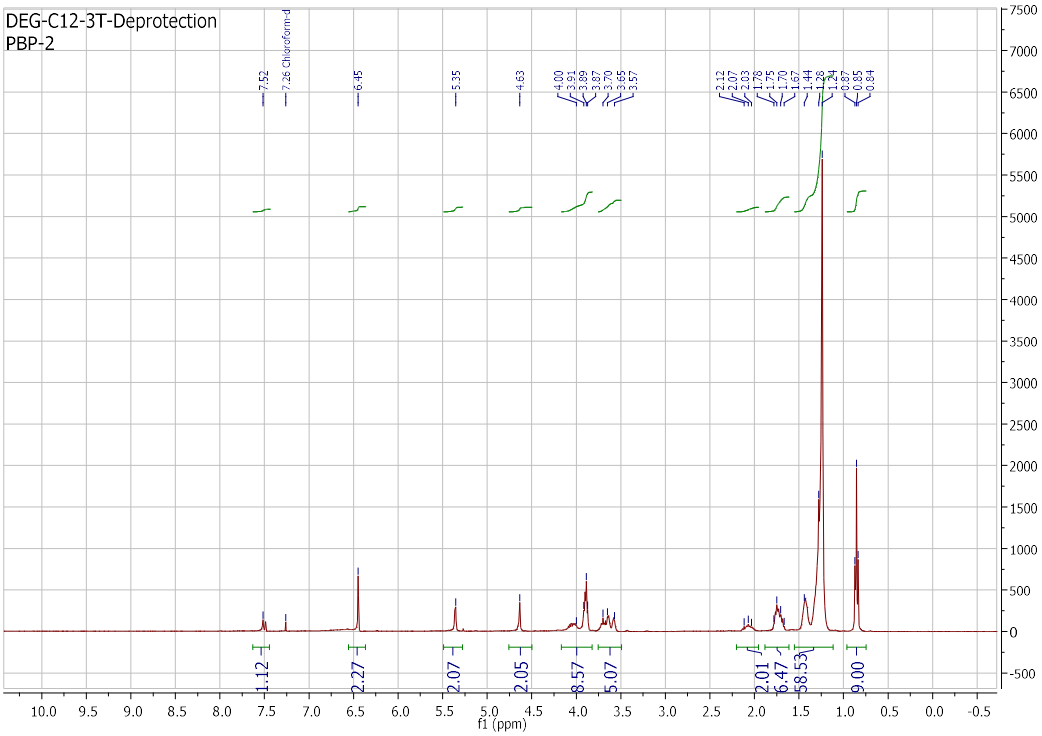


^{13}C NMR spectrum

Compound 12a

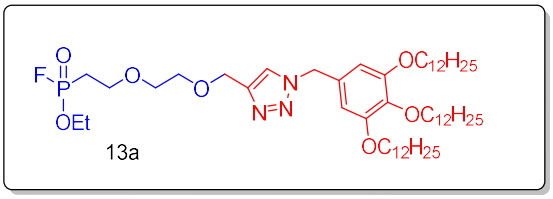


12a

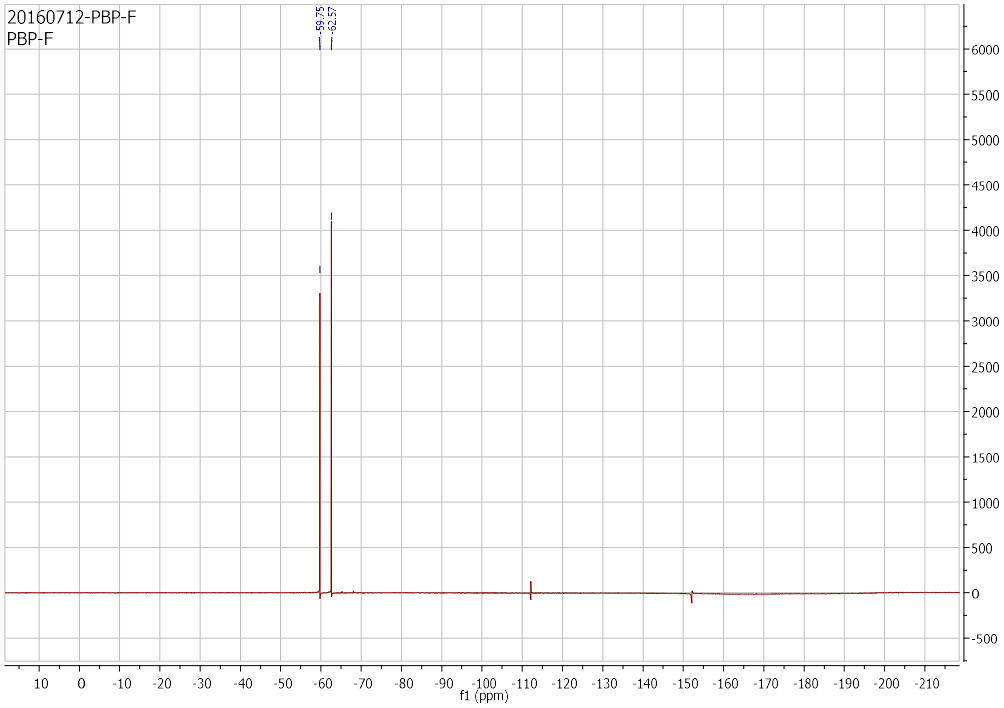


^{1}H NMR spectrum

Compound 13a

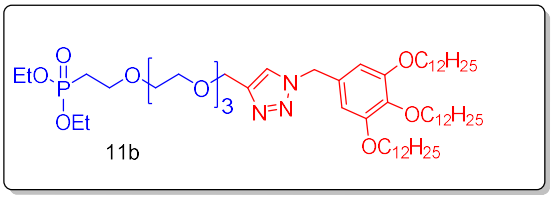


13a

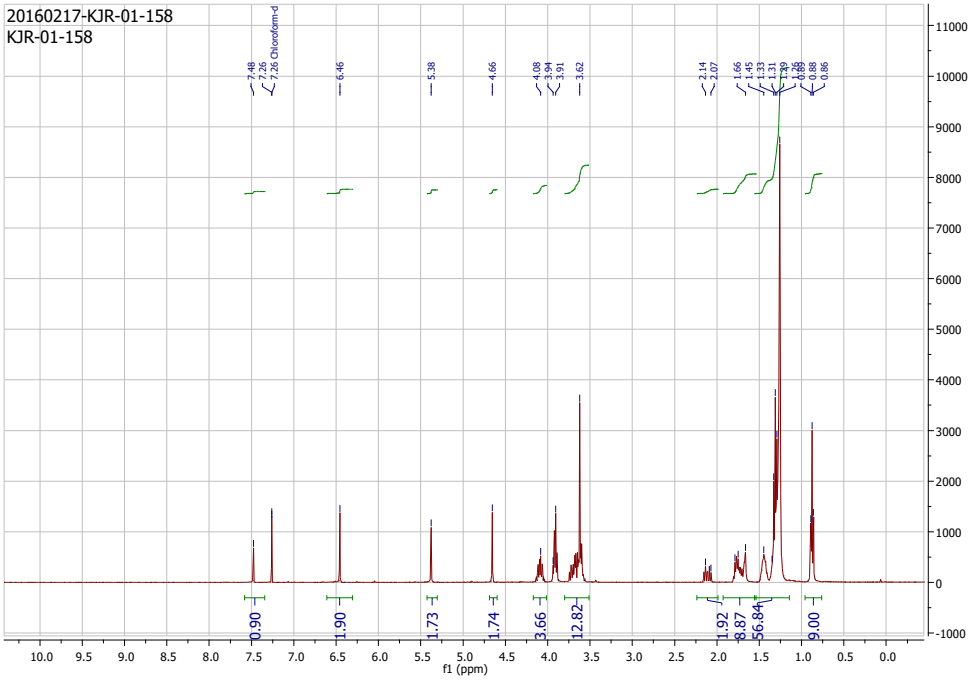


^{19}F NMR spectrum

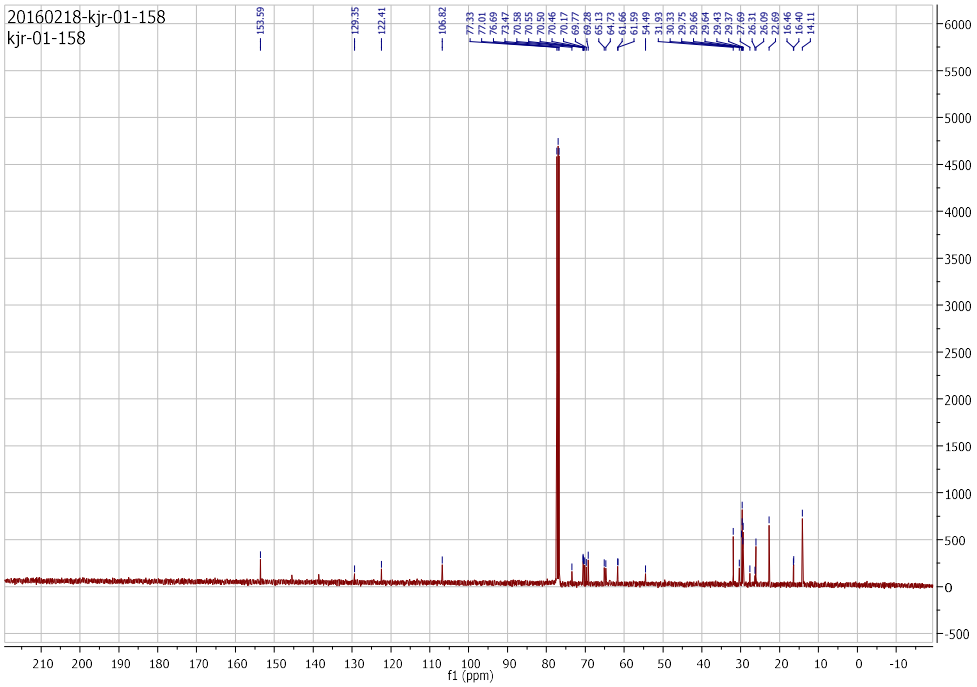
Compound 11b



11b

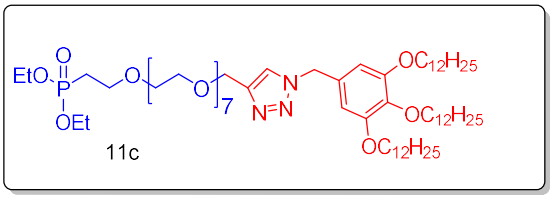


^{1}H NMR spectrum

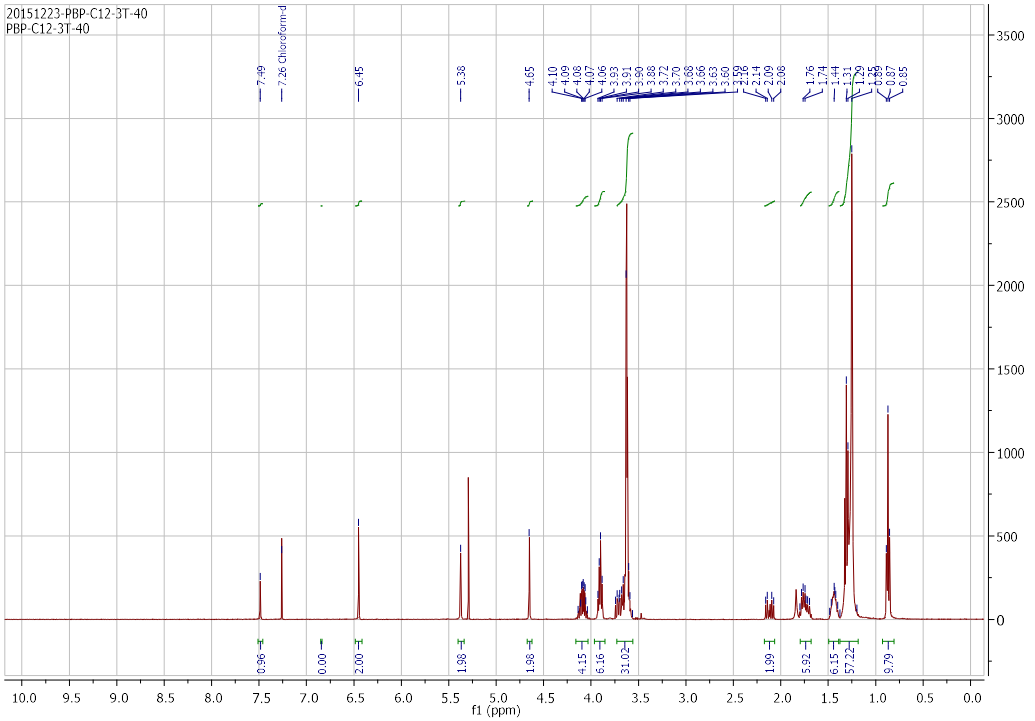


^{13}C NMR spectrum

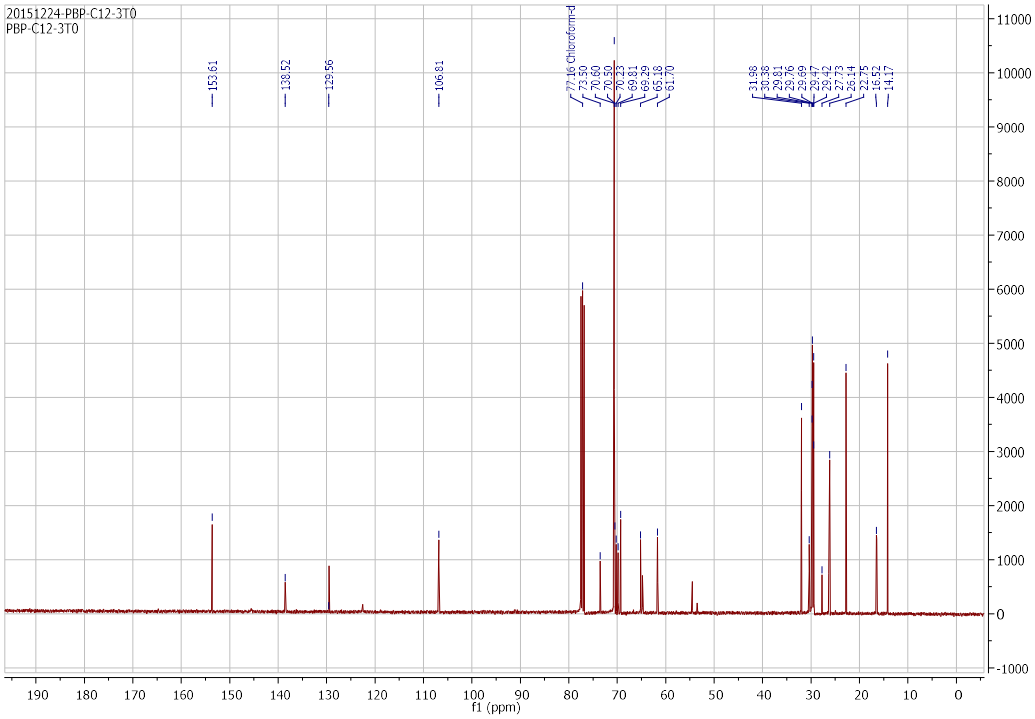
Compound 11c



11c

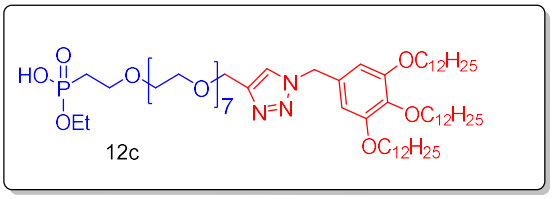


^{1}H NMR spectrum

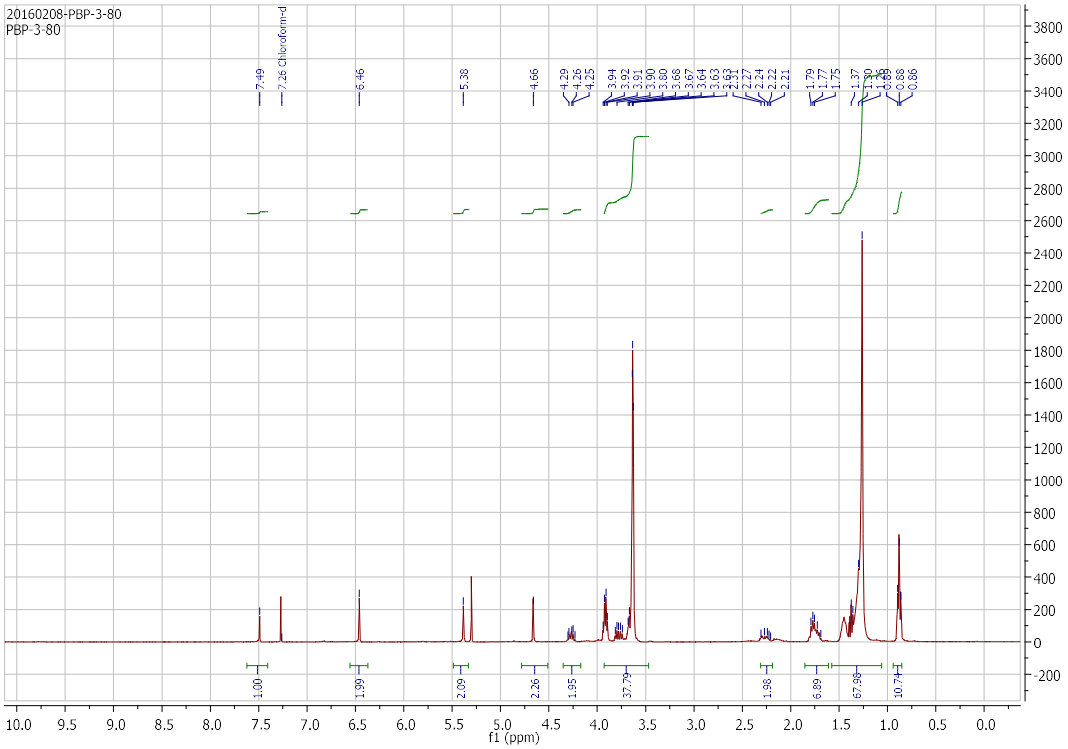


^{13}C NMR spectrum

Compound 12c

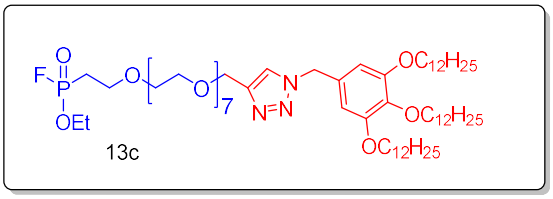


12c

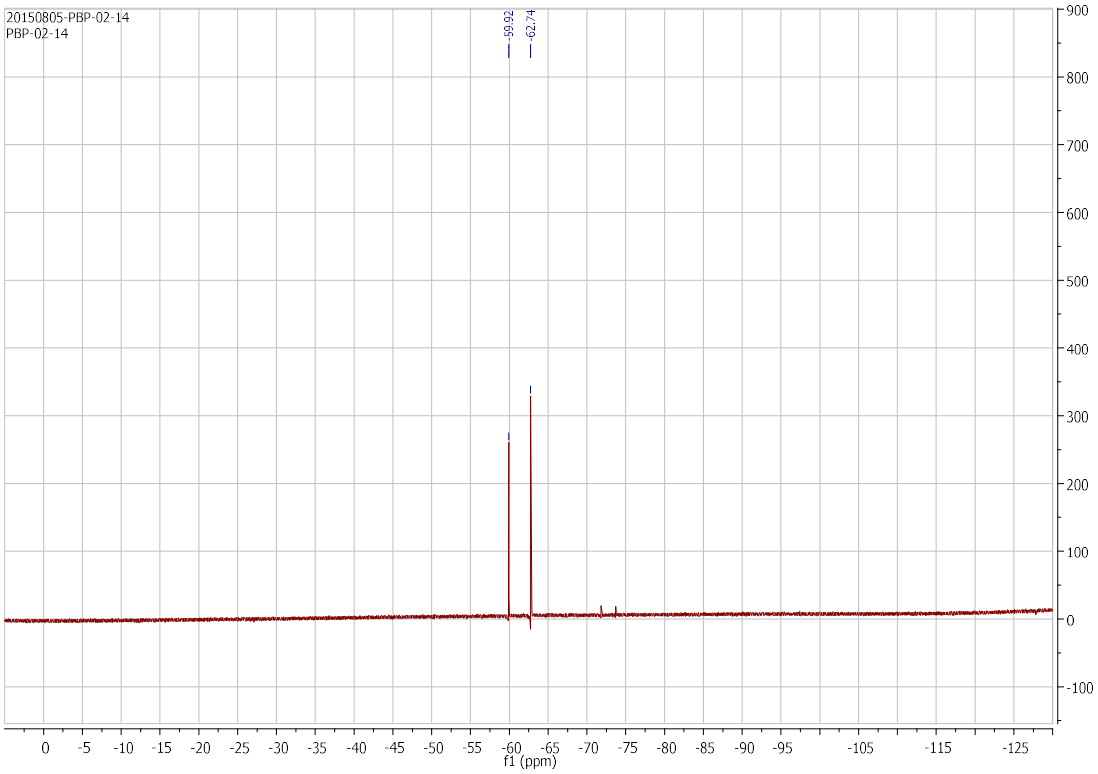


^{1}H NMR spectrum

Compound 13c

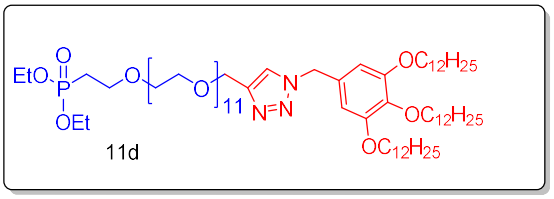


13c

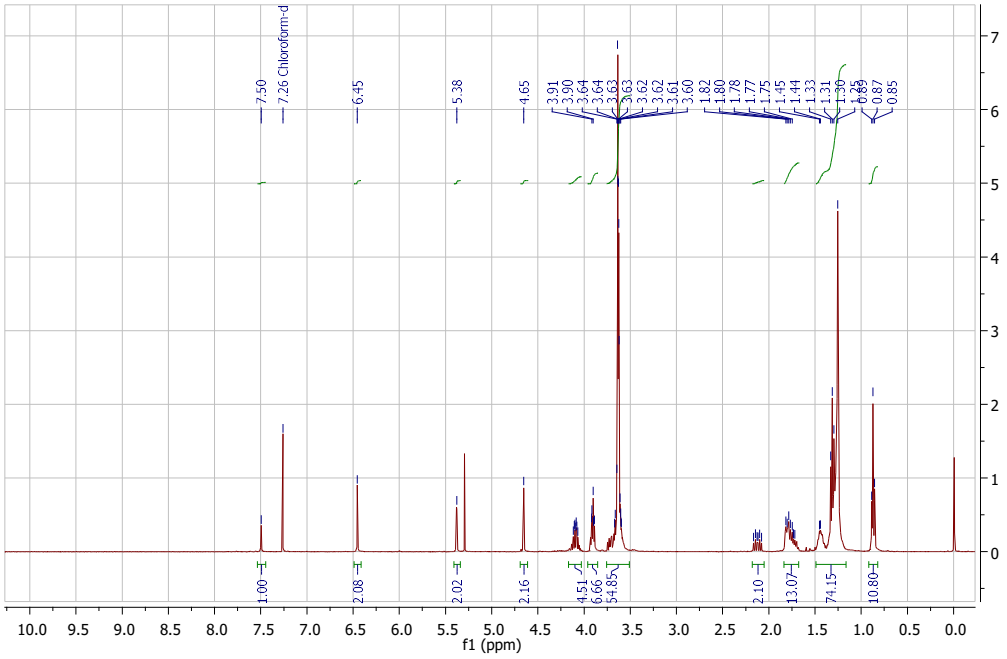


^{19}F NMR spectrum

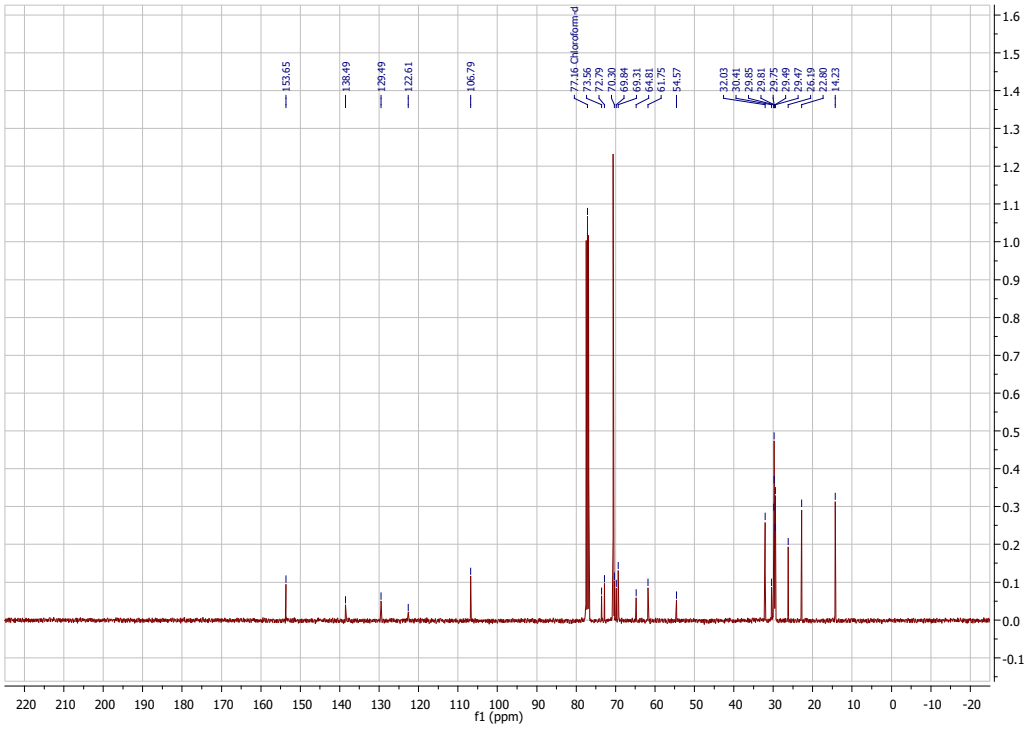
Compound 11d



11d

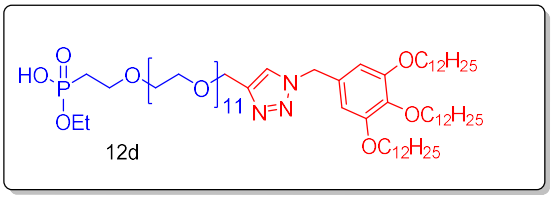


^{1}H NMR spectrum

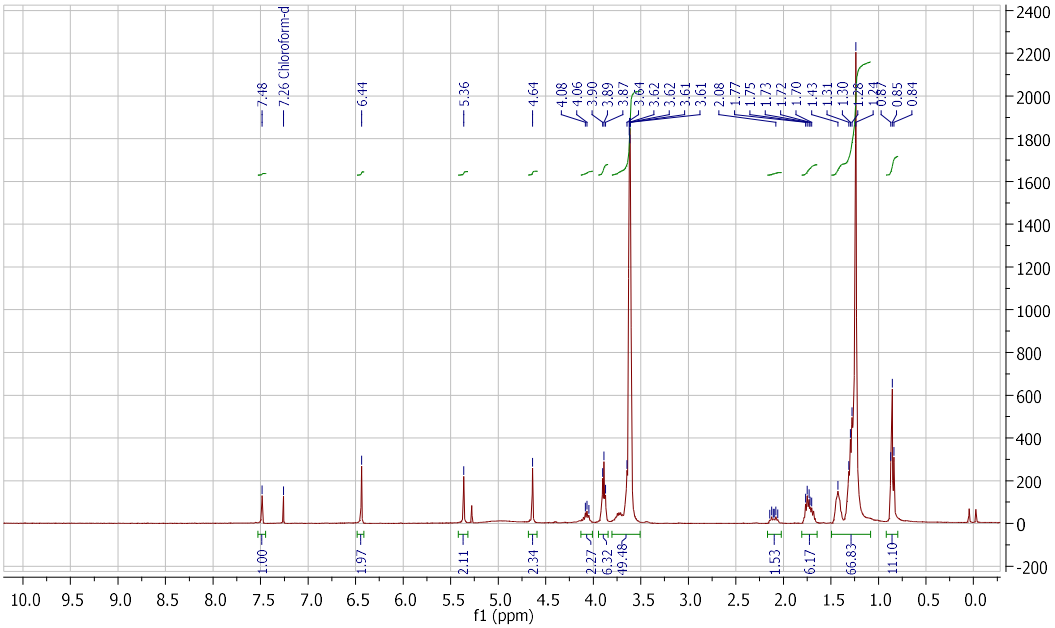


^{13}C NMR spectrum

