Engineering multimeric CTB_5 complexes for membrane cross-linking and fusion

Wenyue Dai

Submitted in accordance with requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Chemistry

March 2023

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Acknowledgments

Completing my PhD has been an invaluable life experience for me, and I wish to express my gratitude to everyone who supported me during this journey I am deeply grateful to the staff members of 1.49 lab for providing a peaceful laboratory environment that allowed me to conduct my scientific research. I extend special thanks to Charlotte Stevens and Alexandra Tamasa for their warm assistance in the biolab. Secondly, I would like to acknowledge my husband and colleague, Tomasz Kaminski, whose guidance and advice have been instrumental in advancing my projects and personal growth. Working alongside him, I have learned many invaluable qualities that make for a successful peer scientist. Furthermore, I am thankful to James Ross for inspiring me with his passion for computational biology. I am also grateful to my collaborators, Chunyue (Spring) Wang and Ralf Richter, for their unwavering support when I needed it most. During my secondment, Erik Kempmann and Francesca Rosato provided me with invaluable and time-consuming guidance. I also want to thank Andrew Booth for his intuitive advice on lipid research. I wish to express my appreciation to Winfried Roemer and Paul Beales for granting me the opportunity to study in their labs, which proved to be invaluable experiences. It was a pleasure working with Natalia Danielewicz, George Heath, and Maria Nikolova, and I am grateful for their direct contributions to my research.

Thanks to my family, especially to my brother Wenzhao Dai, who occasionally meets with me on the continent far from China and travel with me.

I wish to express my profound gratitude to my supervisors, Bruce Turnbull and Michael Webb, for their unwavering support throughout my PhD journey. Bruce has been an invaluable source of guidance, providing valuable ideas and expertise while going above and beyond to assist me in overcoming any challenges that I encountered during my research. His willingness to leverage his vast network of contacts to facilitate my collaborations was particularly noteworthy. Michael's insightful thoughts and clear communication style greatly aided me in navigating bottlenecks in my research. I appreciated his direct approach to research discussions, which allowed for efficient communication and problem-solving.

I am grateful for the Marie Skłodowska-Curie ITN scholarship, which provided me with a generous salary that allowed me to fully focus on my research without financial concerns. Additionally, I would like to express my appreciation to Sarah King for her assistance with the synBIOcarb ITN programme.

Abbreviations

AF	Alexa fluor
AFM	Atom force microscopy
AP	Antiparallel coiled coil
AUC	Sedimentation velocity analytical ultracentrifuge
BoNT	Botulinum neurotoxin
BSTG	BioScreening Technology Group
CBD	Carbohydrate binding domain
CD	Circular dichroism
CTA1	Cholera toxin A1 subunit
CTA2	Cholera toxin A2 subunit
CTB_5	Cholera toxin B pentame
CTx	Cholera toxin
DOL	Degree of labelling
DOPC	1,2-Dioleoyl-sn-glycero-3-phosphocholine
DOPE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine
DSC	Differential scanning calorimetry
ER	Endoplasmic reticulum
FC	Flow cytometry
FCS	Foetal calf serum
FPLC	Fast protein liquid chromatography
FRET	Förster resonance energy transfer
GFP	Green fluorescent protein
GIDs	Gastrointestinal diseases
GUV	Giant unilamellar vesicle
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HUS	Haemolytic ureamic syndrome
IPTG	Isopropyl β-D-1-thiogalactopyranoside

ITC	Isothermal titration calorimetry
Kd	Dissociation constant
KIH	Knob-into-holes
LB	Lysogeny broth
LTx	Heat-labile enterotoxins
LUV	Large Unilamellar vesicle
MBP	Maltose-binding protein
MS	Mass spectroscopy
MW	Molecular weight
NHS	N-hydroxysuccinimide
Ni-NTA	Nickel nitrilotriacetic acid
NSF	N-ethylmaleimide-sensitive factor
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDI	Protein disulfide isomerase
PEG	Polyethylene glycol
pI	Isoelectric point
PP	Parallel coiled coil
PT	Pertussis toxin
r.p.m	Revolutions per minute
r.t.	Room temperature
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SEC-MALS	Size exclusion chromatography – Multiple light scattering
SLB	Supported lipid bilayer

SNAP25	Synaptosomal-associated protein 25
SNARE	Soluble NSF attachment protein
ST	Shiga toxin
STEC	Shiga toxigenic *E.coli* strain
SUV	Small unilamellar vesicle
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
UV	Ultra-violet

Abstract

A comprehensive understanding of cell-cell interactions is critical for elucidating various cellular processes, including the complexities of membrane fusion. Natural proteins, such as SNARE proteins, mediate membrane fusion by forming coiled-coil bundles that bring two membranes into close proximity. Synthetic analogues inspired by the properties of natural proteins have been developed to mimic SNAREs and trigger membrane fusion. However, limited attention has been devoted to reengineering lectins, such as bacterial toxins for this purpose, despite their potential for developing drug delivery systems that can selectively target specific sugars expressed on cell surfaces. Previous researchers in the Turnbull and Webb labs have found that a streptavidin-$(CTA2B_5)_n$ construct, with two (or more) CTB_5 assembled together through streptavidin-biotin interaction, could successfully lead to membrane fusion between two giant unilamillar vesicles (GUVs). However, the factors that could explain the membrane fusion caused by this protein complex have not been fully understood.

In this project, I elaborated on the mechanisms of membrane fusion induced by the streptavidin-$(CTA2B_5)_n$ construct. Furthermore, to build a more controllable CTB_5 assembly, the coiled-coil motif was selected as the mediator to construct a CTB_5 dimer whose membrane behavior could be more directly related to the architecture of the protein complex. Last, I investigated whether antibody-like proteins (Affimers) selected against CTB_5 could also form a complex capable for membrane fusion. This approach could in the future be modified to use Affimers selected against lectins other than CTB_5.

Table of Contents

Chapter I Introduction

I.1 Overview

Nanoparticle drug delivery systems possess favorable pharmacokinetic properties that render them highly promising in the field of biomedicine. These properties include the ability to extend drug circulation time within the bloodstream, safeguard against drug degradation, and selectively accumulate within target tissues for enhanced drug uptake[1]. Lipid-based nanoparticles, such as liposomes, offer advantages such as ease of preparation, biocompatibility, ability to carry large payloads, ease of transport across membranes and extended half-life in blood. When creating nanoparticles, various biological and chemical components can also be added to improve vesicle characteristics like mobility and target specificity. These factors make lipid-based NPs the most prevalent class of nanomedicine drugs recognized by the FDA[2].

Despite the fact that the drug delivery system itself has been the subject of extensive research, the precise transport of the drug-loaded liposome to a particular site and the efficient release of the drug remains challenges. In cells, vesicle fusion can efficiently transport molecules across membranes, as in viral invasion and neuronal exocytosis. It would therefore be beneficial if we could mimic the biological membrane fusion phenomenon and develop liposome-based drug delivery systems that could target specific cell type, i.e., neural cell, and in which drug molecules are delivered via membrane fusion.

In nature, interactions between proteins and glycans are frequently used to achieve specific recognition[3, 4]. For instance, sialic acid-containing oligosaccharides are recognized by the receptor-binding domain of the spike protein on SARS-CoV-2, facilitating SARS-CoV-2 entry into cells[5]. The abnormal glycosylation of tumor cells is also a target for some selective therapies for cancer[6].

In this project, we have investigated whether membrane fusion can be achieved by using the cholera toxin – a lectin which interacts with GM1 gangliosides. GM1 is found on the surface of neurons[7], and its fucosyl form is aberrantly present on small-cell lung cancer cells[8]. This will aid in the creation of lectin-based targeted drug delivery systems.

I.2 Introduction to membrane fusion

I.2.1 Membrane fusion induced by natural proteins

Membrane fusion is crucial for many physiological activities including fertilization[9], synaptic transmission[10], vesicle trafficking and invasion of viruses[11]·[12]. Membrane fusion is not a spontaneous process and the energy barrier and hydration shell between two opposing bilayers needs to be overcome to achieve it. Natural fusion machinery contributes to membrane fusion by ensuring membrane intermediate stability or to help to overcome the energy barrier by protein-lipid interplay[13]. Several natural amphiphilic proteins are able to promote membrane fusion by inducing local disorder of phospholipids. Annexin, for example, can produce local disruption in the hydrophilic-hydrophobic boundary and also bridge membrane surfaces by forming disulfide-linked dimers[14]. Furthermore, some small molecules[15-17] and peptides[18], can also disorder membrane organization.

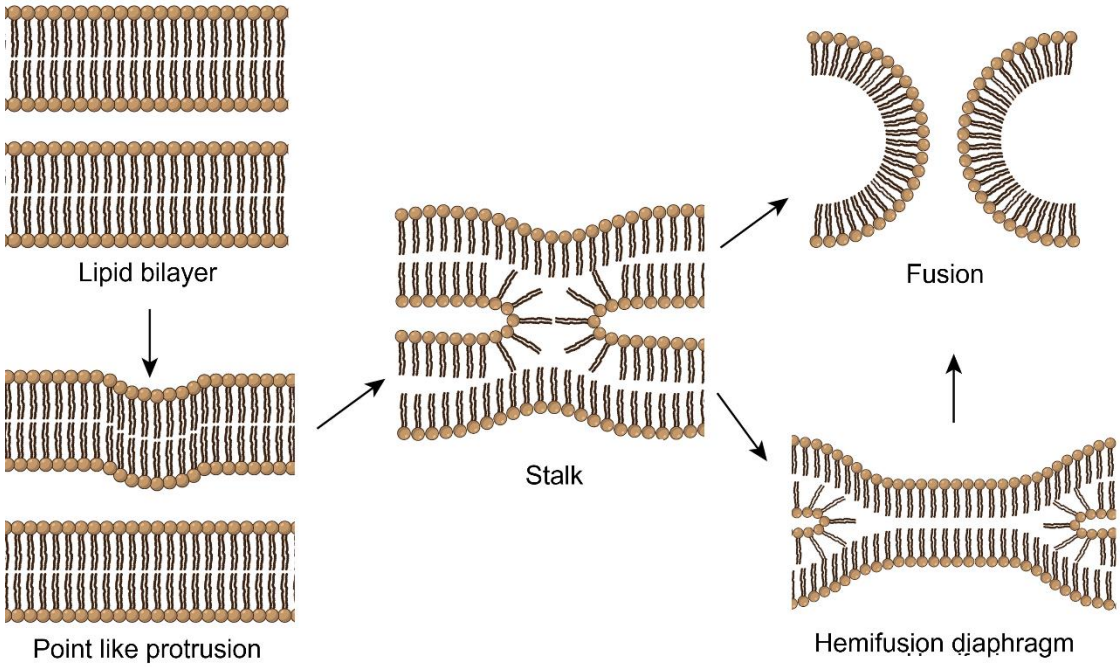


Figure I-1 According to the lipid stalk model, a schematic representation of the transitional states during membrane fusion. Figure based on Ref[19].

By far the most popular theory of membrane fusion is the stalk mechanism, which refers to the kinetic process of membrane fusion as stalk formation (Figure I-1)[20]. In accordance with this theory, the emergence of point-like protrusions trigger the start of membrane fusion because they make it easier for two nearby lipid bilayers to approach

one another by reducing the hydration repulsion between them. Next, the outer leaflets of the membranes merge, but the inner leaflets of the membranes do not, resulting in the formation of a stalk. This process involves the lipid membrane going from a lamellar to a hexagonal phase, which is a crucial prerequisite for stalk formation. Certain lipid molecules that exibit negative spontaneous curvature, such as DOPE (dioleoylphosphatidylethanolamine), facilitate the stalk formation and stabilization of this transition (Figure I-2). In the middle of the stalk, radical expansion eventually takes place, resulting in a phenomenon known as a hemifusion, in which the outer leaflets of the membranes fuse, but inner leaflets of the membrane form a new lipid bilayer. The two vesicles eventually combine into one as the hemifusion diaphragm experiences pore formation.

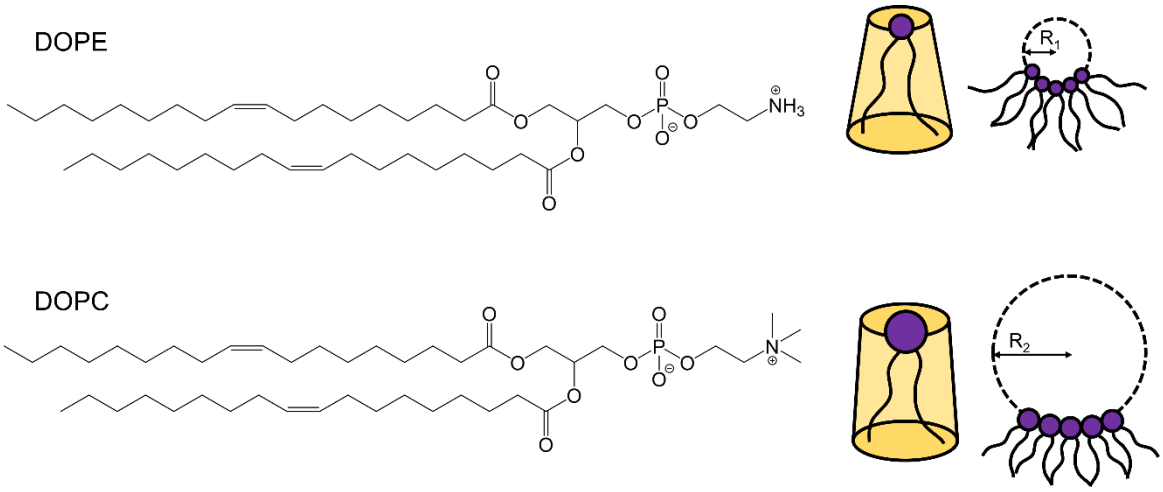


Figure I-2 The chemical composition of DOPE and DOPC (left) and the shape of the lipid molecule that contributes to the formation of intrinsic curvature (right). DOPE could favor negative curvature with a monolayer formed with higher negative curvature stress (R1 is smaller compared to R2 formed by DOPC).

Fusion of two vesicles can be mediated by natural proteins, among which SNAREs (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) are well characterized. SNAREs belong to a superfamily with more than 100 subtypes spread across animals, yeast and plants[21, 22]. These proteins play a critical role in facilitating vesicle-associated membrane fusion events during both exocytosis and endocytosis processes, which involve organelle compartmentalization[23]. SNARE proteins have a 60–70 amino acid cytosolic conserved motif, which forms into a coiled coil. The neuronal SNAREs core complex is composed of syntaxin-1 (Qa-SNARE), SNAP-25 (Qb and Qc-SNARE), and synaptobrevin (R-SNARE) in which SNAP-25 supplies two α-helices and the others each provide one helix for coiled coil bundle formation (Figure I-3). Synaptobrevin, contains three portions: C-terminal transmembrane domain (TMD) that

anchors into a synaptic vesicle, a short linker, and N-terminal SNARE motif which can form coiled coil with its partners. Syntaxin-1 is located in the neuronal terminal plasma membrane. Neuronal SNAP-25 lacks a transmembrane portion but replaces it with palmitoylation post-translational modification for membrane tethering[24, 25].

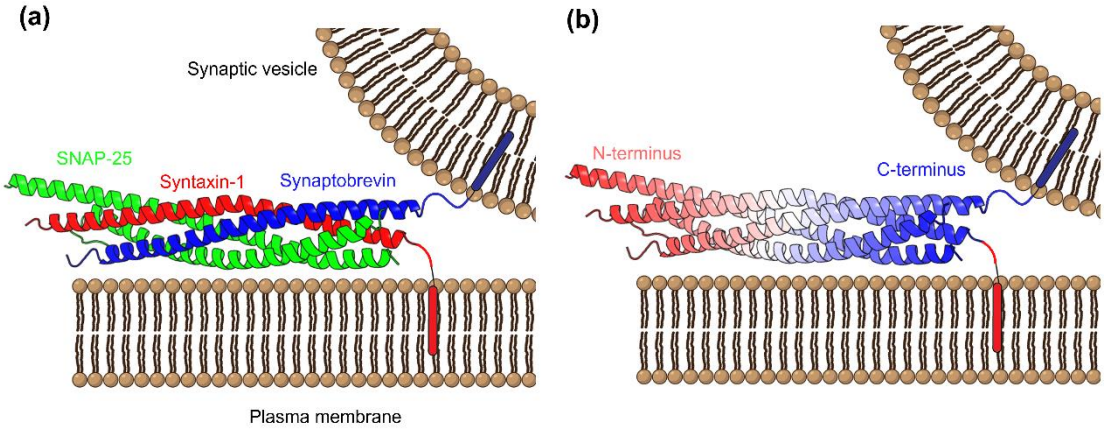


Figure I-3 Neuronal SNARE core complex (pdb: 1SFC). (a) Crystal structure of the neuronal SNARE core complex, which is composed of syntaxin-1 (red), SNAP-25 (green), synaptobrevidin (blue). (b) The orientation of the bundle is from N-terminal (red) to C-terminal (blue).

The isolated SNARE proteins are unstructured in solution[26]. However, under the regulation of Sec1/Munc18-like protein, the synaptobravin interacts with SNAP25 and syntaxin 1 and compose a four-helix bundle called a loose *trans*-SNARES core complex. Thus, the distance between vesicle and presynaptic membrane is decreased and thereafter a zipper-like interaction brings the proteins (and thus membranes) even closer together. The fusion stalk and hemifusion intermediate are then formed with water excluded between two membrane surfaces. Next, late regulatory proteins such as complexin and synaptotagmin can then participate in forming tight *trans*-SNARES complexes. Two alternatives might occur after this point: The vesicle might choose to “kiss and run” with a little neurotransmitter being transferred out of the presynaptic membrane or finish the full fusion process with complete content exchange.

Based on the knowledge of how SNARE proteins are assembled for membrane fusion, several simple model systems and SNARE mimics have been synthesized to help in understanding the mechanism of membrane fusion[27, 28].

I.2.2 Synthetic tools to understand membrane fusion

Many constructs that mimic the tetramer bundles of SNARE proteins have emerged to exploit the benefits coiled coil brings to vesicle fusion through zipper-like oligomerisation or recognition complexes. For example, molecular recognition between a boronic acid and a *cis*-diol together with a coiled coil was used to induce membrane fusion[29-31], with one such example shown in Figure I-4. The functional groups are linked with a pH-sensitive coiled coil which was anchored to the membrane through a stearic acid group. No fusion was observed under physiological pH because the peptides were unstructured and act as a spacer between pilot and target vesicles. However, when the pH dropped to 5, a coiled coil can be formed. Hence the distance between two vesicles was shortened and fusion occurred[29, 30]. These experiments emphasized the importance of distance control in membrane fusion.

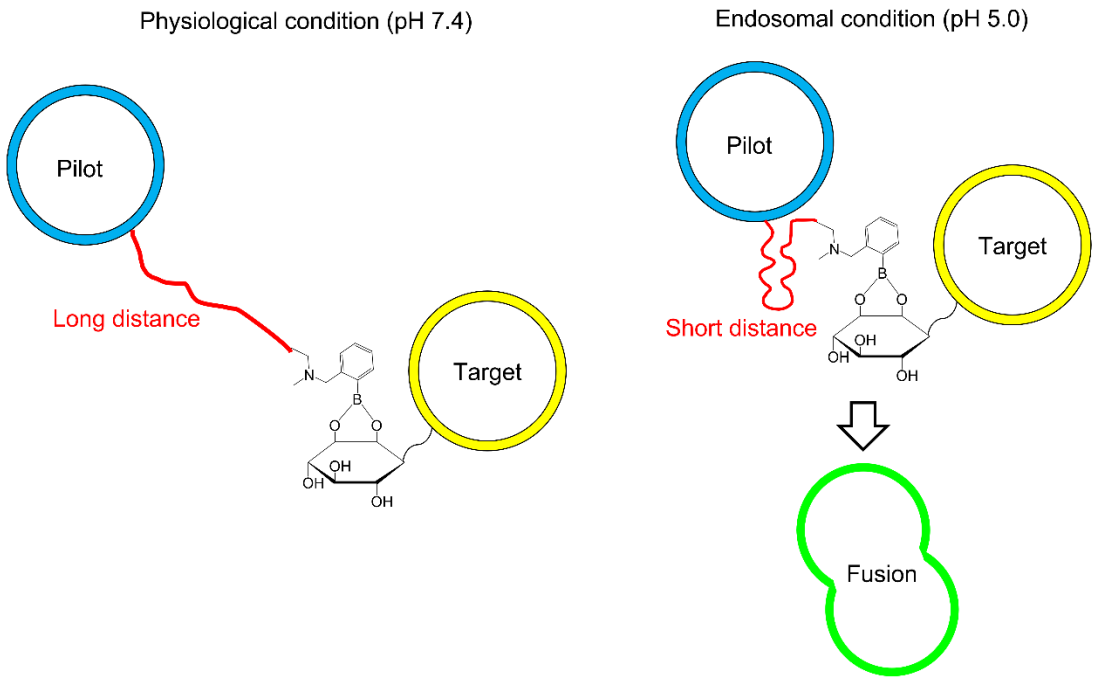


Figure I-4 Scheme of membrane fusion using the interaction between boronic acid and *cis*-diol, and the formation of coiled coil which is pH-sensitive. (Redrawn from Ref [30]) The peptide (red) are incorporated in pilot vesicles through stearic acid groups at the N-terminus, and there is a phenylboronic acid located at the CC-terminus of peptide to recognize the cyclic *cis*-diol group from a different vesicle. The peptide forms random coils at physiological pH, whereas it is rearranged to coiled coil under endosomal pH conditions and drag the pilot vesicle close to target liposome before inducing them to fuse.

DNA-lipid hybridization has also been popular for membrane fusion, in which lipidated DNA[32-34] was used to incorporate the DNA into soft materials. Höök *et al.* controlled

membrane fusion (Figure I-5) by using a target miRNA to interact with its partnered DNA hairpin, which further reveals the sticky end of the double stranded DNA (ds-**A**/**B**) incooperated in the pilot vesicle. The sticky end of one single strand from the pilot vesicle can interact with its complemented DNA strand from target vesicle, and during the displacement the two vesicles can be dragged in close proximity[35]. This study is based on their previous study on using DNA to mediate vesicle fusion on liposome[36, 37] and supported lipid bilayers[38].

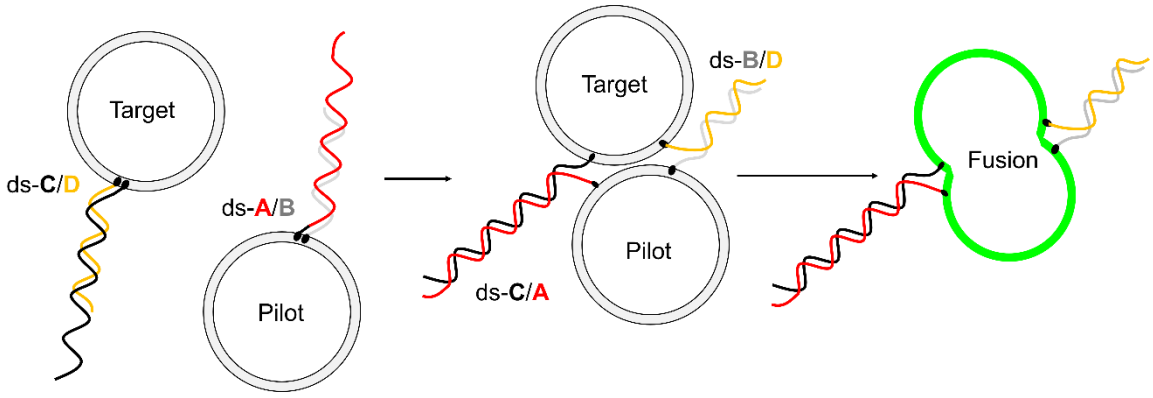


Figure I-5 An illustration of how miRNA can be used to speed up DNA-mediated membrane fusion (Redrawn from Ref [35]). The sticky end of the double strand (ds)-A/B can interact with the ds-C/D to bring two liposomes together and ultimately cause membrane fusion.

A heterodimeric coiled coil comprising the negative charged peptide E $(EIAALEK)_3$ (E_3) and complementary cationic peptide $K(KIAALKE)_3$ (K_3), was investigated as simplified SNARE core domains to achieve efficient and leakage-free fusion. Membrane fusion was observed by combining more than ten liposomes with diameter around 100 nm each into a single liposome with around 10 μm size[39]. Each E_3 and K_3 peptide was linked with a DOPE lipid anchor and a PEG_{12} spacer, forming two types of lipopeptides (*e.g.,* LPE and LPK) which are mutually compatible with each other and ready to be incorporated into vesicles (Figure I-6). The formation of coiled coils was able to bring the membranes closer and initiate membrane fusion, which has been demonstrated to be a promising drug delivery system for targeting living cells[39]. Atomistic molecular dynamics simulation showed that the LPE and LPK lipopeptides not only interact with each other by hydrophobic interaction by hydrophobic and ionic interactions, but also with membrane head groups[40]. The peptide-lipid interaction is especially emphasized in LPK, which could create positive local curvature and disrupt membrane order by interacting with a negatively-charged membrane under neutral conditions[41]. Furthermore, the efficiency of E_3 binding to K_3 was enhanced by increasing the spacer

length. For K_3, both PEG_4 and PEG_8 can assist in destabilizing the outer leaflet of the membrane while PEG_{12} would assist in the fusion process[42]. Apart from DOPE, Kros and co-workers also used cholesterol as the lipid anchor to build CPK_4 and CPE_4, where the coiled coil was elongated to four heptad units and more efficient membrane fusion was observed[43]. Then they did a similar experiment by changing the length of spacer[44] and lipid anchor position[45] to study different roles of K and E peptide in membrane fusion.

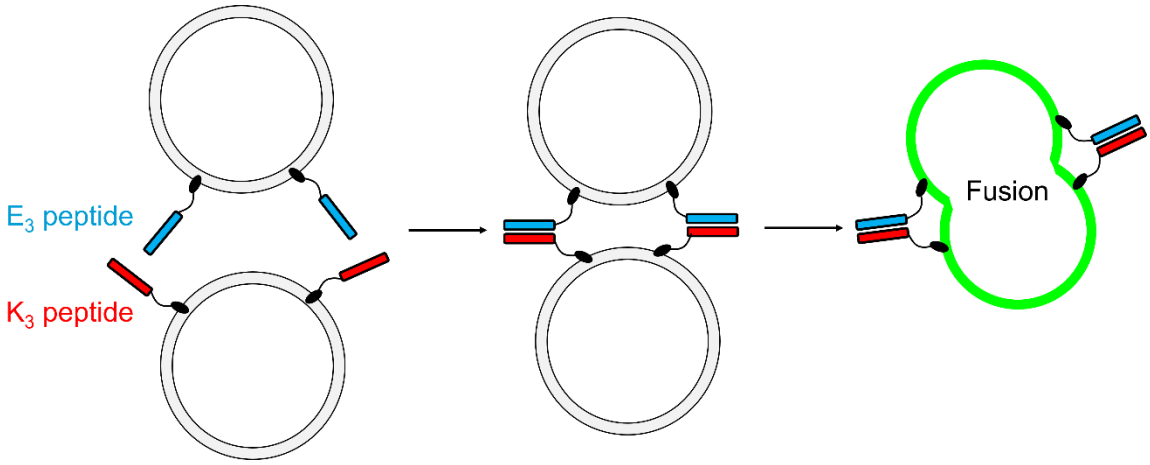


Figure I-6 Schematic representation of targeted membrane fusion mediated by coiled coil formation between E_3 and K_3 peptide. This image is redrawn from Ref[39]. The peptide can be fused with PEG linker at the C -terminus, followed by either DOPE or cholesterol as lipid component for incorporating in vesicles.

Inspired by how SNARE proteins associate by zipping, it was thought that the parallel orientation of two coiled coils is the essential condition to induce membrane fusion. The importance of coiled coil orientation on membrane fusion was also tested by other groups. Pahler *et al.* found that membrane fusion occurred in vesicles by forming a three-heptad parallel coiled coil[46]. Whereas for the antiparallel counterpart, there was only slow, but consistent, lipid mixing. The conclusion of this experiment is similar to that of the Simonsson group[38] who utilized DNA hybridization as a recognition motif. However, some SNARE mimics were found to lead to membrane fusion in both antiparallel non-zipper and parallel zipper arrangement[47].

In addition to directly mimicking SNARE core domains, some studies anchor β-peptide nucleic acids in the membrane using the transmembrane domain from synaptobrevin and syntaxin to achieve complete fusion[48, 49], which is an alternative way to apply SNARE mimics in membrane fusion. Small molecule recognition systems, such as the vancomycin/D-Ala-D-Ala motif[50] and the cyanuric acid/melamine interaction which

exploits hydrogen bonding[17], have also been used to bring two membranes into close proximity for fusion.

I.3 Introduction of AB_5 toxins

I.3.1 The AB_5 toxins

AB_5 toxins are notorious for causing various human diseases including gastrointestinal diseases with life-threatening symptoms. Different classes of AB_5 toxin possess distinct *modus operandi* to disrupt cellular machinery. There are four AB_5 categories according to their origin, sequence homology and biological activity[51]. The cholera toxin family, the members of which are cholera toxin (CTx) from *Vibrio cholerae* and heat-labile enterotoxins (LTx) from *Escherichia coli,* is malicious for inducing copious watery diarrhea. The shiga toxin family (STx) expressed from *Shigella dysenteriae* and Shiga toxigenic *E. coli* strain (STEC), the members of which are Shiga toxin 1 (Stx1) and Shiga toxin 2 (Stx2), is the culprit of haemolytic uremic syndrome (HUS). Pertussis toxin (Ptx) produced by *Bordetella pertussis* can bring about whooping cough. Furthermore, a novel AB_5 toxin named subtilase cytotoxin (SubAB) secreted from O113:H21 STEC strain was defined to be the reason for HUS in South Australia in the late 20th century[52].

Each AB_5 protein contains an A-subunit and five B proteins assembled in a pentameric ring (Figure I-7). The A-subunit of AB_5 toxins usually consists of two domains: A1 and A2 linked by disulfide bonds and non-covalent interactions. After the holotoxin is endocytosed by binding to their corresponding glycolipid receptor (Gb3 for STx and GM1 for CTx), the B-subunit enables retrograde trafficking to the Golgi and Endoplasmic reticulum (ER). The disulfide bond between A1 and A2 is presumed to be reduced by protein disulfide isomerase (PDI) once in the ER, and the A1 segment is therefore released from the rest of the toxin complex into the cytoplasm to proceed with its catalytic activities. For CTx and LTx, the A1 can be translocated into the cytosol through the Sec61p channel, , where it disrupts G-protein-coupled receptors by ADP-ribosylating them[53]. For the A-subunit in Stx, its RNA N-glycosidase activity digests an adenine base from 28S ribosomal RNA that hence blocks normal protein synthesis and finally leads to cell death[54]. STx was divided by two subclasses: Stx1 and Stx2, which share 55% identity by protein sequence alignment. Interestingly, Stx2 is prone to be more virulent by making people with more fragile microvascular endothelial cells more susceptible to HUS[52].

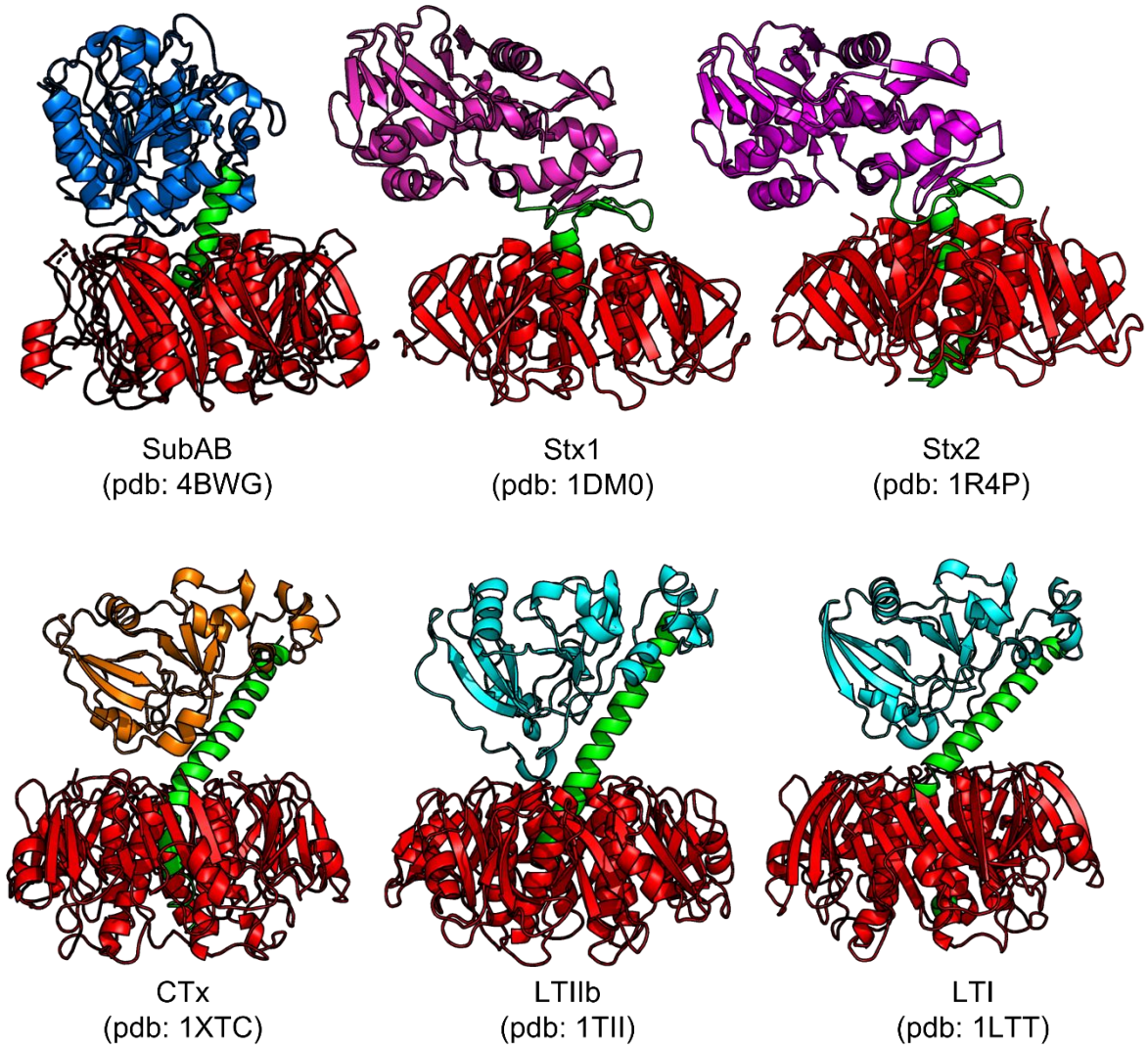


Figure I-7 Comparison of three characterized AB_5 toxin families. The B-subunit is shown in red cartoon. The A2 α-helix is in green. The A1 portion of different AB_5 family has discriminated activities – the SubAB family has subtilase activity (blue), Stx family has RNA N-glycosidase (purple) and Cholera toxin (orange) and LTx (cyan) have ADP-ribosyl transferase activity.

Varied in structure and function, the toxicity of different AB_5 toxins not only depends on the special catalytic activity exerted by their own catalytic A1 domain, but also relies on how the holotoxins, especially B-subunit, interact with glycans expressed on cell membranes and hence induce endocytosis and translocation intracellularly. The B-pentamer from Shiga toxin type 1 (Stx1) and cholera toxin (CTx), named $Stx1B_5$ and CTB_5 respectively, are the main topics of the sections that follow.

Both $Stx1B_5$ and CTB_5 can bind glycolipids in lipid microdomains before being transported into the endosome. The B-subunit is connected to the catalytic A1 domain

via the A2 subunit. In CTx this A2 peptide (CTA2) is a single continuous alpha-helix which threads through the pore in the middle of the CTB_5 pentamer. In contrast, the A2 domain in Stx1 (Stx1A2) is made up of an alpha-helix and a beta-sheet combined. Although the structure and stability of CTB_5 pentamers are not altered by the binding of CTA2[55], the process of CTB monomer assembling to form pentamer can be accelerated by a factor of three when CTA2 is present[56].

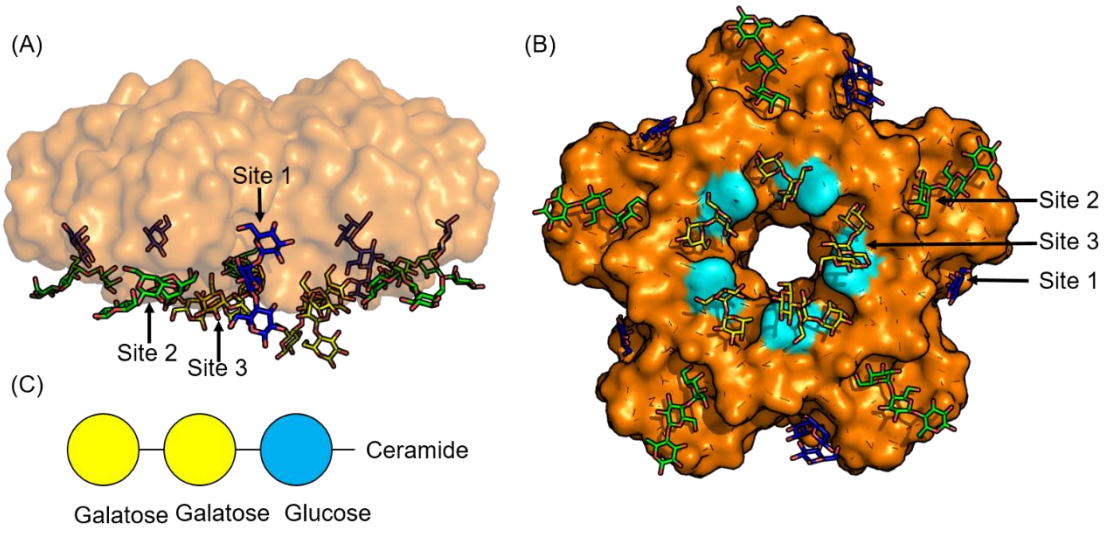


Figure I-8 (A) The side view of the crystal structure of Shiga toxin B subunit with its receptor Gb3 (pdb: 1bos). The receptors binding with three different binding sites are colored blue (site 1), green (site 2) and yellow (site 3) separately. (B) The bottom view of Shiga toxin B subunit with its receptor, with W34 colored cyan. (C) Symbolic representation of Shiga toxin receptor globotriaosylceramide (Gb3).

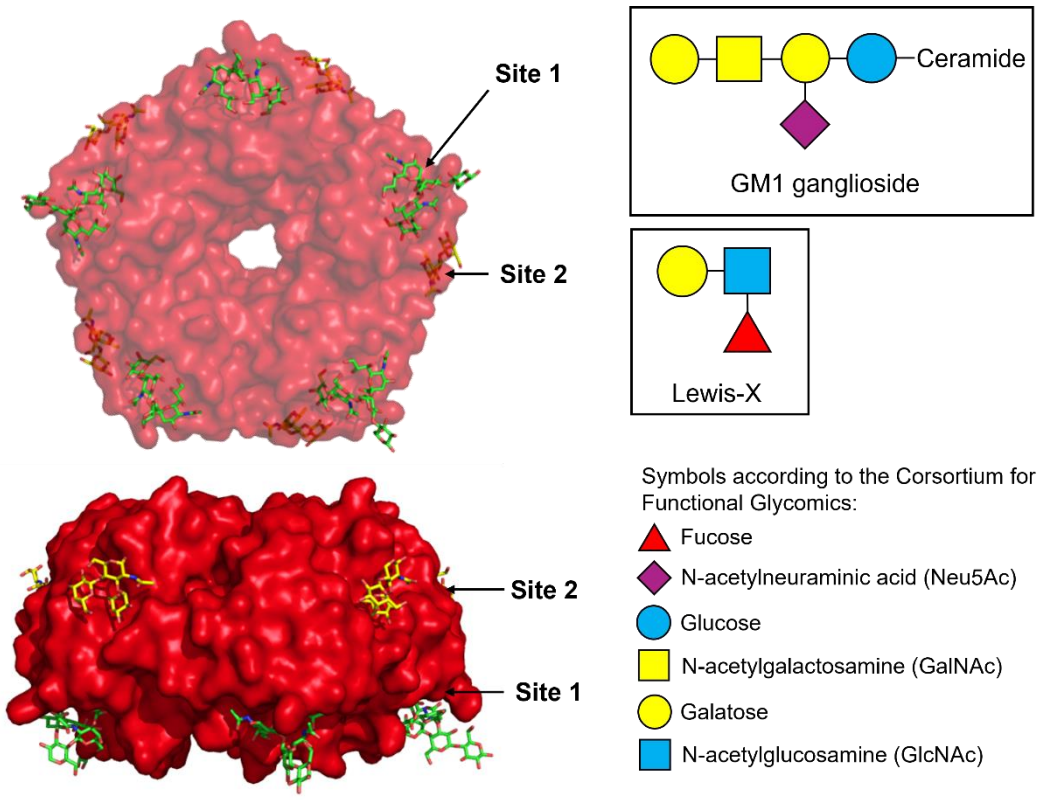


Figure I-9 The bottom view of crystal structure of Cholera toxin B subunit with its receptor GM1 pentasaccharide (green) and Lewis^x (yellow) (pdb: 2chb, 6hjd) binding in different sites, with symbolic representation of GM1 ganglioside and Lewis^x on the side.

The monomer from CTB_5 and $STx1B_5$ differ in size (10.6 kDa and 7.7 kDa respectively, Figure I-8 and Figure I-9), but both contain two three-stranded beta-sheets and one α-helix, while CTB_5 has an additional α-helix at its N-terminus. CTB_5 can bind the GM1 ganglioside headgroup at the bottom[57] and Lewis^x and Lewis^y at the side[58], whereas the $Stx1B_5$ has three binding sites per monomer with different affinity for glycosphingolipid Gb3[52].

I.3.2 Membrane curvature and cross-linking induced by AB_5 protein toxins

CTB_5 and $STxB_5$ subunits have multiple carbohydrate binding domains (CBD) facing in the same direction, hence they can exploit and cluster the glycosphingolipids to induce curvature on cell membrane and further achieve endocytic internalization. The topology of CTB_5 and $StxB_5$ was recently considered to be the main driving force for forming spontaneous curvature in an all-atom molecular dynamics simulation[59, 60]. CTB_5 and $StxB_5$ are similar in size, with radii of 3.6 nm and 3.0 nm respectively. CTB_5 has five binding sites for GM1 ganglioside on is lower face, with a single CTB-GM1 interaction having a 40 nM dissociation constant (K_d)[61]. In contrast, $StxB_5$ has three binding sites

for Gb3 per monomer (fifteen in total) which allows it to reach a strong avidity in the nanomolar range even though binding to single Gb3 receptor has a 1 mM $K_{d[62]}$. It is supposed that the interaction between toxin and membrane exerts an asymmetric pressure on the bilayer [59] and hence helps in deforming the membrane.

The shape of these toxins also assist in deforming the membrane. The bottom of CTB_5 has around 1.5 nm of protruding α-helix located around the central pore - which provide a downward pushing force. For $StxB_5$, the height difference between binding site I, II and binding site III (Figure I-8-a) to Gb3 provide reshaping force for membrane. The height difference between binding site III and binding site I in $StxB_5$ should also lead to the center of the toxin having a deeper insertion to the membrane in than at its edge[59, 60].

The arrangement of B-subunit binding sites is also important for formation of membrane curvature. A StxB-W34A mutant (Figure I-8-b), with the binding site III changed, had similar binding affinity in HeLa cell and giant unilamellar vesicle (GUV) experiments compared with original Stx, but with cytotoxicity decreased[63]. More importantly, StxB-W34A cannot lead to invagination anymore. Also, StxB-W34A could decrease the height difference between membrane and protein from 2.2 ± 0.4 nm to 1.6 ± 0.3 nm. All of these observations indicated that the binding site III is an additional element to recognize Gb3 after Stx is bound to membrane already and it plays a vital role in assisting membrane curvature by recruiting Gb3 and creating membrane curvature by the topology in the binding interface[64, 65].

Interestingly, one study demonstrated that a heterogenous CTB_5 complex capable of binding to only one GM1 ganglioside could still be endocytosed, but at far lower efficiency than wild-type CTB_5[66]. It is thus possible that multivalency and induction of membrane curvature are not strictly necessary for CTB_5 to function as a cell entry vehicle. Note that this phenomenon could also be explained by the increasing local concentration of weaker binding receptors (such as GM2, fucosyl-GM1 and GD1b) which can provide effective attachment to CTB_5[67].

After inducing membrane curvature, toxins cluster in lipid rafts[68], a dynamic microdomain on membranes accumulated with glycosphigolipids, cholesteol and receptors, which then form membrane tubules. The $StxB_5$ could cluster on cell membrane, exerting asymmetric compressive stress[59] on the outer leaflet of the bilayer, and causing membrane bending. The mechanisms behind it, such as lectin-induced local curvature from lipid chain length mismatch, are still under debate. Another

proposed driver of membrane curvature is line tension, present at the boundary between lipid rafts and the rest of the cell membrane – but this hypothesize is under questioned by computational simulation[59]. Instead, the author promoted that it is the membrane fluctuation-induced force that leads to lectin, at least Shiga toxin, clustering[59]. Also, in this paper, the authors questioned the conventional hypothesis of lectin clustering and claimed that line tension which is the interfacial energy between lipid domains formed by compositional mismatch, is not responsible for toxin clustering. Pezeshkian further tested practically the line tension hypothesis by using 30% Gb3 and 70% DOPC under low membrane tension and still found $StxB_5$ clustering and tubulation formed, which proves the inadequacy of the line tension theory[59, 69]. Moreover, he concluded that the successful clustering of $StxB_5$ is driven by mechanical force arising from tight coupling between toxin and membrane.

Furthermore, the length and saturation of glycosphingolipid[63, 70], and the rigidity of membrane[71] should also be considered as factors to affect membrane curvature and formation of membrane tubules. First, the composition of lipids in the membrane matters. The decrease in membrane rigidity by cholesterol and the asymmetrical distribution of cholesterol between outer and inner membrane layer were found to be crucial in $StxB_5$-induced invagination[72]. Similarly for cholera toxin, experiments showed that switching plasma membrane sphingomyelin to ceramide or deletion of cholesterol would attenuate the endocytosis process[73], which all demonstrate the significance of membrane composition on protein-lipid interplay. Secondly, the structure of the lipid receptor cannot be ignored. When Gb3 C22:1 with an unsaturated acyl chain was changed to lyso-Gb3 in which the fatty acyl chain was not present, or Gb3 22:0 with a saturated acyl chain, the tubules cannot be formed although the binding between GUVs and $StxB_5$ still works[63]. However, the saturation state of GM1 gangliosides exerts less effect on CTB_5 binding, as the radius of curvature formed by the interaction between GM1 and CTB_5 is still around 36 nm with both saturated and unsaturated lipids in all-atom dynamic simulations[60]. Therefore, while $StxB_5$ and CTB_5 use a similar mechanism to induce membrane curvature, the saturation state of the lipid receptor might have different influence on toxin clustering and membrane tubule formation for $StxB_5$ and CTB_5.

I.3.3 Changing the architecture of membrane-binding lectins to achieve new membrane modifying behaviour

Multivalency is an intrinsic property of most carbohydrate-binding proteins (lectins), including AB_5 toxins, and it plays an important role in achieving higher affinity to

glycosylated membrane receptors and changing membrane behavior by redistributing glycosphingolipids. Some studies have focused on modifying the number of binding sites in lectins and hence observed how it influenced membrane behavior. RSL (*Ralstonia solanacearum* lectin) is a trimeric lectin with a six bladed β-propeller structure and is able to induce membrane invagination in GUV experiments in a similar way to CTB_5[74]. Interestingly, it was found that some RSL mutants with only two carbohydrate-binding sites could also induce membrane invagination[75]. The topology of proteins can also determine membrane behaviour. Some lectins with binding sites that point in opposing directions can facilitate membrane crosslinking and form prototissues. For example, tetravalent fucose-binding LecB and LecA from *P. aeruginosa*, and VVL (*Vicia villosa* lectin)[76] are able to achieve crosslinking (Figure I-10-b) while $StxB_5$ (Figure I-10-a) and an antibody dimer cannot[74]. A chimeric lectin with bispecific binding ability called a 'Janus' protein was engineered by combining RSL (fucose binding domain) and CAM40 (sialic acid binding domain) connected by a flexible linker sequence GGGGSGGGGS[77]. This construct was proven to be able to crosslink heterogenous GUV vesicles to form tissue-like structures[78] (Figure I-10-c).

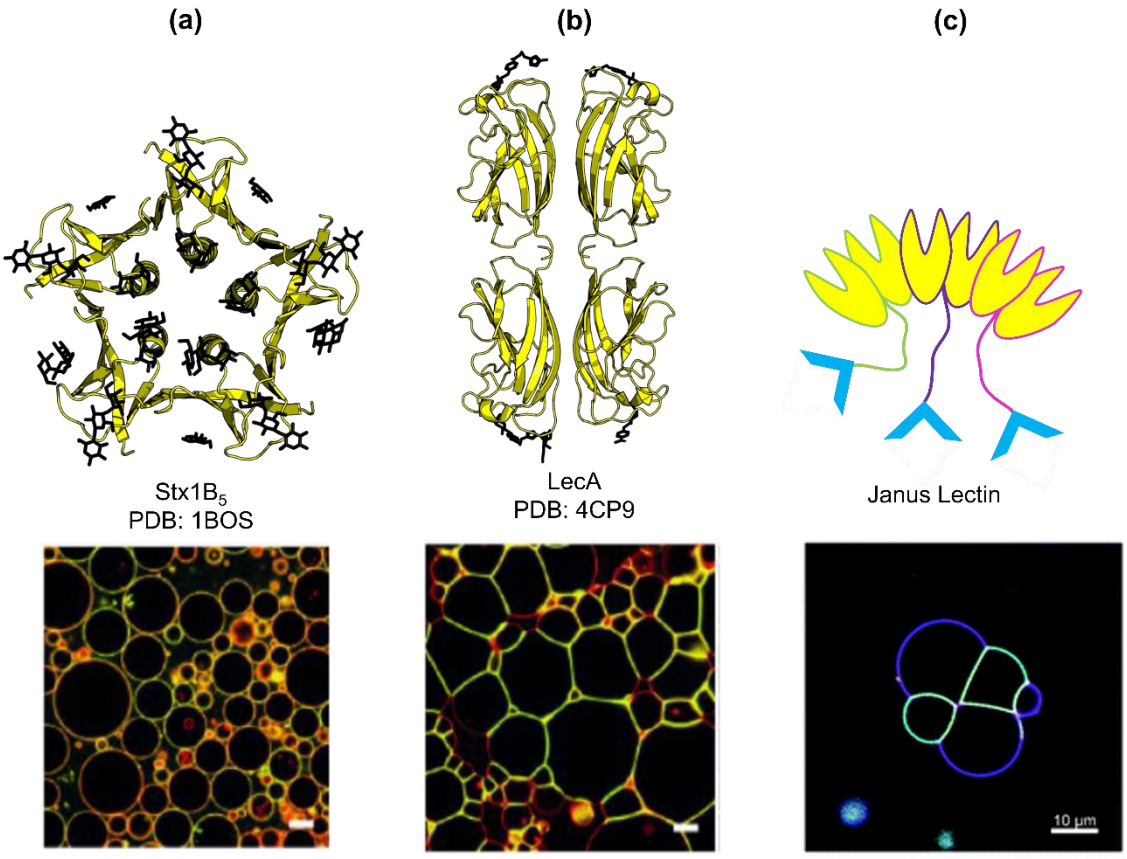


Figure I-10. Different carbohydrate binding sites distribution leads to different membrane behaviors, with their corresponding GUV behavior cited from Ref[76, 77, 79]. (a) Shiga toxin type 1 B-subunit ($Stx1B_5$), with all 15 Gb3 binding sites arranged on one side of protein, leads to homogenous binding to GUV coated with Gb3. (b) LecA with two Gb3 binding sites on each side of the rigid body, contributes to heavy crosslinking of Gb3-GUVs. (c) Janus lectin (image redrawn from Ref[77]) which combines RSL (fucose binding domain) and CAM40 (sialic acid binding domain) through flexible glycine-rich linker, mediated GUV coated with different carbohydrate crosslinking.

Previous researchers in the Turnbull lab (Daniel Williamson, University of Leeds) and Römer lab (Sarah Wehrum, University of Freiburg) were pioneering in identifying that a simple change in CTB_5 arrangement could lead to different membrane behavior by building a multimeric CTB_5 assembly mediated by streptavidin-biotin interactions[80] (Figure I-11).

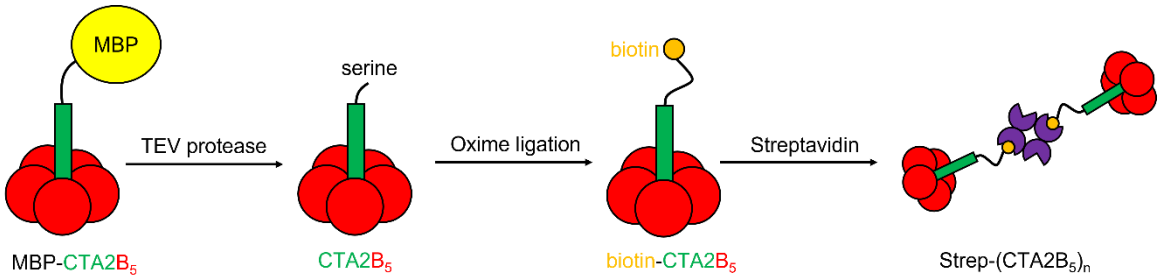


Figure I-11 Construction of a strep-$(CTA2B_5)_n$ complex. The $CTA2B_5$ protein was expressed as a MBP fusion which was later removed by TEV protease. The N-terminal serine was exposed to be oxidized to an aldehyde. The alkoxyamine-PEG_4-biotin group, as a carbonyl-reactive cross-linker, reacted with the aldehyde to attach the biotin group to form biotin-$CTA2B_5$ which was then mixed with commercial streptavidin (labelled with Alexa Fluor 488) to build strep-AF 488-$(CTA2B_5)_n$ construct.

In this strep-$(CTA2B_5)_n$ construct, multiple CTB_5 proteins were assembled into one protein complex through streptavidin-biotin interactions, and this protein assembly was used to interact with GM1-functionalized giant unilamellar vesicles (GUV). In addition to cross-linking vesicles, as seen previously for the Janus lectin, the strep-$(CTA2B_5)_n$ construct was also found to enable hemi-fusion and fusion of GUVs (Figure I-12). Figure I-12-a shows transfer of fluorescent lipid (red) from a labelled to an unlabelled GUV indicating hemifusion occurred. This was followed by exclusion of the protein complex (green) at the interface between two vesicles (Figure I-12-b), which eventually lead to full fusion (Figure I-12-d)[80].

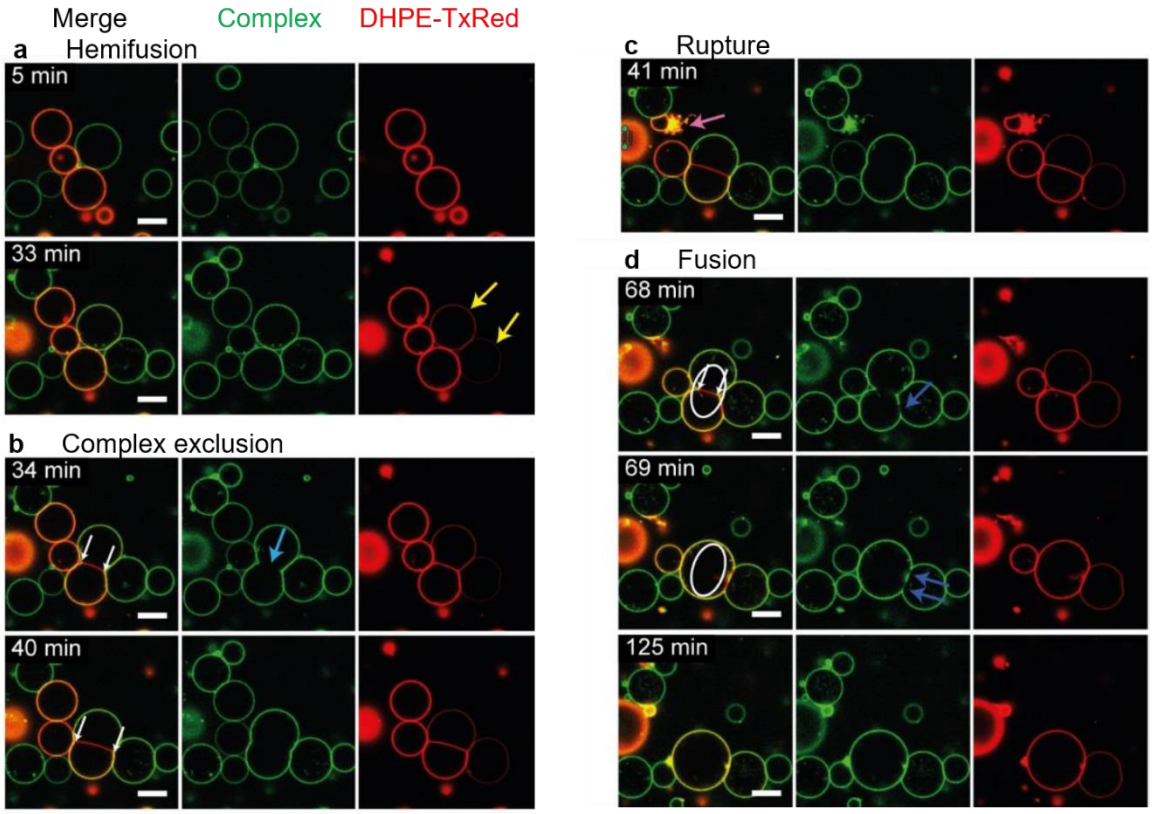


Figure I-12 The hemifusion, complex, rupture and fusion process induced by strep-$(CTA2B_5)_n$ (fluorescently labelled with Alexa Fluo 488). The vesicles contained 2.5 mol% GM1 and either have 0.5 mol% DHPE-TxRed (red) dye or not. After treating the vesicles with strep-$(CTA2B_5)_n$, different vesicle behaviors were observed. (Figure reproduced from Ref[80] under a Creative Commons CCBY4.0 licence[1]. (a+b) Hemifusion occurred between labelled lipids and unlabeled vesicles (yellow arrows), which resulted in an interface (b, white arrow) in which protein complexes were excluded (blue arrow). (c) During this process, rupture was observed as well. (d) Finally, two vesicles were fused into one (white circle), and some protein complex excluded from interface was seen (blue arrow).

Even though both the Janus protein and the strep-$(CTA2B_5)_n$ construct contain multiple carbohydrate biding domains connected by flexible linkers, they ultimately result in different membrane phenomena. This means that we must use more controlled methods to investigate the circumstances necessary for membrane fusion. Even proteins with rigid structures are not always capable of fusing with membranes. LecA, for instance, has the opposing Gb3 binding site, but it can only cause heavy membrane crosslinking, not membrane fusion. This implies that, either alone or in combination, endogenous membrane curvature, linker flexibility, distance, *etc.* brought about by the re-engineered lectin may be involved in membrane fusion.

[1] http://creativecommons.org/licenses/by/4.0/

I.4 Tools to study membrane fusion

I.4.1 Liposomes as models for studying membrane fusion

Unilamellar vesicles of various sizes, such as gaint unilamellar vesicles (GUV, diameter >1 µm), large unilemellar vesicles (LUV, 100 nm – 1 µm), and small unilamellar vesicles (SUV, diameter < 100 nm), can be exploited in membrane fusion studies[81]. The most direct effect of the different vesicle sizes is to produce different membrane curvatures, which is one factor important for inducing membrane fusion[12, 82]. The available techniques also influence the choice of different vesicle sizes; GUV is frequently combined with confocal microscopy because its size makes it easier to observe directly. However, preparation of GUVs is relatively difficult, as neither the commonly used hydration nor electroformation methods guarantees completely unilamellar and uniformly sized vesicles. LUVs and SUVs, on the other hand, can be used in plate assays to measure the fusion of vesicle populations quantitatively or in fluorescence correlation spectroscopy to study protein binding to phospholipid vesicles[83], hence the information such as fusion efficiency and protein-liposome interaction could be obtained. However, one drawback of LUVs and SUVs is that it is difficult to observe vesicle fusion in real time due to the limitation of the resolution of microscopy in order to learn more about the mechanism of membrane fusion; this challenge can be conquered by combining LUV/SUV studies with GUVs studied under confocal microscopy[84, 85]. Other membrane-mimic platforms, like supported lipid bilayers (SLB) or lipid monolayers (LMs), have the advantage of being compatible with Colloidal Probe Microscopy (CPM)[86], Quartz Crystal Microbalance with Dissipation monitoring (QCM-D)[87], Atom Force Microscopy (AFM) or Electrochemical Impedance Spectroscopy (EIS)[88] to assess protein-membrane interactions. The limitation is that an accurate representation of how proteins interact with curved vesicles cannot be achieved due to an insufficient curvature and high membrane rigidity of plenary lipid layers. Finally, lipid nanodisks could also been employed. An experiment demonstrating membrane fusion using SNARE mimic proteins showed that lipid nanodiscs provide a more efficient membrane fusion than using LUVs[89, 90]. Nevertheless, the drawback of this system is that it requires some amphipathic proteins, such as membrane scaffold proteins (MSPs)[91], to wrap the lipid membrane, which makes it more challenging to prepare comparing to a pure lipid-composed system.

I.4.2 Techniques to study membrane fusion

Studies on the process of membrane fusion can be roughly divided into two categories: one involves developing a kinetic model of the fusogen by analyzing the general behavior of the vesicle population. The alternative method involves analyzing the behavior of individual vesicles in order to deduce the precise mechanism of membrane fusion that takes place. Lipid mixing, content mixing, and content leakage assays based on fluorescence are popular for the former, and these will be discussed in the Chapters that follow, where they will be used. The latter will rely more on the higher precision of fluorescence microscopy for the study of individual vesicles. Confocal microscope, for instance, can used to observe the membrane behavior of GUVs. However, confocal microscopy detection is difficult for some studies involving the physiological membrane fusion phenomena, such as the fusion of synaptic vesicles and plasma membranes, as the signal on the cell surface is easily obscured by fluorescence from the main body of the cell. Here, total internal reflection fluorescence (TIRF) microscopy can be applied, which has the benefit of selectively activating fluorophores attached close to the cell surface while suppressing fluorescence from intracellular regions[92].

I.5 Project Objectives

The main objectives of the project as a whole were: to assemble multiple fusogenic CTB_5 proteins in various ways; to investigate the phenomenon of membrane fusion from the resulting protein complexes; and to determine which characteristics of protein architecture facilitate membrane fusion more efficiently.

To start off, we looked more closely at the mechanism by which the streptavidin-$(CTA2B_5)_n$ complex directs membrane fusion, which was inspired by the membrane fusion experiments described in section I-3.3. In the original study[80], the researchers attempted to combine streptavidin and biotin-$CTA2B_5$ into a complexes in a 1-to-2 ratio, as they believed that to be the optimal fusogenic species. This assertion was further examined in my project. Furthermore, instead of using the commercial streptavidin tetramer, I made use of a series of defined streptavidins which had one, two, three, or four alive monomers per tetramer[93]. By mixing such streptavidin species with biotinylated $CTA2B_5$, the different species of streptavidin-$(CTA2B_5)_n$ complexes were obtained with a more controllable number and orientation of CTB_5 units. In turn, this allowed investigation of whether the number of CTB_5 per streptavidin and the binding

orientation had an impact on the membrane fusion effect. This work was described in Chapter II.

Furthermore, the use of a coiled coil as a linker presents an attractive motif to assemble multiple $CTA2B_5$. The coiled coil is a well characterized protein structure, and its conformation is stabilized by the interaction between hydrophobic interactions and salt bridges between α-helices. The length, orientation, dissociation energy and oligomerization could be controlled by rational arrangement of amino acids. Therefore, it was served as an oligomerization unit in my project to combine two cholera toxins in a non-covalent way to achieve controllable lectin complexes.

Last but not least, I used non-antibody scaffold proteins called Affimers[94] as the mediator to link two CTB_5 proteins together. An Affimer is a small protein (~14 kDa) with two hypervariable loops that can be generated via phage display against target proteins. Matthew Balmforth in Turnbull's lab has screened out several Affimers targeting CTB_5 with different binding affinity[95]. In this project, I used Matthew's screened Affimer to assemble a CTB_5 dimer complex, and the obtained super-lectin was used for studying its possibility for membrane fusion. I intended to lay the groundwork for using Affimers to combine different lectins together as in future project, which will be useful for fusing vesicles coated with various types of carbohydrates.

Chapter II Structure-activity relationships of defined streptavidin-$(CTA2B_5)_n$ complexes

As described in section I.3.3, Daniel Willamson (University of Leeds) used chemical ligation to biotinylate $CTA2B_5$ and exploited the streptavidin-biotin interaction to build protein complexes. SEC-MALS and ITC results revealed the formation of a multimeric $CTA2B_5$ complex, and the GUV experiments indicated the ability of such a protein complex to induce membrane fusion. However, there are a couple caveats to this study: Firstly, it was unclear exactly how many $CTA2B_5$ participated in each complex, and how heterogeneous the population of protein complexes was. Although the ITC result indicated an "average" of 2-3 $CTA2B_5$ binding per streptavidin tetramer, the membrane fusion observed from the GUV experiment induced by such a product might only be driven by certain species in the mixture. Interpreting the mechanism behind membrane fusion can be challenging without quantifying and understanding the composition of the protein mixture.

Moreover, since only qualitative GUV experiments were performed, we lacked the solid quantitative data, which limited the conclusions we can draw. To determine which component from the streptavidin-$(CTA2B_5)_n$ complex is responsible for membrane fusion, we need to separate these species and study their membrane behaviour individually. Hence, in this chapter, different streptavidins, possessing either 1, 2 or 3 active binding pockets per streptavidin tetramer, were expressed and purified (Figure II-1). These subspecies of streptavidin are essential for constructing different streptavidin-$(CTA2B_5)_n$ complexes for subsequent tests (Figure II-2). Instead of using ITC or SEC-MALS to characterize binding stoichiometry, mass photometry was exploited to determine the protein composition in a higher resolution. The extraordinary high affinity of biotin to streptavidin means dissociation of the biotin-$CTA2B_5$ complex is negligible under the working concentration of mass photometry (<200 nM). The difference between the mass of complexes with one or two of the biotin-$CTA2B_5$ complexes bound is 65 kDa, resulting in a suitable resolution to distinguish the composition of the complexes.

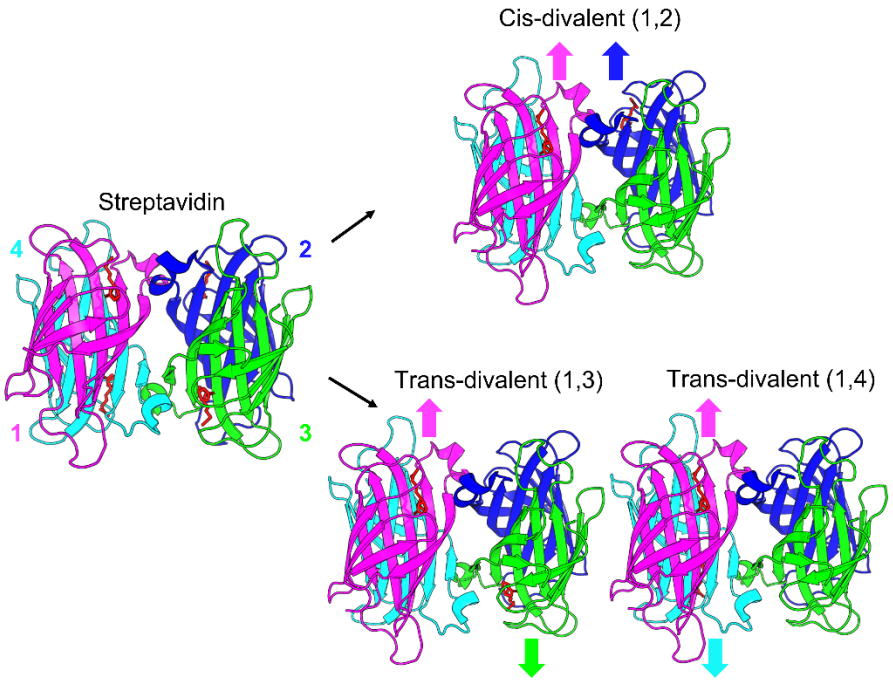


Figure II-1 The diagram illustrates the differentiation between cis- and trans- streptavidins. Streptavidin (PDB: 1LCV), which comprises four active binding pockets, interacts with norbiotin (depicted in orange as sticks). The cis and trans orientation of streptavidin enables norbiotin to be allocated in different directions.

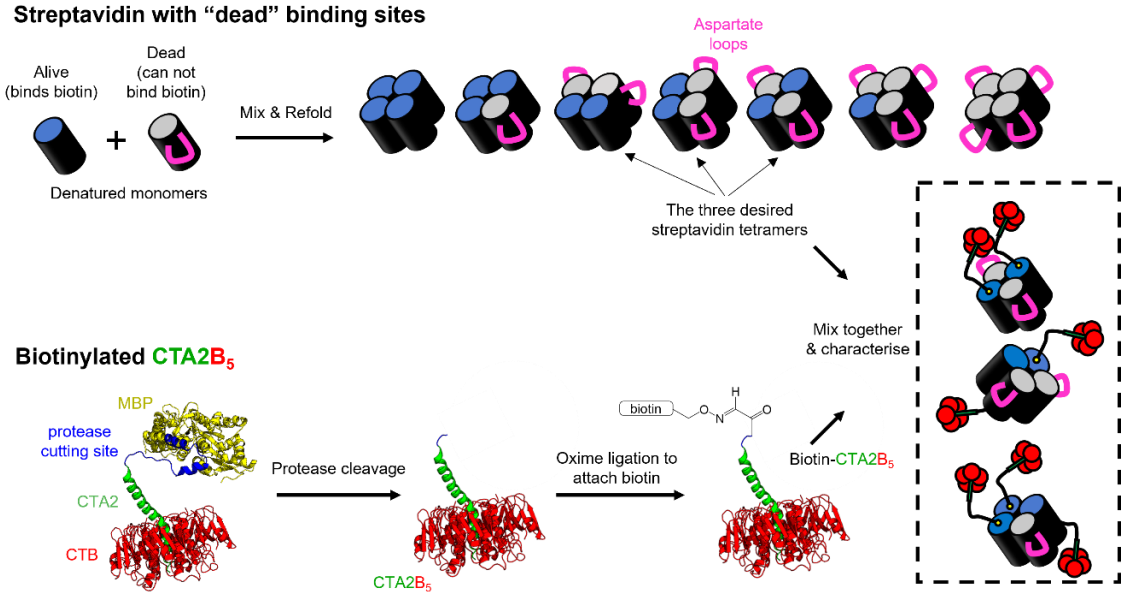


Figure II-2 An illustration of how biotinylated $CTA2B_5$ and defined streptavidin were prepared. The main interest is in the *cis*-, *trans*-, and *trivalent*-streptavidins, which would be combined with biotinylated $CTA2B_5$ to produce corresponding streptavidin-$(CTA2B_5)_n$.

II.1 Assembly of defined Streptavidin with biotinylated CTA2B$_5$

The following chapter will demonstrate how to make streptavidin-(CTA2B$_5$)$_n$ with a defined number and orientation of CTB$_5$.

II.1.1 Preparing biotin-CTA2B$_5$

Initially MBP-CTA2B5 was expressed by me using the plasmid[80] previously engineered by James Ross (Plasmid sequence in Appendix VIII.1.1a), and the details about the expression and purification of MBP-CTA2B$_5$, TEV protease digestion and oxime ligation can be found in the Method & Material sections VII.4.3, VII.5.1, and VII.6.1 respectively. Briefly, the protein purification was conducted using consecutive amylose and nickel resin chromatography (Figure II-3).

In parallel with my project, Natalia Danielewicz (enGenes Biotech, Austria) worked on optimizing the expression of MBP-CTA2B5 and provided purified protein for some of the experiments including GUV experiment (section II.2) and Lipid mixing assay (section II.3). The provided MBP-CTA2B$_5$ was first expressed from a customized plasmid[96] in a *E. coli* V1[97] strain originated from BL21(DE3) and exported to extracellular media. By re-engineering the plasmid provided by our lab[80] which could express the MBP-CTA2B$_5$ protein, Natalia increased the yield of this protein by 11-fold (168 mg/L vs. 15 mg/L). In order to confirm that the expressed protein was correctly folded, Dr. Iain Manfield used ITC to check whether the binding affinity between MBP-CTA2B$_5$ and GM1 expressed from different plasmids and cell strains varied, and I perform the ITC analysis (Appendix VIII.4). I then biotinylated CTA2B$_5$ protein by cleaving off the MBP domain with TEV protease, exposing an N-terminal serine residue on the CTA2B$_5$ in the process. This residue is further oxidized with periodate, and derivitised with the biotin-PEG$_4$ group via oxime ligation (Figure II-4).

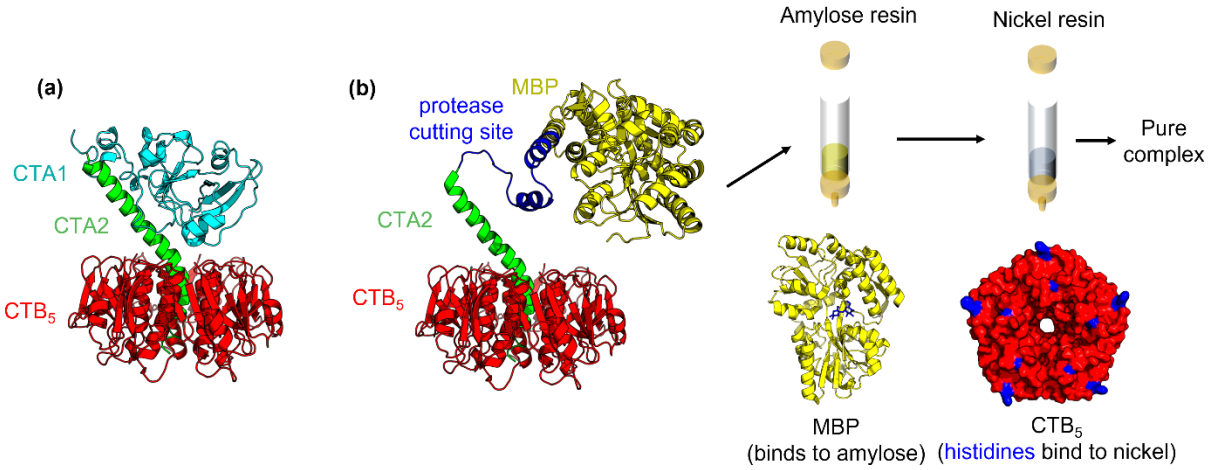


Figure II-3 Scheme representing the structure of MBP-CTA2B_5 and the purification process. (a) Wild-type Cholera toxin holotoxin (PDB: 1xtc). The toxic CTA1 subunit and the CTA2 peptide linker, whose C-terminus enters the cavity of the CTB_5 subunit, are joined by a disulfide bond. The catalytic CTA1 subunit is switched out for an MBP-tag in order to allow the complete CTA2B_5 complex to be purified using Amylose resin (b). For the second stage of purification, CTB_5 can be further captured by nickel resins using native accumulated histidines. A flexible linker with Factor Xa and TEV protease cutting sites is present between MBP tag and CTA2 helix. Following purification, the MBP tag can be removed with TEV protease, exposing the N-terminal serine residue for the subsequent oxime ligation process.

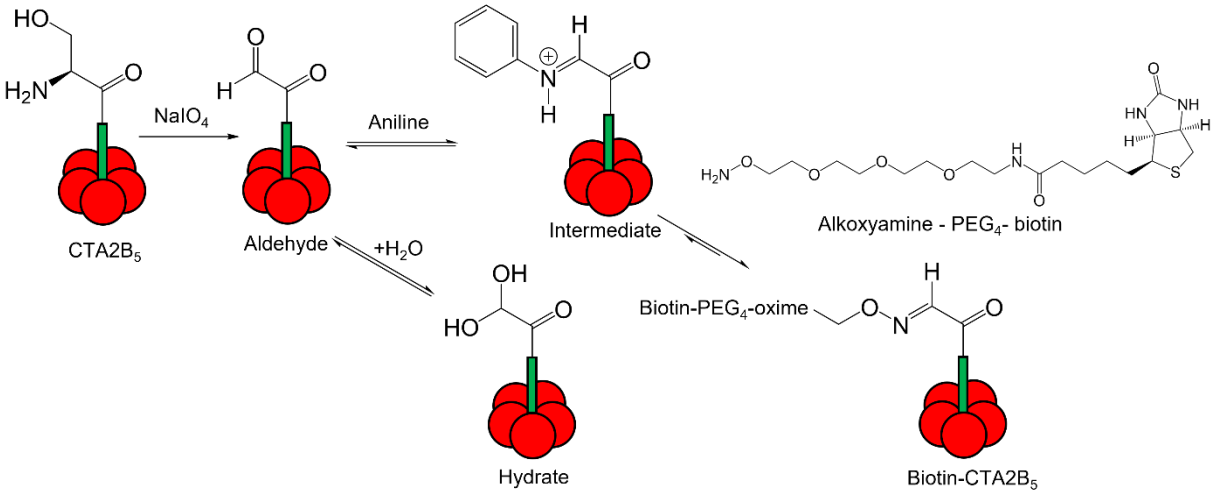
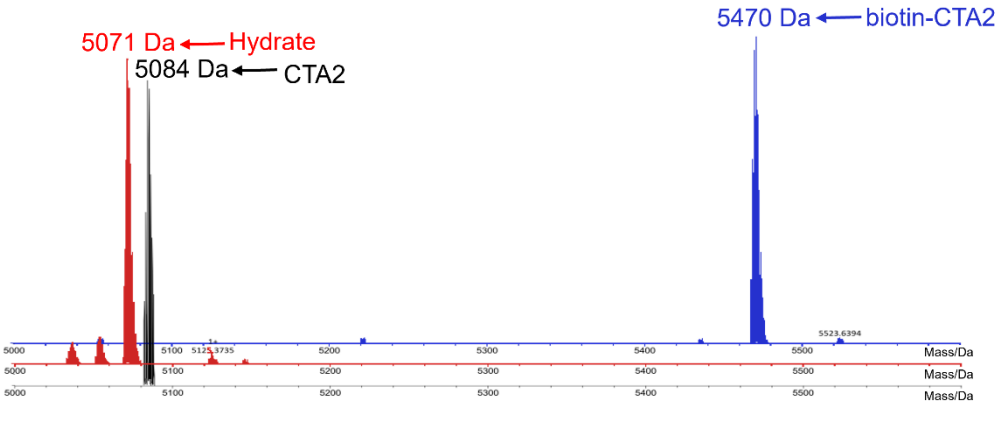


Figure II-4 The scheme of oxime ligation on the N-terminus of CTA2 peptide. The terminal serine residue was oxidised by sodium periodate into aldehyde, which reversibly forms a hydrate (diol) with the participation of water. The aldehyde further interacts with aniline and forms iminium ion as an intermediate, which reacts with the biotin aminooxy group in the next step to add the biotin group on CTA2B_5.



m

Figure II-5 The ES-MS spectrum of oxime ligation on CTA2 peptide.

The process was monitored by electrospray mass spectrometry (Figure II-5), and the molecular weight of CTA2 shifted from the original unmodified 5084 Da to 5071 Da after the N-terminal serine was oxidised and finally reached 5470 Da when the biotin-PEG_4 group was attached. During the oxime ligation, a side product might develop when the sodium periodate concentration is too low or after long reaction times (Figure II-6). To prevent this, it is necessary to use a concentration of at least 40 µM CTA2 and limit the reaction time to 5-10 min.

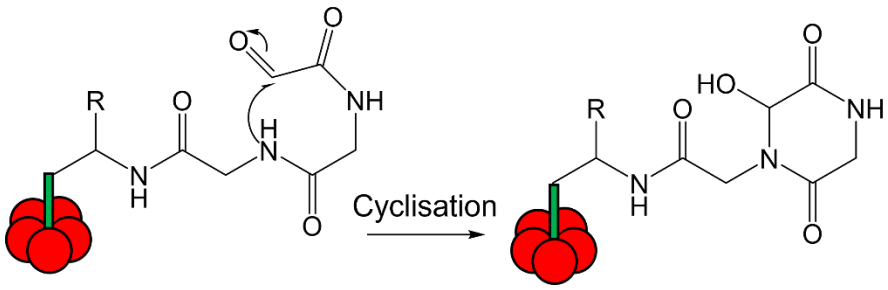


Figure II-6 The amide nitrogen in the glycine residue reacts as a nucleophile with the aldehyde group, which results in an irreversible side product.

After biotinylation was verified, the processed protein was exchanged to HEPES buffer using a PD-10 desalting column. The biotin-$CTA2B_5$ complex needed to be fluorescently labelled with NHS ester (N-hydroxysuccinimide ester) Alexa Fluor™ 488 before mixing with streptavidin. Labelling the biotin-$CTA2B_5$ rather than streptavidin was essential to allow the degree of labelling (DOL) in the final protein conjugate to reveal the number of $CTA2B_5$ per streptavidin, as will be demonstrated in section II.1.4. Given that NHS ester bioconjugation could target both the primary amine group at the N-terminus and the

many lysine residues (Figure II-7), it was necessary to prevent excessive labelling, which could cause a problem for both microscopy images and binding to the GM1.

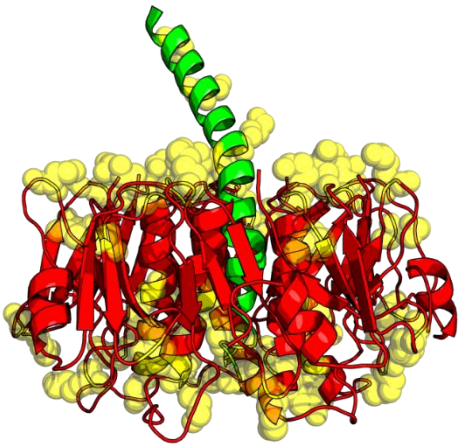


Figure II-7 Distribution of lysine residues (yellow sphere) in the $CTA2B_5$ protein complex.

NHS ester bioconjugation is highly dependent on pH which determines the amine groups' protonation states[98]. A pH of 8.3-8.5 is ideal for this modification reaction since the amine has been deprotonated and is ready for nucleophilic attack. If the pH is too high (>9), the NHS ester will rapidly undergo hydrolysis and lose its function.

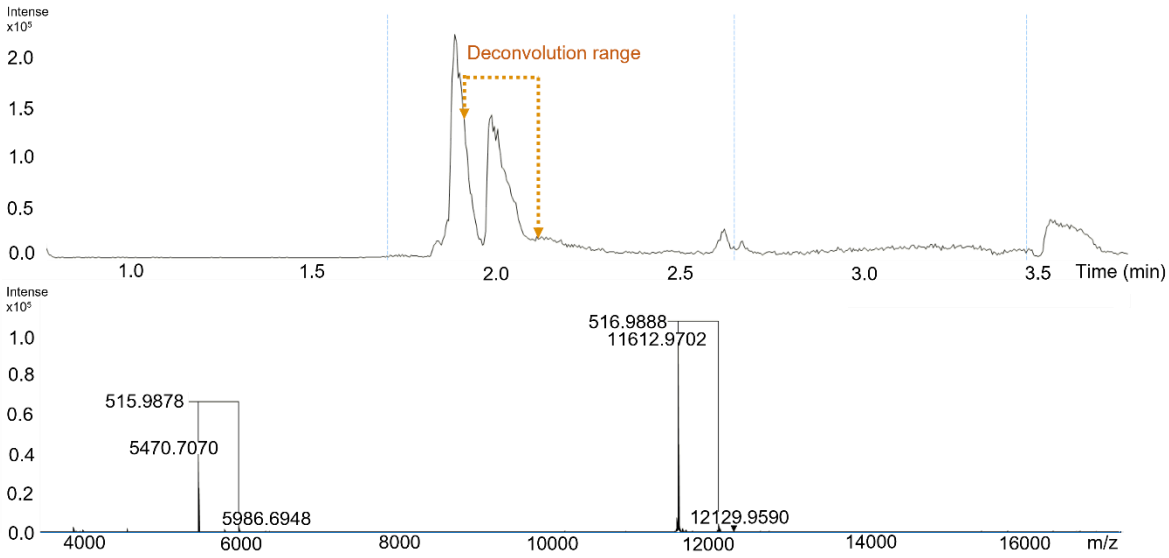


Figure II-8 Total ion mass chromatogram and mass spectra of biotin-$CTA2B_5$ fluorescently labelled with NHS ester-AF488.In the elution from ion mass chromatography, there are two main peaks visible. The first peak, with a deconvoluted mass of 11612 Da, is primarily made up of CTB monomer, and the second peak, with a deconvoluted mass of 5470 Da, is made up of biotinylated CTA2 peptide. The increase of about 515 Da showed that the labeling with NHS eater Alexa fluor 488 was successful. I only choose a small portion of the

first peak to deconvolute, along with the entire second peak, in order to display both subunit in the same panel with a discernible peak height.

Mass spectrometry was used to check if the NHS ester labelling was successful and whether there was excessive labelling (Figure II-8). The labelled protein was collected and exchanged into HEPES buffer to remove excess fluorescent reagent before the absorbance under 280 nm and 495 nm were collected using Nanodrop. Then the degree of labelling per biotin-CTA2B$_5$ was further calculated as 0.42 according to the method described in section VII.6.2.

II.1.2 Preparation of streptavidins with defined numbers of binding sites

Howarth and co-workers have previously described a strategy to prepare streptavidin proteins complexes with different numbers of active binding sites[93]. To obtain streptavidin with a defined orientation and a certain number of alive pockets, *e.g.*, *cis*-divalent, *trans*- divalent and *trivalent* streptavidin, the alive monomer (SA) and dead monomer with insertion of aspartate loop (Dd) were expressed separately by transforming the plasmids obtained from Addgene (reengineered from plasmid #46367, original plasmid #46368) to *E. coli* BL21-CodonPlus (DE3)-RIPL cell, denatured from inclusion bodies with guanidine hydrochloride, mixed in a one-to-one ratio, and refolded. Details of the gene and protein sequence are in Appendix VIII.1.2.a and VIII.1.2.b, and the procedure for protein expression and purification are in Appendix VII.4.5. Ion exchange chromatography using a 5-mL HiTrap® Q Fast Flow resin (Cytiva) was applied to separate distinct streptavidins by their Isoelectric Point (pI), which was heavily influenced by the aspartate loop. SDS-PAGE analysis of the eluted peaks (Figure II-9) indicated the successful separation of *trivalent*- streptavidin (SA3Dd1), *trans*-streptavidin (1,3/1,4-SA) and *cis*-streptavidin (1,2-SA) by comparison of their relative electrophoretic mobilities with the published paper[93].