Compound 12d

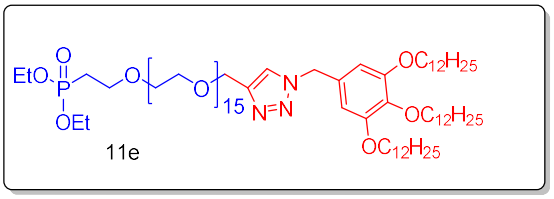


12d

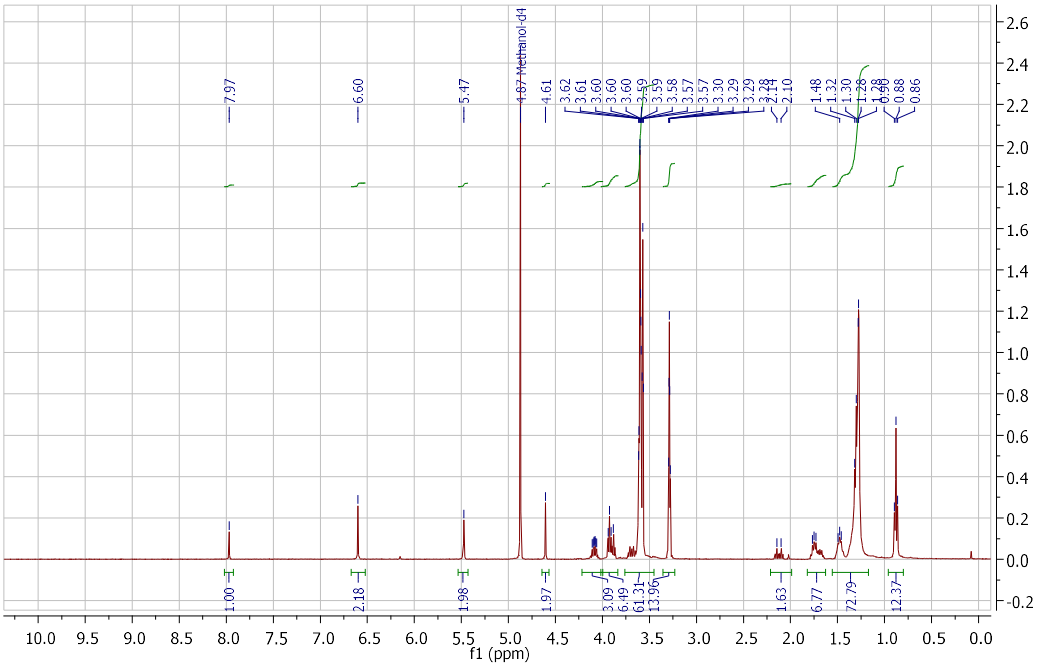


^{1}H NMR spectrum

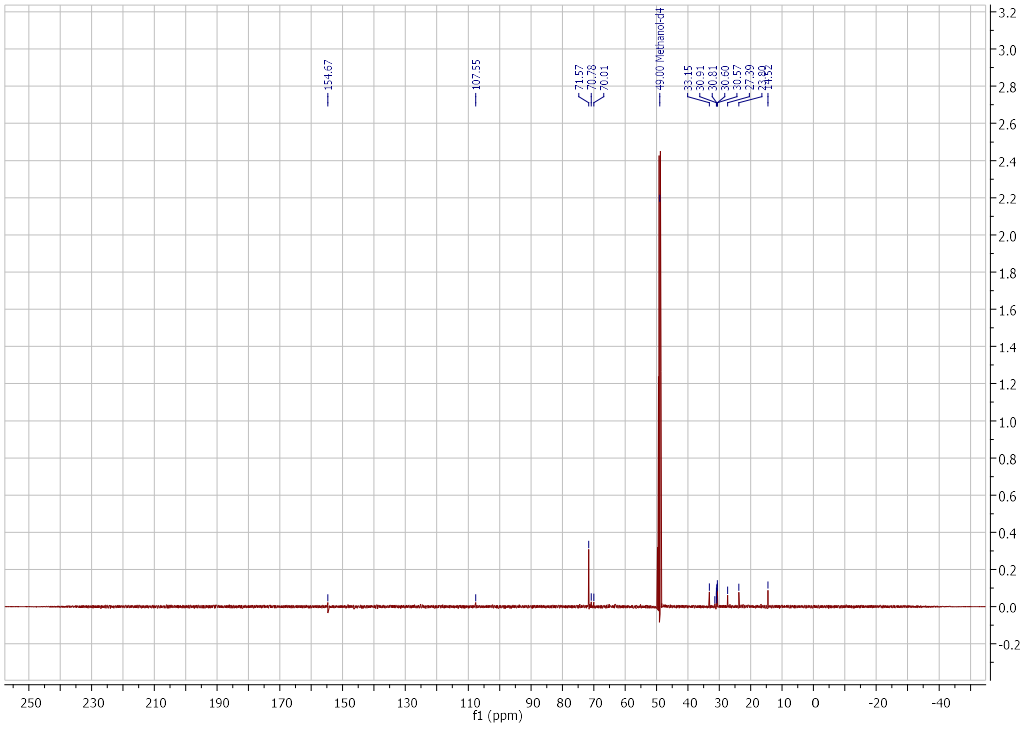
Compound 11e



11e

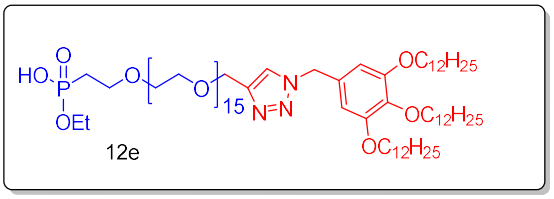


^{1}H NMR spectrum

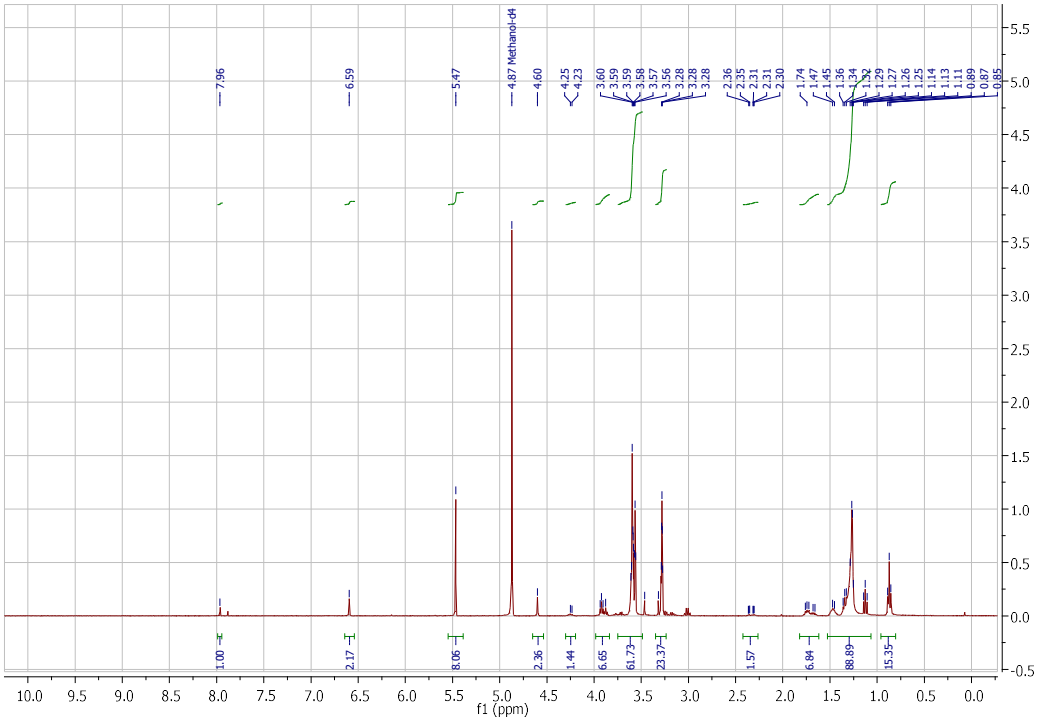


^{13}C NMR spectrum

Compound 12e

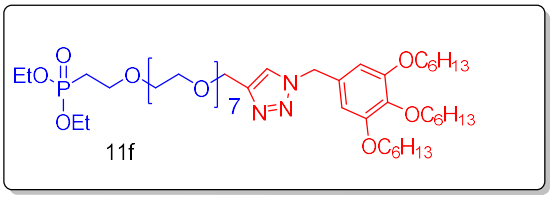


12e

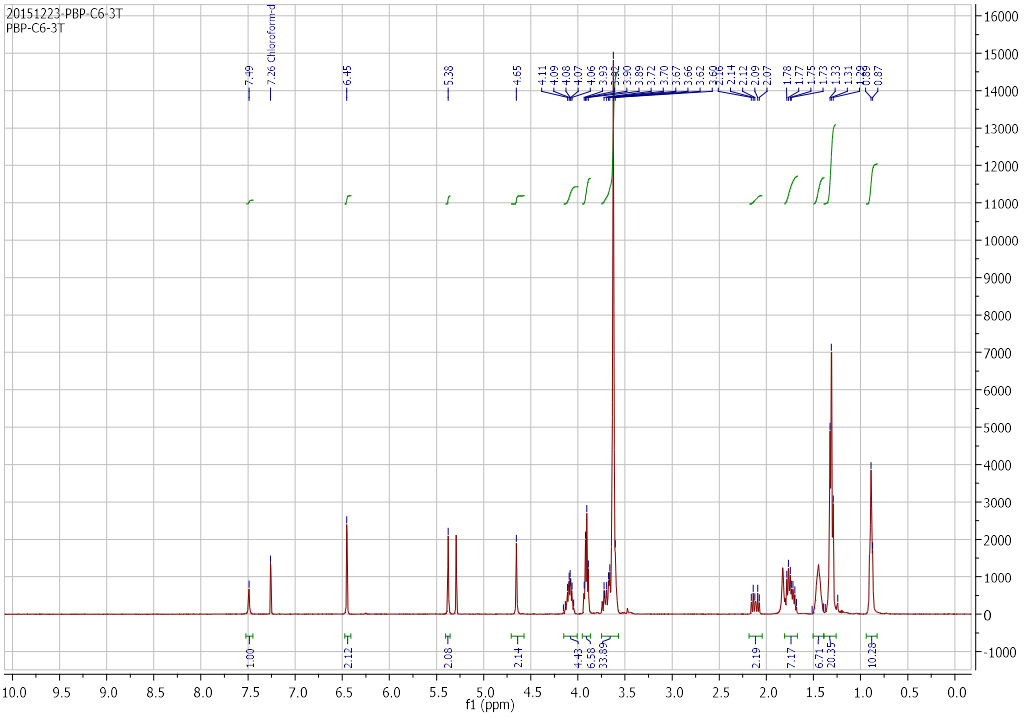


^{1}H NMR spectrum

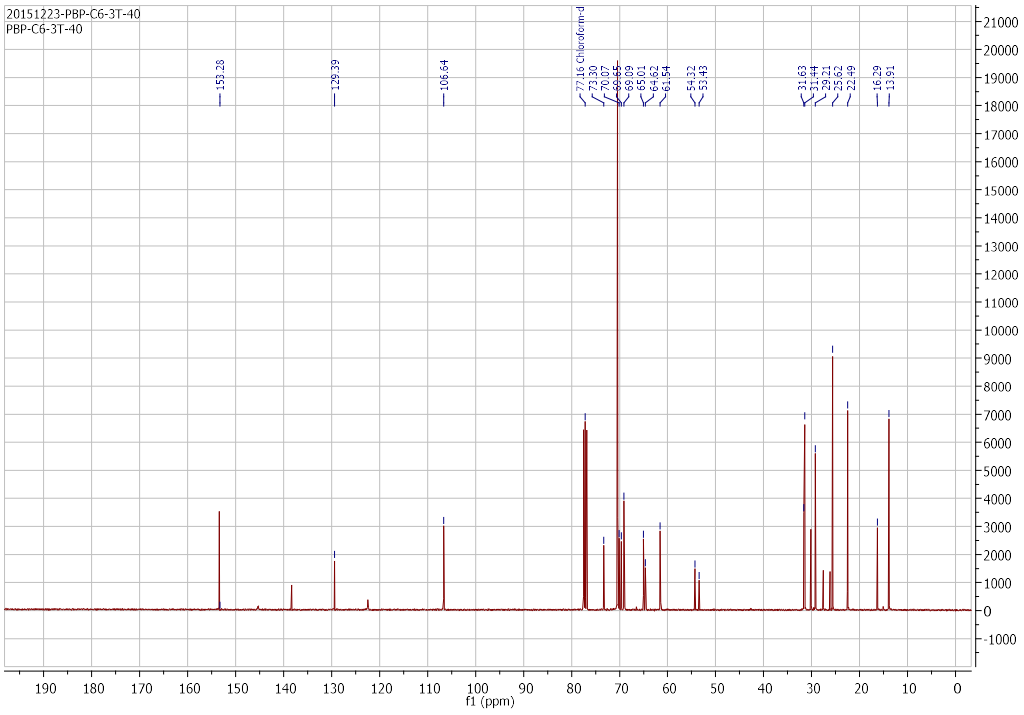
Compound 11f



11f

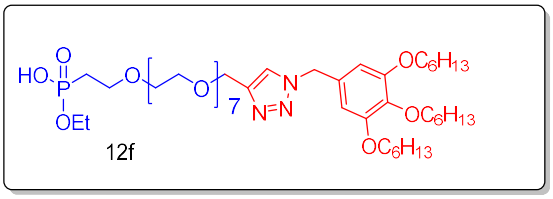


^{1}H NMR spectrum

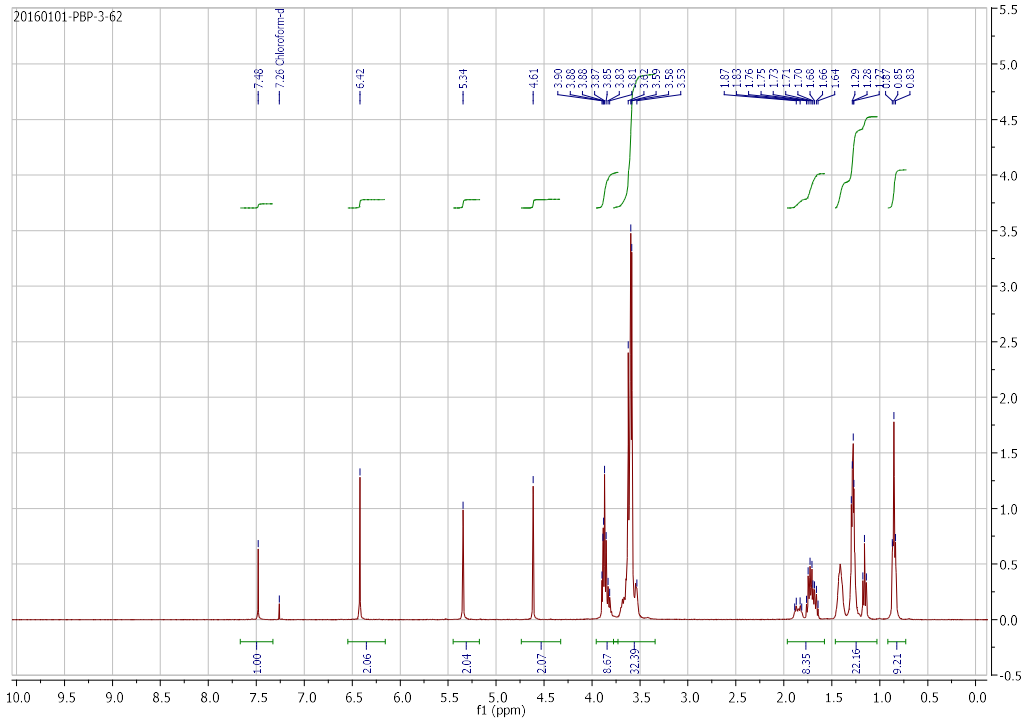


^{13}C NMR spectrum

Compound 12f

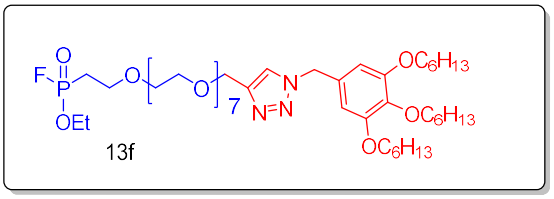


12f

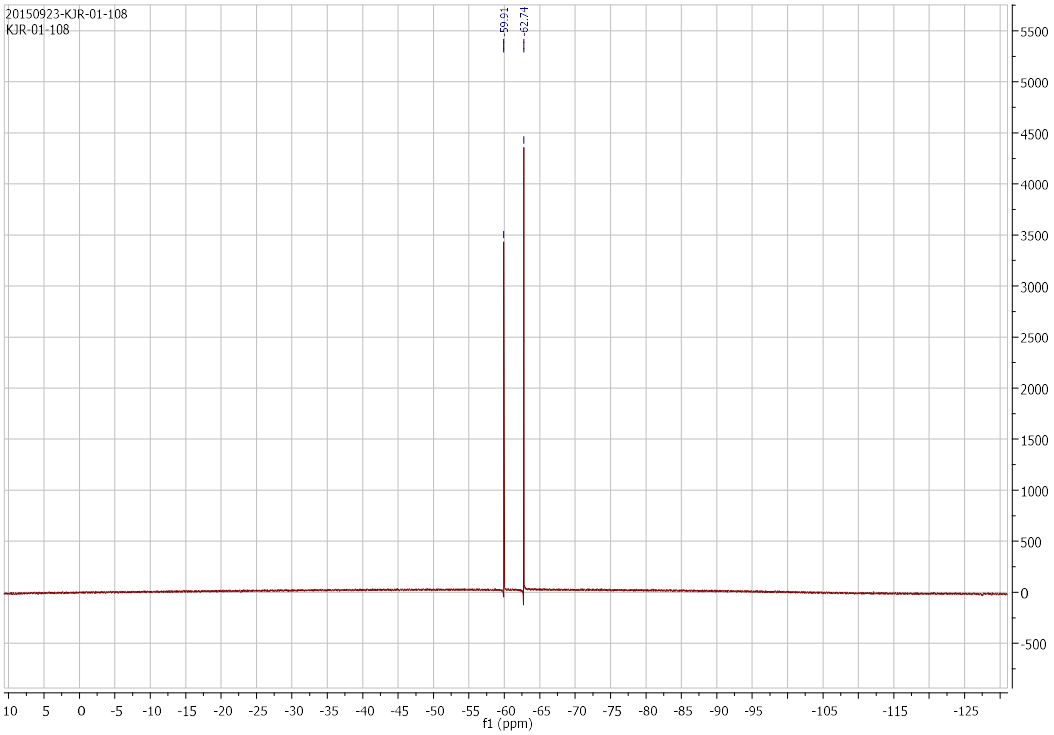


^{1}H NMR spectrum

Compound 13f

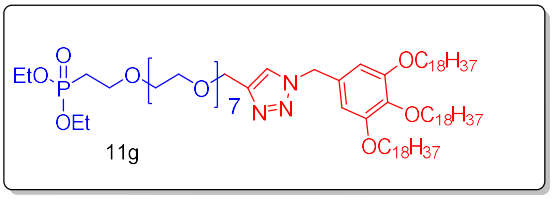


13f

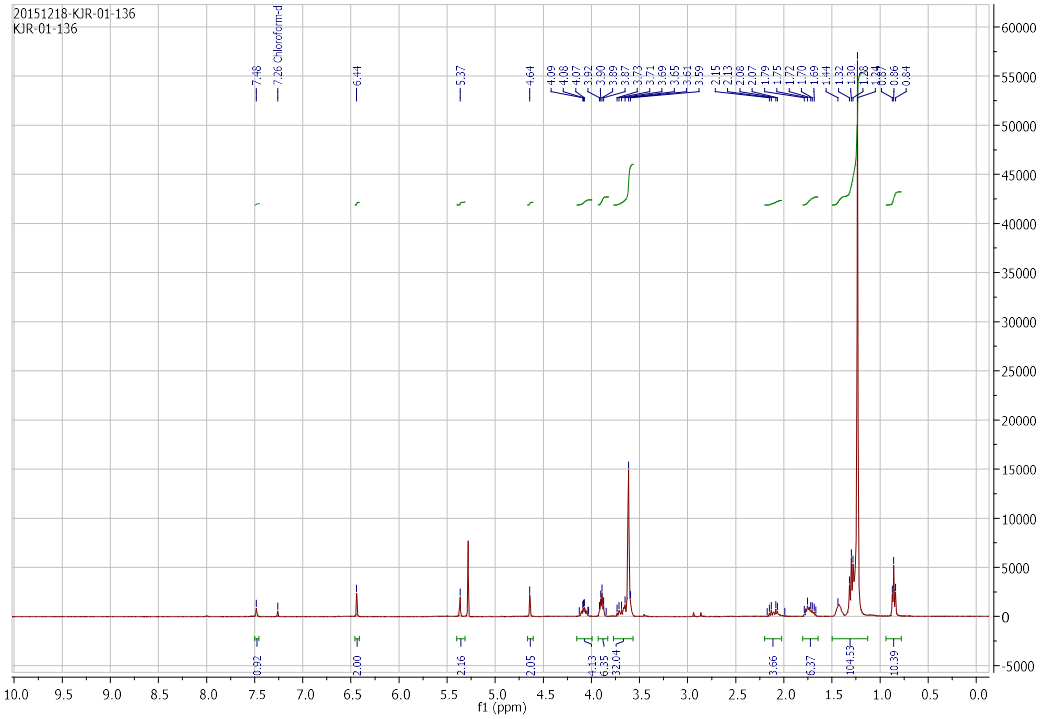


^{19}F NMR spectrum

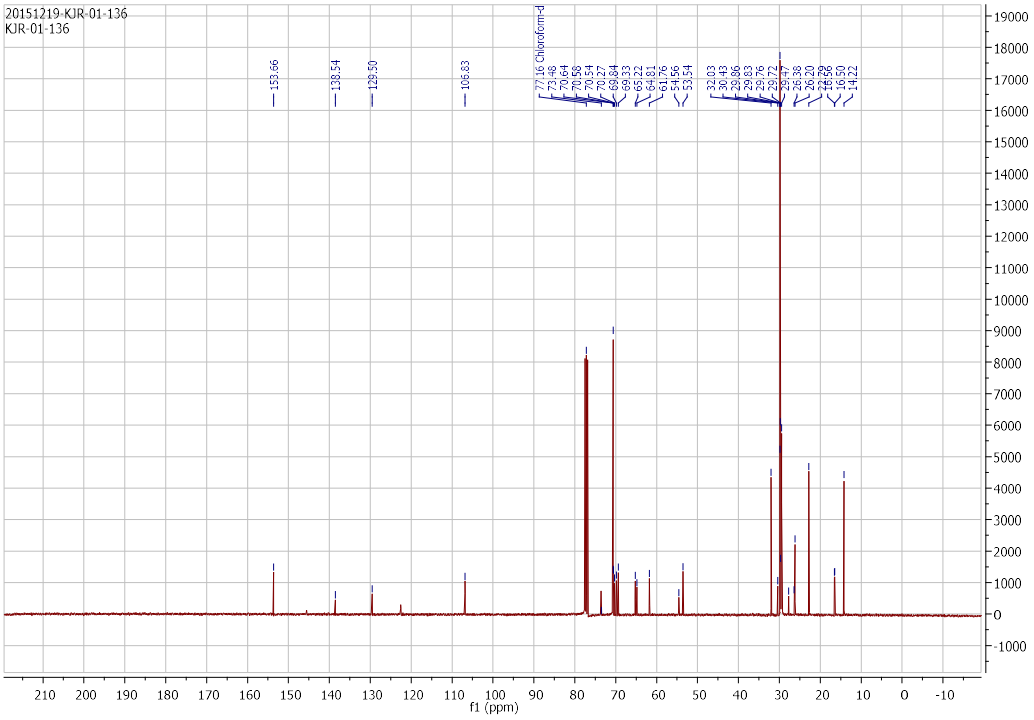
Compound 11g



11g

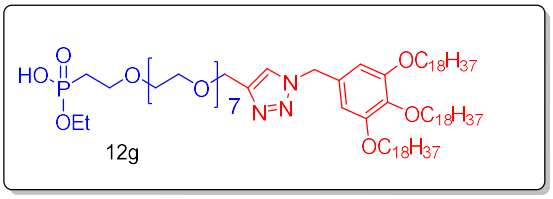


^{1}H NMR spectrum

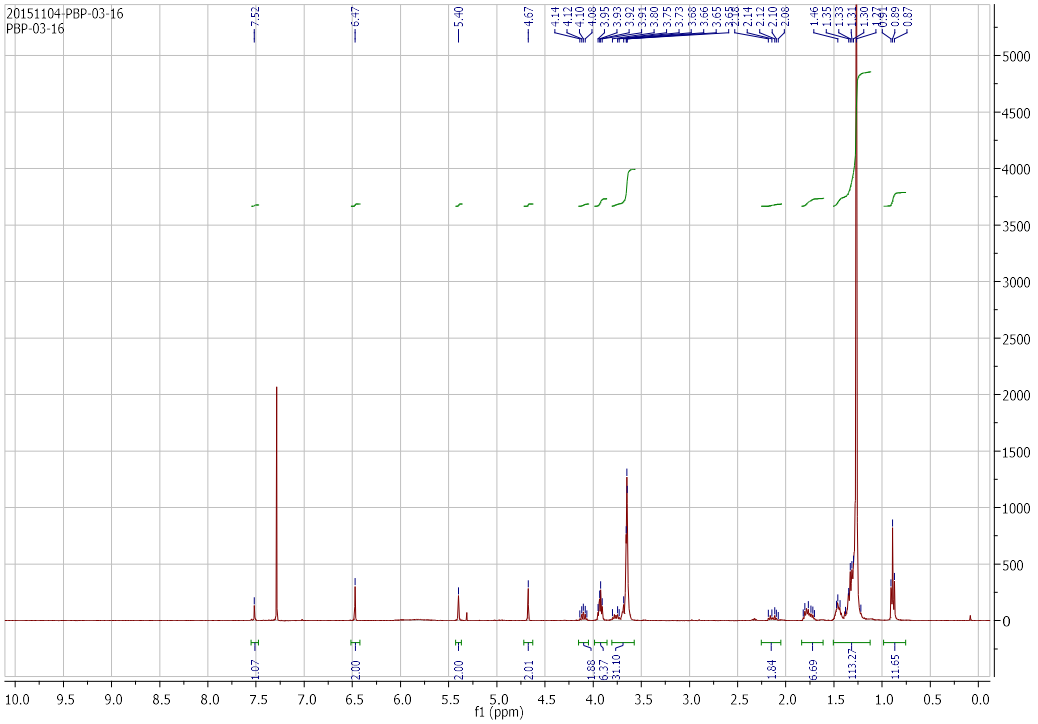


^{13}C NMR spectrum

Compound 12g

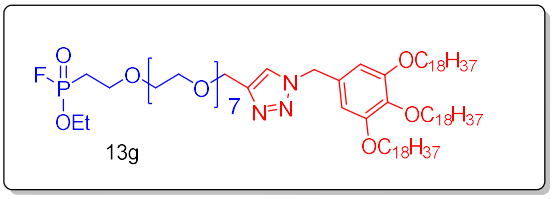


12g

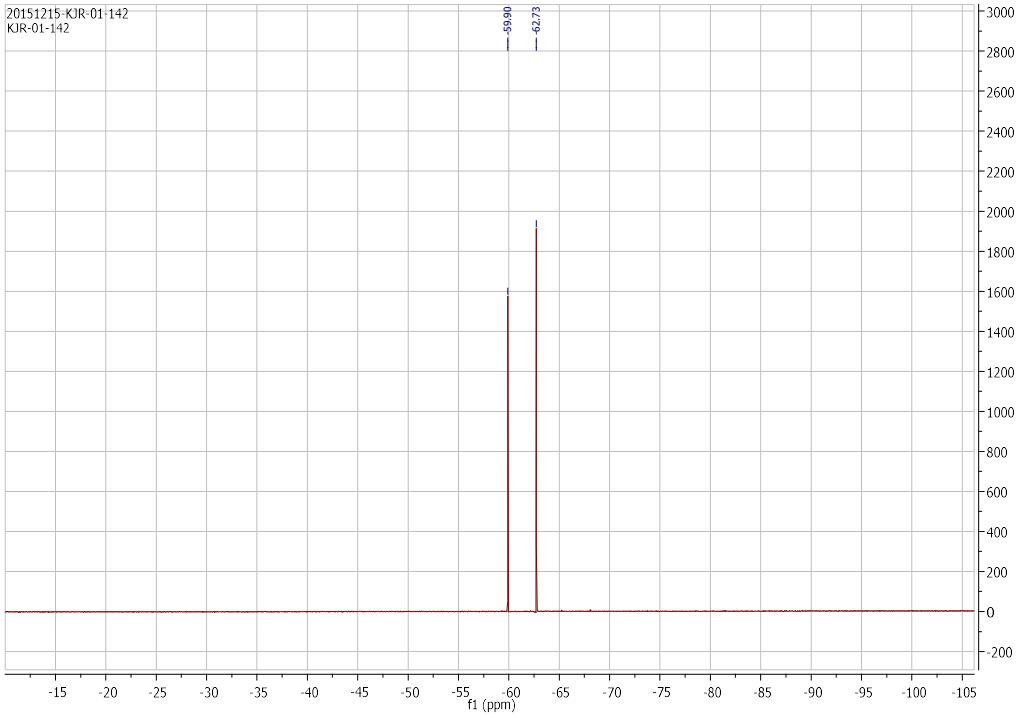


^{1}H NMR spectrum

Compound 13g

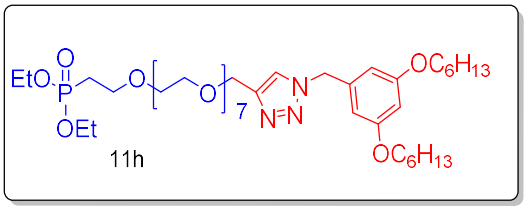


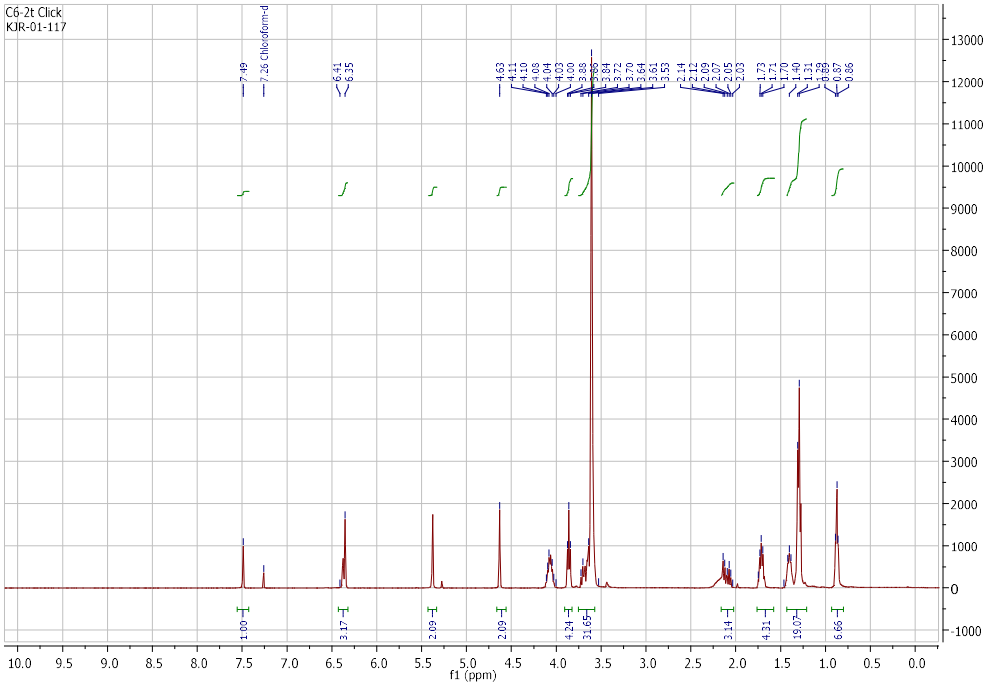
13g



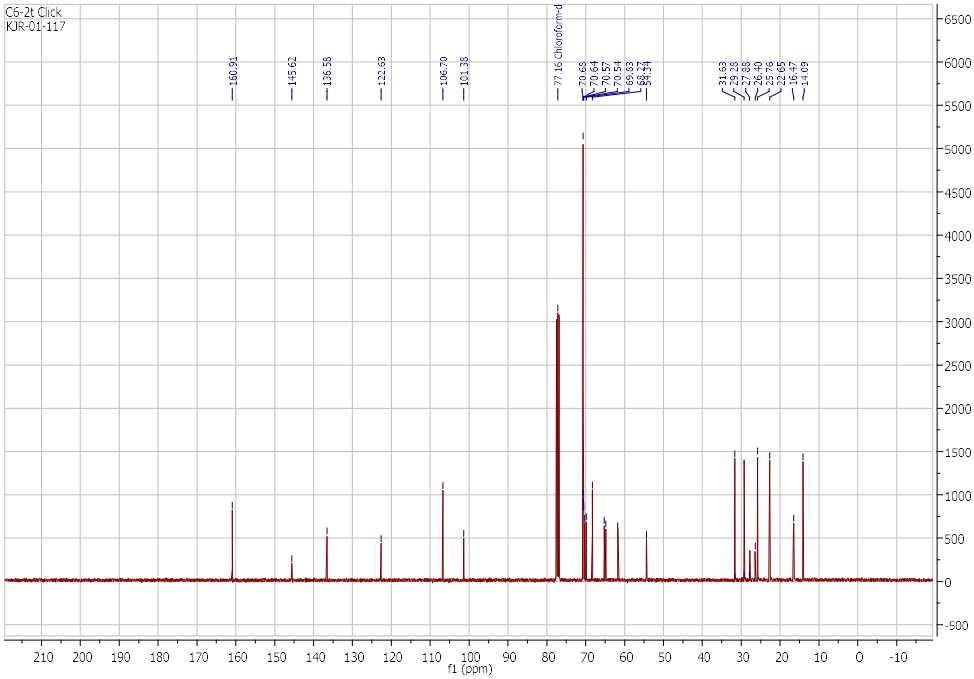
^{19}F NMR spectrum

Compound 11h



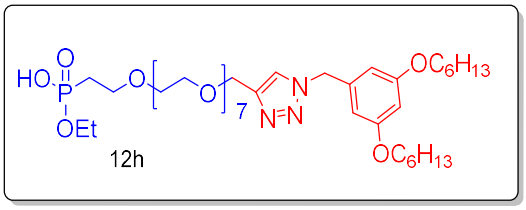


^{1}H NMR spectrum

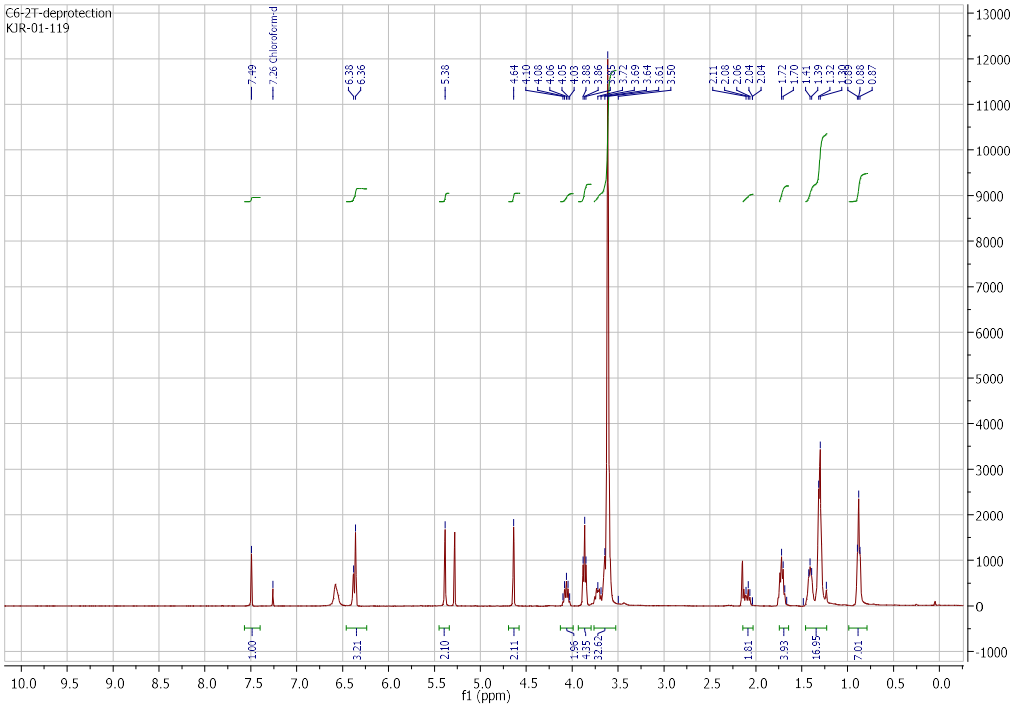


^{13}C NMR spectrum

Compound 12h

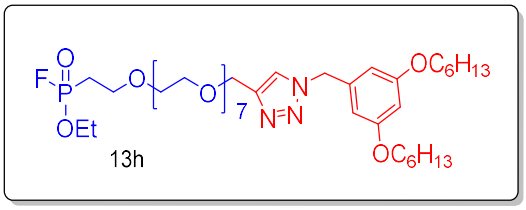


12h

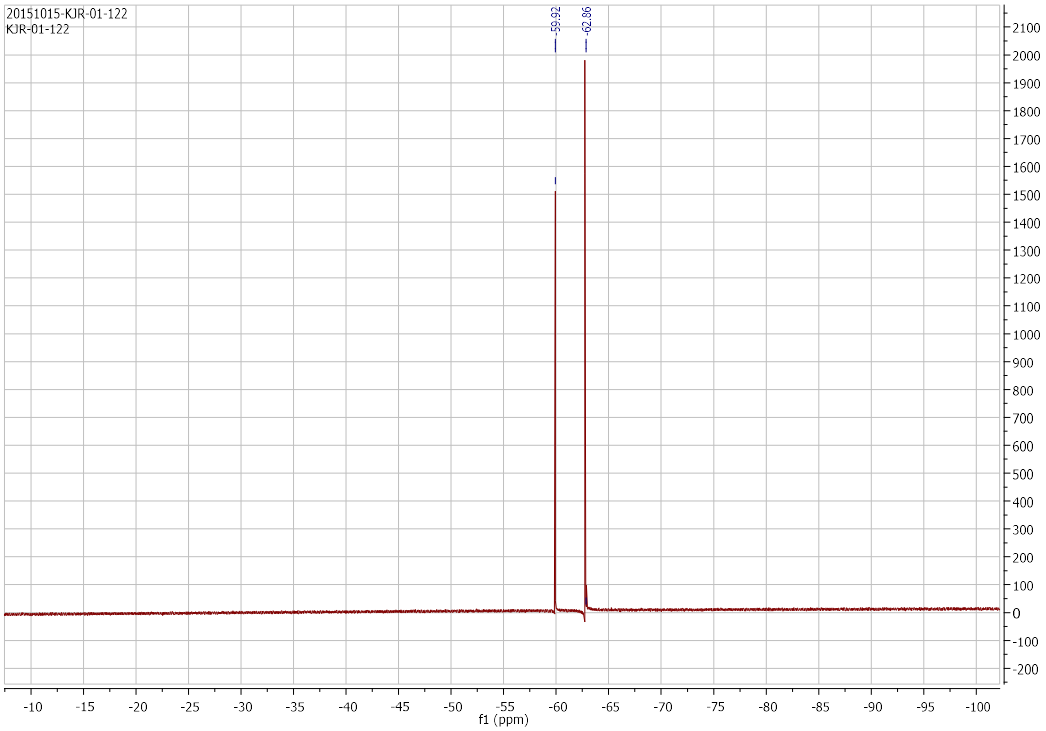


^{1}H NMR spectrum

Compound 13h

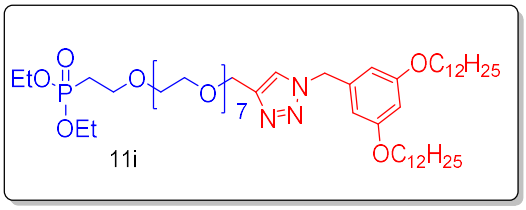


13h

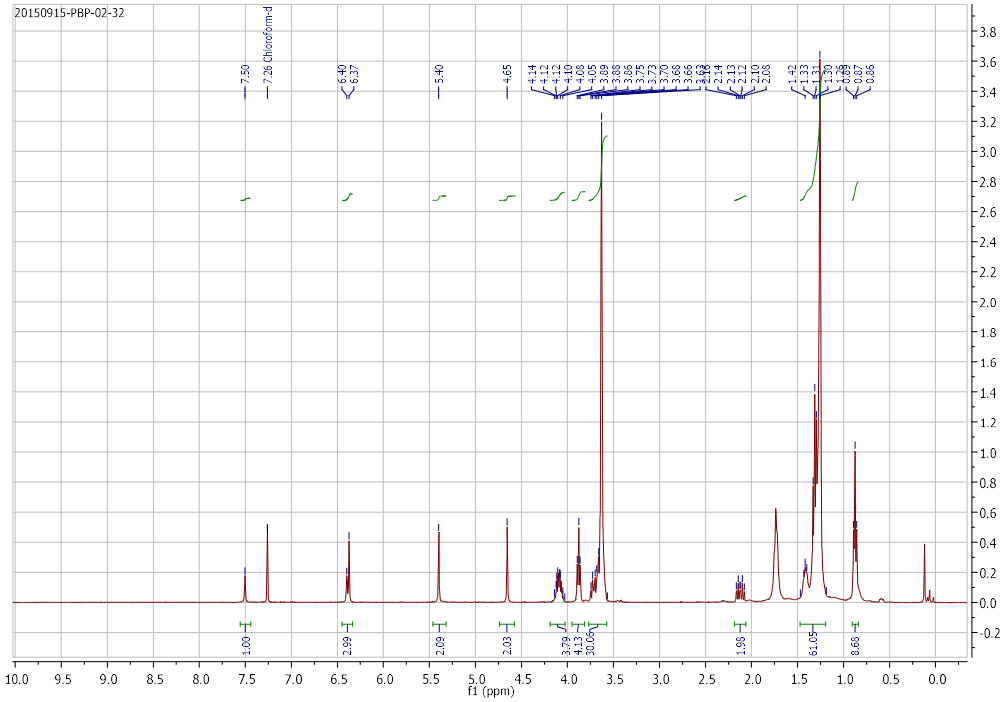


^{19}F NMR spectrum

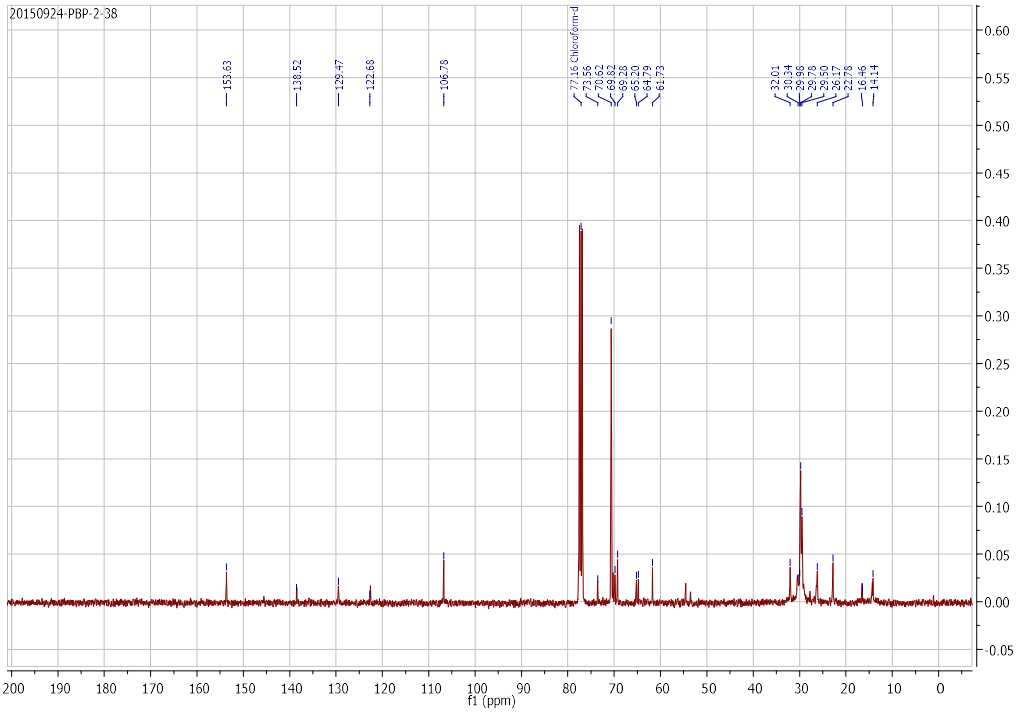
Compound 11i



11i

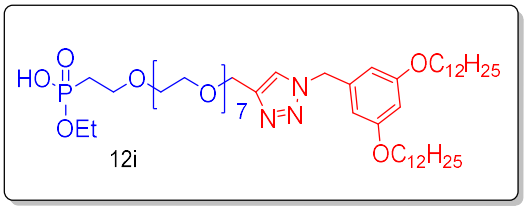


^{1}H NMR spectrum

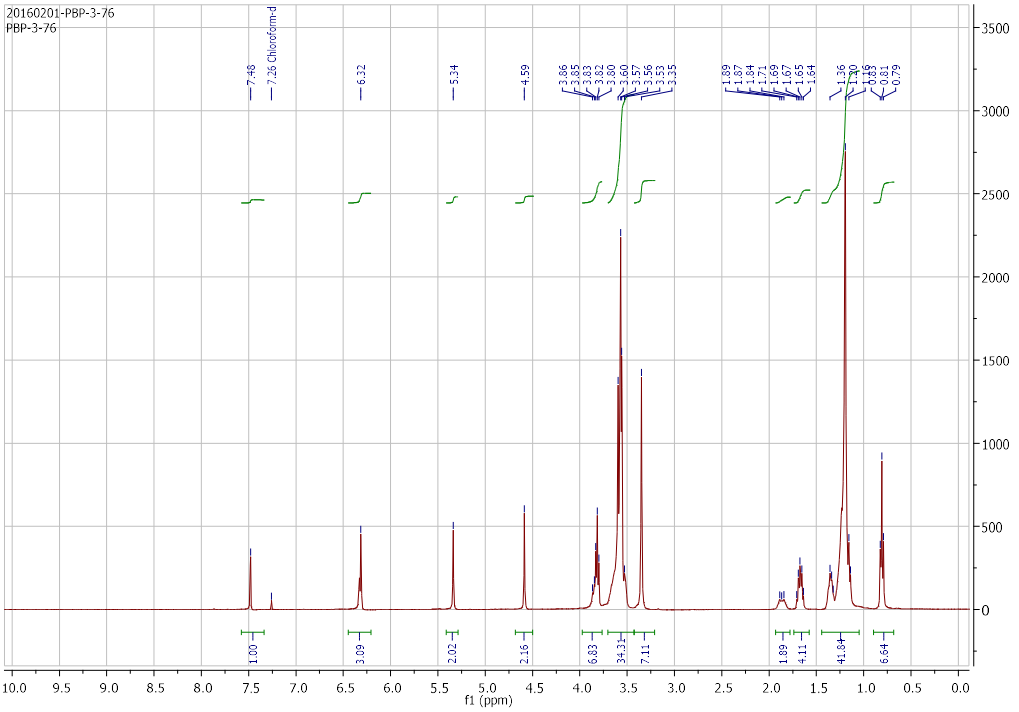


^{13}C NMR spectrum

Compound 12i

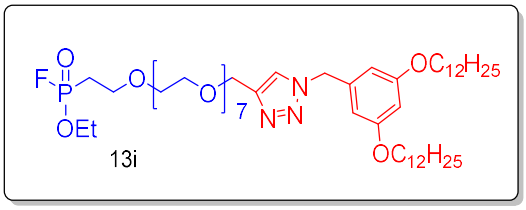


12i

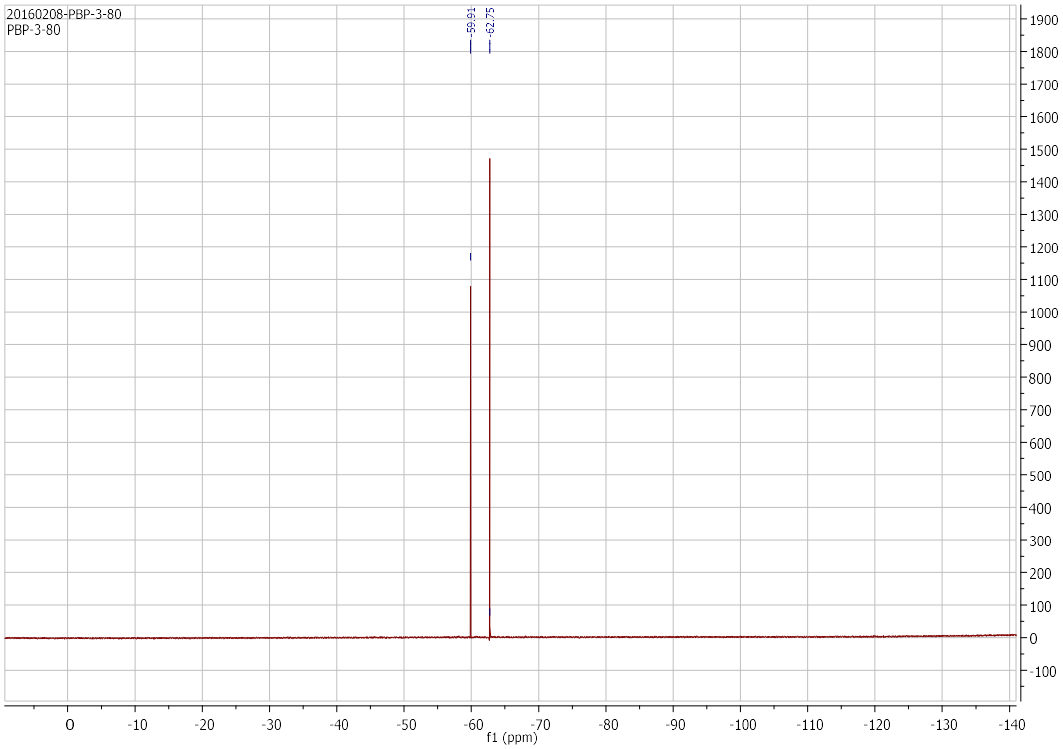


^{1}H NMR spectrum

Compound 13i

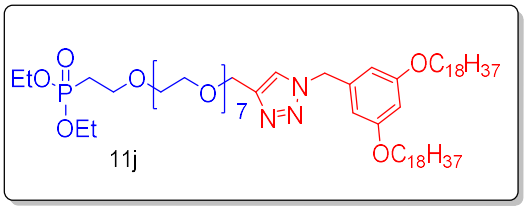


13i

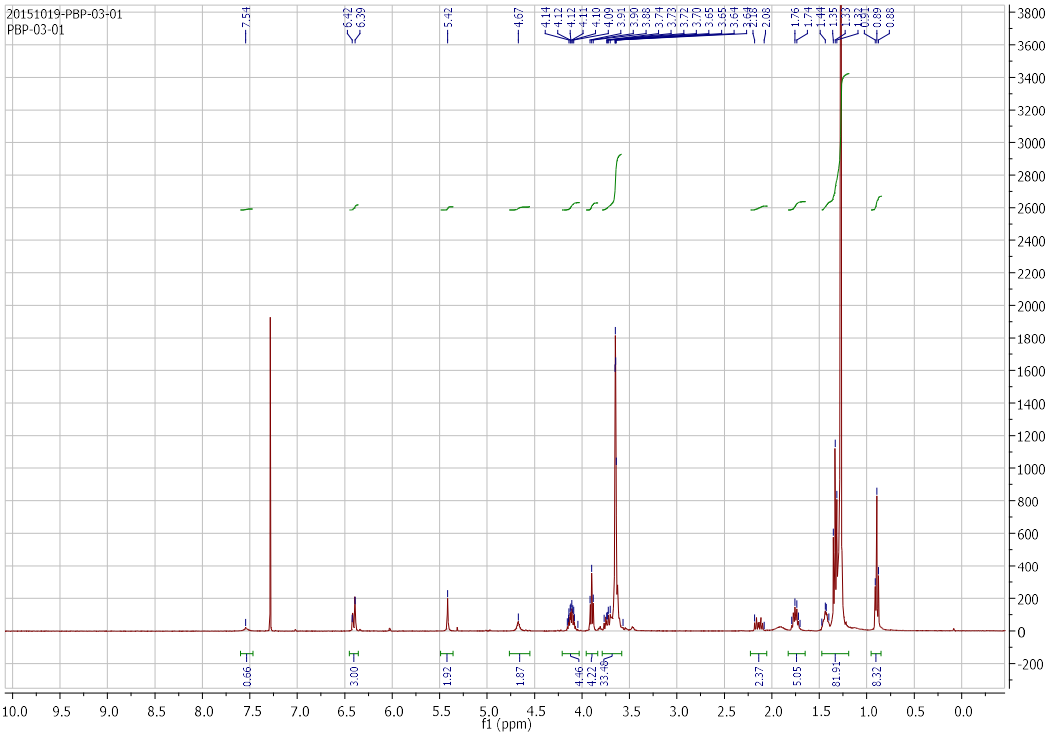


^{19}F NMR spectrum

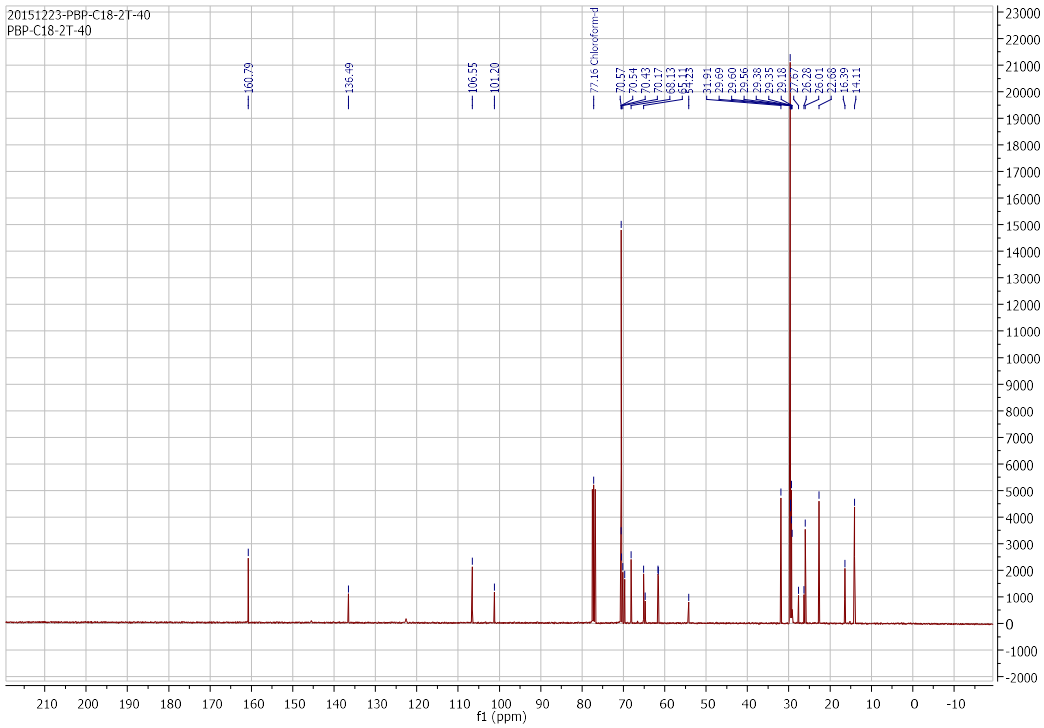
Compound 11j



11j