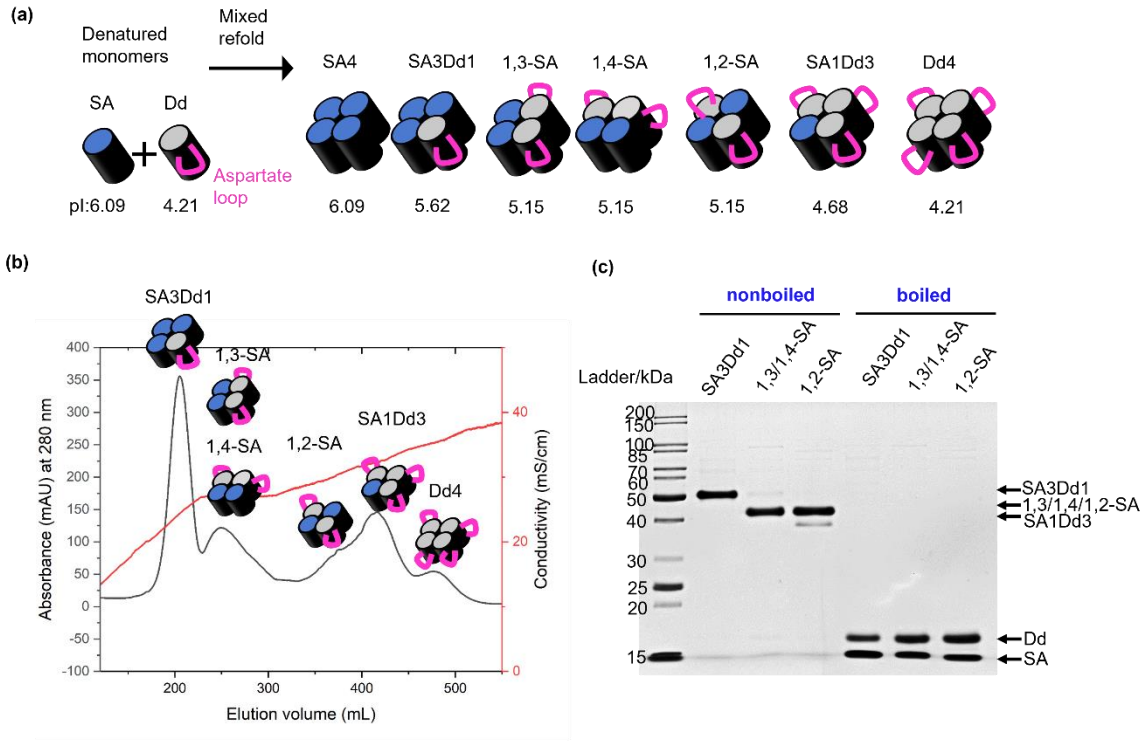


Figure II-9 (a) Figure redrawn based on the cartoons from the literature[93]: denatured alive streptavidin monomer (SA) and dead streptavidin monomer with aspartate loop (Dd) were mixed and refolded into streptavidin tetramer with different compositions. (b) Ion exchange chromatography on HiTrap® Q Fast Flow resin (Cytiva). (c) 16% Commassie stained SDS-PAGE of the *trivalent*- (SA3Dd1), *trans*- (1,3/1,4-SA) and *cis*- (1,2-SA) streptavidin fractions comparing the mobility of proteins with and without boiling the samples prior to electrophoresis. Despite having molecular weights of 13.3 kDa and 14.5 kDa for SA and Dd monomer, respectively, the migration of tetramers did not correspond to their molecular weight in non-boiled protein samples due to their intact tetrameric structure. This phenomenon was attributed to the negatively charged aspartate loops present in Dd monomers, which allowed divalent streptavidin with two Dd molecules to migrate faster than tetravalent streptavidin. Boiling the protein samples with SDS resulted in the linearization of individual monomers and uniform negative charging, enabling their migration according to their molecular weight. SDS-PAGE analysis revealed a slightly higher proportion of SA in the SA3Dd1 boiled sample relative to other divalent streptavidin samples.

II.1.3 Optimisation of conditions for Streptavidin:biotin-CTA2B$_5$ assembly

After successfully obtaining *cis*-, *trans*-, and *trivalent*-streptavidin, the subsequent objective was to identify the ideal temperature and incubation time for saturating the binding pockets with biotinylated CTA2B$_5$. While theoretically, increasing the biotin-CTA2B$_5$: streptavidin monomer ratio could improve saturation, this approach is not feasible due to the intricate production of biotin-CTA2B$_5$. Consequently, the emphasis should be on maximizing the binding efficiency by determining the optimum temperature and incubation time. A mixture was prepared by combining various types of defined

streptavidin that were purified through ion exchange chromatography in section II.1.2, along with the commercial product of streptavidin tetramer. The mixture was then treated with 8 equivalents of biotin-CTA2B$_5$, resulting in a final concentration of 200 nM defined streptavidin tetramer and 1600 nM biotin-CTA2B$_5$.

The protein mixtures were initially incubated at 37 °C overnight to determine whether the defined streptavidin species could reach saturation using mass photometry (Figure II-10-a). Although *trans*- streptavidin was fully saturated in both binding sites with minimal 1-bound complex remaining, *cis*- streptavidin exhibited a higher proportion of 1-bound impurities. *Trivalent*- streptavidin was capable of forming both divalent and trivalent complexes, with little monovalent impurity observed. To investigate the feasibility of achieving a desirable saturation level with a lower temperature of 25°C and shorter incubation time of 1 hour, the procedure was repeated and analyzed using mass photometry (Figure II-10-b). This was done with the aim of optimizing the speed of producing different streptavidin and biotin-CTA2B5 complexes. After comparing the results of the two incubation conditions, it was found that they produced similar levels of saturated defined streptavidin. Therefore, the more efficient procedure of incubating for 1 hour at 25 ºC was chosen for the preparation of the protein complex.

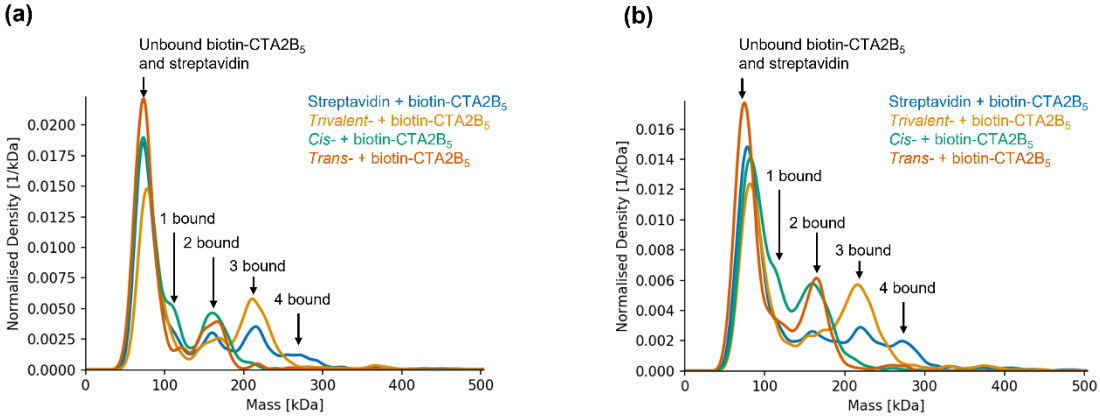


Figure II-10 Mass photometry profiles of different defined streptavidin-(CTA2B$_5$)$_n$ complexes following incubation of different defined streptavidin (200 nM tetramer concentration) with biotin-CTA2B$_5$ (1600 nM) at (a) 37 ºC overnight, or (b) 25 ºC for 1 h. Shown here are the results of a single experiment.

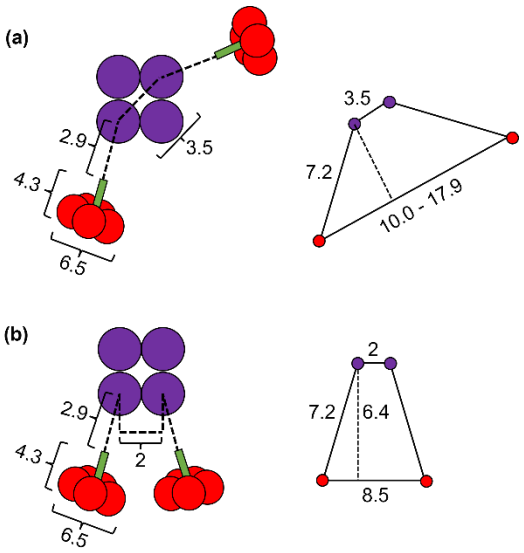


Figure II-11 (a) The diagram shows some ideal case distances for (a) *trans*- and (b) *cis*- streptavidin-$(CTA2B_5)_2$ – some measurement of the distance was obtained by using Pymol. The maximum distance of the streptavidin (pdb: 4bx5) monomer is approximately 3.5 nm, while the distance between the center-of-mass of two streptavidin monomers is approximately 2 nm. To determine the distance of $CTA2B_5$, CTA1 was abstracted from the holotoxin structure (pdb: 1xtc). Based on the assumption that CTA2 is a helix, the distance from the N-terminal end of CTA2 to the GM1 binding surface of CTB_5 was measured to be 4.3 nm, whereas the width of CTB_5 is approximately 6.5 nm. In addition, the length of the Alkoxylamie-PEG4-biotin linker is 2.9 nm according to the statement on the Thermofisher website. In the end, the combination of this information gives us a rough idea of the size of the streptavidin in complex with CTA2B5.

What is the underlying cause of the difference in saturation levels between the binding of biotin-$CTA2B_5$ to *trans*- streptavidin and *cis*- streptavidin? The discrepancy in binding stoichiometries is attributed to the architecture of *trans*- and *cis*- streptavidin, which dictates whether the initial attachment of $CTA2B_5$ would impede the addition of a second molecule. For *trans*-streptavidin (Figure II-11-a), binding of the first biotin-$CTA2B_5$ to the alive pocket will not inhibit the second since it is located at the other side of the streptavidin tetramer. However, for *cis*-streptavidin (Figure II-11-b), the binding of the first biotin-$CTA2B_5$ can potentially block the second protein binding since it can interfere with the space near the second alive pocket. Furthermore, different binding directions result in products with a variety of sizes. After interacting with biotin-$CTA2B_5$, it makes sense that *trans*-streptavidin can form a longer structure than *cis*-streptavidin. Additionally, the $CTA2B_5$ in *trans*- has a greater freedom for space exploration when compared to the $CTA2B_5$ in *cis*-.

II.1.4 Preparation and characterization of streptavidin-(CTA2B$_5$)$_n$ assemblies

As discussed in section II.1.3, 8 equivalents of biotin-CTA2B$_5$ was used to incubate with either *cis*-, *trans*-, or *trivalent*-streptavidin. The resulting mixtures were further purified using a size-exclusion chromatography system (SD200, 16/60).

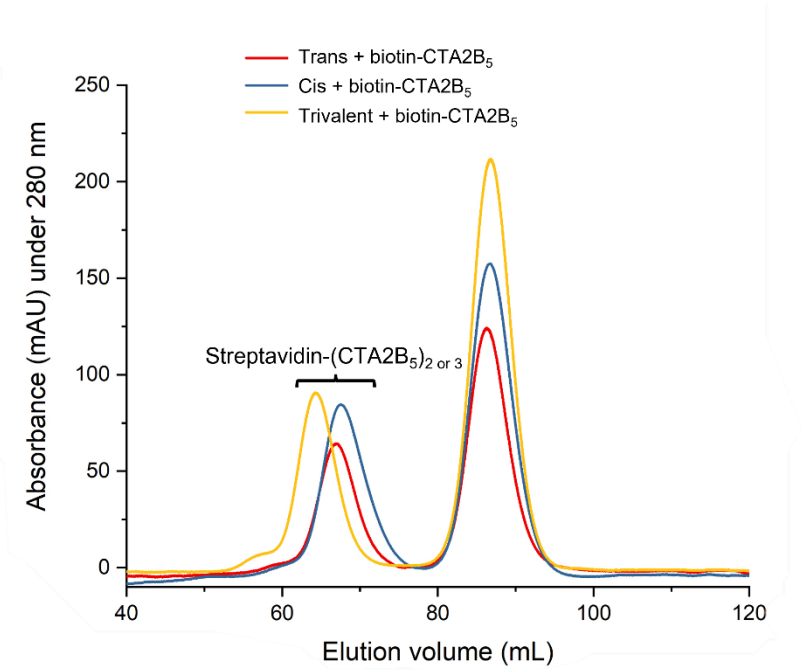
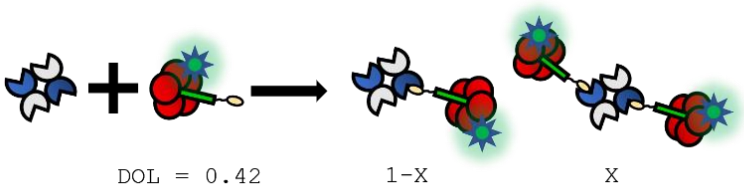


Figure II-12 Size exclusion chromatography (SD200, 16/60) of different types of streptavidin interacting with biotin-CTA2B$_5$.

The slight shift of the elution volume for different types of streptavidin-(CTA2B$_5$)$_n$ is consistent with the different saturation states of the final protein complex. Comparing the mixtures of biotin-CTA2B$_5$ with *trans*- and *cis*-streptavidin, the slightly earlier elution volume of *trans*-complexes indicates more dimeric occupation compared to the *cis*-complex, which also corresponds to the mass photometry results (Figure II-10). Furthermore, the disparity in hydrodynamic radius between *trans*-complexes and *cis*-complexes can also be used to explain this phenomenon. The previous illustration (Figure II-11) shows that the distance between the two CTB$_5$ in *trans*-complex can even get as large as 18 nm, which is roughly twice the size of *cis*-complex. Hence, the trans- might retain for a shorter time in column and eluted earlier.

The eluted streptavidin-(CTA2B$_5$)$_n$ complexes were concentrated using a spin concentrator, and their absorbances at 280 nm and 495 nm were measured to calculate the concentration of each final protein complex.

We have previously calculated the degree of labelling (DOL) of fluorescent biotin-CTA2B$_5$ alone to be 0.42, allowing us to use determine the overall DOL of the *trans*-complex or *cis*-complex. We need to first assume that the final *trans*-complex or *cis*-complex product only contains two mixtures—one 1-bound and the other 2-bound—and that their respective contents percentages can be set to 1-x and x as the unidentified component (Figure II-13). The DOL of the complex (named DOL1) can be calculated by detecting the absorbance at 495 nm and 280 nm using Nanodrop and applying the values in the given formula (Method section VII.6.2, Equation 3, Equation 4). The alternative way of calculating DOL requires to define the DOL of complex mixture as 0.42*(1-x)+0.42*2x using the DOL of biotin-CTA2B$_5$, which is then taken as DOL2. Remember that DOL1 and DOL2 are equal, hence we can deduce the value of x. While still applicable to *trivalent*-complex systems, this method is less accurate because it counts on the final solution to only have 3-bound and 2-bound components, not 1-bound ones.



$$Complex\ concentration = \frac{A_{280} - A_{495} \cdot CF}{X \cdot \varepsilon_{2-bound} + (1-X) \cdot \varepsilon_{1-bound}}$$

$$DOL1 = A_{495} \frac{\varepsilon_{label}}{Complex\ concentration}$$

$$DOL2 = 0.42 \cdot (1 - X) + 0.42 \cdot 2X$$

$$DOL1 = DOL2$$

$$\therefore X = 0.916$$

Figure II-13 Example of calculating the percentage of different compositions of streptavidin-(CTA2B$_5$)$_n$ complex. The Correction factor (CF) from NHS ester Alexa fluor 488 is 0.11. The extinction coefficient of streptavidin tetramer at 280 nm is 167760 M^{-1} cm^{-1}, CTA2B$_5$ is 60905 M^{-1} cm^{-1}. Accordingly, the extinction coefficient for monovalent complex ($\varepsilon_{1\text{-bound}}$) is 228665 M^{-1} cm^{-1}, divalent complex ($\varepsilon_{2\text{-bound}}$) is 289570 M^{-1} cm^{-1}, trivalent complex ($\varepsilon_{3\text{-bound}}$) is 350475 M^{-1} cm^{-1}. ε_{label} is equal to 73,000 M^{-1} cm^{-1}, as the extinction coefficient of the NHS ester - Alexa fluor 488 label.

According to the calculation method in Figure II-13, the percentage of fully saturated streptavidin can be obtained (Table II-1). The majority of the binding pockets for *trans*-streptavidin were occupied, compared to only 61% occupation for *cis*-streptavidin – which agrees with the mass photometry result shown in and the elution volume from size exclusion chromatography .

Table II-1. The quantity of monovalent, divalent, and trivalent protein complexes. The various protein complexes are obtained using SEC purification (section II.1.4). The ratios of the different valent complexes were calculated by employing the earlier described method (Figure II-13). Finally, the DOL of the entire protein complex is counted, which shows the level of labelling per molecule. The calculation is based on the product obtained from the previous section utilizing a ratio of 1:8 for streptavidin tetramer to biotin-CTA2B$_5$, followed by size exclusion chromatography. It does not reflect the maximum purification capability of the stipulated streptavidin and biotin-CTA2B5 system.

Streptavidin	monovalent complex	divalent complex	trivalent complex	DOL of final complex
trans-	8%	92%	-	0.81
cis-	39%	61%	-	0.68
trivalent-	-	29%	71%	1.28

II.2 GUV experiment

The preliminary experiments comparing different defined streptavidin-(CTA2B$_5$)$_n$ complexes was done by Erik Kempmann and me while I was on secondment in the University of Freiburg. After I left, Erik further completed the GUV replication tests with the supervision of Lina Siukstaite (University of Freiburg). The images used in the following sections are from his experimental results during experiment replications. GUVs functionalised with 5 mol% ganglioside GM1 were made by the electroformation method as described in section VII.12, and a negative control was employed to demonstrate that the GUVs exist as discrete spherical vesicles in the absence of protein (Figure II-14). The *cis*-, *trans*- or *trivalent*-streptavidin-(CTA2B$_5$)$_n$ fluorescently labelled with Alexa Fluor 488 was collected directly from size exclusion chromatography shown in Figure II-12. According to the calculation in Table II-1, the protein complexes from different species were diluted to a final CTB$_5$ concentration of 400 nM to ensure the behavior of GUV could be analysed based on the amount of CTB$_5$ present.

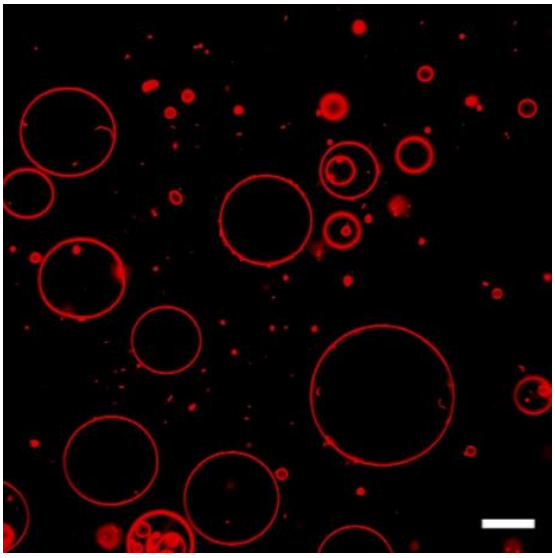


Figure II-14 A negative control was performed by solely incorporating GUVs (comprising of 0.3 mol% Atto647-labelled lipid, 64.7 mol% DOPC, 30 mol% cholesterol, and 5 mol% GM1) without the presence of any protein.

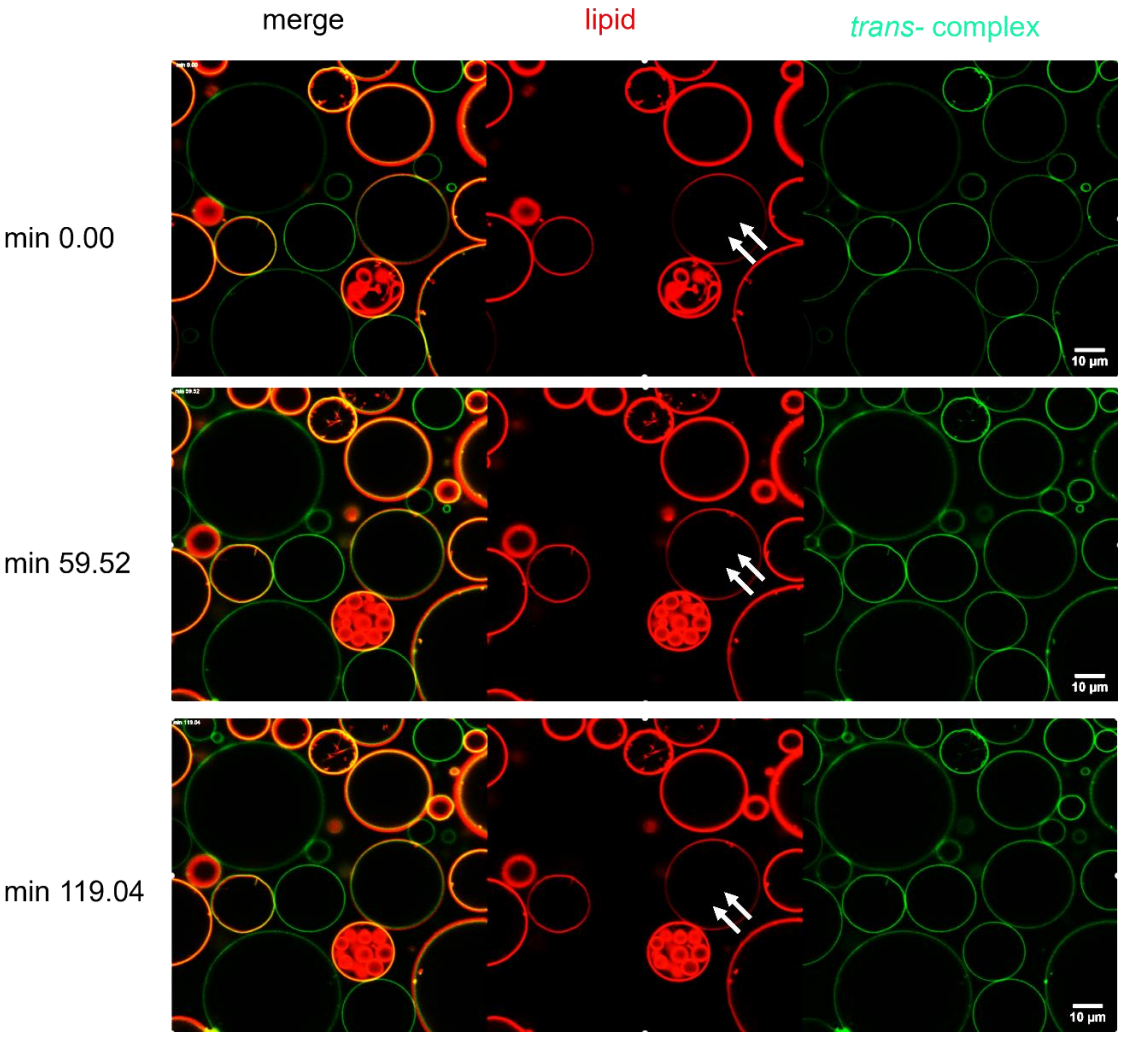


Figure II-15 Confocal microscopy images of GM1-GUVs at three time points (t = 0, t ≈ 1 hour and t ≈ 2 hours) after the addition of *trans*-complex to GM1-GUV (final CTB_5 concentration was 400 nM). Two types of GUV (either 65 mol% DOPC, 30 mol% cholesterol, and 5 mol% GM1; or 0.3 mol% Atto647-labelled lipid, 64.7 mol% DOPC, 30 mol% cholesterol, and 5 mol% GM1) were mixed before loading in the microscope chamber. The red channel shows the fluorescently-labelled lipid and the green channel shows the Alexa488-labelled *trans*-streptavidin-$(CTA2B_5)_n$. The arrow points to an initially no dye-GUV with red fluorephore transferred evidenced by obviously weak red fluorescence compare with the neighboring dye-GUV. Presented here are one of the result image obtained from four times of repeats.

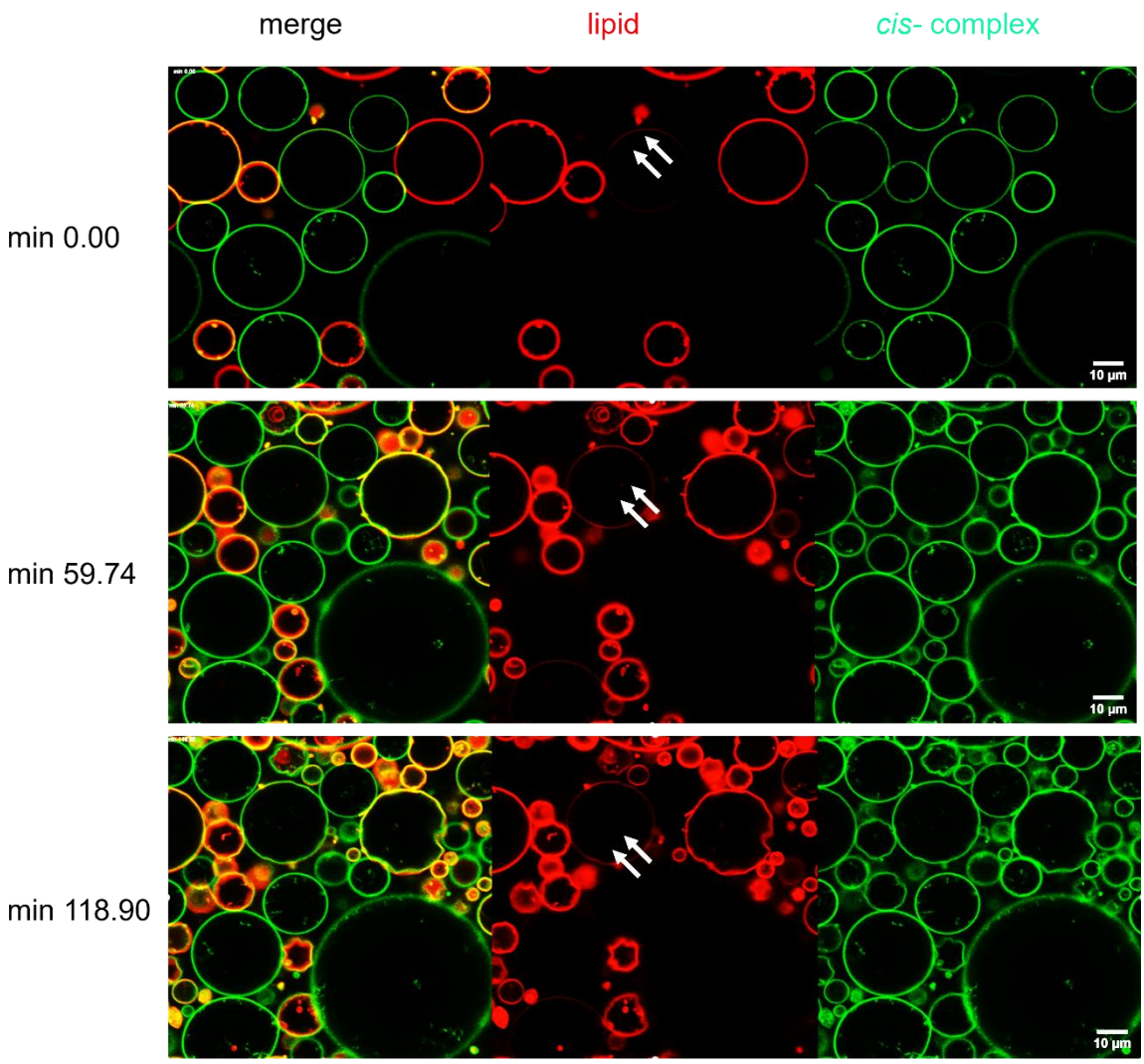


Figure II-16 Confocal microscopy images of GM1-GUVs at three time points (t = 0, t ≈ 1 hour and t ≈ 2 hours) after the addition of *cis*-complex to GM1-GUV (final CTB_5 concentration was 400 nM). The red channel shows the fluorescently-labelled lipid and the green channel shows the Alexa488-labelled *cis*-streptavidin-$(CTA2B_5)_n$. The arrow points to an initially no dye-GUV which gradually turns red, indicating its lipid exchange with the neighboring dye-GUV. Presented here are one of the result image obtained from four times of repeats.

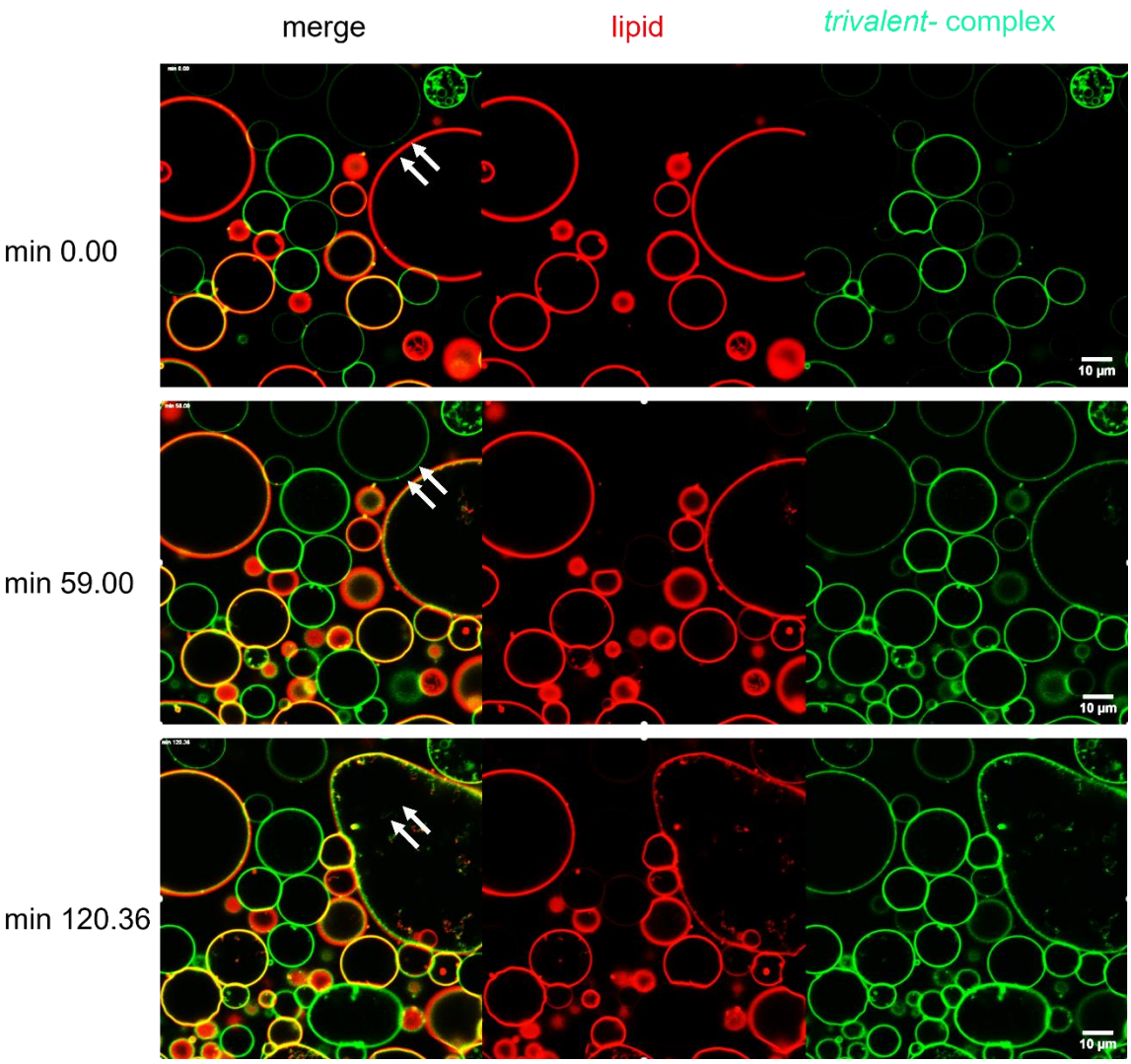


Figure II-17 Confocal microscopy images of GM1-GUVs at three time points (t = 0, t ≈ 1 hour and t ≈ 2 hours) after the addition of *trivalent*-complex to GM1-GUV (final CTB_5 concentration was 400 nM). The red channel shows the fluorescently-labelled lipid and the green channel shows the Alexa488-labelled *trivalent*-streptavidin-$(CTA2B_5)_n$. The arrow points to a fusion event occurred between a dyed GUV and no dye GUV. Presented here are one of the result image obtained from four times of repeats.

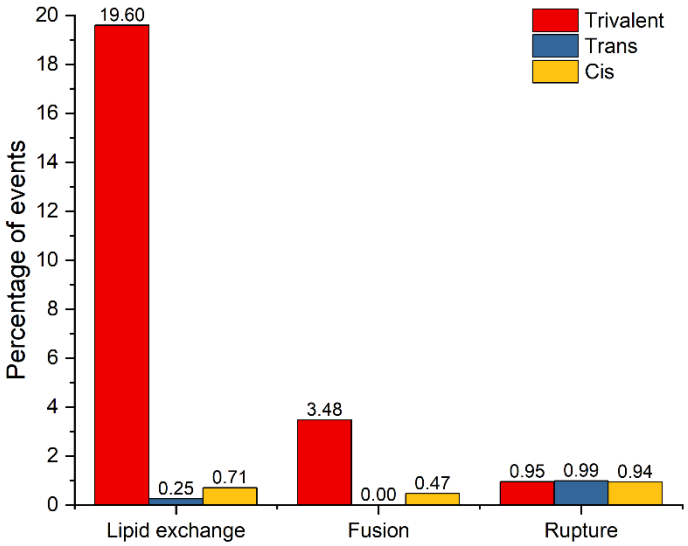


Figure II-18 The percentage of events triggered by different defined streptavidin-$(CTA2B_5)_x$ complex, e.g. *trivalent*- (red), *trans*- (blue), and *cis*- (yellow) respectively, determined through quantifying the frequency of events in a GUV population. Data was extracted from a unified dataset of 316, 403, and 425 GUVs from 4-time repeat experiment for *trivalent*-, *trans*-, and *cis*- streptavidin-(CTA2B5)x, respectively. Specifically, upon addition of the *trivalent*- complex, 19% of the 316 GUVs underwent lipid exchange compared to only 0.24% and 0.70% for the *trans*- and *cis*- complexes, respectively. Additionally, 3.47% of GUVs incubated with the *trivalent*- complex underwent fusion, while only 0.4% of GUVs experienced fusion with the *cis*- complex, and no fusion events were observed for the *trans*- complex.

Comparing the vesicle performances among *trans*-, *cis*- and *trivalent*-streptavidin-$(CTA2B_5)_n$ showed that the *trivalent*-species (Figure II-17) allowed more crosslinking than either *cis*- (Figure II-16*)* or *trans*-complexes (Figure II-15) on the same time scale. In contrast to *trans*- or *cis*-complexes, which resulted in individual and separated vesicles, the *trivalent*-complex contributed to more crosslinking, as evidenced by the flattened interfaces between vesicles. Although *trans*- and *cis*- *complexes* successfully bound to the vesicles, there did not seem to be much interaction between the liposomes as the incubation time increases.

After 2h incubation time, the *trivalent*-streptavidin-$(CTA2B_5)_n$ eventually provoked membrane fusion, while such an event was not observed for either *cis*- or *trans*-complexes during the experiments. Nevertheless, although no noticeable crosslinking was witnessed for either *trans*- or *cis*-streptavidin complexes, lipid exchange was still observed in both experiments which increased with time, as indicated by the white arrow in Figure II-15 and Figure II-16. These observations were further assessed by quantifying the total number of GUVs and the incidence of events such as lipid exchange, fusion,

and rupture over a 2-hour incubation period in repeated experiments (Figure II-8). Specifically, the *trivalent*- complex demonstrated a significantly greater number of lipid exchange and fusion events compared to the *trans*- and *cis*- complexes, while rupture events were equally distributed among the three species.

Since a mediator is required for lipid exchange between liposomes, *cis*- and *trans*-complexes, despite their inability to cause cross-linking, still hold the promise that lipid exchange can take place without the need for lengthy, flattened cross-linked interfaces.

It is considered that for both *cis*- and *trans*-complexes, a maximum of two $CTA2B_5$ are linked to one streptavidin tetramer through a biotin-PEG_4 flexible linker. Intuitively, we expected these two CTB_5 could interact with different vesicles and drag the liposomes close enough for fusion. Moreover, the *trans*-complex would be expected to exercise the function more efficiently than *cis*-complex since the CTB_5 could explore the opposing position more readily than the *cis*-complex, but what if both CTB_5, no matter whether *cis*- or *trans*-complex, predominantly bind to the same vesicle plane? This would result in little unbound CTB_5 facing towards the solution and available to interact with a neighbouring liposome and could potentially explain why there is little crosslinking for *cis*- and *trans*-streptavidin-$(CTA2B_5)_n$. On the contrary, it would be much easier for *trivalent*-streptavidin-$(CTA2B_5)_n$ to possess at least one free CTB_5 for membrane crosslinking. Building on this point, the fusion efficiency could be expected to be maximal when all four binding pockets of streptavidin are occupied since the fourth biotin-$CTA2B_5$ would need to overcome an even stronger geomatric restraints and conformational energy barrier to be able to interact with the same membrane plane in parallel with the other three CTB_5.

II.3 Lipid mixing assay

Confocal microscopy experiments allow a visual understanding of the behaviour of the membrane, but quantitative analysis of the data is more informative. Therefore, to obtain a quantitative measure of the relative ability of each complex to fuse and hemi-fuse vesicles, the complexes were evaluated in a lipid mixing assay (Figure II-19). The lipid mixing assay was performed based on the measurement of Fluorescence Resonance Energy Transfer (FRET) between NBD and Rhodamine. At least two types of 800 nm Large Unilamellar Vesicles (LUV) were made by extruding method: The first is FRET-LUV with NBD-DOPE and Rhodamine-DOPE incorporated in membrane with 0.25 mol% for each. The second is blank-LUV without any fluorescence dye; Both LUVs contain 5 mol% GM1, which allows CTB_5 to interact and bind to membrane. The FRET-LUV and

blank-LUV were mixed in a 1:4 ratio to achieve a final concentration of 133 µM lipid concentration (6.65 µM GM1 concentration). The mixed LUVs were then exposed to different concentrations of protein from 0 h to maximum 8 h. The value for full lipid mixing control was obtained by treating the LUVs mixture with 0.2% Triton-X100, which represents the maximum dilution of the NBD and Rhodamine in membrane for our experiment. Furthermore, the LUVs mixture without the addition of any protein was observed as a control for the liposome self-fusion, and only the time period during which the percentage of self-fusion is low are used to calculate the lipid mixing rate induced by protein candidates (usually within 5h).

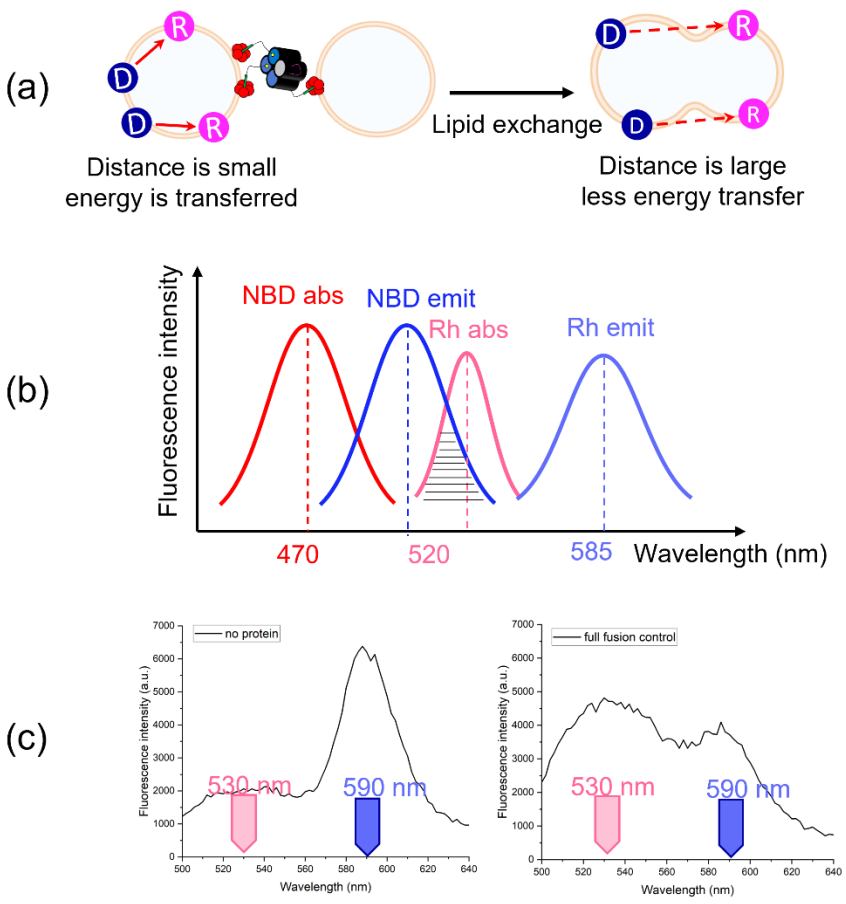


Figure II-19 (a) Schematic illustration of *trivalent*-streptavidin-$(CTA2B_5)_n$-mediated liposome fusion based on proximity-dependent fluorescence resonance energy transfer (FRET). After lipid exchange with probe-free liposomes, the distance between the NBD (Donor) and Rhodamine (Receptor) in FRET-LUV is lengthened. (b) A diagram showing the FRET signal. The NBD has a 470 nm absorbance and a 520 nm emission wavelength. This falls within the rhodamine absorbance range and thus contributes to the signal that rhodamine emits, which peaks at 585 nm. (c) A comparison of the fluorescence signal between the pre- and post-lipid exchange states at wavelengths between 500 and 640 nm. The relative fluorescence intensity at

530 nm rises as a result of quenching because the fluorescence emitted by NBD is not effectively transmitted to excite rhodamine.

When FRET-LUV is fused or hemifused to blank-LUV and the average distance between NBD and Rh molecular increases (Figure II-19-a), the intensity of NBD's emission light at 520 nm increases while Rh's emission light at 585 nm decreases in the detected fluorescence region (500 nm - 650 nm), leading to a dequenching phenomenon. The dequenching percentage can hence be deduced the relative fluorescence intensity at 520 nm and 585 nm, which serves as key value for calculating the lipid mixing percentage and rate. (Details on the calculations and experiment setting are in the Method and Material section VII.14).

It's critical to remember that the lipid mixing assay, as opposed to measuring the efficiency of membrane fusion, evaluates only the efficiency of lipid exchange – which can be caused from both hemifusion and fusion events. The difference between hemifusion and fusion, however, is whether the inner membrane has been disrupted and pore formation to allow the content interchange between two vesicles, which would need to use either content mixing or by quenching the outer membrane signal to distinguish.

Different defined streptavidin (*tetra*-, *trivalent*-, *trans*- and *cis*-) were mixed with biotin-CTA2B$_5$ in a 1-to-1 ratio of alive streptavidin monomer to biotin-CTA2B$_5$ (Figure II-20-a). After the protein mixtures were incubated at 25 ºC for at least one hour, each was diluted to 200 nM streptavidin tetramer concentration and 2 µl of protein solution was mixed with 18 µl HEPES buffer on top of a mass photometry glass chip to detect the composition (Figure II-20-b). It should be emphasized that the protein complexes used in lipid mixing assay were simply prepared by mixing streptavidin and biotin-CTA2B$_5$, and they were not further purified by SEC, hence there was unbound biotin-CTA2B$_5$ in solution. For *cis*- and *trans*-streptavidin, there were maximumly two biotin-CTA2B$_5$ attached to one streptavidin tetramer, the *trans*-streptavidin having more of the divalent component. For *trivalent*- and *tetravalent*-streptavidin, the maximum binding sites occupied per tetramer was three and four, respectively. Overall, compared with the previous GUV incubation test where biotin-CTA2B$_5$ was used at a higher concentration (Tabel II-1), the percentage of full saturation for each four types of streptavidin was decreased. Nevertheless, the mixtures used here still allowed a comparison of the lipid mixing abilities of complexes with the same maximum numbers of biotin-CTA2B$_5$.

(a)

Protein mixture	Streptavidin (μM tetramer)	Biotin-CTA2B$_5$ (μM)
Streptavidin + biotin-CTA2B$_5$	0.75	3
Trivalent- + biotin-CTA2B$_5$	2	2
Cis- + biotin-CTA2B$_5$	3	2
Trans- + biotin-CTA2B$_5$	3	2

(b)

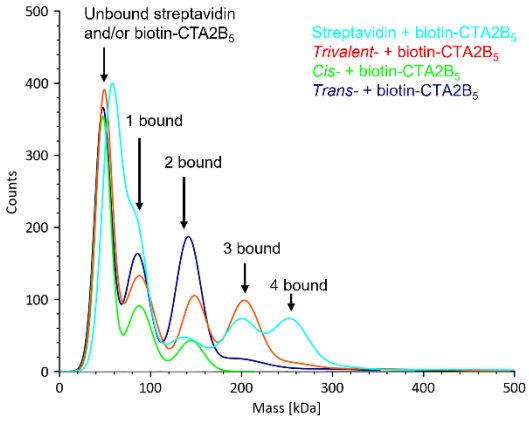


Figure II-20 (a) The final concentration of different protein mixtures. (b) Overlay of mass photometry results for all four different defined streptavidins interacting with biotin-CTA2B$_5$. Each colored curve corresponds to a single mass photometry experiment.

To quantify the lipid exchanging rate from different streptavidin-(CTA2B$_5$)$_n$ species, the *tetra*-, *trivalent*-, *cis*- and *trans*-streptavidin-(CTA2B$_5$)$_n$ protein mixtures were prepared as in Figure II-20-a and diluted to 6 μM CTB$_5$ concentration. Large Unilamellar Vesicles (LUVs) were obtained by extrusion according to the description in Method section VII.14.1. 5 μl of 400 μM LUV, containing the mixture of no dye-LUV and FRET-LUV in a ratio of 4:1, was loaded into the low volume 384-well plate manually, and the solution was topped up with 10 μl of protein mixture to reach a final concentration as 133 μM lipid (6.65 μM GM1) and 4 μM CTB$_5$. A series of protein samples with decreasing concentrations of CTB$_5$ from 4 μM to 0.0625 μM were created by successive two-fold dilution, and the protein was mixed with a final concentration of 133 μM lipid (6.65 μM GM1). While a negative control replacing protein solution with buffer was set to monitor the self-fusion of the liposome.

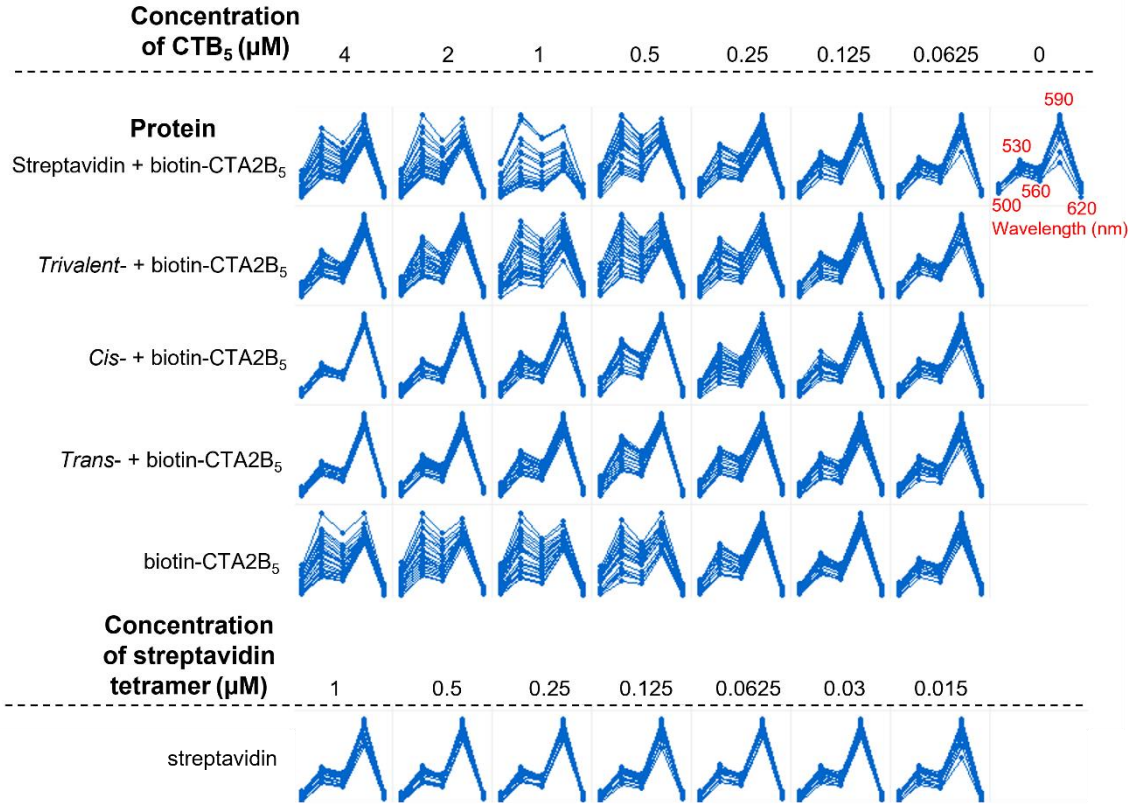


Figure II-21 Lipid mixing results were obtained by exciting at 460 nm and fluorescence detection at 500 nm, 530 nm, 560 nm, 590 nm and 620 nm, respectively. The total data collection lasted for 6.5 h, and the ratio of fluorescence intensity under 530 nm to 590 nm could reflect the lipid mixing efficiency. The image presented here is one representative result from three-time repeats.

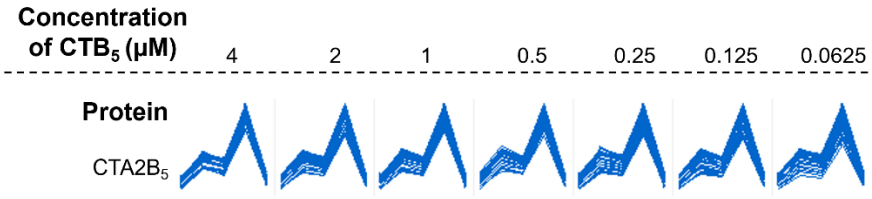


Figure II-22 Lipid mixing results of El Tor CTB_5 and $CTA2B_5$, obtained by exciting at 460 nm, and fluorescence detection at 500 nm, 530 nm, 560 nm, 590 nm and 620 nm respectively. The total data collection lasted for 6.5 h. The image presented here is one representative result from three-time repeats.

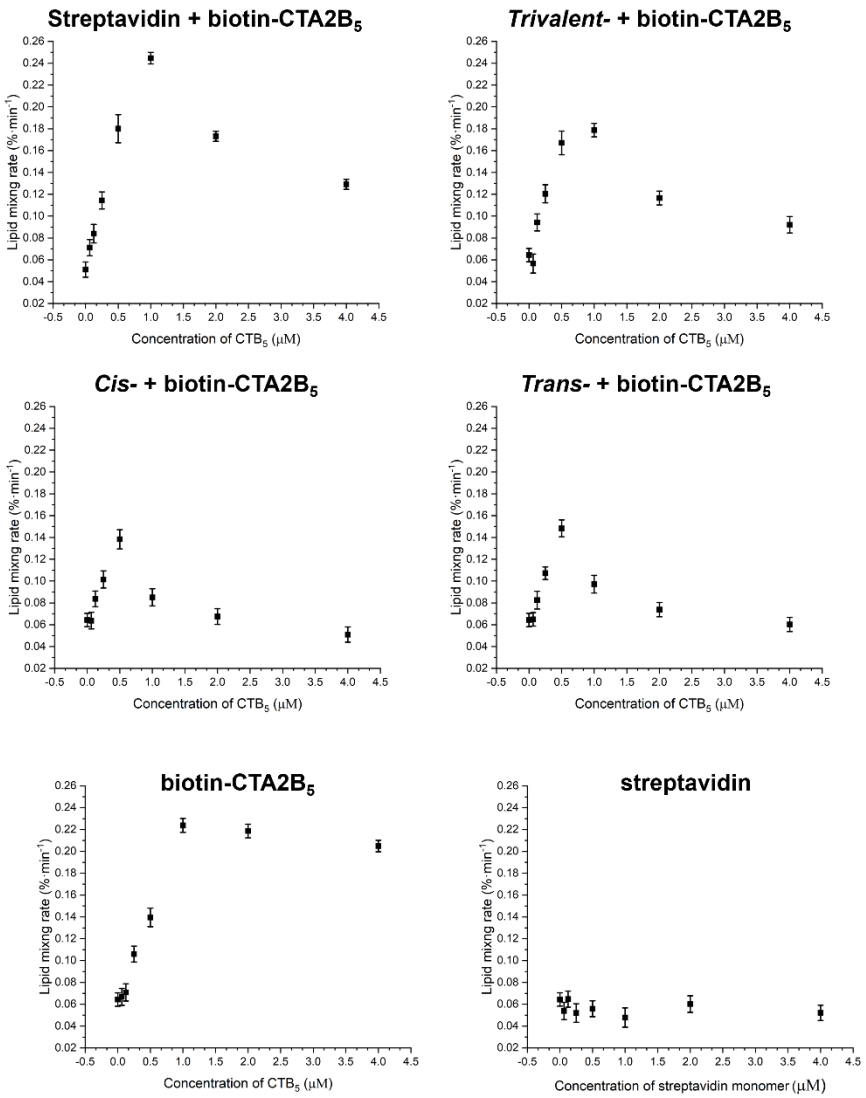


Figure II-23 Lipid mixing rate of different protein mixtures under various protein concentrations. Calculated by using the timepoints within 5 h incubation time from three repeat experiments.

The fluorescence intensity measured at five wavelengths from 500 to 620 nm was overlaid for each well across 6.5 h detection time (Figure II-21, Figure II-22). Detailed information related to the lipid exchanging rate can be obtained by comparing the change in the ratio of fluorescence intensity at 530 nm and 590 nm. The lipid mixing rate (*i.e.*, the increase of lipid mixing percentage per minute) for each sample was calculated using a global fit of three repeats of each experimental condition (Appendix VIII.5.2). The resulting lipid mixing rate for each protein species across different protein concentrations is plotted in Figure II-23, to emphasise how the lipid mixing rate differs as a function of protein concentration and when comparing one protein to the other. The rate of lipid mixing for each streptavidin-$(CTA2B_5)_n$ complex was found to be concentration-

dependent. Depending on the complex, the maximum rate of lipid mixing was attained at CTB_5 concentrations of 0.5–1 µM. Lower lipid mixing rates were observed at higher protein concentrations. In the GUV experiments, the lipid concentration was around 50 nM (including 2.5 nM GM1), while the protein complex's CTB_5 concentration was 400 nM, finally resulting in a protein-to-GM1 ratio as 160. If we want to achieve the same protein-to-GM1 ratio in the lipid mixing experiment as used in GUV, given that the lipid concentration in this experiment was 133 µM (including 6.65 µM GM1), the solution protein concentration would need to be about 1 mM. Because there were so much more protein used in GUV experiment than we currently use in lipid mixing experiments, the lipid mixing rate in GUV expriments should be very low, which would further explain why it is more challenging to observe hemi-fusion and fusion events for all protein complexes.

When only streptavidin was incubated with the LUV, the lipid mixing rate across different streptavidin concentrations was not significantly increased, which means streptavidin itself cannot induce lipid exchange between FRET-LUV and blank-LUV. However, when biotin-$CTA2B_5$ was used individually as a control, the lipid mixing rate increased in a concentration-dependent manner (Figure II-23). Furthermore, the unmodified $CTA2B_5$ cannot induce significant lipid exchange by itself (Figure II-22, Figure II-23). This fact implies the addition of a biotin-PEG_4 linker might assist the lipid exchange, albeit there is no evidence of what is the exact membrane behaviour it induced. It is important to remember that to trigger lipid exchange from one vesicle to the other, the external mediator must be able to connect to both liposomes in a way that can disrupt the local lipid bilayer. As $CTA2B_5$ cannot do so, the biotin-PEG_4 linker presumably can weakly interact with the neighbouring liposome or even penetrate through the bilayer as the mediator. A disturbance on the lipid bilayer, even if it is a subtle force for individual biotin groups, could accumulate in a lipid-rich environment and finally induce membrane fusion. A negative control experiment could be set up by adding the same amount of biotin-PEG_4 alone to explore whether it is due to the existence of a biotin-PEG_4 linker or the fact that biotin-PEG_4-$CTA2B_5$ conjugate leads to the increase of lipid mixing efficiency.

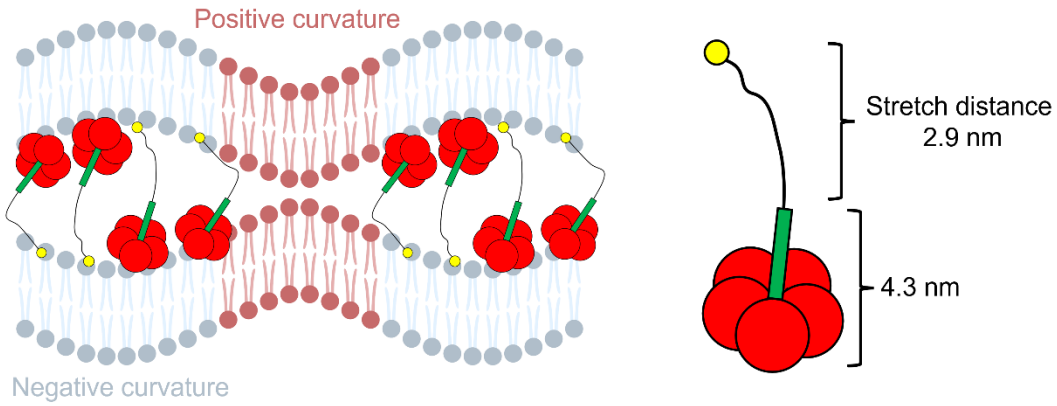


Figure II-24 Scheme demonstrating how biotin-CTA2B$_5$ could interact between two lipid bilayers and induce lipid exchange.

Since the formation of a minute fusion pore depends on the distance between two lipid bilayers less than 3 nm[99], biotin-CTA2B$_5$ need to reduce the space between two membranes for membrane fusion or lipid exchange to occur. A local negative curvature caused by CTB$_5$-liposome interaction will be accompanied by a positive curvature nearby, which is presumably able to bring two lipid bilayers closer together. For biotin-CTA2B$_5$, the overall length of the CTA2B$_5$ protein region is around 4.3 nm, and the accumulation of negative curvature from multiple copies of biotin-CTA2B$_5$ on the neighbouring region could further sink the protein inwards and drag the opposing liposome closer – which creates the chance for lipid exchange.

However, as indicated by the few fusion events seen in the GUV experiments, *cis*- and *trans*-streptavidin-(CTA2B$_5$)$_2$ appear to have more difficulty overcoming the distance barrier and induce membrane fusion (Figure II-15, Figure II-16). A possible explanation for why their lipid mixing rates are lower than those of their *tetra*- or *trivalent*- complex counterparts: despite the fact that the final concentration of biotin-CTA2B$_5$ is the same in all samples, both CTB$_5$ from the same *cis*- or *trans*-streptavidin could bind to the same membrane surface, preventing an additional CTB$_5$ from reacting with the neighboring surface. Additionally, *cis*- or *trans*-streptavidin crosslinking between two liposomes creates a longer distance for excess biotin-CTA2B$_5$ to stretch between neighbouring vesicles, which could attenuate any fusion or hemifusion induced which could be potentially induced by biotin-CTA2B$_5$ (Figure II-25).

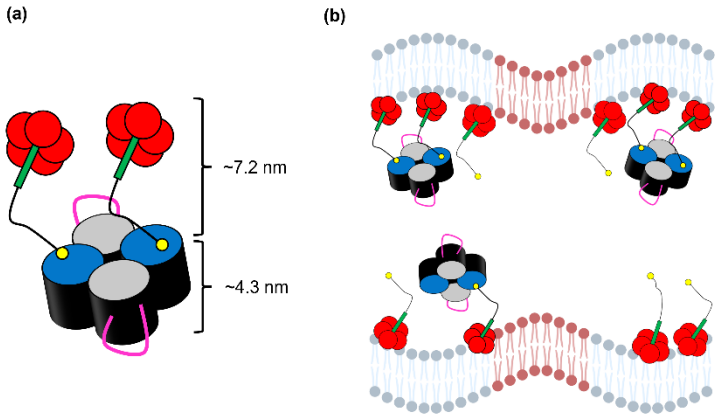


Figure II-25 Scheme showing why *cis*-streptavidin-$(CTA2B_5)_2$ cannot induce high lipid exchange rate. (a) Distance measurement of the biotin-$CTA2B_5$ having a length of 7.2 nm, and the length of streptavidin is 4.3 nm. (b) The protein-membrane interaction is induced by the *cis*-streptavidin-$(CTA2B_5)_2$ mixture, with excess biotin-$CTA2B_5$ in the solution.

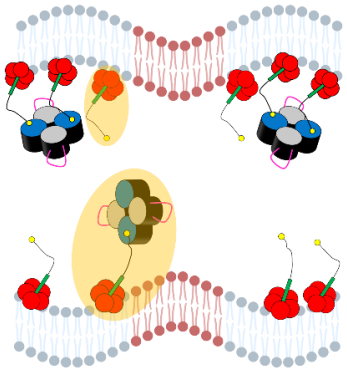


Figure II-26 Lipid exchange could still happen from the reconstitution of divalent $CTA2B_5$ complex. Any partially complexed *cis*- streptavidin could potentially form a cross-link between membranes bound to biotin-$CTA2B_5$.

The protein interplay on the surface of the membrane could be more complicated. Since not all *trans*- or *cis*-streptavidin binding sites are occupied, any free biotin-$CTA2B_5$ from solution bound to the neighbouring membrane could interact with a free streptavidin binding site to connect and bring two liposomes closer (Figure II-26). Nevertheless, if it is preferable for any free biotin-$CTA2B_5$ to interact with the free streptavidin binding sites in the same liposome, then that would disfavor membrane crosslinking and thus lipid exchange. Therefore, even though there might be lipid exchange induced from this mechanism, the proportion is still limited.

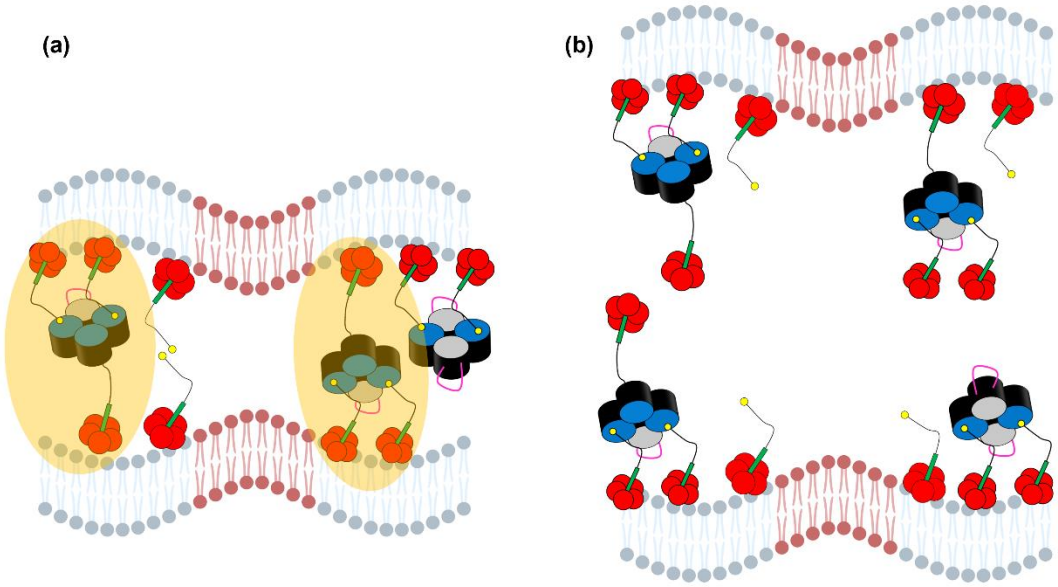


Figure II-27 Scheme showing a mechanism for *trivalent*-streptavidin-$(CTA2B_5)_3$ inducing membrane fusion. (a) The *trivalent*- complex crosslinks the two membranes and induces membrane bending, which creates a closer distance for the third CTB_5 to bind the neighboring liposome and induce lipid exchange. (b) The inhibition effect from having excess CTB_5 in the system will arise from complete binding of the GM1 on the isolated vesicles. Hence even though there might be free CTB_5 binding sites available, they cannot bind to the opposing membrane and induce lipid exchange.

Trivalent-streptavidin-$(CTA2B_5)_3$ can crosslink two liposome with higher efficiency (Figure II-27), and from the lipid mixing result, the *tetramer*-streptavidin-$(CTA2B_5)_4$ can induce even higher lipid exchange rate than the *trivalent*- counterpart. According to the mass photometry result in Figure II-20, it is explainable by the overall larger population of trivalent and tetravalent complexes by incubating tetramer streptavidin with biotin-$CTA2B_5$, which was deduced to be the main driving force for membrane crosslinking and lipid exchange. However, when there is excessive amount of protein, the lipid mixing rate from both *trivalent*- and *tetramer*- complexes decrease. One explanation is the protein complex would preferably land on the vesicles before stretching out the additional CTB_5 for neighboring vesicle, when the GM1 is fully occupied by the overwhelmingly CTB_5, little is avaible to interact with the neighboring CTB_5 for vesicle crosslinking (Figure II-27).

II.4 Conclusion and future works

In this project, various defined streptavidins were purified that only have either one, two, or three alive monomers per streptavidin tetramer, from which the *trans*- divalent, *cis*-divalent, and *trivalent*- streptavidin were used to interact with biotin-$CTA2B_5$ for creating multiple CTB_5 complex. We were able to determine the composition of each complex

through direct measurements using mass photometry and calculations based on the degree of fluorescence labeling of the complexes. Furthermore, we observed how different complexes behave on membranes through GUV experiment; While trivalent compounds can cause hemi-fusion and fusion, divalent compounds are ineffective at fusing membranes. Finally, I quantified lipid exchange rate induced by different protein complexes in lipid mixing experiments, ultimately demonstrating that the divalent complexes, which we had hypothesized to play a significant role in membrane fusion in the previous publication[80], were actually the least active species, whereas the trivalent and tetravalent complexes were more effective at membrane cross-linking and fusion. It is important to remember that in earlier research, we did find evidence in the SEC-MALLS data for the existence of trivalent complexes. As a result, this project examines the mechanism of the membrane fusion brought on by the streptavidin-$(CTA2B_5)_n$ complex in detail and acts as a blueprint for the development of future CTB_5 assemblies that can perform membrane fusion more effectively.

To achieve efficient membrane fusion, it is essential to have an anchor in one membrane and an excess of free bindable CTB_5 that can interact with ganglioside GM1 at opposing membranes in order to crosslink membranes. Divalent CTB_5 complexes assembled with PEG_4 flexible linkers make this process more challenging because the two CTB_5 units in the same streptavidin complex can interact with the same membrane surface and hence cannot be used to interact with the opposite membrane. Furthermore, when the CTB_5: lipid ratio is high, particularly when the concentration of CTB monomers is higher than the total concentration of GM1 on the vesicle surface and the Kd for the CTB-GM1 interaction, one cannot be certain that there is free GM1 on each liposome. This is because GM1 is quickly occupied by the CTB_5 from the same streptavidin-$(CTA2B_5)_n$ complex. In other words, divalent streptavidin-$(CTA2B_5)_2$ complex could not be effectively used for crosslinking neighboring liposomes.

One strategy to overcome this dilemma is to force at least one protein facing towards a solution using the privilege of protein topology, as in the case of *trivalent-* and *tetra-*streptavidin-$(CTA2B_5)_n$. However, this cannot solve the problem that the GM1 in the same liposome can be quickly occupied under high protein concentration since there are always two CTB_5 standing and recruiting the GM1 in the same liposome plane.

To achieve the maximum fusion efficiency per protein construct, the best way is to establish a semi-rigid linker-based, heterodimeric protein complex in which the two connected proteins can possess affinity for different carbohydrates. This way, the two

sides of the complex would never bind to the same vesicle, nor would easily saturate all the glycans on a vesicle make it unable to link and fuse with another one.

The next step involved using coiled coils to construct the CTB_5 complex with increased rigidity to determine whether this linker could be employed by minimizing the second CTB_5 from the same protein assembly interacting with GM1 in the same membrane plane. Theoretically, the antiparallel coiled coil shall make it easier for the second CTB_5 from the same protein assembly facing towards the opposing planes, which could increase the chance of contacting any free GM1 before it is occupied by CTB_5 complexes in solution.

Chapter III Construction of coiled coil mediated CTB_5 assembly

In the last chapter, we discussed several examples of using the biotinylated $CTA2B_5$ and defined streptavidin to mediate the GUV crosslinking and membrane fusion processes. However, the PEG_4 linker used in the above cases is flexible to the extent that both CTB_5 from *cis*- and *trans*-streptavidin-$(CTA2B_5)_2$ appear to be predominantly interacting on the same membrane surface. To further understand if the orientation of carbohydrate-binding sites in CTB_5 dimer affects the ability to achieve membrane fusion, both parallel and antiparallel coiled coils were instead employed to assemble CTB_5 together with the aim they would behave as rigid linkers (Figure III-1).

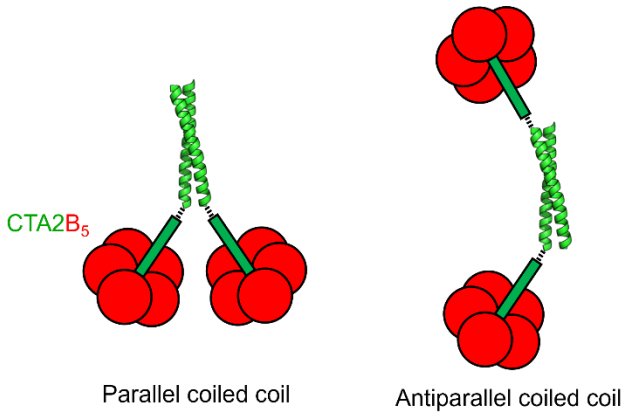


Figure III-1 The coiled coil-$(CTA2B_5)_2$ super-lectin. Protein assemblies composed of multiple $CTA2B_5$ bacterial toxin proteins using coiled coils to achieve different orientation of the carbohydrate-binding domains.

III.1 Introduction to coiled coils

The coiled coil is one of the best understood protein motifs, and the length, flexibility, oligomerization and affinity between helical pairs (*procoils*) can be modified by rational arrangement of amino acid residues. Coiled coils have been used in many applications including acting as molecular spacers, inducing membrane vesicle tethering and fusion, performing catalytic activity, providing allosteric regulation and acting as molecular scaffolds[100]. Furthermore, the incorporation of such a domain into $CTA2B_5$ involves only genetic modification, and chemical ligation is not necessary after obtaining the protein – which simplified the preparation of $CTA2B_5$ and at the same time, assembling the $CTA2B_5$ in a controlled manner.

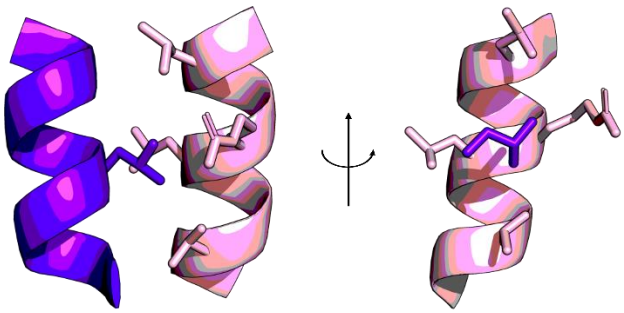


Figure III-2 An example of a parallel coiled coil (pdb: 1D7M) emphasizing the knob-into-holes interaction with a leucine knob packing into a diamond-shape hole.

Parallel dimeric coiled coil

abcdefgabcdefgabcdefgabcdefg
HAAhEEEHAAhEEEHAAhEEEHAAhEEE
abcdefgabcdefgabcdefgabcdefg
HAAhKKKHAAhKKKHAAhKKKHAAhKKK

Antiparallel dimeric coiled coil

abcdefgabcdefgabcdefgabcdefg
HAAhEEEHAAhEEEHAAhEEEHAAhEEE
KKKhAAHKKKhAAHKKKhAAHKKKhAAH
gfedcbagfedcbagfedcbagfedcba

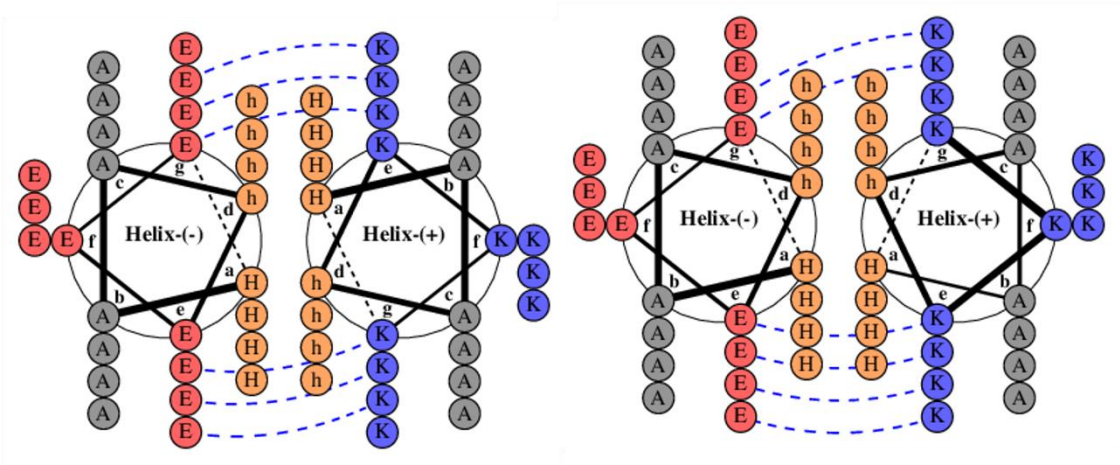


Figure III-3 Examples of helical wheel (drawn by using DrawCoil server[2] from Grigoryan lab) of parallel and antiparallel coiled coil respectively. H and h represent hydrophobic residues in *a* and *d* position separately. The rest of the letters are abbreviations of amino acids.

One characteristic of coiled coils is to pack certain side chains into gaps between other residues in the adjacent α-helix,, which is referred as “knob-into-holes” (KIH) (Figure III-2). The residues involved in this interaction are usually hydrophobic and arranged regularly every seven residues along the procoil sequence (procoil refers to the single helix before interacting with its partner to form coiled coil). This rule has been summarized by Crick[101] as a tandem “heptad repeat” pattern with seven residues as a turn labelled as $[abcdefg]_n$. The coiled coils formed according to the heptad repeat

[2] https://grigoryanlab.org/drawcoil/

pattern are generally right-handed α-helices, which are assembled into a left-handed supercoil-shaped bundle. Each α-helix has 3.6 residues in each turn while the supercoil shrinks the periodicity to 3.5 allowing the *a* and *d* positions to be aligned in accordance with each other[102]. A simplified representation for coiled coil interactions is usually shown as a helix wheel (Figure III-3), from which the main hydrophobic interaction between the *a* and *d* positions and electrostatic interactions between e and g position can be clearly analyzed.

The most frequently occurring residues at *a* and *d* positions for parallel coiled coils are N, V, L, I and L, M, respectively[103]. For antiparallel coiled coils, L, I, M, A are more commonly seen at the *a* position whereas L, Y, M, I are at the *d* position[103]. Moreover, the interaction between the *a* and *d* positions decides the orientation of coiled coil. For dimeric coiled coils, lateral hydrophobic packing (a_n-a_n') and (d_n-d_n') is essential in contributing the stability of a parallel coiled coil[104], whereas (a_n-d_n') interaction is predominant in antiparallel coiled coil[105]. The interaction of g_n-e_{n+1}' is important in facilitating parallel orientation, whereas for antiparallel pairs, the major interactions are e_n-e$_n$' and g_n-g_n'. The residues arranged at *c* and *b* positions are usually alanine, for the purpose of stabilizing the helicity of the structure. While the residues at *f* positions can be Q, K or E to facilitate the protein solubility.

In the rational design of coiled coils, the significance of noncovalent forces can hardly be overestimated. Considering that at least three or four heptads are needed to form a stable coiled coil[106], the first guiding principle is to bury the hydrophobic residues (A, L, I, M, and V) by arranging them at *a* and *d* positions and expose polar residues in the rest of the positions. The Woolfson group figured out a practical rule for designing parallel coiled coil inspired from the hydrophobic core in the GNC4-p1 leucine-zipper peptide and concluded that *a* = *d* = Ile are suitable to specify parallel trimeric coiled coils, and *a* = Leu while *d* = Ile is applicable to assembly of tetrameric coiled coils[107]. However, a single *a* = Ile and *d* = Leu rule is not enough to ensure parallel dimeric coiled coil formation without substituting a central *a* = Ile into Asn[108]. It was found by statistical analysis of the CC+ database, that E, Q, K, and R usually occur at the *e* and *g* sites flanking the core region in parallel homo-oligomers[108, 109]. Whereas R, Q, I, A, M often occur in antiparallel coiled coils from statistical analysis of natural proteins[103]. It should be emphasized that the salt bridge formed between *e* and *g* positions could also guide coiled coil orientation. In this project, the redesigned coiled coils are mostly adopting the most used Lys-Glu interaction at *e* and *g* positions regardless of whether it is for parallel or antiparallel direction.

In this chapter the two CTB_5 were linked to antiparallel or parallel coiled coils either from published sources or rational redesign, to see how the orientation of the second CTB_5 relative to the first one affects membrane fusion. The original plan was to use heterodimeric parallel and antiparallel coiled coils to allow control over the assembly process and to later enable the formation of complexes with two different types of bacterial toxin B-subunit. A characteristic shared by all assemblies was the placement of the MBP at the N-terminus (Figure III-4), followed by a flexible linker that incorporates Factor Xa and TEV cleavage sites, the *procoil*, and either a long or short version of the CTA2 peptide. CTA2 penetrates the cavity formed by CTB_5 and acts as a binder at the C-terminal end to connect CTB_5 to the procoil.

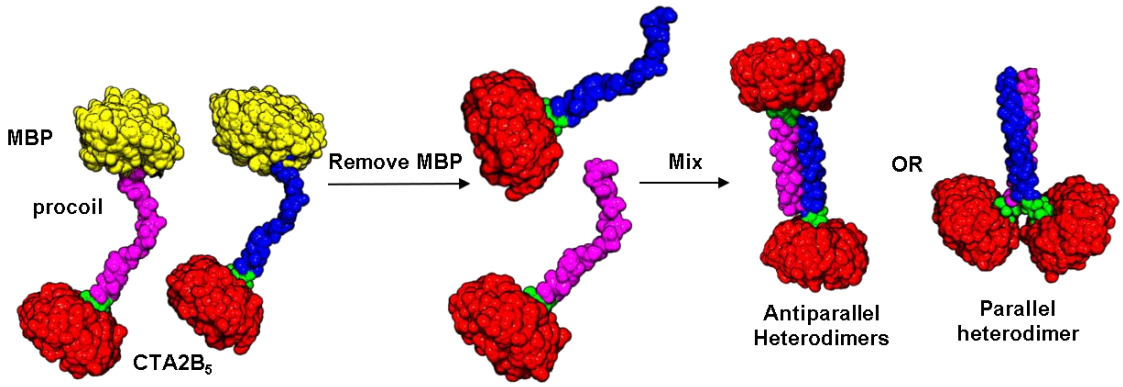


Figure III-4 The workflow to construct parallel and antiparallel heterodimers. The $CTA2B_5$ was linked with a pair of coiled coils which were designed to form either parallel or antiparallel orientation. For purification purposes, a maltose binding protein (MBP) was fused to the N-terminus of each *procoil*, and it was set to be removed by proteolysis of the TEV cleavage site between MBP and the *procoil*.

The inserted *procoil* is expected to be more prone to form a random coil rather than α-helix without an interaction with its partner. However, when the *procoil*-$CTA2B_5$ in the same pair assemble with each other, driven by the coiled coil formation, the *procoil* region would become a rigid helix (Figure III-3). While the relationship between sequence and structure for coiled coils is well understood, it is more difficult to predict whether the design rules still apply when a coiled coil is linked to a globular protein[110]. This is especially true for the parallel coiled coil assemblies, in which steric hinderance exerted by two CTB_5 at one end of the parallel assembly could interfere with the dimer formation.

III.2 Using coiled coil to construct protein assembly

Pioneered by the Yeates group[111], symmetry-based protein assembly has been further studied by multiple researchers. Some experiments have adopted coiled coils as

mediators to connect larger protein domains to build supramolecular assemblies. A heteromeric coiled coil was modified to fuse with CTB_5 in Turnbull's lab to create a tubular supramolecular complex[112], in which each coiled coil CTB_5 stacks with one another by mimicking VP1 capsid protein assembly. Similarly, the Marsh group designed an elaborated octahedral cage by combining esterase and MBP through coiled coils and spacers[113]. The same group also fused a trimeric protein (TriEst) with pentameric coiled coil to assemble virus-like particles[114].

Utilizing coiled coils to assemble multiple protein domains is a potential area. However, when a well-designed coiled coil connected with a protein, it is essential to analyze whether there would be mutual influence on orientation and oligomerization between the coiled coil and the protein, which might lead to the reorganization of structure and produce undesired interactions. Marsh *et al.* considered this point, and designed experiments to detect whether the oligomerization state of the coiled coil follows the theoretical oligomerization state and orientation after fusing with green fluorescent protein (GFP)[110]. It was found that large protein complexes did affect the structure of the coiled coil and only two out of six tested coiled coils were successfully formed in the predict manner. Then the authors made several modifications on the number of coiled coil heptad and length of spacer to enable the desirable interaction to happen by relieving the steric hindrance from GFP globular proteins. That provides us a hint to further discuss the difference between predict and practical results when we are trying to analyzing the coiled coil-$(CTA2B_5)_2$ construct.

III.3 Selection of coiled coil

The majority of the coiled coils used in initial studies came from the published protein sequences, with a small number coming from my rational design. The full-length CTA2 peptide was used for building parallel-oriented coiled coils, since some flexibility is needed reduce the steric hindrance brought on by placing two CTB_5 on the same direction. For the antiparallel coiled coil, however, a shorter CTA2 peptide was used and only the residues from the C terminus of CTA2 peptide required for interaction with the CTB_5 domain were retained. The two *procoils* that make up a coiled coil usually have different protein sequences and are designated *x.1* and *x.2* respectively. As start, we selected two parallel coiled coil and two antiparallel coiled coil, which are stated as PP1.x, PP2.x, AP1.x and AP2.x (Figure III-5). So, as an example, consider the PP1.x

coiled coil, which is made up of the PP1.1 and PP1.2 *procoils*. The complete protein sequences are included in the Appendix VIII.1.1.b.

Ref	Pair	Protein sequence *abcdefg abcdefg abcdefg abcdefg abcdefg abcdefg abcdefg*	Note
[115]	PP1.x	PP1.1 SPEDE IQALKEK NAQLKEK NQALEEK IAQLKEK IQALKEK PP1.2 SPEDE IQALEKE NAQLEKK NQALEKE IAQLEKE IQALEKE	Protein sequences were cited as in original publication.
[120]	PP2.x	PP1.1 NL VAQLENE VASLENE NETLKKK NLHKKDL IAYLEKE IANLRKK IEE PP1.2 AR NAYLRKK IARLKKD NLQLERD EQNLEKI IANLRDE IARLENE VASHEQ	

Ref	Pair	Protein sequence *abcdefg abcdefg abcdefg abcdefg abcdefg abcdefg abcdefg*	Note
[118]	AP1.x	AP1.1 SPEDE LAQLEKE LQAIEKK LAQLEKK AQAREKE LAQLEKE AP1.2 KEK KEKLAQL KEKAQAR KEELAQL KEELAQI KEELAQL SP EDELAQL	Redesigned. Original sequences are listed below as AP1.x (original) and AP2.x (original) seperately
[122]	AP2.x	AP1.1 QLEKE LAQLKEK LQAIEKE LAQLKEK AQALKEK LAQLKEK LQ AP1.2 QŌ EKELAQL EKELAQI EEKLAQL EKEAQAL KKKLAQL EKELQ	
		gfe dcbagfe dcbagfe dcbagfe dcbagfe dcbagfe dcba	

Ref	Pair	Protein sequence *abcdefg abcdefg abcdefg abcdefg abcdefg abcdefg abcdefg*
[118]	AP1.x (original)	AP1.1 (o) AQLEKE LQALEKK LAQLEWE NQALEKE LAQ AP1.2 (o) AQL KKKLQAN KKELAQL KWKLAOL KKKLAQ
[122]	AP2.x (original)	AP1.1 (o) QLEKE LAQLKKK LQAIEKE LAQLKWK LQALKKK LAQLKEK LQ AP1.2 (o) QŌ EKELAQL EKELAQI EWEAQAL EKKLAQL EKELQ
		dcbagfe dcbagfe dcbagfe dcbagfe dcbagfe dcbagfe dcba

Figure III-5 Protein sequences for the chosen heterodimeric parallel and antiparallel coiled coils (PP1.x, PP2.x, AP1.x, AP2.x) that were used to construct CTB_5 dimer complexes. Since AP1.x and AP2.x were redesigned, the original published coiled coil sequences, AP1.x (original) and AP2.x (original), were also listed.

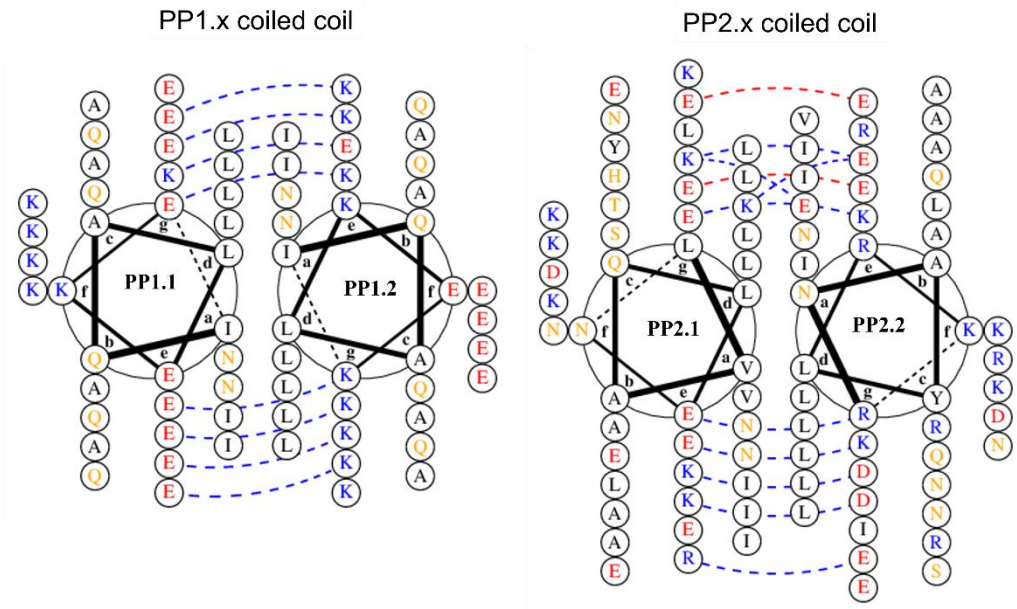


Figure III-6 The helix wheel indicating the formation of two parallel-oriented coiled coil PP1.x (composed of PP1.1 and PP1.2) and PP2.x (made up of PP2.1 and PP2.2).

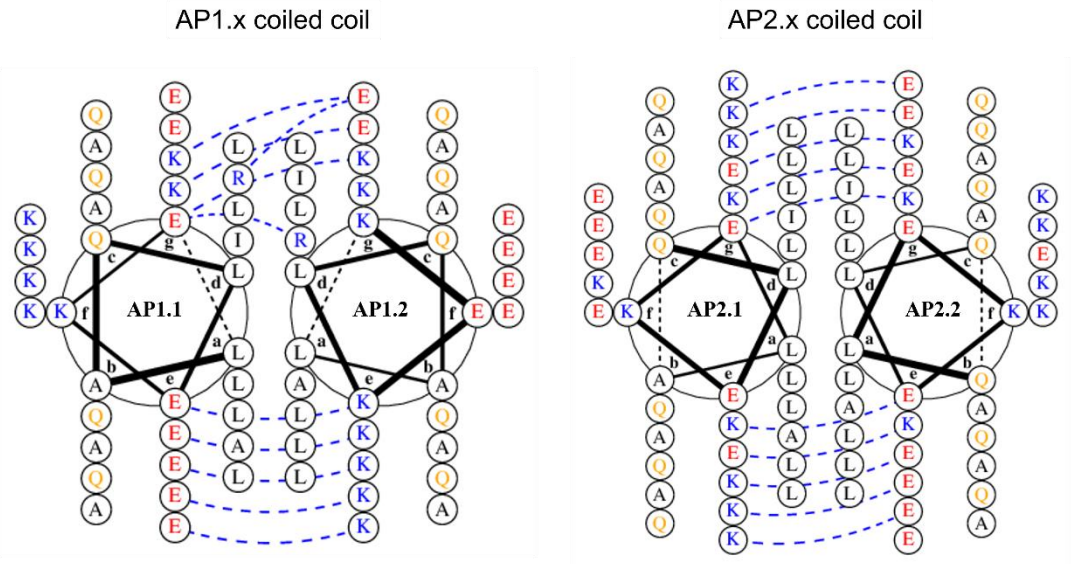


Figure III-7 The helix wheel indicating the formation of two antiparallel-oriented coiled coil PP1.x (composed of PP1.1 and PP1.2) and PP2.x (made up of PP2.1 and PP2.2).

PP1.x (PP1.1:PP1.2) (helix wheel shown as in Figure III-6, Left) is derived from a parallel heterodimeric coiled coil that was designed by the Jerala group[115]. It adopts a parallel orientation by arranging *a* = Ile and *d* = Leu and converting two adjacent Ile[116] in the *a* position to Asn[117, 118], which was shown to produce a more stable coiled coil interaction[115]. Furthermore, a tetrapeptide Ser-Pro-Glu-Asp (SPED) was added as N-cap extension[115], which avoids the untangling of the helix after coiled coil is formed (Figure III-5). Position *b*, *c* and *f* are occupied with Gln, Ala, Lys and Glu which have high helix-forming propensity[119], that could facilitate solubility, and prevent homo-oligomerisation and neutralizing the net charge from *e* and *g* position. The second parallel coiled coil PP2.x (PP2.1:PP2.2) comes from the SYNZIP protein interaction toolbox[120] (Figure III-6, Right), and the Goldman group has used it for constructing protein scaffolds based on parallel heterodimeric coiled coils[121]. The affinity of SYNZIP1-2 pairs had been reported as <10 nM[121], hence it would be ideal for keeping the CTB_5 as a dimer under low concentration range.

Antiparallel AP1.x (AP1.1:AP1.2) (Figure III-7, Left) is derived from a combination of antiparallel heterodimer designed by Oakley[118]. However, the residues arranged in hydrophobic *a-d* positions were modified and an antiparallel orientation was designed to be ensured by the R-I-A (Arginine-Isoleucine-Alanine) combination at the same layer of *a* and *d* sites. The other factor that helps control the orientation is the electrostatic arrangement in *e-g* positions, most of which would turn into repulsion if it adapted a parallel orientation. The second antiparallel coiled coil AP2.x (AP2.1:AP2.2) (Figure III-7,

Right) is a rationally redesigned protein based on a sequence from the Marsh group[122], and the Isoleucine-Alanine arrangement in *a* and *d* region was kept. Instead, I modified the residues in *e*, *g* and *f* position to avoid the accumulation of lysine residues. This was done to prevent the interaction of lysine with mRNA at the ribosomal exit tunnel from impairing protein expression[123].

The gene sequences for the AP1.1-sCTA2 (short CTA2), AP1.2-sCTA2, AP2.1-sCTA2, AP2.2-sCTA2, PP1.1-CTA2 (full CTA2), PP1.2-CTA2, PP2.1-CTA2, PP2.2-CTA2 are in Appendix VII.1.1.b. DNA fragments of selected *procoil*-(s)CTA2 were ordered from TWIST Bioscience, and standard molecular biology procedures were performed to incorporate the *procoil*-(s)CTA2 into pTRBAB5-G1S plasmid[80] as a replacement for the original CTA2 peptide, hence to allow expression of MBP-*procoil*-(s)CTA2-CTB_5 proteins. The MBP-(s)$CTA2B_5$ protein was found be exported to the extracellular media. However, to get the maximum yield, the periplasmic extract was also collected for purification.

The four antiparallel *procoil* constructs were expressed first and the N-terminal MBP tag allowed purification on an amylose resin. 20% α-D-glucose buffer was used to elute the protein instead of using maltose. The weaker binding affinity of α-D-glucose for MBP allowed the ligand to be more easily removed by dialysis so that the MBP tag could be used in a future purification step when MBP needs to be removed following protease digestion. The elution from the amylose column was applied to nickel resin, as the five surface-expose H13 residues in CTB pentamer allows affinity for the nickel resin without the necessity of having a His tag. After elution with 250 mM imidazole, the final yields were 34 mg/L, 0.7 mg/L, 35 mg/L and 55 mg/L for MBP-AP1.1, MBP-AP1.2, MBP-AP2.1, MBP-AP2.2, respectively. Three out of four proteins expressed well, with yields 3-5 times higher than for MBP-$CTA2B_5$. However, it is unclear whether the high yield is due to the characteristic of the protein itself or the fact that both media and periplasmic extract were used to isolate proteins as only the media is processed in our standard protocol for MBP-$CTA2B_5$. Since the number of lysine residues of the inserted AP1.2 *procoil* was well balanced with Glu and the net charge of the inserted helix-A2 region was -0.764 (calculated by Prot pi | Protein Tool[3]), any unfavourable interaction with the ribosome exit tunnel due to positive charge of expressed protein might not be the main reason to explain the low yield. The purified proteins were analyzed by SDS-PAGE (Figure III-8, a

[3] https://www.protpi.ch/Calculator/ProteinTool

and b). Considering the poor expression of AP1.2, this pair of proteins (MBP-AP1.1, MBP-AP1.2) was suspended for further research.

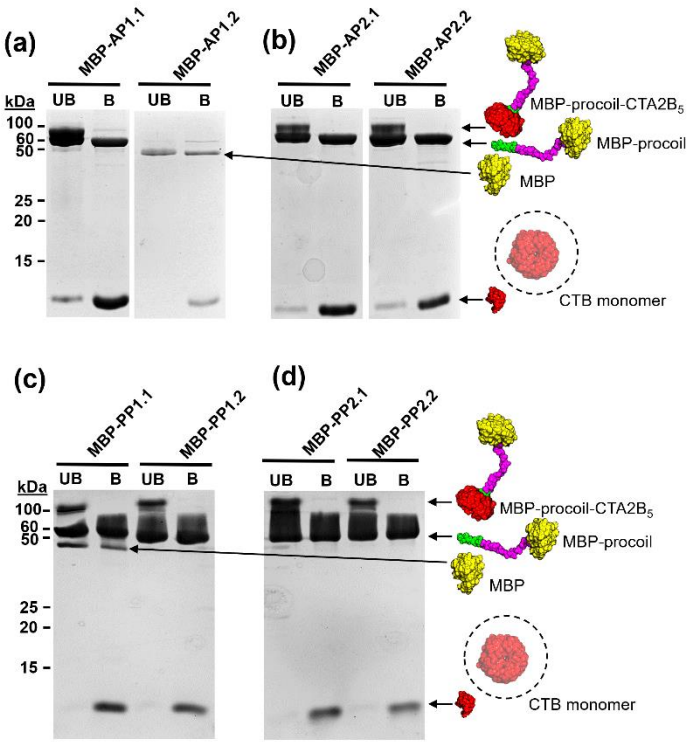


Figure III-8 Coomassie stained SDS-PAGE analysis of ***A.*** MBP-AP1.1-sCTA2B$_5$ (MBP-AP1.1) and MBP-AP1.2-sCTA2B$_5$ (MBP-AP1.2), ***B.*** MBP-AP2.1-sCTA2B$_5$ (MBP-AP2.1) and MBP-AP2.2-sCTA2B$_5$ (MBP-AP2.2), ***C.*** MBP-PP1.1-CTA2B$_5$ (MBP-PP1.1) and MBP-PP1.2-CTA2B$_5$ (MBP-PP1.2), ***D.*** MBP-PP2.1-CTA2B$_5$ (MBP-PP2.1) and MBP-PP2.2-CTA2B$_5$ (MBP-PP2.2) after purification on amylose and nickel resins. Upon boiling (B) of the samples before loading onto the gel, bands corresponding to the CTB monomer and MBP-procoil-CTA2 are observed. Conversely, unboiled (UB) samples partially dissociate on the gel, but also display bands at higher molecular weight corresponding to the complete MBP-procoil-CTA2B5 complex.The MBP-*procoil*-(s) CTA2 portion is around 51.7-54.8 kDa, and the band could be found in all four protein samples with different intensity. When the MBP-*procoil*-(s) CTA2 interacts with CTB$_5$, the overall Mw of the complex reaches 110-112 kDa.

A similar expression and purification method was applied to the four parallel coil constructs: MBP-PP1.1-CTA2B$_5$ (MBP-PP1.1), MBP-PP1.2-CTA2B$_5$ (MBP-PP1.2), MBP-PP2.1-CTA2B$_5$ (MBP-PP2.1), MBP-PP2.2-CTA2B$_5$ (MBP-PP2.2) yielding only 3.8 mg/L, 3.3 mg/L, 2.3 mg/L and 3 mg/L, respectively (Figure III-8, C and D). Electrospray mass spectrometry (ESMS) was used to confirm that the purified products were the correct proteins. Peaks corresponding to the correct size of the different components of MBP-AP2.1, MBP-AP2.2, MBP-PP1.1, MBP-PP1.2, MBP-PP2.1 and MBP-PP2.2 were found after deconvolution (Figure III-9).

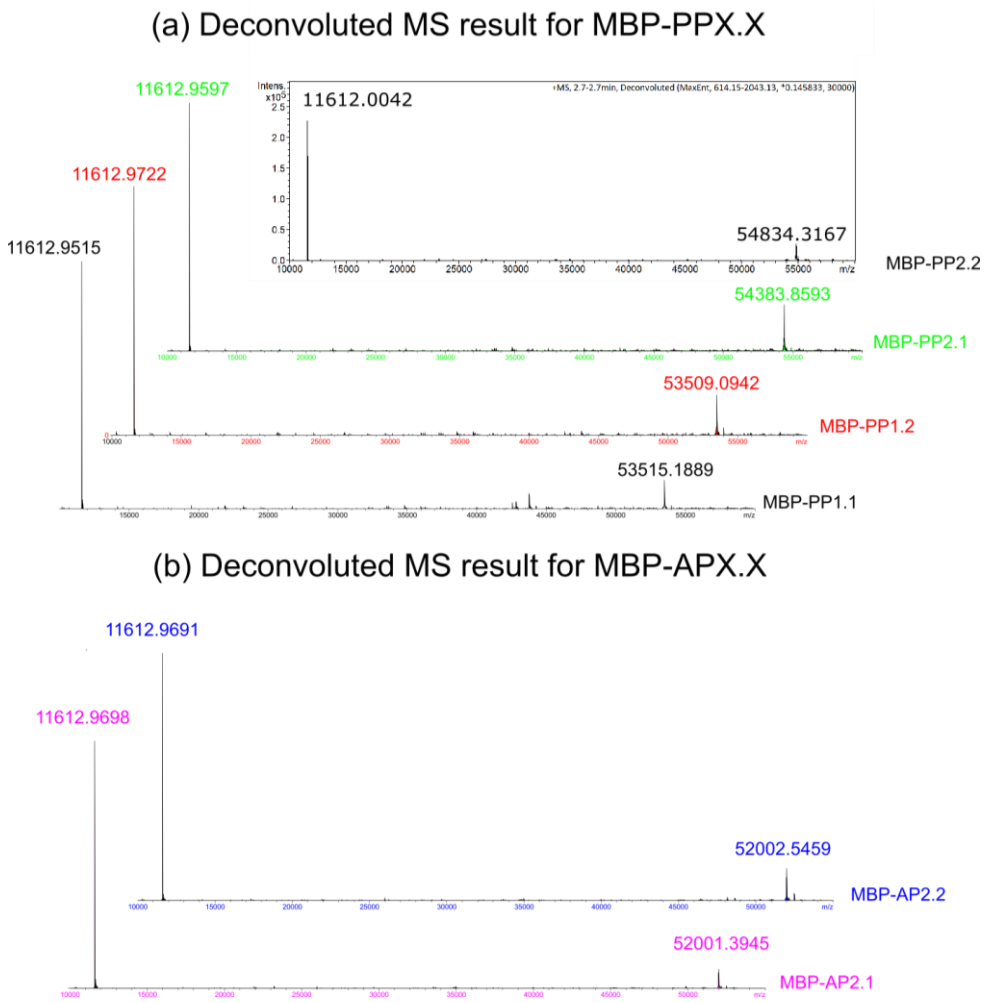


Figure III-9 (a) Deconvolved MS results for MBP-PPX.X proteins. (b) Deconvolution MS results for MBP-AP2.1 and MBP-AP2.2 proteins. The deconvolution results indicate the presence of two major domains within the protein complex. The first domain, consisting of MBP-procoil-CTA2, has an estimated molecular weight of approximately 50 kDa. The second domain comprises the CTB monomer, with a molecular weight of 11612 Da.

Additionally, we can see in the MS that the molecular masses of MBP-AP2.1 and MBP-AP2.2 are remarkably similar. It is challenging to determine whether a homodimer or a heterodimer form in subsequent experiments due to the similarity of the molecular weight of this collection of proteins through either SDS-PAGE or MS, which is a useful lesson to remember.

III.4 Oligomerization state of individual MBP-PP/MBP-AP proteins

III.4.1 Size exclusion chromatography

Prior to combining complementary pairs of proteins, purified individual proteins were analyzed by size-exclusion chromatography to determine their level of oligomerization (Figure III-10). Although proteins with similar molecular weights, such as MBP-PP2.1 and MBP-PP2.2, showed some variation in retention times, one pair of proteins (MBP-AP2.1 and MBP-AP2.2) eluted much earlier than the others even though their molecular weights were 1-3 kDa smaller than the others. We hypothesized that small variations in retention times might be caused by the different helicity of each inserted *procoil*. However, it is more likely that the MBP-AP2.1 and MBP-AP2.2 were forming dimers or even a higher oligomeric state when the differences in retention volume are large, approaching 3 mL. A portion of the protein samples was submitted for SEC-MALS analysis to confirm or refute this hypothesis.

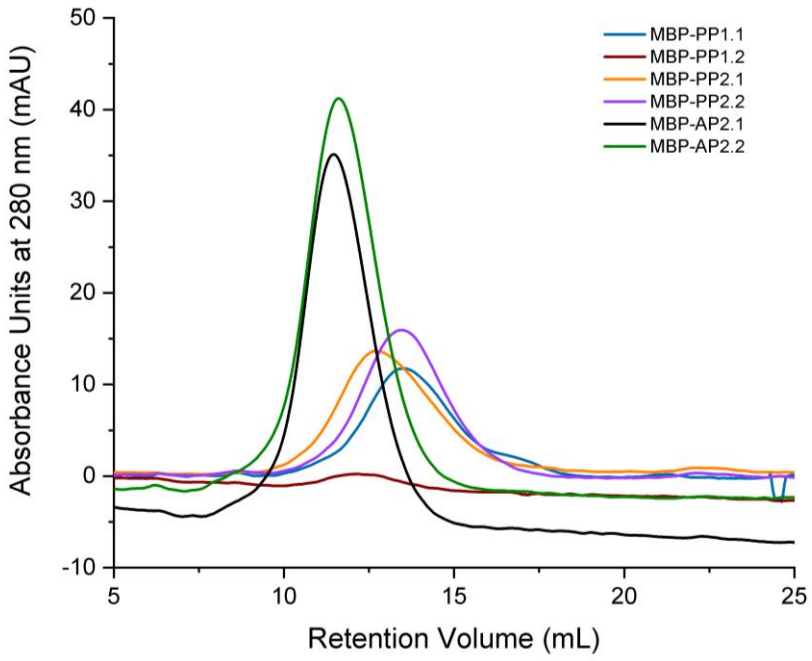


Figure III-10 Overlay of size exclusion chromatograms (SD200 10/200) of MBP-PP1.1 (111.5 kDa), MBP-PP1.2 (111.5 kDa), MBP-PP2.1 (112.4 kDa), MBP-PP2.2 (112.8 kDa), MBP-AP2.1 (110 kDa), MBP-AP2.2 (110 kDa), with the elution volume as 13.5 mL, 12.27 mL, 12.75 mL, 13.5 mL, 11.43 mL, and 11.59 mL, respectively. The manifest shift of MBP-AP2.1 and MBP-AP2.2 indicates the possible formation of the higher oligomers.

III.4.2 SEC-MALS

To get a more accurate prediction of the oligomerization state of the individual proteins, size exclusion chromatography coupled with multiple-angle light scattering (SEC-MALS) was

used by Dr. Maria Nikolova (University of Leeds) to analyze individual proteins of MBP-CTA2B_5 (purified by MChem student Andrew Boswell), MBP-PP1.1, MBP-PP1.2, MBP-PP2.1, MBP-PP2.2, MBP-AP2.1, and 1:1 mixture of MBP-PP2.1 and MBP-PP2.2. The proteins were loaded into SDS-PAGE to check the purity just before being sent for analysis (Figure III-11).

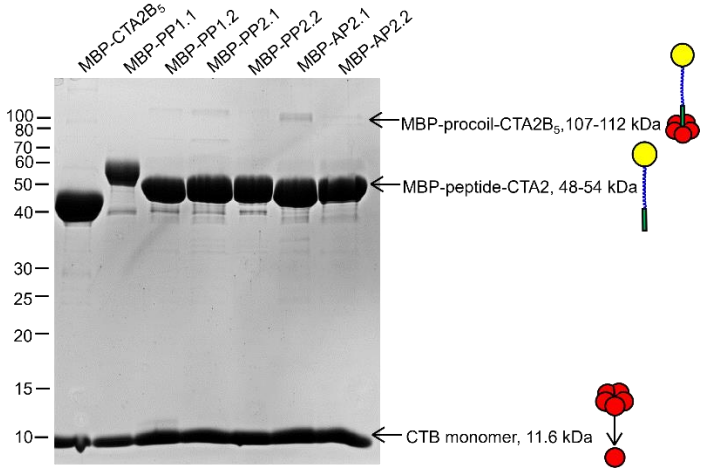
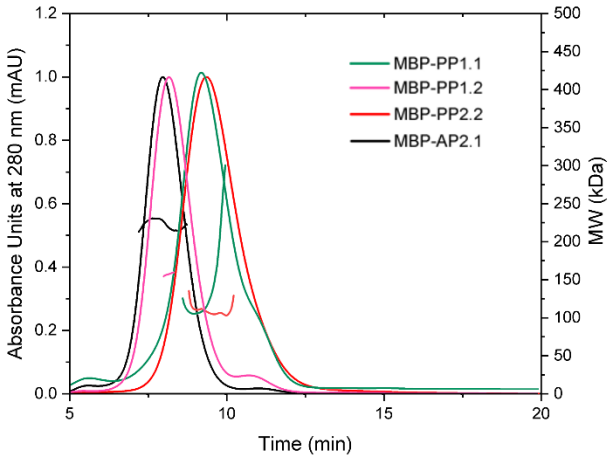


Figure III-11 Coomassie stained SDS-PAGE of the purified proteins (boiled before loading) which were used in SEC-MALS analysis.



Protein	Molecular weight	Mw/Mn
MBP-PP1.1	102 kDa ± 18%	1.022
MBP-PP1.2	157 kDa ± 5%	1.000
MBP-PP2.2	109 kDa ± 19%	1.002
MBP-AP2.1	223 kDa ± 6%	1.001

Figure III-12 SEC-MALS analysis of MBP-PP1.1, MBP-PP1.2, MBP-PP2.2 and MBP-AP2.1 for which the theoretical monomer molecular weights are 111499 Da, 111496 Da, 112822 Da, and 109987 Da respectively. The chromatography displays the UV absorbance at 280 nm curve (continuous lines) together with the molecular weight (MW; discontinuous lines) calculated by MALS (polydispersity < 1.004).

Based on the fitting data from SEC-MALS, the calculated molecular weights and polydispersities (Mw/Mn) for each protein are shown in Figure III-12. According to the

molecular weight determined by SEC-MALS analysis, MBP-AP2.1 forms a dimer, whereas MBP-PP1.1 and MBP-PP1.2 are monomers. The oddity is that MBP-PP2.2 is situated between the monomer and the dimer, most likely as a result of monomer-dimer equilibrium (Figure III-12).

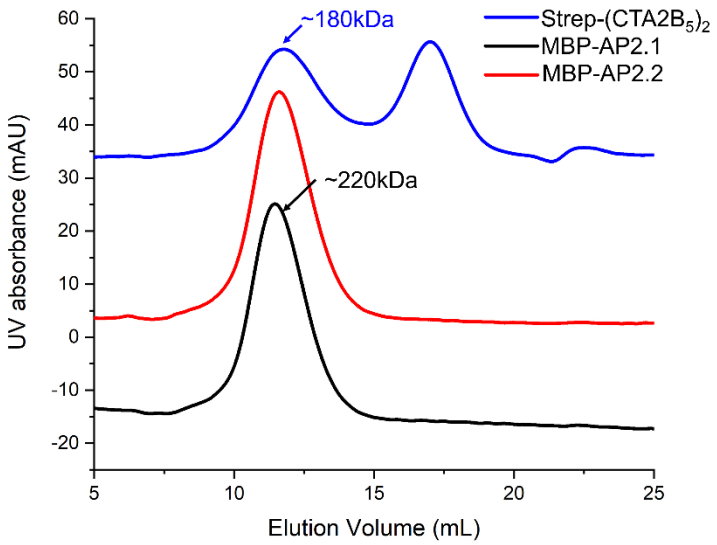


Figure III-13 Size-exclusion chromatography result (using Superdex 200 10/300 GL) of a mixture composed of biotin-CTA2B_5 and streptavidin-(CTA2B_5)$_2$ (blue), MBP-AP2.1 (black) and MBP-AP2.2 (red).

The retention volume of MBP-AP2.2 was directly compared to that of MBP-AP2.1 and Strep-(CTA2B_5)$_2$ to confirm the oligomerization state after the size exclusion chromatogram of MBP-AP2.2 was obtained in our lab using a SD200 10/100 column(Figure III-13). Strep-(CTA2B_5)$_2$, which is known to have an average molecular weight of at least 180 kDa[80], was found to be comparable to the values for MBP-AP2.1 and MBP-AP2.2, which both eluted with similar retention (11.4 mL and 11.6 mL, respectively). Hence, we deduce that MBP-AP2.2 is also a dimer given the SEC-MALS results demonstrating MBP-AP2.1 is a dimer.

However, this created a conundrum because MBP-AP2.1 and MBP-AP2.2 were initially designed as a pair of coiled coils that could interact with one another to create a heterodimer. Since there is no effective experimental technique to identify and separate the potential heterodimer from the homodimer because the MBP-AP2.1 and MBP-AP2.2 have very similar molecular weights, only MBP-AP2.1 was selected for this particular group of coiled coils to further study as a putative antiparallel homodimer.

III.5 Atom force microscopy (AFM) of MBP-AP2.1

Since the MBP-AP2.1 unexpectedly formed a dimer, we intended to explore the orientation of the CTB_5 units using atomic force microscopy (AFM). AFM can produce a high-resolution image of the three-dimensional shape (topography) of the sample surface by measuring the response of the force applied by the probe to the sample. This is accomplished by recording the height of the probe that corresponds to a constant probe-sample interaction, as the tip is scanned across the sample. Therefore, important information about the surface topology can be extracted and reconstructed as a pseudocolour image.

Preliminary AFM experiments were performed by Dr. George Heath (University of Leeds), who examined the topological morphology of MBP-AP2.1. He discovered that MBP-AP2.1 has a cross shape resembling a dumbbell with dimensions of 15–25 nm in length, 15–20 nm in width, and 2.2–4.4 nm in height (Figure III-14-a). Given that MBP and CTB_5 have dimensions of approximately 7 x 5 x 5 nm and 3.5 x 6.5 x 6.5 nm, respectively, the two protein domains with the 4.3 nm height may be CTB_5, while the two proteins with the smaller dimensions are MBP. Figure III-14-B depicts a different field of view in which two subunits on each side have heights of 3 nm versus 2.5 nm. Taken together with the evidence in Figure III-14-A, if these features are CTB_5 and MBP, respectively, the homodimer might adopt an antiparallel orientation. However, there is still insufficient information to conclude definitively the orientation of the coiled coil in the homodimer MBP-AP2.1.

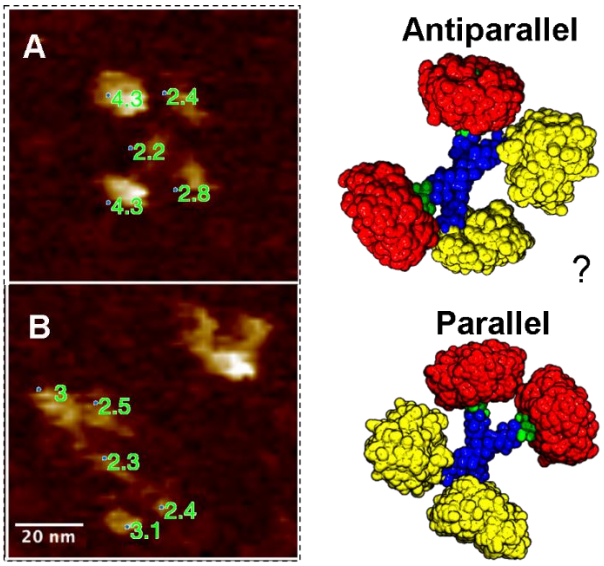


Figure III-14 Topographic AFM images of MBP-AP2.1 assembly into homodimer. (A) The homodimer forms a crosslink with no apparent coiled coil orientation. (B) A confirmation of MBP-AP2.1 homodimer might give a hint of the antiparallel orientation of the coiled coil.

One method to remove ambiguity from the imaging would be to cut off the MBP and then re-examine the topology by AFM to confirm the AP2.1 homodimer orientation. This would prevent any MBP, that was comparable in size to CTB_5, from becoming a problem. The above-mentioned experiments are detailed in section IV.8 alongside the other proteins.

III.6 Attempts to remove MBP tags

Since MBP was only included in the protein in order to perform purification, removing it from the purified proteins would provide a simpler protein and assist in determining how CTB_5 dimer conformations and membrane fusion relate to one another. MBP-TEV protease was first investigated for cleaving the MBP-*procoil*-(s)$CTA2B_5$ proteins because the protein sequence contains a pre-defined TEV cleavage site (Figure III-15).

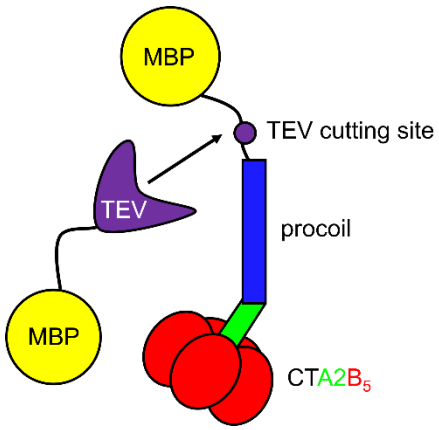


Figure III-15 Scheme illustration of using MBP-TEV protease to digest the TEV cutting site of MBP-*procoil*-(s)$CTA2B_5$.

III.6.1 TEV cleavage test

After the four complexes with parallel coiled coil sequences (MBP-PP1.1, MBP-PP1.2, MBP-PP2.1, and MBP-PP2.2) were expressed and purified, MBP-TEV protease (5 mol% or 10 mol%) was added either at room temperature or 4 ºC to check the optimal conditions required for TEV cleavage. All the proteins were processed and loaded in a 12% SDS-PAGE gel (Figure III-16). The enzyme efficiency was low for the MBP-PPx proteins. Complete cleavage of all four proteins was only achieved when the incubation time was extended from 2h to overnight at r.t. with at least 5 mol% MBP-TEV protease. Remembering that MBP-PP1.1 and MBP-PP2.2 are oligomers while MBP-PP1.2 and MBP-PP2.1 show a certain

degree of assembly (section III.4.2), it is possible that the inserted *procoil* – no matter if its conformation is random coil or helix – has made the attack of MBP-TEV protease toward the cutting site more difficult.

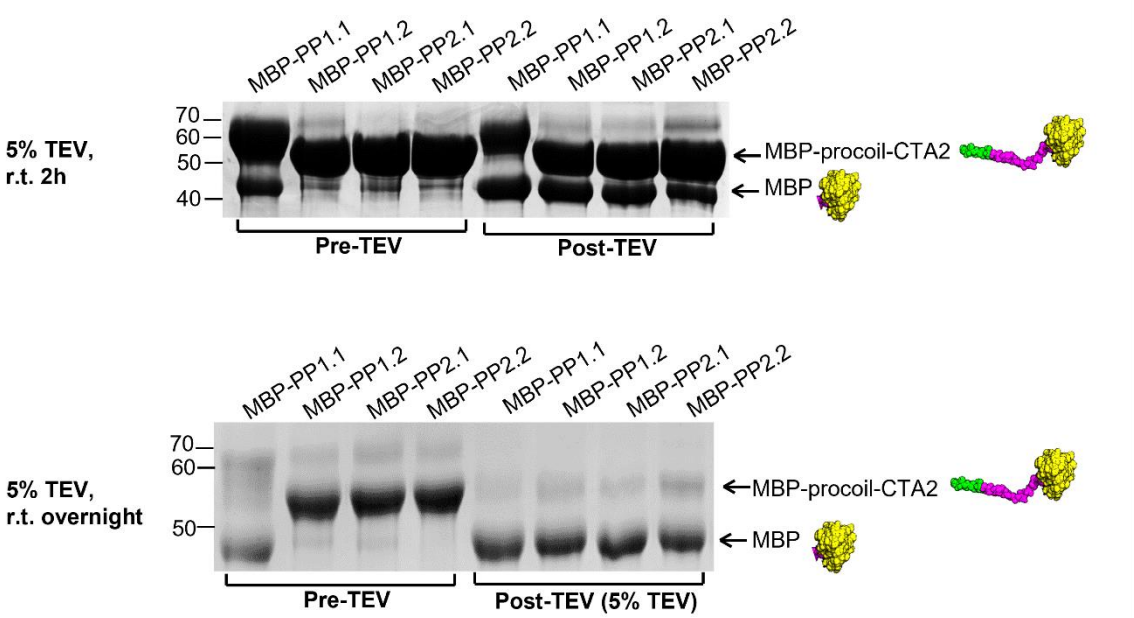


Figure III-16 A Coomassie-stained SDS-PAGE showing the results of TEV cleavage of MBP-PP1.x and MBP-PP2.x. Various conditions were evaluated, including the addition of 5% TEV protease and incubation at room temperature for either 2 hours or overnight. Prior to loading onto the gel, the protein samples were subjected to boiling at 95ºC for 10 minutes.

The reaction conditions that could completely cleave the MBP-PPx protein did not, however, result in complete cleavage of MBP-AP2.1 and MBP-AP2.2 (Figure III-17). Only a small portion of MBP-AP2.1 and MBP-AP2.2 was cleaved even with 10 mol% TEV protease, r.t. overnight conditions, whereas the majority of the MBP-*procoil*-sCTA2 fraction remained unaltered.

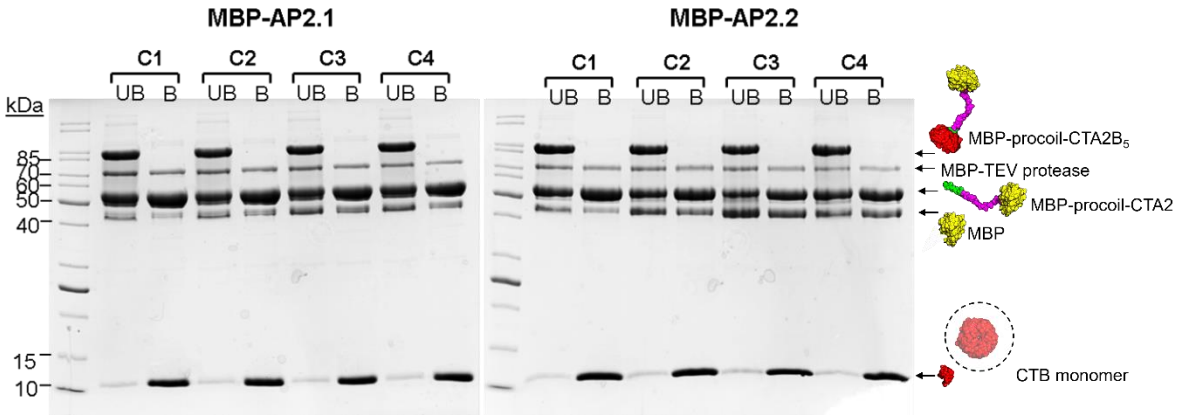


Figure III-17 The TEV cleavage efficiency of MBP-AP2.1 and MBP-AP2.2 was evaluated using Coomassie-stained SDS-PAGE under four conditions. Condition 1 (C1) involved incubating the samples with 5 mol% MBP-TEV protease at room temperature for 2 hours. Condition 2 (C2) involved incubation with 5 mol% MBP-TEV protease at room temperature for 2 hours, followed by overnight incubation at 4°C. Condition 3 (C3) involved overnight incubation at room temperature with 5 mol% MBP-TEV protease. Condition 4 (C4) involved overnight incubation at room temperature with 10 mol% MBP-TEV protease. Each sample was tested with and without boiling prior to gel loading. The boiled samples showed bands corresponding to the CTB monomer and MBP-procoil-CTA2, while the unboiled samples partially dissociated on the gel and showed bands for the full MBP-procoil-CTA2B_5 complex.

Combining the findings from SEC, SEC-MALS, and AFM, we can conclude that MBP-AP2.1 and MBP-AP2.2 undergo homodimerization. As a result, the TEV cleavage site between MBP and the *procoil* may become buried or highly helical because of homodimer formation, making it difficult for the MBP-TEV protease to either reach the digestion site or to complete the digestion.

III.6.2 MBP-AP2.1 and MBP-AP2.2 Factor Xa protease cleavage assay

In order to expose more of the TEV cutting site in solution, it may be necessary to add a flexible linker between the TEV cutting site and the coiled coil region in the case of MBP-AP2.1 and MBP-AP2.2. It may become more rigid if the original TEV cutting site is too close to the coiled coil; while this may not completely thwart the TEV protease attack, it may reduce the cutting efficiency.

The second method involves using the Factor Xa cutting site, which is located six residues further away from the coiled coil region than the TEV cutting site (Figure III-18). The Ile-Glu/Asp-Gly-Arg sequence can be recognized and cleaved by Factor Xa protease which is also a site-specific endoprotease like MBP-TEV protease.

Figure III-18 The protein sequence between MBP and AP2.1 *procoil*-sCTA2B5. The Factor Xa cutting site is colored blue; TEV cleavage site is purple; the *procoil* of AP2.1 is shown as a blue helix; and the CTA2 portion as a green helix. The MBP and CTB_5 are yellow and red, respectively.

According to New England Biolabs' recommendation (available from [80]), the target protein concentration should be at least 1 mg/mL to achieve a clean cut. A variety of conditions were tested to prevent non-specific proteolysis brought on by too much Factor Xa protease and/or too long an incubation time. First, Factor Xa protease (1.2 µL, 0.6 µL and 0.3 µL, 1 mg/mL) was mixed with MBP-AP2.1 (5 µL, 2.4 mg/mL) and incubated at 25 °C overnight. MBP-TEV protease (0.2 µL, 3.5 mg/mL) was added to another tube of MBP-AP2.1 as a negative control, and the reaction was allowed to proceed overnight at 25 °C. In order to check for cleavage, samples were analyzed by SDS-PAGE (Figure III-19). Complete cleavage was observed when 0.6 µL of Factor Xa protease were added, but when only 0.3 µL was added, MBP-*procoil*-sCTA2 was still detected as a faint band. As the Factor Xa protease is smaller than the MBP-TEV protease previously used (37 kDa vs. 95 kDa), the higher cleavage efficiency may be due to the smaller size of the Factor Xa protease or the greater distance between the procoil and the cleavage site, allowing Factor Xa to cleave more efficiently than the MBP-TEV protease.

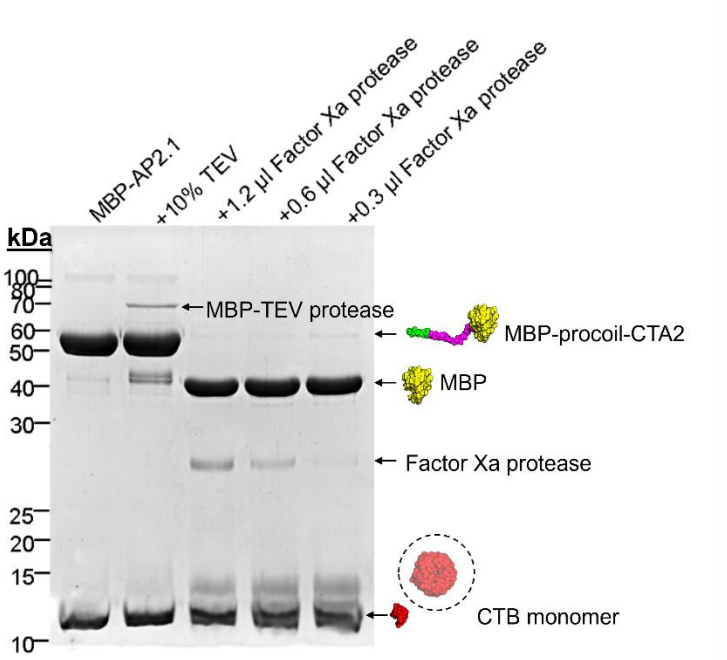


Figure III-19 To assess the efficiency of cleavage of MBP-AP2.1, Coomassie-stained SDS-PAGE was employed, with samples treated with either MBP-TEV protease or Factor Xa protease at various concentrations. Before analysis, all samples were boiled at 95°C for 10 minutes.

While cleaving the protein with Factor Xa can allow removal of MBP, it would still be preferable to cut at the TEV site to expose the SVG at the N-terminus of *procoil*-CTA2B_5 to serve as a site for oxime ligation in case it is needed for specifically linking to fluorophores. Therefore, we also aimed to confirm if MBP-TEV protease attack could occur at its cutting site after MBP removal. Therefore, Factor Xa protease (0.3 µL, 1mg/mL) was added to MBP-AP2.1 (5 µL, 2.4 mg/mL) and incubated at 25 °C overnight, then MBP-TEV protease (0.2 µL, 3.5 mg/mL) was added to react overnight at the same temperature. The reaction was monitored by ESMS (Figure III-20). If the TEV cutting site can be attacked by TEV protease, a peak at 8232 Da would be detected.

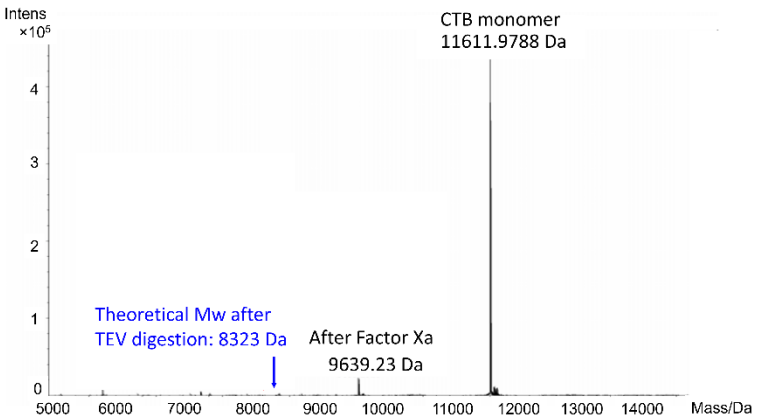


Figure III-20 Deconvolved mass spectrum following the cleavage test using both Factor Xa protease and TEV protease on MBP-AP2.1. The peak at 9640 Da implies only Factor Xa cutting without a further TEV protease cleavage.

However, the anticipated molecular weight (8232 Da) observed following treatment with TEV protease was not found (Figure III-20). This suggests that although Factor Xa protease can successfully digest MBP-AP2.1, further TEV protease attack is still obstructed. One possibility is that the TEV cleavage site changes conformation and becomes part of the coiled coil. Incorporation of a flexible linker between the TEV cleavage site and the *procoil* sequence could interfere with helix formation and allow cleavage to occur without difficulty. Therefore, a flexible 5-glycine linker was introduced immediately following the ENLYFQ-SVG sequence and prior to the initial *procoil*.

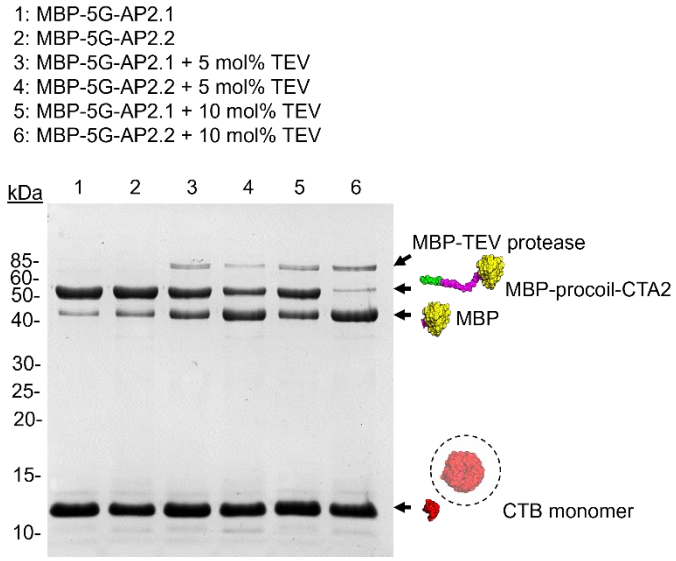


Figure III-21 The results of the proteolysis of MBP-5G-AP2.1 and MBP-5G-AP2.2 using varying concentrations of TEV protease were analyzed by Coomassie stained SDS-PAGE. The protein mixtures were incubated overnight at room temperature with either 5 mol% or 10 mol% MBP-TEV protease. Prior to gel loading, all samples were subjected to boiling at 95 ºC for 10 minutes.

Five glycine residues were introduced into the sequences of both MBP-AP2.1 and MBP-AP2.2 through Q5 mutagenesis[124]. Following expression and purification of MBP-5G-AP2.1 and MBP-5G-AP2.2, the proteins were mixed with MBP-TEV protease (5 mol% or 10 mol%) and allowed to incubate overnight under *r.t.* before analysis of the digestion by SDS-PAGE (Figure III-21). In comparison to the MBP-AP2.1 (Figure III-17), the MBP-5G-*procoil*-sCTA2 subunits showed higher proportions of digestion and nearly complete digestion of MBP-5G-AP2.2 (Figure III-19). This experiment provides evidence that the TEV cutting site has undergone secondary structure changes or steric blocking from *procoil* and it is most

likely the main cause of its resistance to TEV protease cleavage. Moreover, this phenomenon allows us to infer some details about the structure of the coiled coils in AP2.1 homodimers.

First of all, we can deduce whether it is possible for AP2.1 to form a parallel orientation (Figure III-22). The TEV cutting sites are equally inaccessible which implies any arrangement of the coiled coils must be symmetrical (Figure III-22-a). The helix wheel indicates that a generally stable salt bridge formation is possible in this scenario (Figure III-22-b), and CCBuilder[125] calculations provide a Bristol University Docking Engine (BUDE) energy in the coiled coil region (including the TEV cutting site) of -164.9 which indicates a favorable coiled coil formation. However, due to the short CTA2 sequence used in this structure, a significant amount of deformation in CTA2 region is needed to accommodate the two CTB_5 units in close proximity (Figure III-22-c). Therefore, the overall stability of the protein assembly may not be guaranteed.

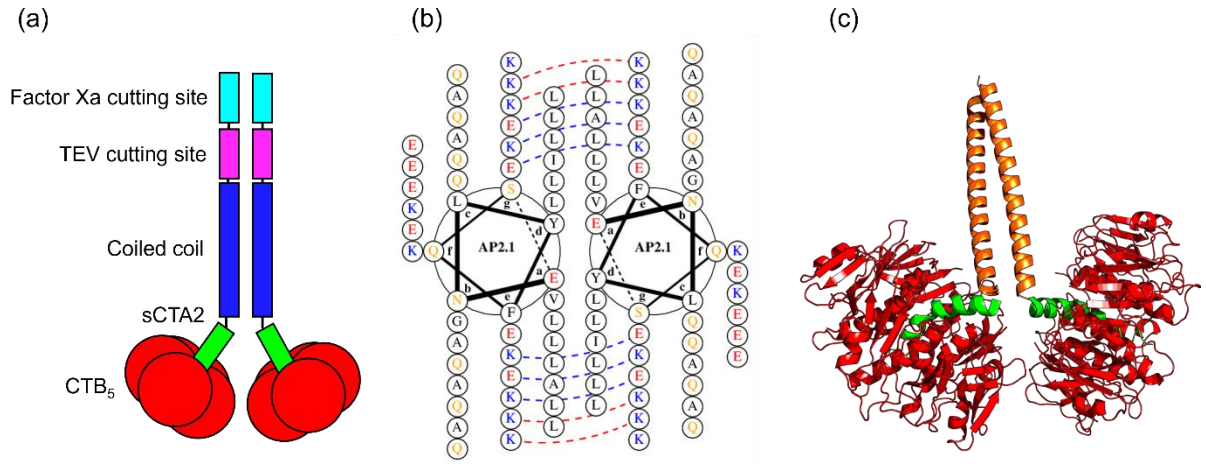


Figure III-22 Scenario one: when AP2.1 is formed as a parallel homodimer. (a) Scheme representing the structure of AP2.1 when it adopts a parallel orientation. (b) Helix wheel of the coiled coil region including the TEV cutting site. Dashed blue lines indicate favourable interactions and red dashed lines indicate unfavourable interactions. (c) A representative image (made in Pymol) mimicking the possible structure.

If AP2.1 is antiparallel, the spatial obstruction created by having two CTB_5 on the same end of the coiled coil is eliminated. However, the helix wheel and BUDE energy score of 31.5 indicate that this coiled coil is not energetically stable if the coiled coil region is a perfectly aligned (Figure III-23).

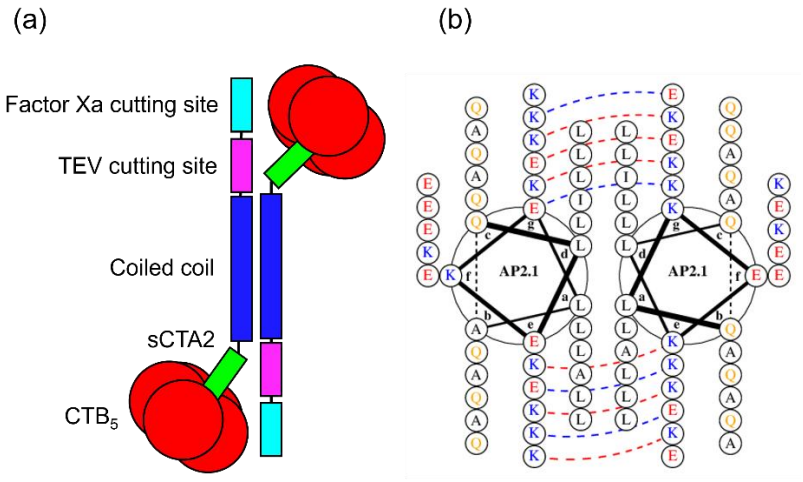


Figure III-23 Scenario two: (a). when AP2.1 adopts an antiparallel orientation while the procoil sequences (blue) are perfectly aligned with each other to form a coiled coil. (b) Helix wheel of the coiled coil region. Dashed blue lines indicate favourable interactions and red dashed lines indicate unfavourable interactions.

So what if the TEV cutting site is involved in the formation of an antiparallel coiled coil? Although the helix wheel indicates that the salt bridge is no longer attracted overall in this instance, the BUDE energy determined by CCBuilder is -190.2 indicating a favourable interaction when the TEV cutting site is included. As a result, the knob-into-hole structure in this section of the coiled coil region is more stable than the two scenarios discussed above. From a structural standpoint, this conformation is more likely to actually happen given that the spatial resistance of having two CTB_5 on the same end of the coiled coil is eliminated.

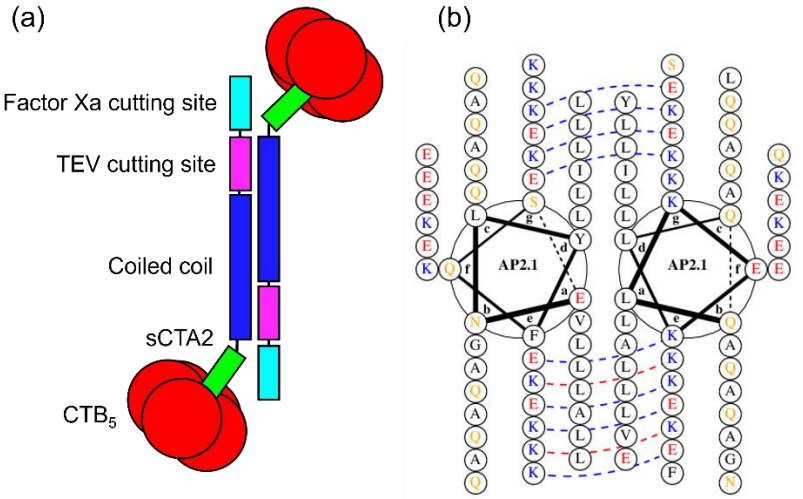


Figure III-24 Scenario three: (a). when AP2.1 adopts an antiparallel orientation while the procoil sequences (blue) are perfectly aligned with each other to form a coiled coil that includes the TEV cutting site. (b) Helix wheel of the coiled coil region. Dashed blue lines indicate favourable interactions and red dashed lines indicate unfavourable interactions.

III.7 A rationally designed homodimeric antiparallel CTB_5 complex

Realizing the need to create an antiparallel dimer with an accessible cleavage site and that is easy to understand, I decided to return to an antiparallel homodimer that was previously published, making any necessary adjustments to guarantee the protein expression yield. In my earlier attempts to express the AP1.x (AP1.1:AP1.2) group, I ran into problems with protein expression; specifically, the MBP-AP1.2, due to its low expression yield, was unable to be used to assemble with MBP-AP1.1. Furthermore, MBP-AP2.1 and MBP-AP2.2, even they were designed to interact with each other to form heterodimer, both formed homodimers by itself. In this case, obtaining a trustworthy antiparallel homodimer had to be the new objective of the work instead of an antiparallel heterodimer. Homodimer could increase the likelihood of successful expression and eventual acquisition of a protein complex in a short amount of time, whereas heterodimers require more time to characterize the oligomerization of the individual proteins and the 1:1 mixture of the two partners. Furthermore, expression issues with one partner protein can undermine use of that heterodimer.

III.7.1 Design strategy

The choice of the coiled coil must adhere to the following standards:

a) a rationally designed coiled coil, rather than natural coiled coil, is preferred to enable optimization of protein yield since coiled coils created by rational design are usually arranged in a predictable way, which simplifies understanding of the optimisation process. Additionally, the rationally created coiled coil will have less overlap with natural biological sequences, minimizing any possible unwanted interactions when performing future biological experiments;

b) the chosen coiled coil should be able to be expressed in prokaryotic cells, which has proven to be problematic for some MBP-*procoil*-$CTA2B_5$, particularly when the coiled coil region is lysine-rich;

c) The homodimer of the coiled coil should have a high binding affinity so that the dimer is intact in the nanomolar concentration range used in membrane fusion experiments. In light of these requirements, I ultimately decided to use AP-homo1[126] as a starting point for additional optimization (Figure III-25).

Only two differences exist between AP-homo1 and AP-homo2: the lysine residue at position *2b* (heptad 2, position *b*) has been changed from lysine to alanine because the original *2b*

lysine does not appear to have a purpose and the presence of an extra lysine could affect how the protein is expressed. Additionally, the tryptophan at position *4f*, which was originally intended to help detect peptide concentrations from their absorbance at 280 nm, was not required because the coiled coil would be fused directly to $CTA2B_5$ for protein co-expression. AP-homo3 has more residue modifications than AP-homo1 particularly regarding the lysine residue at position *f* (Figure III-25-b). This was done to prevent protein expression issues arising from a high overall positive charge in the *procoil* region.

(a)

Refenrence	Pair	Protein sequence *abcdefg abcdefg abcdefg abcdefg abcdefg abcdefg abc* *dcbagfe dcbagfe dcbagfe dcbagfe dcbagfe dcbagfe dcba*	Note
[126]	AP-homo1	QLEKE LKQLEKE LQAIEKQ LAQLQWK AQARKKK LAQLKKK LQA QLEKE LKQLEKE LQAIEKQ LAQLQWK AQARKKK LAQLKKK LQA	Original publication
[126]	AP-homo2	QLEKE LAQLEKE LQAIEKQ LAQLQKK AQARKKK LAQLKKK LQA QLEKE LAQLEKE LQAIEKQ LAQLQKK AQARKKK LAQLKKK LQA	Minor change
[126]	AP-homo3	QLEQE LAQLEAE LQAIEQQ LAQLQAK AQARKQK LAQLKAK LQA QLEQE LAQLEAE LQAIEQQ LAQLQAK AQARKQK LAQLKAK LQA	More changes mainly in f position

(b)

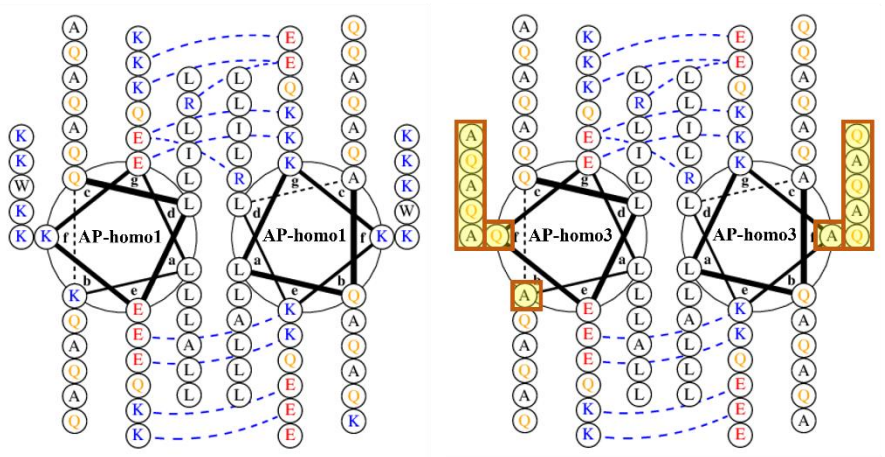


Figure III-25 (a) Sequences for the selected set of antiparallel coiled coil homodimer. (b) Helix wheel comparing AP-homo1 (original coiled coil from publication[126]) and AP-homo3. The modified residues in AP-homo3 were highlighted in yellow.

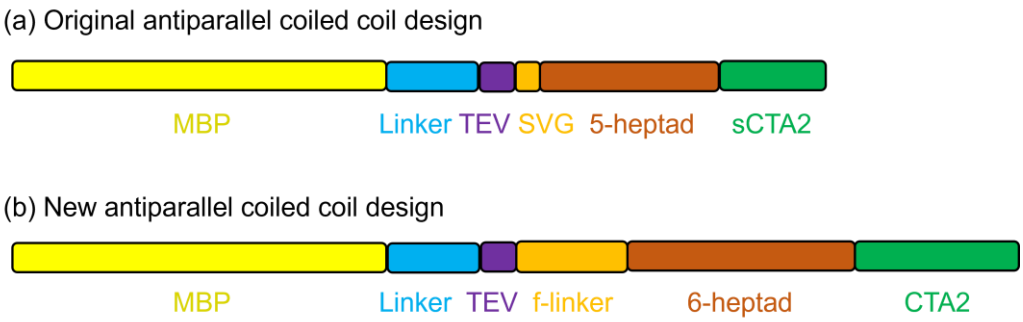


Figure III-26 Scheme representing the new design of the *procoil* region, comparing to the original version. The major changes involve replacing the SVG linker before *procoil* into a more flexible linker (f-linker) composed of GVG-$(GS)_4$-G linker. And an extension of the original *procoil* from 5 heptad to 6 heptad. Furthermore, a longer CTA2 peptide is used at the C terminus rather than the previously used short CTA2 peptide (sCTA2).

Once the protein sequence for the coiled coil region has been established, the next step is to determine whether the structure of the entire protein complex needs to be changed (Figure III-26). First, full-length CTA2 rather than short CTA2 (sCTA2) should be used in the newly designed complex to allow fair comparison with the parallel coiled coil-$CTA2B_5$ assembly. Second, a 12-AA linker, primarily made of flexible glycine and serine, was chosen to be added between the *procoil* and TEV cleavage sites. This was because the original SVG and SVGGGGGG sequences did not provide enough space for the MBP-TEV protease to attack the TEV cutting site. Lastly, the 6-heptad coiled coil was chosen instead of 5-heptad length because it not only offers a higher binding affinity but also makes it possible to directly compare it to the 6-heptad coiled coil in the parallel oriented $CTA2B_5$ dimer, *i.e.* PP2.x(PP2.1:PP2.2).

III.7.2 Protein expression and characterization by SEC

The protein expression and purification process for the newly designed homodimeric antiparallel CTB_5 dimers is described in section VII.4.3. The final yield of purified protein for MBP-AP-homo1 was 0.98 mg/L, while MBP-AP-homo2 and MBP-AP-homo3 were obtained in 0.45 mg/L and 2.6 mg/L, respectively.

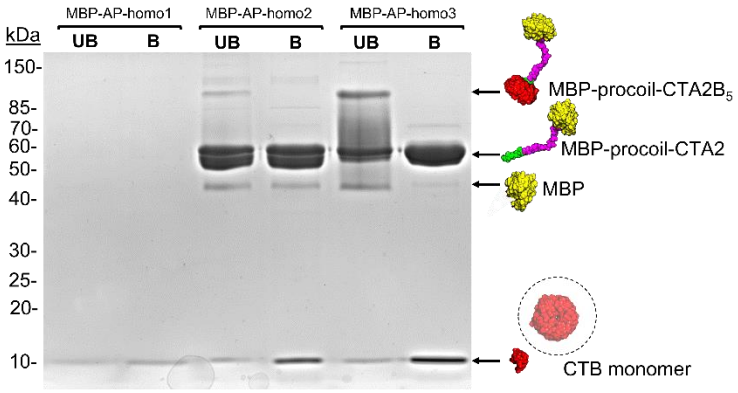


Figure III-27 A 12% SDS-PAGE was conducted with Coomassie Blue staining to analyze the final purified product of MBP-AP-homo1, MBP-AP-homo2, and MBP-AP-homo3. Boiled (B) and unboiled (UB) samples were loaded in parallel to evaluate potential alterations in protein conformation resulting from boiling.

Strangely, no sign of the MBP-AP-homo1 protein band could be seen on the SDS-PAGE gel (Figure III-27) even though UV absorption measurements with the Nanodrop showed that there was protein in solution – a possible explanation is that MBP-AP-homo1 was aggregated before being loaded into the gel. In contrast, the MBP-*procoil*-CTA2 from both MBP-AP-homo2 and MBP-AP-homo3 were found to assemble with CTB_5 as indicated by the presence of a higher mass band on SDS-PAGE (Figure III-27) for unboiled (UB) samples which disassembled into its constituent parts upon boiling the samples (B). However, MBP-AP-homo3 was the better candidate for further characterization because it suffered from less degradation compared to MBP-AP-homo2 based on the number of bands at around 50–60 kDa.

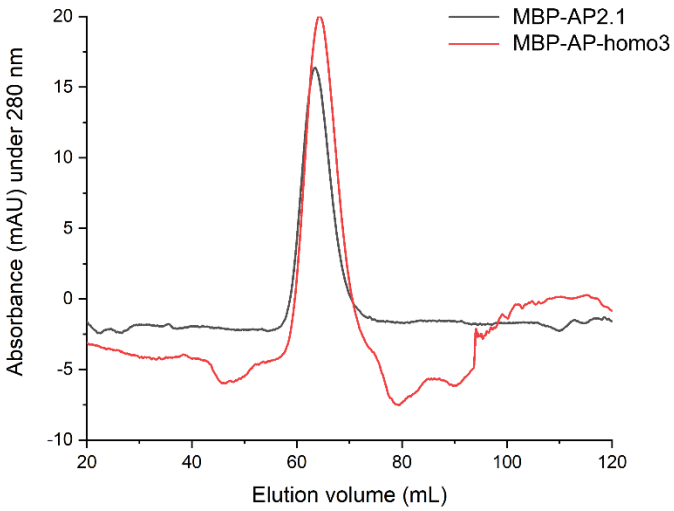


Figure III-28 Size exclusion chromatography (SD200, 16/60) with MBP-AP2.1 and MBP-AP-homo3 elution profiles overlaid.

MBP-AP-homo3 had a similar SEC elution volume to that of MBP-AP2.1(Figure III-28), which provided confidence of it being a homodimer. Even though MBP-AP-homo3 is longer than MBP-AP2.1, MBP-AP-homo3 is overall more flexible due to the glycine-rich linker before coiled coil region and an extended CTA2 peptide, hence the slight difference in retention time cannot directly tell if the MBP-AP-homo3 has dissociated. MBP-TEV protease also efficiently removed the MBP from MBP-AP-homo3 in 2h incubation time at 37 °C (Figure III-29).

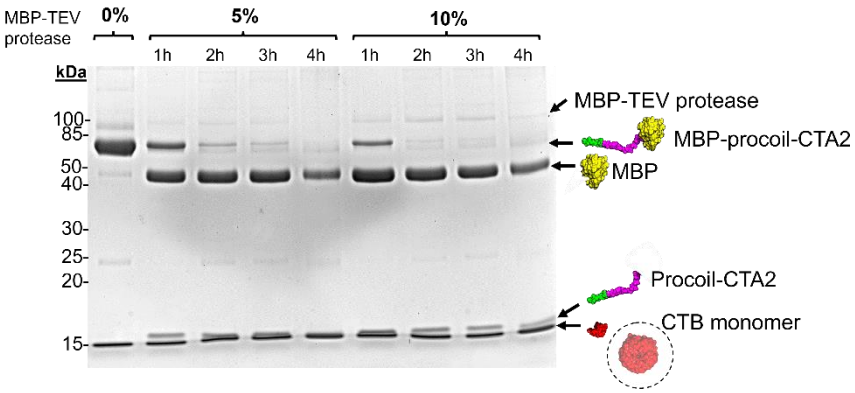


Figure III-29 The time course of MBP-AP-homo3 digestion by TEV protease was evaluated by incubating the protein with either 5 mol% or 10 mol% of MBP-TEV protease at 37°C for 1, 2, 3, and 4 hours. The final products were subjected to boiling and loaded onto a 12% SDS-PAGE gel stained with Coomassie Blue.

All available experimental data point to the ability of MBP-AP-homo3 to form a stable CTB_5 dimer, and this protein will be used in the further characterization with membrane. The following sections will simply refer to it as MBP-AP3.

III.8 Interaction between designed partners in (MBP)-PPx set

In contrast to the challenges encountered in producing the antiparallel coiled coil-$CTA2B_5$ complex, the parallel coiled coil pairs were produced fairly easily. The expression and purification of two sets of parallel coiled coil pairs from the literature[115, 120] were described in section III.2 . Analysis of the individual proteins by SEC and SEC-MALS (section III.4.1 and III.4.2) indicated that they were either monomeric or partially dimerised. In the following chapter, native PAGE and SEC will be used to further characterize the successful formation of parallel coiled coil pairs.

III.8.1 Native PAGE

Native PAGE used to provide an initial judgement for whether the designed partners in each set would interact with each other to form a dimer. If the isoelectric point (PI) value of a protein is smaller than the pH of the solution, the protein should be able to migrate towards the negative electrode. PI and net charge of protein at pH 7.4 for each protein were calculated using the protpi website[5] (Tabel III-1). Considering the pH of the running buffer is around 8.3, all the interested proteins, even those with a very small positive net charge in neutral solution, should be able to migrate towards the negative electrode.

Table III-1. The molecular weight (MW), isoelectric point (PI), and net charge of MBP-PPx protein.

Name	Interested protein	Theoretical MW (Da)	PI	Net charge at pH 7.4
MBP-PP1.1	MBP-*procoil*-CTA2	53514	5.007	-18.757
MBP-PP1.2		53511	5.284	-12.77
MBP-PP2.1		54383	5.283	-13.725
MBP-PP2.2		54835	5.409	-11.714
MBP-PP1.1	MBP-*procoil*-$CTA2B_5$	111499	6.051	-14.439
MBP-PP1.2		111496	6.424	-8.454
MBP-PP2.1		112368	6.377	-9.407
MBP-PP2.2		112821	6.521	-7.396

[5] https://www.protpi.ch/Calculator/ProteinTool/

From the Native PAGE results (Figure III-30), it is difficult to distinguish the identity of each band. Hence the aim was not to determine the oligomerization state of each protein, but to observe if there was a higher oligomer state formed when designed partners (such as MBP-PP1.1 and MBP-PP1.2) were mixed in a 1:1 ratio. For MBP-PP1.1 and MBP-PP1.2, it seems that the 'Mix' lane was more like an overlay of the components from MBP-PP1.1 and MBP-PP1.2 lanes. However, for MBP-PP2.x, a smear formed on the gel when the MBP-PP2.1 and MBP-PP2.2 were mixed as in the 'Mix' lane, which might indicate the formation of higher oligomerization state comparing to MBP-PP2.1 and MBP-PP2.2 individuals.

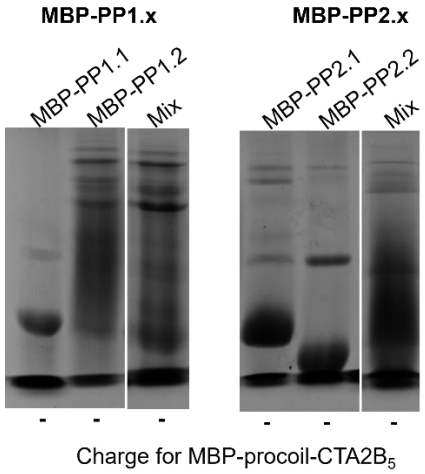


Figure III-30 The Native PAGE of the partners of MBP-PP1.x and MBP-PP2.x. Individual proteins as well as the 1:1 ratio mix of the partners in the same group were loaded into the gel.

The native PAGE indicated that the MBP-PP2.1 and MBP-PP2.2 might be more promising pair for constructing the heterodimeric coiled coil construct and so no further experiments were conducted with MBP-PP1.1 and MBP-PP1.2. However, even though MBP-PP2.1 and MBP-PP2.2 seemed to assemble with each other, there remained a suspicion that the CTB_5 building block might exert steric hindrance on protein complexes with a parallel orientation if the CTA2 and procoil regions formed a continuous helix. Therefore, a proline was added between the *procoil* and CTA2 peptide to generate a twist for both MBP-PP2.1 and MBP-PP2.2. The new set of protein complexes is named MBP-P-PP2.x. Further studies of MBP-PP2.x and MBP-P-PP2.x mixed in a 1:1 ratio were conducted using size exclusion chromatography.

III.8.2 Size exclusion chromatography

An optimized sample processing procedure was developed to unambiguously observe the formation of dimers between mixtures of individual proteins:

a) Individual proteins or protein mixtures were incubated overnight at 37 °C before being injected into size exclusion columns. This allowed any homodimer with weak interactions to dissociate and interact with its partner to create a stable coiled coil assembly;

b) Instead of using a SD200 10/200 size-exclusion chromatography column, a SD200 16/600 was used. This longer column offered better resolution to distinguish between the expected 220 kDa and 110 kDa protein weights even though the larger cross-section requires more protein sample needs to be injected.

The results of assembly experiments using MBP-PP2.1/2.2 and MBP-P-PP2.1/2.2 are shown in Figure III-31.

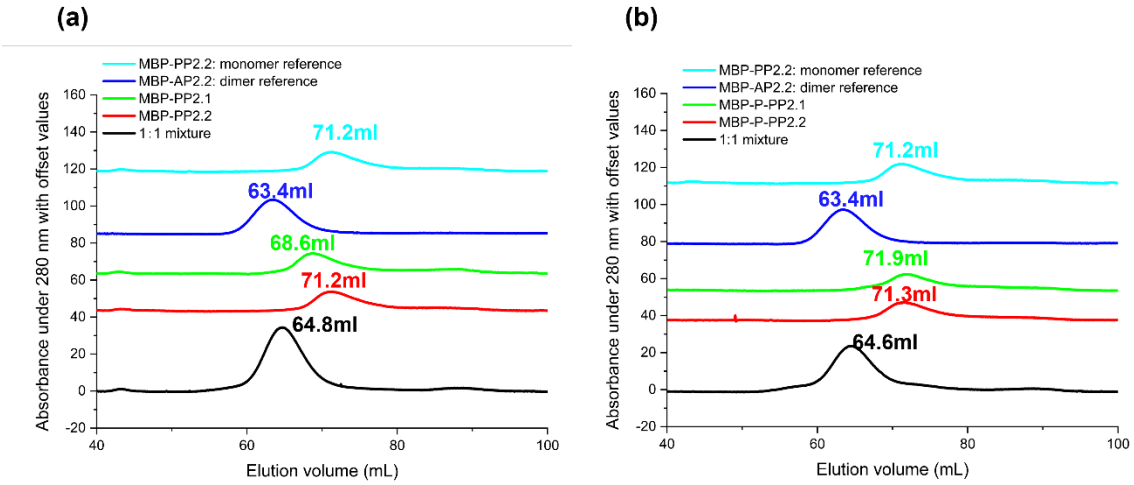


Figure III-31 (a) SEC curve of MBP-PP2.1 and MBP-PP2,2 as individual proteins and the assembled complex. (b) SEC curve of MBP-P-PP2.1 and MBP-P-PP2.2 as individual proteins and the assembled complex. MBP-PP2.2 and MBP-AP2.2 are shown as a reference for where monomers and dimers are expected to elute. For each protein tested in SEC, the concentration of individual was diluted to 10 µM before being incubated overnight at 37 °C and then at 25 °C for at least 2h before injection. For the protein mixtures, the concentration of each component was 10 µM.

The SEC-MALS analysis of MBP-PP2.2 (Figure III-12) had shown that it was a monomeric species in solution. As both MBP-P-PP2.1 and MBP-P-PP2.2 proteins had almost the same elution volumes as MBP-PP2.2, they too must be monomeric. In contrast, MBP-PP2.1 eluted approximately halfway between the values expected for a monomer and dimer indicating that it could be partially dimerized in solution. Another finding is that the 1:1 mixture

consistently eluted more quickly than the protein monomers. Individual proteins from MBP-PP2.x elute between 68.6 and 71.2 mL, whereas the mixture elutes as early as 64.8 mL. MBP-P-PP2.x also showed an earlier elution for the mixture of about 5 mL compared to the individual proteins. Although it is unclear whether the preheating step is required for the formation of heterodimeric protein complexes — weakly formed homodimers may still be driven away by stronger binding partners even at room temperature — this process allows for a more accurate comparison of the various oligomeric states and the 1 to 1 mixture. In terms of the overall performance of individual proteins as well as protein mixtures, the MBP-P-PP2.X pair, rather than MBP-PP2.X, was selected for further study as an example of a parallel coiled coil as there was a clearer shift in elution time from monomers to dimers upon mixing the two proteins.

In conclusion, MBP-P-PP2.x (MBP-P-PP2.1:MBP-P-PP2.2) was therefore selected to join MBP-AP2.1 and MBP-AP3 for further study as described in Chapter IV .

Chapter IV Characterization of coiled coil-mediated CTB_5 assembly

IV.1 Introduction

After several rounds of selection and redesign, the final candidates for future membrane-related experiments were selected. MBP-AP2.1 (Figure IV-1-a) is the accidently formed antiparallel coiled coil. It has a short version of the CTA2 peptide (sCTA2) that only keeps the necessary residues on CTA2 peptide that bind to the CTB_5. The coiled coil region was originally my rational design to be a heterodimer, however it turned out to exhibit stable homodimer formation, which prevented easy analysis of whether it could also form heterodimers due to the same molecular weight of its binding partner. The deduction of its coiled coil orientation has been demonstrated in section III.6.2, and further computational studies also provide more evidence as will be described in section IV.4.

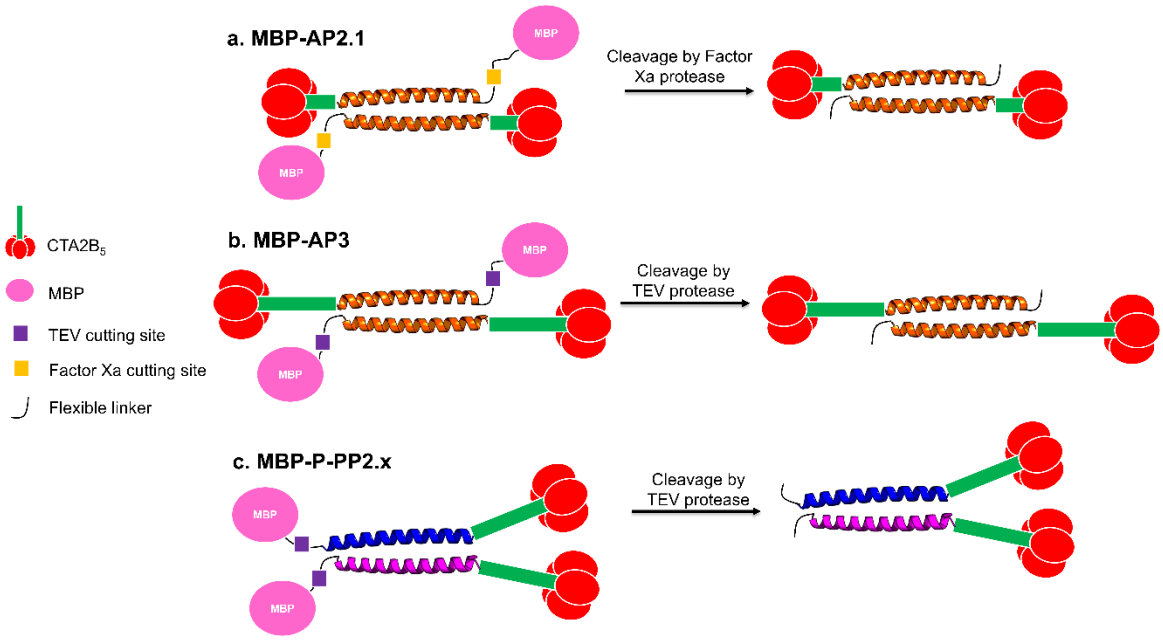


Figure IV-1 Scheme comparing three CTB_5 dimer complexes assembled using different coiled coil linkers.

The second protein picked was MBP-AP3 (Figure IV-1-b), which was rationally redesigned to increase the ease of protein expression primarily in non-interaction surfaces like the *f*-position while maintaining the core interaction residues of the original helical wheel from the published article[126]. For a more accurate comparison with the parallel coiled coil construct, the MBP-AP3 construct uses the entire CTA2 peptide and a 6-heptad coiled coil region. The third protein is formed from the MBP-P-PP2.x pair (Figure IV-1-c), which uses a heterodimeric parallel coiled coil originally selected from SYNZIP[120]. Full-length CTA2 and 6-heptad procoils were inserted to enable a degree of flexibility between the dual CTB_5 on the same end of the coiled coil, as well as keeping a strong binding interaction between partners. In the following sections, parallel experiments were performed to compare any biophysical and membrane-related behaviors induced by the three protein complexes.

IV.2 SDS-PAGE and mass spectrometry

Before introducing the assembled CTB_5 constructs in any membrane-based experiment, the fused MBP was removed to avoid this no longer necessary protein portion from interfering with the interpretation of the results. Different protein cleavage proteases were used according to the cutting efficiency of the target protein, and an amylose column was applied to remove the MBP domain. The final purified protein complexes were finally examined using SDS-PAGE and ESMS to ensure the purity of the sample and the desired clean cutting from the protease (Figure IV-2).

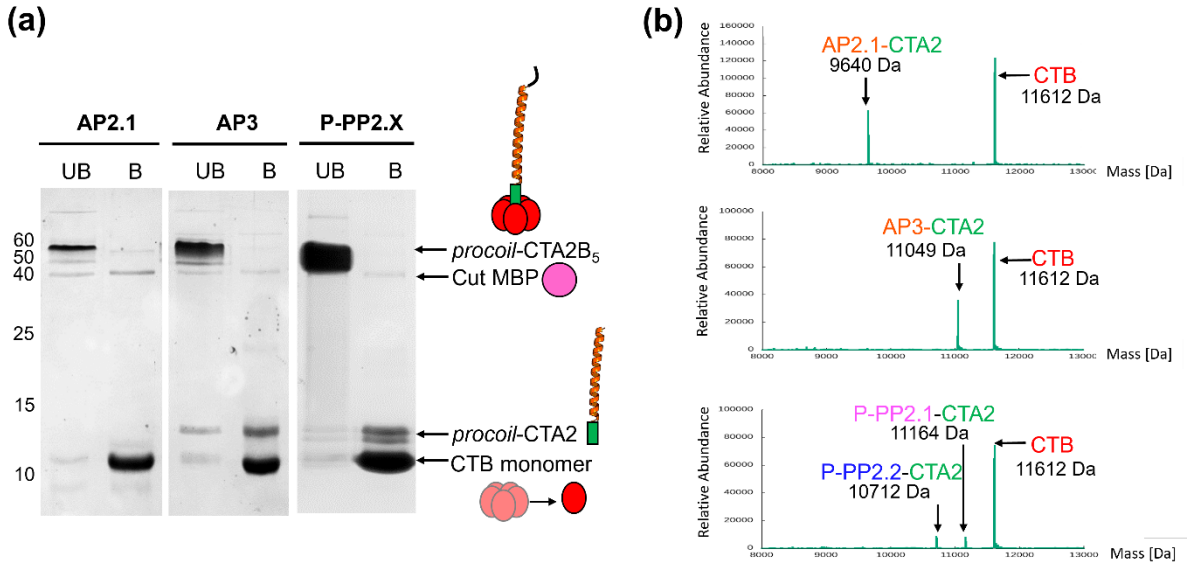


Figure IV-2 After removing the MBP using either MBP-TEV or Factor Xa protease, the cut CTB_5 dimer was purified through an amylose column by collecting the flow-through/wash fraction. (a) Boiled (B) and unboiled (UB) samples from AP2.1, AP3 or P-PP2.x were loaded on a 16% Coomassie stained SDS-PAGE gel. Even though there was a small amount of contamination of the cut MBP, the majority of the protein component is the *procoil*-CTA2 and CTB_5 complex, and the correct cutting site was further verified through (b) mass spectrometry.

After successful proteolysis, the protein mixture containing *procoil*-$CTA2B_5$, cleaved MBP, and protease was purified using amylose resin. For AP3 and P-PP2.x, the *procoil*-$CTA2B_5$ was collected in the flow-through and wash steps because the cleaved MBP and the MBP-TEV protease were retained on the amylose resin. For AP2.1, PMSF (final concentration of 1 mM) was added to the solution following amylose chromatography to stop additional non-specific cleavage because a small amount of Factor Xa protease will co-elute with AP2.1. Depending on the purity required for the experiment, the AP2.1 was also passed through a nickel column for a second purification step. The purified proteins were characterized by ESMS to confirm the molecular mass and verify that the cleavage and purification processes went as planned (Figure IV-2-b).

IV.3 Oligomerization state at nanomolar concentrations

Detecting the oligomerization state of sample proteins at nanomolar concentrations is important for this project. Subsequent experiments involving protein-membrane interactions, including QCM-D and GUV experiments all require protein less than 1 µM – it is therefore necessary to ensure that the test proteins remain stable after MBP is removed and that dimerization occurs even at low protein concentrations.

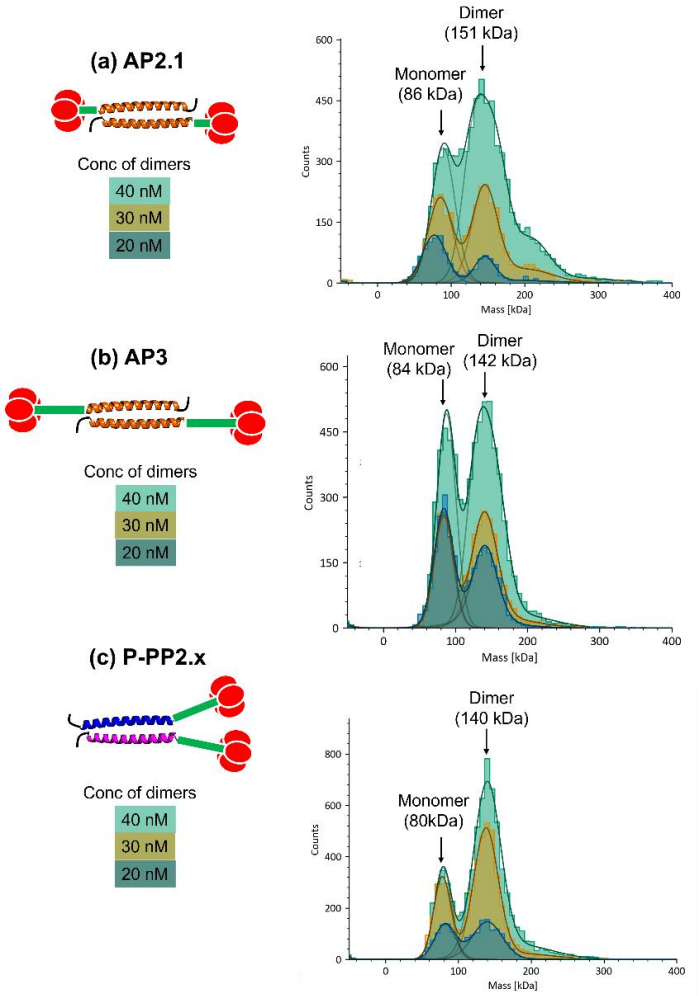


Figure IV-3 Mass photometry of AP2.1, AP3 and P-PP2.x, with final dimer concentration ranging from 20 nM to 40 nM loaded on glass coverslips. Each colored curve, derived from distinct samples, corresponds to a singular experimental trial.

The protein samples were diluted to 200 nM dimer concentration with HEPES buffer and incubated at *r.t.* for at least 1h before executing the mass photometry detection. One way to achieve different final concentrations is to prepare the samples at the target concentrations using LoBind Eppendorf tube, while the mass photometer is blanked without any addition of buffer. However, this experimental method resulted in high background noise which led to unreliable outcomes. The second method is to blank the machine with filtered buffer, while the desired protein is stored at a higher concentration. During detection, the initial droplet used to equilibrate the detector was not removed and the added protein will be quickly diluted to a lower concentration. This is how the experiments were performed here. Since the dissociation rate of different coiled coils is unknown, this operation might contribute to different degree of inaccuracy since the monitoring time of 1 min is probably not enough for

the solution to reach equilibrium and hence may overestimate the oligomerization state of the protein at the given concentration.

At the concentrations studied, all proteins were a mixture of monomers and dimers, with the number of dimers increasing with increasing protein concentration (Figure IV-3). P-PP2.x had the highest level of dimerisation, while AP3 had the lowest. Nevertheless, it is important to note that the lowest dimer concentration used in subsequent studies was 200 nM, indicating that all three proteins are confidently mostly dimerized at this concentration.

IV.4 Prediction of the coiled coil interactions for AP2.1 and AP3 using ISAMBARD

ISAMBARD is a package for parametric coiled coil backbone design with built in optimization strategies[127, 128] (Figure IV-4). A population of parameter sets, such as the radius, pitch (which determines the degree of coiling of the monomeric helix), PhiCA (also known as Crick angle n, the angle corresponding to the registration number of the first residue in the helix), and zshift (relative movement parallel to the second helix), were examined in order to understand the topology of each designed protein complex and the rigidity of the coiled coil linker when it was fused to the native CTA2B$_5$ subunit.

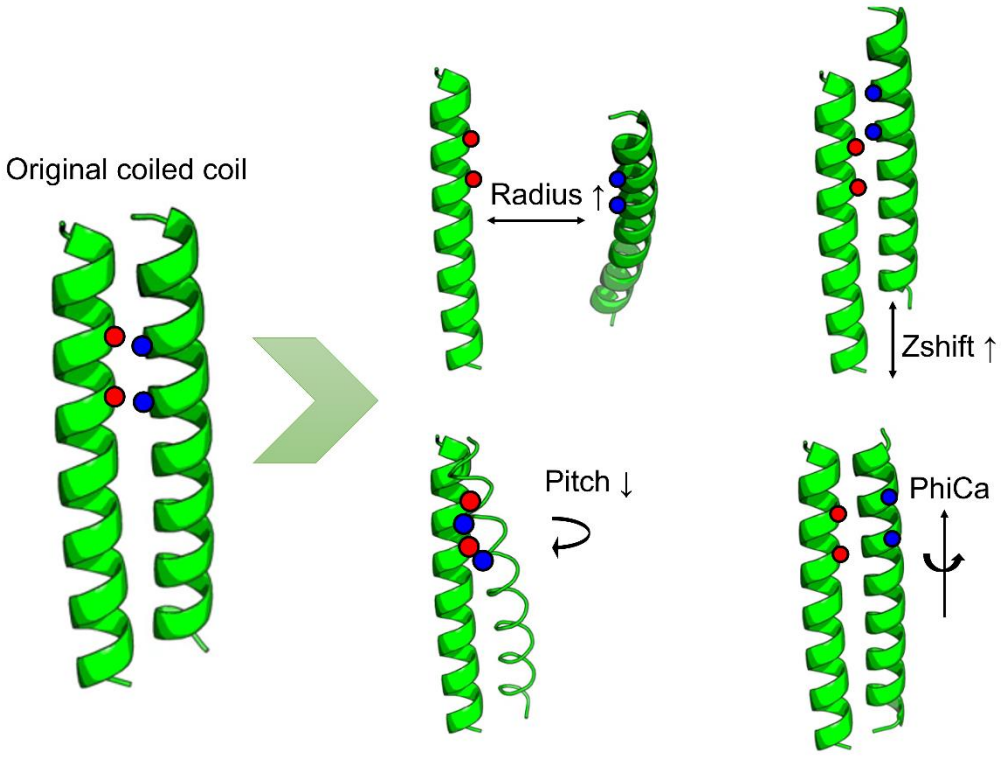


Figure IV-4 Examples of how different parameters influence the coiled coil topology.

Table IV-1 The input coiled coil sequence from each protein complex for ISAMBARD calculation.

Name	Input sequence for ISAMBARD screening and optmization
AP2.1	SENLYFQSVGQLEKELAQLKEKLQAIEKELAQLKEKAQALKEKLAQLKEKLQ
	SENLYFQSVGQLEKELAQLKEKLQAIEKELAQLKEKAQALKEKLAQLKEKLQ
AP3	PAEQLEQELAQLEAELQAIEQQLAQLQAKAQARKQKLAQLKAKLQA
	PAEQLEQELAQLEAELQAIEQQLAQLQAKAQARKQKLAQLKAKLQA
P-PP2.x	LVAQLENEVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEE
	RNAYLRKKIARLKKDNLQLERDEQNLEKIIANLRDEIARLENEVAS

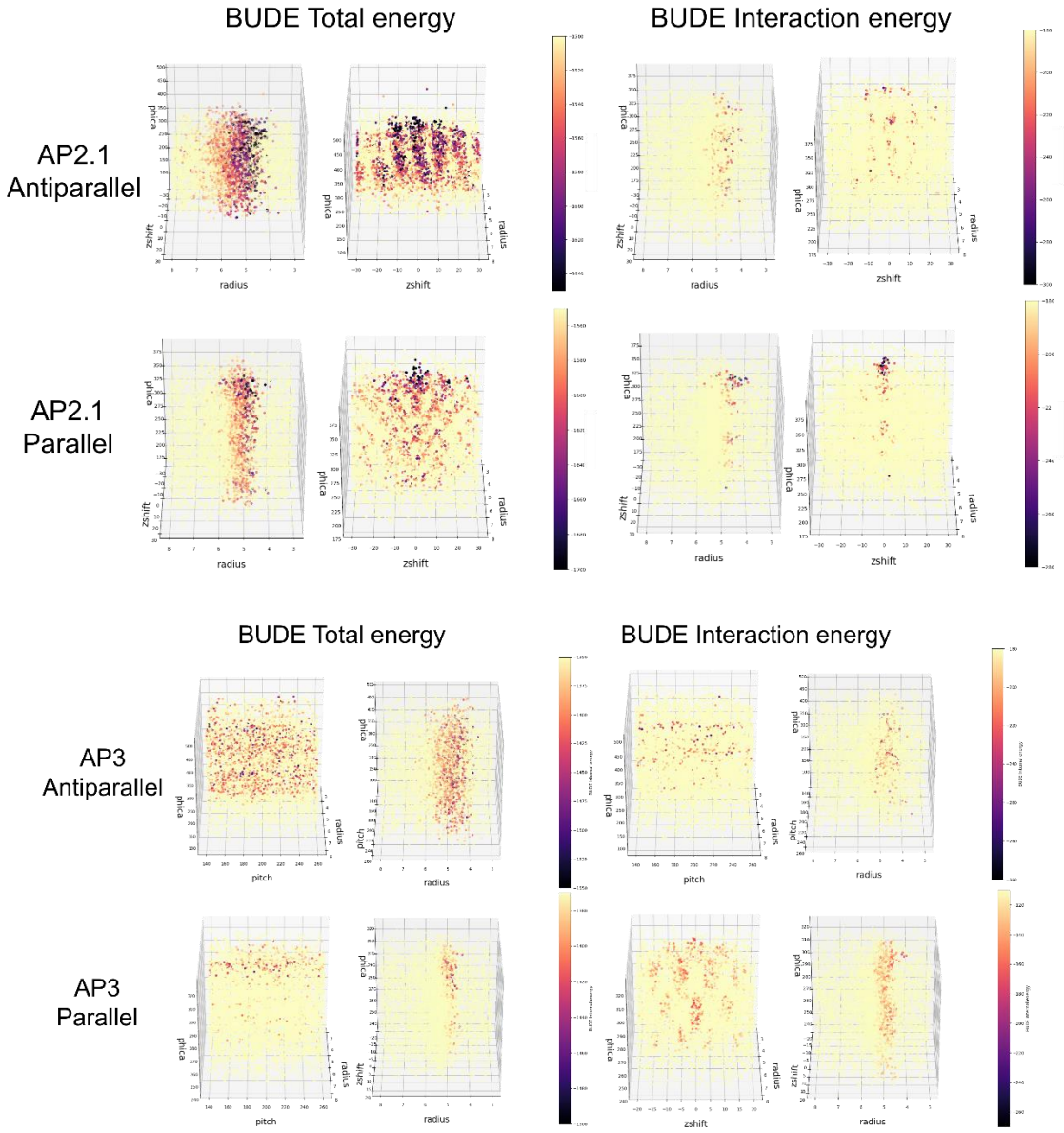
Table IV-1 contains the protein sequences that were used to build the AP2.1, AP3, and P-PP2.X coiled coil models. The models are built in two stages: the first stage is coarse selection, which establishes the general range of parameters that comprise the ideal model. A genetic algorithm is used in the second round to optimize the model for the chosen region. The next section provides a summary and analysis of the outcomes of these two screens.