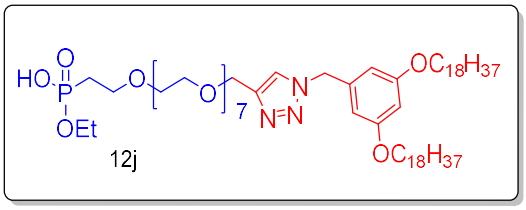


^{1}H NMR spectrum

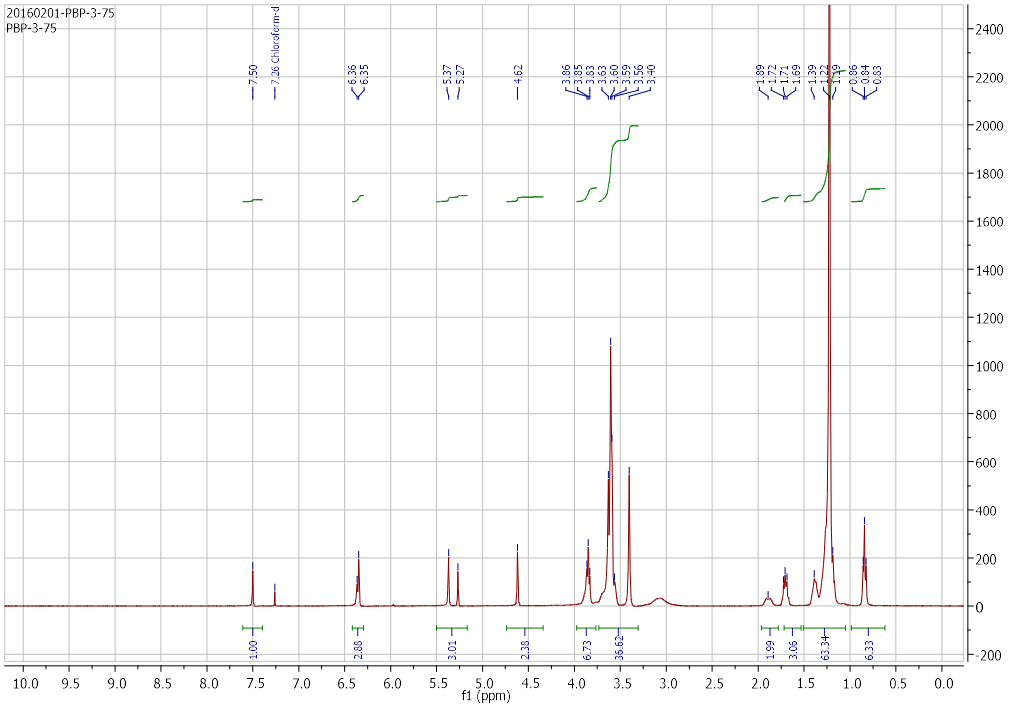


^{13}C NMR spectrum

Compound 12j

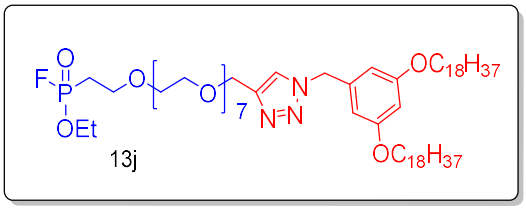


12j

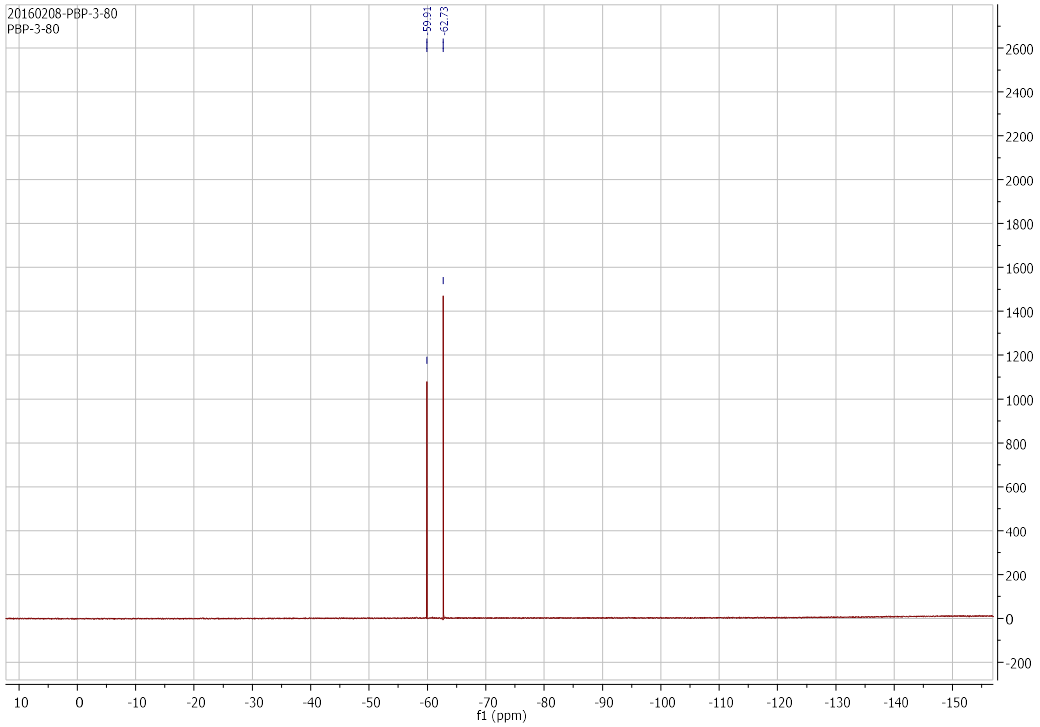


^{1}H NMR spectrum

Compound 13j

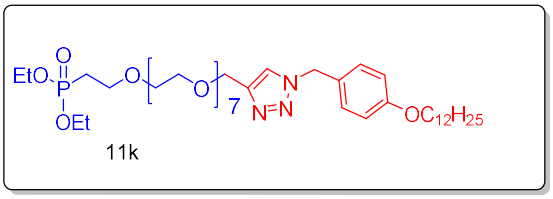


13j

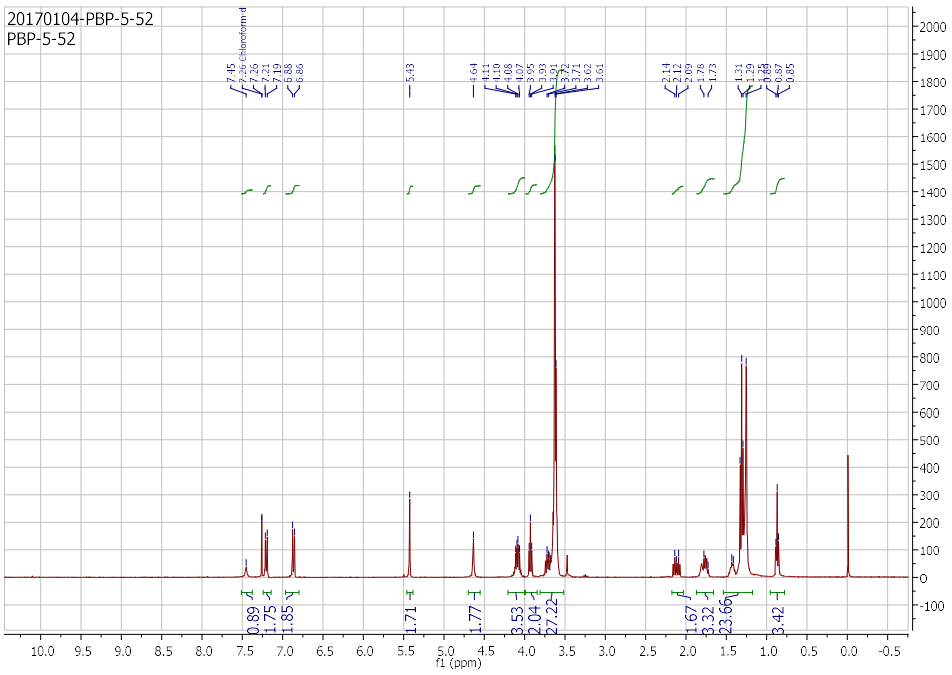


^{19}F NMR spectrum

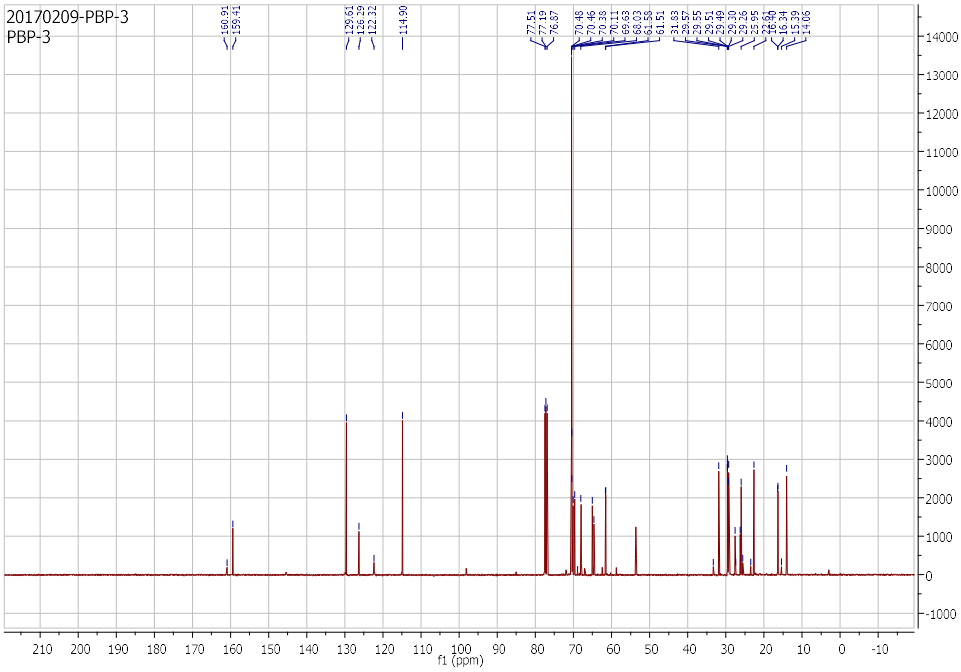
Compound 11k



11k

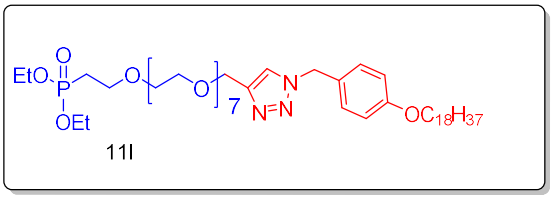


^{1}H NMR spectrum

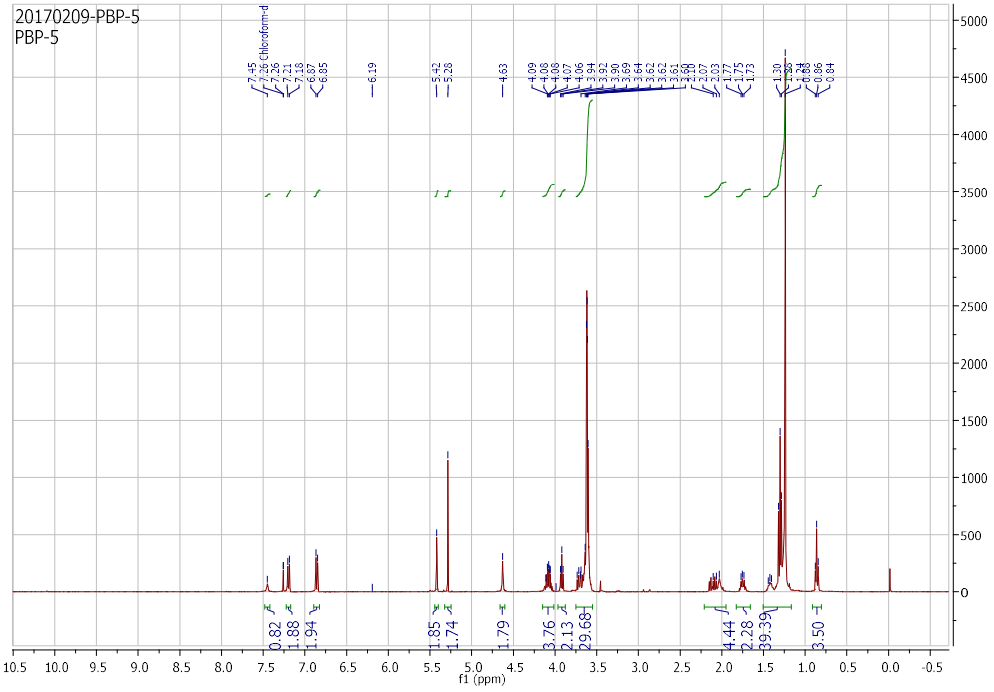


^{13}C NMR spectrum

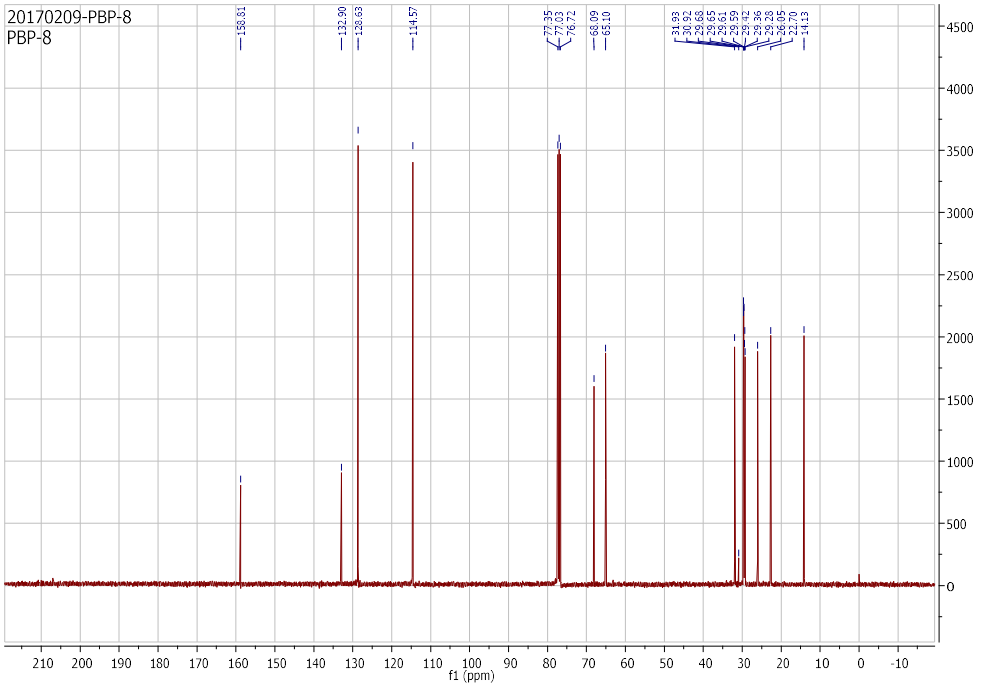
Compound 11l



11l



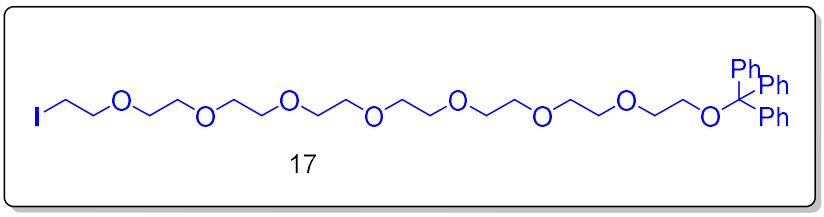
^{1}H NMR spectrum



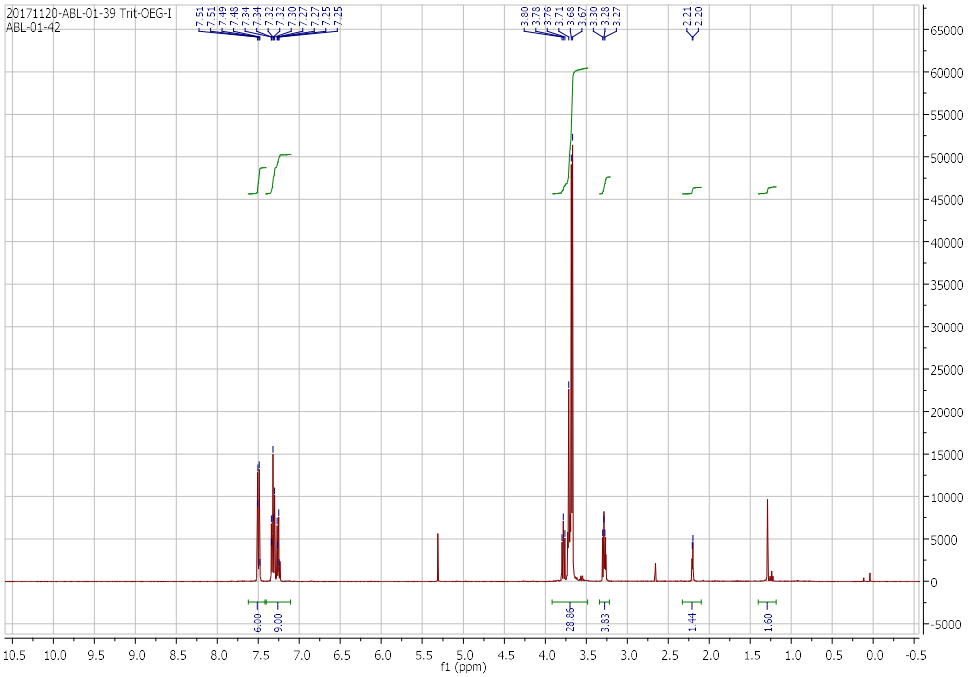
^{13}C NMR spectrum

Chapter 3 Spectral Data

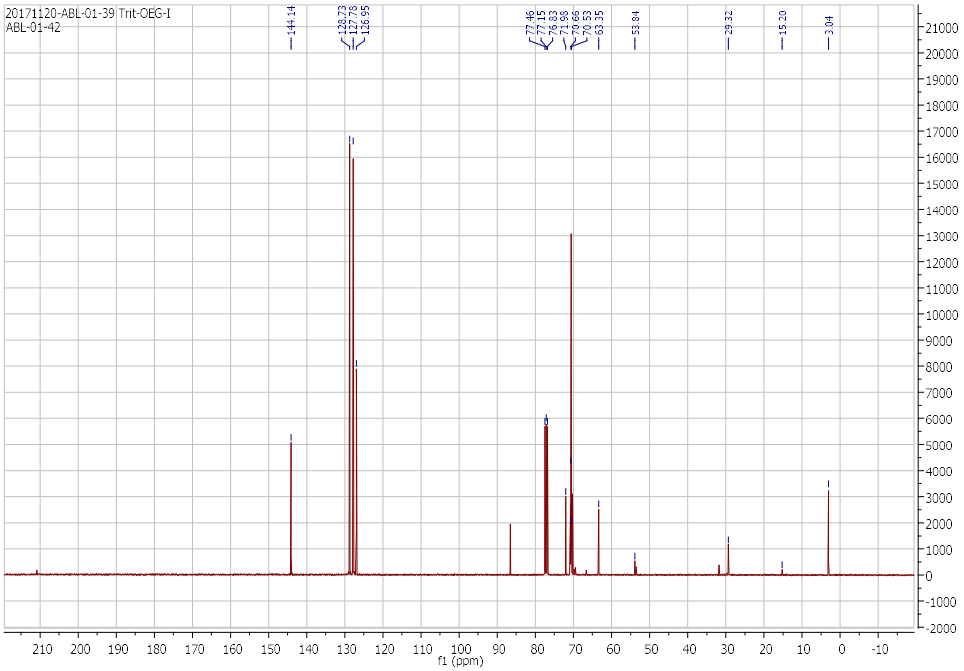
Compound 17



17

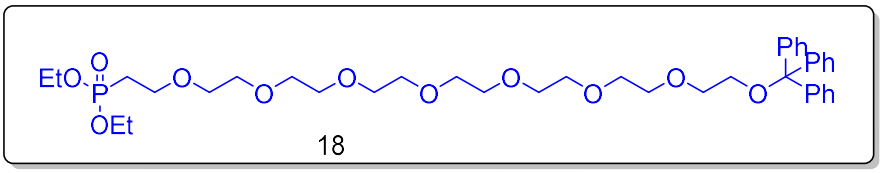


^{1}H NMR spectrum

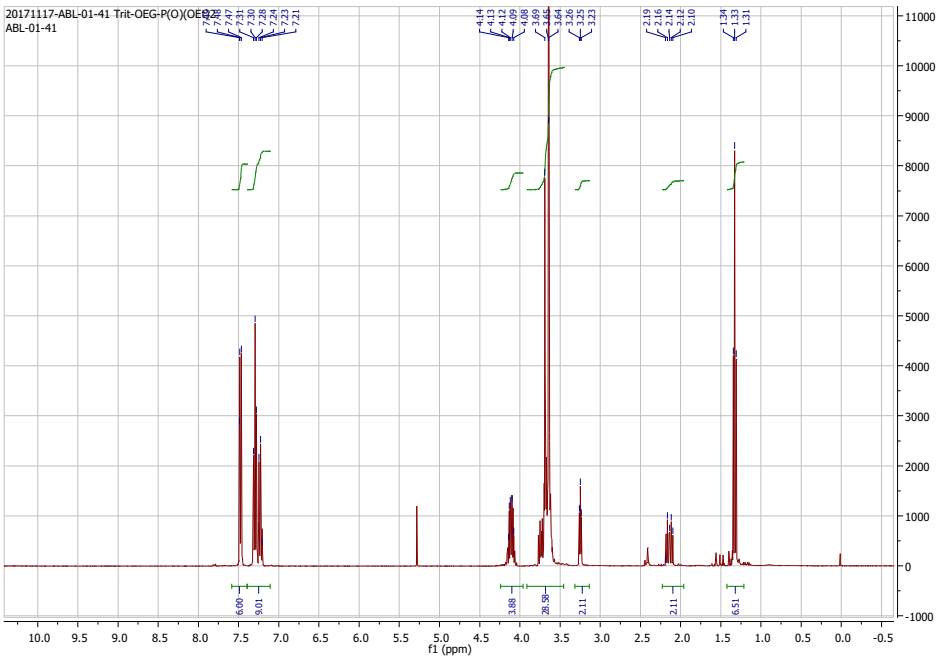


^{13}C NMR spectrum

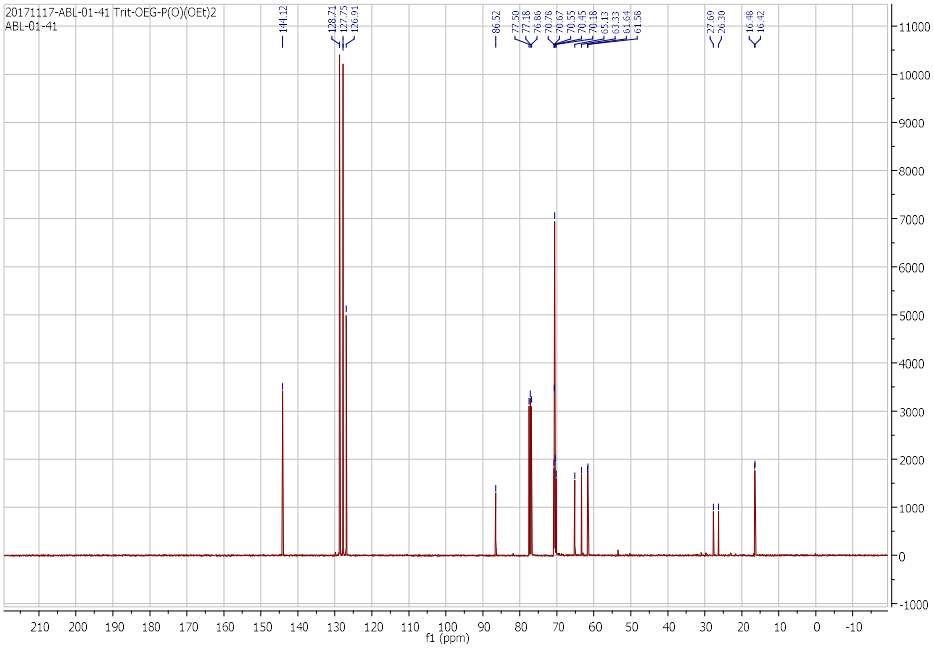
Compound 18



18

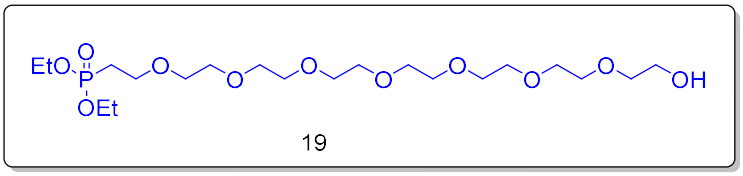


^{1}H NMR spectrum

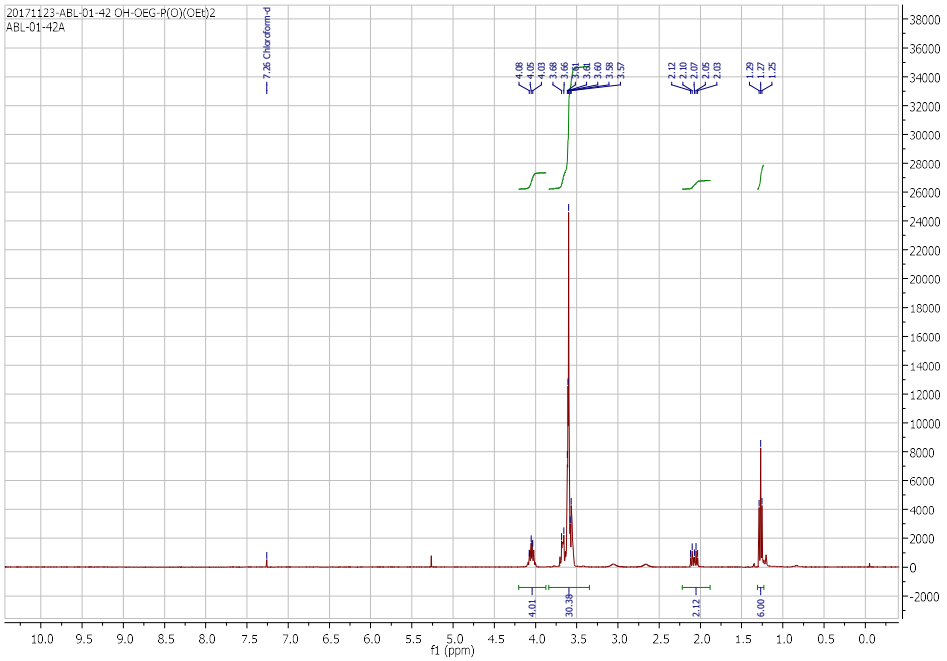