In the initial round, 1000 values were generated within the given parameter range (Tabel IV-2) with either BUDE internal energy or BUDE interaction energy to be calculated. The PhiCA indicates the registration number of the coiled coil, for example, "a," "b," "c," "d," "e," "f," and "g" are each numbered 0, 102.8, 205.6, 308.4, 51.4, 154.2, and 257, respectively. In the recommanded ISAMBARD protocol I followed[6], the PhiCA number was initially set as 283 as start, which roughly corresponds to position g. A higher pitch, on the other hand, denotes a more "straight" helix formation because it indicates the extent at which one helix is wound around the other. Z-shift refers to the movement of one helix with respect to the second helix in the z-axis and is particularly helpful if we want to determine whether the coiled coil exhibits a mismatched interaction. Then, a genetic algorithm (GA) is performed to highlight the region with the best score using the dataset that contains the top 20% of the best scores - this procedure continues to be included in the initial coarse screen step. The first screen was repeated three times simultaneously, and the output datasets were combined into two 3D scatter plots with the Radius, Pitch, and PhiCA acting as the x, y, and z axes (Figure IV-5). The dots were coloured in accordance with the calculated BUDE internal energy or BUDE interaction energy with different levels shown in colour maps. The region where internal energy and interaction energy overlap is taken into account as the narrated parameter range for the following stage of optimization. If the best scored region is not included in the initial coarse-screened sets, the parameter range for genetic algorithm optimization can also be expanded.

[6] https://ISAMBARD-uob.github.io/ISAMBARD/tutorials.html

Table IV-2 The selection of parameters to undergo the coarse screen for building either parallel or antiparallel orientation for each given protein sequence.

		Radius	**Pitch**	**PhiCA**	**Zshift**
AP2.1	P	5.5 ± 2.5 Å	200 ± 60 Å	283 ± 100°	0 ± 30 Å
	AP	5.5 ± 2.5 Å	200 ± 60 Å	283 ± 100°	0 ± 30 Å
AP3	P	5.5 ± 2.5 Å	200 ± 60 Å	283 ± 40°	0 ± 20 Å
	AP	5.5 ± 2.5 Å	200 ± 60 Å	283 ± 200°	0 ± 10 Å
P-PP2.x	P	5.5 ± 2.5 Å	200 ± 60 Å	283 ± 40°	0 ± 20 Å
	AP	5.5 ± 2.5 Å	200 ± 60 Å	283 ± 40°	0 ± 20 Å



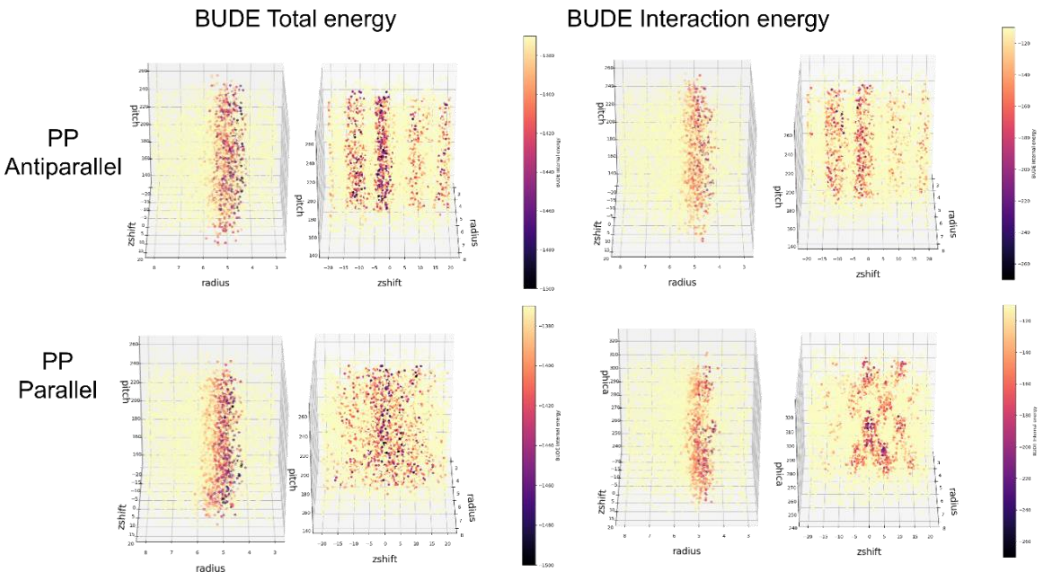


Figure IV-5 The coarse screen scores with designed sequence. PhiCA, pitch, and radius served as the x, y, and z axes, respectively, in a 3D scatter plot. The parameters in the cross-selection between the low-energy region in the two types of energy plots are used for further optimization after either the BUDE internal score or the BUDE interaction score has been calculated and reflected in the colormap.

Table IV-3 Following the coarse screen, the narrowed range for each parameter was used in the second round of optimization using the genetic algorithm.

		Radius	**Pitch**	**PhiCA**	**Zshift**
AP2.1	P	5 ± 0.5 Å	200 ± 60 Å	283 ± 40°	0 ± 10 Å
	AP	5 ± 0.5 Å	200 ± 60 Å	283 ± 40°	0 ± 15 Å
AP3	P	5 ± 1 Å	200 ± 60 Å	310 ± 15°	0 ± 10 Å
	AP	5 ± 0.5 Å	200 ± 60 Å	300 ± 100°	0 ± 10 Å
P-PP2.x	P	5 ± 1 Å	200 ± 60 Å	283 ± 30°	0 ± 10 Å
	AP	5 ± 1 Å	200 ± 60 Å	283 ± 30°	0 ± 10 Å

The population size drops to 400 during the second stage of optimization with narrowed parameter sets (Tabel IV-3), and the best candidate was created using a genetic algorithm and output as the starter for the subsequent round of 50 iterations (Figure IV-6). This candidate was chosen based on the RMSD calculation to the model with the lowest score. To ensure that the final selection was not biased, this process was repeated three times in parallel. For each round of optimization, the avarage value of BUDE interaction energy for each generation as well as the standard deviation was obtained to evaluate the quality of selection. Finally, for each optimization repeat, the best model with the lowest BUDE interaction energy was generated for rational selection and analysis.

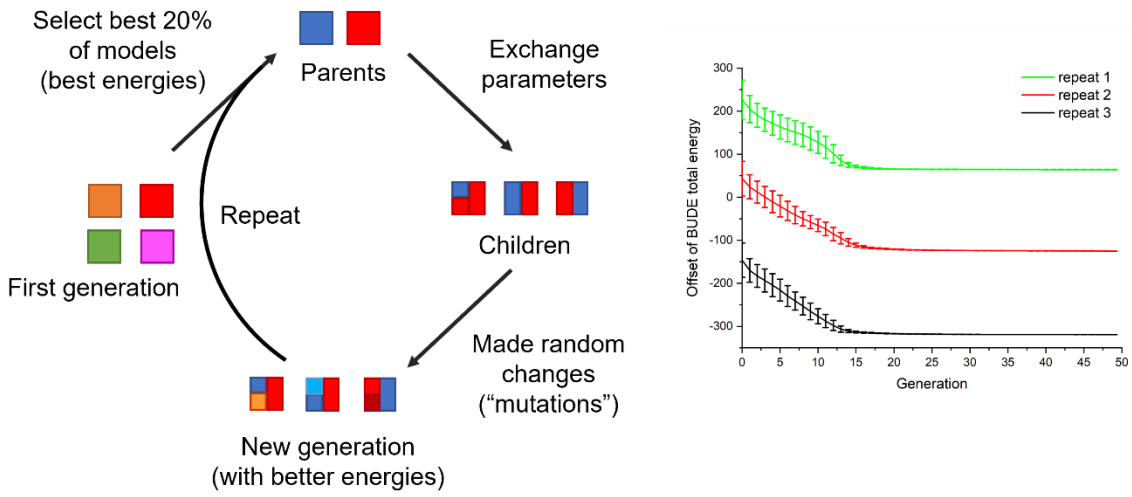


Figure IV-6 Diagram showing the application of a genetic algorithm (left). A line plot of the BUDE total energy of an AP2.1 coiled coil sequence optimized by a genetic algorithm serves as an example of the optimization process. After roughly 15 generations, the average BUDE interaction energy score dropped to its lowest range (Right).

In order to assess the optimization outcome achieved by using GA in ISAMBARD, energy funnel plots with RMSD (Å) to the best scoring model as the x-axis and BUDE internal energy as the y-axis were created for each repeat of optimization (Figure IV-7). Yellow represents the most recent generation while purple denotes the first generation. The majority of the resulting energy funnels clearly show that the later generations have higher scores, but some outliers, like the third repeat from AP2.1 in antiparallel orientation, can still be recognized. In this case, it would be safer to skip the third repeat when deciding on the structure to use for the models that will be built for the MD simulation in the following step.

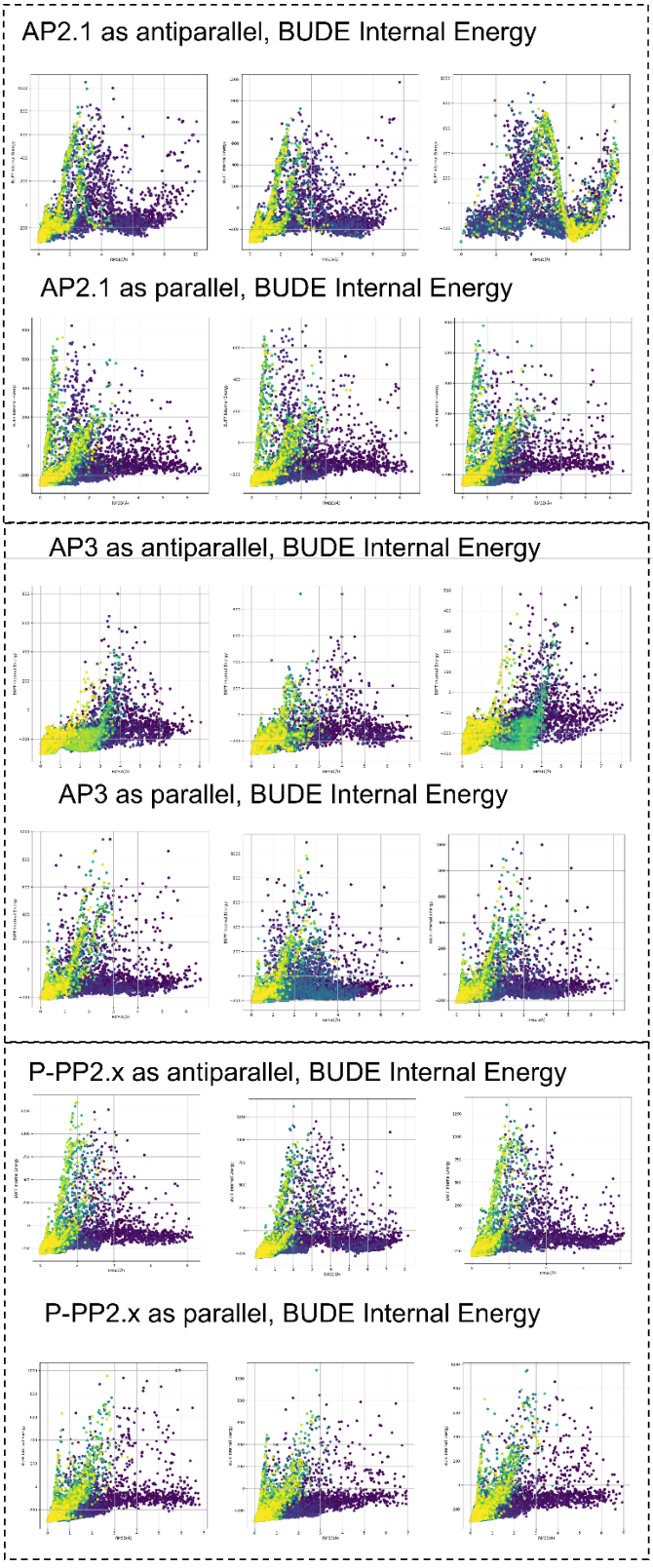


Figure IV-7 Energy funnel of the three-repeat genetic algorithm optimization. The targeted coiled coil sequences were examined in both parallel and antiprallel orientation, and corresponding models were optimized.

In addition to constructing an ideal structure from a given sequence, ISAMBARD can also use parametric packing to determine the direction of a coiled coil[129]. The lowest BUDE interaction energy is finally identified after numerous iterations of a genetic algorithm (Figure IV-8), but there is always a further comparison between parallel and antiparallel orientation through which one can obtain an even lower score. The prediction made by ISAMBARD for AP3 (Figure IV-8) suggests that the core-interaction leads to an antiparallel orientation (Black line) rather than the parallel orientation (Red line), when the core residues used in AP3 at the a-d-e-g position are maintained, which is in agreement with the experiments in the original paper[126].

P-PP2.X, which was shown by X-ray crystallography to favour a parallel orientation, is predicted by ISAMBARD to not have a clear preference for antiparallel or parallel orientation (Figure IV-8). Could P-PP2.x favor antiparallel instead? We hypothesized that the TEV cutting site is involved in the coiled coil of the antiparallel orientation of MBP-AP2.1 (section III.6.2), which would explain why, even though Factor Xa removes MBP, it is difficult to remove the extra N-terminal residues with TEV protease (Figure III-20). However, TEV protease also fails to completely digest MBP-5G-AP2.1 (Figure III-21), indicating that the steric hindrance of MBP attached to an antiparallel coiled coil also lessens the efficiency of TEV protease. because In contrast, MBP-P-PP2.x has a relatively high TEV protease digestion efficiency even though there is no extra flexible linker between the coiled coil and the TEV cutting site (Figure III-16), therefore, P-PP2.x ought to be in parallel conformation. This result illustrates that a full and solid deduction based solely on ISAMBARD calculations is not generally desirable. The BUDE forcefield is not optimal due to the solvation of the hydrophobic interface and its inability to predict the entropy change accurately[130]. One possible solution is to cross-examine the models with different forcefields or to draw conclusions based on experimental data. In the case of AP2.1, ISAMBARD predicts the coiled coil to prefer an antiparallel orientation, which in this case agrees with the TEV digestion experiment, offering yet more support for the deduction of AP2.1 having an antiparallel orientation.

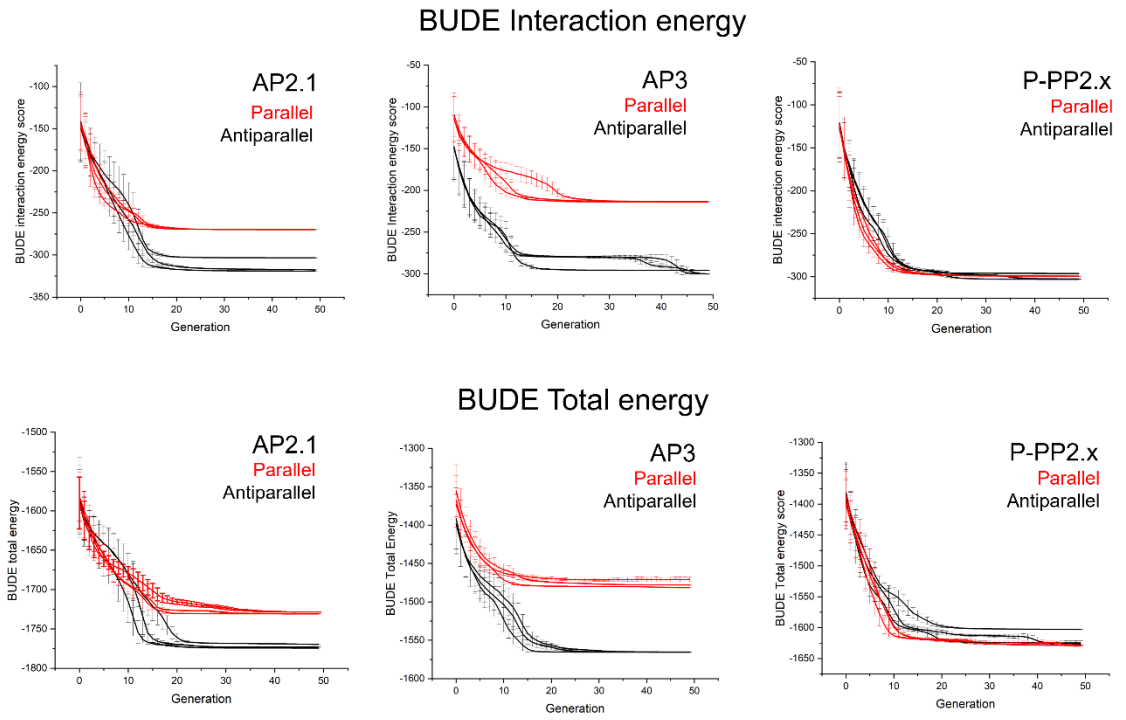


Figure IV-8 Utilizing BUDE Interaction energy or BUDE Total energy as a selection criterion, use genetic algorithms to optimize the parameter sets of the coiled coil in AP2.1, AP3, and P-PP2.x when it adapts either parallel orientation or antiparallel orientation.

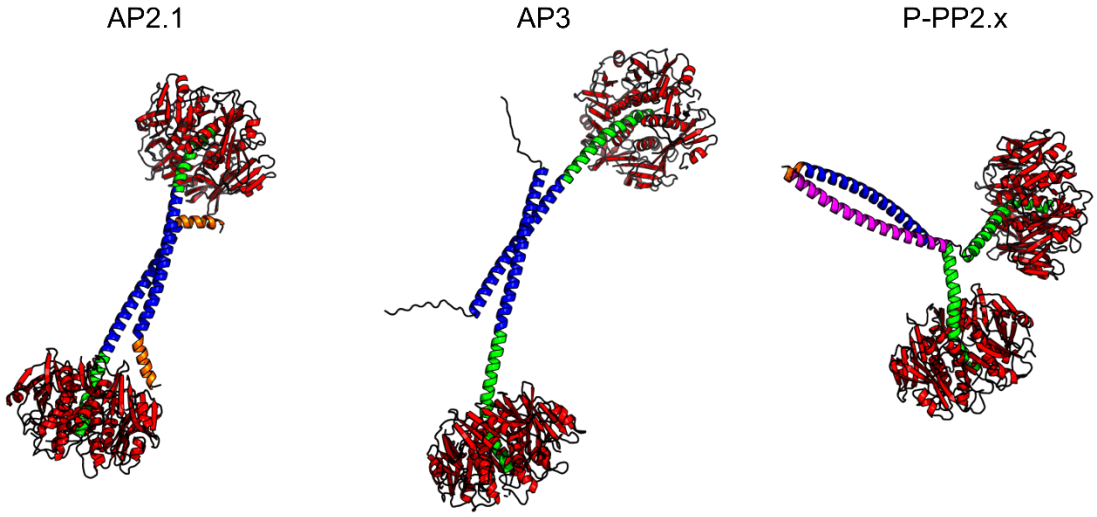


Figure IV-9 AP2.1, AP3, and P-PP2.x initial structures were created using Pymol and Alphafold.

After using ISAMBARD to build the coiled coil model and optimize the parameters, the coiled coil region was fused with the rest of the CTA2B$_5$ subunit using Pymol. The full linker region (*i.e.* the N-terminal flexible linker, *procoil*, and CTA2 region) was modelled with Alphafold, and structural alignment was used to determine the relative position of two full linkers in the

same CTB_5 dimer. In this manner, the complete dimer complexes for AP2.1, AP3, and P-PP2.x were constructed for additional MD simulation (Figure IV-9).

IV.5 MD analysis of the AP2.1, AP3 and P-PP2.x complexes

Simulations were performed with the AMBER14 package and AMBER-ff14SB force field[131]. All three structures — AP2.1, AP3, and P-PP2.x — were first dissolved in a cubic box of TIP3P water[132] that also contained 150 mM NaCl salt and neutralizing ions. The minimum separation between the protein and the water box's edge was set at 12 Å. The Particle-Mesh Ewald (PME)[133] method was used to calculate electrostatic interactions, with a grid spacing of 0.16 nm and a cutoff of 1.0 nm. The simulation system was set up using the AMBER tleap and parm modules. In the presence of 2.0 kcal/mol/Ang^2 restraints on protein backbone atoms, solvent atoms were minimized for 2500 steps of steepest descent minimization, followed by 2500 steps of conjugate gradient minimization. Following minimization, the system was heated in 25 ps using 1 femtosecond steps from 10.0 K to 303.15 K. The constraints were reduced to 0.1 kcal/mol/Ang^2 for heating and equilibration. The production simulations were run with 4 femtosecond steps, no restraints, and a constant temperature of 303.15 K and pressure of 1 bar. The production run for all the systems was carried out for more than 400 ns till it reached a stable RMSD. All hydrogen atom bonds were constrained by SHAKE[134]. The AMBER ptraj module was used to analyze the MD simulation trajectory.

Due to time constraints, the AP2.1, AP3, and P-PP2.x complexes each received a different simulation time. To determine whether the systems had reached convergence, the root mean square deviation (RMSD) of the trajectory with respect to the initial structure was calculated. It can be seen in Figure IV-10 that the fluctuation of RMSD differed from one structure to another. P-PP2.x explored a transition from one topology (100-300 ns) to another (450-950 ns). AP3 experienced the largest conformational shift comparing to the starting structure and no obvious transition state was observed at least during the simulation time. Instead, AP2.1 appears to switch between two conformation states roughly every 200 ns. How are the structures altered by these changes? To answer this, we need to find proper parameters to describe the topology change in a quantitative manner.

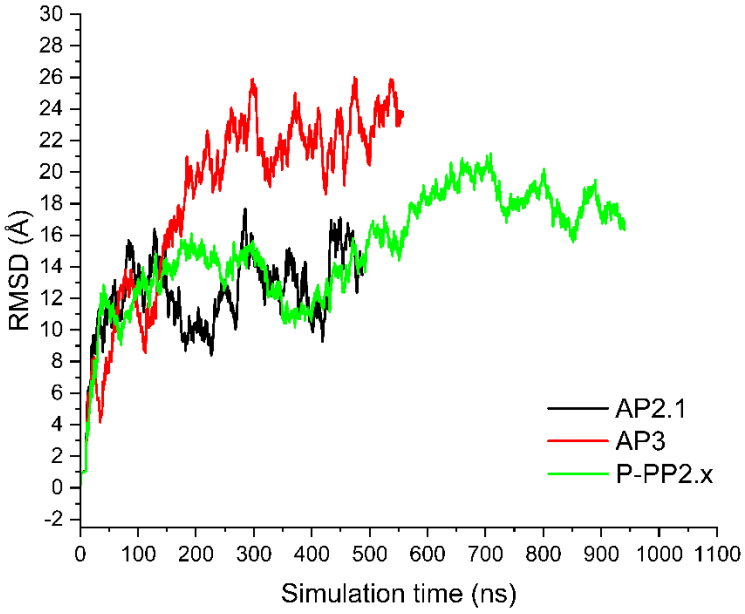


Figure IV-10 The root mean square deviation (RMSD) plot of the AP2.1, AP3, and P-PP2.x models during the respective simulation times of 487 ns, 557 ns, and 940 ns. The backbone atoms were used to calculate the RMSD values.

One of the most important factors in determining how a protein region fluctuates is its root mean square fluctuation (RMSF) value. This measurement gives information about how much a protein's shape changes over time, which can be used to forecast how the protein will behave. A geometric change can be used in place of RMSF to better illustrate how the secondary structure of the given protein changes over the course of the simulation. The geometric variation of the coiled coil - CTA2 linker, which joins the two CTBs, is the one in which we are most interested. There are various calculation methods to indicate the helicity of a protein domain[135], from which we selected the HELO algorithm[136], as not only it can describe more complicated geometry with a simple calculation, but it has also been used previously to compare the helicity of the A2 peptide between cholera toxin and heat-labile enterotoxin[137], which provides a precedent and insight into how to interpret the results.

IV.6 Helicity calculation with simulation states

In this section, we use the HELO algorithm to present a quantitative representation of the helicity of the linkage region. The HELO algorithm is relatively straightforward to apply because it only needs the coordinates of four nearby residues in the structure. Professor Suren Tatulian (University of Central Florida) kindly provided the precise calculation in Excel, and my task was to use Python to automatically apply this algorithm to a given MD trajectory. There are various parameters that can be used to represent the secondary

structure of a protein[136]. Along the linker region, the α-helix was divided into overlapping quadruplets, from i to i+3. The x, y, and z coordinates of each residue's Cα atom were used to calculate the three direction cosines ($cos\ \zeta$) of each quadruplet, and 1.5 Å rise was chosen as the distance between each residue and the helical axis. To calculate the local helicity, $1 - (cos^2\ \zeta_1 + cos^2\ \zeta_2 + cos^2\ \zeta_3)$ will be used as the criteria, and a non-helical geometry is indicated by a significant deviation of more than 0.2.

Before testing the CTB_5 dimer complex, a more straightforward question was addressed first: how flexible is the native CTA2 peptide following its association with CTB_5? This information cannot be directly retrieved from the PDB database because the uploaded holotoxin (PDB: 1xtc) contains a CTA1 that can interact with the intact CTA2 peptide and act to immobilize it. Therefore, it is unknown if CTA2 can continue to be rigid without CTA1. After performing MD simulations of the $CTA2B_5$ structure, we attempted to quantitatively represent the conformational changes of CTA2 using the HELO algorithm in order to confirm the flexibility of CTA2 itself in the $CTA2B_5$ complex.

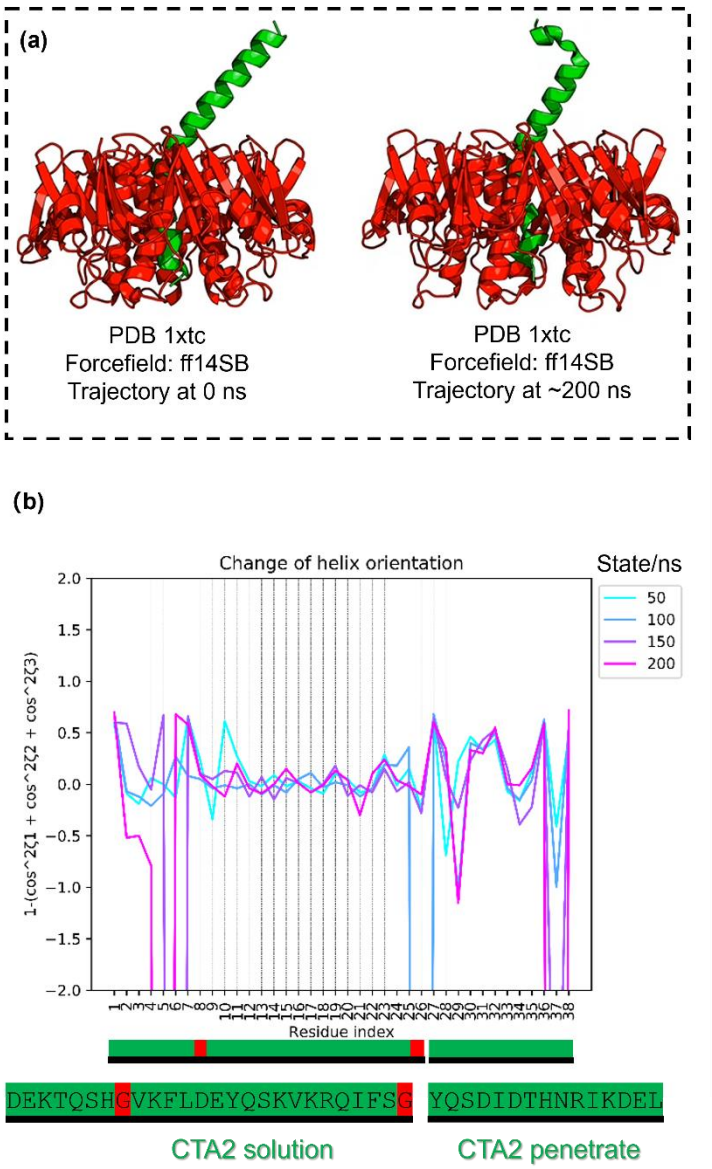


Figure IV-11 Description of the CTA2 peptide's degree of helicity in quantitative terms. (a) Two different simulation times with different $CTA2B_5$ structures. 0 ns represents the structure directly taken from the PDB database (PDB: 1xtc), with CTA1 and part of the CTA2 peptide removed. 200 ns represents one trajectory from MD simulation time. (b) A plot of $1 - (cos^2 \zeta_1 + cos^2 \zeta_2 + cos^2 \zeta_3)$ against the number of residues shows the conformational change at simulation times of 50 ns, 100 ns, 150 ns, and 200 ns. The two glycine residues at residue index 8 and 26 respectively contribute the tilt of linker regions. The protein sequence of used CTA2 peptide is also labelled, with the portion penetrate into the CTB_5 pore labelled as CTA2 penetrate, while the rest is labelled CTA2.

Firstly, two different MD simulations of the protein structure clearly show that the CTA2 peptide has some flexibility and, contrary to what we initially thought, its conformation in solution is not a rigid helical structure (Figure IV-11-a). More specifically, the N-terminal end of the helix disassembles into a random coil structure. When computational mapping is

carried out utilizing the HELO algorithm, this is even more obvious. $1 - (cos^2 \zeta_1 + cos^2 \zeta_2 + cos^2 \zeta_3)$ significantly deviates from 0 at the N- and C-ends (pink line) at 200 ns MD simulation time, demonstrating that they form a random coil structure (Figure IV-11-b). Following testing with the CTA2B$_5$ structure, the helicity of the linker regions of AP2.1, AP3, and P-PP2.x can be determined using the HELO algorithm.

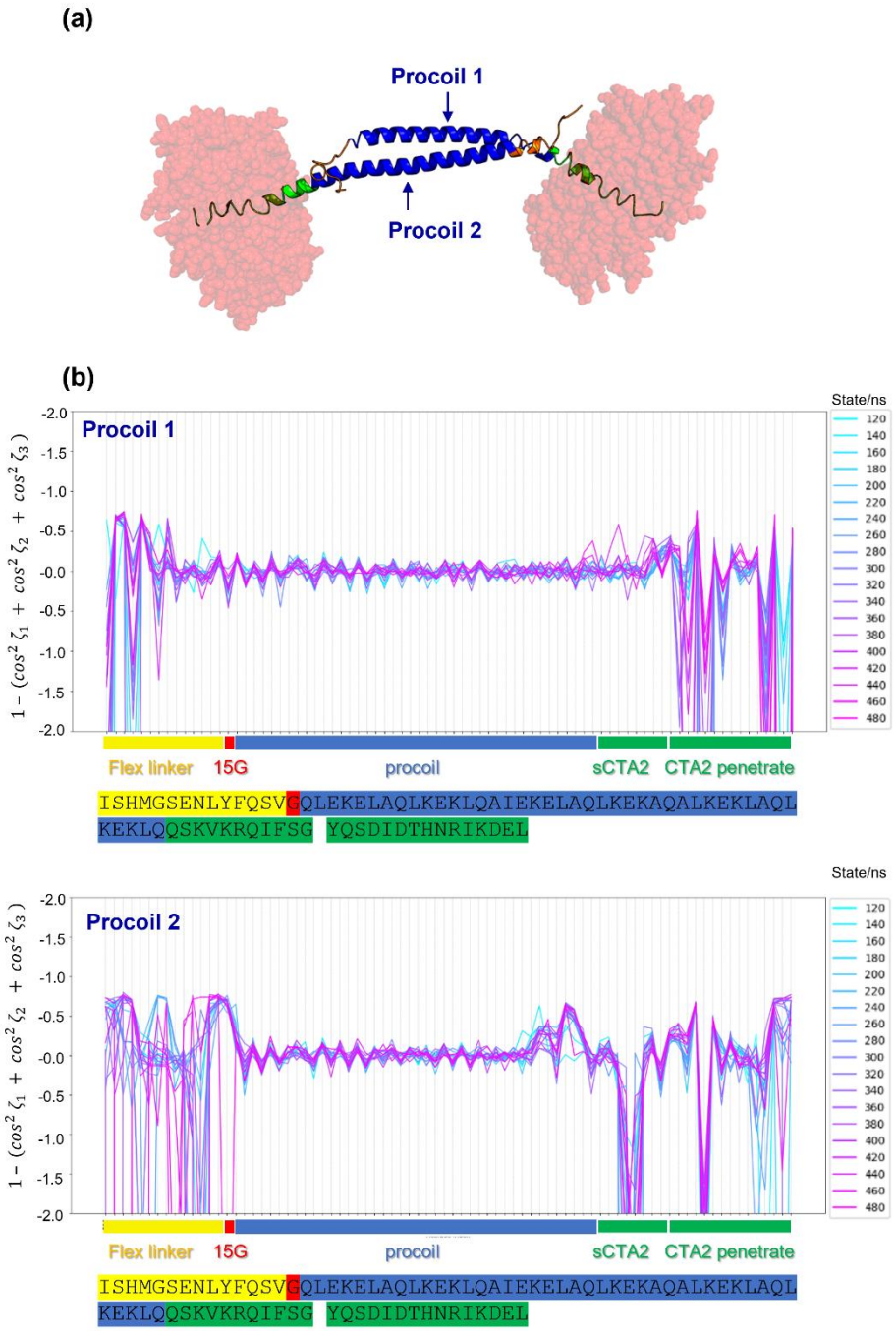


Figure IV-12 Quantitative depiction of the helicity of the *procoil*-CTA2 linker region from AP2.1 during the MD simulation. (a) One trajectory of AP2.1, in which the CTB$_5$ is depicted in sphere form (red) while the linker region

is the main focus and is depicted in cartoon form. The blue marked procoil 1 and procoil 2 regions come from the inserted coiled coil protein sequence. Short CTA2 peptide (green, sCTA2) is used for AP2.1. (b) $1 - (cos^2 \zeta_1 + cos^2 \zeta_2 + cos^2 \zeta_3)$ value along the linker region identifies the helixity and deviation from canonical α-helix.

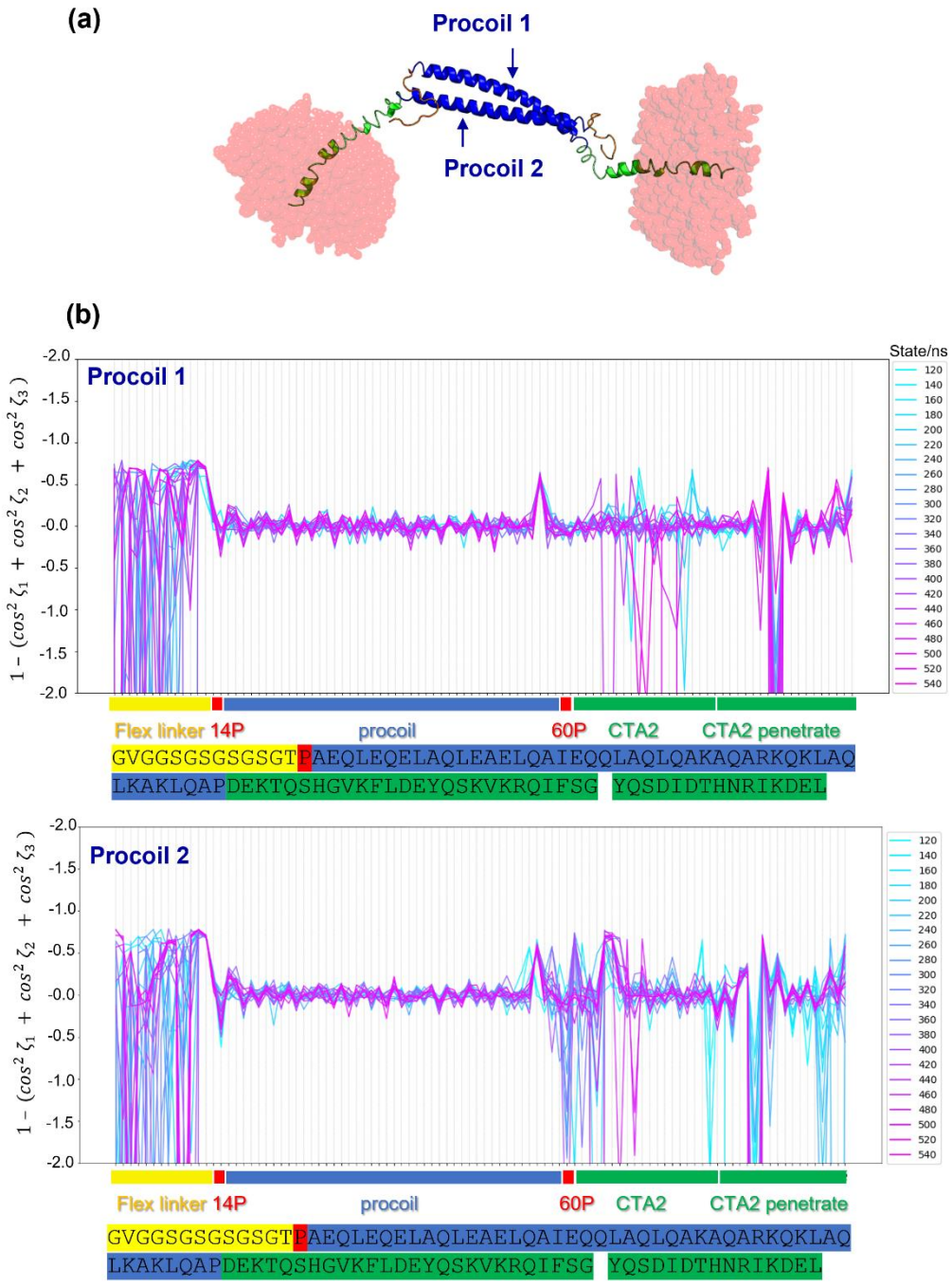


Figure IV-13 Quantitative depiction of the helicity of the *procoil*-CTA2 linker region from AP3 during the MD simulation. (a) One trajectory of AP3, in which the CTB_5 is depicted in sphere form (red) while the linker region is the main focus and is depicted in cartoon form. The blue marked procoil 1 and procoil 2 regions come from the inserted coiled coil protein sequence. Full-length CTA2 peptide (green, CTA2) is used for AP2.1. (b)

$1 - (cos^2 \zeta_1 + cos^2 \zeta_2 + cos^2 \zeta_3)$ value along the linker region identifies the helixity and deviation from canonical α-helix.

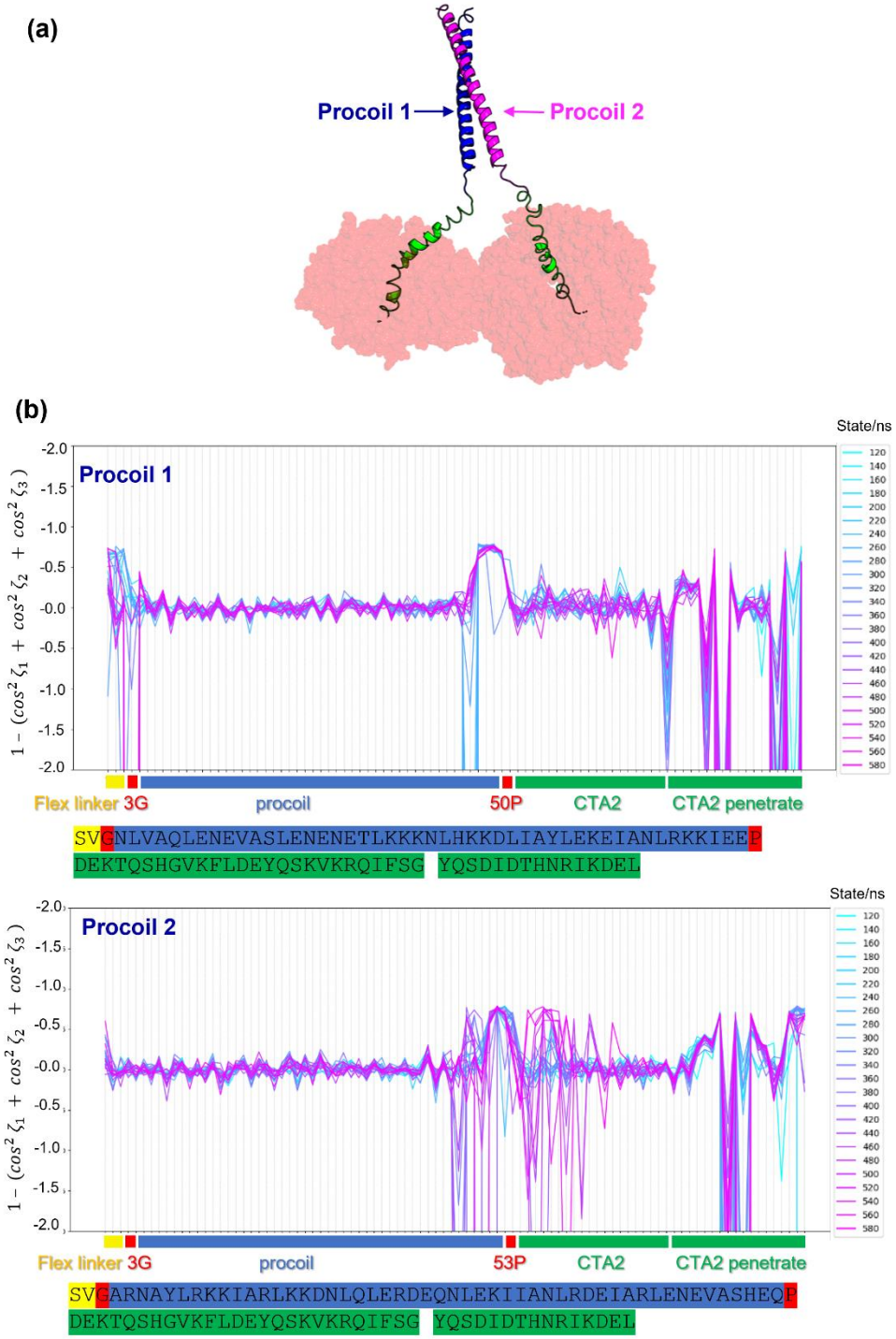


Figure IV-14 Quantitative depiction of the helicity of the *procoil*-CTA2 linker region from P-PP2.x during the MD simulation. (a) One trajectory of P-PP2.x, in which the CTB_5 is depicted in sphere form (red) while the linker region is the main focus and is depicted in cartoon form. The blue marked procoil 1 and procoil 2 regions come

from the inserted coiled coil protein sequence. Full-length CTA2 peptide (green, CTA2) is used for AP2.1. (b) $1 - (cos^2 \zeta_1 + cos^2 \zeta_2 + cos^2 \zeta_3)$ value along the linker region identifies the helixity and deviation from canonical α-helix.

Based on Figure IV-12, Figure IV-13 and Figure IV-14, the coiled coil region (highlight in blue) for all three proteins complexes remained as a helix throughout the MD simulation time. Hence, it is possible to use the constructed model to represent the conformation of these CTB_5 dimer complexes given that the coiled coil does not become dissociated during the simulation period. Moreover, since the flexible linker still remains at the N-terminus of the AP2.1, AP3 and P-PP2.x after the MBP was removed by protease, this portion show high flexibility with $1 - (cos^2 \zeta_1 + cos^2 \zeta_2 + cos^2 \zeta_3)$ varies significantly and larger than 0.5 – that indicate the formation of random coil. The variation at the CTA2 region is more intriguing because it exhibits varying degrees of random coil or broken helix formation in the CTB_5 dimer constructs during the 480-580 ns simulation time, despite the fact that the CTA2 peptide was largely maintained as a helix during the 200ns MD simulation of the $CTA2B_5$ complex (Figure IV-11). In addition to lengthening the simulation time, the attachment of each coiled coil may also have an impact on the CTA2 helicity. These outcomes, however, are only each from a single MD simulation run. Rebuilding the model (to eliminate differences in model construction) and running the MD simulation twice more are required to gain a deeper understanding of the relationship between CTA2 flexibility and various protein topologies.

IV.7 Distance and angle calculations based on the MD simulations

All of AP2.1, AP3, and P-PP2.x are made up of clear domains, which include CTB_5, CTA2 or short CTA2 peptide, and the coiled coil linkers. As a result, the differences between the three constructs can be seen in the angle and distance change between the two CTB_5. The angle between CTB_5 units was determined using vectors created by joining the centers of the two planes that were defined by the five **57H** and **79E** residues at the top and bottom of CTB_5, respectively (Figure IV-15). Additionally, the distance between the centroids of the five **57H** from each of the two CTB_5 is used to calculate the distance between two CTB_5 (Figure IV-15).

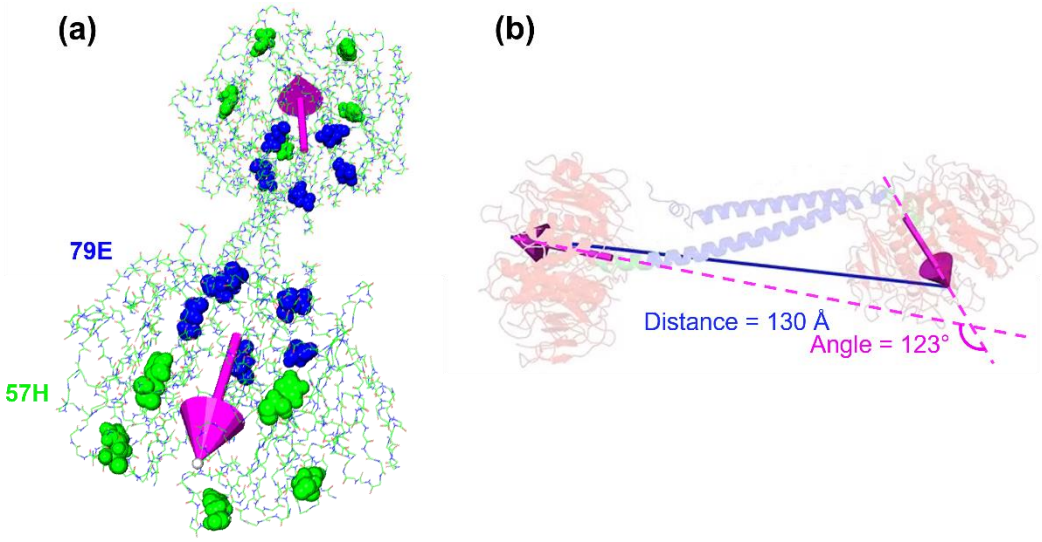


Figure IV-15 (a) The vectors for calculating the angle between CTB_5 pentamers was generated by connecting the centroids of two planes defined by the 57H and 79E residues highlighted with green and blue spheres, respectively. (b) A diagram showing how to calculate the angle and distance of AP2.1 for a single trajectory state during MD simulation.

It is more obvious to see how the CTB_5 complexes move in solution based on how the CTB_5 angles and distances alter over the course of an MD simulation (Figure IV-16). P-PP2.x is the most rigid protein complex among the three. The angle between two CTB_5 switches from 37 ± 4.3° (100-400 ns simulation) to 56 ± 3.1° (600-900 ns simulation) while the distance remains at around 70 Å (70.2 ± 1.2 Å during 100-400 ns and 72.2 ± 1.1 Å during 600-900 ns simulation time); this indicates a twist of the CTB_5 units relative to one another and does not necessarily increase their separation. However, the two CTB_5 units will now be more likely to form a 60-degree angled protein complex rather than being parallel to each other. This understanding keeps us from overly idealizing the parallel orientation, which is crucial for further membrane analysis. After 100 ns of simulation time, both AP3 and AP2.1 experienced an angle fluctuation (95-180° for AP2.1, 80-160° for AP3) and a distance fluctuation (~135-150 Å for AP2.1, ~145-175 Å for AP3). In fact, when two CTB_5 units point in the same direction, as in the case of P-PP2.x, each CTB_5 restricts the possible motion of the other. However, when both CTB_5 are arranged pointing in opposing directions, this restriction is removed, and it comes as no surprise that this arrangement results in a more flexible protein structure.

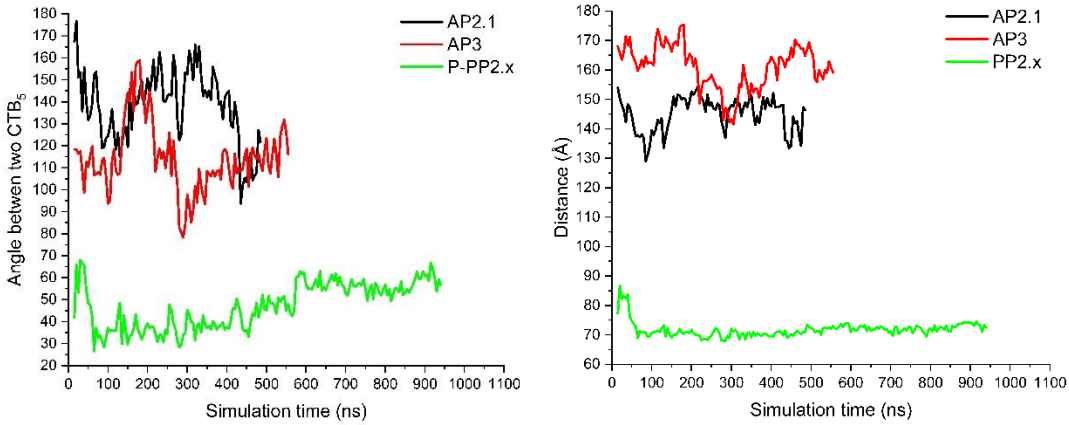


Figure IV-16 The angle and distance change comparing AP2.1, AP3 and P-PP2.x across simulation times.

The distance and angle distribution of the three protein complexes were calculated using the trajectory after initial equilibration in a 100 ns simulation run (*i.e.,* starting at 105 ns), and the results can be represented in histogram plots (Figure IV-17). We can infer that while the AP2.1 is more fixed in distance than AP3, it also has the greatest angle variation of all three complexes. AP3 can move in a more flexible manner in terms of the distance change of two CTB_5. P-PP2.x is the least flexible, and the two CTB_5 are fixed at the same plane with limited freedom of movement in terms of both distance and angle comparing to AP2.1 and AP3. These findings demonstrate that despite the use of parallel or antiparallel coiled coils as linkers, the resulting CTB_5 complex cannot be arbitrarily assumed to be parallel or antiparallel because there is still some flexibility between the two CTB_5 units.

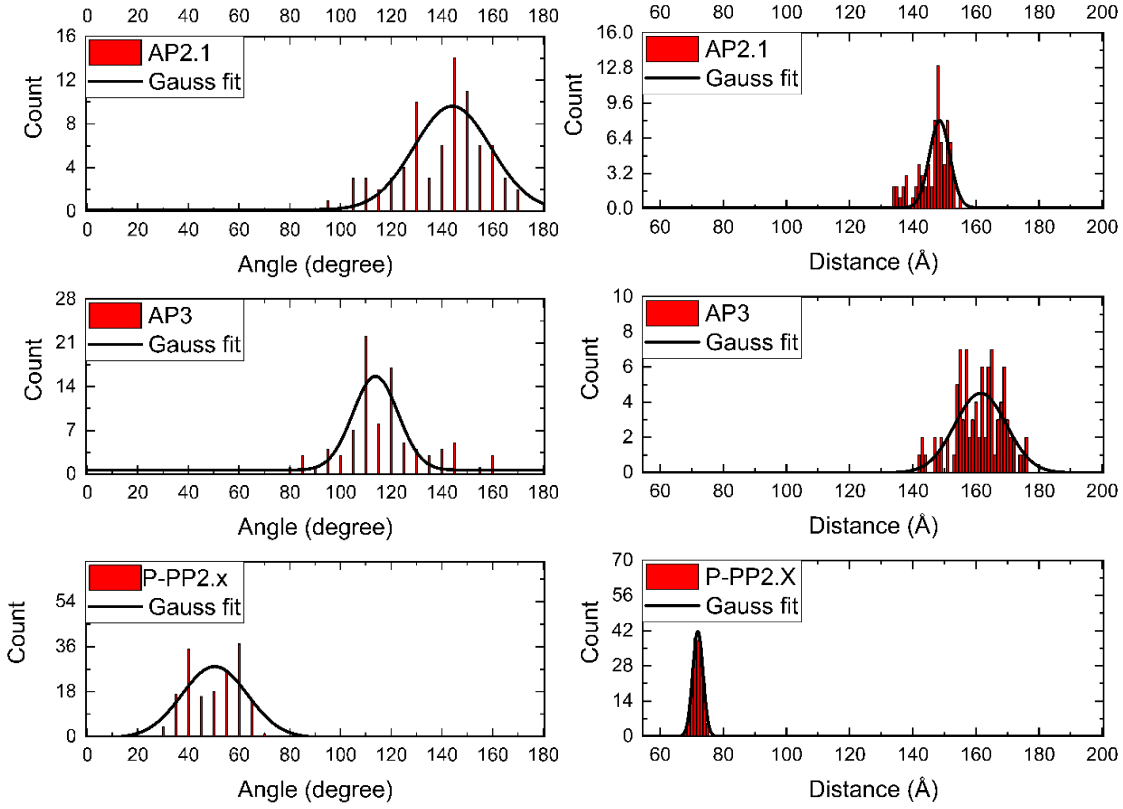


Figure IV-17 The Gaussian distributions of angle (left) and distance (right) calculation for the AP2.1, AP3 and P-PP2.x complexes modelled during MD simulations. The distance bin of 5 Å was used for the range 55-200 Å, and an angle bin of 5° was used for the range 0-180°.

This analysis gives us a general idea of how the three constructs move in solution. It is important to emphasize that it does not reflect any conformational changes that may take place when they contact the membrane. The interaction between CTB_5 and GM1 could potentially change the conformation of protein on the membrane. The subsequent section will discuss these possibilities.

IV.8 Atomic force microscopy

Atomic force microscopy was used to image the topology of the AP2.1, AP3 and P-PP2.x. The AFM analysis result was provided by Dr. George Health (University of Leeds). Pre-treatment of the mica surface with nickel ions allowed the CTB_5 to bind from that aqueous solution through interactions with the CTB_5's native histidine residues on the GM1-binding face. It is possible that various conformational results may arise depending on the concentration of nickel ions, since nickel ions are not evenly distributed on the surface.

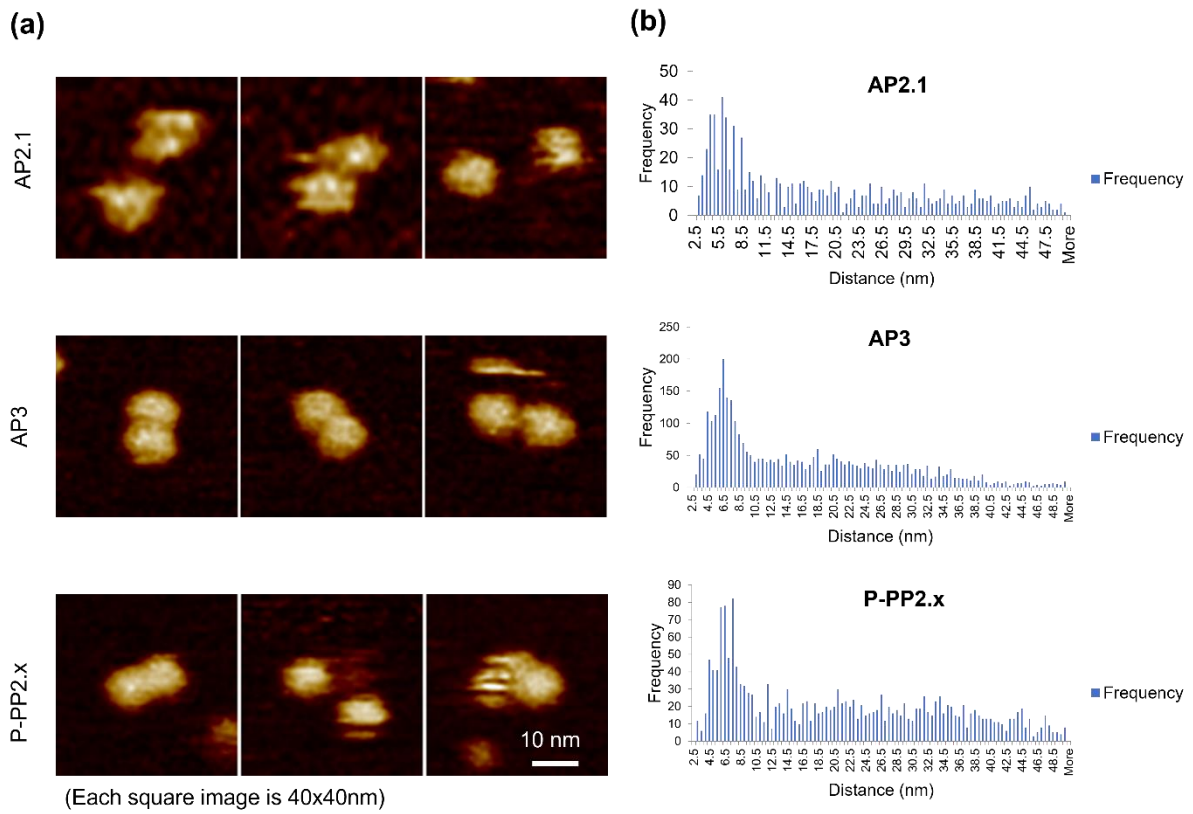


Figure IV-18 Atomic force microscopy results for AP2.1, AP3 and P-PP2.x. (a) Representative images for AP2.1, AP3 and P-PP2.x, respectively, bound to the mica surface under an aqueous solution. (b) Distribution of distances between surface-bound particles of AP2.1, AP3 and P-PP2.x. The frequency was determined by analyzing protein distribution from a singular experiment.

Using the auto-detection tool from George's lab, he was to calculate the distances between surface-bound particles of all three constructs from the AFM topology images (Figure IV-18; example calculation shown in Figure IV-19). The histogram used in the calculation included a distance range of 2.5 nm to 50 nm, and all three samples peaked at roughly 6 nm, which is the width of CTB_5 pentamer. This demonstrated that all complexes, regardless of whether they were initially linked with parallel or antiparallel coiled coils, tend to let both CTB_5 bind to the surface in close proximity to one another.

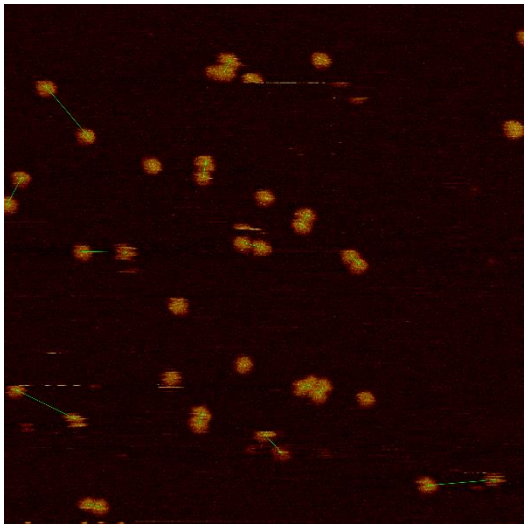


Figure IV-19 Example image showing distances measured between neighbouring CTB_5 particles according to the given AFM image.

The antiparallel coiled coil complex, such as AP2, is thought to interact with the mica surface in one of two ways: either by binding to nickel ions with one CTB_5 and leaving the second to stand in the solution – which might be further disrupted by AFM tip when it passes, or by each CTB_5 interacting with several histidine-nickel interactions on the surface and laying down the entire complex. These hypotheses, however, are predicated on the idea that the coiled region is sufficiently stable to prevent linker dissociation when the protein complex interacts with the membrane. What if it is more favorable for each CTB5 to bind to the nickel ions rather than for the coiled coil to remain intact? This involved a complicated interplay between the coiled coil binding strength, the flexibility of the linker region, and the interaction of CTB_5 with nickel ions. The AFM findings, however, suggested that coiled coils may break when they bind to a surface.

IV.9 Flow cytometry: binding of AP3 and AP2.1 to Jurkat cell membranes

After using AFM to obtain images of how proteins bind to solid mica planes, we unexpectedly found that two CTB_5 from the same construct seem to bind to the plane simultaneously, regardless of whether a parallel coiled coil or an antiparallel coiled coil is used. We therefore sought to learn more about how AP3 and AP2.1 interact with biological membranes. Flow cytometry (FC) is a technique to detect the physical or chemical characteristics of a population of cells or particles. It uses a laser beam as light source to produce fluorescence signal to detect fluorescent cell or particles. Using this technique, we want to figure out if the binding between protein and cell can be detected and quantified. Francesca Rosato (University of Freiburg) performed the flow cytometry detection by

incubating the Jurkat cell with AF 488-AP2.1 and AF488-AP3. Since Jurkat cells have GM1 ligand, they were chosen as the model cell line to test their binding affinities to various CTB_5 complexes. Nine different protein concentrations (0.25 nM, 0.5 nM, 1 nM, 2.5 nM, 5 nM, 10 nM, 15 nM, 20 nM, 40 nM), each labeled with Alexa Fluor 488, were used in the flow-cytometry analysis of Jurkat cells. Protein was added to cells and incubated with them for 30 min on ice; afterwards, the cell was washed to remove any unbound protein. The EC50, which represents the protein concentration required to occupy half of the GM1 in a given population of Jurkat cells, can be calculated based on the fluorescence intensity variation between various protein concentrations. Considering that NHS-Alexa Fluor 488 was used to label AP2.1, AP3, and $CTA2B_5$ with comparable degrees of labeling per CTB_5 (1.01, 0.41, and 0.4, respectively), a direct comparison of the fluorescence intensity across various incubation concentrations can therefore reveal the binding pattern. Because the proteins are not excessively fluorescently labelled, the instrument will not be able to reach the limitation of detection only in the presence of a small amount of protein.

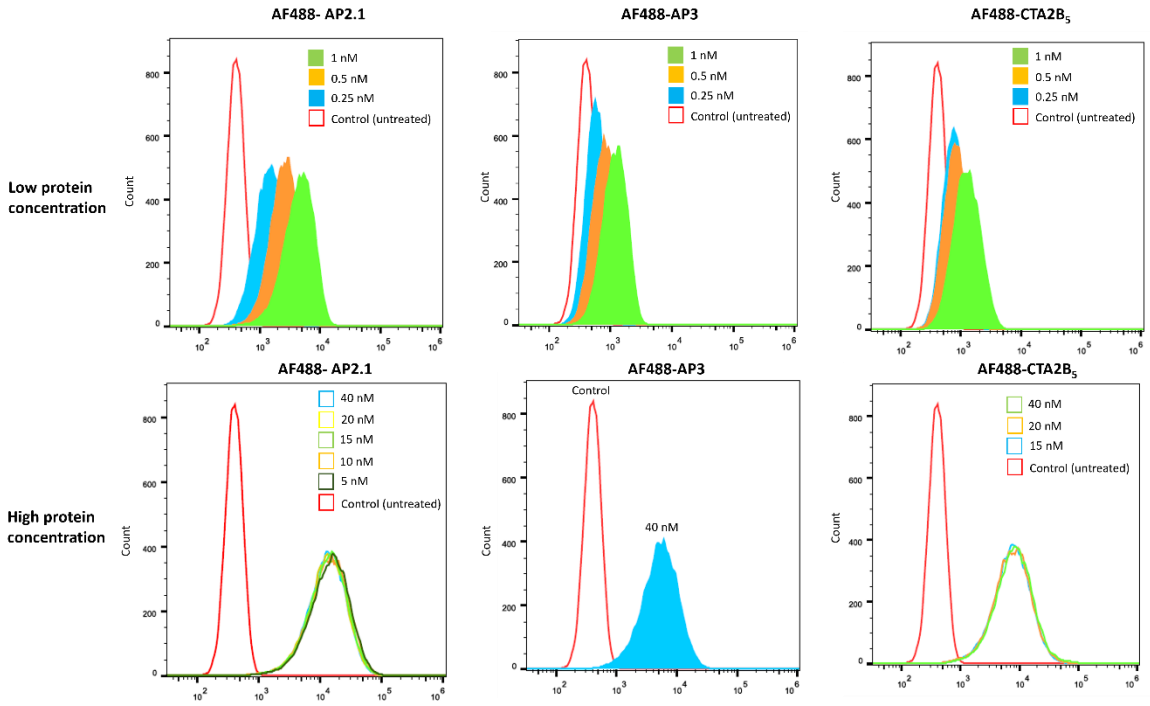


Figure IV-20 Jurkat cells that have been treated with AF488-labeled proteins were analyzed using flow cytometry. The binding of the detected protein complexes to the cell surface under low protein concentration is revealed by the histograms of florescence intensity. A threshold of fluorescence intensity was found for each protein when the protein concentration was greater than at least 5 nM (CTB_5 concentration). The presented plot is a representative outcome obtained from a series of four repeated experiments for each protein sample.

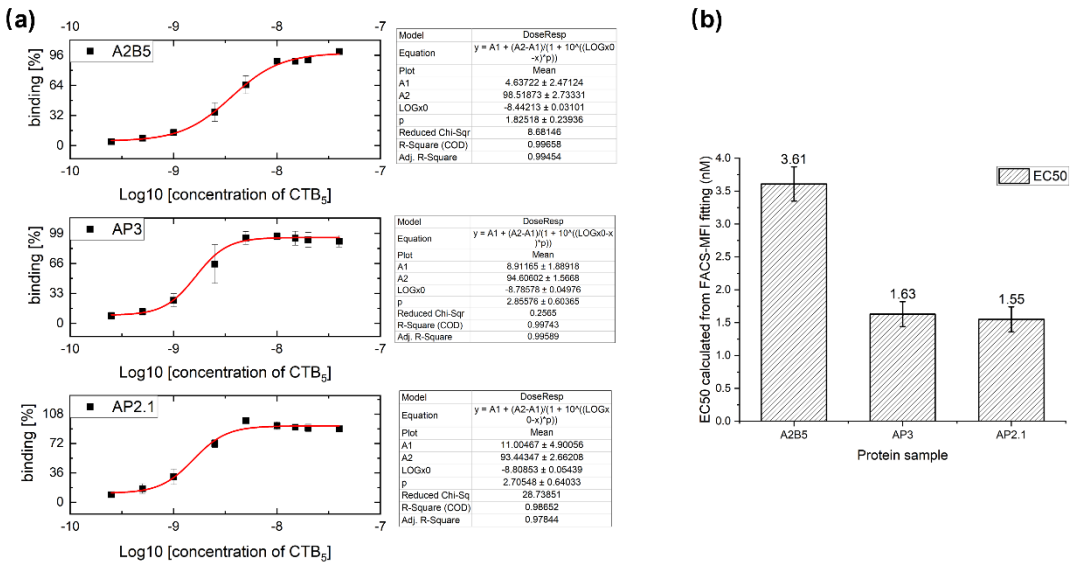


Figure IV-21 (a) The percentage of binding of Jurkat cells to each complex as a function of CTB_5 concentration (unit = mol/L). Jurkat cell was treated with different concentration of AF488-labelled protein before detecting by flow cytometry. The mean fluorescence intensity was extracted for each protein under given treated concentration and normalized to untreated control (repeat number = 4, mean ± s.d.), which was then further plotted against the lg value of different concentrations of protein (mol/L, calculated as CTB_5 concentration). A dose-response curve was further used to fit the obtained datasets. (b) The EC_{50} of $CTA2B_5$, AP3, and AP2.1 binding to Jurkat cells.

When incubated with 10 nM CTB_5, all three samples — $CTA2B_5$, AP2.1, and AP3 — reached binding saturation (Figure IV-20, Figure IV-21). Nevertheless, there are differences in the binding patterns. AP3 and AP2.1 demonstrated a higher binding affinity than $CTA2B_5$ despite the protein concentration being calculated equally based on the molarity of CTB_5, with EC50 values of 1.84 nM and 1.55 nM compared to $CTA2B_5$'s 3.61 nM. This shows that less CTB_5 is required for the CTB_5 dimers to occupy 50% of the GM1 in Jurkat cells, and it also provides clear evidence that the presence of coiled coils either helps the second CTB_5 attach to GM1 at the cell surface or prevents GM1 from interacting with other CTB_5. However, the second speculation is very unlikely because Figure IV-20 reflects that the cellular fluorescence intensity after incubating with high-concentration AF488-AP3 or AF488-AP2.1 is not significant smaller than that of AF488-$CTA2B_5$. Considering that the degree of labelling of AF488-CTA2B5 is higher than AF488-AP2.1 and AF488-AP3, we can speculate that even if the presence of the coiled coil inhibits the binding of other CTB_5 to the cells and thus reaches the saturation point earlier, the effect is minimal.

IV.10 Lipid mixing assay

The next investigation was to assess whether AP2.1, AP3 and P-PP2.x can lead to membrane fusion using lipid mixing assays and to determine which protein has the highest membrane fusion/hemifusion efficiency. The dequenching efficiency of the FRET signal produced by combining each protein with a population of liposomes that included LUV without dye and LUV containing 0.25 mol% NBD-DOPE and 0.25 mol% rhodamine-DOPE allowed for the determination of each protein's membrane fusion efficiency. A more detailed method concerning how the lipid mixing assay was performed is recorded in the Method section VII.14.

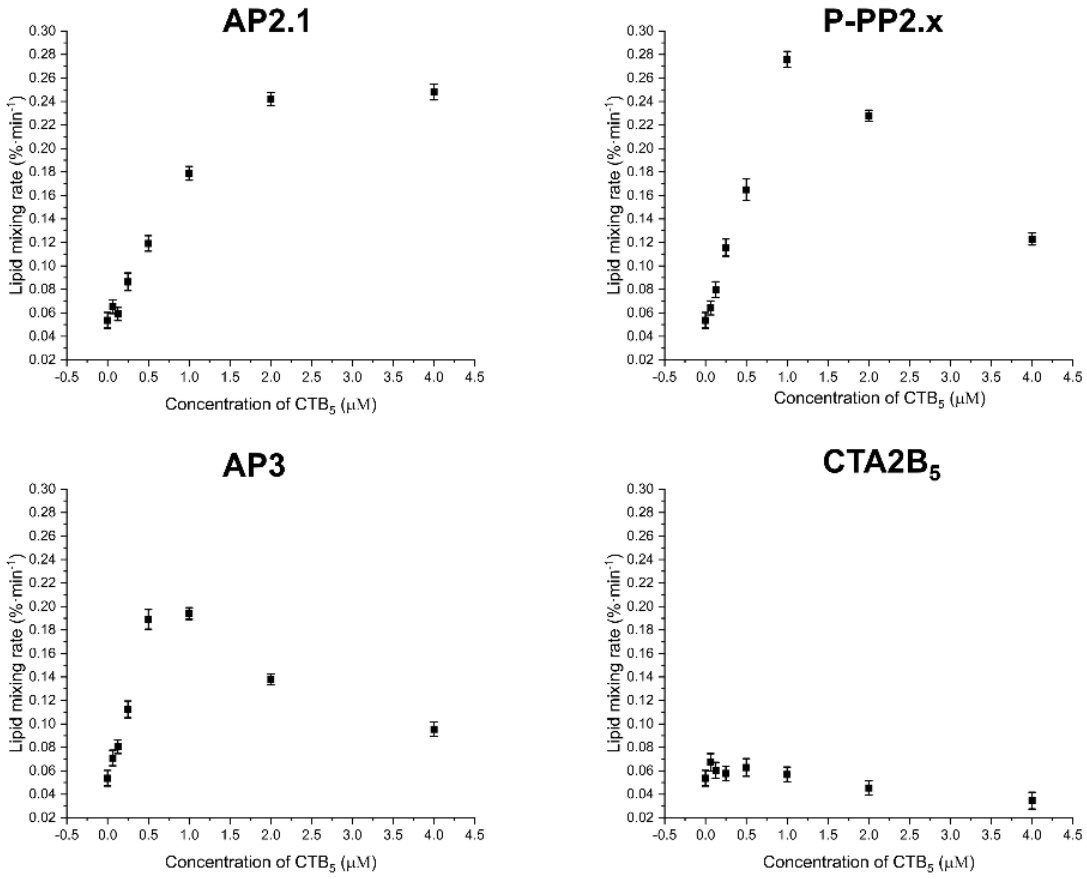


Figure IV-22 The protein-induced lipid mixing rate represents the percentage change per minute in dequenching of FRET between NBD-DOPE (0.25 mol%) and rhodamine-DOPE (0.25 mol%) in LUVs with a constant total lipid concentration of 133 μM (containing 5 mol% GM1). Protein concentrations are normalized by the CTB_5 concentration. The generated plots were derived from three repeated measurements (mean ± s.d.).

The efficiency of lipid mixing between unlabeled LUVs and LUVs labeled with NBD-DOPE and rhodamine-DOPE was poor in solution devoid of protein, as shown in the Lipid mixing rate when the concentration of protein is zero. The overall lipid mixing efficiency of the

liposome was similarly low after exposure to $CTA2B_5$ (Figure III-23). However, lipid mixing efficiency was significantly increased when exposed to optimal concentrations of AP2.1, AP3, and P-PP2.x. Even so, when compared at any specific CTB_5 concentration, these three proteins produced varying levels of lipid mixing efficiency.

It is critical to note that the negative control, $CTA2B_5$, did not improve the efficiency of lipid mixing between vesicles, which is a significant finding. $CTA2B_5$ was unable to perform membrane fusion/hemifusion while CTB_5 and biotin-$CTA2B_5$ can – as shown in section V.9 and II.3 respectively. The primary cause of this, in my opinion, is that CTA2 prevents direct interaction of CTB_5 with the adjacent liposomes. Without the interference of CTA2, direct contact or even minimal interaction between CTB_5 may be sufficient to induce lipid exchange. As a result, CTB_5 might have been self-dimerized on the membrane. Even though such dimers might be weak, it's possible that they could be strong enough due to the multivalency on the vesicle surface. After all, if two membrane surfaces bound by CTB_5 cannot be brought close to one another, membrane fusion should not be possible.

It is also intriguing to compare the rates of lipid exchange of different CTB_5 dimer complexes. While the lipid mixing efficiency of P-PP2.x and AP3 both decline after a certain optimal concentration for lipid exchange, AP2.1 induces faster lipid mixing in a concentration-dependent manner even at the highest concentrations tested. How does this situation make sense? It has been demonstrated that the rate-limiting step of lipid exchange is priming and docking between two vesicles in experiments involving membrane fusion triggered by SNARE proteins[138] or synthetic coiled coils[139]. When $CTA2B_5$ binds to the vesicle surface, the CTA2 peptide projects away from the vesicle surface, potentially preventing contact between nearby liposomes and thereby creating a shield that blocks lipid exchange. Since the coiled coil linkers permit the assembly of multiple CTB_5 in a relatively rigid manner, it is possible for all three CTB_5 dimer to communicate between two vesicles, regardless of how quickly or slowly the exchange takes place. The architecture of the protein assemblies, including the length and flexibility of the linkers, may be the key to understanding the different results of lipid mixing.

As we deduced from the AFM and flow cytometry results (section IV.8 and section IV.9 respectively), it is still possible for both CTB_5 in the same protein assembly to interact with the same membrane surface even when an antiparallel coiled coil is used, as in the case of AP2.1 and AP3. In the aforementioned example, the interfaces for protein binding are the nickel-treated mica chip and the cell, respectively. If this discovery holds true for LUVs as well, the two CTB_5 units from the same protein assembly should occupy the GM1 of the

same vesicle first and only bind to the neighboring liposome when there is a free GM1 or free GM1 binding site. This theory explains why P-PP2.x and AP3 have optimal concentrations for lipid exchange because when the protein concentration is too high, the CTB_5 dimers prefer to wrap around the liposome and saturate all the GM1 from the same liposome – which leaves limited spare GM1 molecules for interaction with CTB_5 on nearby vesicles (Figure IV-23).

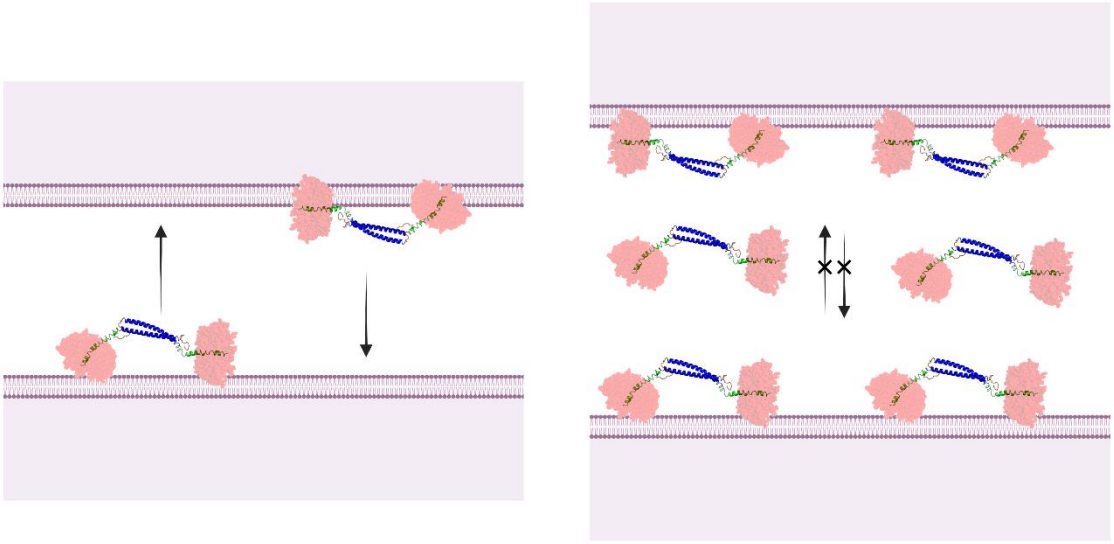


Figure IV-23 Scheme demonstrating the reason behind decreasing speed of lipid mixing when the concentration of coiled coil CTB_5 dimer increases beyond its optimal value from low (Left) to high (Right).

The behavior of AP3 and P-PP2.x can be well explained by this theory since both proteins reach their maximum lipid exchanging rates at about 1.3 μM CTB_5; at this concentration, the surface GM1 on the 5 mol% LUV should be almost completely occupied. AP2.1, on the other hand, behaves differently. Would it be possible, given that AP2.1 is more rigid than AP3 due to its adaptation of a very short CTA2 peptide linker, that the coiled coil is untangled because it is not flexible enough to bind to the surface with both CTB_5?

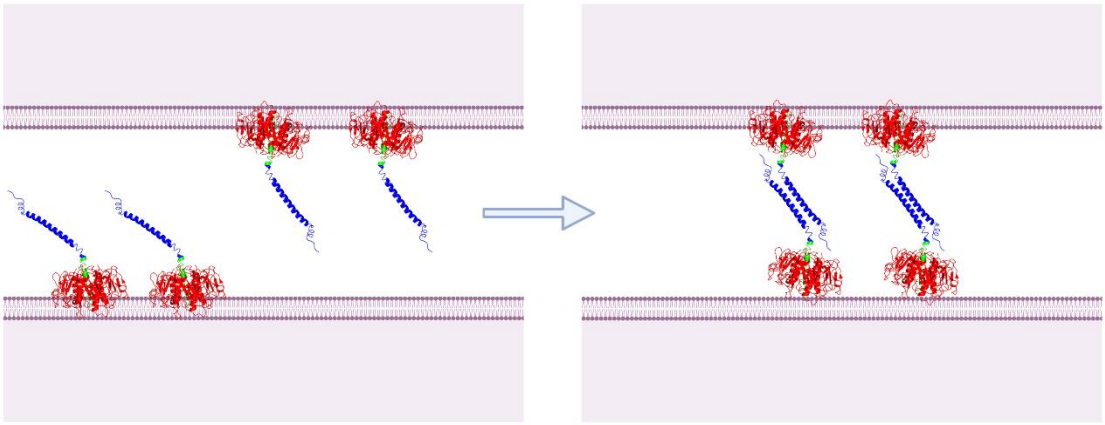


Figure IV-24 Scheme demonstrating a potential mechanism behind membrane contact driven by AP2.1. The coiled coil belonging to AP2.1 breaks when both CTB_5 of the same protein complex interacts with the membrane. The untangled *procoil* can further form an antiparallel coiled coil with another *procoil* from the opposing membrane.

If the coiled coil from AP2.1 dissociates when it interacts with GM1 on the membrane, the resulting *procoil* may combine with similarly unraveled *procoil* on the adjacent membrane to create a new coiled coil (Figure IV-24). If this is the case, then membrane fusion may not occur at an optimal concentration beyond which the rate decreases. Instead, it would result in a reaction that is still concentration-dependent but reaches a plateau that also depends on the affinity of the coiled coil that provides direct contact between adjacent membranes. The AP2.1 outcomes of the lipid-mixing experiment agree with this hypothesis. To further support the validity of this conclusion, QCM-D experiments will be conducted as the next step.

Additionally, despite the fact that AP3 and P-PP2.x have similar optimal concentrations at around 1 µM CTB_5, P-PP2.x can result in faster lipid mixing at its optimal concentration. This can be explained by a closer distance between two CTB_5 in P-PP2.x, which would enable a closer membrane contact. This is consistent with the outcomes of other studies on the rate of lipid mixing that incorporating heterodimer coiled coils in membranes[140].

IV.11 QCM-D analysis

Quartz Crystal Microbalance with Dissipation (QCM-D) allows real-time monitoring of surface phenomena, including mass change and energy dissipation on the sensor surface. This information can reflect changes in the structure at the surface, including mass uptake or loss, softness change of material etc. The outcomes are traces of the frequency and dissipation changes of the Quartz sensor, as detected in real time (Figure IV-25).

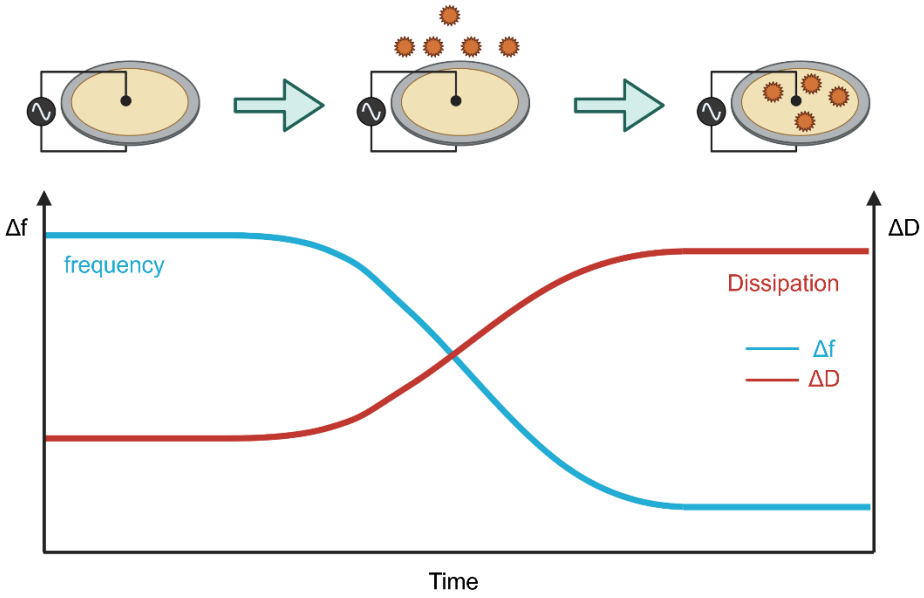


Figure IV-25 Schematic illustration of the frequency and dissipation change when material is absorbed to a quartz crystal surface. As the material is attached to the sensor surface, the decrease of frequency indicates the mass uptake, while the change of dissipation reflects the layer softness.

The mass change is calculated from the frequency shift (Δf) based on the Sauerbrey equation (Equation 1). Furthermore, for rigid films such as supported lipid bilayers or proteins, the mass change can be converted to thickness, which can reflect how materials are attached to the surface (Equation 2). This model assumes that the added layer can be considered as part of the oscillating crystal, hence the only the material that is attached on the sensor will be reflected in the machine – which allows the use of the Sauerbrey equation for calculating the mass change based on frequency shift.

$$\Delta m = -\frac{C \cdot \Delta f}{n}$$

Equation 1 Sauerbrey equation used to convert frequency change (Δf) to mass change (Δm). C is the mass sensitivity constant which relates to the property of quartz, and equals to 18.06 ± 0.15 ng·cm^{-2}·Hz^{-1} for the sensors used here. n represents overtone number, e.g., 1, 3, 5, 7.

$$d_{eff} = -\frac{\Delta m}{P_{eff}}$$

Equation 2.d_{eff} indicates the calculated thickness of the applied layer, with unit of angstrom. P_{eff} represents the effective density of the adhering layer, 1 for lipid and 1.4 for protein.

In the following QCM-D analysis, supported lipid bilayers (SLBs) are used as the model membrane system for the following reason: The lipid composition can be varied during the

small unilamellar vesicles (SUVs) preparation step, which involves only changing the ratio of lipids in solution. Although it is still difficult to make sure the lipid components are distributed evenly on the layer.

The aim of performing QCM-D was to understand how the proteins interact with the membrane through GM1-CTB_5 interactions, and further deduce whether AP2.1 has undergone a coiled coil break. I planned a protocol in which three layers of materials are created on the QCM-D to compare the different behavior of the protein samples in terms of their frequency and dissipation change. First was the SLB layer, that was formed by passing SUVs (50 µg/mL, 5 mol% GM1 in DOPC kindly provided by Chunyue Wang, University of Leeds) over the cleaned quartz sensor. After washing out the excess liposomes, different protein samples (constant CTB_5 concentration of 400 nM) were added to allow the interaction between protein and GM1 glycolipids, thus forming a second layer. The third layer was created by passing an extra layer of SUVs (50 µg/mL 5mol% GM1-SUVs) – the idea being the attachment of a third layer would imply the existence of free CTB_5 at the surface, and hence demonstrate the relative orientation of two CTB_5 in the same protein assembly. Before performing the third step with GM1-containing vesicles, SUVs composed of only DOPC was used as a negative control for the third layer (Figure IV-26).

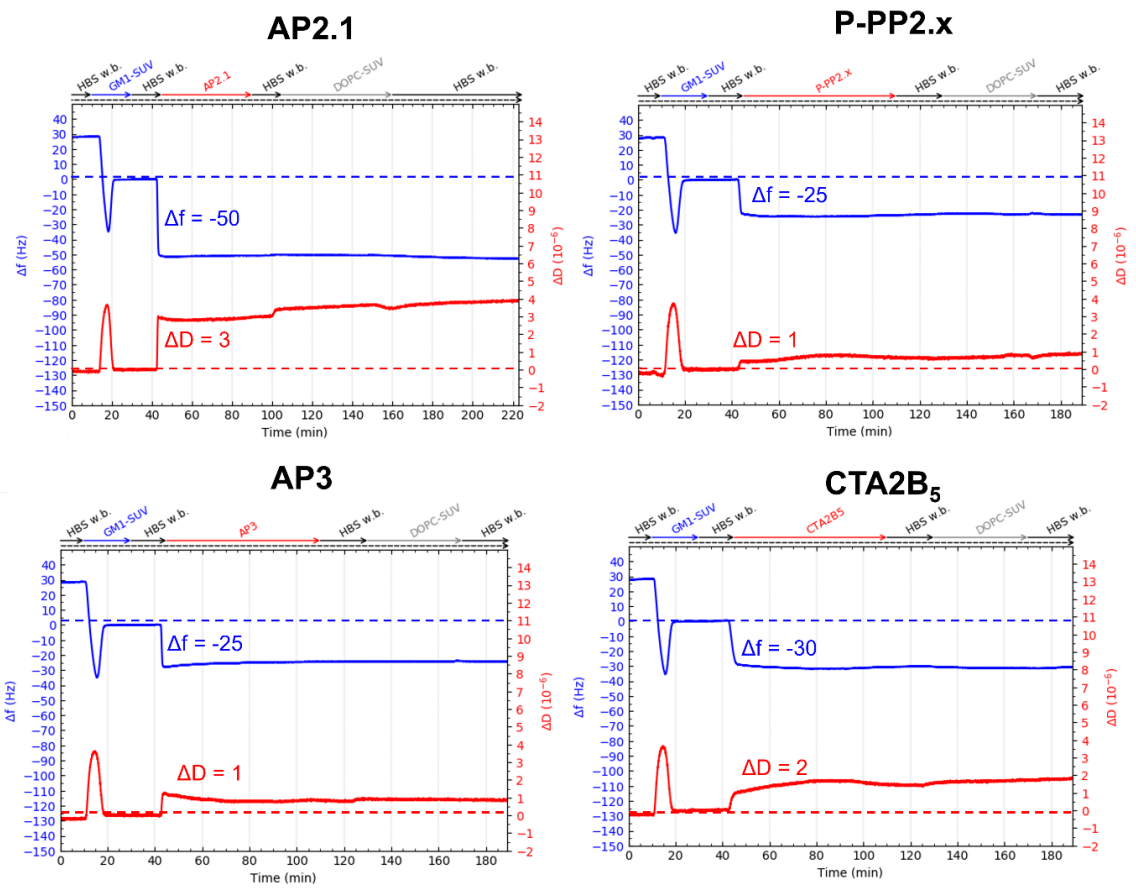


Figure IV-26 QCM-D result of attaching different proteins on 5 mol% GM1-coated supported lipid bilayer, and then passing another layer of DOPC-only SUV to check if any non-specific binding occurred between the attached proteins and DOPC-only SUV. The experiment depicted here is a single-repeat trial.