^{13}C NMR spectrum

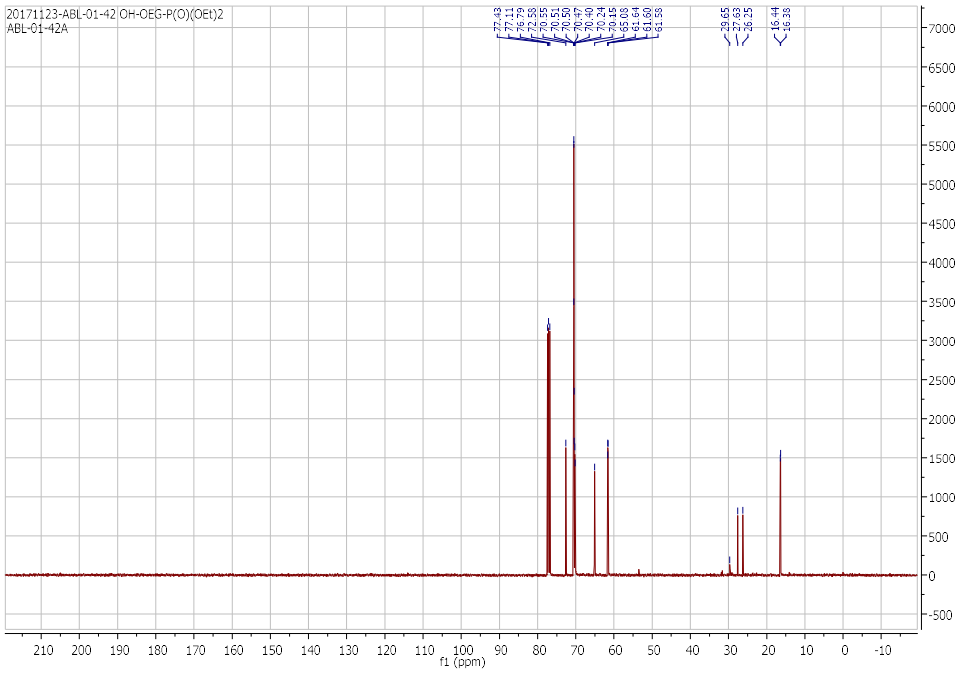
Compound 19



19

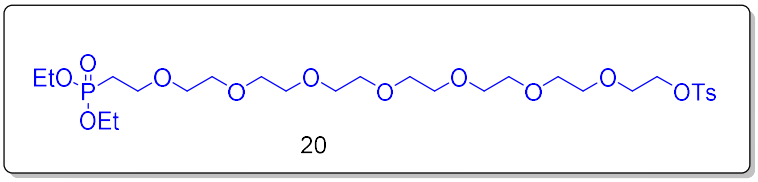


^{1}H NMR spectrum

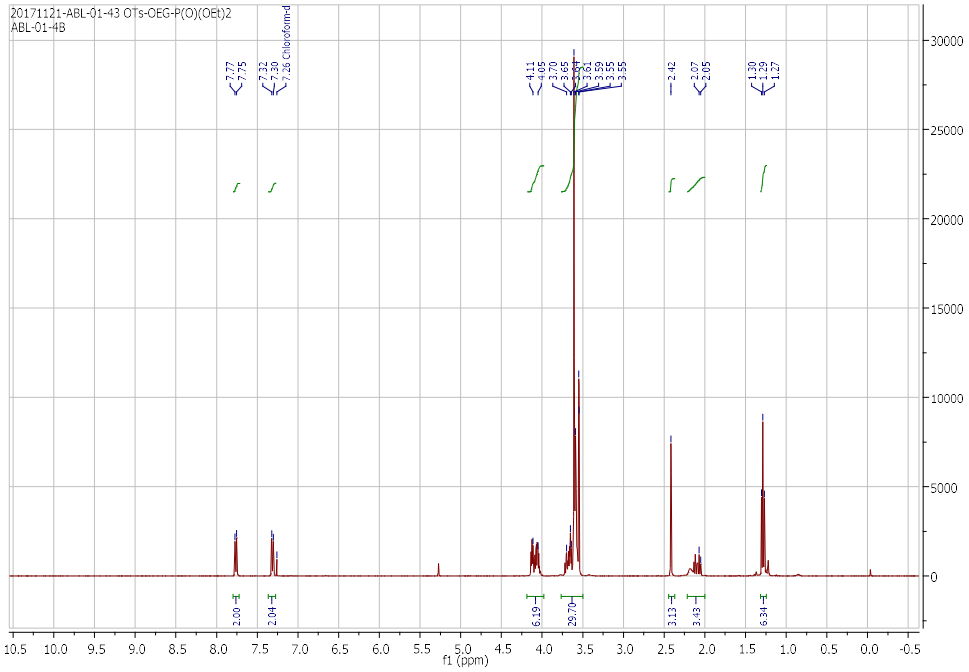


^{13}C NMR spectrum

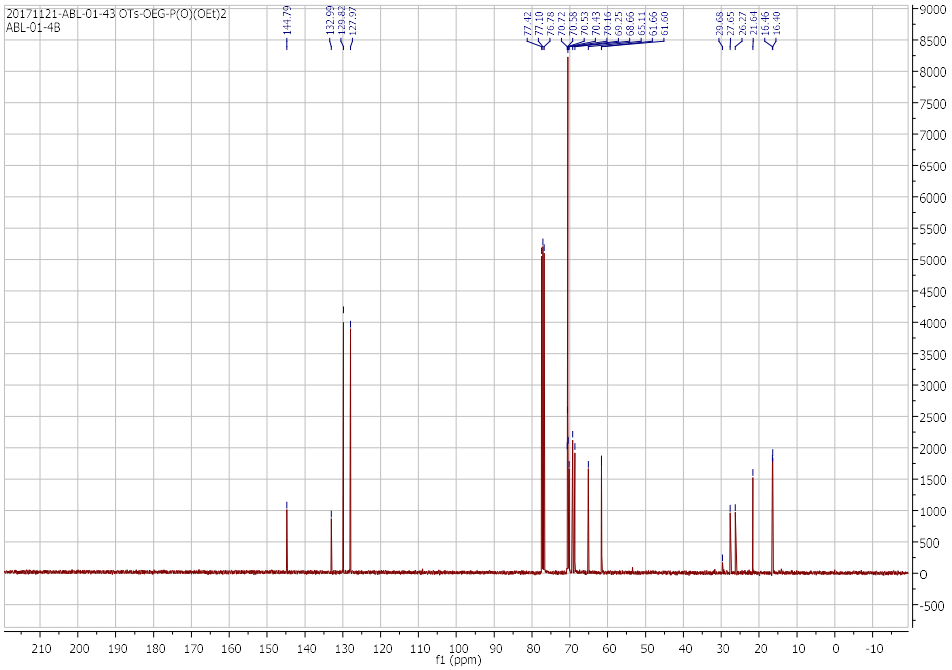
Compound 20



20

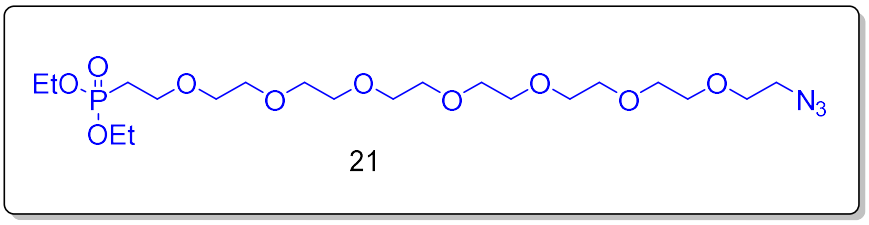


^{1}H NMR spectrum

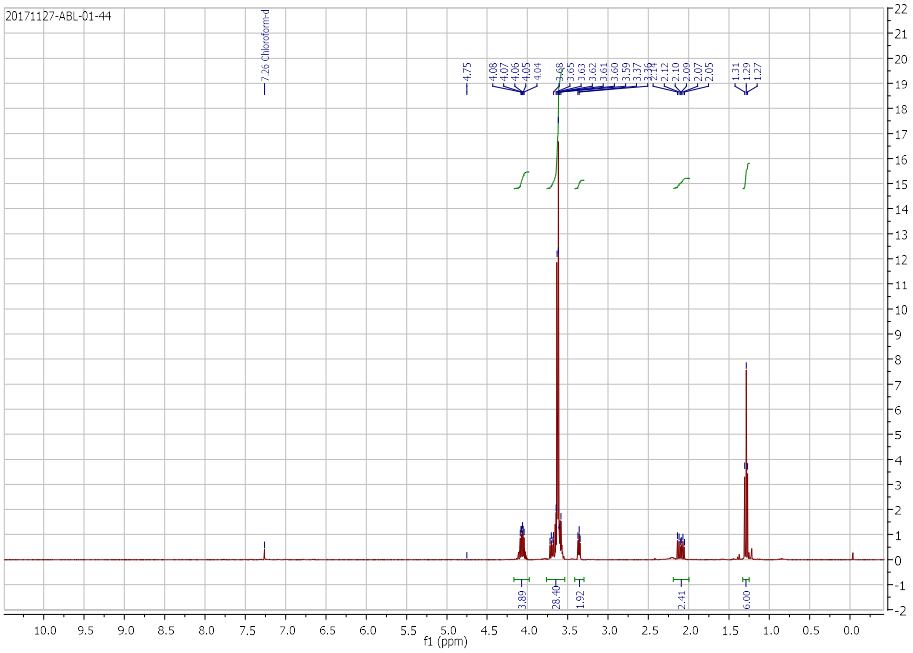


^{13}C NMR spectrum

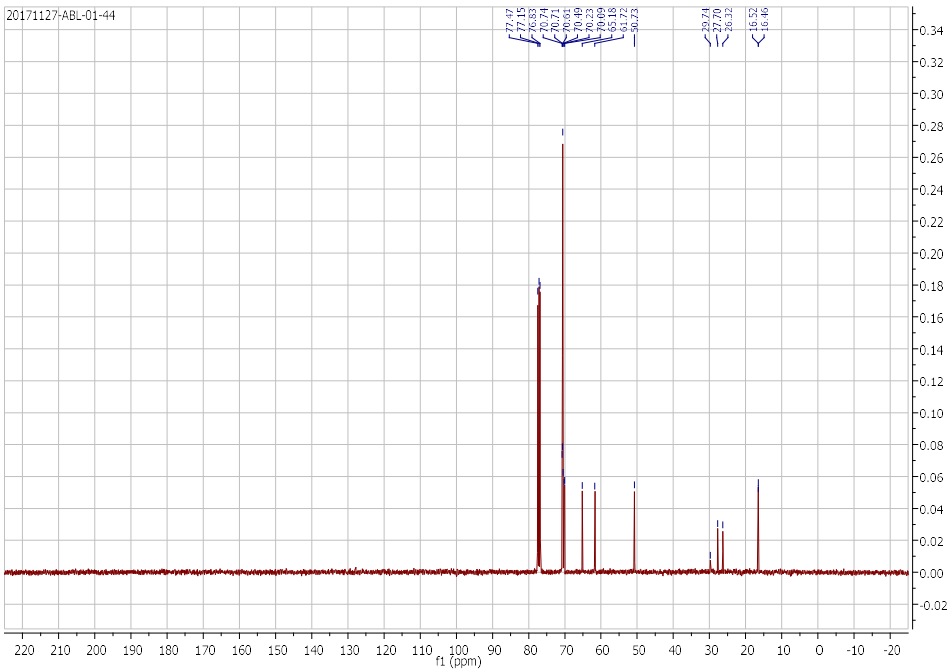
Compound 21



21

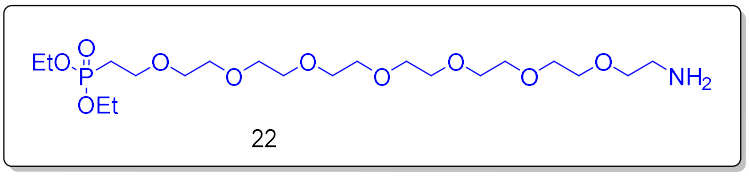


^{1}H NMR spectrum

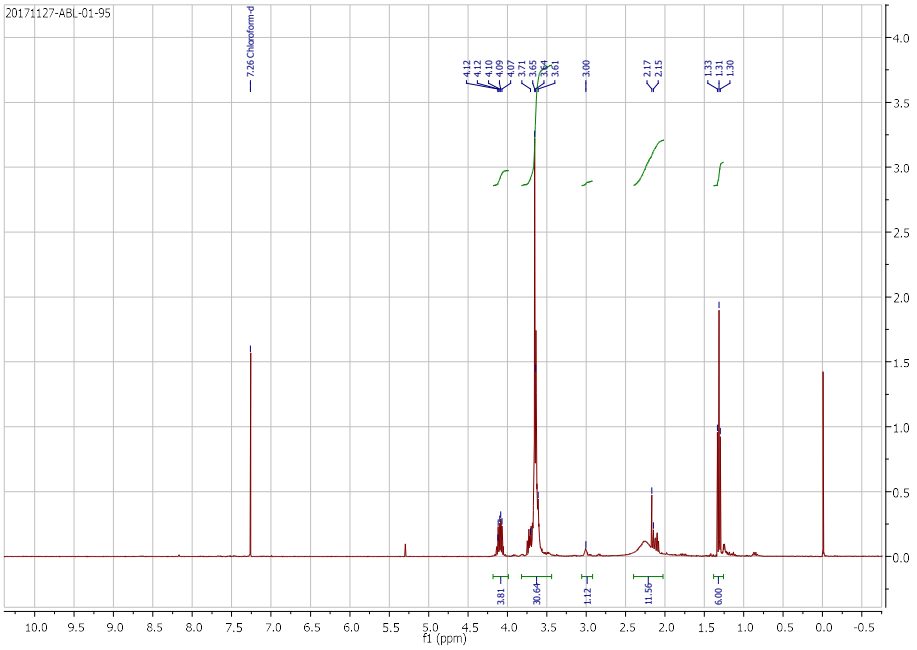


^{13}C NMR spectrum

Compound 22

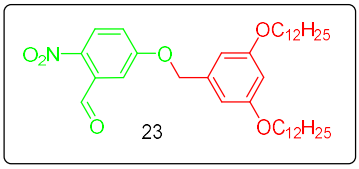


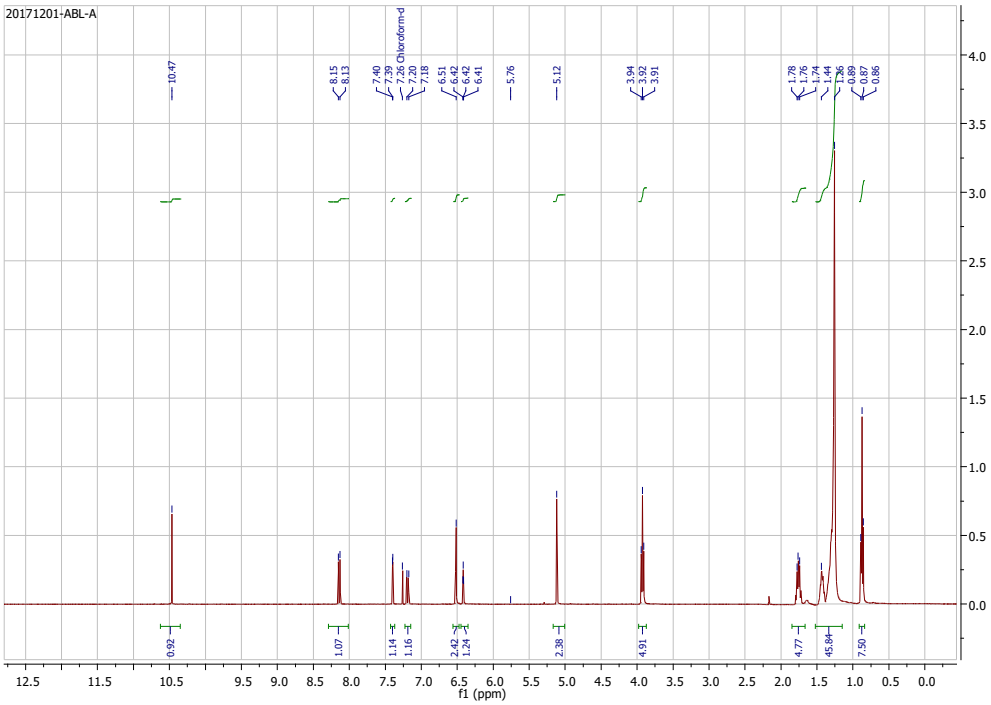
22



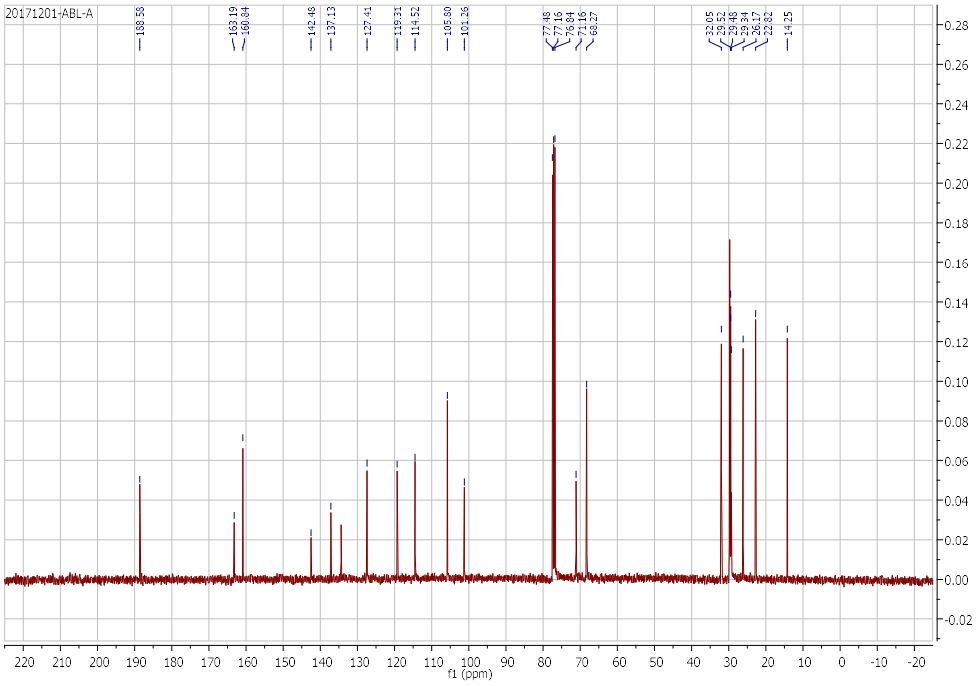
^{1}H NMR spectrum

Compound 23



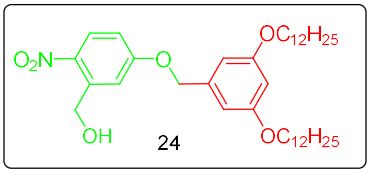


^{1}H NMR spectrum

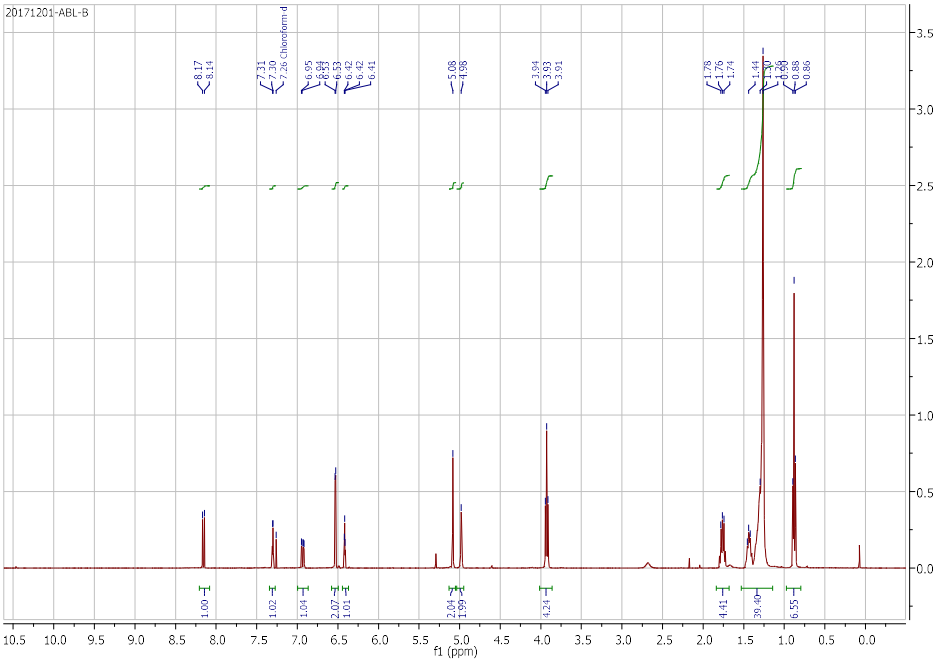


^{13}C NMR spectrum

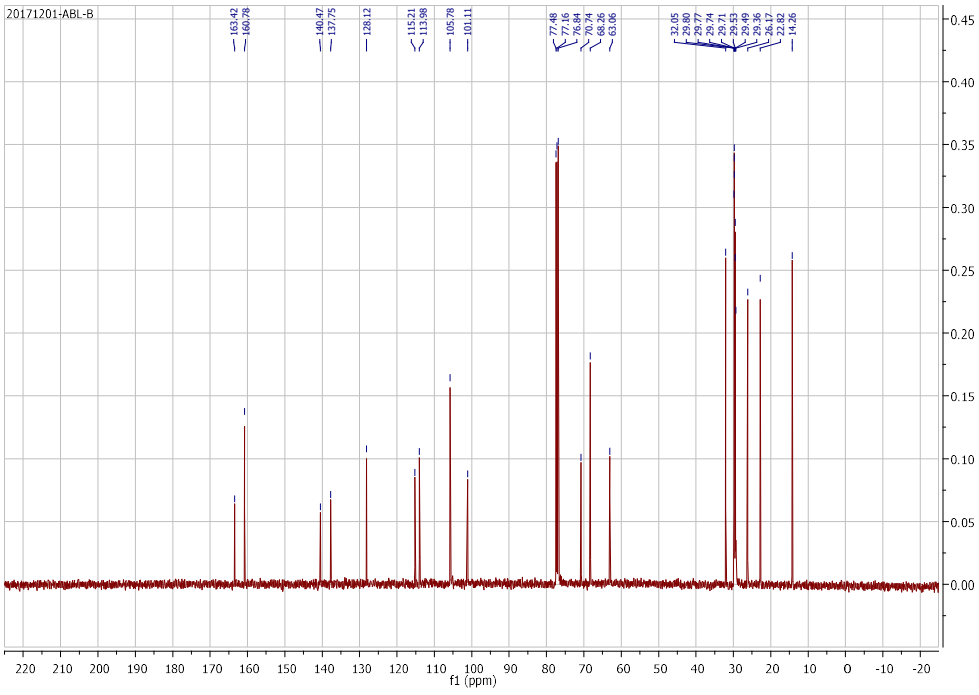
Compound 24



24

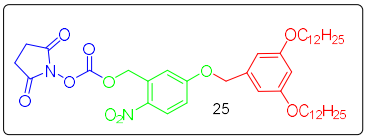


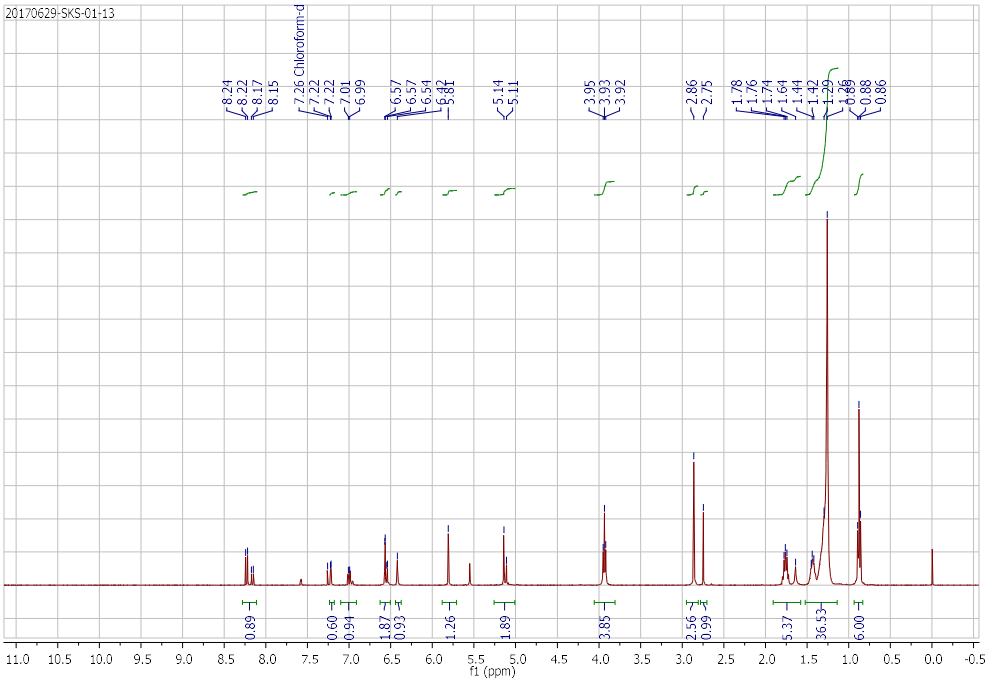
^{1}H NMR spectrum



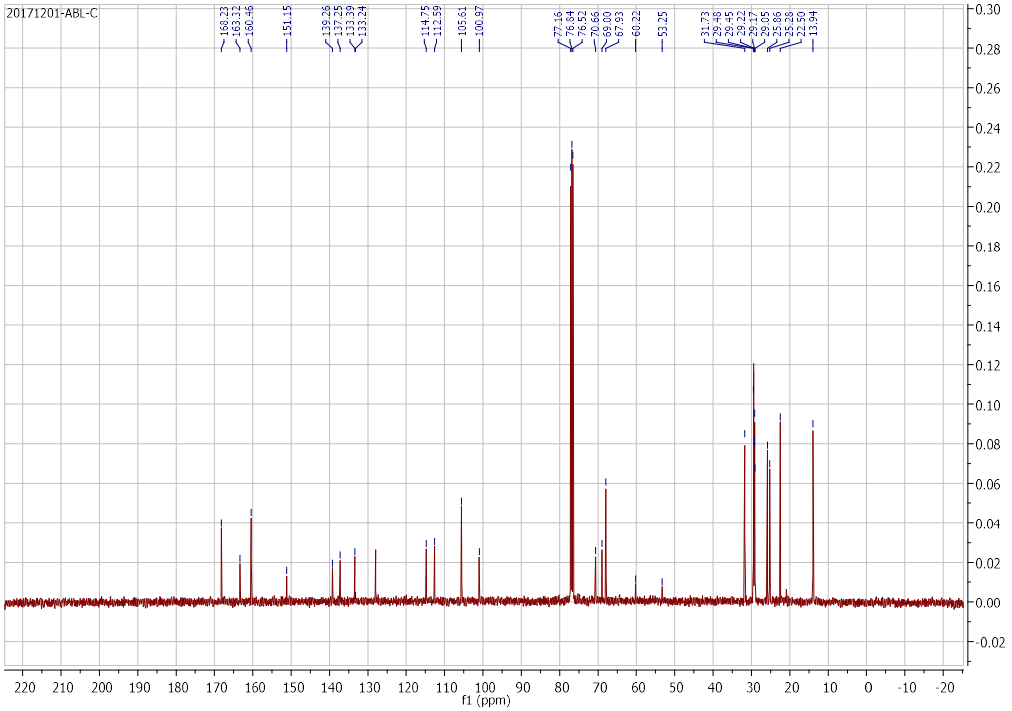
^{13}C NMR spectrum

Compound 25



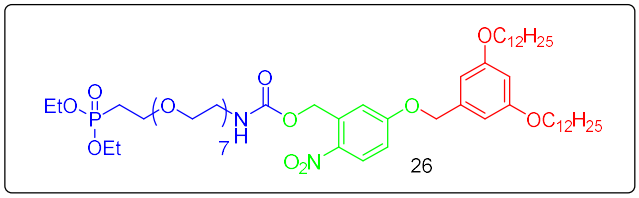


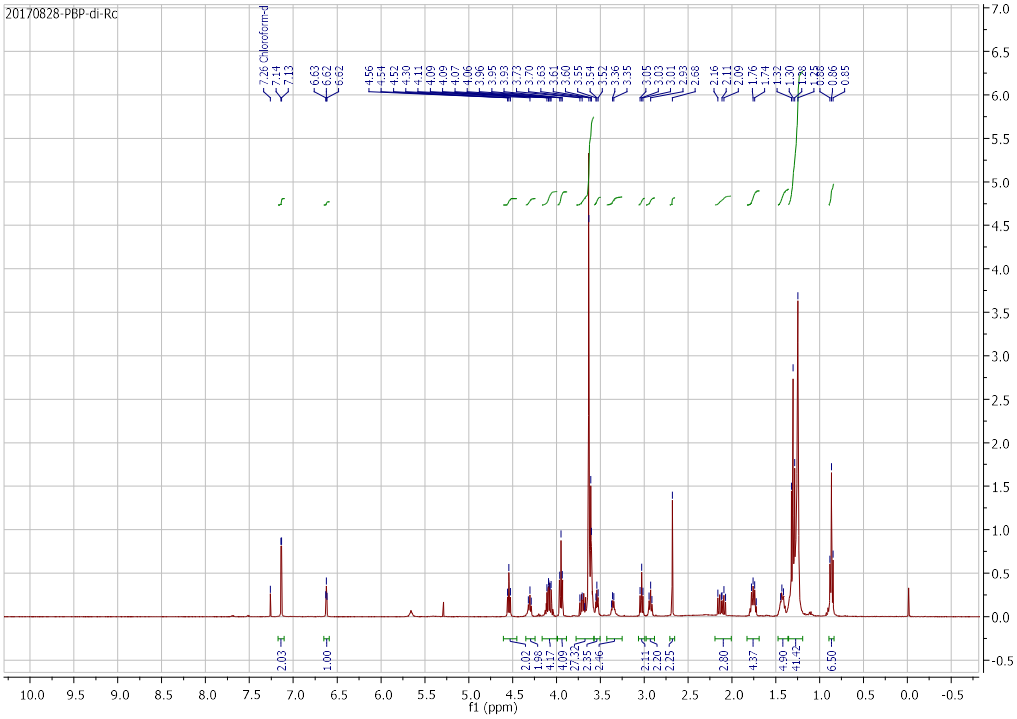
^{1}H NMR spectrum



^{13}C NMR spectrum

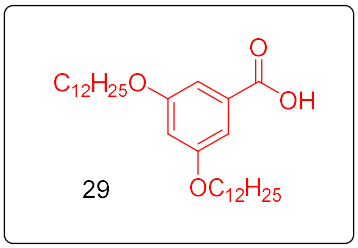
Compound 26

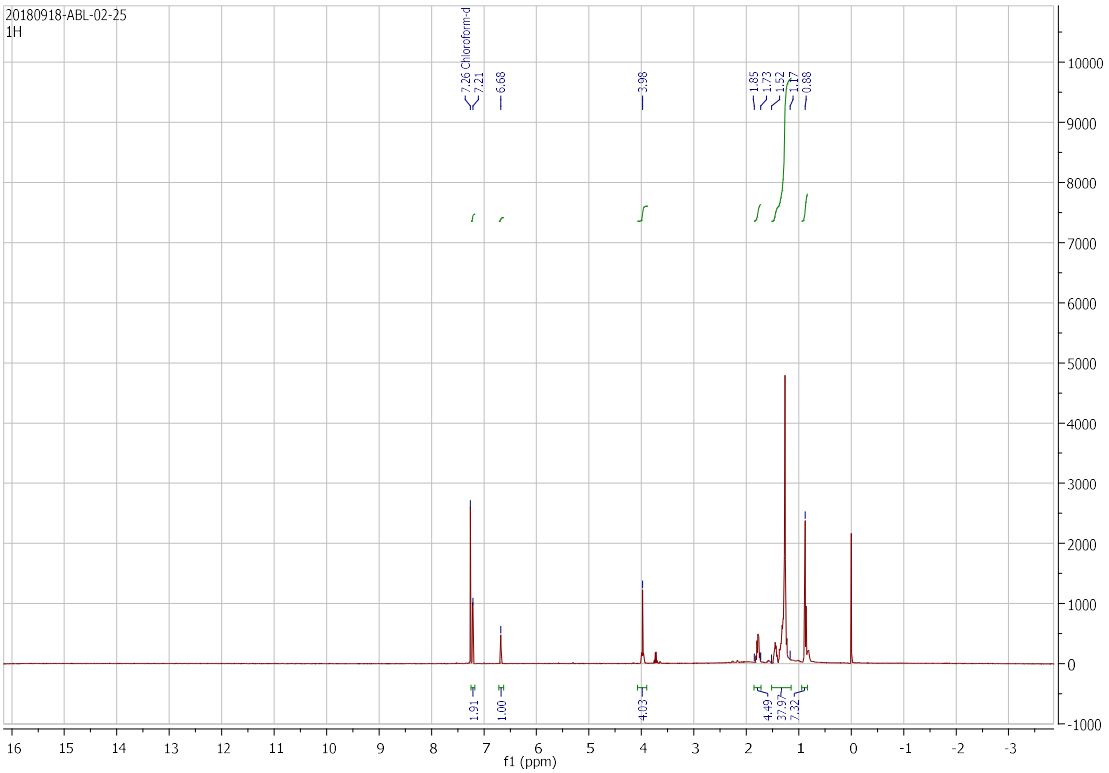




^{1}H NMR spectrum

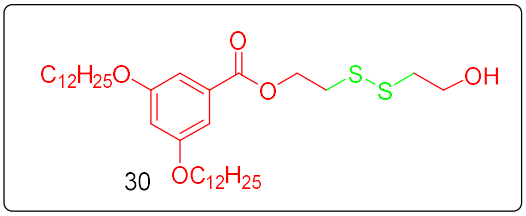
Compound 29



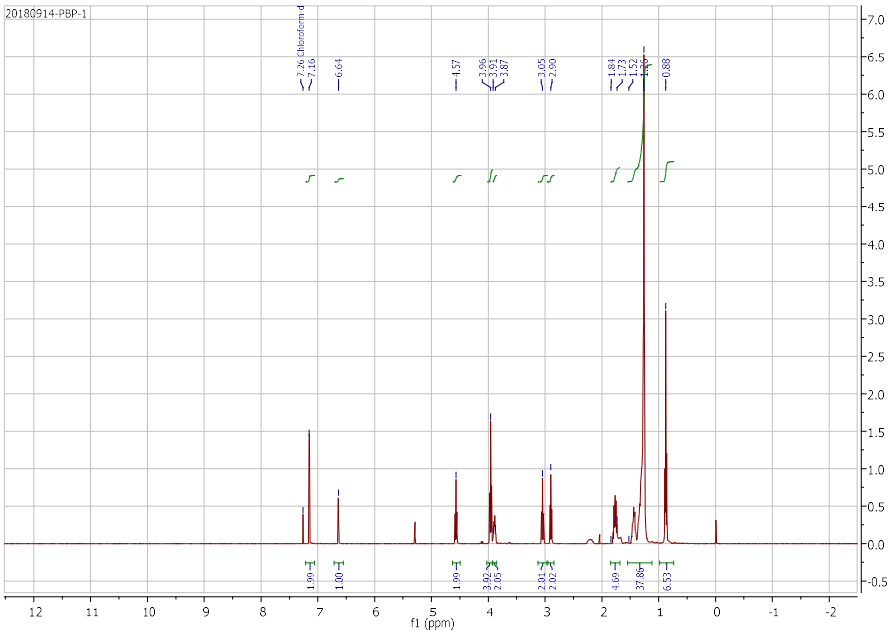


^{1}H NMR spectrum

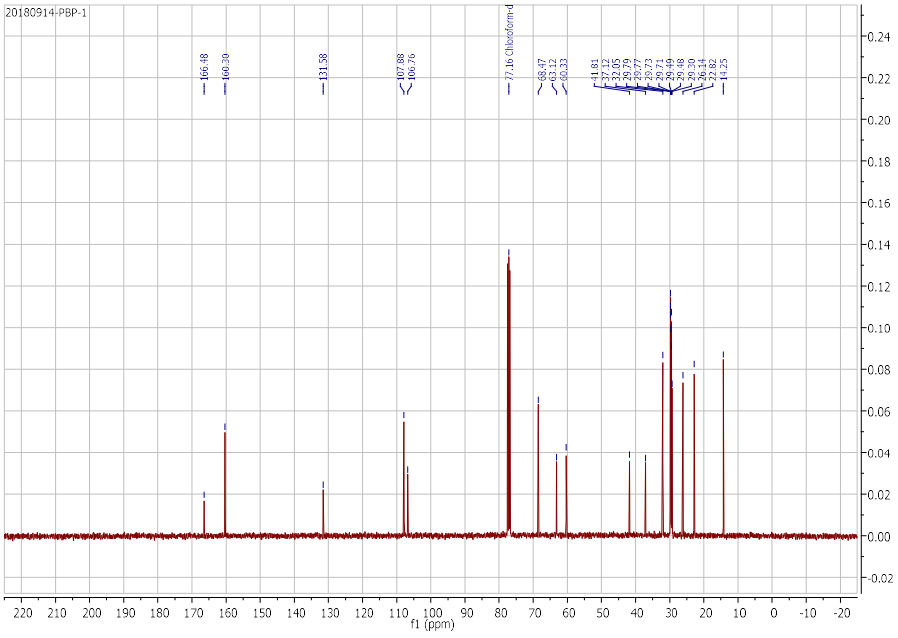
Compound 30



30

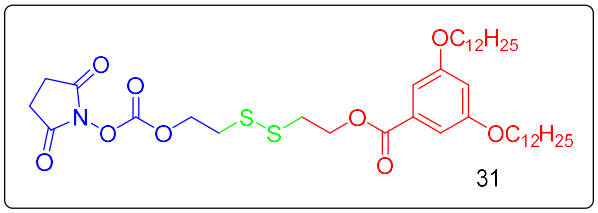


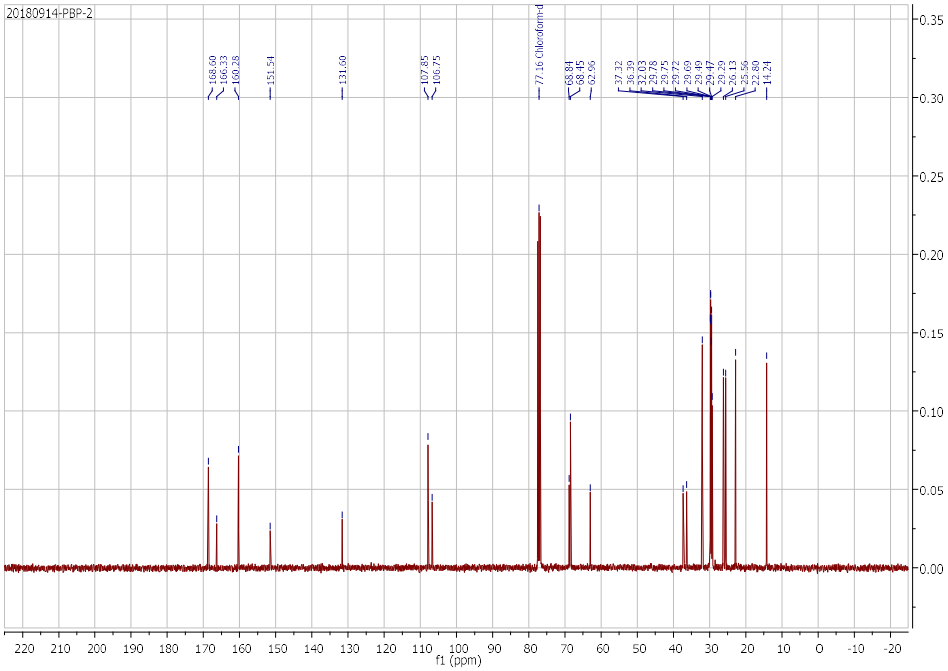
^{1}H NMR spectrum



^{13}C NMR spectrum

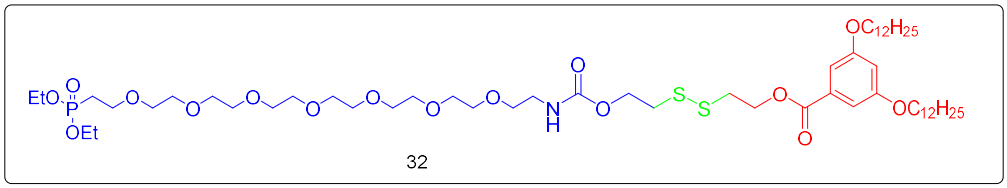
Compound 31



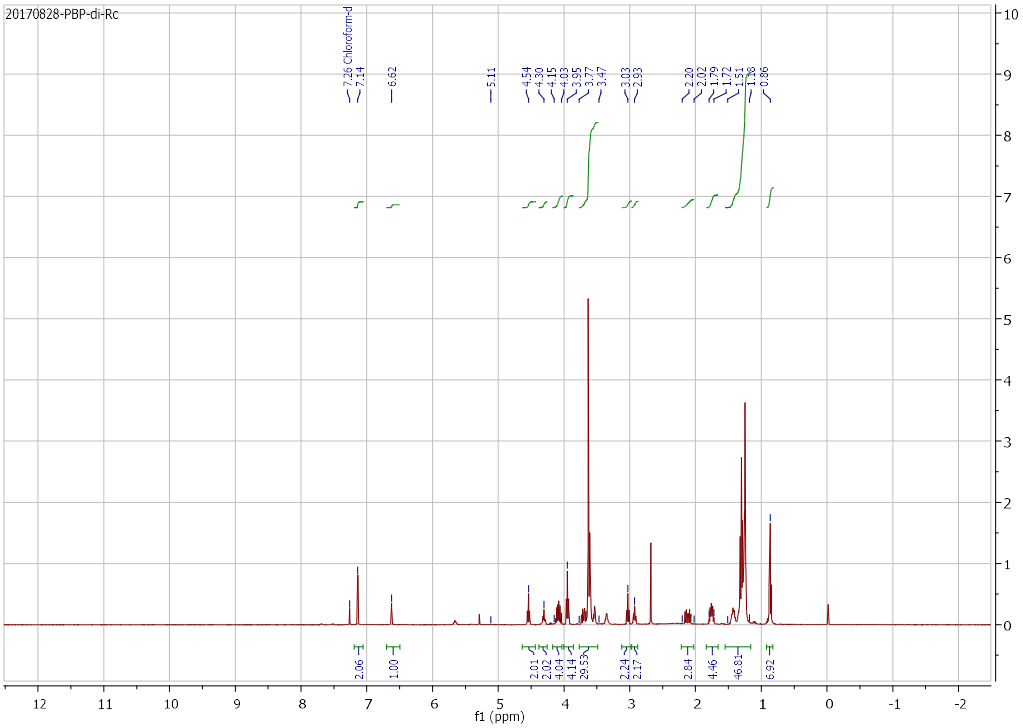


^{13}C NMR spectrum

Compound 32

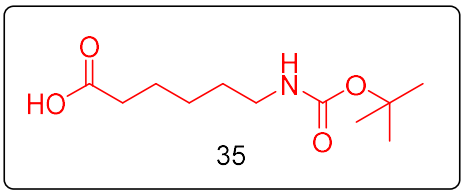


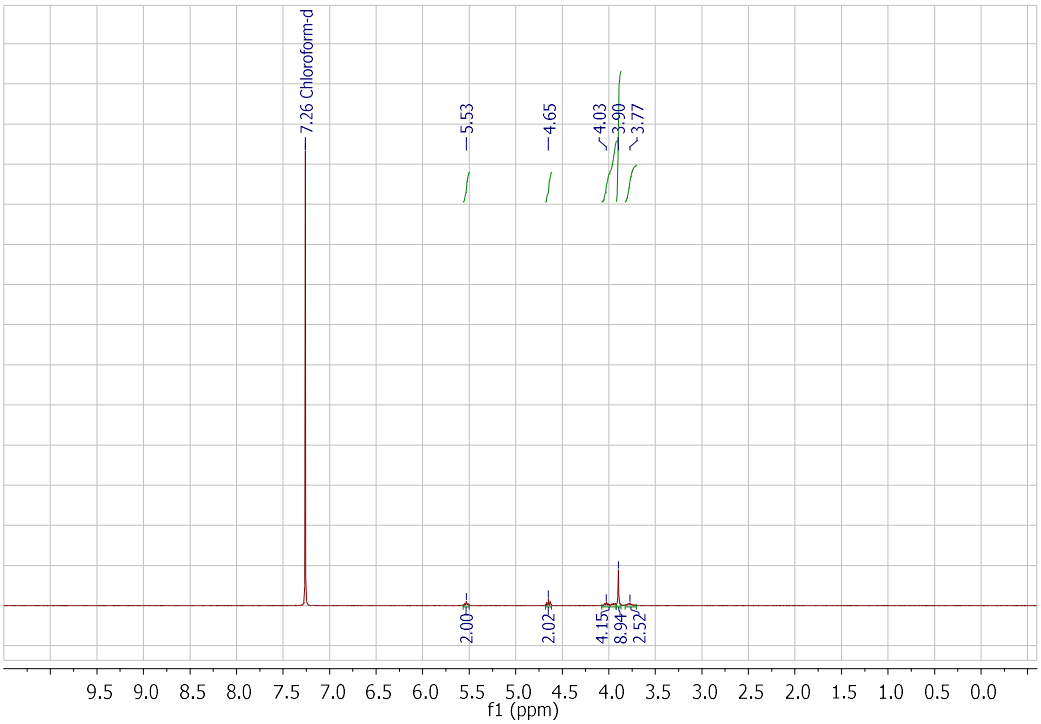
32



^{1}H NMR spectrum

Compound 35

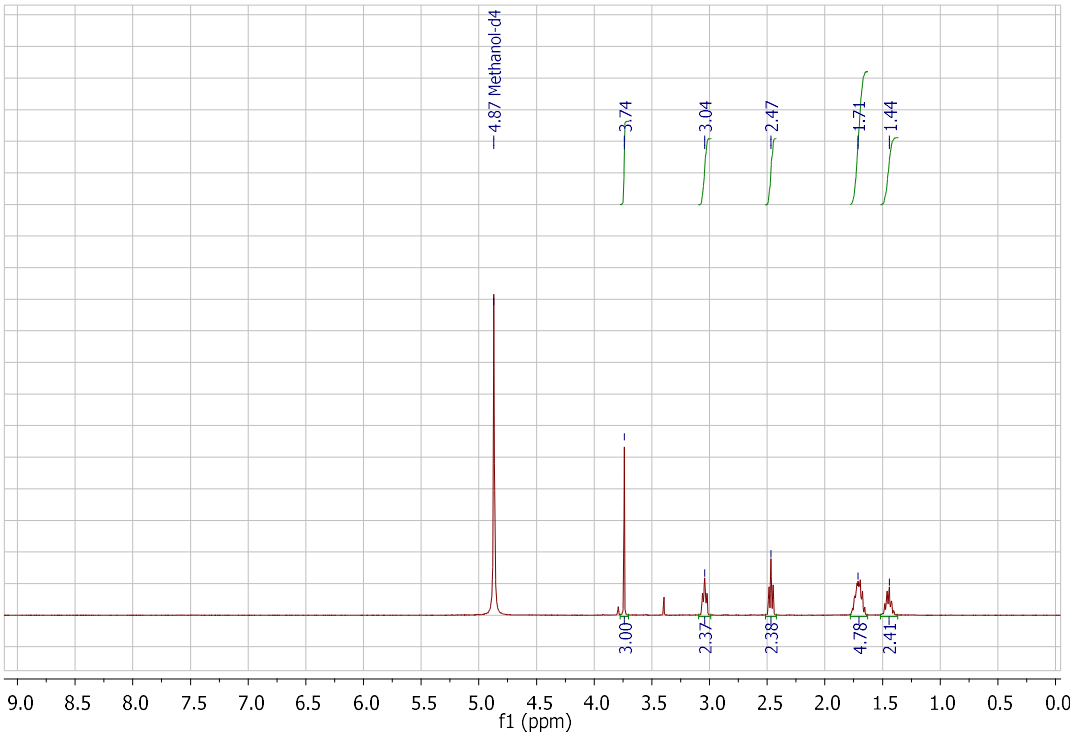




^{1}H NMR spectrum

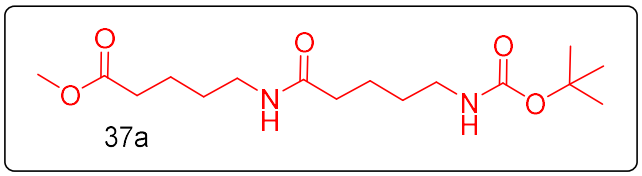
Compound 36

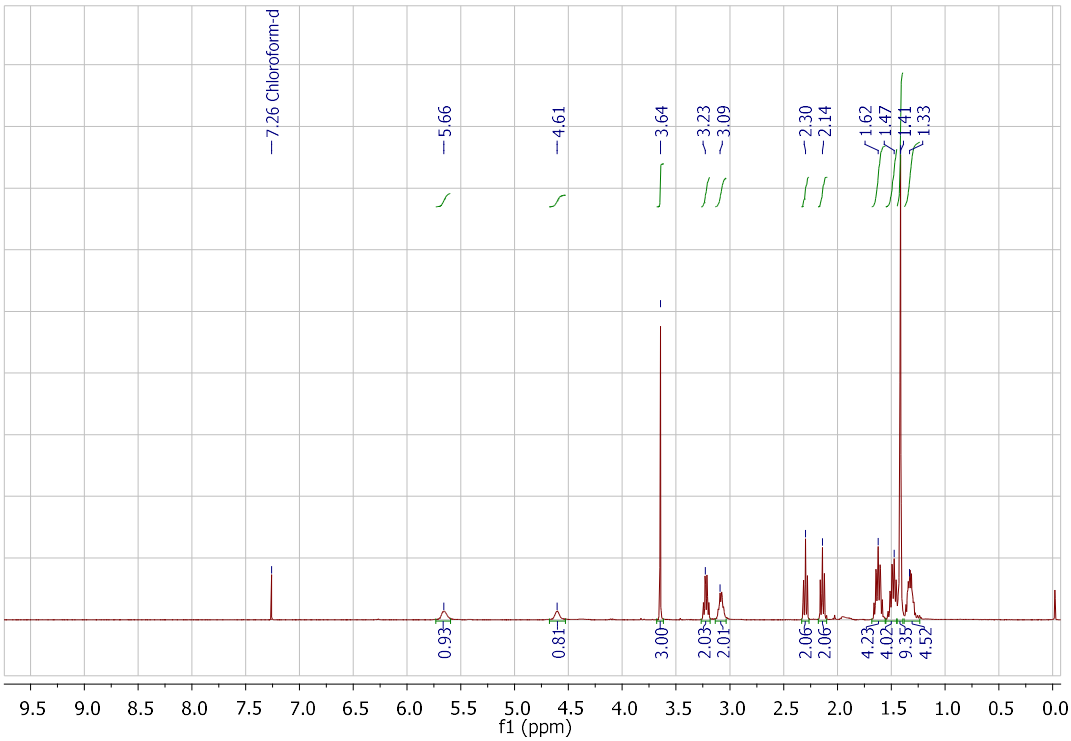




^{1}H NMR spectrum

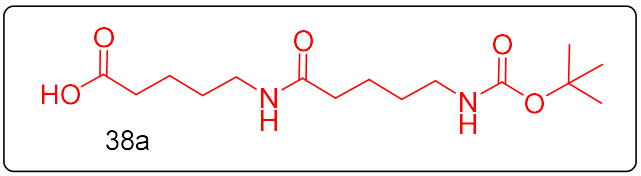
Compound 37a



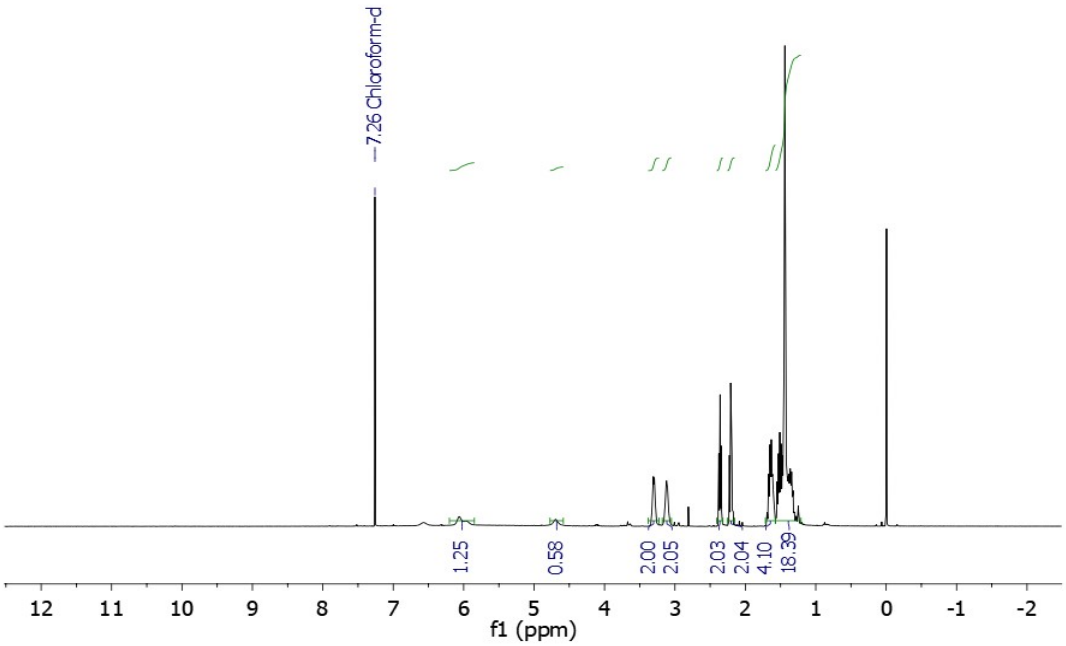


^{1}H NMR spectrum

Compound 38a