Exposure of 5 mol% GM1-SUV to the sensor surface resulted in a two-phase change for both frequency and dissipation (Figure IV-26, 10-20 min). The absolute value of Δf first increased and then decreased, which indicates the initial attachment of SUV on the surface and then the rupture of vesicles to form a supported lipid bilayer (SLB). The dissipation changes also reflected the conversion of a “soft” liposome material into a “rigid” bilayer. In the end, the Δf had around -28 Hz shift while ΔD was less than 0.3 × 10^{-6}, which implies the successful formation of SLB in good quality. As such, the origins were reset as the plateau after the SLBs were formed to allow a more straightforward comparison between proteins.

$CTA2B_5$ served as the control to check if QCM-D could be used to determine values for parameters such as the thickness of the layer after the protein was passed over the SLB. According to Equation 1 and Equation 2, the thickness of the protein layer was calculated to be 77.4 Å, while the direct measurement of $CTA2B_5$ from the crystal structure (PDB: 1xtc) was around 61 Å. Remembering that the CTA2 peptide can also unravel into a more flexible

structure based on the MD simulation result (Figure IV-11), the detected thickness from QCM-D is within the acceptable range. Furthermore, even if there is certain P_{eff} value to represent protein from Equation 2, it does not count into the solvation shell formed on the surface. Hence the final calculation could be biased.

P-PP2.x and AP3 had a similar frequency and dissipation shifts to $CTA2B_5$, which implies that both of the CTB_5 are attached to the membrane surface, rather than one CTB_5 binding while the other is facing toward the solution. However, AP2.1 behaved differently from AP3, P-PP2.x and $CTA2B_5$ from the perspective that it caused a much larger frequency shift (-50 Hz). The calculated thickness of the AP2.1 layer is 12.9 nm which corresponds to the computationally calculated thickness of AP2.1 (Figure IV-17). Moreover, AP2.1 has a larger dissipation change – even though it is unfair to directly compare only the dissipation value among samples with different mass change, this phenomenon can still reflect AP2.1 forming a softer layer on the surface, which either resulted from the unravelling of its coiled coil, or that AP2.1 is standing on the surface with only one CTB_5 rather than with both CTB_5 laying on the same bilayer.

A further application of DOPC-only SUVs was used to check for any unspecific protein-lipid interaction (Figure IV-26). Indeed, AP2.1 showed a slight attachment to the liposome, which added credit to the explanation that AP2.1 might have unraveled since it would allow the exposure of the hydrophobic interface between *procoils*. Amphipathic helices are known to be able to cause different kinds of membrane bending, but only when they are in contact with the membrane. They do this because they have hydrophobic and charged residues between their two faces. Since the hydrophilic face will rest on the exterior of the vesicle as the hydrophobic residues can insert into the lipid molecular layer[141].

With the background test in hand, the three-layer experiment was repeated with the third layer changed to 5 mol% GM1-SUVs (Figure IV-27). Surprisingly, the $CTA2B_5$ managed to interact with the GM1-SUVs, although it had failed to bind DOPC only-SUVs. This result indicated the binding of protein on the 5 mol% GM1-SLB surface might be more complicated than originally thought: CTB_5 may not orient all its binding sites toward to the membrane, which means that its five GM1 binding sites may not be fully occupied by GM1. It is more likely that at least some of the CTB_5, under the conditions implemented, have only a few GM1 binding pockets occupied, resulting in the CTB_5 binding to the membrane laterally and thus exposing excess GM1 binding pockets toward the solution. Interestingly, the frequency and dissipation shifts after passing GM1-SUV over AP2.1 and $CTA2B_5$ are similar but yet smaller than those for P-PP2.x and AP3; the reason behind this might be the coiled coil

decreases the freedom of CTB_5, makes the interaction of the CTB_5 with GM1 more difficult, and therefore more CTB_5 will be exposed to the solution to interact with the GM1-SUV in the third layer. Furthermore, the possibility that the second CTB5 from AP2.1 is facing the solution and only one CTB5 is left standing on the SLB is eliminated because little GM1-SUV was bound to AP2.1 in the third layer. In conclusion, the unraveling of the coiled coil is the most logical explanation for the interaction between AP2.1 and GM1-SLB. Triplicate QCM-D experiments gave very similar results, both in terms of the size of frequency and dissipation changes, and also the overall shapes of the curves.

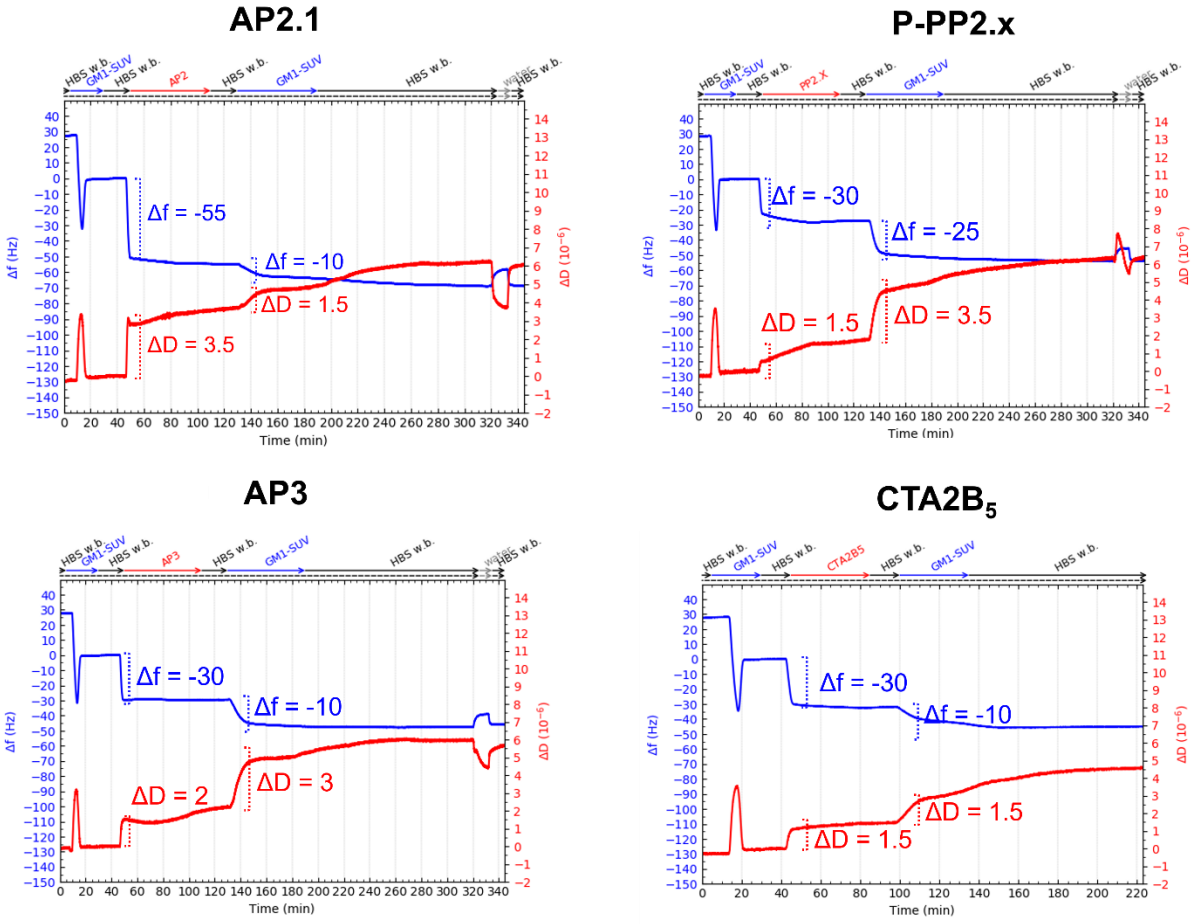


Figure IV-27 QCMD result of attaching different protein on 5 mol% GM1-coated supported lipid bilayer, and passing a second layer of 5 mol% GM1-SUV to check for further GM1-dependent binding between the protein and second SUV. The plot portrays the representative result attained from multiple repetitions of the identical experiment carried out on each protein sample.

Another experiment can also demonstrate that CTB_5 binds to the membrane laterally and that this binding facilitates the adsorption of more vesicles in the third layer. Comparing the QCM-D between classical variant CTB_5 and $CTA2B_5$ (Figure IV-28), the classic CTB_5 binds

more GM1-SUVs despite the fact that it has smaller frequency and dissipation shifts upon binding to the SLBs. The reason for this phenomenon could be that the presence of the CTA2 peptide either prevents the CTB_5 from binding to the membrane in a perpendicular manner or it provides spatial interference blocking the specific binding of the third layer of GM1-SUVs to the protein (Figure IV-28).

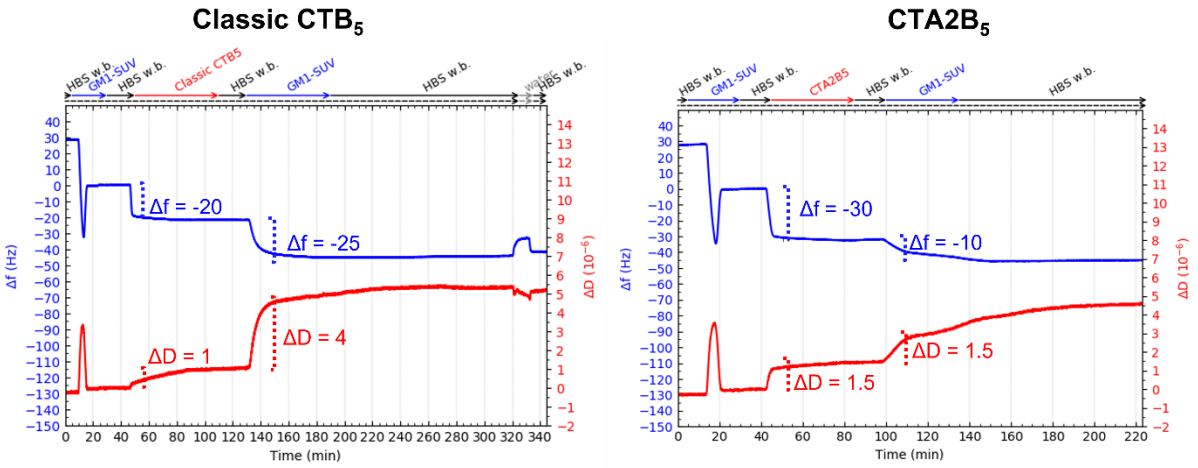


Figure IV-28 QCM-D result of attaching classic CTB_5 on 5mol% GM-coated SLB. When passing the second layer of 5 mol% GM1-SUV over the bound classic CTB_5, the surface properties underwent significant change. The experiment depicted here is a single-repeat trial.

Typically, we pass HBS buffer to wash away any excess in solution after attaching a substance onto the surface of QCM-D sensor. During the washing process, the frequency and dissipation curve usually tend to stabilize. Since the washing buffer is exactly the same as the storage buffer for the previous attached substance, there should not be any frequency changes or dissipation when the system is switched to the washing buffer, because there are no extra ion to interfere with the solvation shell. As a result, the variation of frequency and dissipation observed can only come from the absorbed substance.

Hence, one phenomenon that cannot be ignored: the Δf and ΔD changed further during the rinsing of the extra GM1-SUV vesicles with buffer after the third layer of SUV coverage (Figure IV-27– period labelled as HBS w.b.). This indicates that during buffer rinsing step, vesicles underwent some sort of rearrangement, but it does not necessarily imply that vesicle rupture, hemi-fusion, fusion, or other vesicle-vesicle interactions are occurring, since there is no obvious two-phase change of Δf and ΔD as in the case of when SUVs rupture to form an SLB. What is more intriguing for classic CTB_5 is that there is no obvious signal for this kind of rearrangement – only a small increase of ΔD. Perhaps vesicle rearrangements are induced by the presence of CTA2 and its derivatives. When CTB_5 is perpendicularly

bound to vesicles, the presence of many vesicles in the surrounding area could cause CTA2 to remain tense. However, this rigidity could be reduced when the environment is devoid of an excess of vesicles, leading to increased flexibility of CTA2. Vesicles could therefore rearrange themselves as if they were fish swimming freely through algae.

The conclusion drawn from this study is that the QCM-D results are in line with the hypothesis drawn from the lipid mixing experiment: when AP2.1 encounters the membrane, it undergoes a rearrangement and the coiled coil separates into *procoils*, as opposed to AP3 and P-PP2.x, which preserve the integrity of their coiled coils. Additionally, there is always a portion of CTB_5 bound perpendicularly to the membrane for both AP2.1 and other CTB_5 dimers at a protein concentration of 400 nM (CTB_5 concentration) and a 5 mol% GM1 concentration in the bilayer. Particularly for AP3 and P-PP2.x, this binding mode is essential for determining the interaction with nearby liposomes. Additionally, given that there were fewer vesicles bound to $CTA2B_5$ than there were to Classic CTB_5 (Figure IV-28), CTA2 reduces the efficiency of binding to the neighbouring membranes. This may explain why $CTA2B_5$ does not result in lipid exchange during lipid mixing experiments, while CTB_5 does. However, because the SLB is stiffer than the vesicles, it does not completely reflect the protein-membrane interactions on vesicles, which is a limitation of QCM-D. Therefore, even though CTB_5 binds perpendicularly to SLB, this does not necessarily mean that the response on vesicles will be the same. The QCM-D assay can be used as a tool to help understand what is happening rather than providing absolute proof for nature of the protein-vesicle interactions.

IV.12 GUV fusion test

Finally, confocal microscopy was used to determine whether membrane fusion was taking place for all of the CTB_5 dimers. Erik Kempmann (University of Freiburg) carried out the AP2.1 preliminary experiments with the assistance of Lina Siukstaite (University of Freiburg), I carried out the preliminary GUV experiments of AP3 and P-PP2.x with Erik during my secondment, and Erik then carried out all of the GUV experiment repeats under the supervision of Lina Siukstaite. In the experiment, two vesicle compositions were used: one contained 0.3 mol% Atto647, 64.7mol% DOPC, and 30 mol% cholesterol 5 mol% GM1, while the other contained 65 mol% DOPC, 30 mol% cholesterol, and 5 mol% GM1. This enabled us to determine whether lipid exchange was taking place. All proteins were used at 400 nM CTB_5 final concentration. As soon as AP2.1 was added to the solution containing the GUVs, significant cross-linking between membranes was seen right away, as well as GUV

fusion. For AP3, a portion of GUVs fused as well, but membrane fusion in P-PP2.X was much less common.

Figure IV-29 depicts a series of two AP2.1-induced fusion events taking place between three nearby GUVs. The vesicle at position p1 initially merges with a vesicle that is not in the same viewing plane, creating a larger vesicle and appearing to transfer the membrane to the progressively larger adjacent GUV. This is followed by the fusion of two vesicles at the p2 position.

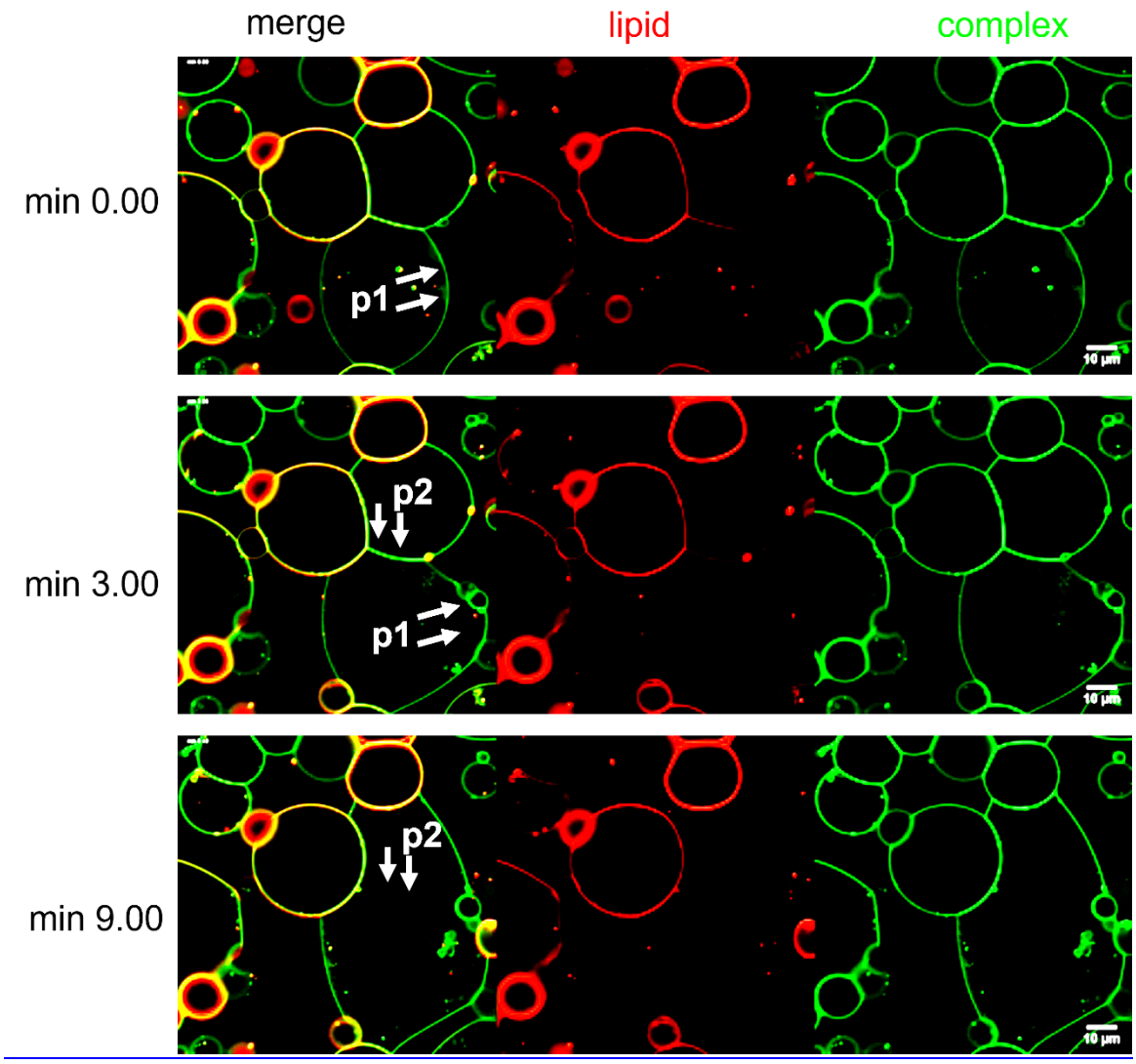


Figure IV-29 Time-lapse of GUVs after the addition of AF488-AP2.1 (final CTB_5 concentration is 400 nM). Two types of GUV were mixed – one with composition of 65 mol% DOPC, 30 mol% Cholesterol, 5 mol% GM1; the other contained 64.7 mol% DOPC, 0.3 mol% Atto647, while the rest of the components remained the same. Presented here are one of the result image obtained from four times of repeats.

In Figure IV-30, membrane fusion caused by AP3 is presented as a more direct process in which the interface at the p1 position gradually elongates over time, leading to membrane fusion in the end.

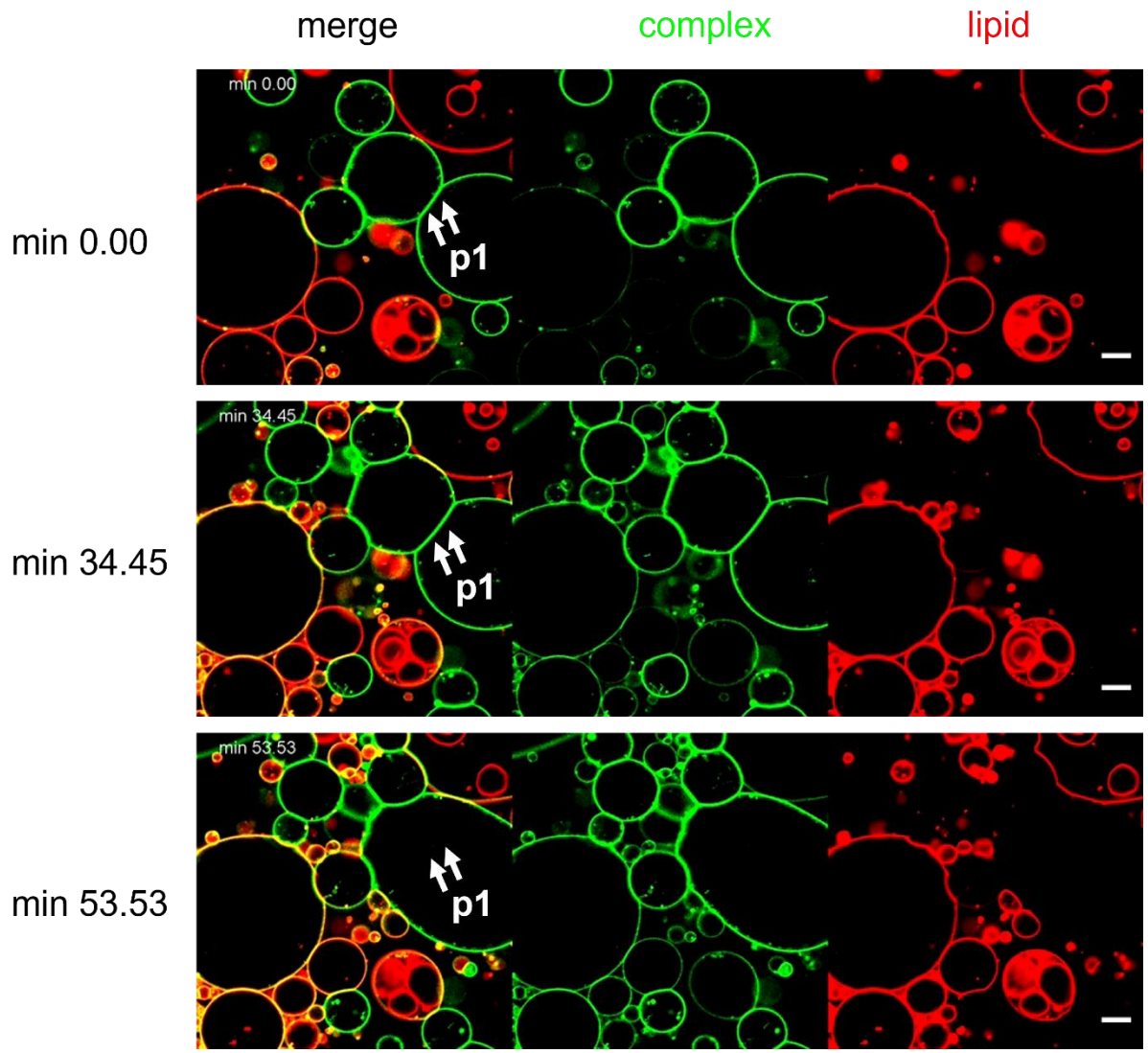


Figure IV-30 Time-lapse of GUVs after the addition of AF488-AP3 (final CTB_5 concentration is 400 nM). Two types of GUV were mixed – one with composition of 65 mol% DOPC, 30 mol% Cholesterol, 5 mol% GM1; the other contained 64.7 mol% DOPC, 0.3 mol% Atto647, while the rest of the components remained the same. Presented here are one of the result image obtained from four times of repeats.

P-PP2.x is shown in a representative image going through a procedure that does not involve membrane fusion (Figure IV-31). The gradual reddening of the vesicles in the middle panel demonstrates a lipid exchange between GUVs containing Atto647 and those lacking the dye (position p1, Figure IV-31). Even though P-PP2.x can result in extensive membrane cross-

linking, the fusion process is less obvious when comparing to AP2.1 and AP3 under the same conditions.

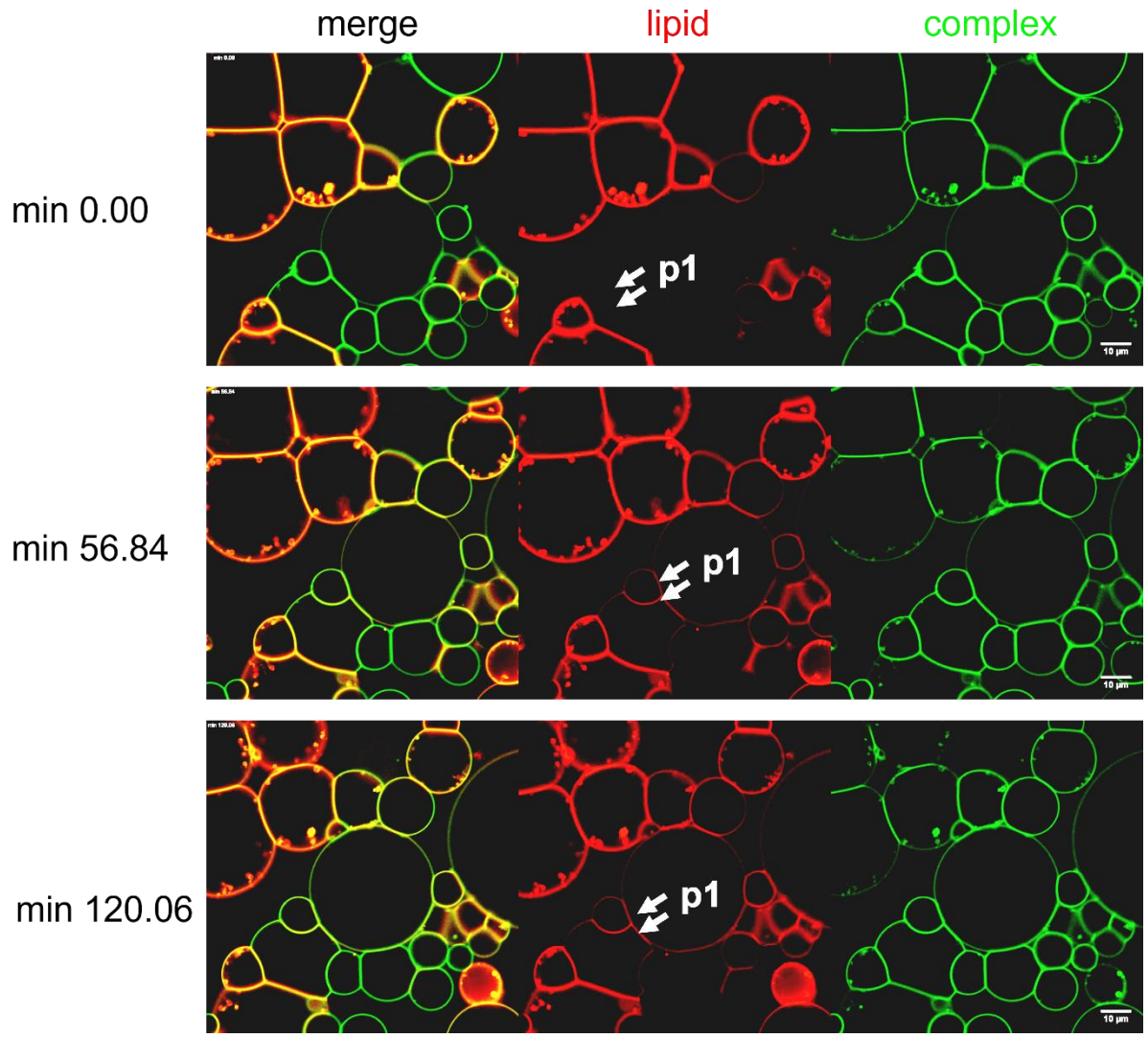


Figure IV-31 Time-lapse of GUVs after the addition of AF488-P-PP2.x (final CTB_5 concentration is 400 nM). Two types of GUV were mixed – one with composition of 65 mol% DOPC, 30 mol% Cholesterol, 5 mol% GM1; the other contained 64.7 mol% DOPC, 0.3 mol% Atto647, while the rest of the components remained the same. Presented here are one of the result image obtained from four times of repeats.

What accounts for the difference? A subjective difference is the first thing that cannot be ruled out because the GUV experiments involve observations made in specific areas at a specific time. As a result, its objectivity must be confirmed over a significant number of repetitions to ensure the experiment's reproducibility and it is challenging to analyze the data quantitatively. Erik Kempmann has carried out multiple replicate experiments for both AP2.1 and AP3 proteins. P-PP2.x experiments were not repeated as often, as we did not notice membrane crosslinking in the initial experiments. Another issue with GUV experiments is

that the vesicle quality can have a significant impact on the findings of the experiments, making the reproducibility of the results ambiguous. Because of this, GUV observations are best used as an add-on to lipid mixing experiments to clarify whether the lipid exchange phenomenon that results from lipid mixing is more likely to occur through membrane fusion or hemifusion. Although it would be best to combine the lipid mixing experiment with content mixing experiments to reach this conclusion, I was unable to do so due to limited time.

In order to determine how well the results of the lipid mixing experiment can be applied to the GUV, we can first compare the lipid and protein concentrations used in the lipid mixing experiment with those used in the GUV experiment. In lipid mixing experiment, the molarity of GM1 and CTB binding sites were 6.65 nM and 0-20 nM, respectively, resulting in a CTB monomer : GM1 ratio of up to 3. Whereas for the GUV assay, the concentration of lipid was roughly 2.5 nM, and CTB monomer was 2 µM, resulting in a CTB monomer : GM1 ratio of 800. In conclusion, in GUV experiments the ratio of protein vs membrane molecules was much higher than in the lipid mixing experiments. Extrapolating from the lipid mixing data, one would thus expect AP3 and P-PP2.x to show low fusion efficiency in GUV experiments, but high efficiency for AP2.1, as that is the only protein which does not suffer from an efficiency drop off at high protein : GM1 ratios.

IV.13 Jurkat cell-GUV fusion assay

Finally, we investigated whether this membrane-to-membrane fusion could be applied to cellular experiments. To do this, we used Jurkat cells as a cellular model (as already described in section I.8.8) and mixed them with GM1-containing GUVs to amplify any interactions between cellular membranes and synthetic membranes. Francesca Rosato and Erik Kempmann (University of Freiburg) collaborated on this experiment, with Francesca taking care of the cell cultivation and incubation with the protein and Erik performing the GUV preparation. Finally, they worked together to track the cell-protein-GUV response using confocal microscopy. There is no obvious crosslinking between cells and GUVs when there is no protein in the chamber. More specifically, GUVs and cells are remaining apart from one another (Figure IV-32).

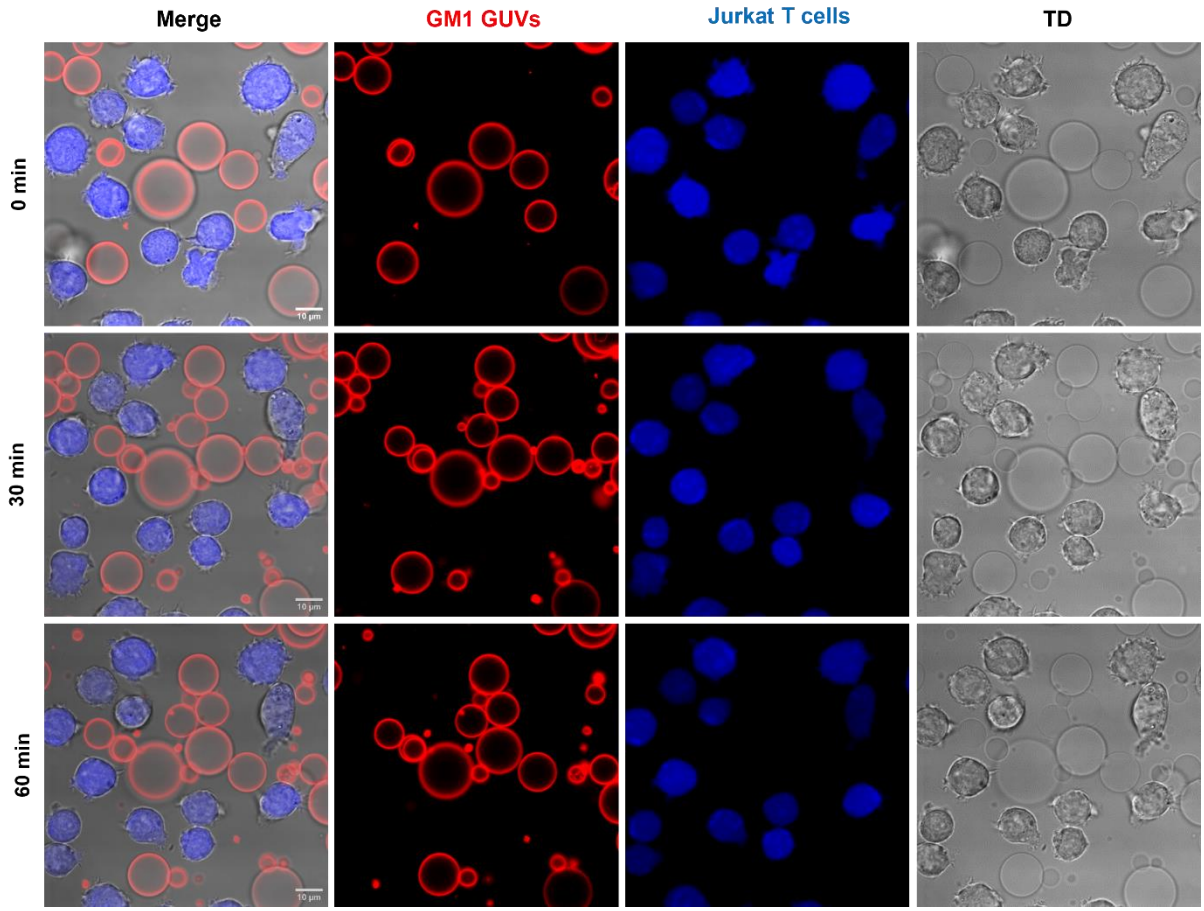


Figure IV-32 Confocal microscope images of 5 mol% GM1-decorated GUVs and Jurkat T cells incubated in the absence of CTB_5. TD stands for transmission detector. The GUVs (red color; labeled with the fluorescent lipid DOPE-Atto 647N) were incubated with the Jurkat cells (blue color; stained with CellTrace™ Violet). Live-cell imaging experiments were performed at 37 °C by using confocal laser scanning microscopy and were recorded for 60 minutes. Scale bars represent 10 µm. Presented here are one of the result image obtained from single experiment.

As shown in Figure IV-33, a distinct crosslinking of the cells to the GUVs was seen after the addition of AP2.1 to the chamber (the final CTB_5 concentration was 400 nM), and this crosslinking was stronger after one hour of incubation, to the point where distorted GUVs could be seen clearly in the field of view. Notably, despite the fact that there is a strong crosslinking between the membrane and the cell, no transfer of membrane molecules from the liposome to the cell was seen, as shown by the absence of any red Atto647 molecules in what was once a cell location after one hour.

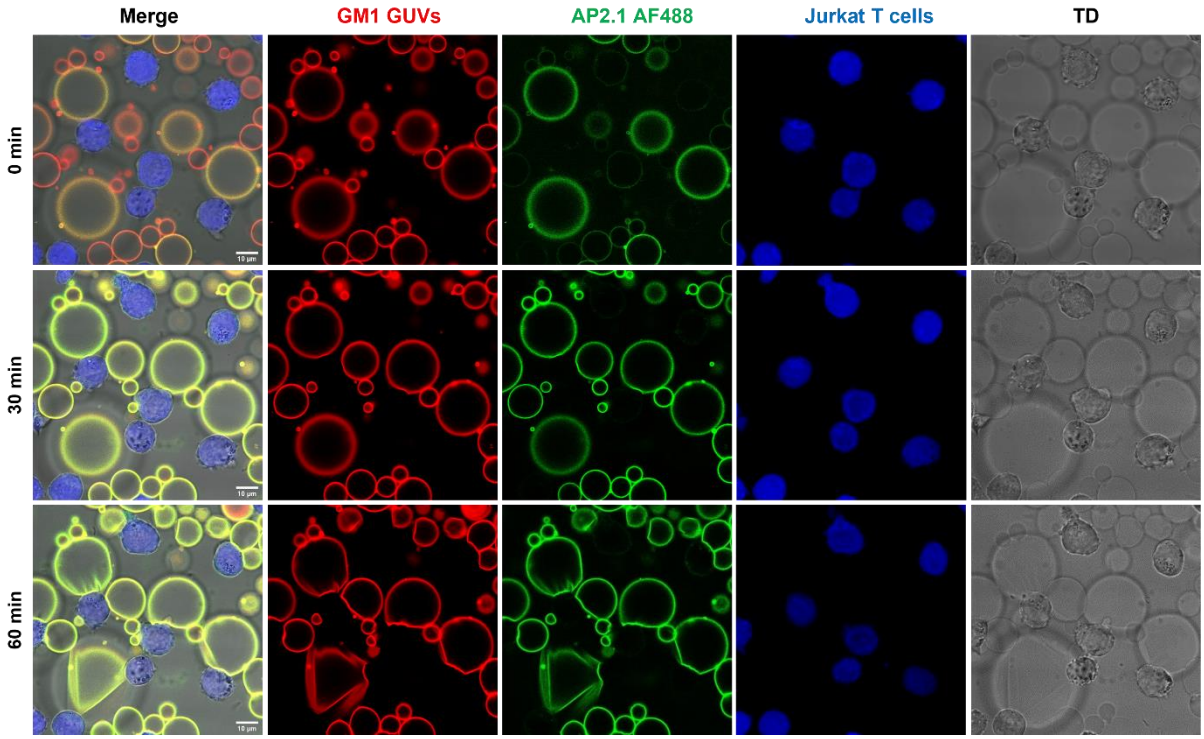


Figure IV-33 Crosslinking of GM1-decorated GUVs and Jurkat T cells is mediated by AF488-AP2.1 (400 nM CTB_5 concentration). TD stands for transmission detector. The GUVs (red color; labeled with the fluorescent lipid DOPE-Atto 647N) were incubated with the Jurkat cells (blue color; stained with CellTrace™ Violet) and AP2.1 (green color; labeled with AF488). Live-cell imaging experiments were performed at 37 °C by using confocal laser scanning microscopy and were recorded for 60 minutes. Imaging of samples started upon addition of protein. Scale bars represent 10 µm. Presented here are one of the result image obtained from four times of repeats.

However, for AP3 (Figure IV-34), we did not observe crosslinking between GUVs, possibly as a result of the relative dispersion of GUVs in the field of view. No interactions between cells and vesicles were also observed, despite there being an adequate number of cells in the field of view. The same thing happened in $CTA2B_5$, *i.e.*, we did not see any interactions between membranes (Figure IV-35).

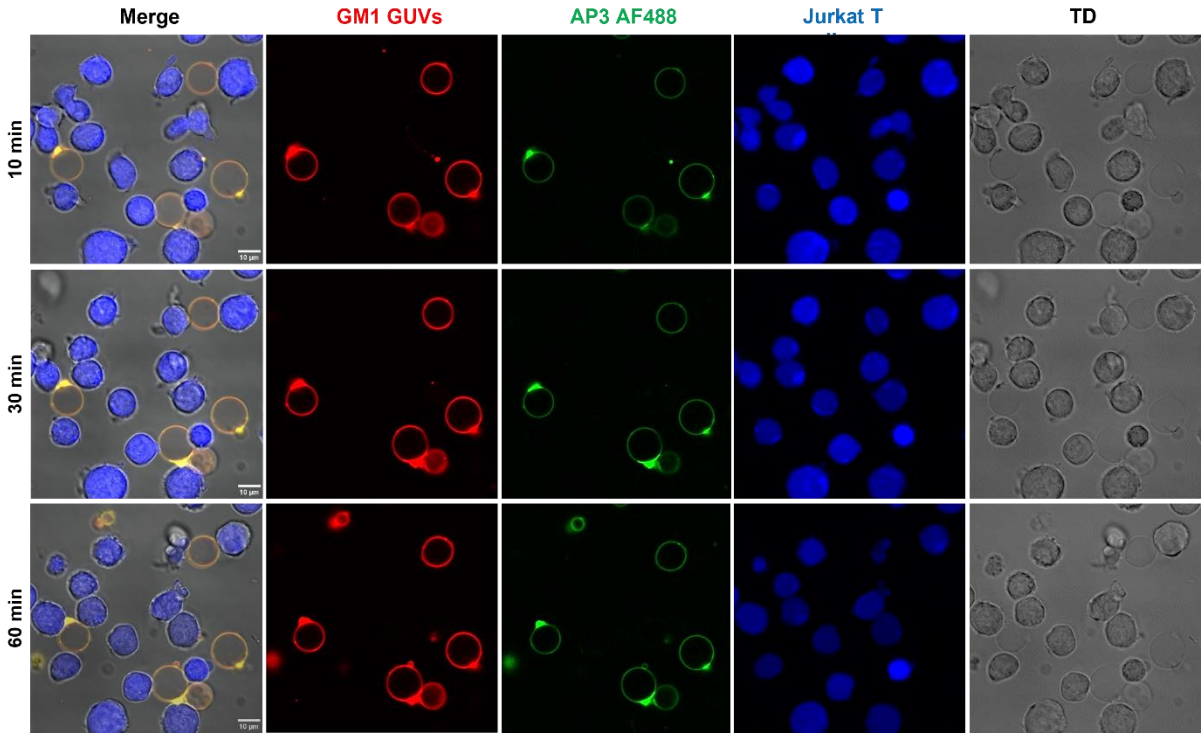


Figure IV-34 Crosslinking of GM1-decorated GUVs and Jurkat T cells is mediated by AF488-AP3 (400nM CTB_5 concentration). The GUVs (red color; labeled with the fluorescent lipid DOPE-Atto 647N) were incubated with the Jurkat cells (blue color; stained with CellTrace™ Violet) and AP3 CtxB (green color; labeled with AF488). Live-cell imaging experiments were performed at 37 °C by using confocal laser scanning microscopy and were recorded for 60 minutes. Imaging of samples started upon addition of protein. Scale bars represent 10 µm. Presented here are one of the result image obtained from four times of repeats.

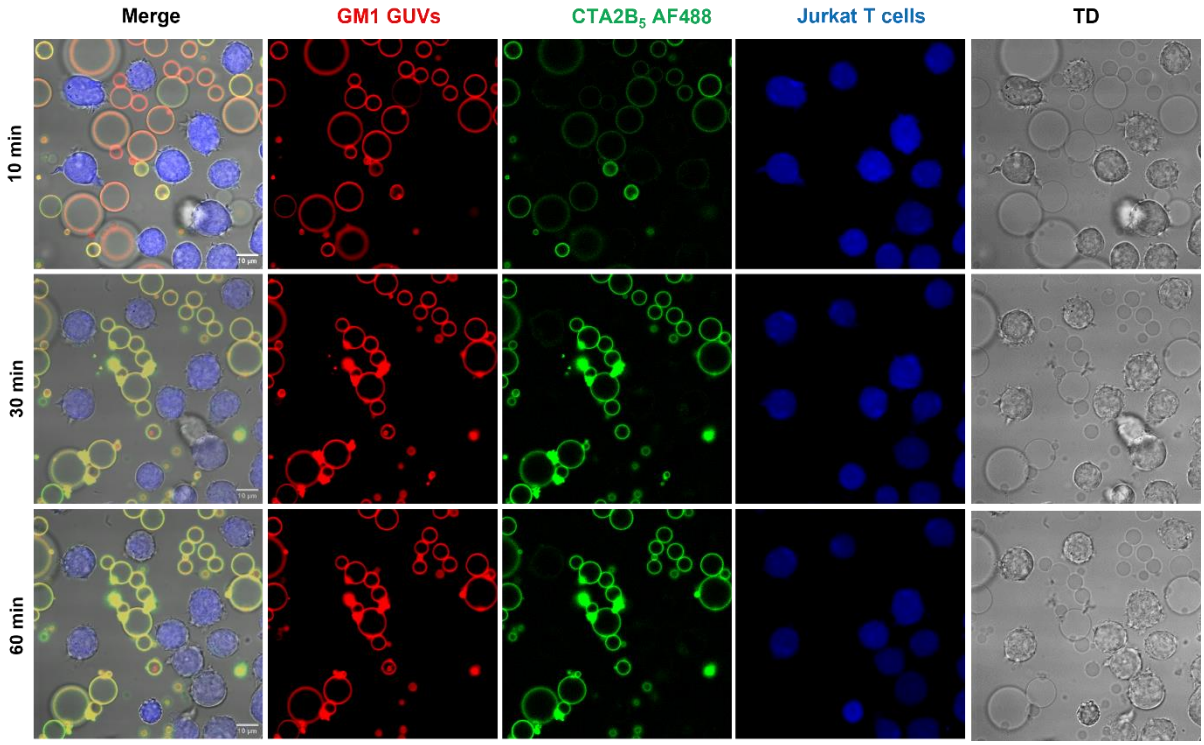


Figure IV-35 Imaging of GM1-decorated GUVs and Jurkat T cells incubated with 200 nM $CTA2B_5$. The GUVs (red color; labeled with the fluorescent lipid DOPE-Atto 647N) were incubated with the Jurkat cells (blue color; stained with CellTrace™ Violet) and A2B5 CtxB(green color; labeled with AF488). Live-cell imaging experiments were performed at 37 °C by using confocal laser scanning microscopy and were recorded for 60 minutes. Imaging of samples started upon addition of protein. Scale bars represent 10 µm. Presented here are one of the result image obtained from four times of repeats.

IV.14 Discussion

The lengths of the coiled coil and CTA2 were varied in this study to produce CTB dimers with varying degrees of flexibility and binding potency. The flexibility of the CTB dimers was then assessed quantitatively using model building and MD simulations. Three proteins, AP2.1, AP3 and P-PP2.x, were subjected to lipid mixing experiments to ascertain whether they are capable of lipid exchange and the relationship between lipid exchange rate and protein concentration. Based on the findings of the experiments, it was hypothesized that the coiled coil of AP2.1 may have unraveled during its interaction with the membrane.

We used the QCM-D as a secondary tool to demonstrate changes in the height and softness of proteins binding to the SLB surface in order to better understand the mechanism of protein-membrane interactions. Through parallel comparison, we came to the conclusion that AP2.1 did experience coiled coil dissociation on the membrane surface. Following this, GUVs were prepared and used to visualize the behavior of membranes with the various proteins. AP2.1 caused the most frequent membrane fusion, followed by AP3 and P-PP2.x. The dissociation of AP2.1 may be the reason why its corresponding lipid mixing rate is not inhibited at high protein concentration, leading to a more pronounced membrane fusion phenomenon in the GUV experiments. Finally, the interaction between Jurkat cells and GUV induced by proteins was also studied, especially comparing AP2.1 with AP3 which could previously cause more intense GUV-GUV interactions, with AP2.1 finally winning out through strong GUV distortion.

IV.15 Future experiments

A promising idea is to measure the influence of flexibility and coiled coil binding strength on the rate of membrane fusion. According to the article written by Wolfson's lab [116] , we could specifically investigate this issue using coiled coils with various binding strengths from nM to mM range and by including various types of flexible linker between short CTA2 and coiled coil. One implication is that the lipid mixing outcome of CTB_5 dimer with a weaker coiled coil and a more rigid linker is more likely to resemble AP2.1. The combination of a

more durable coiled coil and a flexible linker is more likely to produce a phenomenon similar to AP3 or P-PP2.x.

Chapter V Assembly CTB_5 complex through Affimer

The use of coiled coils to construct heterogeneous lectin assemblies (Chapter III and Chapter IV) proved to be more challenging than anticipated for several reasons. Firstly, generating heterogeneous coiled-coil assemblies is limited by the protein expression yield of each *procoil* species – when one protein is not successfully expressed, its partner cannot be utilised for building the heterodimer. Secondly, separating the heterodimeric coiled coil-$CTA2B_5$ from their homomeric counterparts was difficult. Streptavidin was another tool we used to construct lectin assembly (Chapter II). Although it could efficiently assemble multiple copies of the same $CTA2B_5$, it was not suitable for building heterogenous assemblies, due to the lack of fine control over how the proteins combine to form a complex. To establish a system for building heterodimeric lectins in future, using a chemical linker to connect multiple Affimers was considered.

Affimers are single-domain non-antibody proteins (around 15 kDa) initially engineered in University of Leeds[142]. It originated from consensus sequence of Cystatin A, a plant-derived antagonist of cysteine proteases. Affimer contains four β-strands and one α-helix with two flexible loops extending from the two adjacent antiparallel β-strands as their hypervariable binding domain (Figure V-1). The backbone of Affimer belongs to the class of non-immunoglubulin scaffold and has strong structural stability with a melting temperature of 101 degrees[142]. This allows it to tolerate a broad range of sequences in the hypervariable loops and withstand chemical modification to anchor tags to achieve of protein-protein interaction[94].

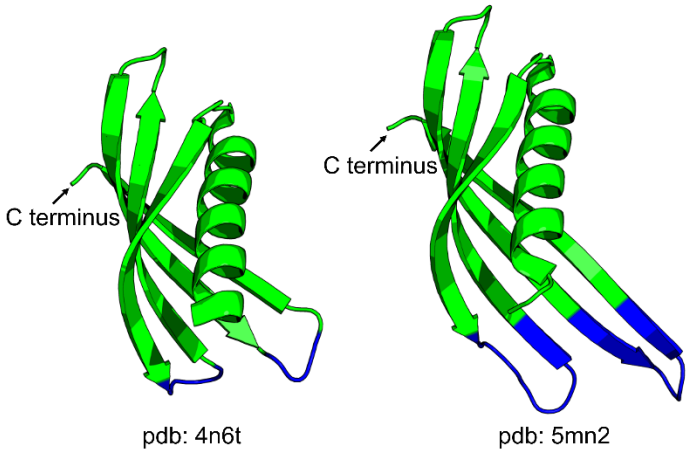


Figure V-1 The crystal structure of two chosen Affimers. The first Affimer prototype, also known as the Adhiron92 scaffold (pdb: 4n6t), is shown on the left. On the right is an Affimer targeting Fc gamma receptor IIIa screened out by phage display (pdb code is 5mn2). Both of the Affimers' two hypervariable loops are highlighted in blue.

Specific Affimers against a selected target can be raised *in vitro* by phage display, followed by subcloning and large-scale expression with purification using a C-terminal His-tag. All these features make Affimer an easy protein to work with. Therefore, if an Affimer that binds tightly to the desired lectin can be produced, this binder shall have a great potential since it allows for constructing super-lectins protein assembly more freely. With this basic idea in mind, I used Affimer as the binder protein for the non-carbohydrate binding face of CTB_5 and attempted to covalently join two Affimers (Figure V-1). The presence of a single cysteine in the Affimer at the C terminus makes it easy to chemically modify in a controlled manner, at a position far away from the binding loop. Using Affimer as an approach can produce another benefit: CTB_5 can be added by itself to a sample to bind GM1 and the Affimer can be applied later, once inducing membrane fusion is desired.

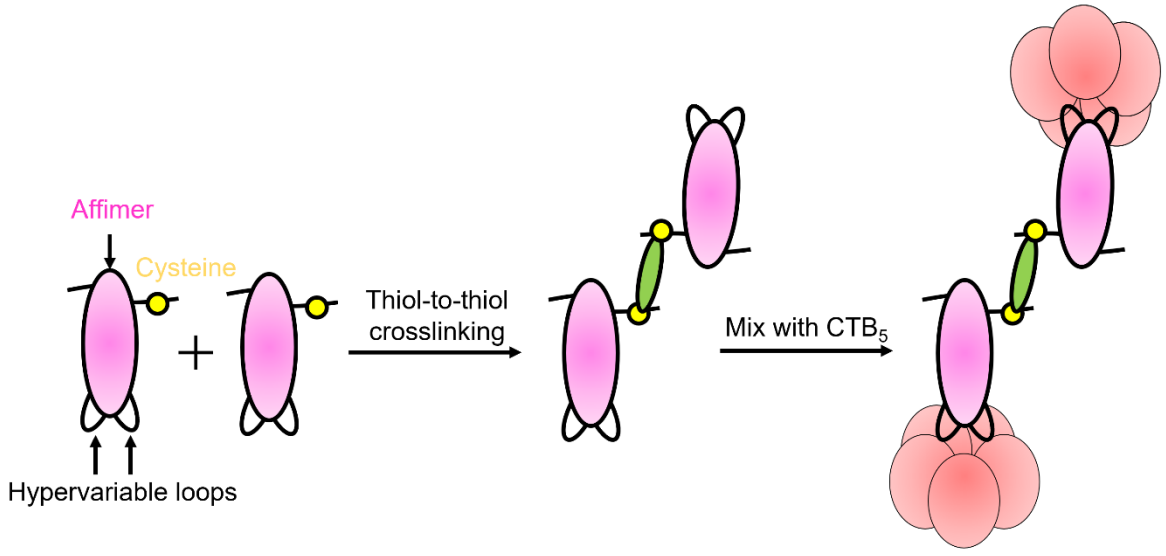


Figure V-2 A schematic illustration that exhibits the fundamental concept of constructing an Affimer dimer via thiol-to-thiol chemical ligation, which possesses the capacity to non-covalently assemble two CTB_5 into a single protein complex

V.1 Introduction to 3C6 Affimer

The work of constructing the Affimer dimer-CTB_5 assembly (Figure V-3) is based on the results obtained by two former PhD students from Turnbull and Webb lab. Matthew Balmforth has previously identified two Affimers (referred to as 3A2 and 3C6), which can specifically target CTB_5 [95]. Tomasz Kaminski subsequently developed a strategy to link two different Affimers through cysteine labelling using the chemical APN-Maleimide (APN-M)[143]. There are several points to consider regarding the specific Affimer to use. Firstly, the selected Affimer should be able to bind to CTB_5 at the concentrations used in both GUV experiments (400 nM CTB_5, or 200 nM dimer) and lipid mixing assays (from 100 nM to 4 uM CTB_5 concentration). Secondly, the linker between two Affimers should not be too flexible, which might hinder the process of membrane fusion, as in the case of *trans*- or *cis*-streptavidin-$(CTA2B_5)_2$ complex (Chapter II).

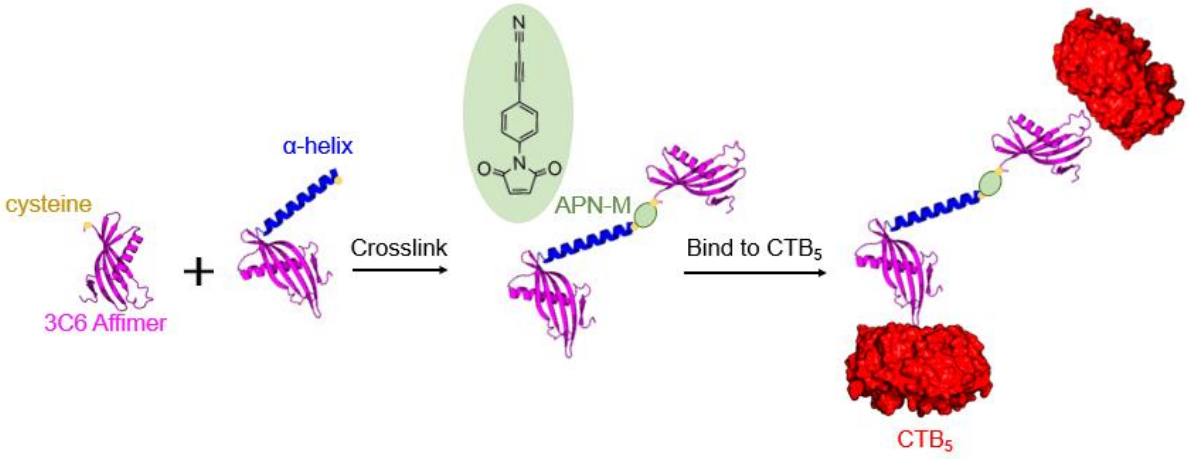


Figure V-3 A schematic illustration depicting the utilization of an APN-M linker for constructing a 3C6 Affimer dimer, which is subsequently employed for assembling with CTB_5.

The ITC results generated by Matthew Balmforth (Tabel V-1) shed light on which Affimer would be the best candidate to use. Tomasz Kaminski previously employed 3A2 Affimer to bind CTB_5 since he wanted the Affimer to dissociate from CTB_5 after the protein complex entered the Golgi apparatus. I used the 3C6 Affimer instead, because it exhibits a higher binding affinity for CTB_5, ensuring that the protein complex remains intact even at low concentrations in GUV experiments. The difference between 3A2 and 3C6 Affimers is the composition of their hypervariable loops – while 3A2 Affimer contains mainly hydrophobic residues in the binding loops, the residues in 3C6 Affimer are more hydrophilic (Figure V-4).

Table V-1 Thermodynamic parameters for CTB_5, titrated into Affimer 3A2 or 3C6 in ITC analysis (adapted from Matthew Balmforth's thesis).

Affimer	ΔH (kcal/mol)	ΔG (kcal/mol)	Kd (nM)	TΔS (kcal/mol)	n
3A2	-20.45 ± 0.28	-11.86 ± 0.52	820 ± 70	-11.86 ± 0.80	1.20 ± 0.01
3C6	-5.76 ± 0.06	-10.06 ± 0.09	72 ± 11	4.31 ± 0.15	0.97 ± 0.01

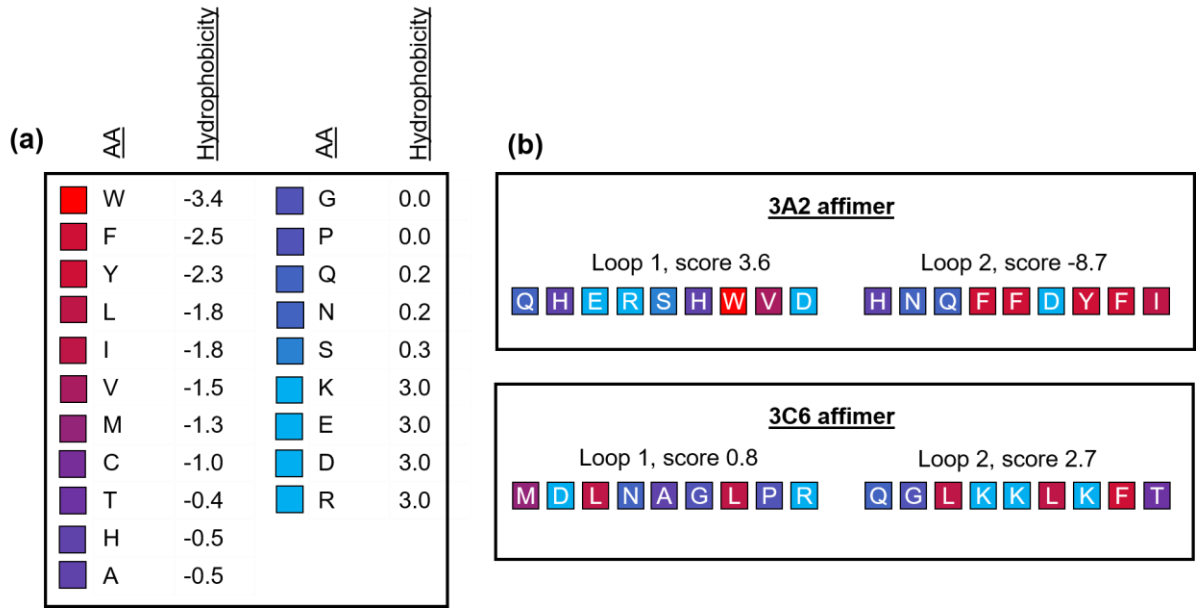


Figure V-4 (a) Hopp and Wood scale describes the hydrophilicity of amino acids. (b) Alignment of the hypervariable loops from 3A2 and 3C6 Affimer, with different residues colored according to Hopp and Wood scale. The total hydrophilicity/hydrophobicity score for the individual loop was calculated as a reference.

V.2 Gene works to insert the 24AA helix at the C terminus

In previous work, Tomasz Kaminski sought to link the 3A2 Affimer to an Affimer against the human enzyme beta-1,4-galactosyltransferase 1 (B4GALT1) (Figure V-5-a), but this resulted in precipitation of the Affimer dimers. To solve the solubility issue, a 24 amino-acid helical linker was grafted onto the C-terminus of the 3A2 Affimer (Figure V-5-b). This α-helix was based on the repeating protein sequence of the central region in smooth muscle caldesmon – one of the highly charged single α-helix domains found in nature [144]. It contained mostly glutamic acid and lysine residues (KKAEEEKKAAEEKKAEEEKKAAEE), which strongly boosted the overall solubility of the Affimer. Tomasz studied the secondary structure of the 24AA helix through Circular dichroism (CD), which demonstrated that it indeed was α-helical at low temperature (5.2 ºC) and gradually unfolded as the temperature increased to 42.8 ºC (Figure V-6).

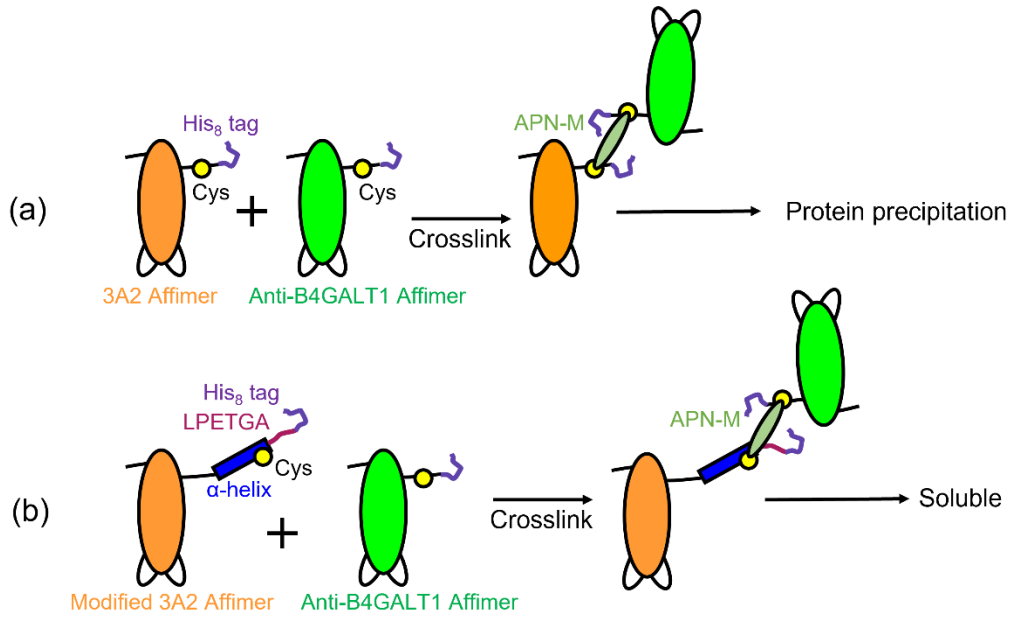


Figure V-5 Scheme illustration which compares different Affimer dimer construction strategies. (a) Directly ligating two cysteines with the bifunctional APN-M linker results in linkage of 3A2 Affimer and anti-B4GALT1 Affimer. It eventually resulted in protein precipitation because the assembled Affimer dimer possesses excessive hydrophobic residues. (b) An α-helix and Sortase labelling tag is fused at the C-terminus of 3A2 Affimer to dimerize with anti-B4GALT1 Affimer, which resulted in a soluble protein during storage.

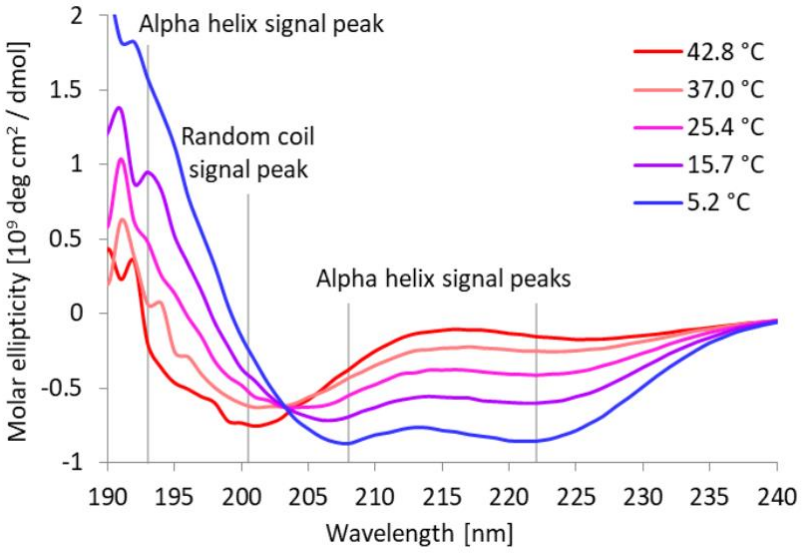


Figure V-6 Circular dichroism of 3A2-24AA-LPETGA-His_8 Affimer (image copied from Tomasz Kaminski's thesis, 2020). Depicted is the net circular dichroism difference between 3A2-24AA-LPETGA-His_8 Affimer and 3A2-Cys-His_8 Affimer with an iodoacetamide capped at cysteine residue. The insertion of 24AA-LPETGA produced a higher overall helical content at 5 ºC, but the helix unfolded as the temperature rose.

Affimer 3C6 is natively less aggregation-prone than 3A2 thanks to its hydrophilic binding loops (Figure V-4), so transplanting the 24AA helix into it was not necessary from a solubility

standpoint. However, the helix was added to it regardless, as we suspected that increasing the distance between the two Affimers in a dimer might facilitate the membrane fusion process by increasing the chance of docking to the neighboring vesicles with a longer distance between two CTB_5 and also decreasing steric clashes.

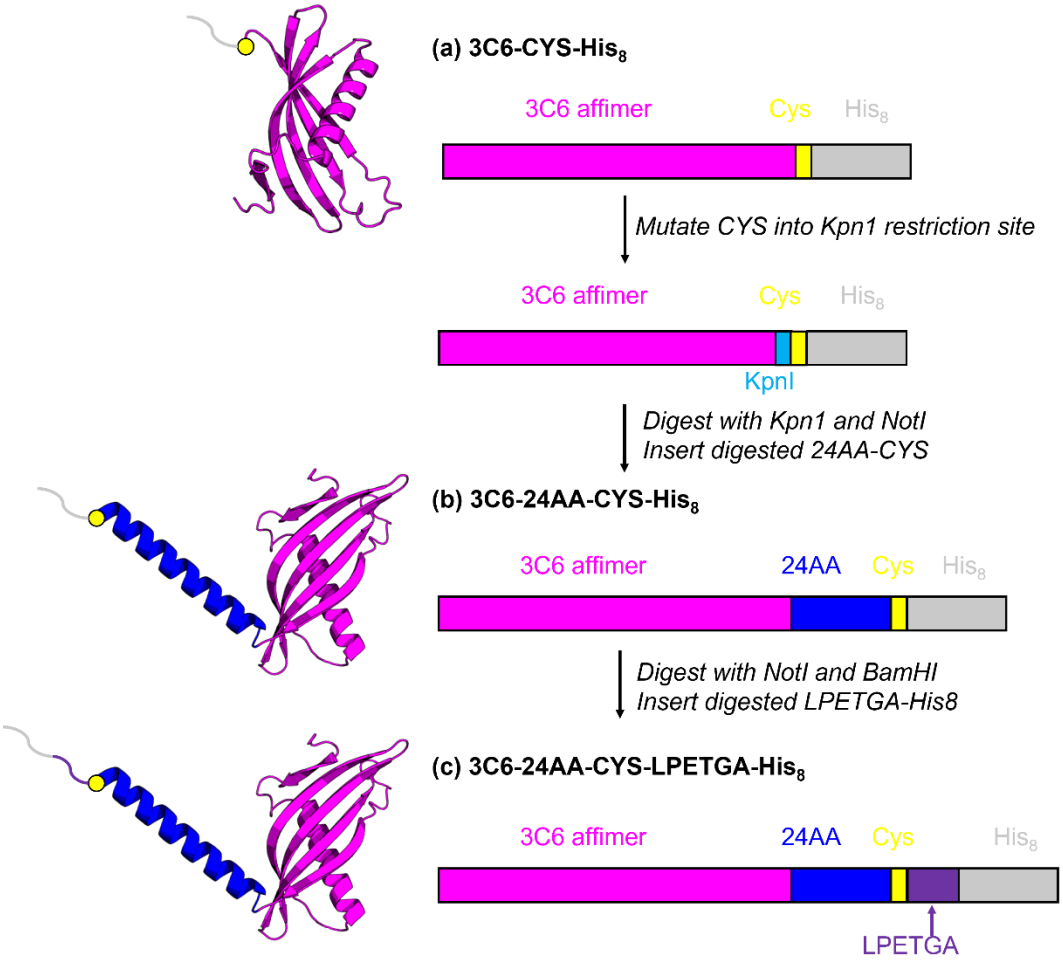


Figure V-7 The scheme of three types of 3C6-based Affimer used in the project. At the C-terminus of the 3C6-CYS-His_8 Affimer, the gene piece is altered to incorporate a Kpn1 digestion site. After that, a gene fragment encoding the 24AA-helix was added, resulting in the creation of the plasmid for expressing the 3C6-24AA-CYS-His_8 protein. The last step is to switch the C-terminus His_8 tag for the LPETGA-His_8 tag to express the finished product, 3C6-24AA-CYS-LPETGA-His_8. (Detailed genetic work is in Appendix VIII.1.3.d

To incorporate the 24AA linker at the C terminus of 3C6 Affimer, first the C terminus of the Affimer was mutated to contain a KpnI digestion site using Quickchange PCR, with the primers Affimer_3C6_FRW_insert_KpnI and Affimer_3C6_REV_insert_KpnI. The plasmid was then digested with KpnI-HF plus NotI-HF and a pair of hybridized (double-stranded) 5'-phosphorylated oligonucleotides encoding the 24AA helix was ligated in. To insert an

LPETGA Sortase labelling tag at the end of the protein, the plasmid was further digested with NotI-HF and BamHI-HF, and another pair of hybridized oligonucleotides was ligated in.

Table V-2 The sequence of primer or inserted DNA pieces for constructing 3C6-24AA-CYS-LPETGA-His_8 protein.

Name	Sequence
Affimer_3C6_FRW_insert_KpnI	GAGTTCAAACCGGTTGGTACCGCTTGTGCGGCCGCGCAT
Affimer_3C6_REV_insert_KpnI	ATGCGCGGCCGCACAAGCGGTACCAACCGGTTTGAACTC
24AA-CYS forward oligonucleotides	CAAGAAAGCTGAAGAAGAGAAGAAGGCTGCCGAAGAAAAGAAAGCCGAAG AGGAGAAGAAAGCCGCTGAAGAGGCTTGTGC
LPGA-His8 forward oligonucleotides	GGCCGCGCTGCCGGAAACCGGCGCGCATCACCATCATCACCACCATCATT GATA

V.3 Express and purify 3C6-24AA-His_8

Tomasz originally added a 24AA helix to the 3A2 Affimer along with the Sortase labeling tag LPETGA at the C terminus (Figure V-5-b). Since Sortase labeling is not urgently needed in my project, I chose to start by inserting only the 24AA at the C terminus of the 3C6-CYS-His_8 Affimer and leave out the Sortase labeling tag LPETGA. This protein was expressed in parallel with 3A2-24AA-LPETGA-His_8 (Methods in section VIII.1.3) for the sake of comparison. Although the latter protein could be obtained in good yields from *E. coli* BL21 (DE3), the 3C6-24AA-CYS-His_8 did not express well – a band corresponding to the overexpressed protein could not be observed in the whole-cell lysate (Figure V-8).

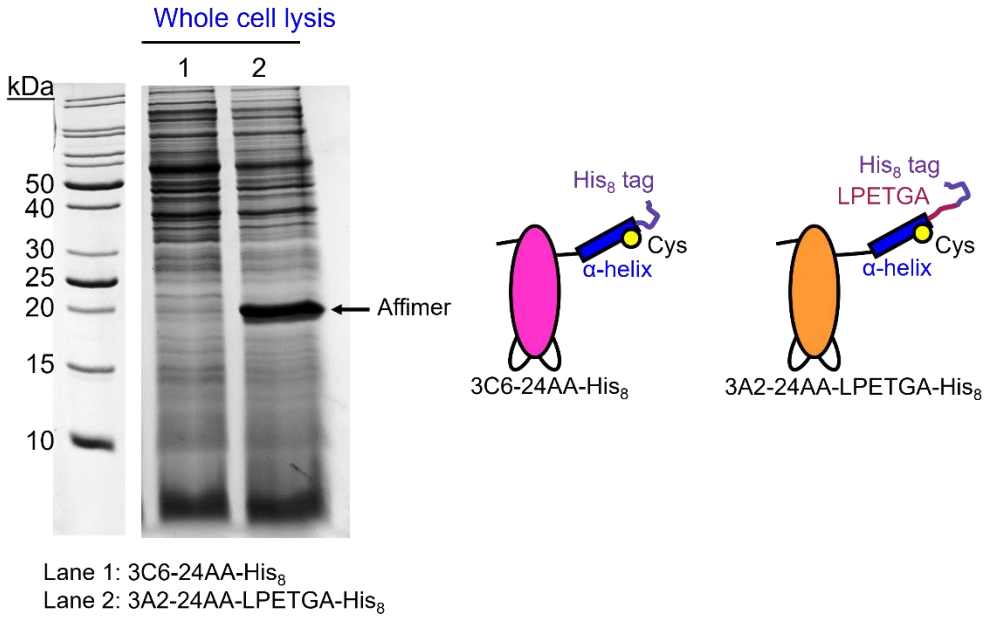


Figure V-8 16% Coomassie stained SDS-PAGE of the whole cell lysis for 3C6-24AA-His_8 and 3A2-24AA-LPETGA-His_8. The targeted Affimer is around 16 kDa. When the successful expression can be found for 3A2-24AA-LPETGA-His_8, the overexpression band was not pronounced for 3C6-24AA-His_8 expression. On the right is

the composition of the two expressed Affimers – the only difference on the protein scaffold is the Sortase labelling tag LPETGA in 3A2-24AA-LPETGA-His_8.

In case the poor expression was caused by lack of the Sortase tag, we performed the genetic work to add it in and repeated the expression tests. Although initial overexpression in *E. coli* C41(DE3) again suggested relatively low yields (Figure V-9-a), scaled-up expression (1.6 L media) and automated purification produced enough pure protein (16 mg/L, Figure V-9-b and Figure V-9-c) for our purposes.

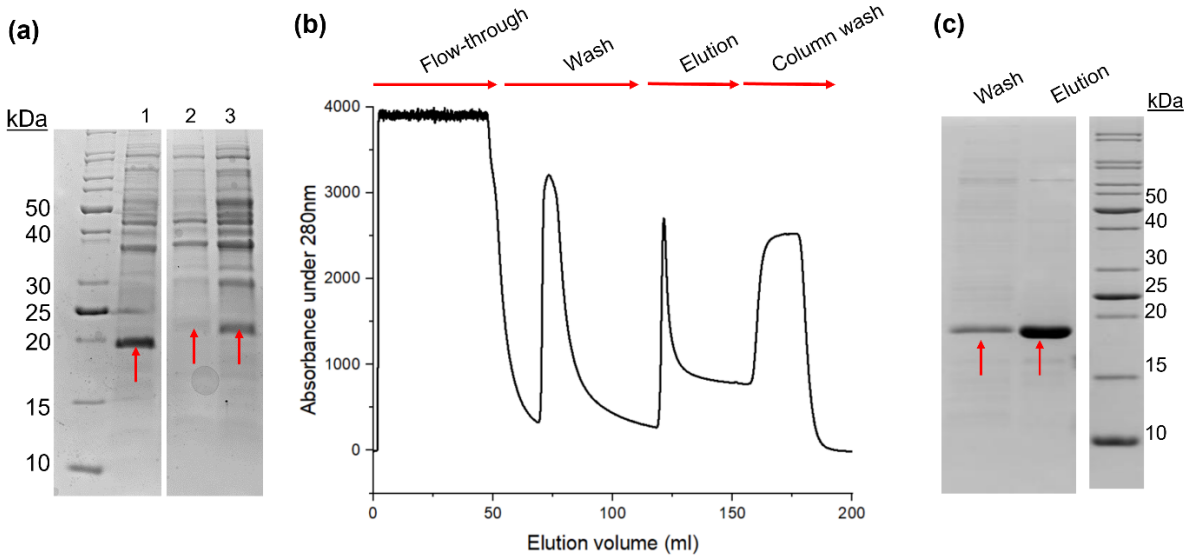


Figure V-9 (a) 16% Coomassie stained SDS-PAGE of the soluble extraction using chemical reagent B-PER buffer from -1: 3A2-24AA-CYS-LPETGA-His_8 (BL21(DE3)) -2 and -3: 3C6-24AA-CYS-LPETGA-His_8 expressed from BL21(DE3) and C41(DE3) respectively. The Affimer protein of interest was indicated with a red arrow. The C41(DE3) cell could produce a larger amount of overall protein than the BL21(DE3) cell for 3C6 Affimer derivatives. (b) Nickel purification of 3C6-24AA-CYS-LPETGA-His_8 protein expressed from C41(DE3) cell. (c) 16% Coomassie stained SDS-PAGE of nickel wash and nickel elution from (b). The Affimer protein of interest was indicated with a red arrow.

After the Affimer was purified via nickel affinity chromatography, its molecular weight was confirmed to be correct through accurate mass spectrometry (Figure V-10). As expected, some of the Affimers had lost their N-terminal alanine residue – this is a common occurrence for Affimer proteins, which Tomasz showed to be the result of cleavage by *E. coli* enzymes (Tomasz Kaminski's thesis, 2020). A pair of minor peaks with a mass shift of +162 Da was also seen, corresponding to cysteine residue glycosylation by an as-yet-unidentified mechanism (Tomasz Kaminski's thesis, 2020).

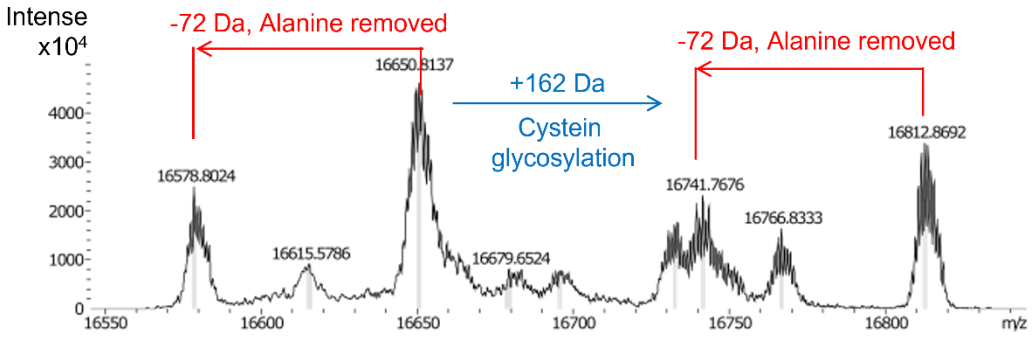


Figure V-10 (a) The mass spectrum of the 3C6-24AA-CYS-LPETGA-His_8 Affimer, following purification via a nickel column, reveals the anticipated molecular weight of 16651 Da, albeit with impurities generated due to N-terminus alanine residue cleavage and glycosylation.

Tomasz discovered significant DNA contamination in the purified 3C6-24AA-CYS-LPETGA-His_8 protein when using the Nanodrop for determining the protein concentration (data not shown). The Affimer-containing nickel chromatography elution was therefore incubated with DNase (Roche) for an hour, following which size exclusion chromatography was utilized to separate the DNA fragments from the Affimers. Unexpectedly, the Affimers in SEC elution fractions were consistently smaller than the protein loaded onto the column, indicating proteolytic degradation. We speculate that the commercial DNase enzyme was contaminated (as sold to us) with proteases, causing the cleavage (Figure V-11).

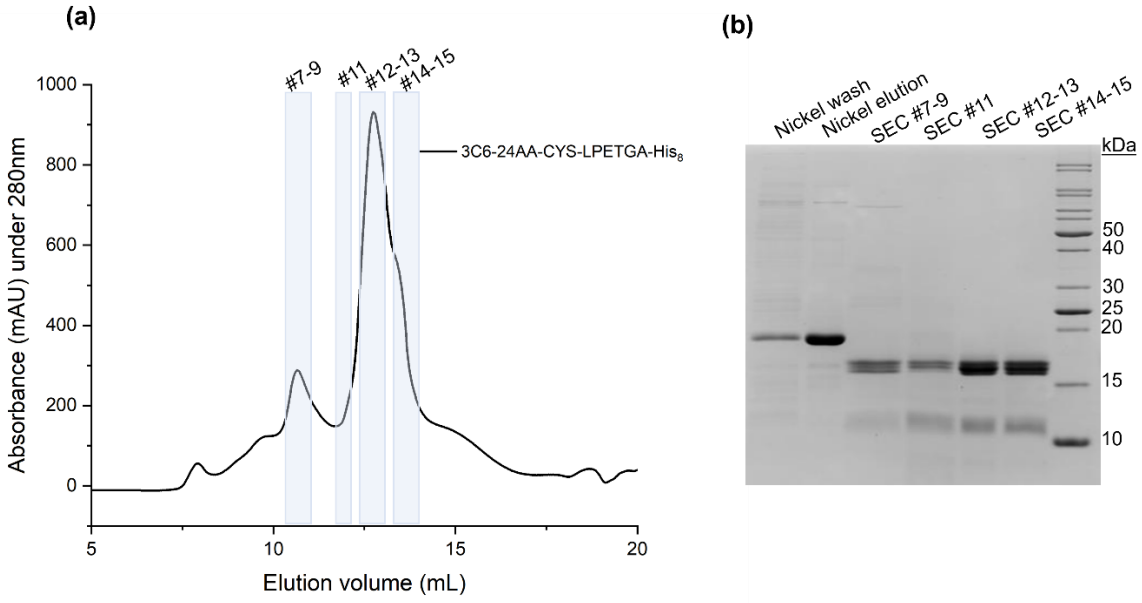


Figure V-11 (a) Size exclusion chromatography (SD75 10/300 column) of DNAase-treated 3C6-24AA-CYS-LPETGA-His_8. Fractions 7-9, 11, 12-13 and 14-15 were collected separately. (b) 16% Coomassie stained SDS-PAGE of the SEC elution fractions from (a). Fraction 7-9 were eluted at the void volume and implied the

existence of a large protein-DNA complex Protein gathered from the central peak (fraction 12-13) had a smaller molecular weight than the original nickel elution, according to Coomassie stained SDS-PAGE. Moreover, multiple bands were seed in the same fraction even though the nickel elution had only one clean band – all signs of nonspecific protein cleavage.

Time pressure prevented us from immediately expressing and purifying a fresh batch of Affimer protein, so we proceeded to label the degraded Affimer with APN-Maleimide and employ it for dimerization. Accurate mass spectroscopy analysis of 3C6-24AA-CYS-LPETGA-His_8 before and after reaction with APN-M thankfully revealed that most of the proteolysis at the C terminus occurred after the cysteine residue, meaning that the Affimer was still capable of forming covalent dimers. It seems likely that the inserted 24AA α-helix was responsible for protecting the adjacent cysteine from being cleaved off, as adding the α-helix decreased the degree of cysteine glycosylation in the 3A2 Affimer (Tomasz Kaminski's thesis, 2020). However, some cysteine loss did still happen, as shown in the MS spectrum (Figure V-12).

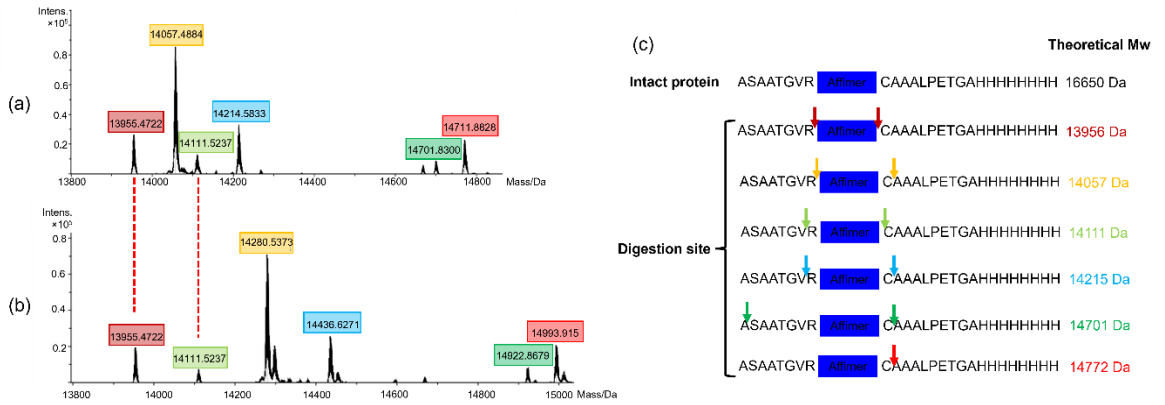


Figure V-12 (a) Mass spectrum of degraded 3C6-24AA-CYS-LPETGA-His_8 eluted from SEC fraction 12-15 in Figure V-11. (b) APN-Maleimide labelling of the SEC elution, with an increase of 222 Da indicating the successful labelling. The solid line indicated an increase of 222 Da after labelling with APN-Maleimide; the dotted line highlights the peak which was not labelled. (c) Deducing the degradation sections of 3C6-24AA-CYS-LPETGA-His_8 according to the molecular weight shifts.

To clarify, dimers created from the degraded Affimer were used only for the GUV experiment (section V.7) and the Jukat cell experiment (section V.8). All the other experiments, including mass photometry (section V.6) and lipid mixing experiments (section V.9), were performed with intact Affimers – freshly expressed and purified as described above, but in the presence of protease inhibitor AEBSF and without addition of DNAse. These two changes were sufficient to obtain a full-length Affimer, as assessed by accurate mass spectrometry (Figure V-13).

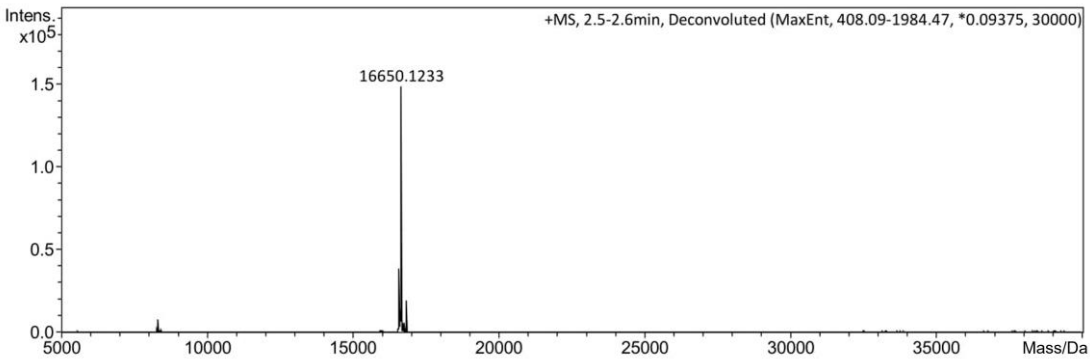


Figure V-13 The mass spectrum of the purified intact 3C6-24AA-CYS-LPETGA-His_8 protein indicates a primary peak with a molecular weight of 16650 Da, in alignment with the theoretical mass of interested Affimer.

V.4 Constructing 3C6-24AA-3C6 and 3C6-3C6 by APN-Maleimide

For simplicity of presentation in the following sections, I will refer to 3C6-24AA-CYS-LPETGA-His_8 and 3C6-CYS-His_8 as 3C6-24AA-Cys and 3C6-Cys respectively. Affimer dimers made by the interaction of 3C6-24AA-Cys and 3C6-Cys will be represented by 3C6-24AA-3C6. Those made by the linking of two 3C6-Cys by APN-M will be represented by 3C6-3C6.

After successfully purifying the intact 3C6-24AA-Cys protein, it was decorated with APN-M bifunctional linker before interacting with 3C6-Cys to form Affimer dimers (Figure V-15). The 3C6-24AA-Cys Affimer was exchanged into a pH 7.22 HEPES buffer and the protein diluted to 50 µM. It was incubated with 100 µM of the APN-Maleimide linker for an hour, followed by a buffer exchange step to remove the excess unreacted linker molecules. Due differences in kinetics, the cysteine thiol preferentially interacted with maleimide groups, leaving the alkyne group of APN-M free (Figure V-14-a). The labelled Affimer was concentrated to 500 µM and mixed with 600 µM 3C6-Cys in a 1-to-1 ratio for an overnight incubation. During this process, the free thiol group of 3C6-Cys interacted with the alkyne group at the other end of APN-Maleimide (Figure V-15-b), creating covalent Affimer dimers.

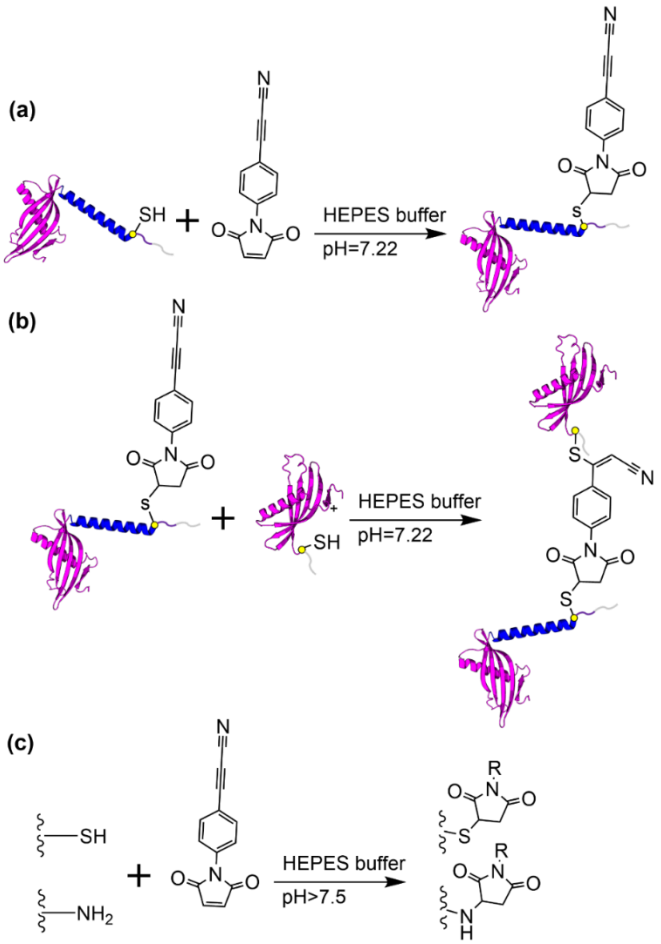


Figure V-14 (a) The functionalization of 3C6-24AA-Cys with APN-M. (b) Reaction of the second 3C6-Cys Affimer with the chemically labelled 3C6-24AA-Cys. (c) Side products caused by high pH (>7.5). Both thiol and amine groups are now capable of interacting with maleimide.

An important point that needs to be considered in this reaction is the pH, which needs to stay below 7.5. More specifically, the reaction rate between maleimide and thiol is 1000 times faster than amine & maleimide, but only under pH 7.5. If the pH is any higher, more lysine sidechains will become deprotonated and therefore rapidly react with the maleimide group in a nucleophilic addition[145] (Figure V-14-c).

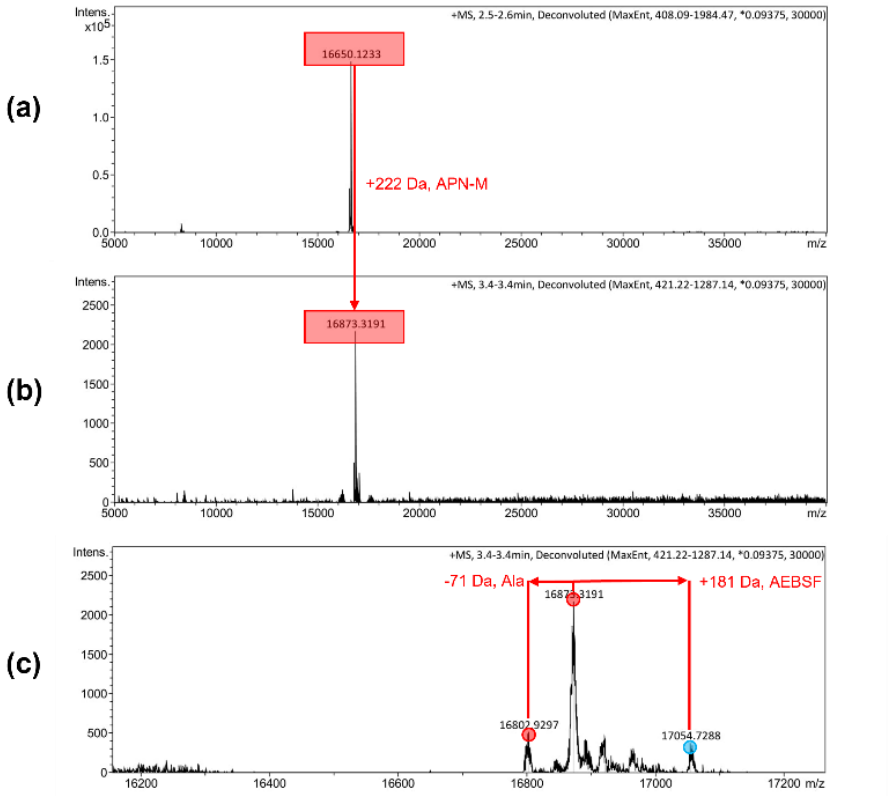


Figure V-15 (a) MS spectra showing intact 3C6-24AA-Cys from nickel elution. (b) An increase of 222 Da from the major peak indicates successful APN-Maleimide labelling without self-dimerisation, which would appear as peak at 33 kDa. (c) A narrowed mass range, zoomed in on the labelled Affimer. Apart from the main peak of the correctly labelled protein, the 17054 Da species indicates the addition of serine protease inhibitor AEBSF, which could be prevented if no more protease inhibitor was added after obtaining the nickel elution.

Direct homo-dimerisation of two 3C6 Affimers via APN-M was also investigated, to check whether the 24AA helix exerts a noticeable effect on 3C6 Affimer dimer solubility and membrane fusion efficiency. In the end, soluble dimers of both 3C6-3C6 and 3C6-24AA-3C6 were obtained (Figure V-16, Figure V-17) and studied in the lipid mixing assays.

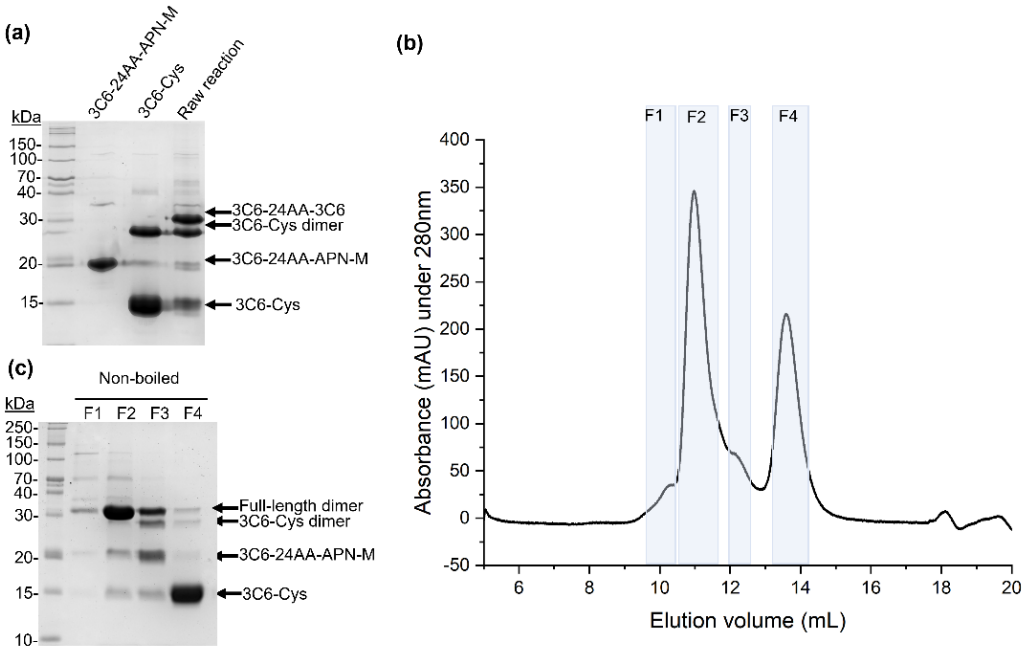


Figure V-16 Dimerizing 3C6-24AA-Cys with 3C6-Cys. SDS-PAGE was used to analyze the dimerization in raw reaction, and size exclusion chromatography (SD75, 10/300) was used to purify the raw reaction. In the end, the purified dimer was subjected to SDS-PAGE to check the purity. (a) 16% SDS-PAGE of APN-M labeled 3C6-24AA-Cys Affimer, 3C6-Cys Affimer, and raw reaction of dimerizing the mentioned two Affimers dimers. (b) Size exclusion chromatography of the raw reaction of forming 3C6-24AA-3C6. The F1, F2,, F3 and F4 fractions were gathered and subjected to SDS-PAGE, as seen in (c). (c) selected SEC fractions including F1, F2, F3 and F4 in different lanes.

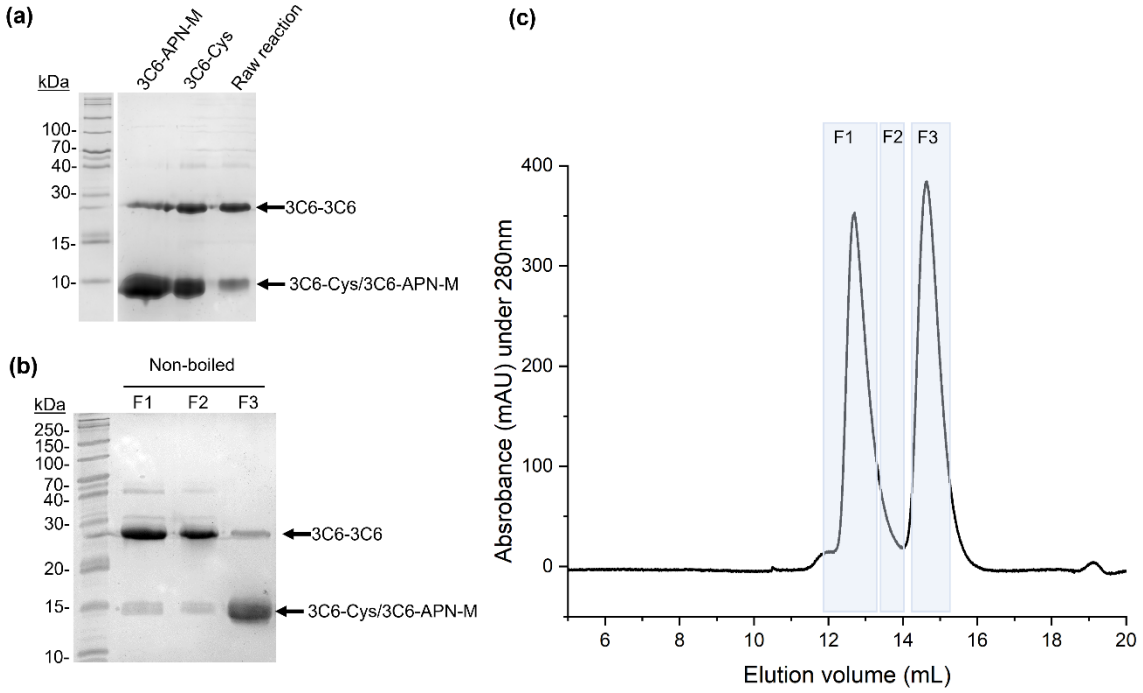


Figure V-17 Dimerizing two 3C6-Cys Affimer through APN-M. SDS-PAGE was used to analyze the dimerization in raw reaction, and size exclusion chromatography (SD75, 10/300) was used to purify the raw reaction. In the end, the purified dimer was subjected to SDS-PAGE to check the purity. (a) 16% SDS-PAGE of APN-M labeled 3C6-Cys Affimer, non-labelled 3C6-Cys Affimer, and raw reaction of dimerizing the mentioned two Affimers dimers. (b) Size exclusion chromatography of the raw reaction of forming 3C6-3C6. The F1, F2, and F3 fractions were gathered and subjected to SDS-PAGE, as seen in (c). (c) selected SEC fractions including F1, F2, F3 in different lanes.

V.5 Impact of the addition of APN-Maleimide to absorbance under 280 nm

Due to its aromatic groups, it was possible that APN-Maleimide (APN-M) would absorb UV light at 280 nm, and thus change the 280 nm absorption coefficient of whatever protein it was linked to. To rule out this possibility, the UV spectra of 3C6-24AA was recorded before and after labeling with APN-M (Figure V-18). To make the two spectra directly comparable, they had to be normalized according to the concentration of the protein sample, but we were unable to directly determine the concentration of the APN-M-labelled Affimer. Instead, we scaled the signal intensity of the two spectra until the net difference between them was minimized, especially for wavelengths between 240 and 270 nm, where we were not expecting any APN-M absorption. Based on this simple test, we found that although APN-M does absorb in the 290 to 350 nm range, it has negligible absorbance at 280 nm. No correction factor is needed to adjust the linker's influence on protein detection.

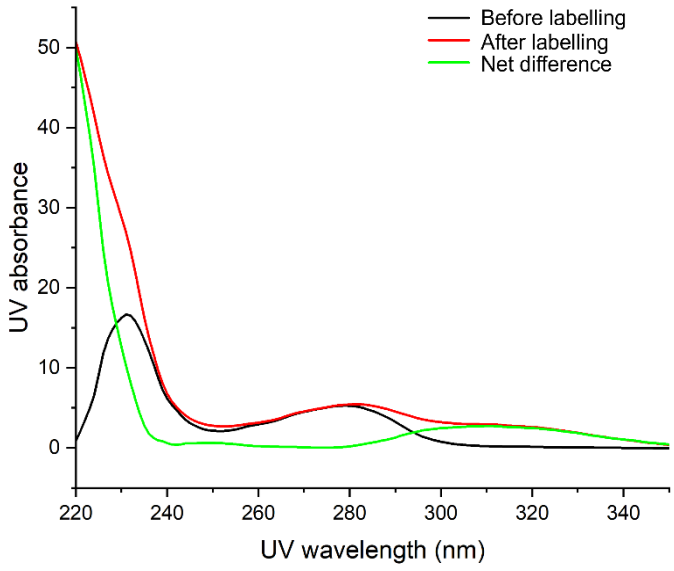


Figure V-18 A comparison of the absorbance levels of the 3C6-24AA Affimer, both prior to and post-labelling with APN-Maleimide, reveals a net difference. This discrepancy suggests that the APN-Maleimide may exhibit absorbance within the range of approximately 290 to 340 nm. The aforementioned data has been provided by Tomasz Kaminski.

V.6 SEC and Mass photometry

Affimer dimers 3C6-24AA-3C6 and 3C6-3C6 were tested for their capacity to bind to Classic CTB_5. 250 µl, 50 µM of Classic CTB_5 and 170 µl, 35 µM Affimer dimer were combined, and the mixture was then eluted through an SD75 (10/300) column to separate the two components. Without boiling, fractions of the central peak were collected for SDS-PAGE, and the band showed that Classic CTB_5 and an Affimer dimer coexisted, demonstrating the formation of a dimer-CTB_5 complex. The main peak's neighbouring small peak may have been a classic CTB_5 peak.

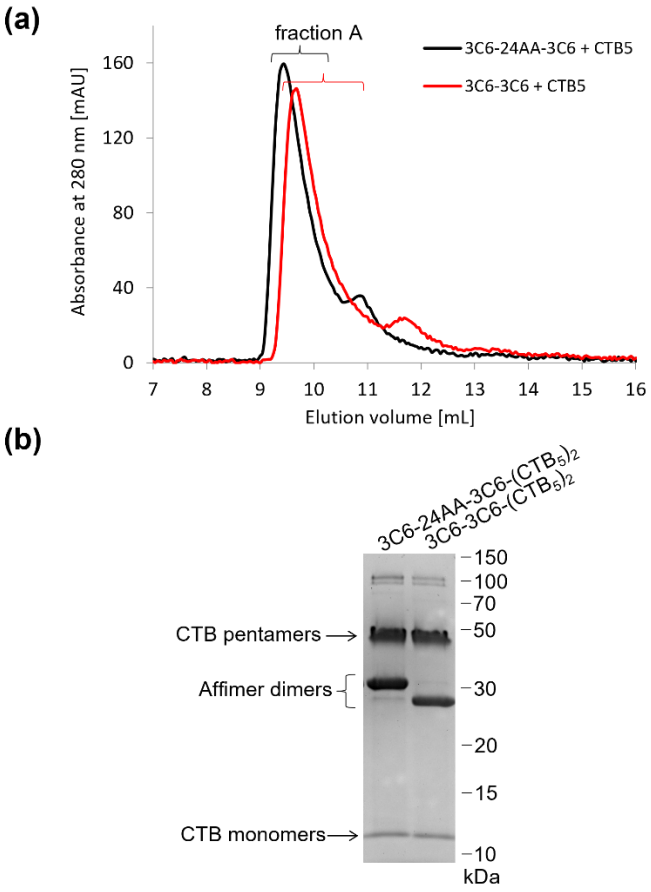


Figure V-19 Affimer dimers bind to CTB_5. (a) Chromatogram from the SEC of purified 3C6 Affimer dimers plus Classic CTB_5, mixed in a 1-to-2 molar ratio, on the Superdex 75 10/300 column. (b) 16% Coomassie stained SDS PAGE gel showing fractions A from the SEC. The Classic CTB_5 and Affimer dimers were found to co-elute.

The dimerization of CTB_5 by the 3C6-24AA-3C6 and 3C6-3C6 Affimer dimer was further studied using mass photometry. We can determine whether the resulting Affimer dimer in the GUV experiment is stable by using mass photometry, which can detect protein components at nanomolar concentrations. The mass of CTB_5 alone (50 kDa) is at the Refeyn OneMP instrument's detection threshold. A control experiment using 40 nM of CTB_5 revealed a

molecular weight of 50 kDa with a minor peak that may have been an 87 kDa dimer (Figure V-20). This can be explained as the misread from the Mass Photometry machine when two CTB_5 landing simutaneouly to the glass chip, so the landing events are mistakenly projected into a larger molecule. Next, both 3C6-24AA-3C6 and 3C6-3C6 Affimer dimer – $(CTB_5)_2$ complex eluted from SEC was loaded to Mass Photometry with a final CTB_5 concentration as 80 nM and Affimer dimer concentration as 40 nM. However, the molecular weight of the main peak was still about 50 kDa – an indication that the Affimer dimers were unbinding to CTB_5 under the detection concentration, and the correct complex mass was not predominant (146 kDa for 3C6-24AA-3C6-$(CTB_5)_2$ and 142.7 kDa for 3C6-3C6-$(CTB_5)_2$). Considering that the Kd for 3C6 Affimer to CTB_5 is 72 nM, the dissociation is falling into expectation.

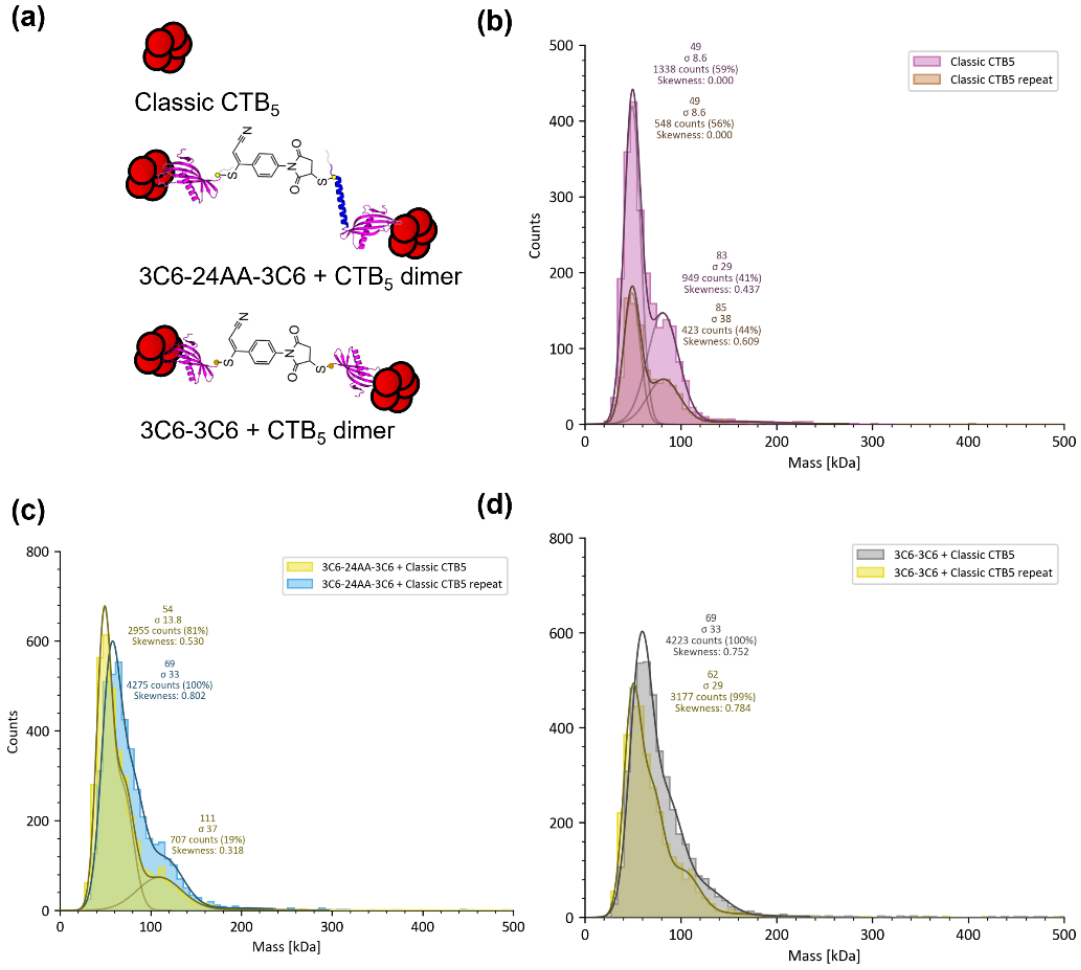


Figure V-20 (a) Scheme represent the three-time repeats mass photomoretry results of Classic CTB_5, 3C6-24AA-3C6-$(CTB_5)_2$ and 3C6-3C6-$(CTB_5)_2$. (b) Two repeats of landing Classic CTB_5 on a Mass photometry machine. The protein stock was prepared in 200 nM CTB_5 concentration, and 4 ul protein mixture was mixed with 16 ul buffer on top of the detection chip before monitoring lasts for 1 h. The molecule size was calculated based on a mass calibration curve undertaken by Biomolecular interactions facility (University of Leeds). (c) Landing 3C6-

24AA-3C6-$(CTB_5)_2$ eluted from the size exclusion chromatography, with final CTB_5 concentration as 80nM and Affimer dimer as 40nM. (d) Landing 3C6-3C6-$(CTB_5)_2$ eluted from the size exclusion chromatography, with final CTB_5 concentration as 80nM and Affimer dimer as 40nM. SEC: 3C6-24AA-3C6 and 3C6-3C6 interacts with CTB_5

The protein-membrane interaction experiments will continue even though we were unable to detect the complex formed by Affimer dimer and CTB_5 at a given concentration in Mass photometry because we were able to successfully detect the formation of a protein complex by SEC.

V.7 GUV experiment with 3C6 Affimer dimer – $(CTB_5)_2$

To examine whether the complex formed between Affimer dimer and CTB_5 can induce membrane fusion in GUV experiment, a negative control was first operated with the 3C6-24AA-3C6 Affimer dimer only (Figure V-21). This was to check whether the Affimer dimer constructs themselves can disrupt the vesicle shape and exert any influence on other detection. With a well-dispersed Affimer dimer protein stock (200nM dimer concentration, fluorescenctly labelled with NHS ester – Alexa fluor 488), GUV-Atto647 was added to a final lipid concentration of around 50 nM – 100 nM and recorded continuously. Since the record occurred immediately after adding the GUVs, the number of vesicles in the observation zone gradually increased due to the vesicles' gradual sinking. However, there is no increase in green fluorescence signal under 495 nm on the vesicle membrane – which indicates no protein binding to the vesicles. Furthermore, GUVs kept separated from each other and spherical during the observation time, and no particular phenomenon was found.

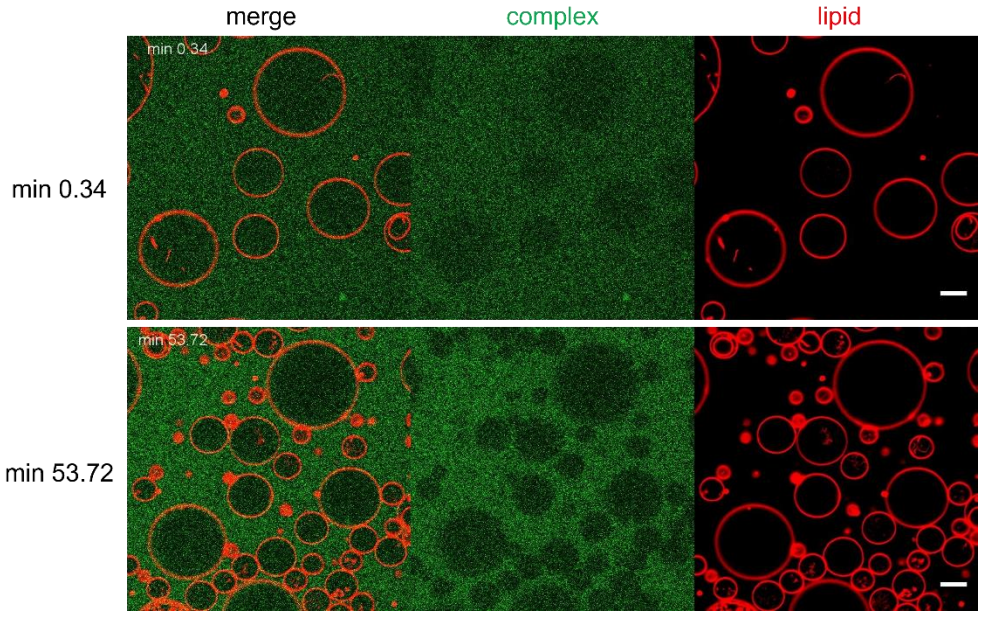
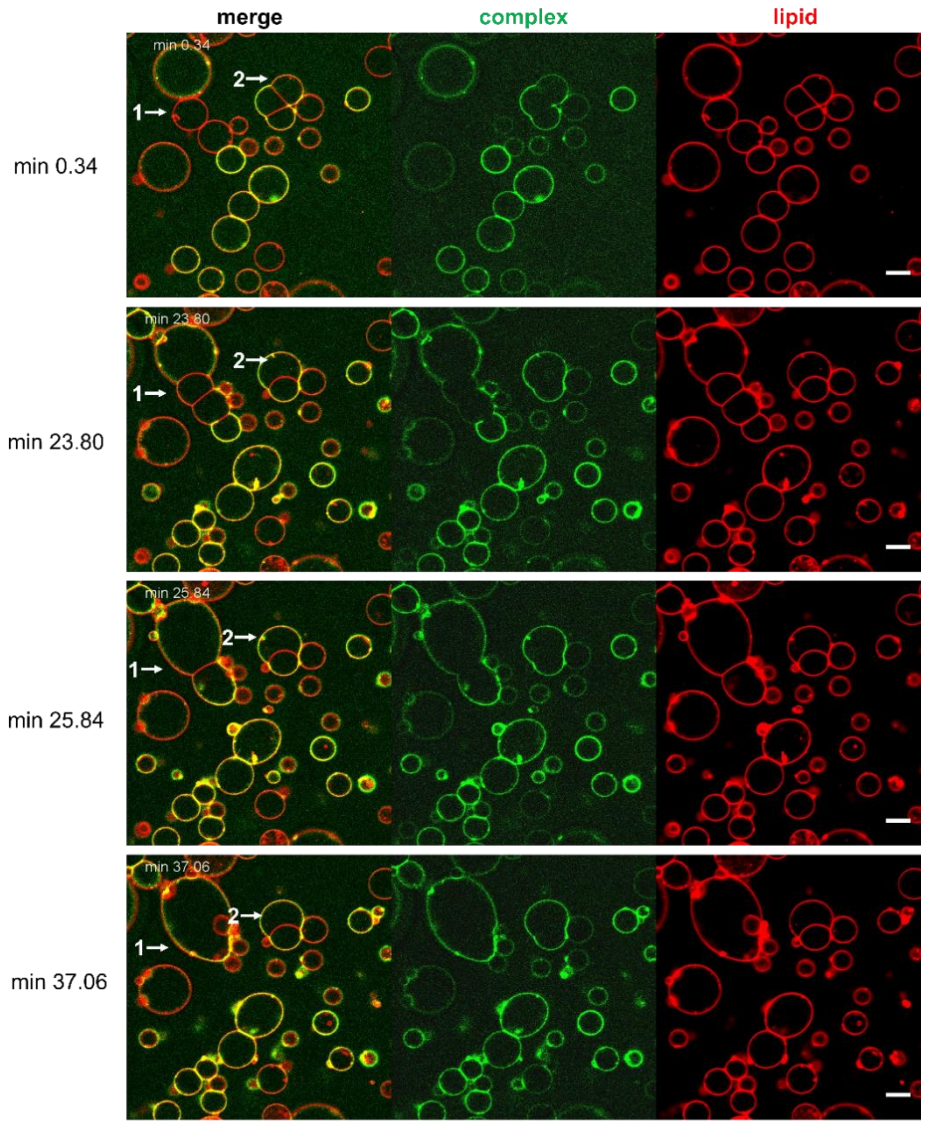


Figure V-21 Incubate Alexa Fluor 488-3C6-24AA-3C6 Affimer dimer with 2.5 mol% GUV-Atto647, with time lapse as.0.34 min and 53.72 min. GUVs composed of DOPC:cholesterol:GM1:Atto647-DOPE (65.2:30:2.5:0.3mol%, respectively). The spherical shape of the vesicles was kept until 50 min. The protein signal in the solution was amplified to indicate any leaking. During the observation time, no crosslinking, leaking or fusion events were detected in the zone. Scale bars represent 10 µm. Presented here are one of the result image obtained from a single experiment.

The GUVs with the same composition and sourced from the same preparation batch were used to detect whether 3C6-24AA-3C6-$(CTB_5)_2$ can induce membrane crosslinking or fusion (Figure V-22).



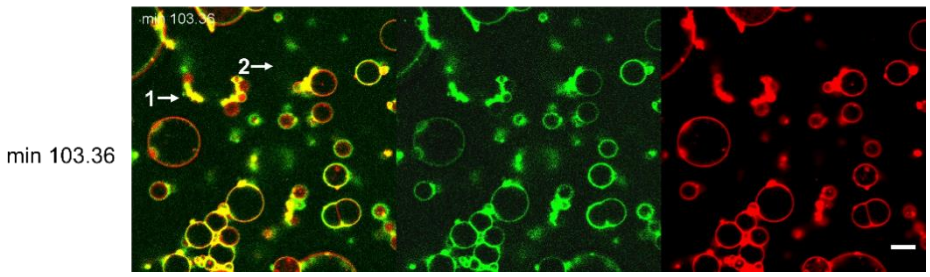


Figure V-22 Time-lapse of 3C6 Affimer dimer – CTB5 complex binding to GM1-functionalized GUVs, inducing hemifusion, fusion and rupture in successive order. GUVs composed of DOPC:cholesterol:GM1:Atto647-DOPE (65.2:30:2.5:0.3mol%, respectively). Protein signal accumulated on the vesicles, with the interface elongated as time passed, and the protein was excluded from the contact face. Fusion occurred following the hemifusion, with rupture as a consequence. Scale bars represent 10 μm. Presented here are the representative result from two repeats.

The protein stock to enable a final concentration of 3C6-24AA-3C6 Affimer dimer (200 nM dimer concentration) and Classic CTB_5 (400 nM) was diluted into GUV stock before recording. Compared with the negative control, the fluorescence of protein was accumulated on the vesicle surface indicating the attachment of Affimer dimer on the membrane mediated by the interaction between Classic CTB_5 and GM1. Membrane hemifusion was observable in the first recorded images evidenced by the exclusion of protein at the GUV interface with only the Atto647 signal exposed at position 2 in frame 1. Several different membrane phenomena were seen in the p1 position, starting with the mutual hemifusion of four conjugated vesicles. With the exclusion of proteins at the interface, the contact point between the two GUVs gradually grew longer. The hemifusion ultimately results in a fusion event and rupture. P2 position shows a hemifusion between three vesicles, along with fusion and rupture at the end of the recording period.

V.8 Crosslinking between cell and GUV

To investigate whether the 3C6-24AA-3C6-$(CTB_5)_2$ complex can further induce the internalization of GUV to cell, the Jurkat was prepared with the help of Francesca Rosato (University of Freiburg) and preincubated for 1.5 hours at 4 ºC with 100 nM 3C6-24AA-3C6-$(CTB_5)_2$ (50 nM 3C6-24AA-3C6, 100 nM CTB_5). This temperature was set to 4 ºC to prevent protein internalization. Before recording, the GUV-Atto647 (with 2.5 mol% GM1) was added at a suitable vesicle density (to around 50-100 nM lipid concentration) without removing the free protein in the solution by exchanging buffer or washing the cell. The recording lasted for two hours, and several phenomena were observed. First is the increased fluorescence intensity on the GUV membrane and partial cell domain, indicating at least some attachment between GUV and cell with protein complex (Figure V-23). In position 1, there is a red line

indicating the deformation of the GUV membrane; after this appeared, the cell seemed to be wrapped with the GUVs, which further led to cell death.

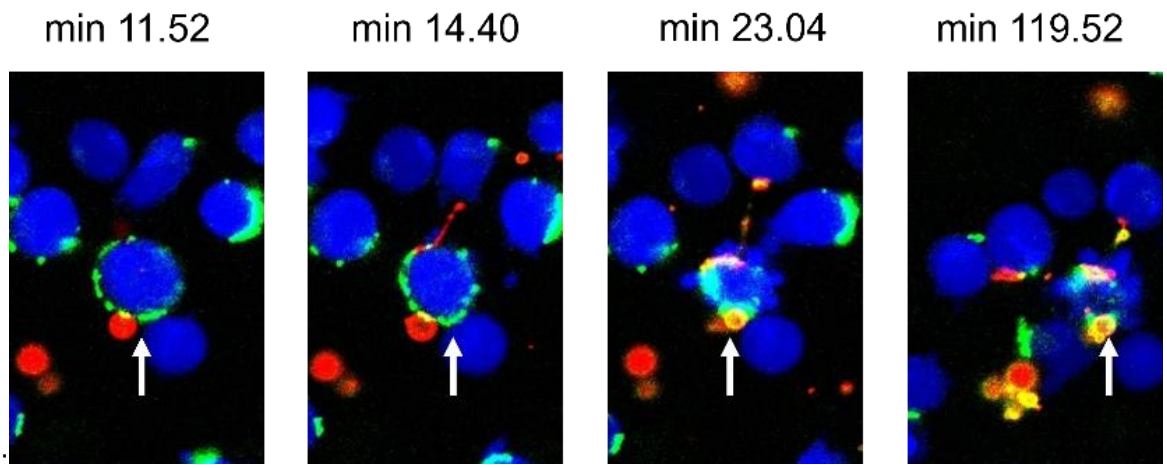


Figure V-23 Timelapse microscopy images indicated one case of GUV internalized into suspended Jurkat cells. 200 nM classic CTB5 was fluorescently labelled with Alexa fluor 488 (green) and mixed with 100 nM 3C6-24AA-3C6 Affimer dimer – this protein mixture was further incubated with Jurkat cell (blue) and Atto 647-GUV containing 2.5mol% GM1 (red) at different time points. The internalization of GUV to cell happened at 23 min, which finally led to cell death. Presented here are one of the result image obtained from a single experiment.

The interaction between GUV and cells was more evident in a different observation area (Figure V-24). Following the provision of both cells and GUVs, it was demonstrated how cell movement interfered with the shape of the GUVs during an incubation period at 37 ºC. The exclusion of protein (green) at the p1 position on the contact surface between the vesicle (red) and the cell (blue) shows that GUVs can attach to the cell and that the GUV explosion is caused by the push and pull of the two cells. Following the explosion, GUV fragment pieces bound to the cell. During the two hours of observation, no cells were found to be actively able to remove these GUV fragments.

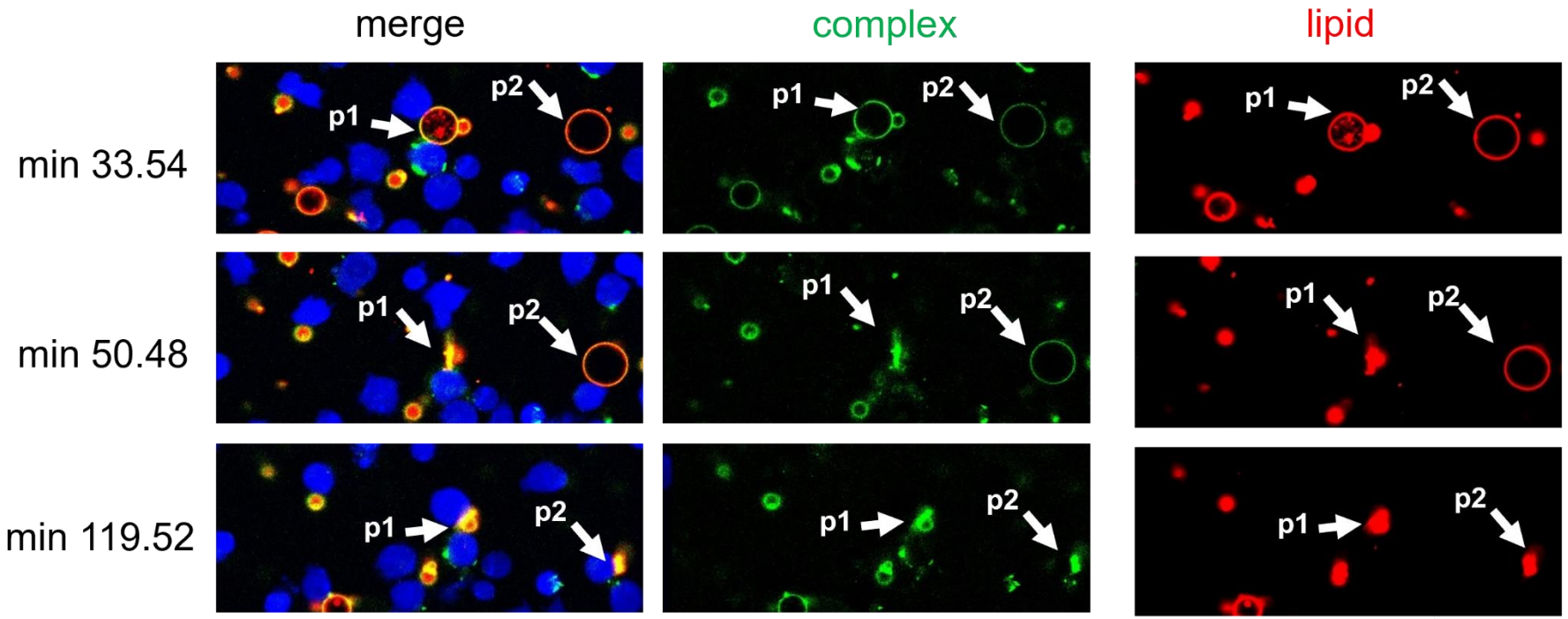


Figure V-24 A confocal microscopy image showed the crosslinking between the suspended Jurkat cell and Atto 647-GUV (2.5mol%). At 33.84 min, protein exclusion from the cell-GUV interface was observed. Two cells were contacted with the same GUV and moved in the opposite direction to deform and eventually explode the GUV at 60.48 min. After GUV was destroyed, the membrane was attached to the cell's surface until the recording's end. Another GUV rupture was also observed at position 2 at the end of 2h recording. Presented here are one of the result image obtained from a single experiment.

V.9 Lipid mixing assay

A lipid mixing assay was carried out to quantify the lipid exchange rate induced by Affimer dimer-$(CTB_5)_2$ protein complex after GUV fusion events were observed using microscopy. The Affimer dimer is made of Affimer and APN-Maleimide, which differs from a protein that only consists of native amino acids and can alter the appearance of the absorbance spectrum. Since the NBD-Rhodamine lipid mixing assay uses a detection wavelength range of 500–600 nm, it will be problematic if APN-Maleimide absorbs in that range. In order to determine whether 3C6-24AA-3C6 had absorbance in the 400–600 nm range, a straightforward experiment was conducted to look at the absorbance region of the molecule (Figure V-25).

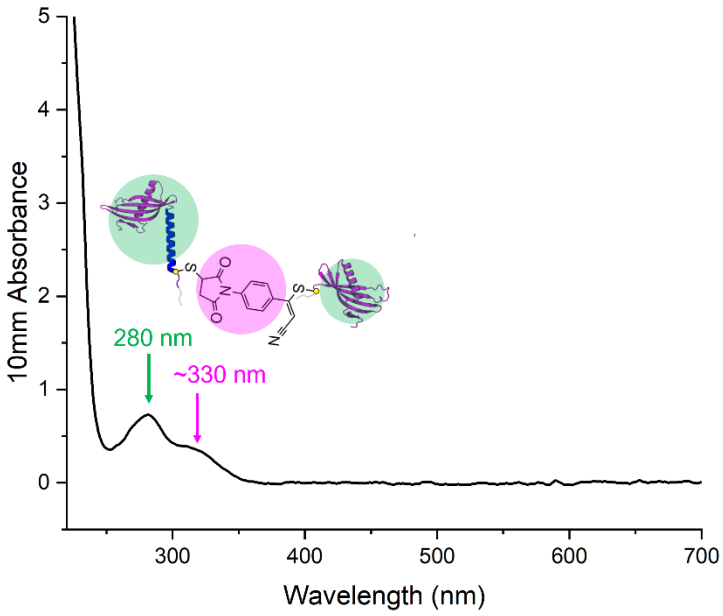


Figure V-25 Absorbance spectrum of 3C6-24AA-3C6 detected by Nanodrop. The signal of 280 nm indicated the presence of tryptophan and histidine amino acid residues. An extra peak with around 330 nm absorbance was observed, as supposed to be produced from the APN-Maleimide linker.

No significant absorbance signal was observed during 400-600 nm wavelength, so the APN-Maleimide linker will not exert an unexpected influence on the lipid mixing assay. With this basic knowledge, the NBD-Rhodamine energy transfer lipid mixing should be reliable from signal detection.

The lipid mixing assay was performed as described in section VII.14. Briefly, the test liposome was made containing 1-to-4 molar ratio of 800 nm FRET-LUV and blank-LUV. *El Tor* CTB_5 was premixed with 3C6-24AA-3C6 or 3C6-3C6 Affimer in 2:1 ratio (CTB_5 pentamer: Affimer dimer) and loaded into plate to achieve a final concentration of 4 µM

CTB_5. For each sample, a two-fold dilution series of the protein was used while the lipid concentration was kept the same for all wells (133 µM). In parallel, Affimer dimer and *El Tor* CTB_5, were investigated separately as the negative control. The lipid-mixing was observed over 8h with sample reads every 8 minutes (Figure V-26). Since the liposome's self-fusion rate was too high for the data to be relied upon after five hours, the lipid mixing rate at various protein concentrations was further calculated using the fluorescence data from the only first five hours (Figure V-27).

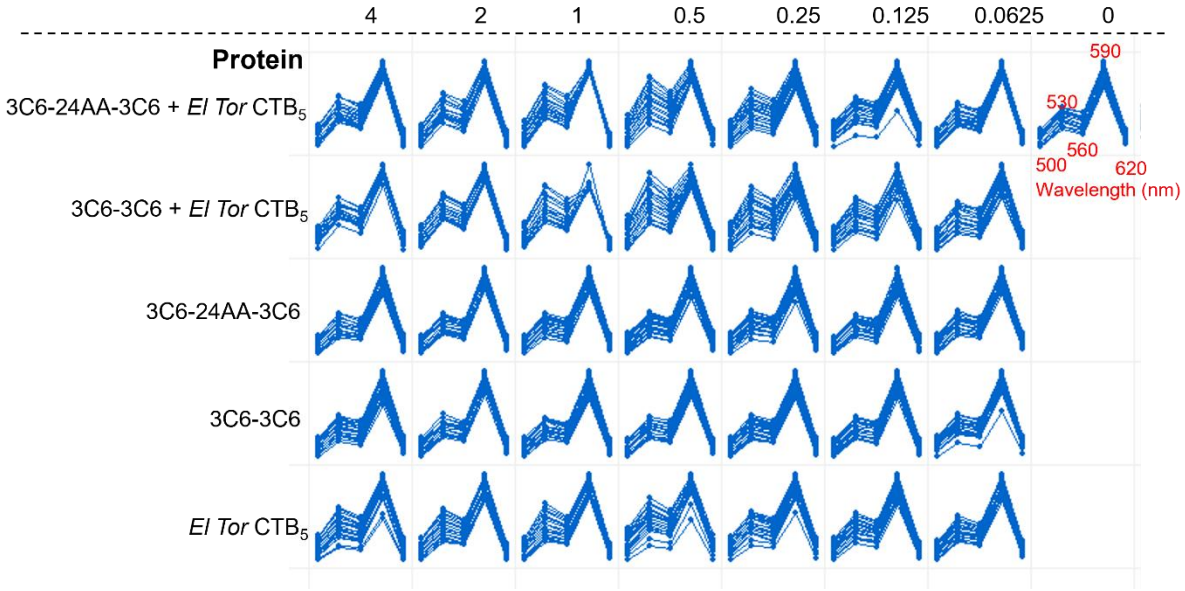


Figure V-26 Five different wavelengths' fluorescence intensities—500 nm, 530 nm, 560 nm, 590 nm, and 620 nm—overlapped during the 8-hour detection period. respectively. The overall test lasts 8 h, and the fluorescence intensity ratio under 530 nm to 590 nm could reflect the lipid mixing efficiency. Presented here are one of the result image obtained from three-time repeats.

Figure V-26 compares the lipid mixing efficiency between different protein complexes under various concentrations. Lipid exchange was observed for both complexes of Affimer dimers – 3C6-3C6 and 3C6-24AA-3C6, after they were mixed with *El Tor* CTB_5 separately. However, the Affimer dimers alone did not induce significant lipid mixing, as evidenced by seemingly unchanged ratio between absorbance under 590 nm to 530 nm during the 8h incubation time (Figure V-26). Surprisingly, *El Tor* CTB_5 itself did also cause lipid exchange. One possibility is that both concentrations of lipid and protein used in the lipid mixing experiment are much higher (133 µM for lipid and 4-0.06 µM for CTB_5) than those used compared to the GUV experiment (~100 nM for lipid and 400 nM for CTB_5). Hence, *El Tor* CTB_5 dimer could be potentially formed natively under such crowded environment. However, compared to the lipid mixing result from $CTA2B_5$ (section IV.10, Figure IV-22), in which there was no significant lipid exchange observed, one could deduce the *El Tor* CTB_5 dimer

formation happened at the non-sugar binding site, and the process can be interrupted by CTA2 peptide due to the steric hindrance exerted from the protruding segment.

To gain more insights into the lipid mixing speed among different protein constructs, the lipid mixing rate was calculated, which indicates how much the lipid mixing percentage changes per minute. Global fit for three repeats of lipid mixing results was performed (section VIII.5.1), and the slope and the deviation were collected to represent the lipid mixing rate. As observed in Figure V-26, a minimal lipid mixing rate was obtained when the system did not contain *El Tor* CTB_5 – for these tests. The liposome self-fusion contributes to the lipid exchange rate. However, when *El Tor* CTB_5 participate in the system, no matter is premixed with Affimer dimer or present alone, a significant increase in the lipid mixing rate was observed at an optimal concentration.·

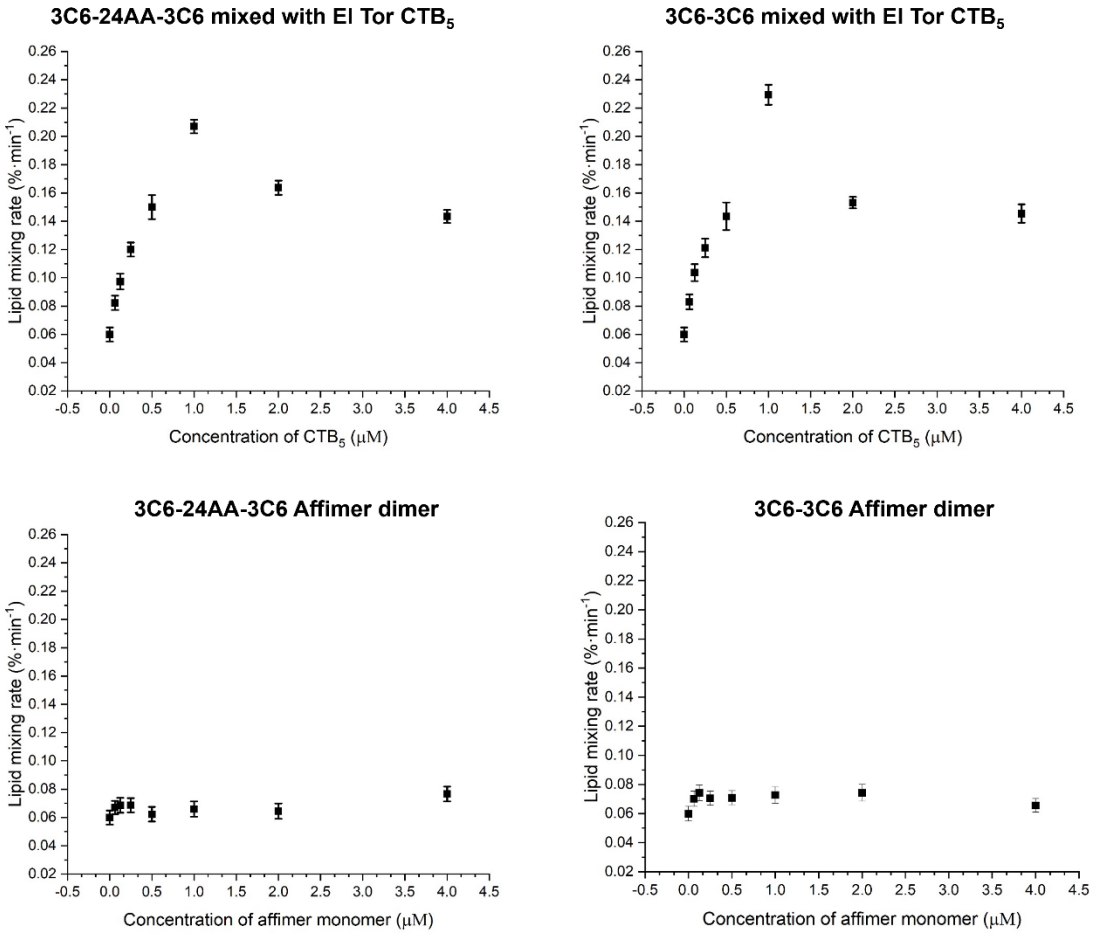


Figure continues

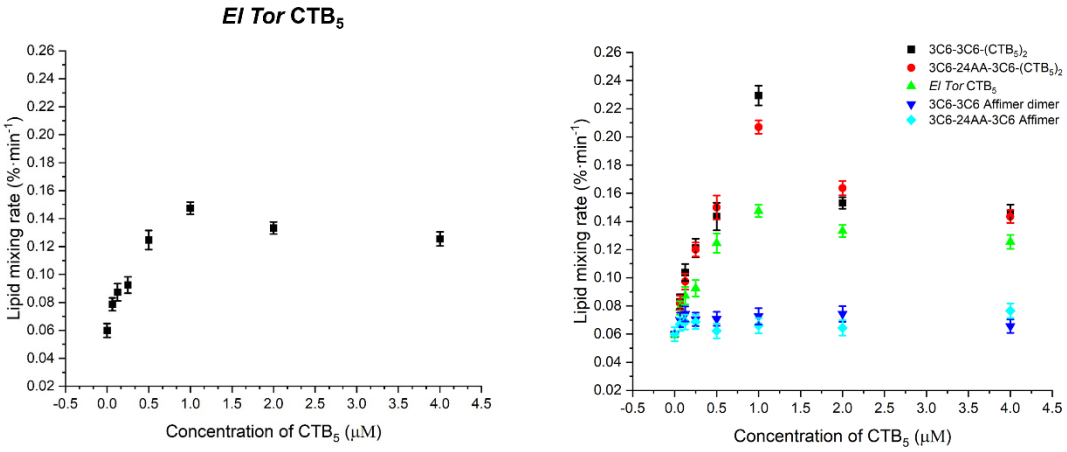


Figure V-27 Lipid mixing rate (%· min^{-1}) of selected protein mixtures under different protein concentrations. The outcome was calculated based on three-time repeats.

Based on the lipid mixing rate pattern in Figure V-27, one typical pattern for the lipid mixing rate caused by *El Tor* CTB_5 as well as its mixture with Affimer dimer is: they all possess an optimal concentration to achieve the maximum lipid exchange efficiency, and the sweet point is always when the CTB_5 concentration is around 1 μM. However, the actual optimal concentration might be 0.5-2.0 μM since the data points are far away from each other. Comparing the lipid exchange efficiency induced by *El Tor* CTB_5 as well as its mixture with Affimer dimers when the CTB_5 concentration is within the range of 0.0-0.5 μM, one can figure that the combination of 3C6-24AA-3C6 or 3C6-3C6 Affimer dimer with *El Tor* CTB_5 could exert higher lipid exchange rate comparing to using *El Tor* CTB_5 alone under the same CTB_5 concentration. Furthermore, even though the lipid exchange underwent inhibition when the protein concentration is higher than 2 μM, the Affimer dimer premixed with the CTB_5 group can still end with a 0.14 %·min^{-1} lipid exchange rate. At the same time, *El Tor* CTB_5 alone could only achieve 0.12 %·min^{-1}.

However, their behaviour more significantly differentiates three protein mixtures when the CTB_5 concentration is at 1.0 μM. When 3C6-24AA-3C6 mixed with *El Tor* CTB_5, the lipid exchange rate reached 0.21 %·min^{-1}, while for 3C6-3C6 with CTB_5 mixture, the value increases to 0.23 %·min^{-1} and for *El Tor* CTB_5 itself is 0.15 %·min^{-1} only. It is important to understand how different protein mixtures interact with the membrane, which can shed light on further guidance on designing the protein construct to induce efficient membrane fusion.

To understand how CTB_5 interacts with the membrane, we need to understand if the resource per liposome, *e.g.*, the number of GM1 and liposome area, could allow the CTB_5 to lay on the surface with all five binding pockets occupied. First, we consider the liposome a

simplified and rigid model without taking into account for the lipid exchange between the inner and outer membrane layer (*i.e.,* Flip-Flop). Since the lipid mixing assay was conducted using the 5 mol% GM1 LUV, there will be 5 GM1 for every 100 lipid molecules. 5 GM1 can be found in an area of about 72 nm^2 because the headgroup area for DOPC is about 0.72 nm^2. However, because CTB_5 only has a surface area of about 33.2 nm^2, it can only cover about half of the liposome's surface. This suggests that CTB_5 can saturate the liposome's surface completely under 5 mol% GM1 concentration.

What is the least amount of CTB5 required to achieve such a saturation? Given that the lipid concentration in each well of the lipid mixing assays is 133 µM, 0.67 µM CTB_5 could completely occupy the surface of the 800 nm, 5 mol% GM1 liposomes. In fact, it appears that the maximum lipid mixing rate can be reached when the concentration of CTB_5 is ranged from 0.5 µM to 1.5 µM; the discrepancy between calculation and observation may be due to inaccurate measurement during liposome preparation, dynamic inner-outer membrane interaction, and CTB5 endocytosis.

If complete saturation of GM1 is the factor leading to a maximum lipid exchange rate, how can the phenomenon be explained more intuitively? From the lipid mixing result, when the concentration of CTB_5 is less than 1 µM, the lipid exchange rate increases proportionally to protein concentration. Since $CTA2B_5$ did not induce such a lipid exchange, it is possible that the lipid exchange was happening through the CTB_5 dimer formation due to local aggregation (CTB_5 self-dimer theory), or the CTB_5 bind to one liposome still have free GM1 binding site for the second liposome (lateral CTB_5 theory). If CTB_5 self-dimer theory stands, when there is little protein, less dimer is formed, and the connection between two neighbouring liposomes is not as strong as a higher concentration – such a relationship keeps when all the GM1 on liposomes are allocated to interact with CTB_5. However, it fails to explain why the lipid exchange drops when the concentration of protein is too high. On the other hand, the lateral CTB_5 theory can also explain the phenomenon – when there is little protein compared to the available GM1 on membrane, the CTB_5 will predominantly interact with the GM1 on the same vesicle and reach out to the GM1 on the neighboring membrane – hence the lipid mixing can occur through the communal connection between two liposomes through one CTB_5. However, when there is excessive CTB_5, the GM1 on the same liposome will be quickly occupied by the CTB_5, leaving no space for the neighboring CTB_5 from the other vesicle to bind to.

Maybe that can explain why CTB_5 itself can induce lipid exchange. However, what happens when Affimer dimer coexists with CTB_5 in a lipid environment, and why can the interaction

between Affimer dimer and CTB_5 boost the lipid exchange rate, especially near the GM1-saturation concentration? When the CTB_5 concentration is enough to occupy all the GM1 in the lipid system, the Affimer dimer creates more restraints for two CTB_5 to bind to the same membrane than free CTB_5. This, in turn, creates more chance for the second CTB_5 to bind to the neighbouring vesicles. The Affimer dimer could also work as a stronger linker to enable a stable hemifusion structure as mediate. Based on the CD result (Figure V-6), the inserted 24AA peptide could exert a degree of flexibility, but overall a rigid helix structure under 25 degrees, which could explain why it has less lipid exchange rate at the saturation point compared to 3C6-3C6 Affimer without the 24AA inserted – only a shorter distance is needed to conquer to disrupt two neighbouring liposomes for 3C6-3C6 Affimer dimer.

V.10 Conclusion and future works: Affimer dimer as a potential and powerful fusion tool

The advantages of screening specific Affimer to bind the target protein open a more expansive world for constructing different heterodimer lectins to achieve membrane fusion. Moreover, the functionality of APN-Maleimide was examined, and it can also be used as a general tool to link heterodimeric Affimers together. Suppose the Affimer meets aggregation issues when two of them are directly linked through APN-M. In that case, a hydrophilic helix can be inserted at the C terminus to gap them away, resulting in a more stable protein cluster. The longer distance between two Affimers, even though it slightly decreases the lipid exchange rate at saturation point, could still achieve a much better lipid mixing ratio than CTB_5. Moreover, the Affimer dimer complex used in this project can be mixed and matched with another linker used in another project, e.g., coiled-coil and PEG flexible linker, hence creating a more dynamic system to examine whether the rigidity of the linker could exhibit influence on the lipid exchange rate.

The future project can be established based on the 3C6-3C6 Affimer dimer by cooperating CTB-targeted Affimer and VTB-targeted Affimer and study whether such a heterodimeric Affimer dimer can be used to build hetero-lectins to achieve membrane fusion.

Chapter VI Conclusion and future works

VI.1 Summary

The purpose of my project is to determine which protein architecture is most effective for this CTB_5 dimer in order to achieve more efficient membrane fusion, as well as whether membrane fusion can be achieved through the interaction between CTB_5 and GM1 by assembling various CTB_5 dimers. In addition, we compare several methods of composing the desired protein assembly and show the difficulties that may be encountered in the preparation process. Further, the interaction of each protein system with the membrane is examined, either quantitatively by lipid mixing and QCM-D or qualitatively by confocal microscopy. In this way, we attempt to explain how the protein interacts with the membrane and whether or not this interaction promotes membrane fusion. Last but not least, we conducted cellular tests on the CTB_5 dimer system, which consists of a coiled coil or an Affimer dimer as a linker, to determine whether it can be used as a drug delivery system in more intricate physiological objects.

To achieve this, two or more CTB_5 have been put together using three different systems. The first system is to assemble two CTB_5 in a *trans*- or *cis*- configuration using the interaction between streptavidin and biotin, so that we can test whether membrane fusion is already possible when two CTB_5 are self-assembled in this manner (Chapter II). Upon determining that two CTB_5 molecules were not efficient enough to cause membrane fusion, a third CTB_5 molecule was added to create a protein architecture that contained all three CTB_5 molecules assembled together. Our findings revealed that when only two CTB5 molecules were used and connected by a flexible linker such as PEG4, these molecules would bind together first in the same plane of the vesicle, thus preventing any additional CTB5 molecules from reaching out to bind to nearby vesicles. However, with the presence of a third CTB5 molecule, the communication efficiency between the two vesicles was increased. This was due to the fact that when two CTB5 molecules bind in the same plane, the third CTB5 molecule always faces in the opposite direction due to structural limitations. As a result, the likelihood of binding to the adjacent liposome was increased, thus promoting membrane fusion efficiency. Additionally, we discovered that when the ratio of CTB_5:GM1 in the solution is gradually increased, the lipid mixing rate tends to rise and then fall (*optimal-concentration tendency*). It's possible that this happens because there isn't enough GM1 in the neighboring membrane to bind to CTB_5 when the protein concentration is too high, disrupting the communication between the two membranes.

We further questioned if it would be possible to assemble the two CTB_5 in either a parallel or antiparallel configuration using just a coiled coil as the linker. It would be more challenging to allow two CTB_5 to lay on the same vesicle when using the coiled coil because its rigidity is higher than that of the PEG4 linker (Chapter III). Additionally, we could look into whether the linker's orientation affects how well the membrane fuses. The results of the lipid mixing and GUV experiments demonstrate that, at specific range of CTB_5:GM1 ratios, coiled coils can indeed result in greater lipid exchange and even membrane fusion. However, there is also an optimal concentration tendency observed for two out of three examined protein assemblies. Only AP2.1 exhibited a different pattern of membrane behavior, as its rate of lipid exchange increased as the protein concentration of CTB_5 increased (and hence the CTB_5:GM1 ratio is increased). According to the QCM-D experiments, this might happen because the coiled coil dissociated when AP2.1 interacts with the membrane, always leaving the spare protein available to engage with the adjacent membrane and create a new antiparallel coiled coil.

Both of the aforementioned systems use the native CTA2 peptide as the CTB_5 binder, but this has the drawback that CTA2 must be expressed alongside CTB_5 in order to mosaic into the cavity created by CTB_5. As a result, there are some restrictions on its use, including the inability to easily switch linkers and to control when to begin the membrane fusion. Affimer dimer serves as a linker between two CTB_5 in a third system. It is more adaptable because one available Affimer named 3C6 can bind to the CTB_5 surface with high binding affinity (Kd is around 72 nM). Using 3C6-based Affimer dimer as linker to connect two CTB_5 achieved good membrane fusion efficiency, but it also displays the aforementioned *optimal-concentration tendency*.

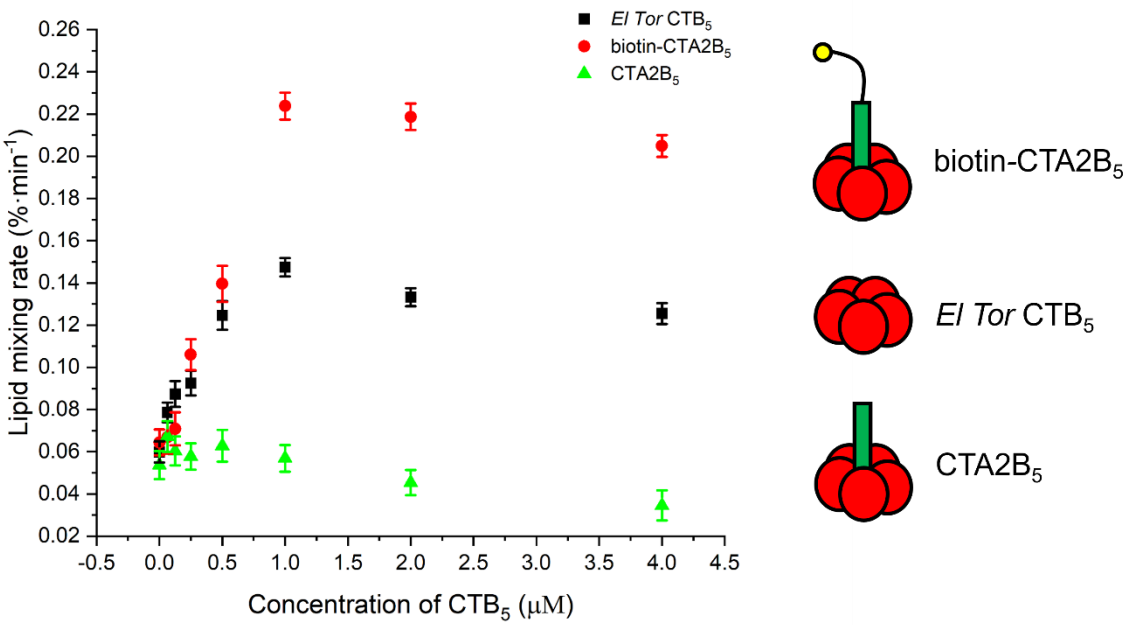


Figure VI-1 Overlay of the lipid mixing rate induced by El Tor CTB_5, biotin-$CTA2B_5$ and $CTA2B_5$. The final lipid concentrations are controlled to 133 µM, using 800 nm liposome containing 5 mol% GM1.

The different lipid exchange patterns induced by *El Tor* CTB_5, $CTA2B_5$, and biotin-$CTA2B_5$ are another phenomenon that should not be disregarded. *El Tor* CTB_5 unexpectedly also induces some lipid mixing, but this lipid exchange phenomenon is blocked by CTA2. Additionally, when the CTA2's N-terminal was decorated with biotin-PEG4 linker, it was more effective than *El Tor* CTB_5 at recreating the lipid mixing phenomenon. The cause of this is unclear, but one theory postulates that CTB_5 either forms dimers or interacts with neighboring vesicles via lateral binding, and that CTA2 disrupts this interaction. As a result, either the CTB_5 dimer cannot be formed, or the CTA2 peptide blocks its lateral GM1 binding pocket, reducing the interaction between vesicles. Through weak interactions with the membrane, the biotin group addition might be able to restore this connection.

VI.2 Future works

First, for the CTB_5 dimer with the coiled coil as a linker, the length, binding strength, and orientation of the coiled coil, as well as the length of the CTA2 as an extended linker, all affect how the CTB_5 dimer behaves on the membrane. By using a well-researched set of coiled coil pairs designed by Wolfson's lab [116] with various binding affinities from pM to µM, it is possible to assemble CTB_5 dimers in various strength. The lipid exchange rate induced by such protein assembles can be further measured by lipid mixing and content mixing assay. It is possible to conduct these experiments with heterodimeric coiled coils, hence heterodimeric lectins can be achieved which may have even more practical uses.

The second subject that can be thoroughly investigated is the use of Affimer as a binder tool to link two or even more lectins together. Despite being a very effective method for creating Affimer dimers, chemical ligation requires more processing steps, which adds time. It would be possible to create a super-lectin system if we could combine coiled coil and Affimer by joining the procoil from heteromeric coiled coil at the C-terminus of Affimer. This system would benefit from being more easily produced and flexible in how it could be put together.

In my project, only lipid mixing assay was performed – but it functions by measuring the lipid exchange rate rather than membrane fusion, which makes it unable to distinguish hemifusion from fusion. There are two ways to differentiate between fusion and hemifusion[81]. One goes along with the content mixing assay. A popular method relies on the development of fluorescent $Tb^{3+/}$dipicolinic acid complexes[146, 147]. This assay uses two populations of vesicles that have been loaded with either $TbCl_3$ or dipicolinic acid inside. The vesicles combine to form a Tb^{3+}/dipicolinic acid complex, which possesses a fluorescence intensity about 10,000 times more than free Tb^{3+}. EDTA can further acts as a quencher to examine the leakage of liposome, since it can form non-fluorescent Tb^{3+}/EDTA if the liposome has been leaking. The second involves measuring the lipid mixing rate only occurs at the inner membrane[148, 149]. To achieve this, sodium dithionite is added in solution to get rid of the NBD florescence from the outer membrane. Since sodium dithionite cannot penetrate through the lipid layer, any dequenching in the inner membrane could be used to explain the presence of fusion events. These experiments can serve as supplementary test in future.

Chapter VII Method and Material

VII.1 Gene construct: PCR, digestion, and ligation

Two types of genetic methods were used in this project to perform mutagenesis depending on how many mutations or the type of mutation to be executed. One was amplifying and digesting a long DNA piece and religating it into a new plasmid – this was usually aiming to introduce a new piece of protein sequence as in the case of coiled coil–CTA2B_5, or to change the plasmid used for protein expression. The other type of mutagenesis is less dramatic, modifying one or two bases by QuickChange mutagenesis or inserting/deleting/mutating a short piece of DNA fragment through Q5 mutagenesis. The detailed protocols for representative examples are listed in the following sections.

VII.1.1 Preparation of plasmids for expression of MBP–procoil–CTA2B5

Based on the pTRBAB5-G1S plasmid[80] designed in the Turnbull lab (see section VIII.1.1 for details), from which the MBP-CTA2B_5 complex can be expressed from a single promotor, the CTA2 region can be replaced with procoil-CTA2 by genetic digestion and ligation reactions.

VII.1.1.a Preparation of the insert

DNA fragments (Figure VII-1) were ordered from TWIST BIOSCIENCE with additional filler DNA added at both 5' and 3' ends to reach 300bp length to minimise the price of the DNA synthesis. Eva_geneblock_forward primer (GTA CCA TGC ATA TGG GAT CCG) and Eva_gene_block_reverse primer (GCT AAT CGA ATT CAC CTC GAG) were ordered separately from Genescript to clone the fragment.

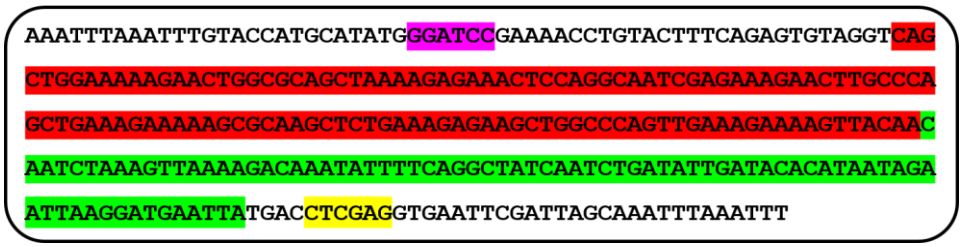


Figure VII-1 The DNA fragment was ordered from TWIST BIOSCIENCE to replace the original CTA2 peptide with the AP2.1-short CTA2 peptide. BamHI and XhoI cutting sites are located from the 5' to 3' end of the DNA pieces, with a procoil region and short CTA2 peptide.

A standard PCR reaction was set up in 50 µL, with the composition and reaction conditions listed below in Table VII-1 and Table VII-2. If not stated explicitly, the buffers and enzymes were from New England Biolabs (NEB). The amplified product was subjected to DNA electrophoresis to confirm successful amplification before purification.

Table VII-1. The content of PCR reaction to clone the procoil-CTA2 gene fragment.

Ingredient	**Volume**
Water	makeup to 50 µL volume
10 * PCR buffer [7]	5 µL
Template DNA fragment	50 ng
Forward primer	0.5 µL, (use 100 µM stock)
Reverse primer	0.5 µL, (use 100 µM stock)
dNTPs	0.5 µL (use 25mM of dATP dCTP dGTP dTTP each as stock)
Phusion High Fidelity (PWO) polymerase	0.2 µL

Table VII-2 Conditions for the PCR reaction. X is the lowest T_m (melting temperature was calculated from thermofisher web tool available from[8]) between the primer pair minus 0.5 °C. The extension time Y was counted in relation to the length of the DNA fragment to be replicated, with each 500 bp adding 1 min to the initial 30s.

Time	**Temperature (°C)**	**Aim**	**Repeat**
30s	95	Denaturation	Repeat 30 times
30 s	X(52.4)	Annealing	
Y (64 s)	72	Extension	
2Y (128 s)	72	Final extension	Final and storage
-	4	Storage temperature	

After the PCR process, the amplified DNA fragment was separated from contaminants before digestion. The QIAquick PCR purification kit (QIAGEN) was used: 50 µL PCR product

[7] 10*PCR buffer: 200 mM Tris HCl, 100 mM potassium chloride, 100 mM $(NH_4)_2SO_4$, 10% v/v Triton X-100, 10 mg/mL BSA, 15 mM $MgSO_4$, pH 8.8

[8] https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.htmL#

was loaded in the QIAquick spin column (with a 2 mL collection tube) and centrifuged. Further, 500 µL DNA binding buffer (PB) and 750 µL ethanol-containing PE buffer were used to wash the column successively, with centrifugation between each step. It should be emphasized that the PE buffer must be removed entirely by extensive centrifugation, since the trace of ethanol will decrease the solubility of DNA in EB buffer (10 mM Tris-Cl, pH 8.5) of which 50 µL was applied to elute the purified DNA fragment.

VII.1.1.b Preparation of the vector

E. coli XL10 cells containing the pTRBAB5-G1S plasmid were grown in 5 mL sterilized Luria-Bertani (LB) media with 100 µg/mL ampicillin at 37°C for 20 h. Then the cells were harvested by centrifuging at 4000 x g in a Heraeus 3S-R centrifuge for 10 min. Buffers from QIAGEN miniprep kit including P1, P2, N3, PE and EB were used to extract the plasmid. Centrifuged cell pellet was resuspended in 250 µl Buffer P1 from QIAGEN miniprep kit before transferring to a 1.5 mL Eppendorf tube. 250 µl buffer P2 was mixed until the solution became clear. Afterwards, 350 µl Buffer N3 was added to lysate the cell into colorless but viscous suspension. After centrifugation, the supernatant was poured into the QIAprep spin column and washed with Buffer PB and Buffer PE successively, before finally eluting with 50 µl buffer EB.

pTRBAB5-G1S contains BamHI (GGATCC) and XhoI (CTCGAG) restriction sites between the MBP-TEV cleavage site and CTB subunit; hence these restriction enzymes were used to obtain the desired linear vector for inserting new DNA pieces. The digestion process was: 7.5 µg plasmid was mixed with 5 µL 10 X Cutsmart buffer (NEB) and 0.75 mL of the corresponding restriction enzymes (NEB). The reaction mixture was diluted to 50 µL with nuclease-free water and allowed to react at 37 °C for 4h. The final product was evaluated by nucleic acid electrophoresis, the digested vector band with the correct migration distance was cut out with a sterilized blade, and purified using a QIAquick Gel Extraction Kit.

VII.1.1.c Ligation

After the digested pTRBAB5-G1S vector and DNA insert were prepared, the following solutions were mixed to obtain the complete plasmid with the targeted gene fragment inserted:

Table VII-3 Composition of the ligation reaction.

Ingredient	**Volume**
Water	Make up to 10 µL

Plasmid	50 ng
Insert DNA	X ng
T4 ligase buffer (NEB #B0202)	1 µL
T4 ligase (NEB #M0202)	0.5 µL

The value of X in Table VII-3 was calculated using the Ligation calculator[10] as the molarity for the inserted DNA fragment to vector in ligation reaction is 3:1. The mixture was incubated at r.t. for 50 min and then transferred to a 65°C water bath for 20 min to inactivate the enzymes. The ligated plasmid was used directly to transform XL10 cells by the following protocol.

VII.1.1.d Transformation

To prepare enough of the vector sequence for sequencing, the amount of the ligated plasmid needs to be amplified in XL10 cells. The steps of transformation involved first thawing the competent cells on ice for around 10 minutes, while chilling the ligation reaction mixture/purified plasmid at the same time. For ligation reactions, 15 µL of competent cells was mixed with 1 µL of ligation reaction mixture, gently flicking the tubes without touching the bottom of the Eppendorf tube to avoid warming the cells to the temperature that might already induce transformation. On the other hand, 10 µL cells is enough for 1 µL of purified plasmid at 100 ng/µL. The cell-DNA mixture was placed on ice for 15 minutes, and then heat shocked at 42 °C for the 30 s before transferring back onto the ice for another 5 minutes. 200 µL of sterilized LB media was added and the tube incubated at 37 °C for 30 minutes (for purified plasmid) or 1 hour (for ligation reactions). The media was spread evenly over an agar plate, which was incubated at 37 °C overnight or 25 °C over the weekend with the agar side on the top. A single colony was used to inoculate the starter culture, and the plasmid was extracted using QIA DNA extraction kit as described in section VII.1.1.b . The pMAL_reverse primer (0.25 µL, 100 ng/µL; TGT CCT ACT CAG GAG AGC GTT CAC) was added to a solution of the vector (9.75 µL, 40 ng/µL) and labelled with a Eurofin barcode before sending out for sequencing.

[10] http://www.insilico.uni-duesseldorf.de/Lig_Input.htmL

VII.1.2 Preparation of plasmids for expression of Affimers against VTB_5

VII.1.2.a Preparation of the insert

Forward primer (pET11_Adhiron_forward: TTC TGG CGT TTT CTG CGT CTG C) and reverse primer (pDHis_cys_C_term_rev: TTA CTA ATG CGG CCG CAC AAG CGT CAC) were used to clone the Affimer fragment by the PCR reaction from the Affimer phagemid. A standard PCR reaction was set up in 50 µL, with the composition and reaction conditions listed below in Table VII-4 and Table VII-5.

Table VII-4 Composition of the PCR reaction.

Ingredient	**Volume**
Water	make up to 50 µL volume
10 * PCR buffer (containing 10 mM $MgSO_4$)	5 µL
Affimer phagemids	50 ng
Forward primer	0.5 µL, 100 µM
Reverse primer	0.5 µL, 100 µM
dNTPs	0.5 µL (25mM of dATP dCTP dGTP dTTP each)
Phusion High Fidelity (PWO) polymerase	0.2 µL

Table VII-5 Conditions for PCR reaction.

Time	**Temperature (°C)**	**Aim**	**Repeat**
30s	95	Denaturation	Repeat 30 times
30 s	54	Annealing	
50 s	72	Extension	
100 s	72	Final extension	Final and storage
-	4	Storage temperature	

DpnI enzyme (1 µL) was added to a 50 µL PCR reaction to digest the phagemid DNA for 1.5 h, 37 °C, before a QIAquick PCR purification kit (QIAGEN) was used for purification of the amplified DNA. After eluting the digested PCR product with 35 µL EB buffer, the concentration was around 70 ng/µL. Each purified PCR sample was diluted with nuclease-

free water (14 µL) and Cutsmart buffer (5 µL) before adding NotI-HF (0.5 µL) and NheI-HF (0.5 µL). The digestion was performed overnight at 25 °C and an extra 2h at 37°C for a thorough cut.

VII.1.2.b Preparation of the vector

pET11a Affimer 3A2 vector (4500 ng) was digested overnight at 37 °C with NheI-HF (0.75 µL), and NotI-HF (0.75 µL) with CutSmart buffer (5 µL) in a total volume of 50 µL. The following day, CIP enzyme (1 µL) was added, and incubation continued at 37 °C for 1h. The digested plasmid was loaded on a 1% agarose gel with SYBR Safe and run at 90 V for 30 minutes to separate DNA pieces of difference sizes. The gel was observed under blue light illumination and the band around 5 kb was cut out with a razor blade and temporarily stored in a 1.5 mL Eppendorf tube. The agarose gel can be imaged using the GelDoc only after the required strand has been removed, to avoid the mutagenicity induced by the UV light source. The digested plasmid was extracted using the QIAquick Gel Extraction kit (QIAGEN).

VII.2 Site-directed mutagenesis

VII.2.1QuickChange site-directed mutagenesis

QuickChange is efficient to mutate 1-3 neighbouring bases simultaneously. The designed primers should complement the opposite strands of the targeted dsDNA and incorporate the mutations within it. After the PCR cycling, the product was treated with DpnI endonuclease, which targets the (hemi-)methylated DNA to eliminate the parental vector. E. coli XL10 cells were transformed with the newly synthesized DNA to prepare DMA for sequencing. No PCR purification step is needed between PCR and transformation. The primers used in Quickchange must be designed carefully based on the desired mutation, and the web server[11] for designing the Quickchange primer can also be used. The standard PCR setting, as described in section VII.1.2.a can be further applied when the primers are ordered.

VII.2.2 Q5 mutagenesis

The Q5 Site-Directed Mutagenesis is superior to Quickchange, especially for long-range deletions (>6 nts), insertions (>18 nts), and substitutions (>4 nts). In this project, Q5 site-directed mutagenesis Kit (#E0554S) was used according to manufacturer's instruction. Rather than designing the primer to "cover" the area of mutation as in the case of QuickChange, the primer design for Q5 mutagenesis involves the complementary region of

[11] https://www.agilent.com/store/primerDesignProgram.jsp

plasmid and an overhang region which will be ligated using either home-made or commercial Kinase, Ligase, and Dpn1 (KLD) treatment. Below is an example using Q5 mutagenesis to insert five glycines between the TEV cutting site and the *procoil* for MBP-AP2.1 and MBP-AP2.2. The primers used for constructing MBP-AP2.1 and MBP-AP2.2 are listed below, and the overhang piece for insertion of five insertion is coloured red.

Table VII-6 Example of primer pairs for performing Q5 mutagenesis.

		Sequence
MBP-AP2.1	Forward primer	TGGCGGTCAGCTGGAAAAAGAACTGGCGCAG
	Reverse primer	CCGCCACCACCTACACTCTGAAAGTACAGGTTTTCG
MBP-AP2.2	Forward primer	TGGCGGTCAGCTGGAGAAAGAACTGGCTCAG
	Reverse primer	CCGCCACCACCTACACTCTGAAAGTACAGGTTTTCG

A standard PCR reaction was carried out to amplify the target plasmid with the insertion. After that, a home-made KLD treatment was performed at 25 °C, overnight using a mixture of 5 µL nuclease-free water, 1 µL of each the following components: PCR product, 10*T4 DNA ligase buffer, T4 Polynucleotide Kinase (PNK), T4 DNA ligase, and Dpn I enzyme. After incubation, the DNA and protein mixture was transformed to XL10 (Gold) for DNA extraction.

VII.3 DNA electrophoresis

1.0% w/v agarose gels were made to separate plasmids and digested gene fragments. Agarose (300 mg) was mixed with TAE (Tris-acetate-EDTA) buffer (30 mL) and heated in the microwave for 1.5 min to dissolve. The mixture was cooled down to r.t. for 10 minutes. SYBR Safe DNA gel stain (3 µL; ThermoFisher) was added into the mixture for visualizing the DNA fragment under blue light and UV excitation. Ethidium bromide could also be used as an alternative for DNA gel stain but is not recommended due to its high mutagenic toxicity. Before loading DNA samples into an agarose gel, 6 X DNA loading buffer (NEB, #B7021S) was added to DNA samples. The gel was electrophoresed at 90 V for 25 min in TAE buffer. Due to the direct DNA damage caused by UV light, the digested pSTAB-1/2 plasmid was first transferred to UV transilluminators for gel isolation using sterilized blade before visualizing under UV light of SYBR Safe mode.

VII.4 Protein expression and purification

VII.4.1Buffers used in this section

HEPES buffer: 50 mM HEPES, 150 mM NaCl, pH=7.5

Amylose elution buffer: 50 mM HEPES, 150 mM NaCl, 20% (wt/vol) glucose, pH=7.5

Nickel elution buffer: 50 mM HEPES, 150 mM NaCl, 250 mM imidazole, pH=7.5

Periplasmic extraction buffer: 50 mM HEPES, 150 mM NaCl, 20% (wt/vol) glucose, pH=7.5

Affimer lysis buffer: 300 mM NaCl, 50 mM HEPES, 10% (wt/vol) glycerol, 0.4% (vol/vol) Triton X-100, 1 mM TCEP-HCl, pH=7.4

Affimer nickel wash Buffer A: 300 mM NaCl, 50 mM HEPES, 10% (wt/vol) glycerol, 25 mM imidazole, 0.3 mM TCEP-HCl, pH=7.4

Affimer nickel elution Buffer B: 300 mM NaCl, 50 mM HEPES, 10% (wt/vol) glycerol, 500 mM imidazole, 0.3 mM TCEP-HCl, pH=7.4

VII.4.2 General tips for protein expression and purification

When a new protein is expressed and purified, it is preferable to pick some sample from the pre-induction cell pellet, post-induction extracellular media, post-induction soluble part, and post-induction inclusion body to get an overall idea of the solubility of the targeted protein. A recommended procedure for 400 mL cell cultures is as follows:

Pre-induction cell pellet: An aliquot (50 µL) of pre-induced cell culture is centrifuged at 13,000 x g, the supernatant is sucked up by blotting with tissue paper and discarded. Then 10 µL 1*SDS loading buffer is added to the cell pellet. The sample is either boiled or kept at room temperature before loading 3 µL onto the SDS-PAGE gel. This should give information of the protein content before the target protein is expressed and check if there is any protein expression leakage.

Post-induction media (Optional): An aliquot of extracellular media (400 µL) is concentrated to 40 µL with an ultrafiltration centrifugal concentrator with appropriate cutoff, and then the concentrated media (10 µL) is added to 6*SDS loading buffer (2 µL). The sample is either boiled or kept at room temperature before loading 10 µL onto the gel.

Soluble fraction: Before harvesting all the cell pellet/extracellular media through centrifugation, transfer 200 µL cell culture into an Eppendorf tube and centrifuge at 13,000 x g to obtain the cell pellet. Add 20 µL 1*Bugbuster® (Merck, #70584) or 1*B-PER reagent (Thermo Scientific, #78243), and mix with a pipette tip to distribute the pellet evenly, before incubating at 25 °C for 30 minutes. The soluble protein fraction can be obtained by centrifuging at 13,000 x g for 10min, of which 15 µL is mixed with 3 µL 6*SDS loading buffer. The sample is either boiled or kept at room temperature before loading 3 µL onto the SDS-PAGE gel.

Inclusion body: The leftover pellet separated from the soluble part is mixed with 15 µL water and 3 µL 6* SDS loading buffer. The sample is either boiled or kept at room temperature before loading 2 µL in SDS-PAGE gel.

VII.4.3 Expression of MBP-procoil-CTA2B5: export to extracellular media

The E. coli C41(DE3) cells harbouring the target plasmid, either from a glycerol cell stock or an agar plate, were inoculated in LB media (5 mL) with ampicillin (5 µL, 100 mg/mL). The starter culture was incubated at 37°C for 20h with shaking at 200 rpm. A 4 mL aliquot of the starter culture was added into sterilized LB media (400 mL) containing ampicillin (100 µg/mL) and shaken at 37 °C, 200 rpm for 2-3h until the OD_{600} reached 0.6-0.8. Following that, IPTG (400 µL of 1.2 mg/mL solution) was added to give a final concentration of 0.5 mM and the temperature was then changed from 37°C to 30°C to induce protein expression for 20 h. The extracellular growth medium was collected by centrifugation at 10,000 x g for 10 minutes in an Avanti JXN-30 centrifuge using a JA10 rotor. The cell pellet was discarded while retaining the extracellular media, which was filtered through 15 mm Whatman® filter paper to get rid of most of cell before being filtered through a 47 mm MF-Millipore® cellulose esters membrane with a pore size of 0.22 µm.

Amylose resin (5 mL) was equilibrated with 5 column volumes (CV) of HEPES buffer. After all the extracellular media had passed through, the column was washed with 10 x CV of HEPES buffer and eluted with 5 x CV amylose elution buffer (HEPES buffer supplied with 20% (wt/vol) glucose). During the washing step, the MBP, MBP-procoil-(s) CTA2, and the MBP-procoil-(s) $CTA2B_5$ remained bound to the resin while the other impurities were washed away. The amylose elution containing MBP derivatives was loaded into a 5 mL nickel column for further purification. Two types of nickel columns were generally used: one is gravity column which the HisPur™ Ni-NTA Resin was packed manually, the second is a 5-mL HisTrap™ HP column which can be used by coupling with an NGC FPLC system

(BioRad). For CTB_5 proteins, including MBP-procoil-$CTA2B_5$, HEPES buffer was used for washing the column before eluting the protein with nickel elution buffer (HEPES buffer containing 250 mM imidazole). Since there are five exposed His13 residues at the GM1-binding face of CTB_5, MBP-procoil-(s)$CTA2B_5$ was bound to the nickel resin and eluted out as the final purified product. The protein was dialyzed into HEPES buffer, either using a dialysis bag (Thermo Scientific SnakeSkin Pleated Dialysis Tubing) or centrifugal concentrator (Amicon.Ultra-15 Centrifugal Filter Units).