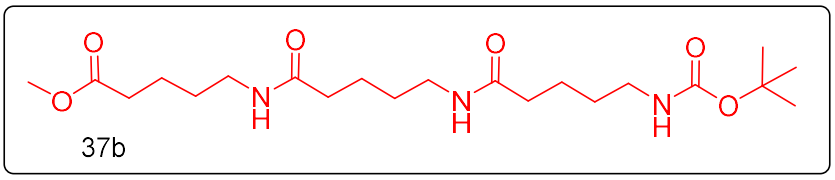


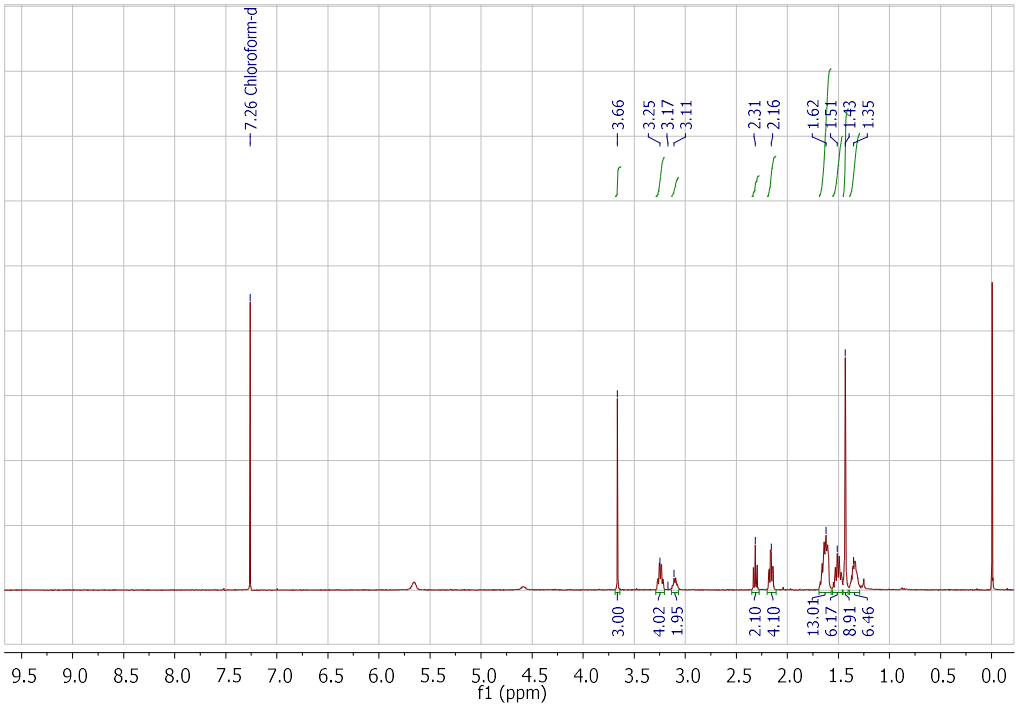
38a



^{1}H NMR spectrum

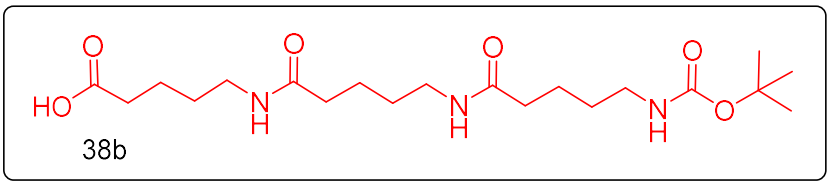
Compound 37b

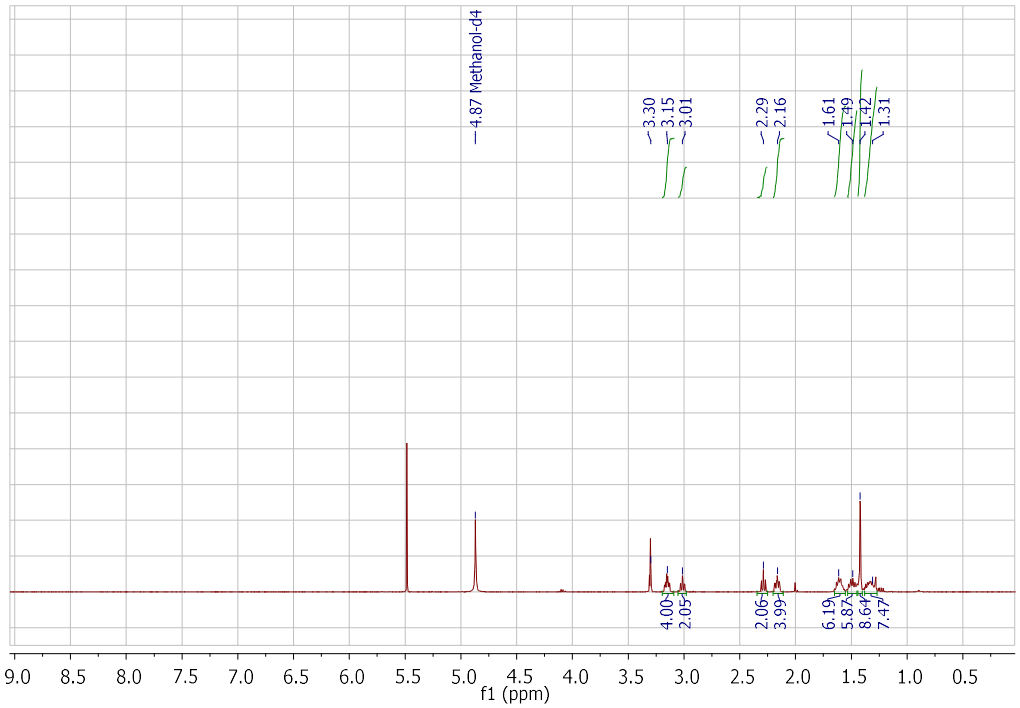




^{1}H NMR spectrum

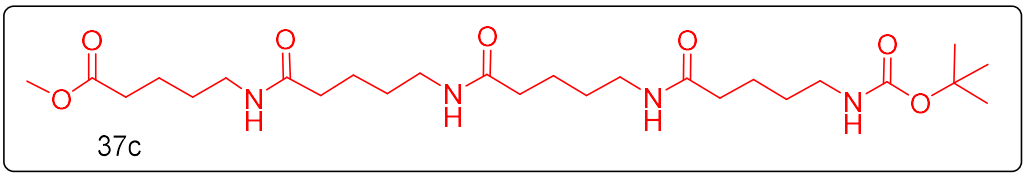
Compound 38b

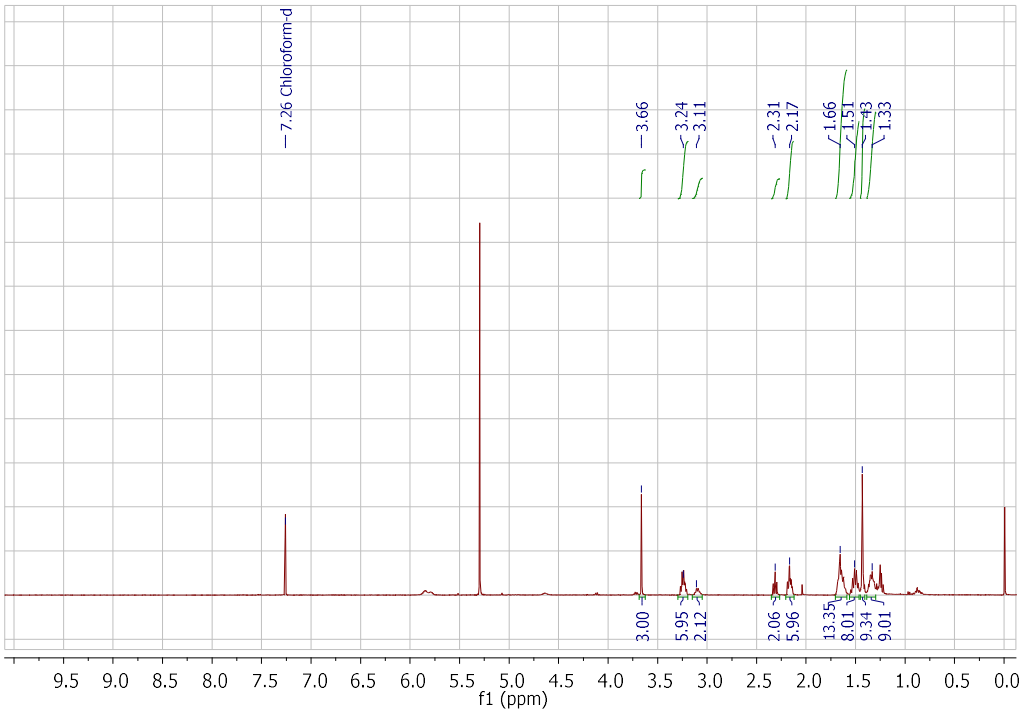




^{1}H NMR spectrum

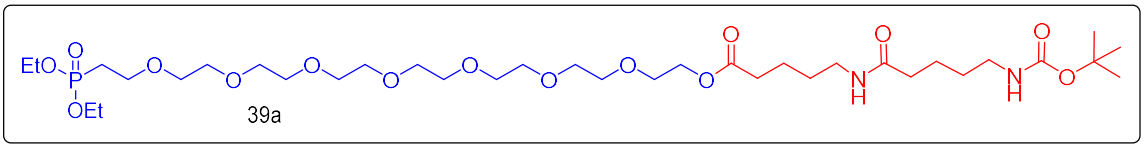
Compound 37c

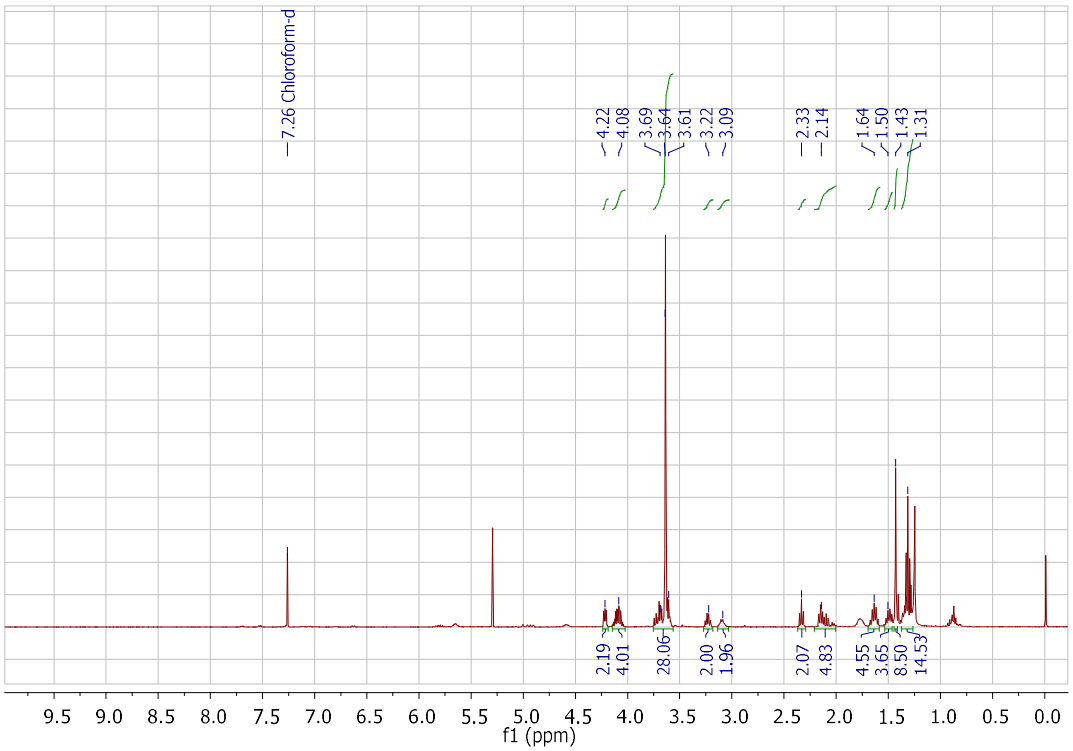




^{1}H NMR spectrum

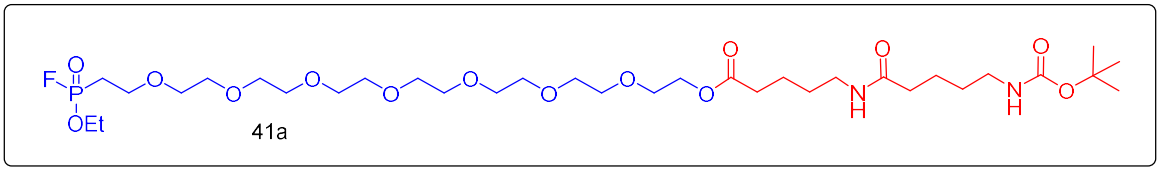
Compound 39a

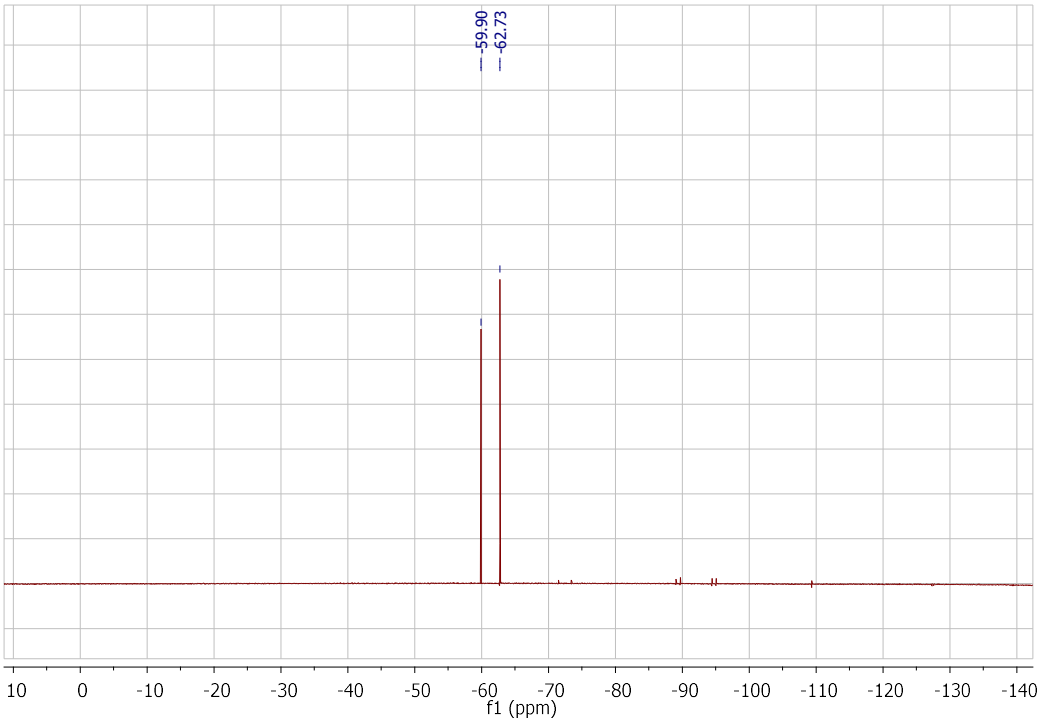




^{1}H NMR spectrum

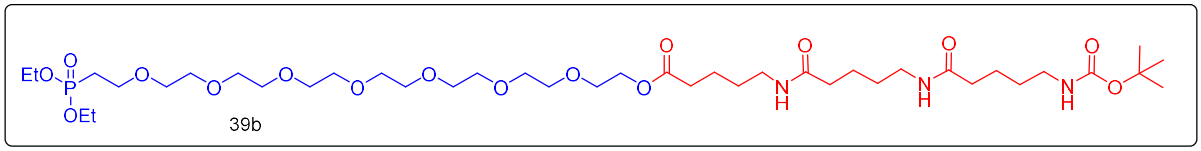
Compound 41a

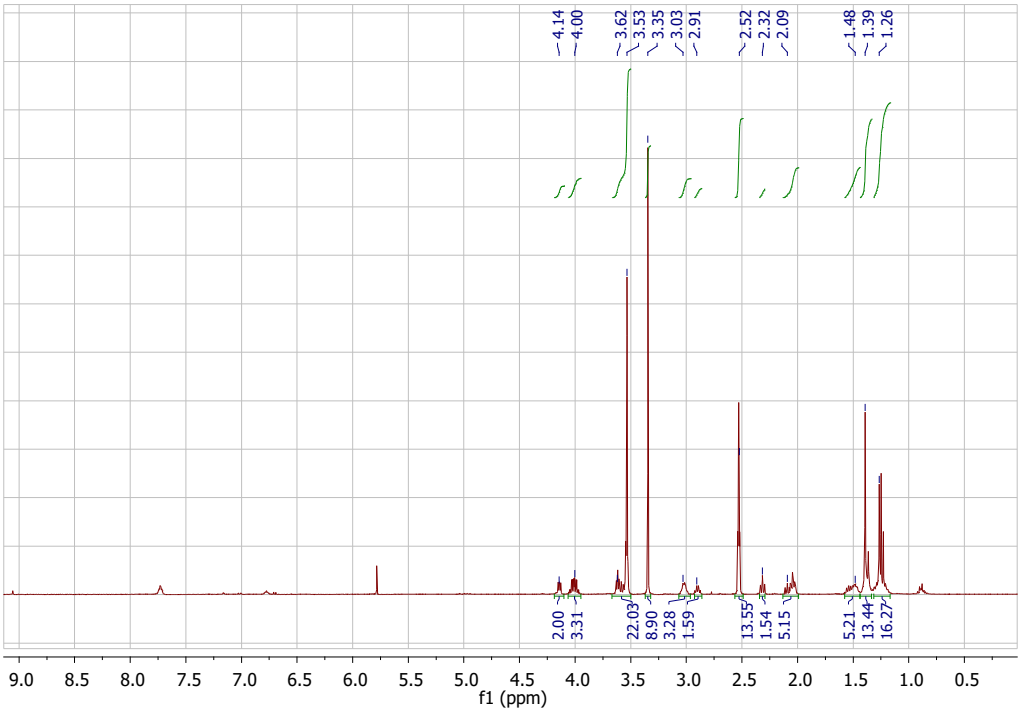




^{19}F NMR spectrum

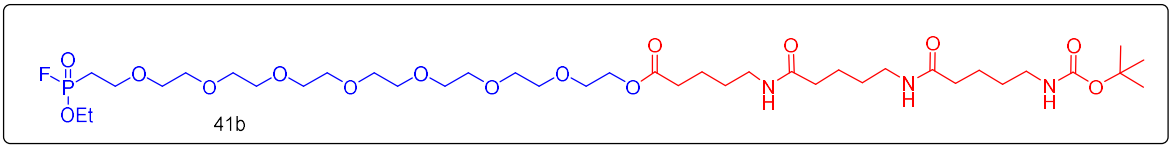
Compound 39b



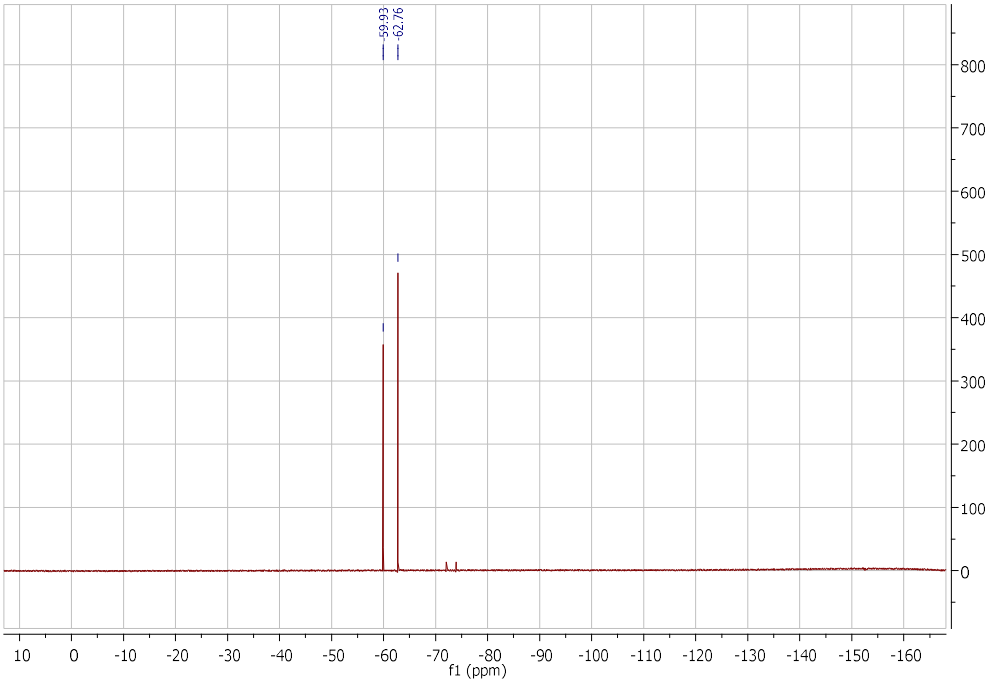


^{1}H NMR spectrum

Compound 41b

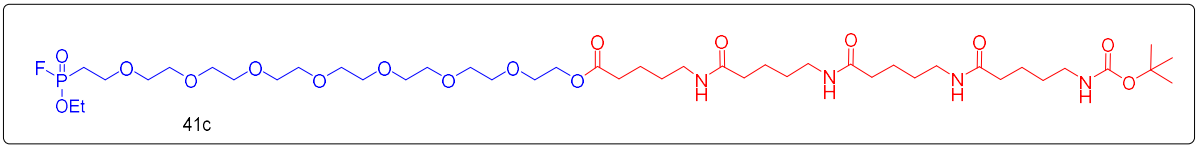


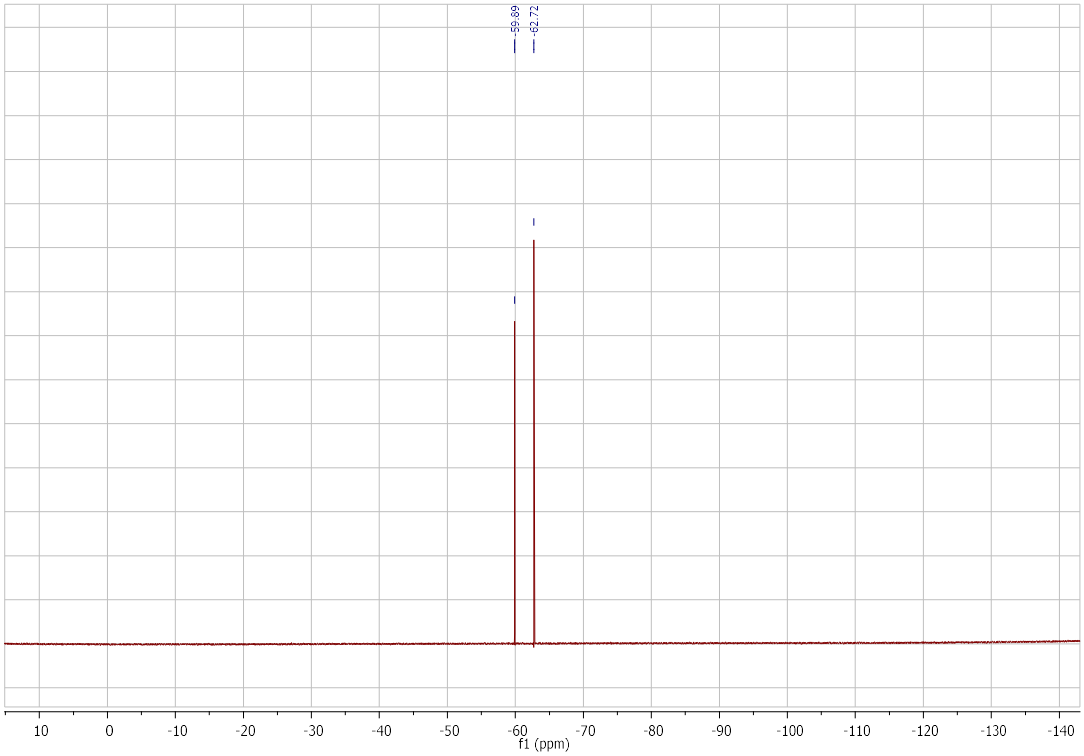
41b



^{19}F NMR spectrum

Compound 41c

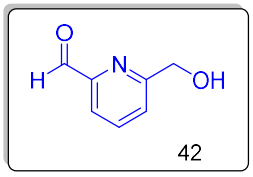


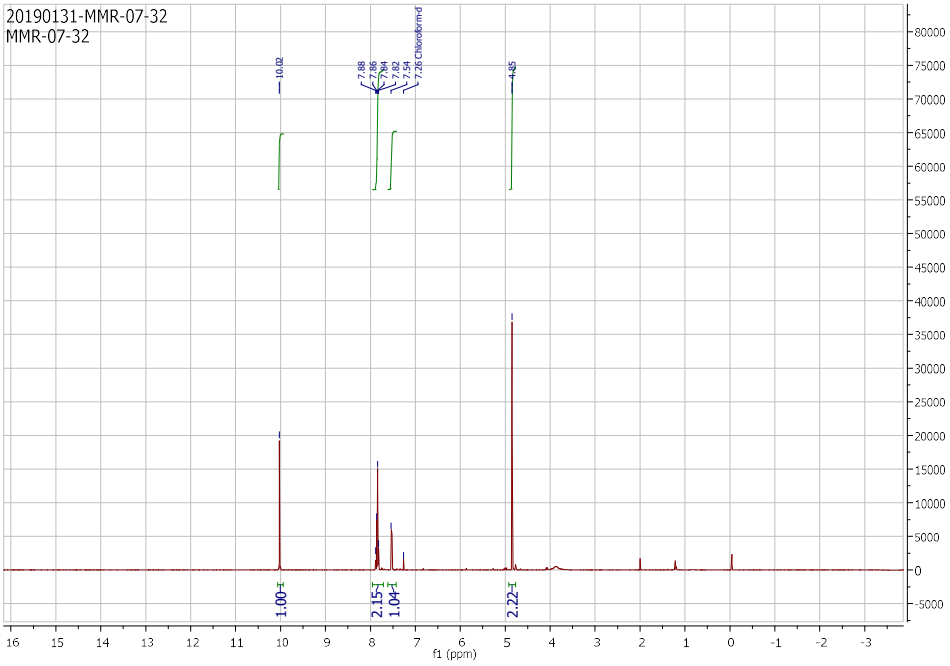


^{19}F NMR spectrum

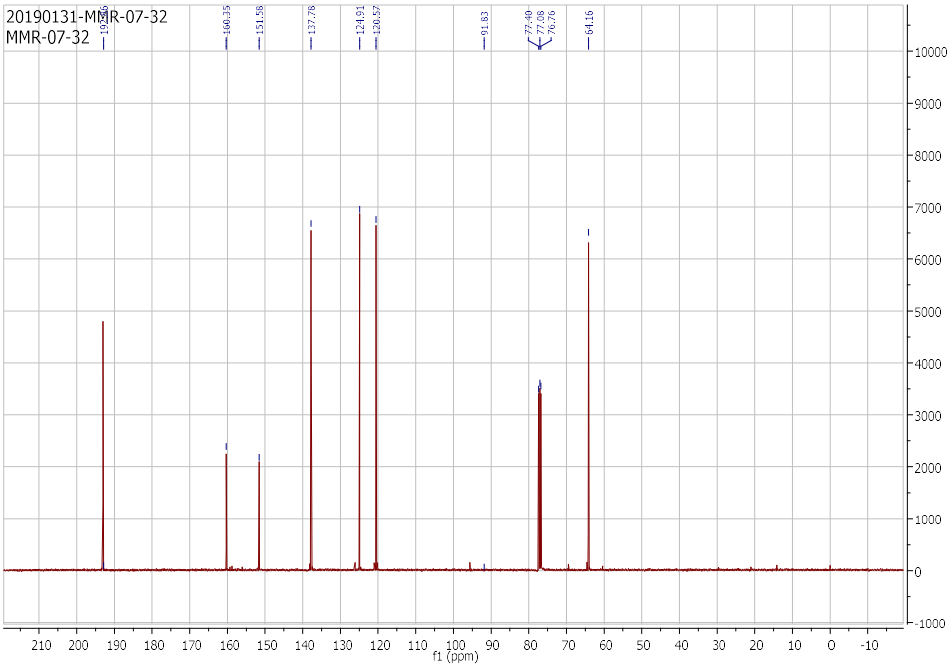
Chapter 4 Spectral Data

Compound 42



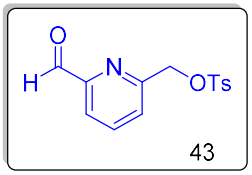


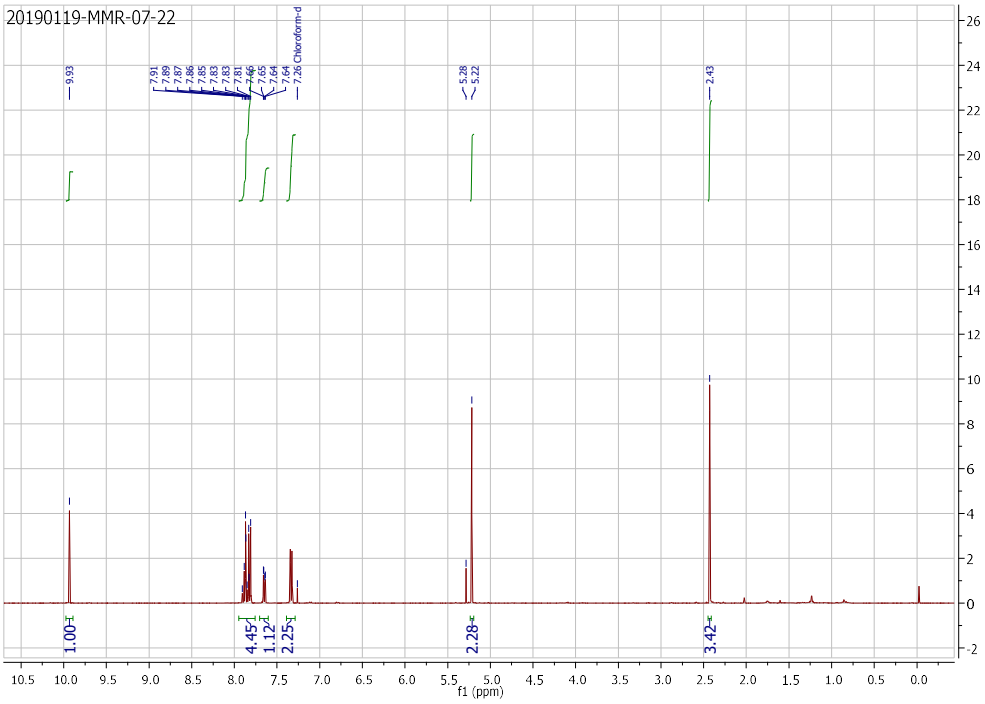
^{1}H NMR spectrum



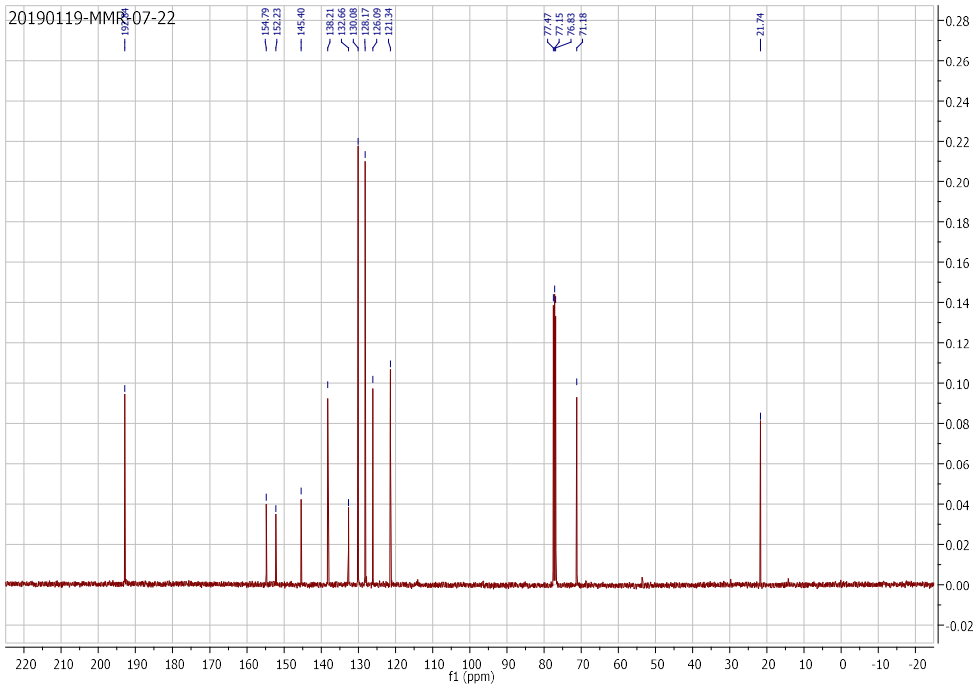
^{13}C NMR spectrum

Compound 43



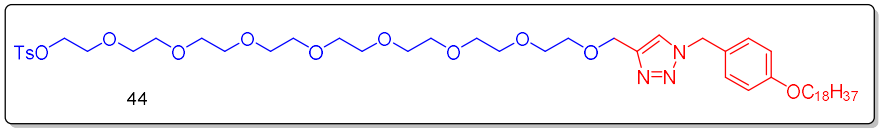


^{1}H NMR spectrum

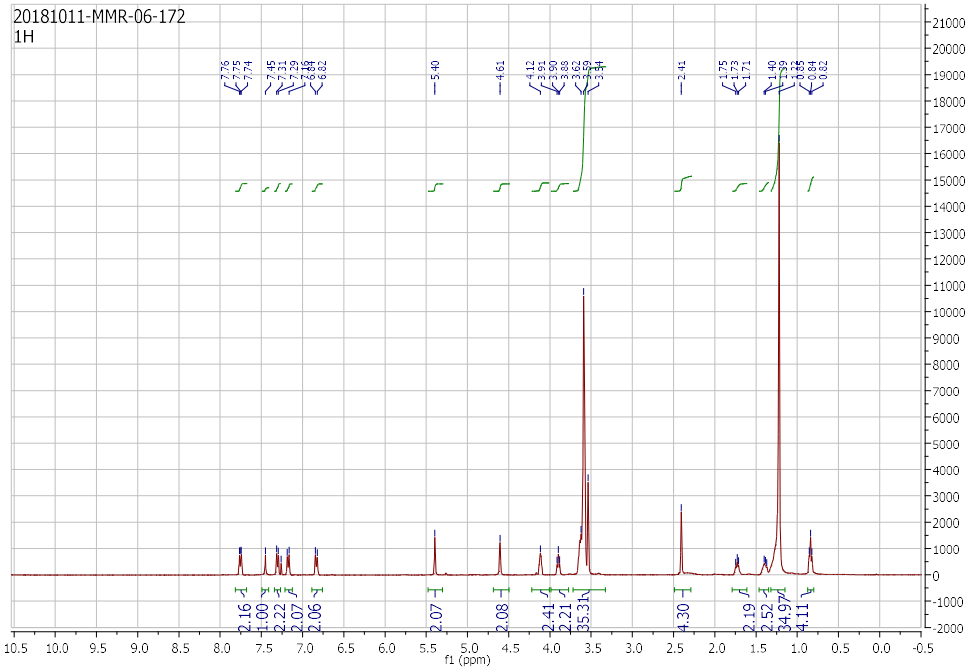


^{13}C NMR spectrum

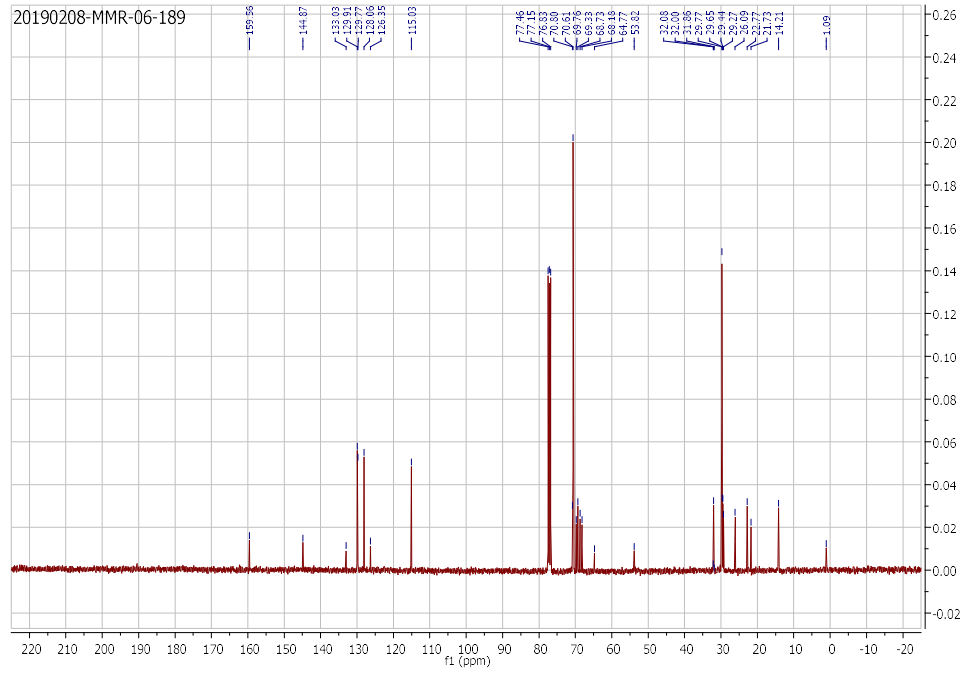
Compound 44



44

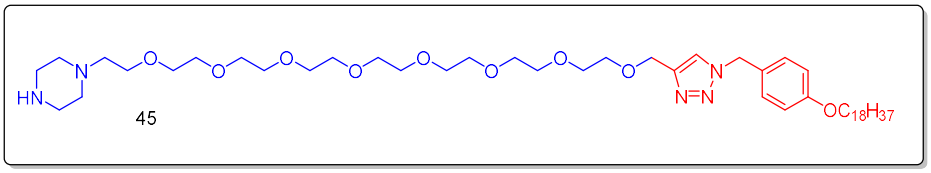


^{1}H NMR spectrum

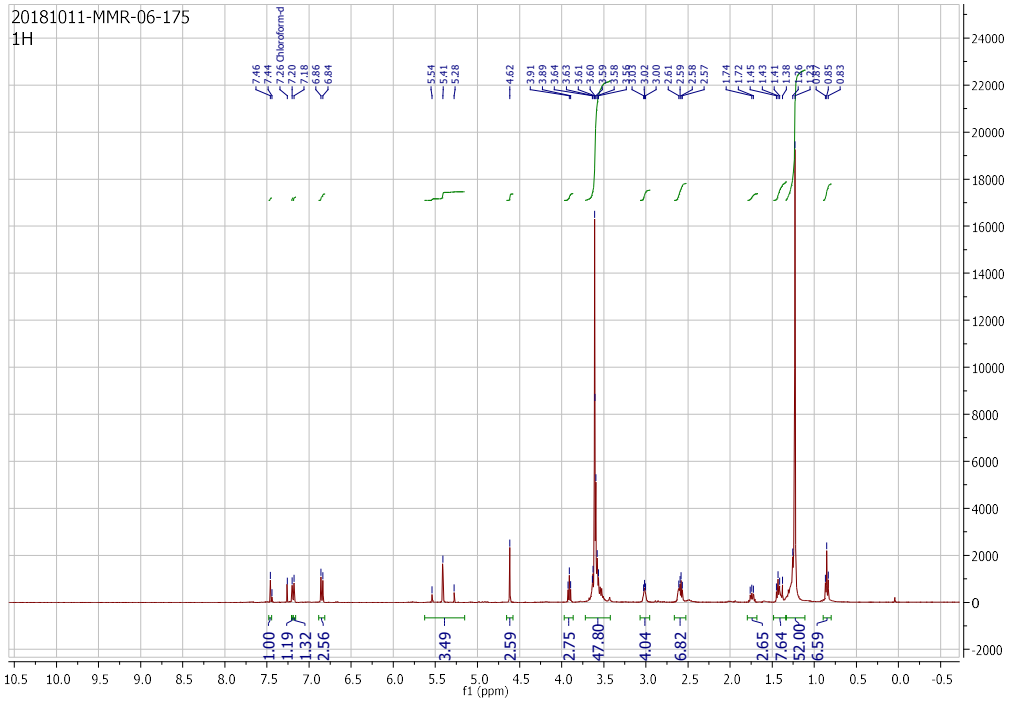


^{13}C NMR spectrum

Compound 45

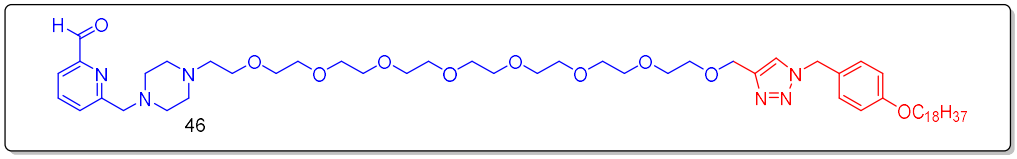


45

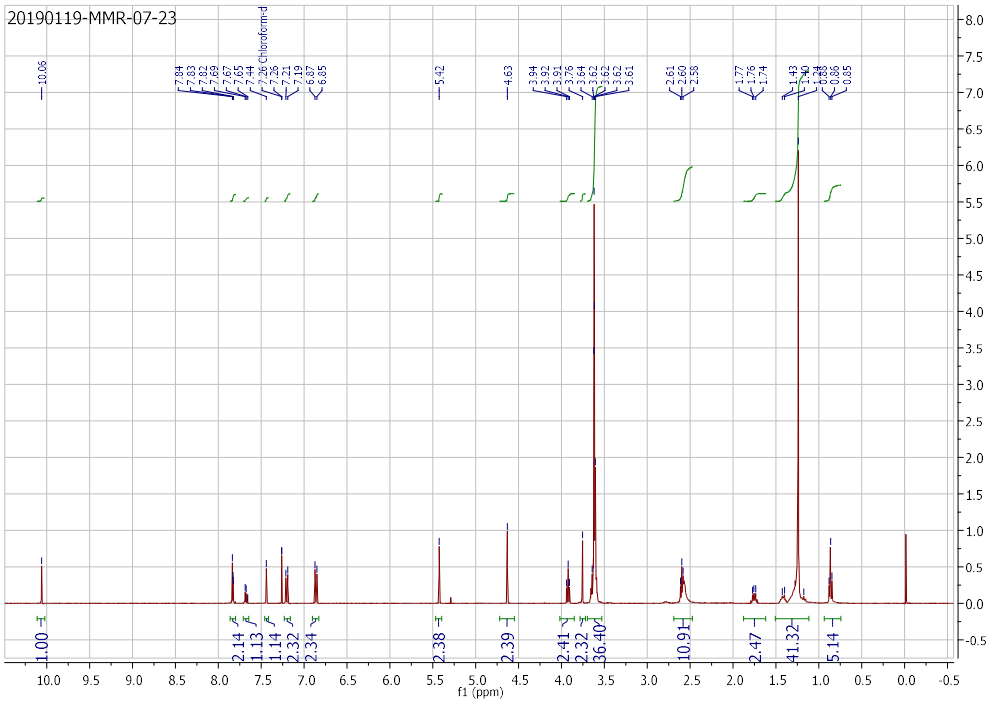


^{1}H NMR spectrum

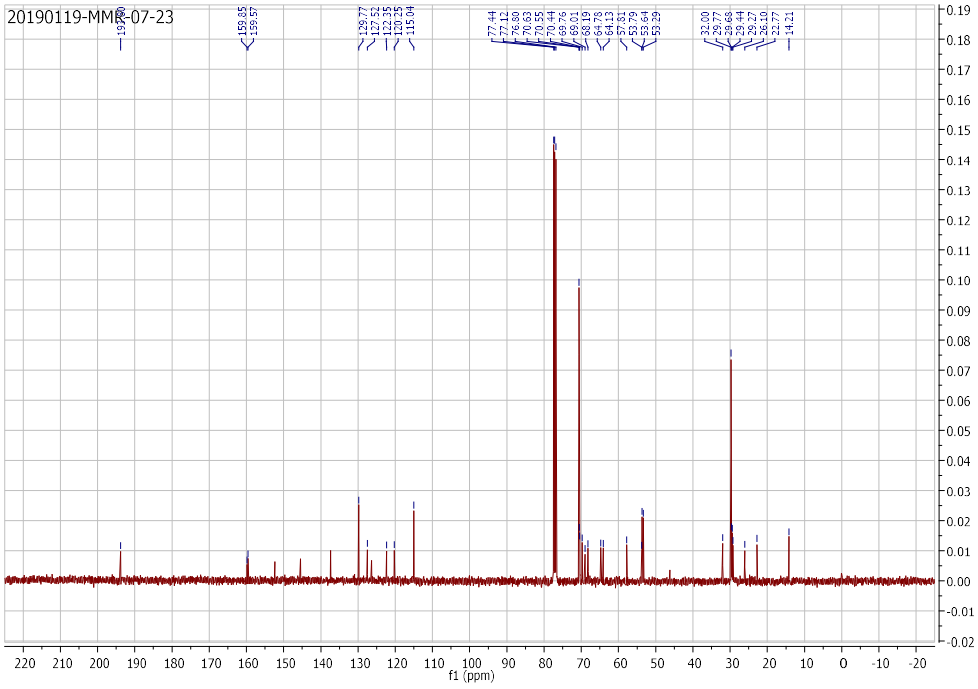
Compound 46



46

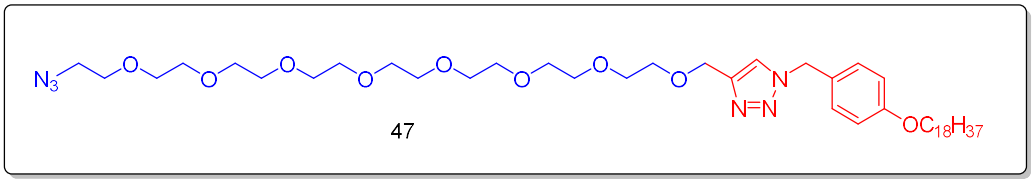


^{1}H NMR spectrum

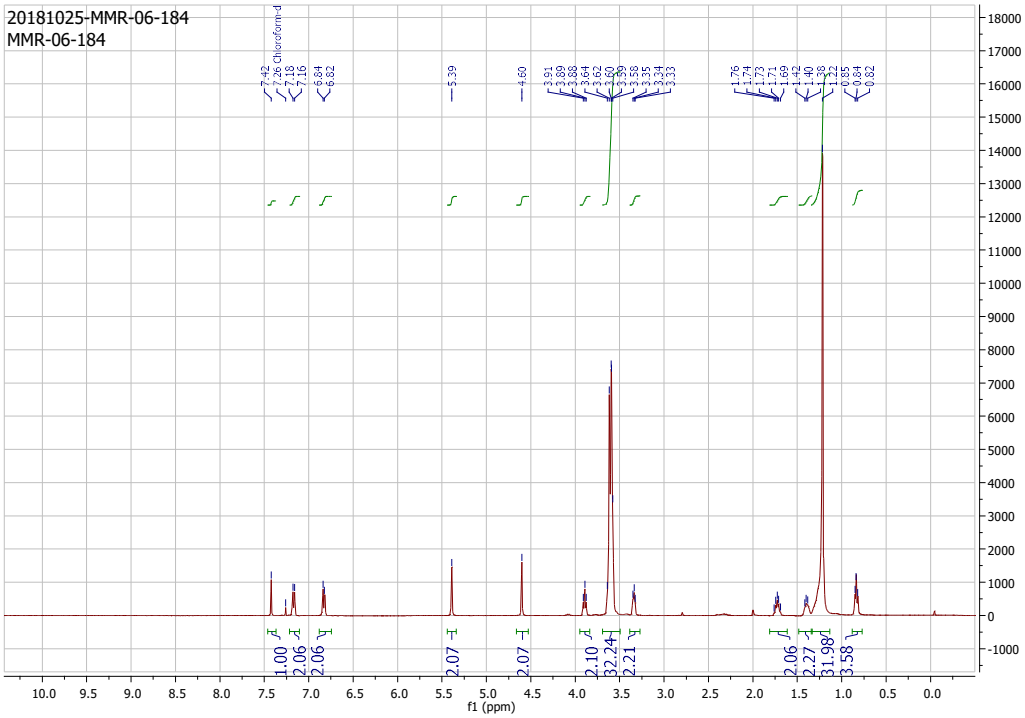


^{13}C NMR spectrum

Compound 47

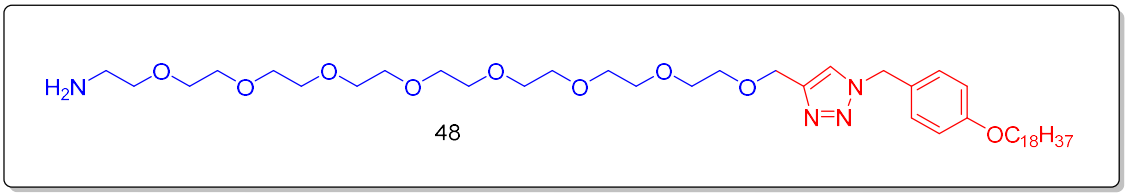


47

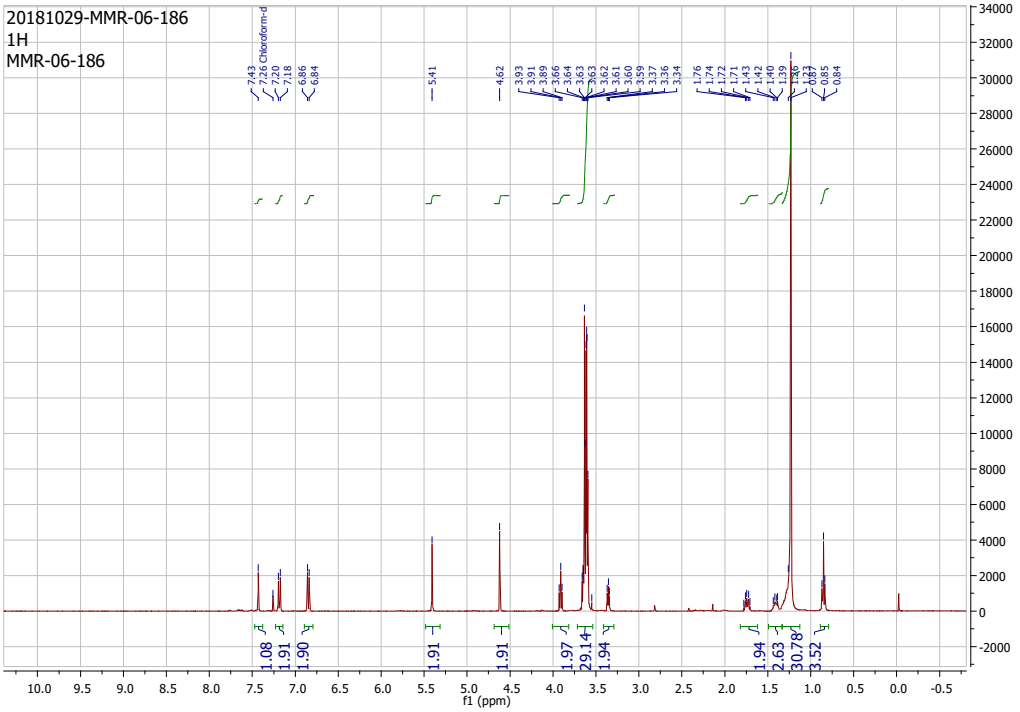


^{1}H NMR spectrum

Compound 48

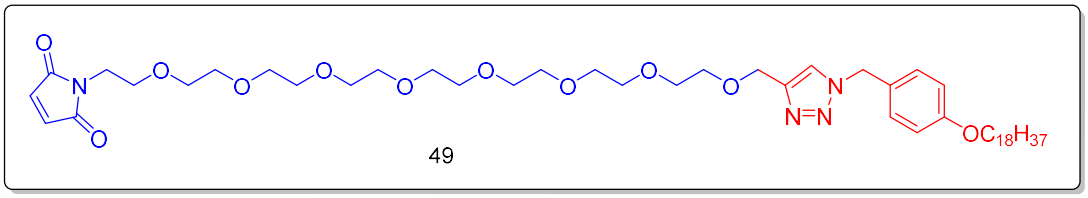


48

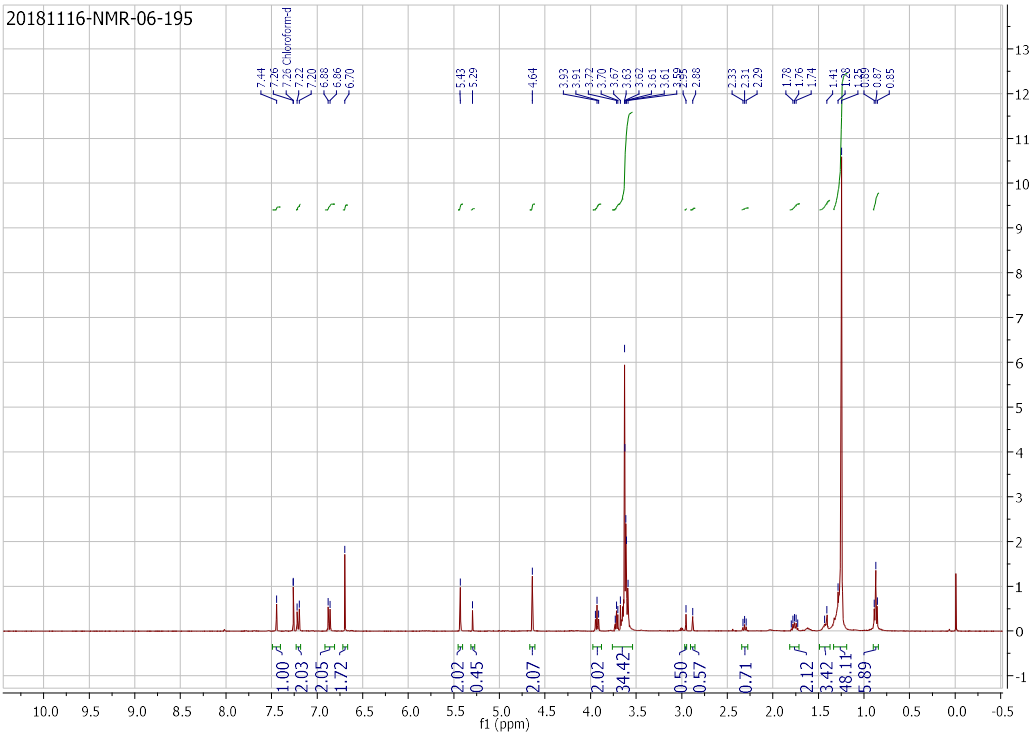


^{1}H NMR spectrum

Compound 49



49



^{1}H NMR spectrum