VII.4.4 Expression of Affimers: purification from the cell pellet

A starter culture comprising E. coli BL21 (DE3) cells carrying the relevent pET11 Affimer plasmid in LB media (5 mL) with ampicillin (5 µL, 100 mg/mL) was incubated for 20h (200 rpm, 37°C). An aliquot (4 mL) of the pre-culture was added into autoinduction media (400 mL) containing ampicillin (100 µg/mL) and incubated for 2h, 37°C, then the temperature setting in the incubator was changed to 30°C for 20 h induction. The cell pellet was harvested the following day by centrifugation at 10,000 x g for 10 minutes and stored at -80 °C for up to several months before cell lysis.

Depending on the amount of cell pellet one works with, chemical lysis (such as B-PER and Bugbuster as lysis agent) or mechanical disruption (e.g., sonication/cell disruptor) can be utilized. Chemical lysis is quick and gentle and advantageous in obtaining multiple different proteins from separate cell pellets in parallel. However, the reagents are commercial products and costly to use often. On the other hand, individually or combined, sonication and the Constant Systems cell disruptor can also be operated to extract total soluble protein efficiently, which is especially useful when the wet weight of cell pellet is heavier than 5 g. Both chemical lysis and mechanical lysis were used in the project to extract the total soluble protein.

Cell pellet from 400 mL culture media was re-suspended in 30 mL Affimer lysis buffer (300 mM NaCl, 50 mM HEPES pH 7.4, 10% w/v of glycerol, 0.4% v/v Triton X-100, 1 mM TCEP-HCl). The apolar head of Triton X100 can insert into the lipid bilayer and disrupt the membrane integrity; TCEP-HCl serves as a reducing reagent targeting any disulfide bonds between the C-termini of Affimers. Protease inhibitor is not necessary for Affimer purification since the Affimer scaffold is resistant to proteolysis, and it was found that the addition of the cOmplete EDTA-free protease inhibitor tablet can interact with the free cysteine in Affimers (by Tomasz Kaminski, a previous Ph.D. student in the Webb lab). The cell suspension was sonicated on ice using Bandelin Sonopuls, with 50% cycle and 40% power, for 2 minutes

operation, five times with 30 s intervals. To denature protein impurities, the Affimer lysates were then heated at 50 °C for 30 minutes. After that, the lysates were centrifuged at 30,000 x g for 30 minutes, and the supernatant was collected and filtered for the next-step: nickel purification.

5 mL of Histrap FF Crude resin was pre-equilibrated with Affimer nickel wash Buffer A (300 mM NaCl, 50 mM HEPES pH 7.4, 10% w/v of glycerol, 25 mM imidazole, 0.3 mM TCEP HCl). After the supernatant was passed through the column, 10 column volumes of buffer A was passed through the column to wash out the impurities. Affimer Buffer B (300 mM NaCl, 50 mM HEPES pH 7.4, 10% w/v of glycerol, 500 mM imidazole, 0.5 mM TCEP HCl) was used to elute to Affimer. The buffer of the eluted Affimer was changed to either Affimer storage buffer (300 mM NaCl, 50 mM HEPES, pH=7.4, 10% w/v of glycerol, 0.5 mM TCEP HCl) without further concentration, or cysteine labelling buffer (300 mM NaCl, 50 mM HEPES, pH=7.22) and concentrated to 150 µM. The concentration was performed with Centrifuge Filter Unit (Amicon.Ultra-15). For the Affimer exchanged to the cysteine labelling buffer, the removal of TCEP are prerequisites for the success of maleimide-thiol reaction to add a biotin group at the C-terminus.

VII.4.5 Expression of streptavidin mutants: refolding from inclusion body

A starter culture comprising E. coli BL21-CodonPlus (DE3)-RIPL cells transformed with either the Alive streptavidin (SA) pET21 vector, or Aspartate loop-Dead streptavidin (Dd) pET21 vector[93](purchased from Addegne, #46368) in LB media (5 mL) with ampicillin (5 µL, 100 mg/mL) was incubated for 20h at 200 rpm, 37°C. The SA plasmid was derived from the Se pET21 plasmid (purchased from Addgene, #46367) and a Q5 mutagenesis was used to delete the glutamate tag. An aliquot (4 mL) of starter culture was added into sterilized LB media (400 mL) containing 0.8% (wt/vol) glucose and ampicillin (100 µg/mL) and incubated at 37 °C, 200 rpm until OD_{600} reached 0.8-1.0. IPTG was added to a give a final concentration of 100 µg/mL to induce protein expression, and incubation was continued at 37 °C for 4 h. The cell culture was then centrifuged at 10,000 x g for 10 minutes in an Avanti JXN-30 centrifuge using a JA10 rotor; the cell pellet was isolated and resuspended in Tris buffer (15 mL, 100 mM Tris-HCl, pH=8) and stored in a 50 mL falcon tube at -80 °C. The cell pellet was thawed in a room temperature water bath, and lysozyme (15 µL, 100 mg/mL in 50% glycerol stock) was added, and the mixture was incubated at room temperature on a rocker for 45 min. For SA, the amount of lysozyme was doubled, and one tablet of cOmplete protease inhibitor tablet was added. Triton X-100 (750 µl, 10% solution) was added to lyse the cell at room temperature for 10 min. After another round of freeze-thaw, EDTA was

added (final concentration is 1 mM) and incubated at r.t. for 10 min before the sample was frozen again. The lysis mixture was thawed again in the 25 °C water bath and add water (15 mL) was added to reduce viscosity. Sonication was then performed using a Bandelin Sonopuls (50% cycle, 40% power, 4 repeats of 40 sec pulses separated by 60 s intervals). The mixture was centrifuged at 30,000 x g for 30 minutes at 4 °C and the cell pellet was collected for the next washing steps.

A pipette boy was used to add 15 mL PBS buffer and evenly distribute the cell pellet in suspension, which was then centrifuged at 30,000 x g for 20 min at 4 °C. The washing step was repeated 4 times. Sufficient washing is a prerequisite for efficiently refolding the protein with high purity. The protein was dissolved in 15 mL of 6 M guanidinium hydrochloride, and keep mixing with the pipette boy until the pellet is fully resuspended. The sample was centrifuged again at 30,000 x g and the supernatant was kept – this is the denatured streptavidin monomer. Equal molar amounts of SA and Dd were mixed in a total volume of 10 mL. The mixture was added dropwise to ice-cold PBS (200 mL) with a final guanidinium concentration less than or equal to 0.3 M; make sure the stirring is at a sufficient speed so that the guanidinium can be diluted rapidly, but minimizing the amount of foam generated – the intermediate concentration of guanidinium can promote protein aggregation. After all the guanidinium was added, the PBS buffer solution was moved to the cold room to stir at a lower speed overnight.

The following day, ammonium sulfate (120 g for every 200 mL of PBS) was added over 5 minutes, and left to stir for at least 2h at 4 °C. The solution was centrifuged at 30,000 x g for 30 min using a JA 25.50 rotor, and the supernatant was discarded, while making sure there was no PBS buffer left in the pellet. The pellet was dissolved in Tris buffer (10 mL, 20 mM, pH = 8), and centrifuged again at 30,000 x g for 30 min to remove any undissolved debris. Depending on what is needed, an iminobiotin column can be used to purify the refolded streptavidin from the other impurities. A 3 mL iminobiotin column was equilibrated with 10 CV iminobiotin binding buffer (50 mM boric acid, 500 mM NaCl pH=10), then the refolded streptavidin was loaded onto the column, washed with 10 CV iminobiotin binding buffer, and finally eluted with 6 CV iminobiotin elution buffer (20 mM NaH_2PO_4 pH=2). The pH of the sample was adjusted by adding 0.5 CV Tris buffer (1 M, pH=8). The sample was dialysed into 10 mM Tris buffer (pH= 8). The column was washed with 0.1 M NaOH and stored in 20% aqueous EtOH.

After refolding the streptavidin, the reassembled mixture of SA/Dd was loaded on a 5 mL HiTrap® Q Fast Flow resin (Cytiva) at a flow rate of 4 mL/min, using 10 mM Tris (pH=8.0) as

the washing buffer. Different tetramers were isolated using a NaCl concentration gradient. If the streptavidin was not previously purified from the iminobiotin column, a 5% 1M NaCl wash for 5 CV can eliminate SA_4 and most impurities. Then a gradual increase from 5% NaCl to 15% NaCl spanning 20 CV can elute out SA_3Dd_1. The 15% NaCl concentration should be kept for another 12 CV to obtain a mixture of two types of trans-SA_2Dd_2 in a single peak. Finally, increasing the NaCl concentration from 15% to 30% over 48 CV provides the cis-SA_2Dd_2, SA_1Dd_3, and Dd_4, respectively. The elution was collected in 5 mL fractions.

VII.5 Sequence-specific protein digestion

VII.5.1 TEV cutting site

TEV protease can recognize the ENLYFQG/S sequence and cut between Q and G/S residue using a catalytic triad involving its cysteine residue[150]. To maintain the free thiol group, 1 mM TCEP is usually included in the MBP-His-TEV stock. Since a difference in cutting efficiency for the different complexes studied in this project was found to be a result of steric hindrance, a preliminary test is recommended before large-scale digestion. At the start, 5 mol% or 10 mol% of MBP-His-TEV protease (in 100 µM stock) was incubated with 40-50 µM MBP-procoil-CTA2B_5 in HEPES buffer at either 25 °C or 37 °C for 1-4 h, or 4 °C overnight. After the cleavage test, the samples were then loaded onto a SDS-PAGE gel to check for digestion.

VII.5.2 Factor Xa cutting site

Factor Xa protease can recognize the Ile-Glu/Asp-Gly-Arg sequence and cleave after the arginine residue. To achieve a clean cut, the concentration of the interested protein should be at least 0.25 mg/mL, and a series of conditions should be tested to avoid nonspecific proteolysis that can arise from excessive concentration of Factor Xa protease and overlong incubation. To start with, an aliquot of MBP-AP2.1 (5 µL, 2.4 mg/mL) was mixed with Factor Xa protease (1.2 µL, 0.6 µL, and 0.3 µL, 1 mg/mL) and allowed to incubate at 25 °C overnight. After the cleavage test, the samples were then loaded onto a SDS-PAGE gel to check for digestion.

VII.6 Chemical ligation

VII.6.1 Oxime ligation

To a solution of the protein (40 µM, 400 µL) in HEPES buffer, was first added 10 molar equivalents of methionine (4 µL of 50 mM stock) and then 5 molar equivalents of sodium

periodate (4 µL of 25 mM stock), and the solution was kept for 8 minutes at 25 °C. For the N-terminus serine group, the ESMS signal should decrease by 14 Da, indicating that the diol hydrate of the aldehyde has formed. If the signal decreases by 32 Da which would be expected for the aldehyde, then a side reaction has occurred involving cyclisation of aldehyde onto peptide backbone – the free aldehyde will not exist in aqueous solution. After ensuring oxidation had taken place, a GD-25 desalting column was used to remove excess periodate. Since the elution volume is 1 mL, the protein was diluted 2.5 times during the desalting step and reached a final concentration of 20 µM. 10 µL Aniline (neat) and alkoxyamine-PEG4-Biotin (6 µL of 250 mM stock) were added to the desalted 1 mL, 20 µM protein sample. The protein mixture was incubated overnight at 25 °C, and ESMS showed a shift of molecular weight of 386 Da indicating the successful ligation of the PEG_4-biotin group at the N-terminus.

VII.6.2 NHS AF488/AF647 labeling

Alexa Fluor 488/647 succinimidyl ester was used to label proteins by reaction with the primary amines located on either the N-terminus or lysine residues. 100 µL, 50 µM CTB_5 was supplemented with sodium bicarbonate solution (10 µL of a 1 M stock) to raise the pH of the reaction mixture to 8.3 to enable a more targeted labeling at the terminal amine rather than lysine epsilon-amino group. Five molar equivalents of AF488-NHS (dissolved in water-free DMSO, 1 mg/mL stock) was added to the protein and incubated at 25 °C for 1h in an aluminum foil-wrapped tube to exclude light. Excess dye was removed using a GD-10 column, and the final degree of labeling was calculated according to Equation 3 and Equation 4.

$$Protein\ concentration\ (mol/L) = \frac{[A_{280} - (A_{495} \cdot CF)] \cdot dilution\ factor}{\varepsilon}$$

Equation 3. Determining the protein concentration (unit: mol/L) after the protein has been fluorescently labelled with Alexa Fluor 488 fluorescence. ε stands for the extinction coefficient of protein ($M^{-1}cm^{-1}$), that is 60905 M^{-1} cm^{-1} for biotin-CTA2B_5. A_{495} is the absorbance of protein solution at the wavelength equals to 495 nm. The correction factor, or CF, is used to account for the impact of dye on protein absorbance under 280 nm. The CF from NHS ester Alexa fluor 488 is 0.11. Only when the protein solution is diluted during the measurement process is a dilution factor required.

$$Degree\ of\ labelling = \frac{A_{495}}{\varepsilon_{label} \cdot Protein\ concecntration\ (mol/L)}$$

Equation 4. Calculating the degree of labelling (DOL) of protein which has been fluorescently labelled with Alexa Fluor 488. ε_{label} is equal to 73,000 M^{-1} cm^{-1}, as the extinction coefficient of the NHS ester - Alexa fluor 488 label.

VII.6.3 Biotinylation of cysteine with biotin-PEG$_2$-maleimide

The maleimide group can react efficiently with the free sulfhydryl group of Affimers at pH 6.5-7.5 and form thioether bonds. It is essential to control the pH below 7.5 to avoid reaction with primary amines and maleimide group hydrolysis. The removal of TCEP in the buffer is also necessary just before the biotinylation step to prevent reaction with the maleimide group. Twenty molar equivalents of EZ-Link Maleimide-PEG$_2$-Biotin (ThermoFisher; stock solution of 20 mM in DMSO) was added to Affimer (150 µM, 500 µL) in cysteine labeling buffer (300 mM NaCl, 50 mM HEPES, pH=7.22), and incubated at 25 °C, dark for 2h. An increase of 496 Da in ESMS indicated successful ligation. It was important to make sure the majority of the Affimer had been modified since the biotin-Affimer would be used for quantitative measurement through ALBA assays to compare the binding affinity of one Affimer with one another. If less than 95% biotinylation had occurred, a second biotinylation was done after reducing any disulfide groups by adding and removing TCEP. Unreacted maleimide was removed using a GD-10 desalting column.

VII.6.4 Dimerization of Affimer with APN-Maleimide

The dimerization mechanism has two steps governed by the different reaction speeds of the sulfhydryl group to the APN and the maleimide groups. The reaction between cysteine and APN-maleimide (Merck) was performed as described for the maleimide in section I.6.3. After confirming maleimide-thiol interaction had taken place, a centrifugal concentrator was used to remove excess APN-maleimide and the Affimer-S-S-Maleimide-APN solution was concentrated to 385 µM. If the protein was not immediately used, it should be quickly flash-frozen and stored at -80 °C in case the protein precipitation occurs. For conjugation to the second Affimer, Affimer-S-S-Maleimide-APN was mixed with 1.25 molar equivalent of helix-3C6 Affimer (385 µM, 250 µL) and incubated at 25 °C in the dark for a minimum of 18h. Wrapping in foil is not

Immediately before purifying the dimerized protein through SEC, TCEP-HCl was added to give a final concentration of 1 mM to reduce any disulfide bonds, ensuring the dimerization peak from SEC indicated the dimer formed through APN-Maleimide. The protein was concentrated to 110 µL and injected into an Increase Superdex 75 10/300 column through a 100 µL sample loop, and eluted with Affimer storage buffer (50 mM HEPES pH=7.22, 200 mM NaCl, 0.5 mM TCEP).

VII.7 SDS-PAGE and Native PAGE

Based on the need, either denaturing SDS-PAGE or Native PAGE gels can be prepared. The volumes necessary for making different percentages of SDS-PAGE can be calculated through an online website[12]. For making native PAGE, the recipe for resolving gel and stacking gel is in Table VII-7 and Table VII-8. The composition of gel running buffer and sample buffer for either SDS-PAGE or Native PAGE are in

Table VII-9. SDS contained in SDS-PAGE can denature proteins and coated linearized protein with a different number of negative charges, and heating can speed up the denature process. β-Mercaptoethanol is used to reduce disulfide bonds in proteins and can alternatively be replaced with either DTT or TCEP. In Native PAGE, protein samples are prepared in a non-reducing and non-denaturing buffer. Hence the migration is dependent on the tertiary structure and mass-to-charge ratio of the folded protein.

Table VII-7 Recipes for making Native PAGE stacking gel.

5 mL Native PAGE stacking gel	
0.375 M Tris-HCl pH=8.8	4.275 mL
Acrylamide/Bis-acrylamide (30%/0.8% w/v)	0.67 mL
*10% (w/v) ammonium persulfate (AP)	0.05 mL
*TEMED	5 µl

Table VII-8 Recipes for making the different percentages of native PAGE resolving gel.

10 mL Native PAGE resolving gel					
Acylamide percentage	6%	8%	10%	12%	15%
Acrylamide/Bis-acrylamide (30%/0.8% w/v)	2 mL	2.6 mL	3.4 mL	4 mL	5 mL
0.375M Tris-HCl (pH=8.8)	7.89 mL	7.29 mL	6.49 mL	5.89 mL	4.89 mL

[12] https://www.cytographica.com/lab/acryl2.htmL

10% (w/v) ammonium persulfate (AP)	100 µL	100 µL	100 µL	100 µL	100 µL
TEMED	10 µL	10 µL	10 µL	10 µL	10 µL

Table VII-9 Comparing recipes of SDS-PAGE and Native PAGE for sample buffer and running buffer.

	For Native PAGE	For SDS-PAGE
Sample buffer	(2x) 62.5 mM Tris-HCl, pH 6.8 25% glycerol 1% Bromophenol Blue	(2 x, (Morris formulation) 125 mM Tris-HCl, pH 6.8 5% SDS 20% glycerol 10% β-mercaptoethanol 0.004% Bromophenol Blue
Running buffer	25 mM Tris 192 mM glycine	25 mM Tris 192 mM glycine 3.5 mM SDS

VII.8 Mass spectrometry

Mass spectrometry was performed on a Bruker Daltonics micrOTOF Mass Spectrometer with electrospray ionization (ESMS). The equivalent of 40 pmol protein was injected onto the machine, and the deconvolution of multiple charge states was analyzed using the Bruker Compass Data Analysis software.

VII.9 Analytical size exclusion chromatography (SEC)

Analytical Size Exclusion Chromatography (SEC) was performed using the Biorad FPLC system at 4°C. All protein samples applied were filtered through a 0.2 µm Sartorius Minisart filter. A minimum amount of 50 µL of 30 µM 100 kDa protein sample was injected onto either a HiLoadTM 10/300 SuperdexTM 200 column (GE Healthcare) or 16/60 SD 200 column, or a 16/60 SD 75 column, based on the molecular weight and the total volume of targeted protein. Elution was usually operated with HEPES buffer at a flow rate of 1 mL/min. The elution fractions were detected via UV/vis absorbance at 280 nm, and 495 nm or 647 nm absorbance if the protein was fluorescently labelled.

VII.10 Size exclusion chromatography with inline multi-angle light scattering (SEC-MALS)

MBP-AP2(aa) or MBP-PP(b)-v1 (35 µL, 3 mg/mL) was injected into a 5/150 Superose 6 column equilibrated with HEPES buffer (50 mM HEPES, 200 mM NaCl, pH=7.5) at a flow rate of 0.2 mL/min. Data were obtained with a DAWN 8+ multi-angle light scattering (LS)

detector, an Optilab T-rEX differential refractive index (RI) detector, and a UV absorbance detector (Wyatt Technology Corporation) and analyzed using Astra6.2 (Wyatt Technology).

VII.11 Mass photometry (MP)

Mass photometry (MP) is a label- and immobilization-free technique; it calculates the molecular weight of protein based on light scattering and measures the species population by counting each single molecular landing event. The Refeyn One^{MP} was used to detect molecules larger than 40 kDa at low concentration from 100 pM to 100 nM.

VII.11.1 Preparation of coverslips and protein samples.

First, 24 mm x 50 mm high precision No. 1.5 coverslips (Thorlabs CG15KH) were rinsed with water, isopropanol, and water. The coverslips were dried with nitrogen gas and ensure the working region in the coverslip is cleaned. A silicon well was attached to the cleaned coverslips, while being careful to press the edge of the silicon to get rid of any air and ensure a complete contact. All buffers used in the MP analysis, must be filtered into a water-rinsed bottle to remove particulates. The protein sample was dluted in filter buffer to 200 nM monomer concentration in a Protein LoBind® Eppendorf tube and incubated for 20 minutes at room temperature to allow the protein-protein interaction to reach equilibrium.

VII.11.2 Acquisition and analysis of data

The acquisition and analysis of data require Refeyn One^{MP}, Refeyn Acquire^{MP} (Version 2.5.1), and Refeyn Discover^{MP} (Version 2.5.0), respectively. A drop of immersion oil (Thorlab MOIL-30) was applied on the top of the objective and the coverslip was placed with the center of the silicon wall aligned. Filtered HEPES buffer (20 µL) was added into one silicon well and the instrument was focused. The signal value at the maximum sharpness position should be below 0.5%, and the color bar for all parameters (sharpness, brightness, saturation, signal, and motion should be in the blue zone). If not, the buffer needs to be filtered again, or there are some impurities in the coverslip, or a round of parameter optimization needs to be undertaken. For each measurement, HEPES buffer (16 µL) was first applied to Find Focus, and the protein sample (4 µL, 200 nM) was then loaded and mixed before the detection began. Even though it is possible to load the protein sample directly on the coverslips without adding protein-free buffer, the result was not very consistent and hence not recommended. Based on requirement, the final protein concentration could be varied from 20 nM to 100 nM. The recorded video was processed to count the landing events and convert it to molecular mass based on the calibration curve

measured with different concentrations of BSA. The mass distribution was fitted with a Gaussian function and a histogram was generated using the MP data processing software. (*Check this protocol [151] for more detailed guidance.*)

VII.11.3 Atom Force Microscopy

5 mm-diameter mica discs glued to a 10 mm Teflon support was fixed on a magnetic sample holder. For each experiment the mica was freshly cleaved with Scotch tape and imaged in 30–50 µL of adsorption buffer to check the cleavage quality; 3 µl of protein solution (100 nM dimer concentration) was then injected into the adsorption buffer drop on the mica surface. After 2 h the sample was carefully rinsed with recording HEPES buffer. Imaging was performed with a commercial Nanoscope III AFM (from Digital Instruments, Santa Barbara, CA) equipped with a 120 µm scanner (J-scanner) and oxide-sharpened Si3N4 cantilevers with a length of 120 µm (k = 0.1 N/m) (from Olympus Ltd, Tokyo, Japan). The AFM was operated in contact mode applying constantly minimal forces (<0.2 nN) at a scan frequency of 4–6 Hz. The instrument was calibrated using layered crystals of MoTe2.

VII.12 Gaint Unilamellar Vesicle (GUV) experiments

VII.12.1Preparation of the lipid mix

Two types of lipid mix were used in the preliminary GUV experiment. One was GUV-Atto647 (DOPC : cholesterol : GM1 glycosphingolipid : Atto 647-DOPE = 64.7 : 30 : 5 : 0.3 mol%) and GUV-no dye (DOPC : cholesterol : GM1 glycosphingolipid = 65 : 30 : 5 mol%). The original lipid stock was either dissolved in chloroform or a methanol-chloroform mixture. A chloroform-cleaned Hamilton syringe was used to aliquot the calculated amount of each component to reach the defined molarity percentage, and solutions were mixed under Argon gas in a glass vial to avoid oxidation of the lipid. The lipid mix was stored at -20 °C.

VII.12.2Preparation of GUV by electroformation

The conductive face of two indium-tin oxides (ITO) slides were cleaned with isopropanol using Precision Wipes (KIMTECH). A circle area was marked at the non-conductive side indicating the lipid spray region. Hamilton syringe was thoroughly cleaned 20 times using the chloroform stored in a cleaned Amber glass vial. An aliquot of lipid mix (7.5 µL) was taken from the stock solution and quickly sprayed in the selected circle; This step was repeated to add 15 µL lipid mix in total on the slide. The ITO slides were placed under vacuum at 15 mb pressure for at least two hours. The Hamilton syringe was cleaned in a new bottle of chloroform after use. A thin strand of sigillum wax was prepared by rolling it on a clean

surface and laying it as a ring around the lipid-coated area on the slide. The electroformation chamber was assembled by carefully pressing the two slides and squeezing the wax to leave only a pore on one side and tape was used to fix at the ends of the chamber formed by ITO slides. The chamber was filled with 0.3 M sucrose and the sigillum wax ring was carefully closed without trapping any air. The crocodile clip was attached from the TG315 Function generator cables to the chips and the preset parameters were used to generate GUVs for 2.5 h. (*This book chapter [152] demonstrates the process in more detail with pictures and graphs.*)

VII.12.3Preparation of the "Open" microscopy chamber.

The microscopy observation chamber was prepared by coating it with 200 µL, 3 mg/mL beta-casein for 30 minutes to prevent vesicles from exploding. The casein was then washed away and rinsed with Ultrapure water two times and finally rinsed with 200 µL PBS buffer. After the recording, the chambers were washed with Brand™ Mucasol™ universal detergent before cleaning with Ultrapure water.

VII.12.4 Confocal microscopy and image processing

After both the microscopy chamber and GUVs were prepared, 10-120 µL GUV (final lipid concentration is around 50-100 nM), protein and PBS buffer (314 mOsm L^{-1}) were added to the chamber and reach a 200 µL, 400 nM CTB_5 dimer concentration. After 5 mins, when the GUVs had sunk to the bottom of the chamber, confocal microscopy (Nikon Eclipse Ti-E with A1R confocal laser scanner, 60× oil objective, N.A. 1.49) was used to record for 2h duration time with 20s interval. Image acquisition was performed with NIS-Elements (Nikon) and further processed with Fiji software.

VII.13 Flow cytometry

For flow cytometry analysis, 1×10^5 cells were counted and transferred to a U-bottom 96 well plate (Sarstedt AG & Co. KG, Numbrecht, Germany). To quantify protein binding to cell surface receptors, cells were incubated with increasing concentrations (0.3 nM, 0.5 nM, 1 nM, 2.5 nM, 5 nM, 10 nM, 15 nM, 20 nM, and 40 nM) of fluorescently labeled protein for 30 minutes at 4 °C and protected from light. Samples were compared to PBS-treated cells as the negative control. Subsequently, cells were centrifuged at 1600 x g for 3 minutes at 4 °C and washed twice with FACS buffer (PBS (-/-) supplemented with 3% FCS v/v). After the last washing, the cells were re-suspended with FACS buffer and transferred to FACS tubes (Kisker Biotech GmbH Co. KG, Steinfurt, Germany) on ice and protected from light. The

fluorescence intensity of treated cells was monitored using a FACS Gallios instrument (Beckman Coulter Inc., Brea, CA, USA) and analyzed using FlowJo V.10.5.3. Curve fitting was performed using the DoseResp model in OriginPro to calculate the EC_{50}.

VII.14 Lipid mixing assay

VII.14.1 Preparation Large Unilamellar Vesicles (LUV) by extrusion.

The lipid mixing assay was performed based on the measurement of Förster Resonance Energy Transfer (FRET) between 7-nitro-2,1,3-benzoxadiazole (NBD) and Rhodamine. Three types of 800 nm Large Unilamellar Vesicles (LUV) were made by extrusion (Table VII-10), with a total lipid concentration as 3.75 mM. Composition 1 was FRET-LUV (DOPC : DOPE : cholesterol : GM1 glycosphingolipid : NBD-DOPE : Rhodamine-DOPE = 47.25 : 23.63 : 23.63 : 5 : 0.25 : 0.25); Composition 2 was blank-LUV (DOPC : DOPE : cholesterol : GM1 glycosphingolipid = 47.5 : 23.75 : 23.75 : 5); Composition 3 was full fluorescence-LUV (DOPC : DOPE : cholesterol : GM1 glycosphingolipid : NBD-DOPE= 47.38 : 23.69 : 23.69 : 5 : 0.25). The solvent for stock solutions was usually chloroform, but GM1 glycosphingolipid was dissolved in chloroform:methanol = 2:1.

Table VII-10 The composition and volume for each type of LUV.

Composition of FRET-LUV			
Components	Stock concentration (mg/mL)	Molar ratio (%)	Taken volume (μL)
DOPC	25.00	47.25	74.29
DOPE	25.00	23.63	35.16
Cholesterol	25.00	23.63	18.29
GM1	5.00	5.00	78.10
NBD-PE	1.00	0.25	10.90
Rh-PE	1.00	0.25	16.27

Composition of blank-LUV			
Components	Stock concentration (mg/mL)	Molar ratio (%)	Taken volume (μL)
DOPC	25.00	47.50	74.68

DOPE	25.00	23.75	35.34
Cholesterol	25.00	23.75	18.38
GM1	5.00	5.00	78.10

Composition of full fluorescence-LUV			
Components	Stock concentration (mg/mL)	Molar ratio (%)	Taken volume (µL)
DOPC	25.00	47.38	14.90
DOPE	25.00	23.69	7.05
Cholesterol	25.00	23.69	3.67
GM1	5.00	5.00	15.62
NBD-PE	1.00	0.25	2.18

After preparing the lipid mixture in a glass vial according to the composition above, the chloroform and methanol were evaporated under a stream of nitrogen gas and placed under vacuum to remove the residual solvent overnight. The following day, 1 mL HEPES buffer was added into the glass vial to rehydrate the lipid mix for 1h before distributing the lipid more evenly by vortexing. The vial lid was replaced with a layer of parafilm with an open pore on top. The solution was subjected to freeze-thaw using liquid nitrogen at least 5 times to alleviate the formation of multilamellar liposomes. The extruder was assembled with a 800 nm Whatman Nuclepore Track-Etch membrane and the liposomes were passed through the extruder 11 times.

VII.14.2 384-Well plate assay with different lipid-protein ratios

For all the fluorescent measurements, a 384-well plate (Small Volume™, CELLSTAR®, Greiner bio-one) was used under r.t.. NBD can be excited at 460 nm and the emission at 538 nm could be further utilized by Rhodamine as excitation source which emits at 590 nm. Hence the fluorescence ratio between 530 nm and 590 nm is the evaluation value to determine the lipid mixing rate. The FRET-LUV and blank-LUV were mixed in 1:4 ratio to achieve a final lipid concentration of 400 µM, and then 5 µL of this liposome solution was added to the 384-well plate. The protein samples were prepared in Protein LoBind® Eppendorf tube (usually 6 µM CTB_5 concentration, 20 µL) and 10 µL was added per well to the liposomes. The final concentration of lipid and CTB_5 protein were 133 µM and 4 µM, respectively. The remaining 10 µL protein sample was diluted with an equal amount of

buffer, mixed by vortex and centrifuged at 13,000 rpm for 10 seconds, before adding 10 µL of diluted protein to the second lane of liposomes. The dilution process was repeated to finally achieve a two-fold dilution series to test the protein concentration effect on lipid mixing efficiency. Triton-x100 was added to the test liposome mixture (FRET-LUV : blank-LUV = 1:4) to achieve a final concentration of 0.2% detergent as the control for full fusion. A control experiment to monitor the liposome self-fusion rate used only the test liposomes. Fluorescence measurements were made using a Perkin Elmer EnVision plate reader with the monochromator set at 470 nm for excitation, and emission collected from 500-630 nm with a step of 30 nm. The number of flashes was set to be 2000 to make ensure a stable signal while the time needed for one round of assay was minimized.

VII.14.3 Data analysis

The maximum fluorescence intensity of NBD (I_{NBD} at 530 nm) and Rh (I_{Rh} at 590 nm) were used to calculate the FRET ratio (R) of each sample as $R = I_{Rh} / I_{NBD}$. The FRET ratios for full fusion and self-fusion, $R_{full\_fusion\_control}$ and R_{self_fusion}, were obtained using the values from the wells containing test liposome destroyed by 0.2% Triton, and test liposome only. Three repeats for each sample were performed in separated plates, and the $R_{full\_fusion\_control}$ and R_{self_fusion} values were averaged within the time scale when the liposome self-fusion is minor (usually within 5h).

The maximum fluorescence intensities of NBD (I_{NBD} at 530 nm) and Rh (I_{Rh} at 590 nm) were used to calculate the FRET ratio (R) for each sample, $R = I_{Rh} / I_{NBD}$. The timepoints up to 5h were used for calculation during which liposome self-fusion was minimal. The percentage of lipid mixing was then calculated by normalising the FRET ratios of each sample (R_n, R value at incubation time n min) between (R_0, R value at incubation time 0 min) and $R_{full\_fusion\_control}$:

$$\%\ Lipid\ mixing = (R_n - R_0) \times 100 / R_{full\_fusion\_control} - R_0$$

The percentage of lipid mixing for all protein samples, including the self-fusion control, were calculated and plotted against the incubation time. A global fit to the straight line was performed in OriginPro with the slope of the fitted curve shared among three repeats while intercept is separated – the slope was extracted as the lipid mixing rate and plotted against different concentrations of proteins.

VII.15 Quartz crystal balance with dissipation monitoring (QCM-D)

A Q-Sense E4 system (Biolin Scientific) with four independent flow modules was used to conduct QCM-D measurements on silica-coated sensors (QSX303; Biolin Scientific, Västra

Frölunda, Sweden). This system was coupled to a syringe pump (Legato; World Precision Instruments, Stevenage, UK) to deliver a fluid flow of 20 $\mu L \cdot min^{-1}$. Six overtones (i = 3, 5, 7, 9, 11 and 13) were used to measure changes in resonance frequency (Δf) and dissipation (ΔD). Except where otherwise noted, results are shown for overtone i = 5. QSoft410 (Version 2.5) was used to monitor the measurement of QCM-D. The workflow involved in the following steps.

VII.15.1 Preparation of the sensor

A silica-coated sensors (QSX303; Biolin Scientific, Västra Frölunda, Sweden) was used to detect the signal of sample interacting with the chips. It consists of a quartz disk and electrodes with gold surface layer on one side. The sensor needs to be carefully treated and always held with clean tweezers at the quartz disc edge without touching the electrodes or gold surface. Before each round of experiment, the sensor was submerged in 2% SDS buffer in a weighing boat for 30 min with the gold layer facing upwards in solution – this step was used to remove any residue lipid component from the last operation on the sensor surface. After that, the sensor was cleaned successively with water, ethanol and ultrapure pure water, and finally dried with nitrogen gas. Make sure the sensor is completely dried before placing them under UV zone cleaner (Procleaner™) to destroy any virus or bacteria for 30 min.

VII.15.2 Preparation of buffer and protein samples

Unless stated otherwise, the buffer used in QCM-D experiment was HBS buffer (10 mM HEPES, 150 mM NaCl, pH 7.4). A sonication bath was used to degas the HBS buffer for at least 30 min before the prepared buffer was used to dilute and prepare the protein solutions. Degassing was done at a temperature of around 30 °C, to avoid the formation of air bubbles when the solution passed through the QCM-D working chamber for which the temperature was 23 °C.

VII.15.3 Set up and running measurements

The core of QCM-D measurement is the flow module. When liquid flows through the inlet, it adopts a snake-shape path to stabilize the working temperature before entering the chamber space where sensor is anchored. The chamber was dried with nitrogen gas and connected with inlet and outlet tubes before mounting the sensor on the surface. Measurement were set up by first finding all resonance for chambers, after which the detection was started from sensing the ambient air in the chamber. At this point, the baseline for frequency (f) and

dissipation (D) should be stable and drift no more than 1 Hz/min. After observing for 5 min, the HBS buffer was passed through at a rate of 200 µL/min. At this point, it is essential to check if D3 signal experience an air-to-buffer exchange signal and finally reaches 162 ppm – the value is buffer dependent and might vary if different buffer system is used. After the initial setting was ready, the running speed was reduced to 20 µL/min and start to measure the samples. After detection, the chamber was cleaned by passing through water, 2% SDS, water, ethanol and water consecutively and finally dried with nitrogen gas.

VII.15.4 Data analysis

Q-Sense Dfind software was used to convert the data of f5 and D5 generated from Q-Sense to an excel sheet, and the further plot was performed by using matplotlib package from Python. Annotation can be added by the same script to indicate the components passing through the sensor.

VII.16 Computational method

VII.16.1MD simulation

All three structures — AP2.1, AP3, and P-PP2.x — were first dissolved in a cubic box of TIP3P water that also contained 150 mM NaCl salt and neutralizing ions. The minimum separation between the protein and the water box's edge was set at 12 Å. The Particle-Mesh Ewald (PME) method was used to calculate electrostatic interactions, with a grid spacing of 0.16 nm and a cutoff of 1.0 nm. The simulation system was set up using the AMBER tleap and parm modules. In the presence of 2.0 kcal/mol/Ang^2 restraints on protein backbone atoms, solvent atoms were minimized for 2500 steps of steepest descent minimization, followed by 2500 steps of conjugate gradient minimization. Following minimization, the system was heated in 25 ps using 1 femtosecond steps from 10.0 K to 303.15 K. The constraints were reduced to 0.1 kcal/mol/Ang^2 for heating and equilibration. The production simulations were run with 4 femtosecond steps, no restraints, and a constant temperature of 303.15 K and pressure of 1 bar. The production run for all the systems was carried out for more than 400 ns until it reached a stable RMSD. The AMBER-ff14SB force field for proteins in Amber (2020) was used for all simulations. The long-range electrostatic interactions were handled by Particle-Mesh Ewald sum, and all hydrogen atom bonds were constrained by SHAKE. The AMBER ptraj module was used to analyze the MD simulation trajectory.

VII.16.2 ISAMBARD

In order to investigate the coiled coil structure based on a given protein sequence, ISAMBARD is used[127]. Backbone populations were created within the range of the radius, pitch, phica, and zshift Crick parameters. The SCWRL4 embedded in ISAMBARD is also used to pack side chains onto the backbone structure[153]. The primary tool used to assess the model's quality was BUFF (Bristol University Docking Engine Force Field)[154], an empirical free capacity field originally developed to predict the free energy of binding between proteins and ligands. The size of the structure space increases exponentially as more parameters are used to describe it. Genetic algorithms, which were implemented in ISAMBARD, were used to optimize the generated coiled coil in this project to enable efficient exploration of the structure space. Three parallel sets of genetic algorithms were applied to each set of coiled coil sequences. Finally, Pymol was used to further examine the three coiled coil structures that the Genetic algorithm had optimized.

VII.16.3 Analysis of helixity change across MD simulation

The HELO (HELix Orientation) algorithm was used to model the helicity of the CTA2 peptide linker to the coiled-coil structural domain[136]. In brief, the α-helix was split into overlapping quadruplets along the linker region, ranging in number from i to i+3. The three direction cosines ($cos\,\zeta$) of each quadruplet were calculated using the x, y, and z coordinates of each residue's C atom, and a 1.5 Å rise was selected as the separation between each residue and the helical axis. The local helicity is calculated using the formula $1 - (cos^2\,\zeta_1 + cos^2\,\zeta_2 + cos^2\,\zeta_3)$, and a non-helical geometry is denoted by a significant deviation of more than 0.2. After the trajectory of the simulated protein is imported into Pymol, a custom Python script is used to analyze the helixty of selected amino acids in different molecular simulation states, and the matplotlib package is used to plot the resulting local helicity.

VII.17 Affimer-Lectin binding assay (ALBA)

The Affimer-Lectin binding assay was used to compare binding affinities of the Affimers selected against VTB_5. FSL-Gb3 (Merck; 50 µl of 2 µg/mL in PBS) was loaded by multi-channel pipette to high binding 96 well microplates (Greiner Bio-one REF 655077) and incubated for 3h at room temperature. After that, the plate was washed with PBS containing 0.2% BSA to remove any unattached Gb3. Blocking buffer (100 µL, 20% v/v 10 × Casein Blocking Buffer (Sigma-Aldrich, B6429) in PBS-T, pH=7.4) or BSA (1%) was added to each well using a Combi Reagent Dispenser to block the plate for 1h at 37 °C. After that, the blocking buffer was removed and the wells were washed three times with 100 µl PBS (0.2%

BSA) using a Combi Reagent Dispenser and standard tube dispensing cassette. Then VTB (50 µL of 1 µM solution) was transferred into wells to incubate with Gb3 for 1h at room temperature before repeating the washing step three times. of The biotinylated Affimer (50 µl of 1 µM solution) was added to interact with VTB_5 for 1h. The wells were washed three times to remove any unbound Affimer. Ultra Streptavidin-Horseradish peroxidase (Life Technologies, N504; 50 µl diluted at 1:1000 in PBS buffer) was added to interact with biotinylated Affimer for 0.5 h at room temperature with lid covered on the plate. The plates were washed three times. A solution of Amplex red (4 µM 10-Acetyl-3,7-dihydroxyphenoxazine, Cayman Chemical Company, 10010469) and hydrogen peroxide (4 µM) was freshly prepared. The Amplex red/peroxide solution (50 µL per well) was added rapidly by Combi Reagent Dispenser, and the plate was instantly loaded into the plate reader for fluorescent detection at 585 nm. The mirror module used in the Perkin Elmer EnVision plate reader was BODYPY TMR, while the excitation and emission filters were BODYPY TMR FP531 and Cys595, respectively.

Chapter VIII Appendix

VIII.1 Plasmid and protein sequence

This section contains parent plasmid encoding MBP-procoil-$CTA2B_5$, MBP-$StxA2+B_5$, and Affimer, with each mutation highlighted with the replaced gene region.

VIII.1.1pTRBAB5-G1S

pTRBAB5-G1S[80] was based on a pSAB2.1 plasmid of which the maltose binding protein (MBP) – CTA2 region and a El Tor strained CTB (T1A) were cloned into pMalp5x between the NdeI and PstI sites. MKIKTGARILALSALTTMMFSASALA serves as leader sequence for MBP, and there is a Factor Xa recognition sequence and a tobacco etch virus (TEV) protease cutting site between the MBP and CTA2 peptide, with serine-glycine-glycine exposed at the N terminus after the TEV protease is operated. Between CTA2 and CTB gene region, there is a ribosome binding site to allow expressing both protein fragment in the same plasmid. MSFKKIIKAFVIMAALVSVQAHA is the LTIIb-signal peptide in front of CTB peptide to allow its transportation to the periplasm.

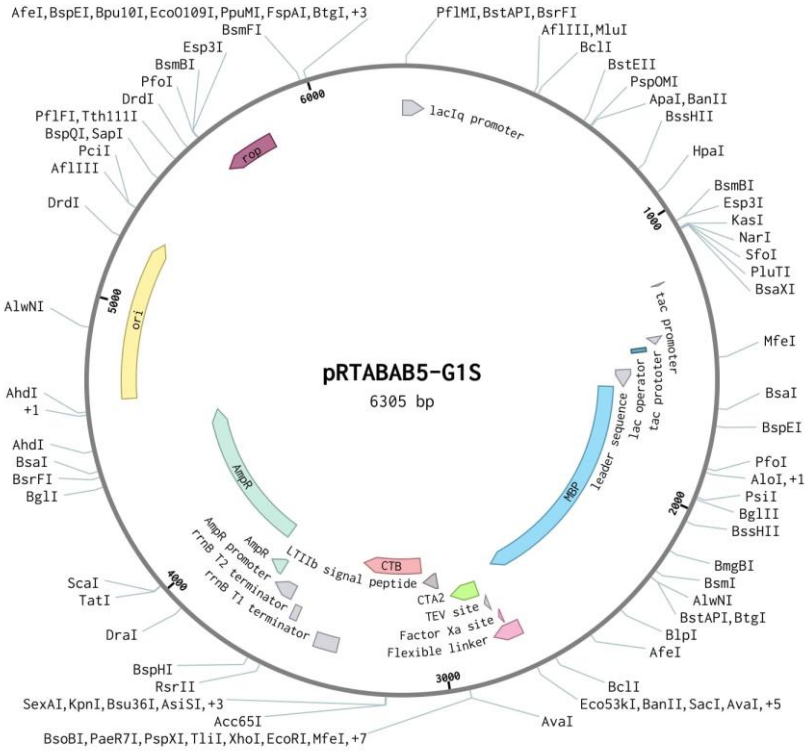


Figure VIII-1 Plasmid map for pTRBAB5-G1S.

Full sequence 5’ – 3’:

CCGACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTC

AGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGAC

CGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGA

TGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATT

GGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGC

CGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGG

CGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGAT

GCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACC

CATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGG

GTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGC

TGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCAT

GTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCA

ACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATT

TCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACA

GGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGA

AGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACC

GCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGG

GCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACAATTCTCATGTTTGACAGCT

TATCATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTG

TGCAGGTCGTAAATCACTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTT

GCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGT

ATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCCAGTCCGTTTAGGTGTTTTC

ACGAGCAATTGACCAACAAGGACCATAGATTATGAAAATAAAAACAGGTGCACGCATCCTCGCATTAT

CCGCATTAACGACGATGATGTTTTCCGCCTCGGCTCTCGCCAAAATCGAAGAAGGTAAACTGGTAATC

TGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGG

AATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCG

ATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCT

GAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAA

CGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGC

CGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGC

GCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGC

GTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGG

GTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCA

GAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGA

CACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCG

TTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAA

AACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCT

GAAGTCTTACGAGGAAGAGTTGGTGAAAGATCCGCGTATTGCCGCCACTATGGAAAACGCCCAGAAAG

GTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAAC

GCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAA

CAACAACAATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTTCACATATGGGATCCGAAAACC

TGTACTTTCAGAGTGGCGGTGATGAAAAAACCCAAAGTCATGGTGTAAAATTCCTTGACGAATACCAA

TCTAAAGTTAAAAGACAAATATTTTCAGGCTATCAATCTGATATTGATACACATAATAGAATTAAGGA

TGAATTATGACCTCGAGGTGAATTCACGAGCAATTGACCAACAAGGACCATAGATTATGAGCTTTAAG

AAAATTATCAAGGCATTTGTTATCATGGCTGCTTTGGTATCTGTTCAGGCGCATGCAGCTCCTCAAAA

TATTACTGATTTGTGCGCAGAATACCACAACACACAAATATATACGCTAAATGATAAGATCTTTTCGT

ATACAGAATCGCTAGCGGGAAAAAGAGAGATGGCTATCATTACTTTTAAGAATGGTGCAATTTTTCAA

GTAGAGGTACCAGGTAGTCAACATATAGATTCACAAAAAAAAGCGATTGAAAGGATGAAGGATACCCT

GAGGATTGCATATCTTACTGAAGCTAAAGTCGAAAAGTTATGTGTATGGAATAATAAAACGCCTCATG

CGATCGCCGCAATTAGTATGGCAAACTAAGTTTTCCCTGCAGGTAATTAAATAAGCTTCAAATAAAAC

GAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGT

AGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACG

CCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTC

TACAAACTCTTTCGGTCCGTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAA

TAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGC

CCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAA

AAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATC

CTTGAGAGTTTTCGCCCCGAAGAACGTTTCCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGC

GGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACT

TGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGT

GCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGA

GCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGA

ATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAA

CTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAA

AGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCG

GTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTT

ATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTC

ACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTCCTTAGGA

CTGAGCGTCAACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTG

CTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC

TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAG

TTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT

GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG

CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAA

CTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA

TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC

TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG

CGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGC

TCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTG

ATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTG

ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATAAGGTGCACTGTGACTGGG

TCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCA

TCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACC

GAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGCAGCGATTCACAGATGTCTGCCT

GTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCC

ATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGG

GGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGT

TACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAG

GGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGAT

GCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAAC

CGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCG

CGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAG

GAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAAATT

VIII.1.1.a MBP-CTA2B5

KIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFG

GYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALD

KELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHM

NADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPN

KELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPQMSAFW

YAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIEGRISHM**GSENLYFQSGGDEKTQSH**

GVKFLDEYQSKVKRQIFSGYQSDIDTHNRIKDEL

APQNITDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAIITFKNGAIFQVEVPGSQHIDSQKKAIERM

KDTLRIAYLTEAKVEKLCVWNNKTPHAIAAISMAN

Figure VIII-2 MBP-CTA2B_5 expressed from plasmid pTRBAB5-G1S. Individual region, e.g. MBP, flexible linker, Cholera toxin A2 peptide and CTB monomer are colored differently. The area being replaced with a procoil-CTA2 is highlighted by bold and underline, with different replacement pieces listed in the following chapter.

VIII.1.1.b Protein and gene sequence of MBP-Coiled coil-CTA2B5

The gene fragment and corresponding protein sequence of the TEV cutting site-procoil-CTA2 used to replace the original TEV cutting site-CTA2. Each gene fragments can be amplified with Forward primer (GTA CCA TGC ATA TGG GAT CCG) and reverse primer (GCT AAT CGA ATT CAC CTC GAG), digested with BamHI (GGATCC) and XhoI (CTCGAG) site, and ligated to the treated pTRBAB5-G1S.

Table VIII-1 Gene fragments and protein sequence of the replaced region to construct MBP-procoil-CTA2B$_5$

Class	Coiled coil	Individuals	Protein sequence (xxx-procoil-CTA2) [a]	Gene sequence (BamHI-procoil-CTA2-XhoI) [b]
Parallel heterodimer	PP1	PP1.1	GSENLYFQSVG SPEDEIQALEKENAQLEKKNQA LEKEIAQLEKEIQALEKEDEKT QSHGVKFLDEYQSKVKRQIFSG YQSDIDTHNRIKDEL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTAGCCCGGAAGATG AGATCCAGGCGTTGGAGAAAGAAAACGCGCAGCTTGAGAAAAAAAA TCAAGCACTGGAAAAGGAAATTGCCCAATTAGAAAAAGAAATTCAG GCTCTGGAGAAAGAAGATGAAAAAACCCAAAGTCATGGTGTAAAAT TCCTTGACGAATACCAATCTAAAGTTAAAAGACAAATATTTTCAGG CTATCAATCTGATATTGATACACATAATAGAATTAAGGATGAATTA TGACCTCGAG
		PP1.2	GSENLYFQSVG SPEDEIQALKEKNAQLKEKNQA LEEKIAQLKEKIQALKEKDEKT QSHGVKFLDEYQSKVKRQIFSG YQSDIDTHNRIKDEL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTAGCCCGGAAGATG AGATCCAGGCGCTGAAAGAAAAGAATGCGCAACTTAAAGAAAAGAA CCAAGCATTGGAAGAAAAAATTGCTCAGTTAAAAGAGAAAATTCAG GCCCTGAAAGAGAAAGATGAAAAAACCCAAAGTCATGGTGTAAAAT TCCTTGACGAATACCAATCTAAAGTTAAAAGACAAATATTTTCAGG CTATCAATCTGATATTGATACACATAATAGAATTAAGGATGAATTA TGACCTCGAG
	PP2	PP2.1	GSENLYFQSVG NLVAQLENEVASLENENETLKK KNLHKKDLIAYLEKEIANLRKK IEEDEKTQSHGVKFLDEYQSKV KRQIFSGYQSDIDTHNRIKDEL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTAACCTGGTTGCGC AGCTCGAAAACGAAGTTGCGTCTCTGGAAAATGAGAACGAAACCCT GAAGAAAAAGAACCTGCACAAAAAAGACCTGATCGCGTACCTGGAG AAAGAAATCGCGAATCTGCGTAAGAAAATCGAAGAAGATGAAAAAA CCCAAAGTCATGGTGTAAAATTCCTTGACGAATACCAATCTAAAGT TAAAAGACAAATATTTTCAGGCTATCAATCTGATATTGATACACAT AATAGAATTAAGGATGAATTATGACCTCGAG

		PP2.2	GSENLYFQSVG ARNAYLRKKIARLKKDNLQLER DEQNLEKIIANLRDEIARLENE VASHEQ DEKTQSHGVKFLDEYQSKVKRQ IFSGYQSDIDTHNRIKDEL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTGCGCGTAACGCGT ATCTGCGTAAGAAAATCGCACGTCTGAAAAAAGACAACCTGCAGCT GGAACGTGATGAACAGAACCTGGAAAAAATCATCGCGAACCTGCGT GACGAAATCGCGCGTCTCGAAAACGAAGTTGCGTCTCACGAACAGG ATGAAAAAACCCAAAGTCATGGTGTAAAATTCCTTGACGAATACCA ATCTAAAGTTAAAAGACAAATATTTTCAGGCTATCAATCTGATATT GATACACATAATAGAATTAAGGATGAATTATGACCTCGAG
	PP2-P	PP2.1-P	GSENLYFQSVG NLVAQLENEVASLENENETLKK KNLHKKDLIAYLEKEIANLRKK IEEPDEKTQSHGVKFLDEYQSK VKRQIFSGYQSDIDTHNRIKDE L	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTAACCTGGTTGCGC AGCTCGAAAACGAAGTTGCGTCTCTGGAAAATGAGAACGAAACCCT GAAGAAAAAGAACCTGCACAAAAAAGACCTGATCGCGTACCTGGAG AAAGAAATCGCGAATCTGCGTAAGAAAATCGAAGAACCGGATGAAA AAACCCAAAGTCATGGTGTAAAATTCCTTGACGAATACCAATCTAA AGTTAAAAGACAAATATTTTCAGGCTATCAATCTGATATTGATACA CATAATAGAATTAAGGATGAATTATGACCTCGAG
		PP2.2-P	GSENLYFQSVG ARNAYLRKKIARLKKDNLQLER DEQNLEKIIANLRDEIARLENE VASHEQP DEKTQSHGVKFLDEYQSKVKRQ IFSGYQSDIDTHNRIKDEL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTGCGCGTAACGCGT ATCTGCGTAAGAAAATCGCACGTCTGAAAAAAGACAACCTGCAGCT GGAACGTGATGAACAGAACCTGGAAAAAATCATCGCGAACCTGCGT GACGAAATCGCGCGTCTCGAAAACGAAGTTGCGTCTCACGAACAGC CGGATGAAAAAACCCAAAGTCATGGTGTAAAATTCCTTGACGAATA CCAATCTAAAGTTAAAAGACAAATATTTTCAGGCTATCAATCTGAT ATTGATACACATAATAGAATTAAGGATGAATTATGACCTCGAG
Antiparallel heterodimer	APCC	APCC.1	GSENLYFQSVG LEKE LAQLEKELQALEKKLAQLEKEN QALEKELAQLEKELQALQSKVK RQIFSGYQSDIDTHNRIKDEL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTCTTGAGAAGGAAT TAGCCCAGTTGGAAAAGGAGCTTCAGGCGTTGGAGAAAAAGCTCGC ACAGCTTGAGAAAGAAAACCAAGCATTAGAAAAGGAATTGGCCCAG TTAGAGAAAGAACTCCAAGCACTACAATCTAAAGTTAAAAGACAAA TATTTTCAGGCTATCAATCTGATATTGATACACATAATAGAATTAA GGATGAATTATGACCTCGAG
		APCC.2	GSENLYFQSVG LKKK LAQLKKKLQANKKKLAQLKKEL QALKKKLAQLKKKLQALQSKVK RQIFSGYQSDIDTHNRIKDEL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTTGAAGAAGAAGCT GGCACAACTCAAGAAAAAGTTACAAGCTAACAAGAAAAAATTGGCA CAGCTAAAAAAAGAACTACAGGCCTTGAAAAAGAAACTCGCACAGC TCAAGAAGAAACTCCAAGCACTGCAATCTAAAGTTAAAAGACAAAT

				ATTTTCAGGCTATCAATCTGATATTGATACACATAATAGAATTAAG GATGAATTATGACCTCGAG
	AP1	AP1.1	GSENLYFQSVG SPEDELAQLEKELQAIEKKLAQ LEKKAQAREKELAQLEKEQSKV KRQIFSGYQSDIDTHNRIKDEL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTAGCCCGGAAGATG AACTGGCGCAGCTGGAGAAGGAATTACAGGCGATTGAAAAAAAACT TGCCCAACTGGAAAAAAAGGCTCAAGCACGCGAGAAAGAATTGGCC CAGCTCGAAAAAGAGCAATCTAAAGTTAAAAGACAAATATTTTCAG GCTATCAATCTGATATTGATACACATAATAGAATTAAGGATGAATT ATGACCTCGAG
		AP1.2	GSENLYFQSVG SPEDELAQLKEELQAIKEELAQ LKEKAQARKEKLAQLKEKQSKV KRQIFSGYQSDIDTHNRIKDEL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTAGCCCGGAAGATG AATTAGCTCAGCTCAAAGAGGAGTTGCAAGCCATTAAAGAAGAACT GGCGCAACTGAAAGAAAAAGCACAGGCGCGCAAAGAGAAGCTGGCC CAGCTTAAGGAAAAACAATCTAAAGTTAAAAGACAAATATTTTCAG GCTATCAATCTGATATTGATACACATAATAGAATTAAGGATGAATT ATGACCTCGAG
	AP2	AP2.1	GSENLYFQSVG QLEKELAQLKEKLQAIEKELAQ LKEKAQALKEKLAQLKEKLQQS KVKRQIFSGYQSDIDTHNRIKD EL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTCAGCTGGAAAAAG AACTGGCGCAGCTAAAAGAGAAACTCCAGGCAATCGAGAAAGAACT TGCCCAGCTGAAAGAAAAAGCGCAAGCTCTGAAAGAGAAGCTGGCC CAGTTGAAAGAAAAGTTACAACAATCTAAAGTTAAAAGACAAATAT TTTCAGGCTATCAATCTGATATTGATACACATAATAGAATTAAGGA TGAATTATGACCTCGAG
		AP2.2	GSENLYFQSVG QLEKELAQLEKELQAIEEKLAQ LEKEAQALKKKLAQLEKELQQS KVKRQIFSGYQSDIDTHNRIKD EL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTCAGCTGGAGAAAG AACTGGCTCAGCTAGAAAAAGAACTGCAAGCAATCGAAGAAAAACT CGCCCAGCTGGAAAAGGAGGCGCAGGCCTTGAAAAAAAAGCTGGCG CAACTTGAAAAAGAGTTACAGCAATCTAAAGTTAAAAGACAAATAT TTTCAGGCTATCAATCTGATATTGATACACATAATAGAATTAAGGA TGAATTATGACCTCGAG
		AP2.1_G5	GSENLYFQSVGGGGGG QLEKELAQLKEKLQAIEKELAQ LKEKAQALKEKLAQLKEKLQQS KVKRQIFSGYQSDIDTHNRIKD EL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTGGTGGCGGTGGCG GTCAGCTGGAAAAAGAACTGGCGCAGCTAAAAGAGAAACTCCAGGC AATCGAGAAAGAACTTGCCCAGCTGAAAGAAAAAGCGCAAGCTCTG AAAGAGAAGCTGGCCCAGTTGAAAGAAAAGTTACAACAATCTAAAG TTAAAAGACAAATATTTTCAGGCTATCAATCTGATATTGATACACA

				TAATAGAATTAAGGATGAATTATGACCTCGAG
		AP2.2_G5	GSENLYFQSVGGGGGG QLEKELAQLEKELQAIEEKLAQ LEKEAQALKKKLAQLEKELQQS KVKRQIFSGYQSDIDTHNRIKD EL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTGGTGGCGGTGGCG GTCAGCTGGAGAAAGAACTGGCTCAGCTAGAAAAAGAACTGCAAGC AATCGAAGAAAAACTCGCCCAGCTGGAAAAGGAGGCGCAGGCCTTG AAAAAAAAGCTGGCGCAACTTGAAAAAGAGTTACAGCAATCTAAAG TTAAAAGACAAATATTTTCAGGCTATCAATCTGATATTGATACACA TAATAGAATTAAGGATGAATTATGACCTCGAG
		AP2.1_I_S	IEGRSSHMGSENLYFQSVG QLEKELAQLKEKLQAIEKELAQ LKEKAQALKEKLAQLKEKLQQS KVKRQIFSGYQSDIDTHNRIKD EL	ATCGAGGGAAGGTCTTCACATATGGGATCCGAAAACCTGTACTTTC AGAGTGTAGGT CAGCTGGAAAAAGAACTGGCGCAGCTAAAAGAGAAACTCCAGGCAA TCGAGAAAGAACTTGCCCAGCTGAAAGAAAAAGCGCAAGCTCTGAA AGAGAAGCTGGCCCAGTTGAAAGAAAAGTTACAACAATCTAAAGTT AAAAGACAAATATTTTCAGGCTATCAATCTGATATTGATACACATA ATAGAATTAAGGATGAATTATGACCTCGAG
	AP2-M	AP2.1-Marsh	GSENLYFQSVG QLEKELAQLKEKLQAIKKKLAQ LKEKAQALKEKLAQLKEKLQQS KVKRQIFSGYQSDIDTHNRIKD EL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTCAGCTGGAAAAAG AACTGGCGCAGCTAAAAGAGAAACTCCAGGCAATCAAGAAAAAACT TGCCCAGCTGAAAGAAAAAGCGCAAGCTCTGAAAGAGAAGCTGGCC CAGTTGAAAGAAAAGTTACAACAATCTAAAGTTAAAAGACAAATAT TTTCAGGCTATCAATCTGATATTGATACACATAATAGAATTAAGGA TGAATTATGACCTCGAG
		AP2.2-M	GSENLYFQSVG QLEKELAQLEKELQAIEEELAQ LEKEAQALKKKLAQLEKELQQS KVKRQIFSGYQSDIDTHNRIKD EL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTCAGCTGGAGAAAG AACTGGCTCAGCTAGAAAAAGAACTGCAAGCAATCGAAGAAGAACT CGCCCAGCTGGAAAAGGAGGCGCAGGCCTTGAAAAAAAAGCTGGCG CAACTTGAAAAAGAGTTACAGCAATCTAAAGTTAAAAGACAAATAT TTTCAGGCTATCAATCTGATATTGATACACATAATAGAATTAAGGA TGAATTATGACCTCGAG
		AP2.2-Marsh	GSENLYFQSVG QLEKELAQLEKELQAIEEELAQ LEKEAQALEKKLAQLEKELQQS KVKRQIFSGYQSDIDTHNRIKD EL	GGATCCGAAAACCTGTACTTTCAGAGTGTAGGTCAGCTGGAGAAAG AACTGGCTCAGCTAGAAAAAGAACTGCAAGCAATCGAAGAAGAACT CGCCCAGCTGGAAAAGGAGGCGCAGGCCTTGGAAAAAAAGCTGGCG CAACTTGAAAAAGAGTTACAGCAATCTAAAGTTAAAAGACAAATAT TTTCAGGCTATCAATCTGATATTGATACACATAATAGAATTAAGGA

			TGAATTATGACCTCGAG
Antiparallel homodimer	AP1-homo	GSENLYFQGVGGSGSGSGSGTP AEQLEKELKQLEKELQAIEKQL AQLQWKAQARKKKLAQLKKKLQ APDEKTQSHGVKFLDEYQSKVK RQIFSGYQSDIDTHNRIKDEL	AAATTTAAATTTGTACCATGCATATGGGATCCGAAAACCTGTACTT TCAGGGCGTAGGTGGCTCCGGTTCAGGTTCAGGCTCTGGCACCCCG GCGGAACAGCTGGAAAAGGAACTGAAACAGCTGGAAAAAGAACTGC AAGCGATTGAGAAGCAACTGGCGCAGCTGCAGTGGAAAGCGCAGGC CCGCAAAAAGAAACTGGCCCAGCTGAAAAAAAAACTGCAGGCGCCG GATGAAAAAACCCAAAGTCATGGTGTAAAATTCCTCGACGAATACC AGTCCAAAGTTAAAAGACAAATATTTTCAGGCTATCAATCTGATAT TGATACACATAATAGAATTAAGGATGAATTATGACCTCGAGGTGAA TTCGATTAGCAAATTTAAATTT
	AP2-homo	GSENLYFQGVGGSGSGSGSGTP AEQLEKELAQLEKELQAIEKQL AQLQKKAQAKKQKLAQLKKKLQ APDEKTQSHGVKFLDEYQSKVK RQIFSGYQSDIDTHNRIKDEL	AAATTTAAATTTGTACCATGCATATGGGATCCGAAAACCTGTACTT TCAGGGCGTAGGTGGCTCCGGTTCAGGTTCAGGCTCTGGCACCCCG GCGGAACAGCTGGAAAAGGAACTGGCGCAGCTGGAAAAAGAACTGC AAGCGATTGAGAAGCAACTGGCGCAGCTGCAGAAAAAAGCGCAGGC CAAAAAACAGAAACTGGCCCAGCTGAAAAAAAAACTGCAGGCGCCG GATGAAAAAACCCAAAGTCATGGTGTAAAATTCCTCGACGAATACC AGTCCAAAGTTAAAAGACAAATATTTTCAGGCTATCAATCTGATAT TGATACACATAATAGAATTAAGGATGAATTATGACCTCGAGGTGAA TTCGATTAGCAAATTTAAATTT
	AP3-homo	GSENLYFQGVGGSGSGSGSGTP AEQLEQELAQLEAELQAIEQQL AQLQAKAQARKQKLAQLKAKLQ APDEKTQSHGVKFLDEYQSKVK RQIFSGYQSDIDTHNRIKDEL	AAATTTAAATTTGTACCATGCATATGGGATCCGAAAACCTGTACTT TCAGGGCGTAGGTGGCTCCGGTTCAGGTTCAGGCTCTGGCACCCCG GCGGAACAGCTGGAACAGGAACTGGCGCAGCTGGAAGCGGAACTGC AAGCGATTGAGCAGCAACTGGCGCAGCTGCAGGCGAAAGCGCAGGC CCGCAAACAGAAACTGGCCCAGCTGAAAGCGAAACTGCAGGCGCCG GATGAAAAAACCCAAAGTCATGGTGTAAAATTCCTCGACGAATACC AGTCCAAAGTTAAAAGACAAATATTTTCAGGCTATCAATCTGATAT TGATACACATAATAGAATTAAGGATGAATTATGACCTCGAGGTGAA TTCGATTAGCAAATTTAAATTT
	AP4-homo	GSENLYFQGVGGSGSGSGSGTP AEDIEQELERAKASIRRLEQEV NQERSRMAYLQTLLAKPDEKTQ	AAATTTAAATTTGTACCATGCATATGGGATCCGAAAACCTGTACTT TCAGGGCGTAGGTGGCTCCGGTTCAGGTTCAGGCTCTGGCACCCCG GCGGAAGACATCGAGCAGGAGCTGGAGCGCGCGAAGGCCTCCATTC

		SHGVKFLDEYQSKVKRQIFSGY QSDIDTHNRIKDEL	GGCGCCTGGAGCAGGAGGTGAACCAGGAGCGCAGCCGCATGGCGTA CCTGCAGACGTTGCTGGCCAAGCCGGATGAAAAAACCCAAAGTCAT GGTGTAAAATTCCTCGACGAATACCAGTCCAAAGTTAAAAGACAAA TATTTTCAGGCTATCAATCTGATATTGATACACATAATAGAATTAA GGATGAATTATGACCTCGAGGTGAATTCGATTAGCAAATTTAAATT T

VIII.1.2pET21a-Se6

pET21a-Se6 was ordered from Addgene (Plasmid #46367) to express alive streptavidin with 6 glutamates at the C terminus.

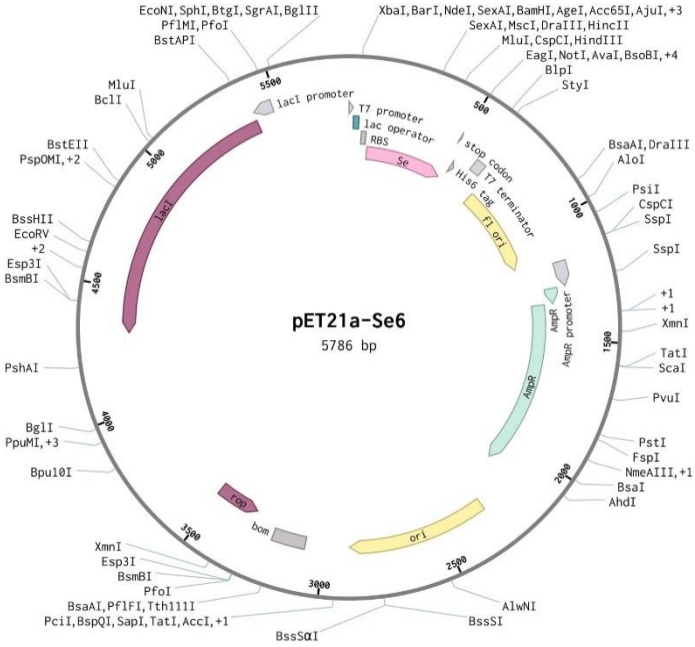


Figure VIII-3 Plasmid map for pET21a-Se6

Full sequence 5’ – 3’:

TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAAC

TTTAAGAAGGAGATATACATATGGCTGAAGCTGGTATCACCGGCACCTGGTACAACCAGCTGGGATCC

ACCTTCATCGTTACCGCTGGTGCTGACGGTGCTCTGACCGGTACCTACGAATCCGCTGTTGGTAACGC

TGAATCTAGATACGTTCTGACCGGTCGTTACGACTCCGCTCCGGCTACCGACGGTTCCGGAACCGCTC

TGGGTTGGACCGTTGCTTGGAAAAACAACTACCGTAACGCTCACTCCGCTACCACCTGGTCTGGCCAG

TACGTTGGTGGTGCTGAAGCTCGTATCAACACCCAGTGGTTGTTGACCTCCGGCACCACCGAAGCTAA

CGCGTGGAAATCCACCCTGGTTGGTCACGACACCTTCACCAAAGTTAAACCGTCCGCTGCTTCCGAAG

AGGAAGAGGAAGAGTAATAAAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCG

GCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACC

CCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGA

ATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTA

CACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC

TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGA

CCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCC

CTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCT

ATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCT

GATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCG

GGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGA

GACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGT

GTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAA

AGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTA

AGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT

GGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAA

TGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTAT

GCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCG

AAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGA

GCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGC

GCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCG

GATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGG

AGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCG

TAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGT

GCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAA

ACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTT

AACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCT

TTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCC

GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTG

TCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCT

CTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAG

ACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGG

AGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAA

GGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCC

AGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTT

TGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTG

GCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTAT

TACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCG

AGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCAAT

GGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTAC

GTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTG

CTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACC

GTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGA

TGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATA

AAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCT

GTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAAC

ATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAA

ATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCA

TCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAA

CACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCAC

GTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCT

CAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGC

CGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCA

AGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGC

CGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCC

GCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAA

TGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTG

CCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTT

TTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAG

CAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATAT

AACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGAC

TCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGAT

GCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCG

CTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACA

GAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAG

TCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATA

ACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATG

ATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCG

TTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTT

GCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGT

TGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTT

CGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTG

CGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCAT

GCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACT

CCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCA

TGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGC

GCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAA

CCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCG

CGAAAT

VIII.1.2.a Se6

AEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSGTALGWTVAWKNNYRN
AHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVKPSAASEEEEEE

Figure VIII-4 Streptavidin-Glutamate6 (Se6) expressed from plasmid pSTAB1.1.

VIII.1.2.b Protein and gene sequence of other streptavidin variants

The gene fragment and corresponding protein sequence of different streptavidin variant used to replace the Se6 region. The plasmids of Traptavidin and Dead streptavidin-Aspartate loop (Dd) were purchased from Addgene, and the Alive streptavidin (SA) and Dead streptavidin (D) were further engineered from the commercial vector to remove the inserted loop.

Table VIII-2 Gene fragments and protein sequence of the replaced region to construct different streptavidin variants.

Name	Protein sequence	Gene sequence
Traptavidin (Addgene #24753)	AEAGITGTWYNQLGSTFIVTAGADG ALTGTYESAVGNAEGDYVLTGRYDS APATDGSGTALGWTVAWKNNYRNAH SATTWSGQYVGGAEARINTQWLLTS GTTEANAWKSTLVGHDTFTKVKPSA ASHHHHHH	GCTGAAGCTGGTATCACCGGCACCTGGTACAACCAGCTGGGATCC ACCTTCATCGTTACCGCTGGTGCTGACGGTGCTCTGACCGGTACC TACGAATCCGCTGTTGGTAACGCTGAAGGCGATTACGTTCTGACC GGTCGTTACGACTCCGCTCCGGCTACCGACGGTTCCGGAACCGCT CTGGGTTGGACCGTTGCTTGGAAAAACAACTACCGTAACGCTCAC TCCGCTACCACCTGGTCTGGCCAGTACGTTGGTGGTGCTGAAGCT CGTATCAACACCCAGTGGTTGTTGACCTCCGGCACCACCGAAGCC AACGCGTGGAAATCCACCCTGGTTGGTCACGACACCTTCACCAAA GTTAAACCGTCCGCTGCTTCCCATCACCATCACCACCAT
Alive Streptavidin (SA)	AEAGITGTWYNQLGSTFIVTAGADG ALTGTYESAVGNAESRYVLTGRYDS APATDGSGTALGWTVAWKNNYRNAH SATTWSGQYVGGAEARINTQWLLTS GTTEANAWKSTLVGHDTFTKVKPSA AS	GCTGAAGCTGGTATCACCGGCACCTGGTACAACCAGCTGGGATCC ACCTTCATCGTTACCGCTGGTGCTGACGGTGCTCTGACCGGTACC TACGAATCCGCTGTTGGTAACGCTGAATCTAGATACGTTCTGACC GGTCGTTACGACTCCGCTCCGGCTACCGACGGTTCCGGAACCGCT CTGGGTTGGACCGTTGCTTGGAAAAACAACTACCGTAACGCTCAC TCCGCTACCACCTGGTCTGGCCAGTACGTTGGTGGTGCTGAAGCT CGTATCAACACCCAGTGGTTGTTGACCTCCGGCACCACCGAAGCT AACGCGTGGAAATCCACCCTGGTTGGTCACGACACCTTCACCAAA GTTAAACCGTCCGCTGCTTCC
Dead Streptavidin (D)	AEAGITGTWYAQLGDTFIVTAGADG ALTGTYEAAVGNAESRYVLTGRYDS APATDGSGTALGWTVAWKNNYRNAH	GCTGAAGCTGGTATCACCGGCACCTGGTACGCCCAGCTGGGAGAC ACCTTCATCGTTACCGCTGGTGCTGACGGTGCTCTGACCGGTACC TACGAAGCCGCTGTTGGTAACGCTGAATCTAGATACGTTCTGACC

	SATTWSGQYVGGAEARINTQWLLTS GTTEANAWKSTLVGHDTFTKVKPSA AS	GGTCGTTACGACTCCGCTCCGGCTACCGACGGTTCCGGAACCGCT CTGGGTTGGACCGTTGCTTGGAAAAACAACTACCGTAACGCTCAC TCCGCTACCACCTGGTCTGGCCAGTACGTTGGTGGTGCTGAAGCT CGTATCAACACCCAGTGGTTGTTGACCTCCGGCACCACCGAAGCT AACGCGTGGAAATCCACCCTGGTTGGTCACGACACCTTCACCAAA GTTAAACCGTCCGCTGCTTCC
Dead Streptavidin-D (Dd) (Addgene #46368)	AEAGITGTWYAQLGDTFIVTAGADG ALTGTYEAAVGDDDGDDDGDDDGAE SRYVLTGRYDSAPATDGSGTALGWT VAWKNNYRNAHSATTWSGQYVGGAE ARINTQWLLTSGTTEANAWKSTLVG HDTFTKVKPSAAS	GCTGAAGCTGGTATCACCGGCACCTGGTACGCCCAGCTGGGAGAC ACCTTCATCGTTACCGCTGGTGCTGACGGTGCTCTGACCGGTACC TACGAAGCCGCTGTTGGTGATGACGATGGTGACGATGACGGTGAT GACGATGGAGCTGAATCTAGATACGTTCTGACCGGTCGTTACGAC TCCGCTCCGGCTACCGACGGTTCCGGAACCGCTCTGGGTTGGACC GTTGCTTGGAAAAACAACTACCGTAACGCTCACTCCGCTACCACC TGGTCTGGCCAGTACGTTGGTGGTGCTGAAGCTCGTATCAACACC CAGTGGTTGTTGACCTCCGGCACCACCGAAGCTAACGCGTGGAAA TCCACCCTGGTTGGTCACGACACCTTCACCAAAGTTAAACCGTCC GCTGCTTCC

VIII.1.3pET11b-3C6 affimer-Cys

The gene fragment of Affimer variant, after being amplified through PCR reaction using Forward primer (pET11_Adhiron_forward: TTC TGG CGT TTT CTG CGT CTG C) and reverse primer (pDHis_cys_C_term_rev: TTA CTA ATG CGG CCG CAC AAG CGT CAC), can harbor a cysteine residue at the C terminus and be subcloned from phage-plasmid into pET21b-3C6-Cys affimer plasmid between NheI and NotI restriction sites.

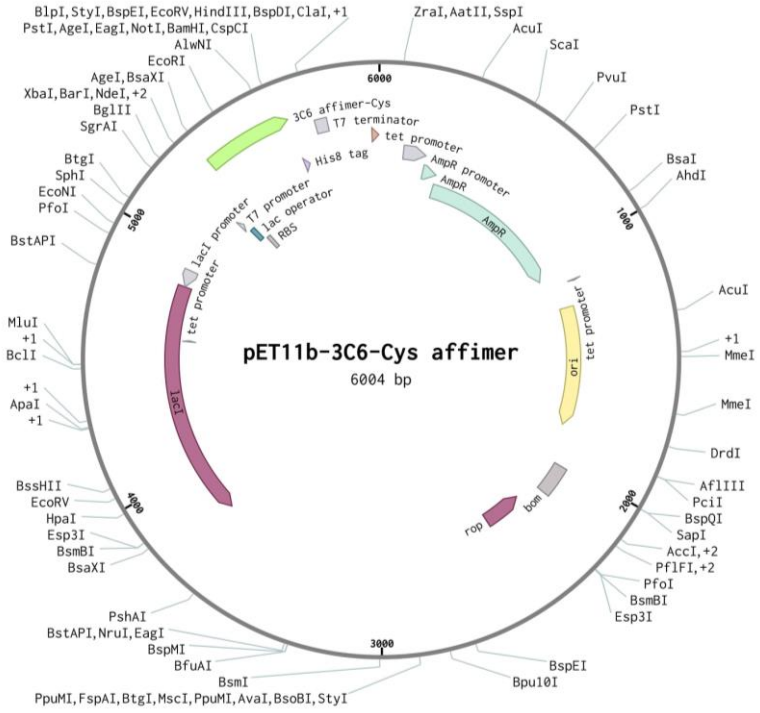


Figure VIII-5 Plasmid map for pET11b-3C6 affimer-Cys.

Full sequence 5’ – 3’:

TTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTT

CTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATA

CATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAA

GAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTT

TTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTAC

ATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGAT

GAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCG

GTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACG

GATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTT

ACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAA

CTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATG

CCTGCAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCA

ACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTG

GCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGG

CCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACG

AAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACT

CATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTT

GATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAA

GATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCAC

CGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTC

AGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTC

TGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT

CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGG

GGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCT

ATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAA

CAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGC

CACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAG

CAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTAT

CCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACG

ACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCA

TCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTT

AAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACC

CGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCG

GGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCA

TCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTC

CAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCA

CTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATG

CTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGC

GGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATG

TAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCT

GACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGC

AGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAA

GGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGAC

CCAACGCTGCCCGAGATGCGCCGCGTGCGGCTGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCA

AGGGTTGGTTTGCGCATTCACAGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATC

CGTTAGCGAGGTGCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCGGCTCCATGCACCGCGACGCAACG

CGGGGAGGCAGACAAGGTATAGGGCGGCGCCTACAATCCATGCCAACCCGTTCCATGTGCTCGCCGAG

GCGGCATAAATCGCCGTGACGATCAGCGGTCCAGTGATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGA

TCCTTGAAGCTGTCCCTGATGGTCGTCATCTACCTGCCTGGACAGCATGGCCTGCAACGCGGGCATCC

CGATGCCGCCGGAAGCGAGAAGAATCATAATGGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGC

AAGACGTAGCCCAGCGCGTCGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGT

GGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGA

TCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCT

ACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAA

GGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAAC

TTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAA

TGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCA

GTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACG

CTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTC

TTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGC

GCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGC

ATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAAT

TTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGC

CCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCT

TCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATT

AGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGA

CGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGAC

ACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTG

CAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGC

GGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGG

CTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAA

CGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAA

AGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAA

GCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGG

CGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCG

AAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGC

GCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACG

ACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGA

AGGAGATATACATATGGCTAGCGCCGCTACCGGTGTTCGTGCAGTTCCGGGTAACGAAAACTCCCTGG

AAATCGAAGAACTGGCTCGTTTCGCTGTTGACGAACACAACAAAAAAGAAAACGCTCTGCTGGAATTC

GTTCGTGTTGTTAAAGCGAAAGAACAGATGGACCTGAACGCTGGTCTGCCGCGTACCATGTACTACCT

GACCCTGGAAGCTAAAGACGGTGGTAAAAAGAAACTGTACGAAGCGAAAGTTTGGGTTAAGCAGGGCC

TGAAAAAACTGAAATTCACCAACTTCAAAGAACTGCAGGAGTTCAAACCGGTTGGTGACGCTTGTGCG

GCCGCGCATCACCATCATCACCACCATCATTGATAAGGATCCGGCTGCTAACAAAGCCCGAAAGGAAG

CTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTG

AGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCCGCAAGAGGCCCGGCAGTACCGGCATA

ACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTT

CATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGA

TAAGCTGTCAAACATGAGAA

VIII.1.3.a 3C6 affimer-CYS-His_8

ASAATGVRAVPGNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQMDLNAGLPRTMYYLTLEAKDGGKKKL
YEAKVWVKQGLKKLKFTNFKELQEFKPVGDACAAAHHHHHHHH

Figure VIII-6 3C6 affimer-Cys expressed from plasmid pET11b-VII.1.3.a 3C6 affimer-CYS-His_8. The main body of 3C6 Affimer is colored green, and the His_8 tag is colored purple.

VIII.1.3.b 3C6 affimer-24AA-CYS-His_8

ASAATGVRAVPGNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQMDLNAGLPRTMYYLTLEAK
DGGKKKLYEAKVWVKQGLKKLKFTNFKELQEFKPVGTKKAEEEKKAAEEKKAEEEKKAAEEACAAAHH
HHHHHH

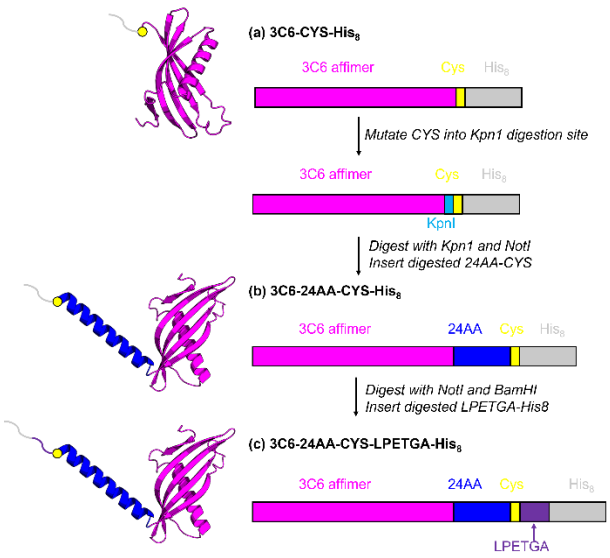
Figure VIII-7 3C6 affimer-24AA-Cys expressed from plasmid pET11b-VII.1.3.b 3C6 affimer-24AA-CYS-His_8. The main body of 3C6 Affimer is colored green, the inserted 24AA α-helix is colored yellow and the His_8 tag is colored purple.

VIII.1.3.c 3C6 affimer-24AA-CYS-LPETGA-His_8

ASAATGVRAVPGNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQMDLNAGLPRTMYYLTLEAK
DGGKKKLYEAKVWVKQGLKKLKFTNFKELQEFKPVGTKKAEEEKKAAEEKKAEEEKKAAEEACAAALP
ETGAHHHHHHHH

Figure VIII-8 3C6 affimer-24AA-Cys expressed from plasmid pET11b-VII.1.3.b 3C6 affimer-24AA-CYS-His_8. The main body of 3C6 Affimer is colored green, the inserted 24AA α-helix is colored yellow, the His_8 tag is colored purple, and the sortase labelling tag LPETGA is colored cyan.

VIII.1.3.d Genetic work to insert 24AA-LPETGA at the C-terminus



Detailed gene shift for constructing 3C6-24AA-CYS-LPETGA-His$_8$

CAGGAGTTCAAACCGGTTGGTGACGCTTGTGCGGCCGCGCATCACCATCATCACCACCATCATTGA
ATCAATGATGGTGGTGATGATGGTGATGCGCGGCCGCACAAGCGTCACCAACCGGTTTGAACTCCT
Q E F K P V G D A C A A A H H H H H H H H *

Mutate CYS into KpnI digestion site

GGTACC
CAGGAGTTCAAACCGGTTGGTGACGCTTGTGCGGCCGCGCATCACCATCATCACCACCATCATTGA
ATCAATGATGGTGGTGATGATGGTGATGCGCGGCCGCACAAGCGTCACCAACCGGTTTGAACTCCT
CCATGG
Q E F K P V G T A C A A A H H H H H H H H *

Digest with KpnI and NotI

KpnI *NotI*

CAGGAGTTCAAACCGGTTGGTGGTACCTGTGCGGCCGCGCATCACCATCATCACCACCATCATTGA
ATCAATGATGGTGGTGATGATCCATGGGCGCGGCCGCACAAGCGTCACCAACCGGTTTGAACTCCT
Q E F K P V G T A C A A A H H H H H H H H *

Insert 24AA-CYS

CAGGAGTTCAAACCGGTTGGTGGTAC **(24AA-CYS)** GGCCGCGCATCACCATCATCACCACCATCATTGA
ATCAATGATGGTGGTGATGATCCATG **(24AA-CYS)** GCCGCACAAGCGTCACCAACCGGTTTGAACTCCT
Q E F K P V G T A **(24AA-CYS)** A A H H H H H H H H *

Digest with NotI and BamHI

NotI *BamHI*

CAGGAGTTCAAACCGGTTGGTGGTAC **(24AA-CYS)** GGCCGCGCATCACCATCATCACCACCATCATTGA
ATCAATGATGGTGGTGATGATCCATG **(24AA-CYS)** GCCGCACAAGCGTCACCAACCGGTTTGAACTCCT
Q E F K P V G T A **(24AA-CYS)** A A H H H H H H H H *

Insert digested LPETGA-His8

NotI *BamHI digestion site destroyed*

CAGGAGTTCAAACCGGTTGGTGGTAC **(24AA-CYS)** **(LPETGA-His8)**
ATCAATGATGGTGGTGATGATCCATG **(24AA-CYS)** **(LPETGA-His8)**
Q E F K P V G T A **(24AA-CYS)** **(LPETGA-His8)**

Figure VIII-9. Genetic work to insert 24AA-LPETGA at the C-terminus of 3C6 Affimer.

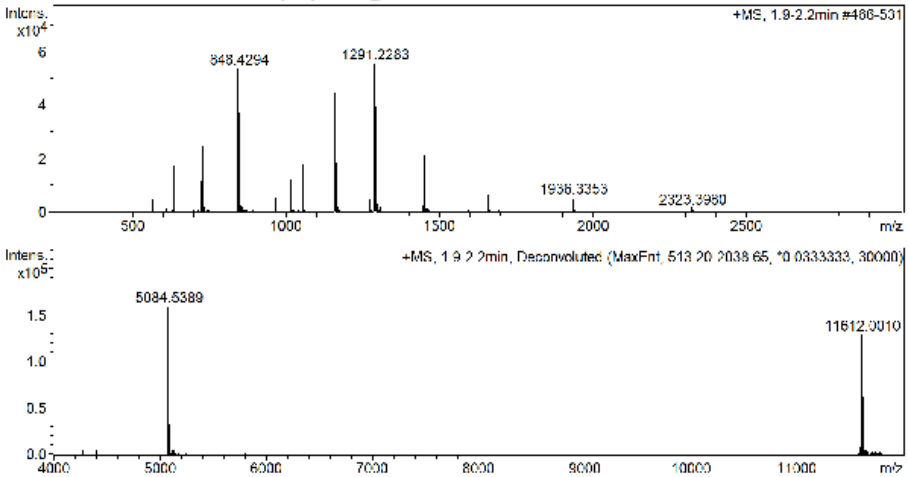
VIII.2 Primer library

Primer name	Primer sequence	Additional Information
Eva_gene_block_forward	GTACCATGCATATGGGATCCG	For amplifying the DNA fragments ordered from TWIST to replace the CTA2 peptide in pTRABAB5-G1S plasmid into procoil-CTA2.
Eva_gene_block_reverse	GCTAATCGAATTCACCTCGAG	
AP2.1_delete_A2_forward	GCTGGCCCAGTTGAAAGAAAAGTTACAATAATCTAAAGTTAAAAGACA AATATTTTCAGGC	Delete CTA2 peptide from AP2.1.
AP2.1_delete_A2_reverse	GCCTGAAAATATTTGTCTTTTAACTTTAGATTATTGTAACTTTTCTTT CAACTGGGCCAGC	
PP2.1_delete_A2_forward	CGCGAATCTGCGTAAGAAAATCGAAGAATAAGAAAAAACCCAAAGTCATGGTG	Delete CTA2 peptide from PP2.1.
PP2.1_delete_A2_reverse	CACCATGACTTTGGGTTTTTTCTTATTCTTCGATTTTCTTACGCAGATTCGCG	
PP2.2_delete_A2_forward	GTTGCGTCTCACGAACAGTAAGAAAAAACCCAAAGTCATGGTG	Delete CTA2 peptide from PP2.2.
PP2.2_delete_A2_reverse	CACCATGACTTTGGGTTTTTTCTTACTGTTCGTGAGACGCAAC	
FWD_AP2.1_S_I	CAACCTCGGGATCGAGGGAACAGTTTCACATATGGGATCCG	Mutate the Factor Xa recognition site in AP2.1 from IEGRI to IEGRS
REV_AP2.1_S_I	TTCGGATCCCATATGTGAAACTGTTCCCTCGATCCCGAGGTTG	
NEW AP2.1 FWD QC (GGGS)3	CGAAAACCTGTACTTTCAGAGTGTAGGTGGTGGCGGTAGCGGT GGCGGTAGCGGTGGCGGTAGCCAGCTGGAAAAAGAACTGGCGCAG	Insert $(GGGS)_3$ between TEV cutting site and procoil in AP2.1.
NEW AP2.1 REV QC (GGGS)3	CTCTTTTAGCTGCGCCAGTTCTTTTTCCAGCTGGCTACCGCCACC GCTACCGCCACCGCTACCGCCACCACCTACACTCTGAAAGTACAGGTTTTCG	
PP2.1-P-FWD	CTGCGTAAGAAAATCGAAGAACCGGATGAAAAAACCC	Insert proline between procoil and CTA2 in PP2.1.
PP2.1-P-REV	GGGTTTTTTCATCCGGTTCTTCGATTTTCTTACGC	
PP2.2-P-FWD	GTCTCACGAACAGCCGGATGAAAAAACCC	Insert proline between procoil and CTA2 in
PP2.2-P-REV	GGGTTTTTTCATCCGGCTGTTCGTGAGAC	

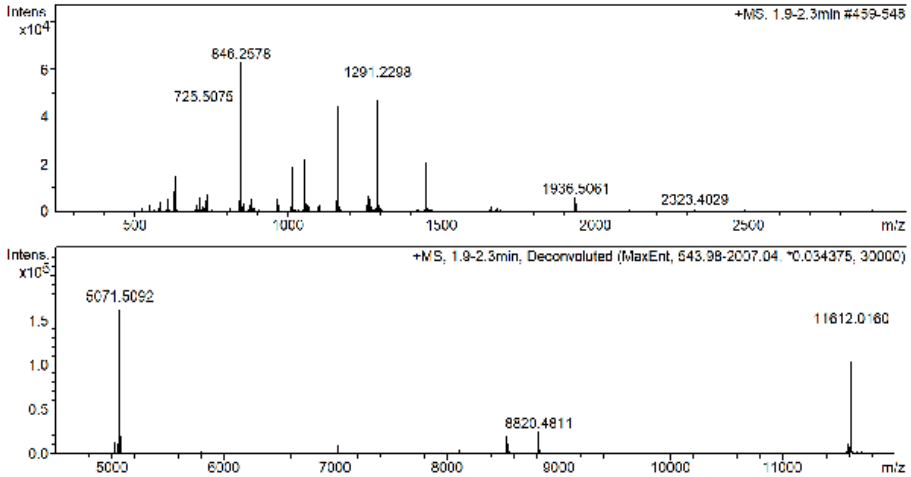
		PP2.2.
Dd-D Forward	CGCTGAATCTAGATACGTTCTGACCG	Q5 mutagenesis to delete the aspartate loop of Dd into Dead streptavidin (D).
Dd-D Reverse	TTACCAACAGCGGCTTCGTAG	
Se-S Forward	TAATAAAAGCTTGCGGCCGCACTC	Q5 mutagenesis to delete the glutamate loop of Se6 into Alive streptavidin.
Se-S Reverse	GGAAGCAGCGGACGGTTTAACTTTG	
StxB_signal_D_FWD	ACGCCTGATTGTGTAACTGGAAAGGTGGAG	To remove the LTIIb signal peptide for StxB in pSAB2.1.
StxB_signal_D_REV	CATAATCTATGGTCCTTGTTGGTCAATTGCTCGTAAG	
P5X into C5X forward-2	AATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGC	To remove the signal peptide for MBP in pSAB2.1.
P5X into C5X reverse-2	TTCATAATCTATGGTCCTTGTTGGTCAATTGCTCGTG	
extreme_cut_F_STX	GGTGTGCACCACCACCAC	Q5 mutagenesis to remove the flexible linker between MBP and Stx1A2 peptide in pSAB1.1.
extreme_cut_STX_reverse	GTTCGAGCTCGAATTAGTCTGC	
pET11_Adhiron_forward	TTCTGGCGTTTTCTGCGTCTGC	To amplify the affimer fragment from phage-vector to pET11b plasmid and carry a cysteine at the C terminus.
pDHis_cys_C_term_rev	TTACTAATGCGGCCGCACAAGCGTCAC	
Affimer_3C6_Fpw_insert_KpnI	GAGTTCAAACCGGTGGTACCGCTTGTGCGGCCGCGCAT	To mutate the C terminus of affimer to carry a KPN1 digestion site to insert the 24AA helix.
Affimer_3C6_REV_insert_KpnI	ATGCGCGGCCGCACAAGCGGTACCAACCGGTTTGAACTC	

VIII.3 Mass spectramatry (MS)

(a) digested CTA2B5



(b) oxidised CTA2B5



(c) biotinylated CTA2B5

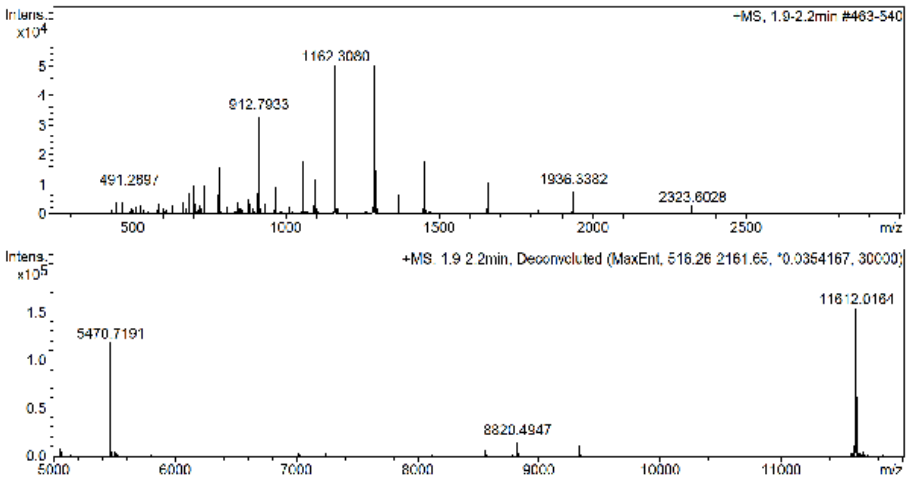


Figure VIII-10 (a) The CTA2B$_5$ with MBP cleavage with TEV protease. (b) Oxidised CTA2 peptide using sodium periodate. (c) CTA2 decorated with PEG4-biotin through hydroxylamine – aldehyde interaction.

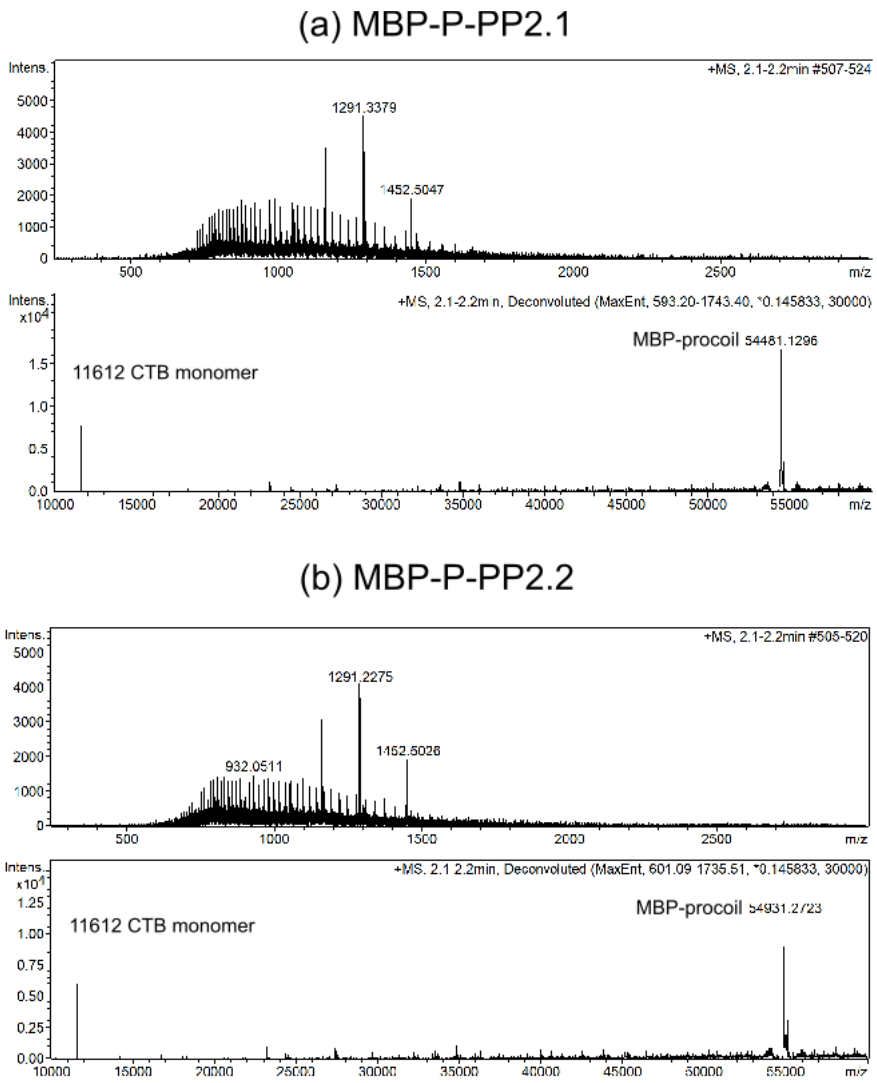


Figure VIII-11 Mass spectrum of MBP-P-PP2.1 and MBP-P-PP2.2.

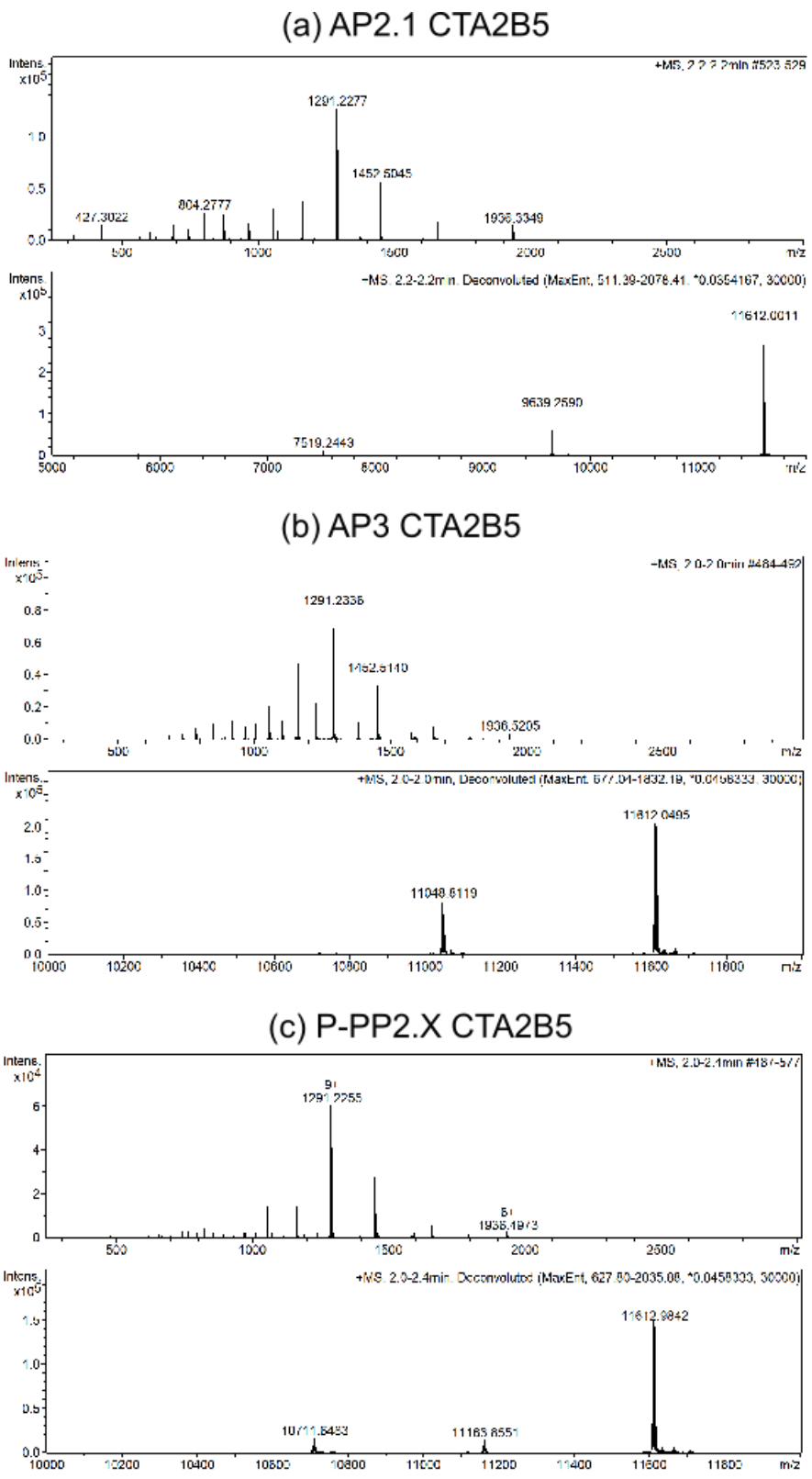


Figure VIII-12 AP2.1, AP3, and P-PP2.X proteins, with MBP either removed through Factor Xa cutting site or by TEV protease.

VIII.4 Isothermal Titration Calorimetry (ITC)

Table VIII-3 Calculated KD for MBP-CTA2B5 : GM1 binding interaction using ITC

Name		K_D_best_fit	K_D low_range	K_D_high_range
Classic CTB		38.82	61.52	23.39
C41 MBP-CTA2B5		39.66	131.37	8.68
S1	X-press (V1) pET30a <MBP-A2 CTxB>Cer	60.35	97.90	36.37
S2	X-press (V1) pET30a T70.09 <MBP-A2 CTxB>Cer	44.03	142.46	7.83
S3	X-press (V2) pET30a <MBP-A2 CTxB>Cer	49.19	91.60	24.65
S4	X-press (V2) pET30a T70.09 <MBP-A2 CTxB>Cer	44.36	80.72	22.23
S5	X-press (V1)ΔThyA pET30a <MBP-A2 CTxB>ThyA_Cer	45.44	80.80	23.85
S6	BL21 (DE3) ΔThyA pET30a <MBP-A2 CTxB>ThyA_Cer	35.72	103.01	7.46
S7	BL21 (DE3) pET30a <MBP-A2 CTxB>Cer	67.14	125.03	32.88
S8	BL21 (DE3) pTRBAB5-GS <MBP-A2 CTxB>	29.70	66.51	11.29

Table VIII-4 Thermodynamic parameters for MBP-CTA2B5 : GM1 binding interaction using ITC

Name	Repeat	Calculated binding site concentration (μM)	Correction factor for binding site	Best-fit ΔH (kcal/mol)	Low range (ΔH)	High range (ΔH)	Best-fit Log10[KAB]	Low range (Log10[KAB])	High range Log10[KAB]	Gained from x round of ITC
Classic CTB	1	3.5 (cuvette)	1.038	-19.83	-21.03	-18.75	7.41	7.21	7.63	2nd
	2		1.046							
C41 MBP-CTA2B5	1	4.6 (Nanodrop)	1.096	-17.99	-20.39	-16.07	7.40	6.84	8.14	2nd
	2		1.086							
S1	1	4.5 (Nanodrop)	1.075	-18.90	-20.21	-17.77	7.22	6.99	7.47	1st
	2		1.043							
S2	1	5.1 (Nanodrop)	1.034	-17.04	-20.08	-14.82	7.36	6.80	8.21	1st
	2		0.909							
S3	1	5.4 (Nanodrop)	0.980	-19.17	-20.78	-17.75	7.31	7.01	7.64	1st
	2		0.957							
S4	1	4.9 (Nanodrop)	1.079	-19.43	-20.80	-18.22	7.35	7.07	7.68	1st
	2		1.099							
S5	1	5.5 (Nanodrop)	0.998	-19.64	-21.25	-18.20	7.34	7.07	7.65	2nd
	2		1.016							
S6	1	5.2 (Nanodrop)	1.032	-17.78	-20.32	-15.72	7.45	6.95	8.22	1st
	2		0.949							
S7	1	4.9 (Nanodrop)	1.213	-18.94	-20.85	-17.37	7.17	6.87	7.51	1st
	2		1.010							
S8	1	4.3 (Nanodrop)	1.034	-18.04	-19.87	-16.44	7.34	7.15	7.99	2nd
	2		0.989							

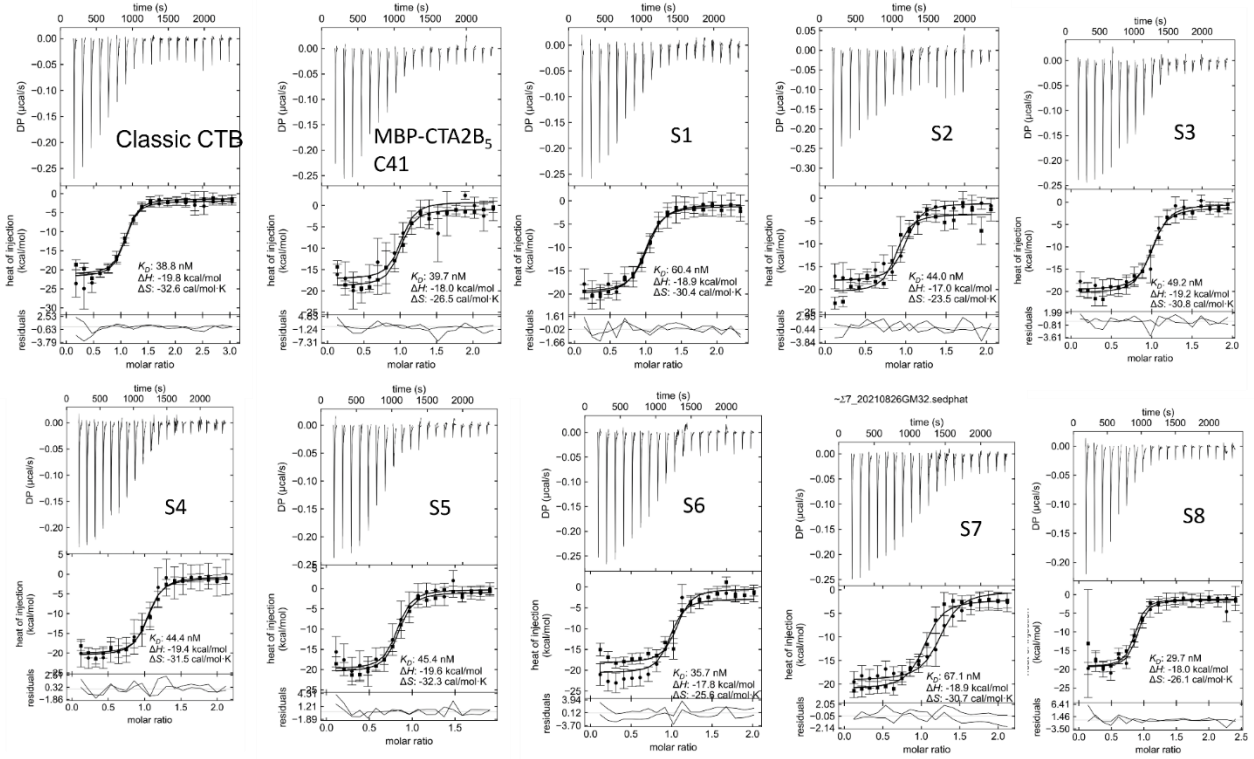
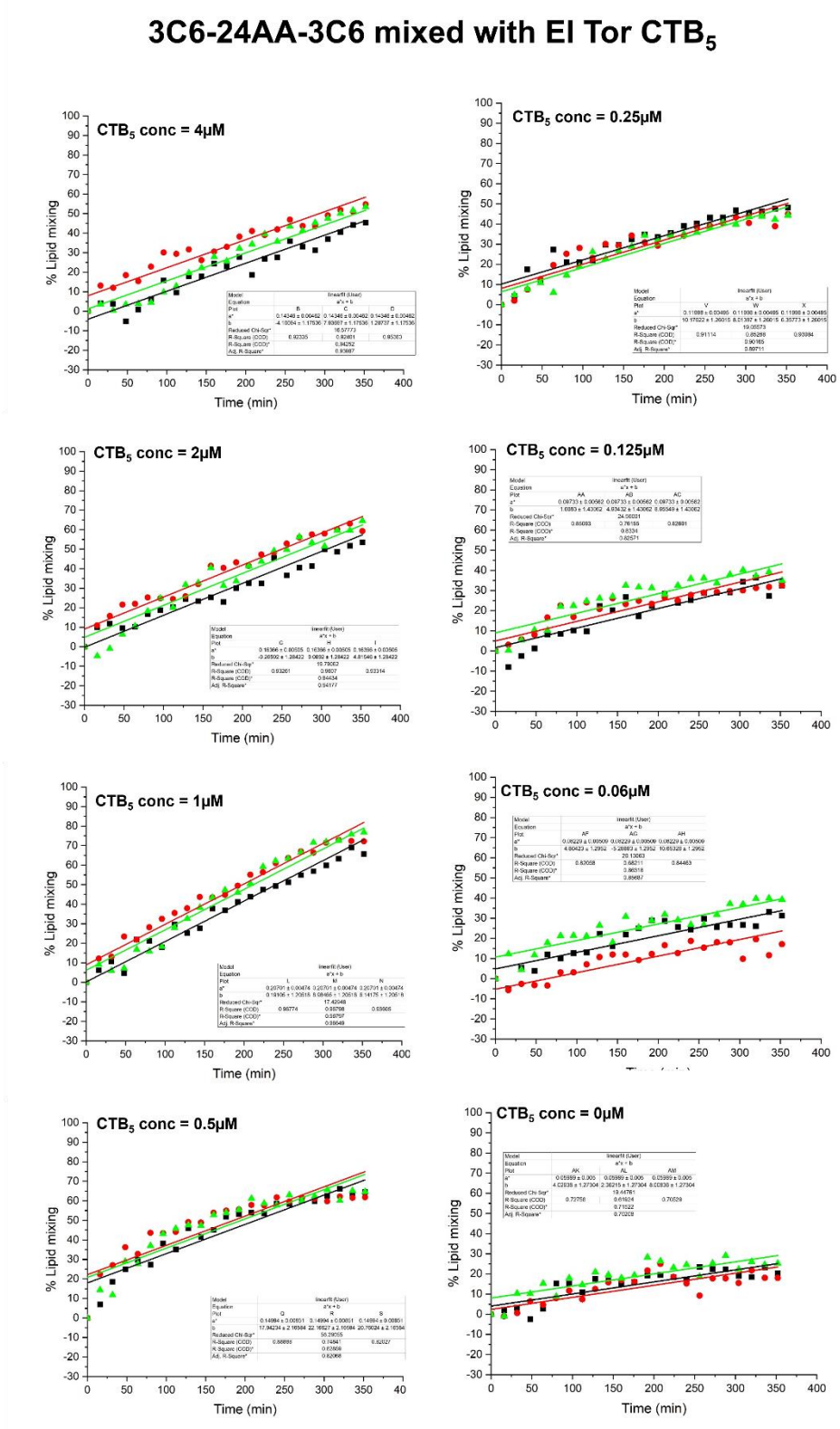


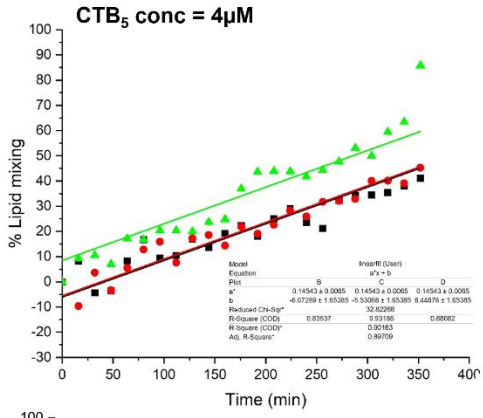
Figure VIII-13 ITC analysis result of MBP-CTA2B$_5$ expressed from different strains or vectors (from enGenes).

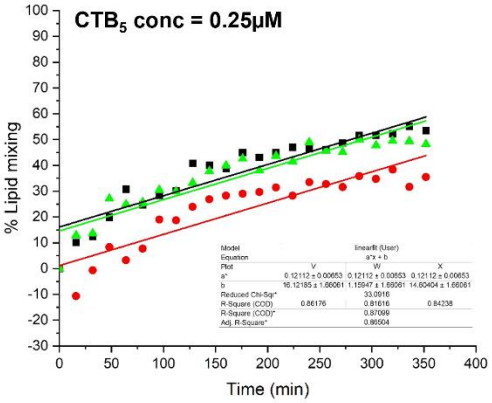
VIII.5 Global fitting of Lipid mixing assay

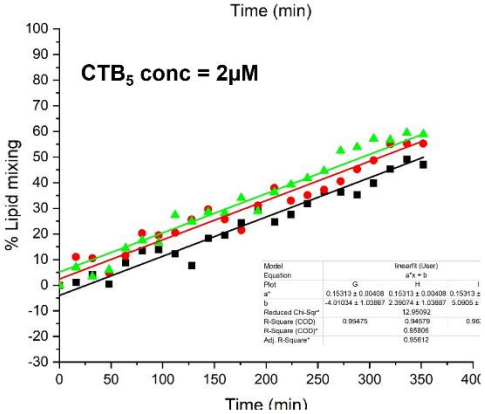
VIII.5.1Affimer dimer

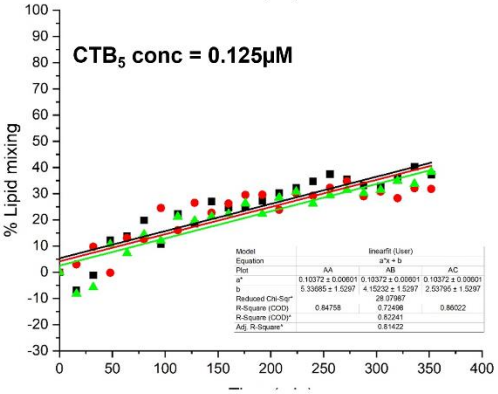


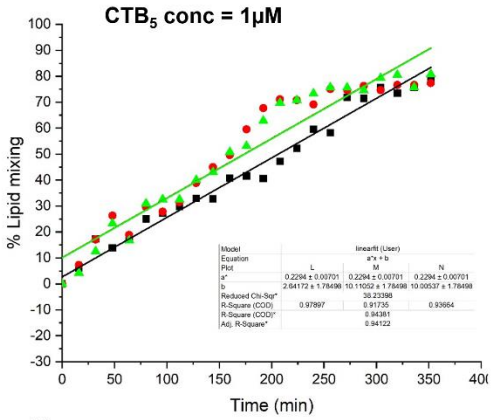
3C6-3C6 mixed with El Tor CTB_5

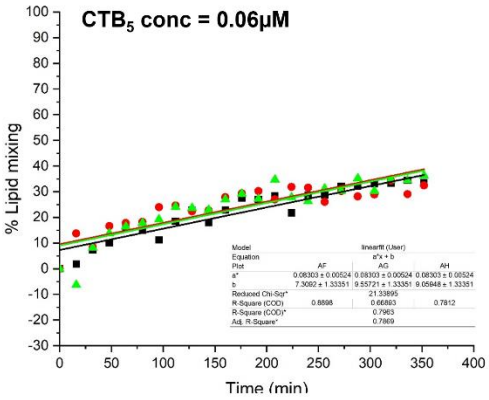


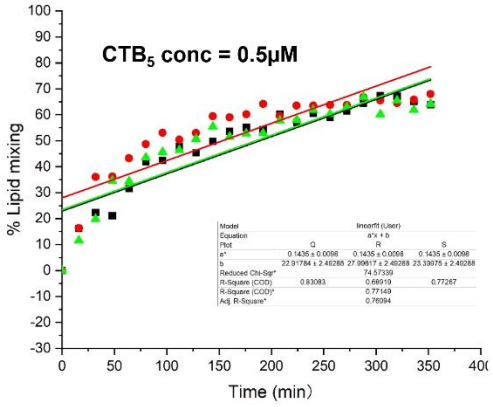


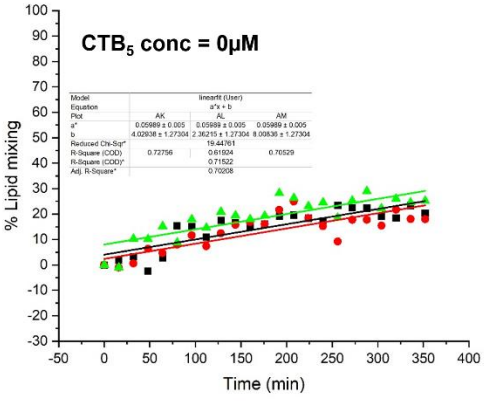




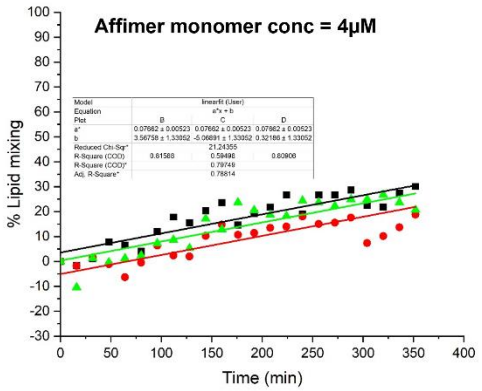


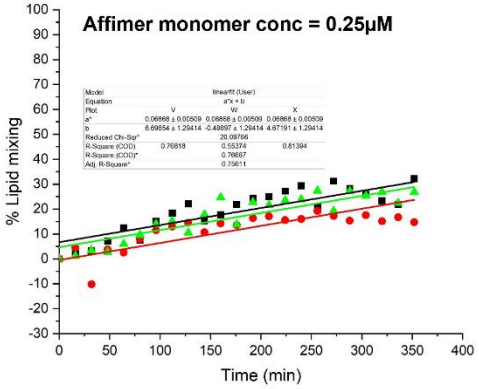


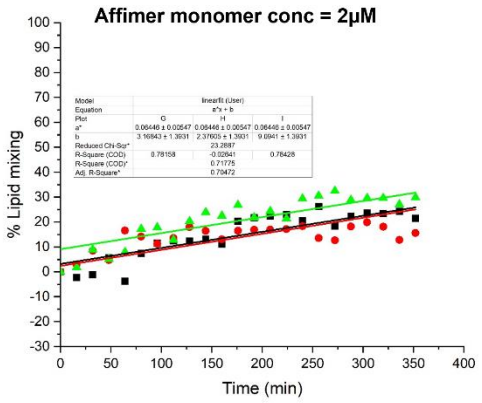


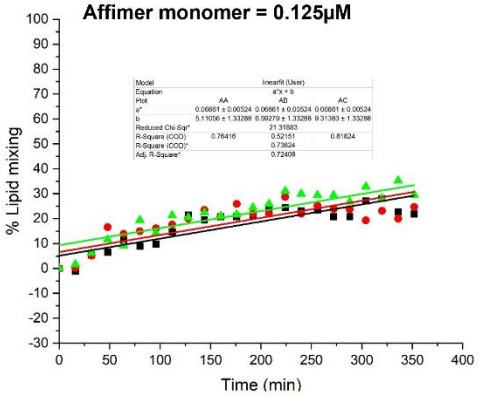


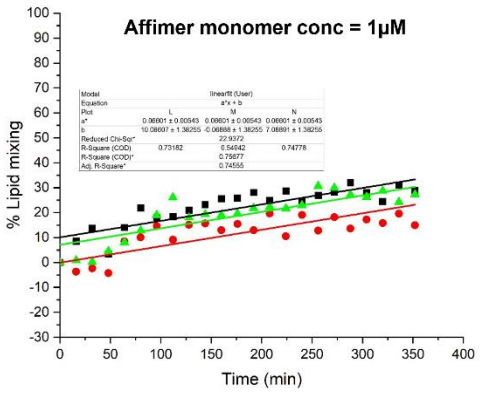
3C6-24AA-3C6

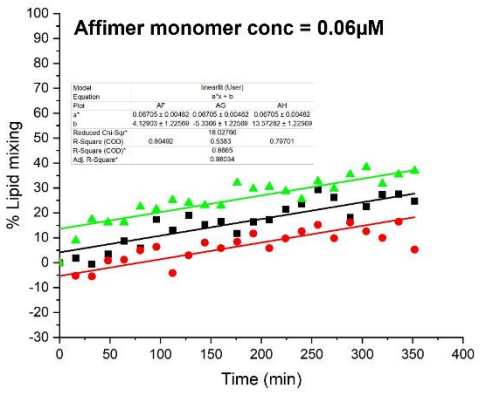

Affimer monomer conc = 4μM
% Lipid mixing
Time (min)

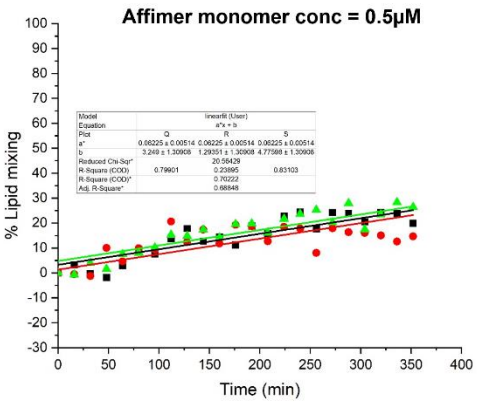

Affimer monomer conc = 0.25μM
% Lipid mixing
Time (min)

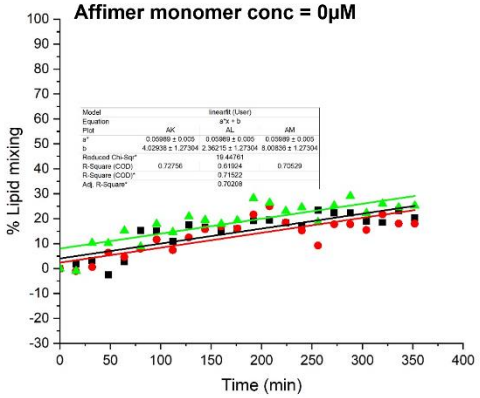

Affimer monomer conc = 2μM
% Lipid mixing
Time (min)


Affimer monomer = 0.125μM
% Lipid mixing
Time (min)

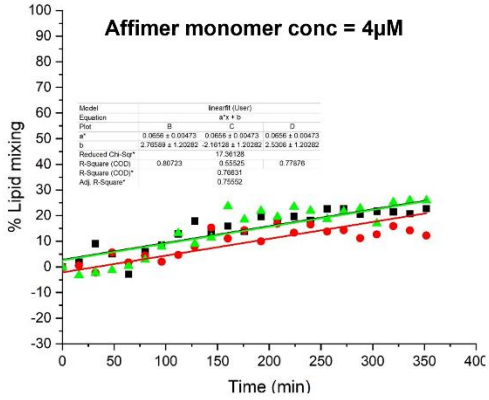

Affimer monomer conc = 1μM
% Lipid mixing
Time (min)

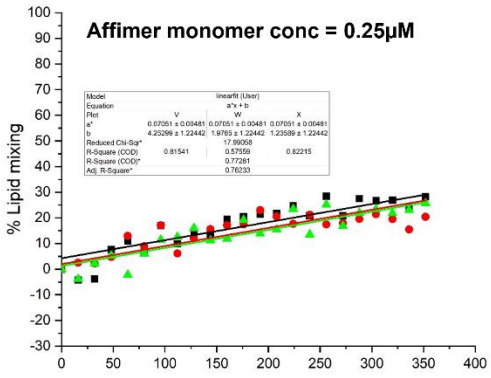

Affimer monomer conc = 0.06μM
% Lipid mixing
Time (min)

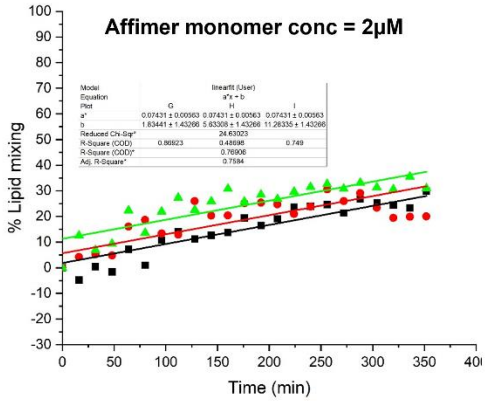

Affimer monomer conc = 0.5μM
% Lipid mixing
Time (min)

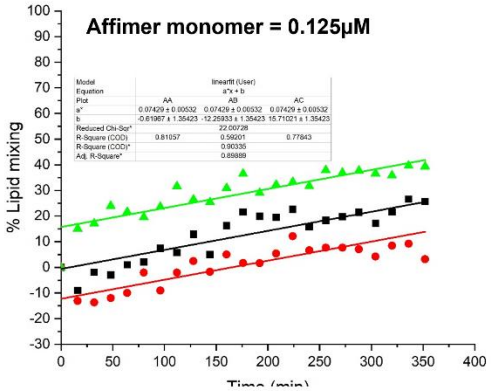

Affimer monomer conc = 0μM
% Lipid mixing
Time (min)

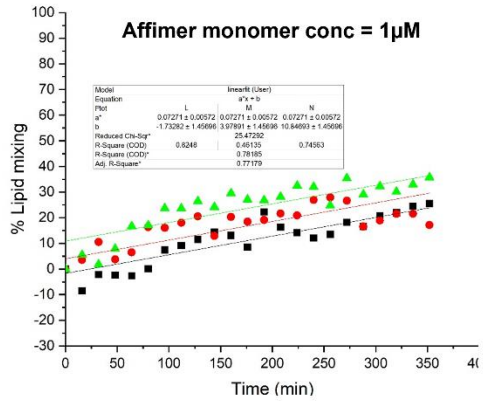
3C6-3C6

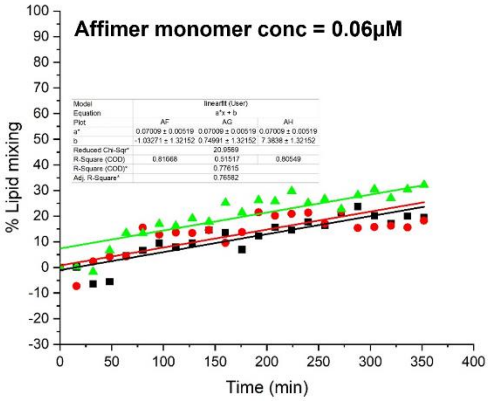

Affimer monomer conc = 4μM
% Lipid mixing
Time (min)

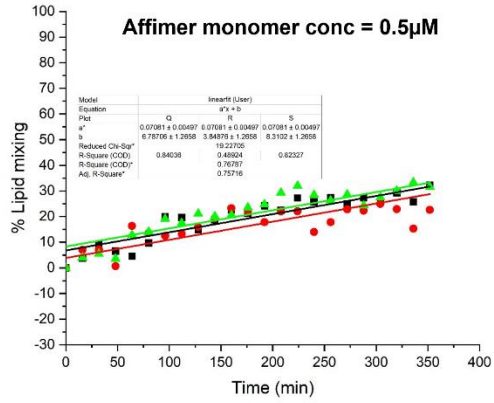

Affimer monomer conc = 0.25μM
% Lipid mixing

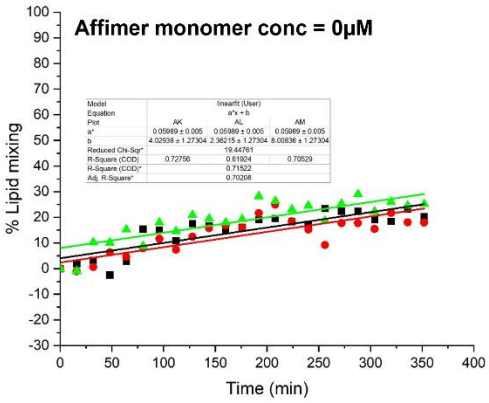

Affimer monomer conc = 2μM
% Lipid mixing
Time (min)


Affimer monomer = 0.125μM
% Lipid mixing

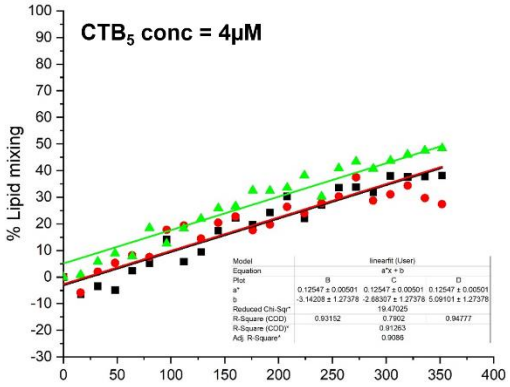

Affimer monomer conc = 1μM
% Lipid mixing
Time (min)

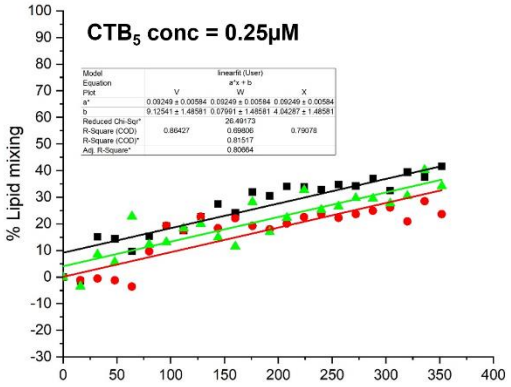

Affimer monomer conc = 0.06μM
% Lipid mixing
Time (min)

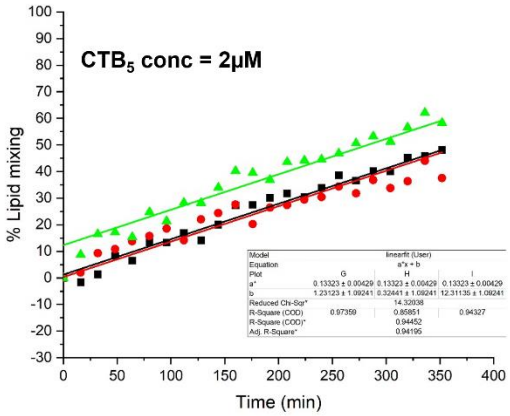

Affimer monomer conc = 0.5μM
% Lipid mixing
Time (min)

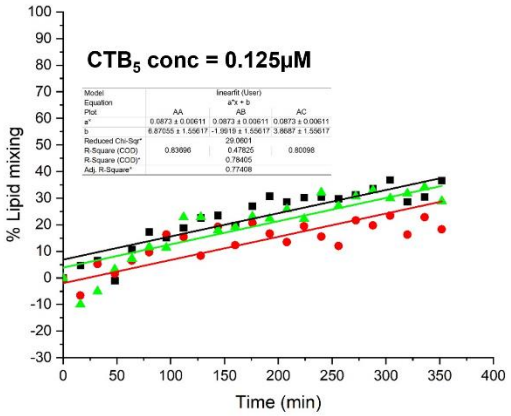

Affimer monomer conc = 0μM
% Lipid mixing
Time (min)

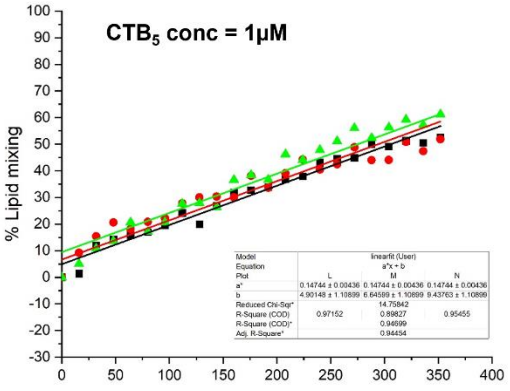
El Tor CTB_5

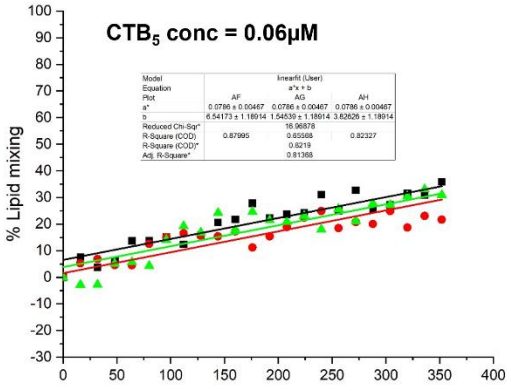

CTB5 conc = 4μM
% Lipid mixing

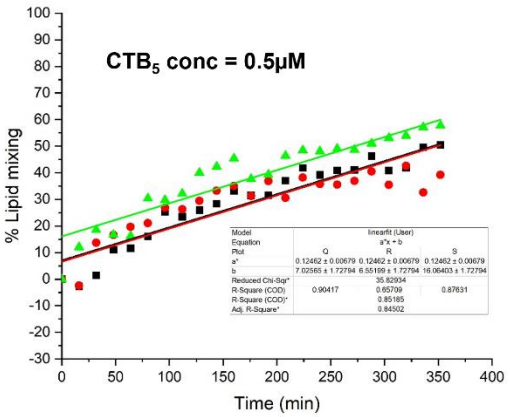

CTB5 conc = 0.25μM
% Lipid mixing


CTB5 conc = 2μM
% Lipid mixing
Time (min)


CTB5 conc = 0.125μM
% Lipid mixing
Time (min)


CTB5 conc = 1μM
% Lipid mixing


CTB5 conc = 0.06μM
% Lipid mixing


CTB5 conc = 0.5μM
% Lipid mixing
Time (min)

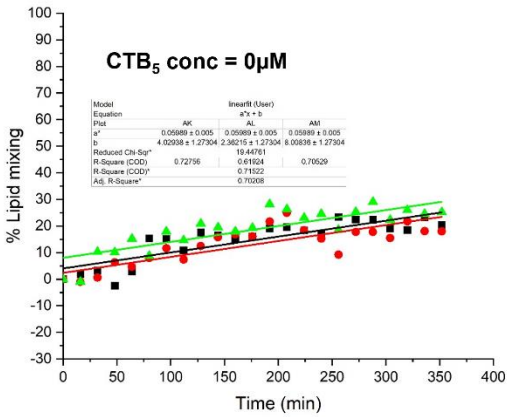

CTB5 conc = 0μM
% Lipid mixing
Time (min)

Figure VIII-14 Global fit of three repeats from lipid mixing results. The slope for each plot represents the percentage of lipid mixing (%) per minute, and the value is used to indicate the speed of lipid mixing rate.

Concentration of CTB_5 (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.14543	0.0065
2	0.15313	0.00408
1	0.2294	0.00701
0.5	0.1435	0.0098
0.25	0.12112	0.00653
0.125	0.10372	0.00601
0.0625	0.08303	0.00524
0	0.05989	0.005

Table VIII-5 Lipid mixing rate for different concentration 3C6-3C6 affimer dimer mixed with El Tor CTB_5.

Concentration of CTB_5 (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.14346	0.00462
2	0.16366	0.00505
1	0.20701	0.00474
0.5	0.14994	0.00851
0.25	0.11998	0.00495
0.125	0.09733	0.00562
0.0625	0.08229	0.00509
0	0.05989	0.005

Table VIII-6 Lipid mixing rate for different concentration 3C6-24AA-3C6 affimer dimer mixed with El Tor CTB_5.

Concentration of CTB_5 (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.12547	0.00501
2	0.13323	0.00429
1	0.14744	0.00436
0.5	0.12462	0.00679
0.25	0.09249	0.00584
0.125	0.0873	0.00611
0.0625	0.0786	0.00467
0	0.05989	0.005

Table VIII-7 Lipid mixing rate for different concentration of El Tor CTB_5.

Concentration of affimer monomer (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.0656	0.00473
2	0.07431	0.00563
1	0.07271	0.00572
0.5	0.07081	0.00497
0.25	0.07051	0.00481
0.125	0.07429	0.00532
0.0625	0.07009	0.00519
0	0.05989	0.005

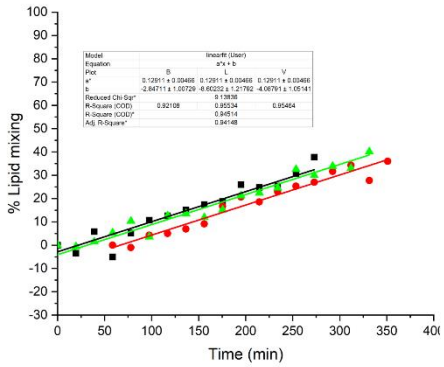
Table VIII-8 Lipid mixing rate for different concentration of 3C6-3C6 affimer dimer.

Concentration of affimer monomer (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.07662	0.00523
2	0.06446	0.00547
1	0.06601	0.00543
0.5	0.06225	0.00514
0.25	0.06868	0.00509
0.125	0.06861	0.00524
0.0625	0.06705	0.00482
0	0.05989	0.005

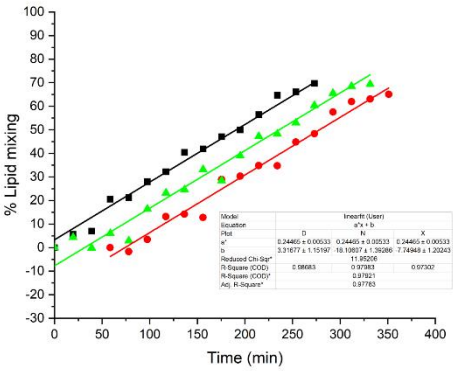
Table VIII-9 Lipid mixing rate for different concentration of 3C6-24AA-3C6 affimer dimer.

Streptavidin + biotin-CTA2B$_5$

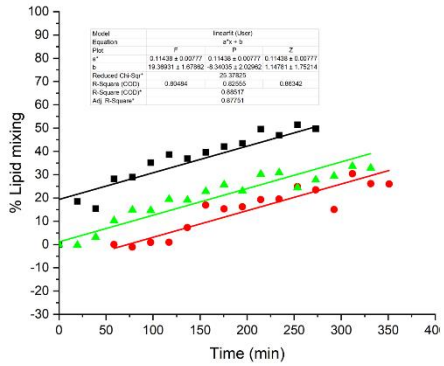
CTB$_5$ conc = 4μM



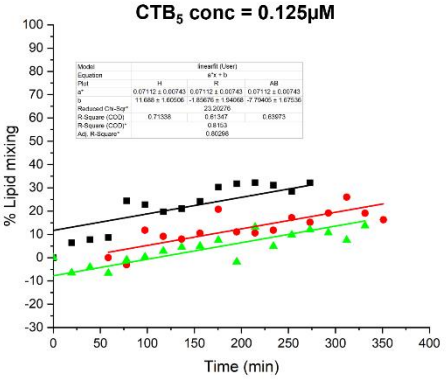
CTB$_5$ conc = 0.25μM



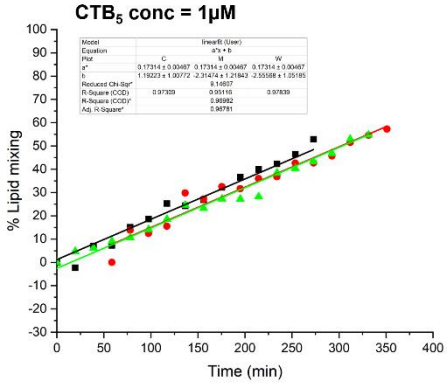
CTB$_5$ conc = 2μM



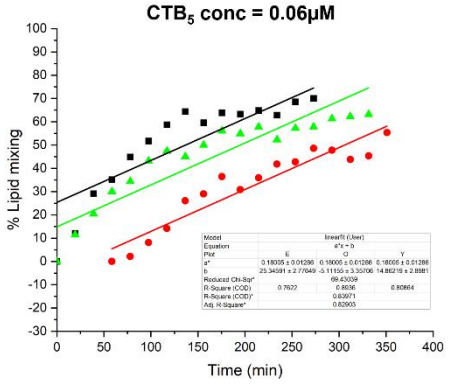
CTB$_5$ conc = 0.125μM



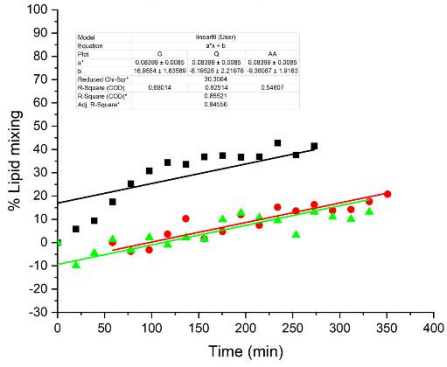
CTB$_5$ conc = 1μM



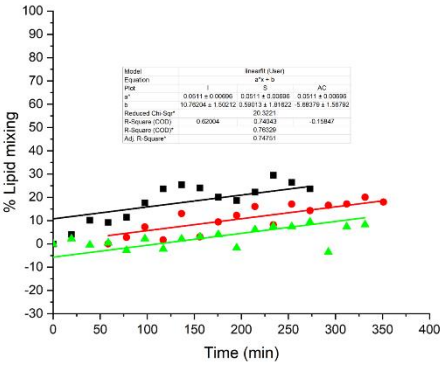
CTB$_5$ conc = 0.06μM



CTB$_5$ conc = 0.5μM

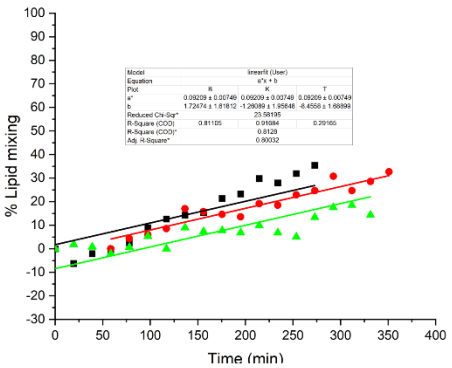


CTB$_5$ conc = 0μM

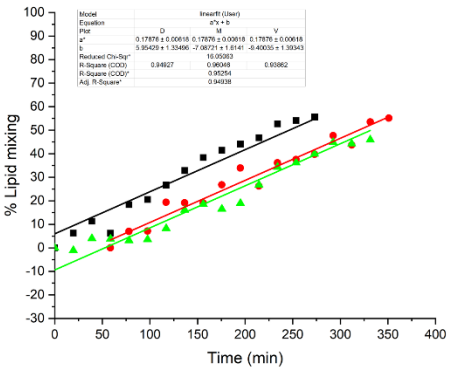


Trivalent- + biotin-CTA2B$_5$

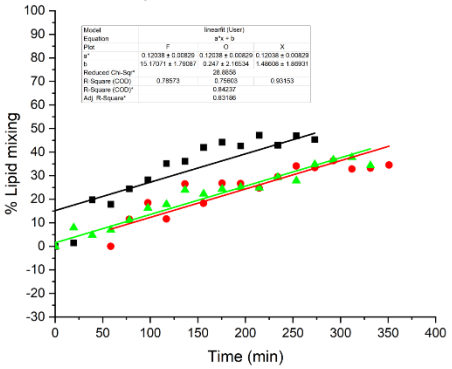
CTB$_5$ conc = 4μM



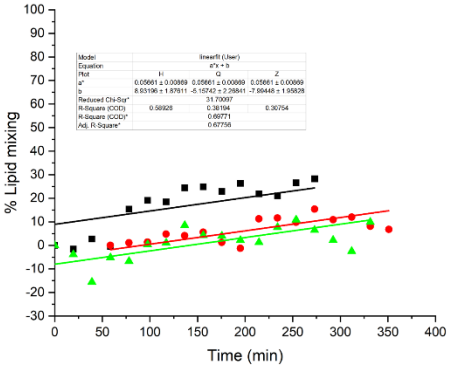
CTB$_5$ conc = 0.25μM



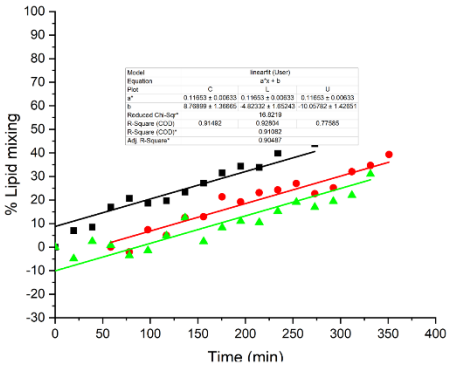
CTB$_5$ conc = 2μM



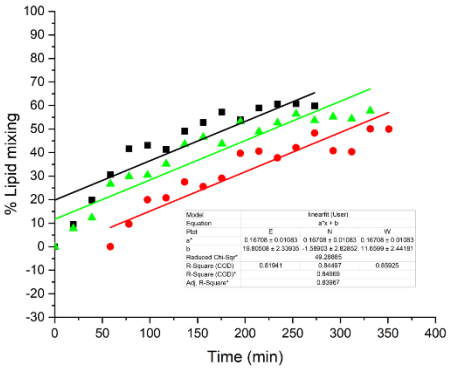
CTB$_5$ conc = 0.125μM



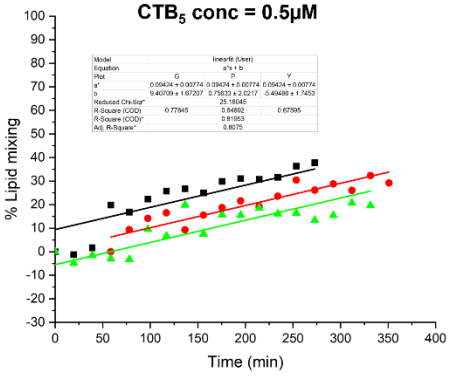
CTB$_5$ conc = 1μM



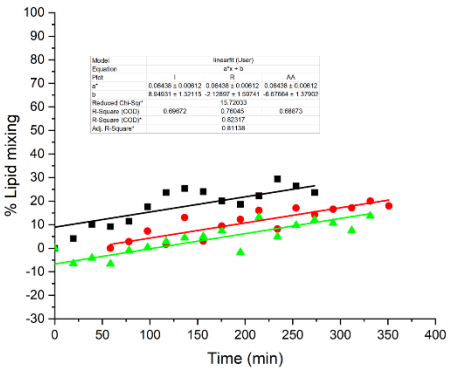
CTB$_5$ conc = 0.06μM



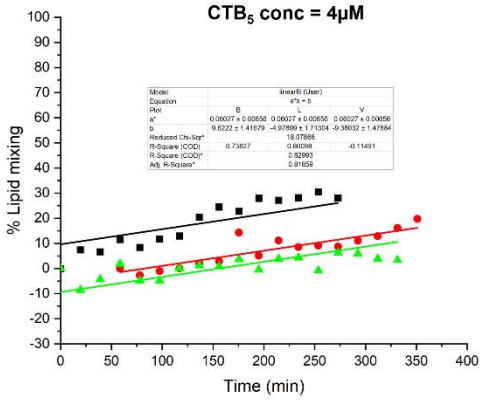
CTB$_5$ conc = 0.5μM

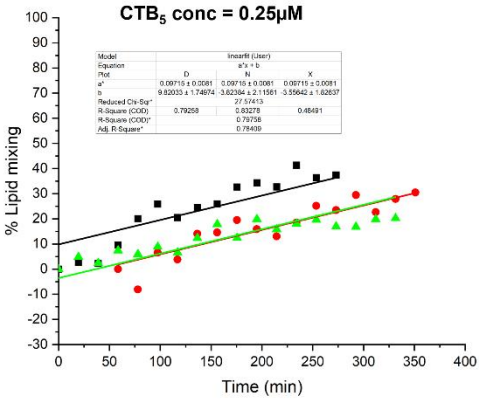


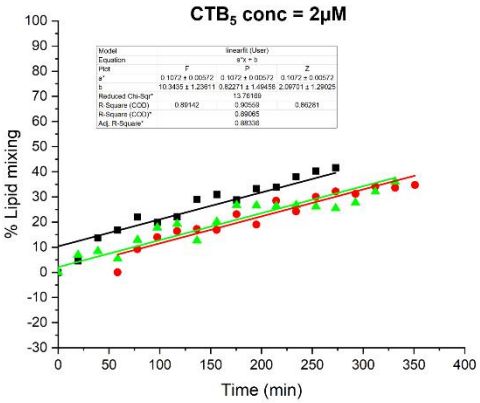
CTB$_5$ conc = 0μM

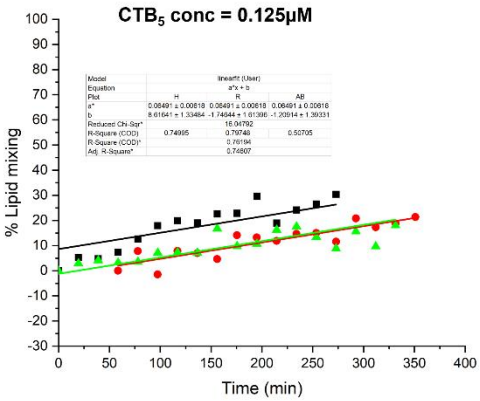


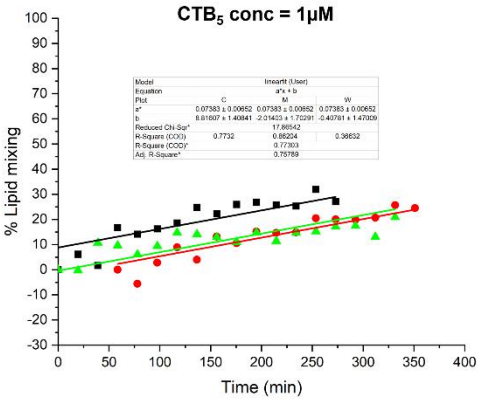
Trans- + biotin-CTA2B$_5$

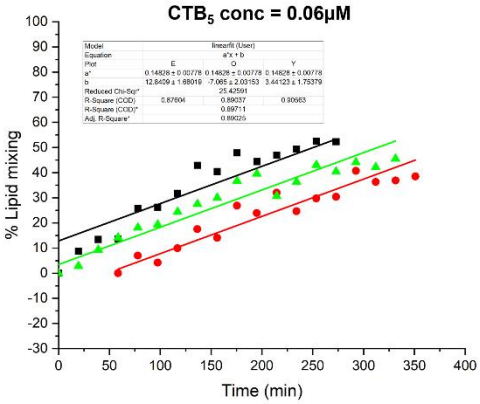


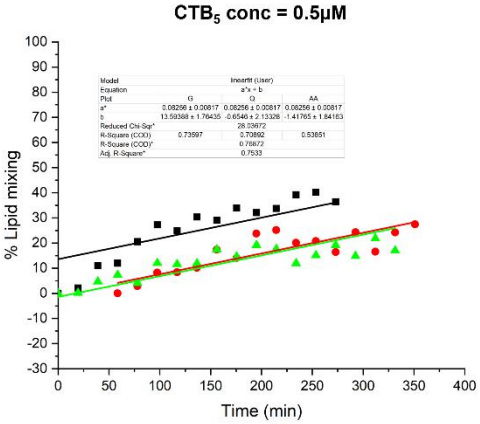


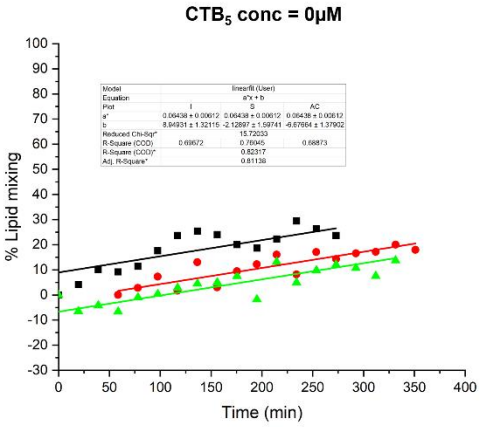




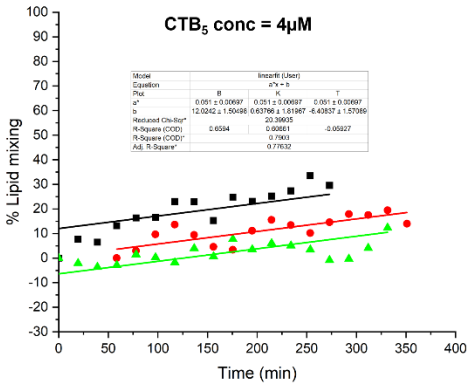


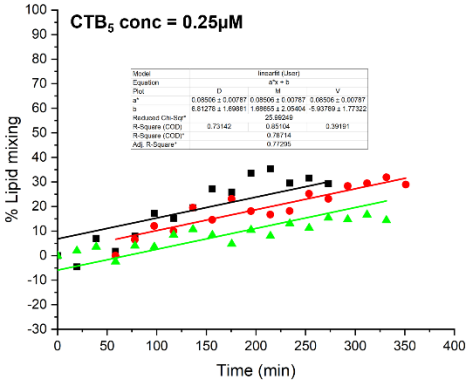


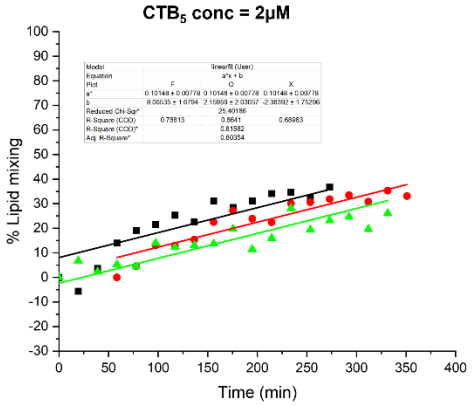


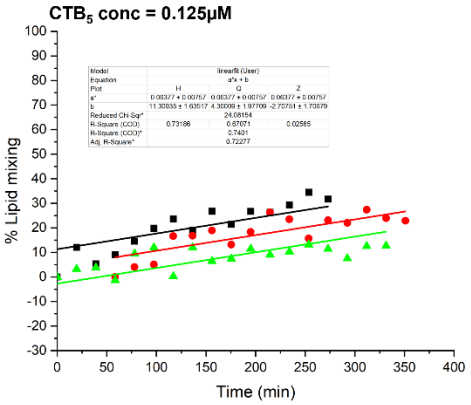


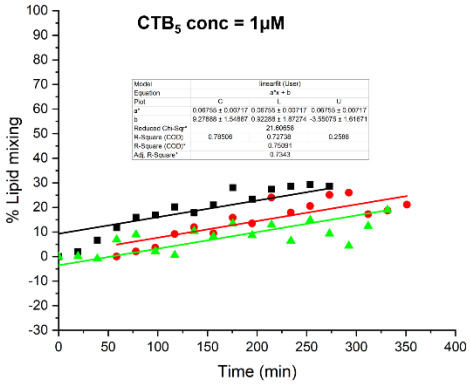
Cis- + biotin-CTA2B$_5$

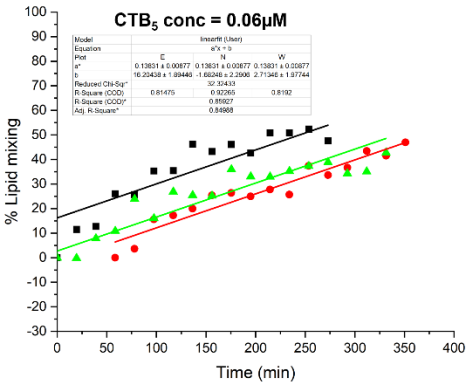


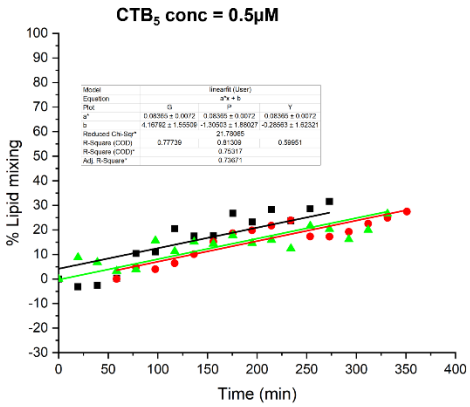


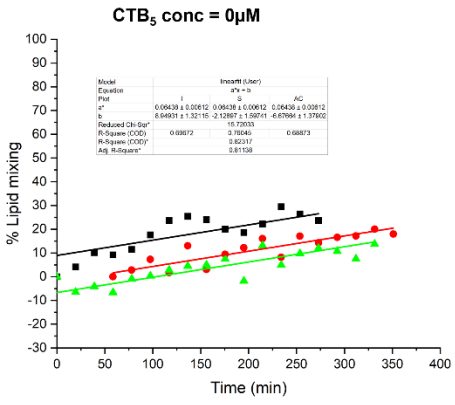




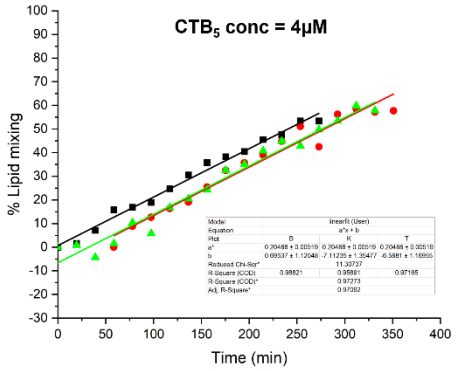


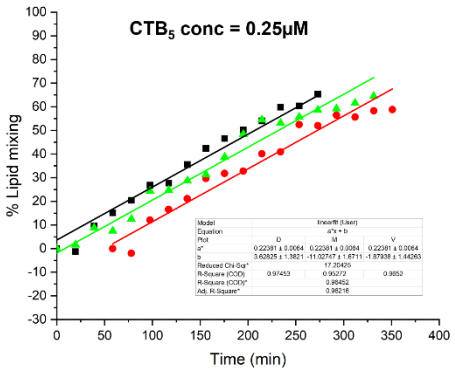


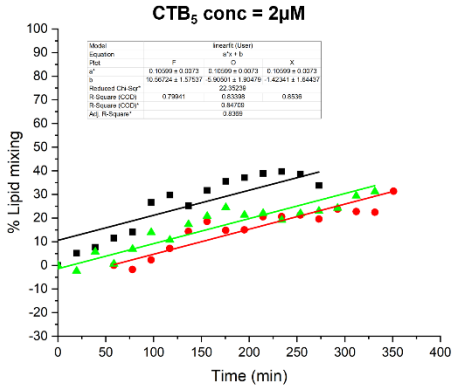


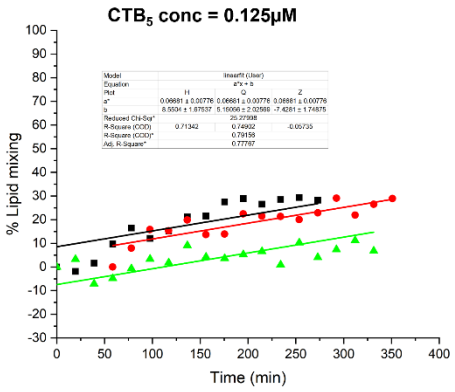


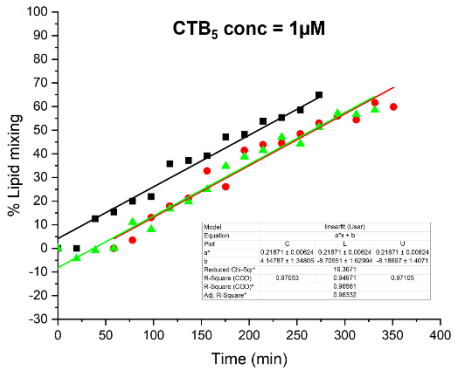
biotin-CTA2B$_5$

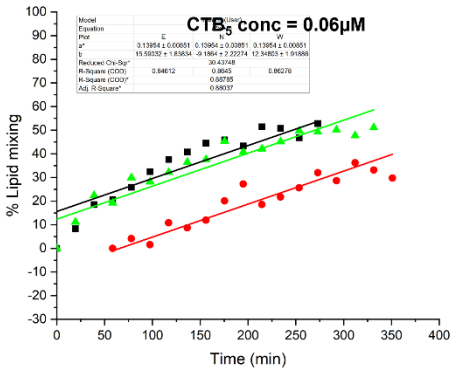


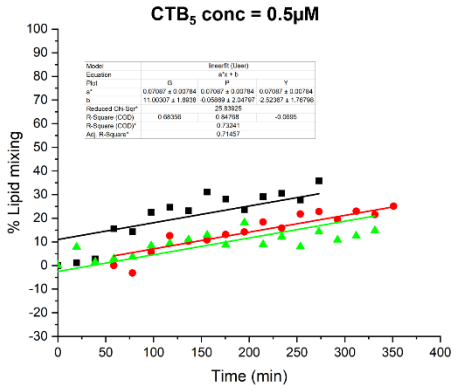


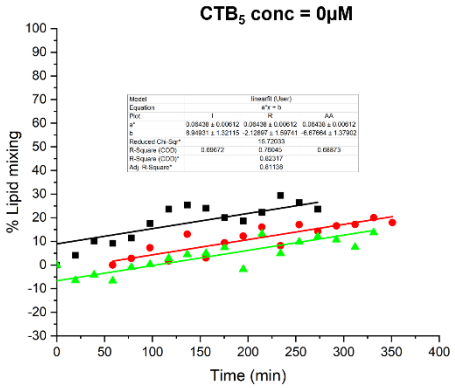




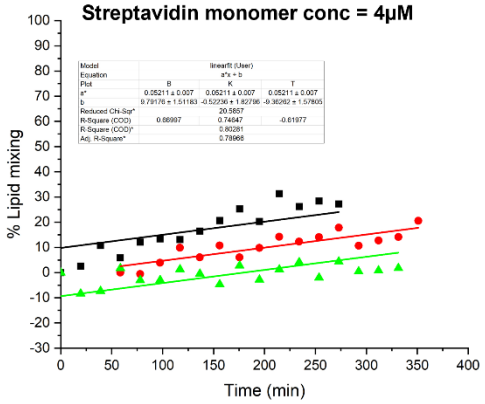


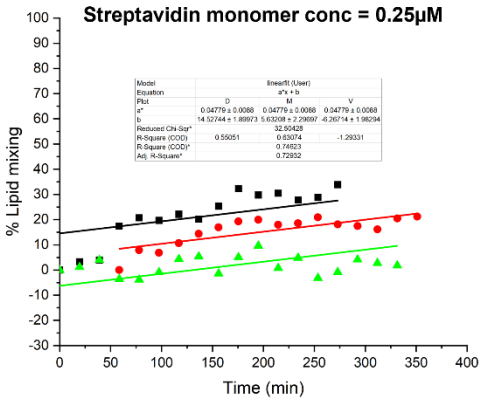


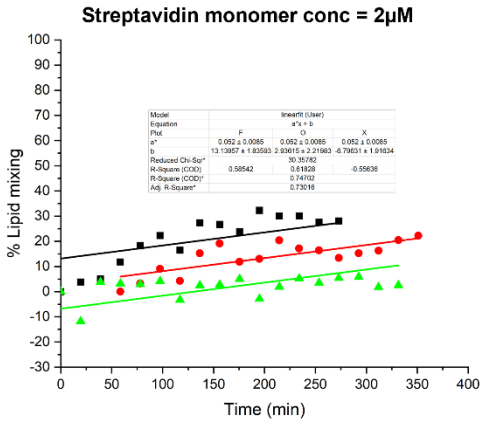


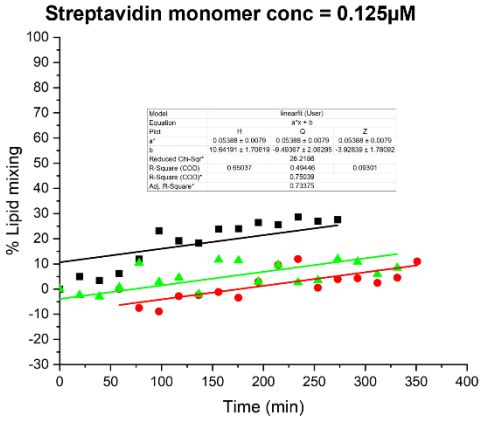


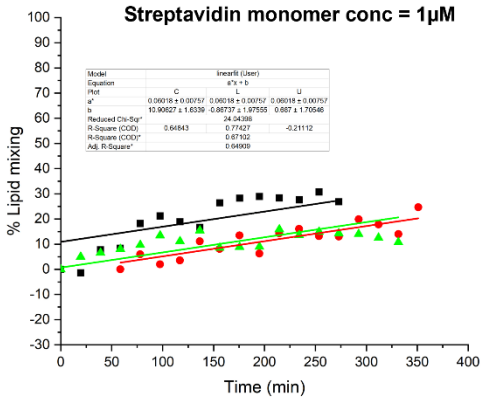
Streptavidin

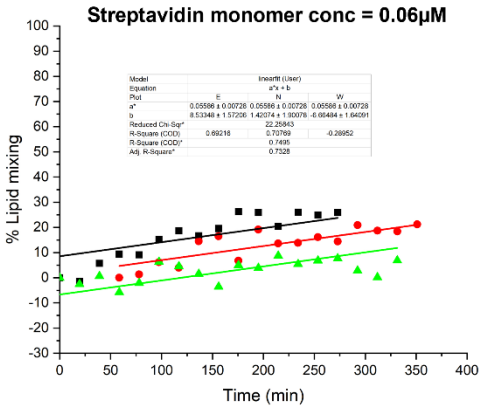


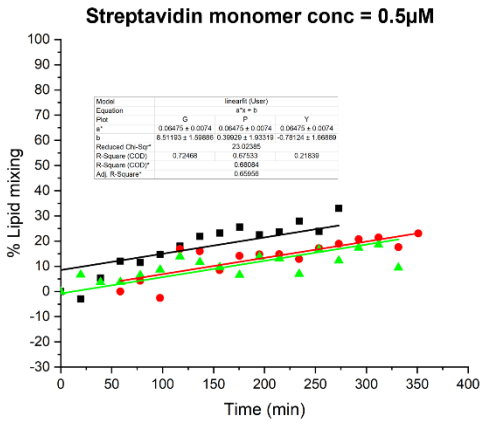












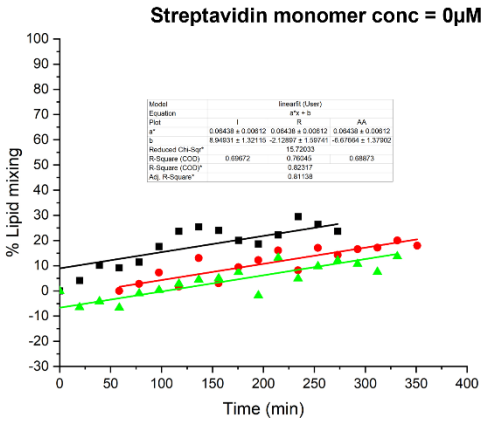


Figure VIII-15 Global fit of three repeats from lipid mixing results. The slope for each plot represents the percentage of lipid mixing (%) per minute, and the value is used to indicate the speed of lipid mixing rate.

Concentration of CTB_5 (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.12911	0.00466
2	0.17314	0.00467
1	0.24465	0.00533
0.5	0.18005	0.01286
0.25	0.11438	0.00777
0.125	0.08399	0.0085
0.0625	0.07112	0.00743
0	0.0511	0.00696

Table VIII-10 Lipid mixing rate for different concentration of mixing streptavidin and biotin-CTA2B_5.

Concentration of CTB_5 (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.09209	0.00749
2	0.11653	0.00633
1	0.17876	0.00618
0.5	0.16708	0.01083
0.25	0.12038	0.00829
0.125	0.09424	0.00774
0.0625	0.05661	0.00869
0	0.06438	0.00612

Table VIII-11 Lipid mixing rate for different concentration of mixing trivalent- streptavidin and biotin-CTA2B_5.

Concentration of CTB_5 (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.06027	0.00656
2	0.07383	0.00652
1	0.09715	0.0081
0.5	0.14828	0.00778
0.25	0.1072	0.00572
0.125	0.08256	0.00817
0.0625	0.06491	0.00618
0	0.06438	0.00612

Table VIII-12 Lipid mixing rate for different concentration of mixing trans- streptavidin and biotin-CTA2B_5.

Concentration of CTB_5 (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.051	0.00697
2	0.06755	0.00717
1	0.08506	0.00787
0.5	0.13831	0.00877
0.25	0.10148	0.00778
0.125	0.08365	0.0072
0.0625	0.06377	0.00757
0	0.06438	0.00612

Table VIII-13 Lipid mixing rate for different concentration of mixing cis- streptavidin and biotin-$CTA2B_5$.

Concentration of streptavidin monomer (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.05211	0.007
2	0.06018	0.00757
1	0.04779	0.0088
0.5	0.05586	0.00728
0.25	0.052	0.0085
0.125	0.06475	0.0074
0.0625	0.05388	0.0079
0	0.06438	0.00612

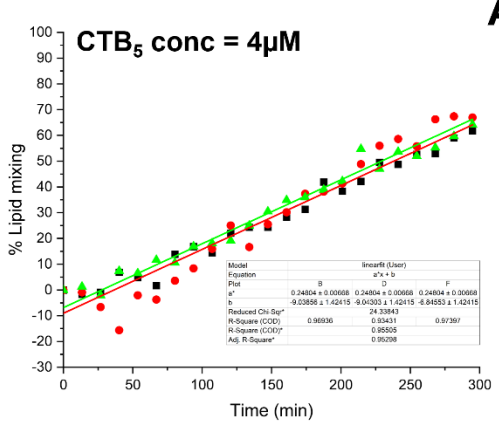
Table VIII-14 Lipid mixing rate for different concentration of streptavidin monomer.

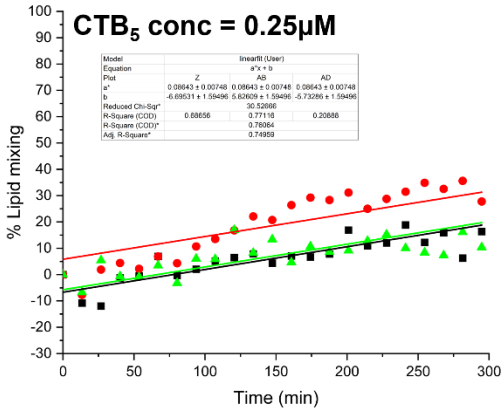
Concentration of streptavidin monomer (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.20488	0.00519
2	0.21871	0.00624
1	0.22381	0.0064
0.5	0.13954	0.00851
0.25	0.10599	0.0073
0.125	0.07087	0.00784
0.0625	0.06681	0.00776

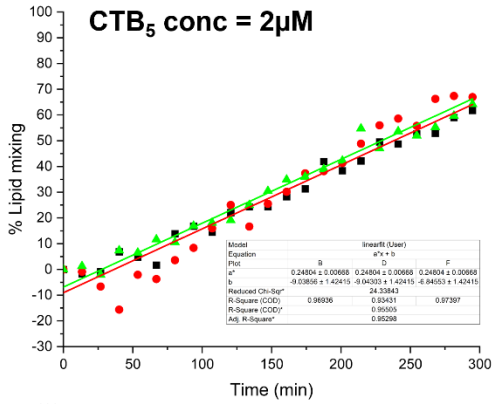
0	0.06438	0.00612

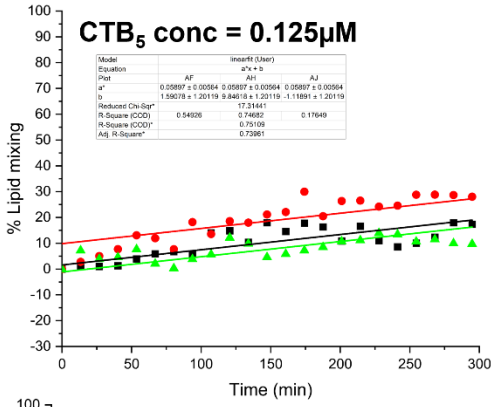
Table VIII-15 Lipid mixing rate for different concentration of biotin-CTA2B$_5$.

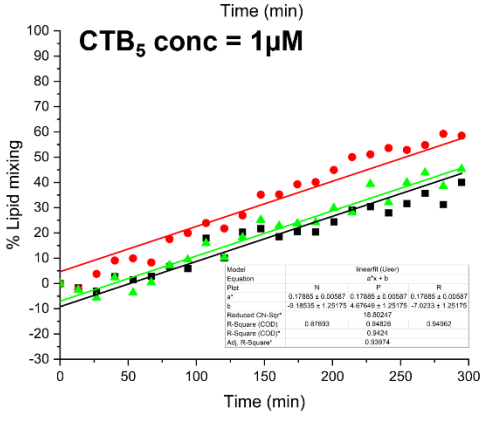
AP2.1

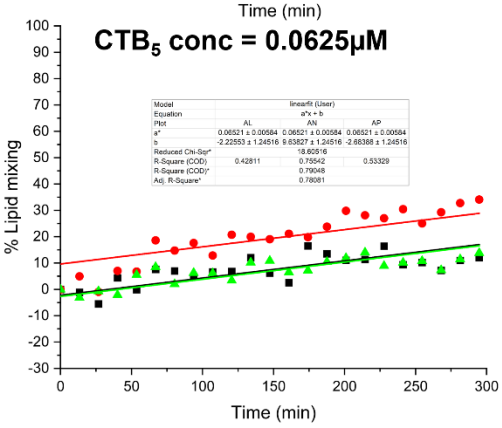

CTB5 conc = 4μM
% Lipid mixing
Time (min)

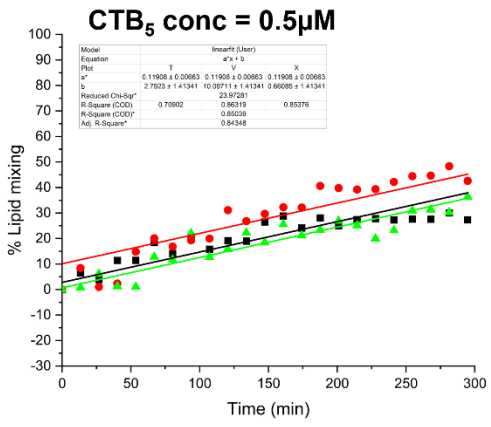

CTB5 conc = 0.25μM
% Lipid mixing
Time (min)

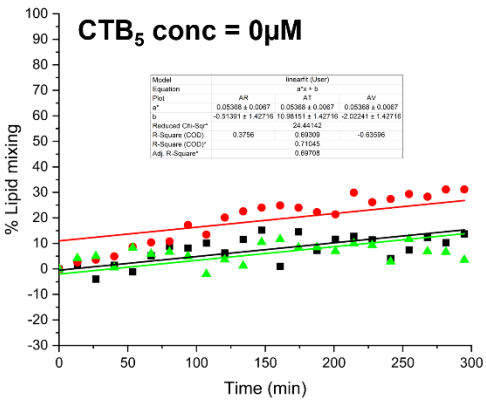

CTB5 conc = 2μM
% Lipid mixing
Time (min)


CTB5 conc = 0.125μM
% Lipid mixing
Time (min)


CTB5 conc = 1μM
% Lipid mixing
Time (min)

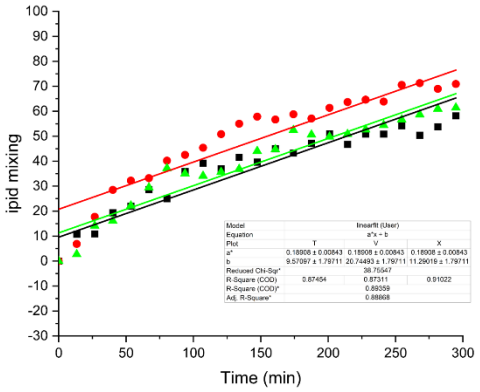

CTB5 conc = 0.0625μM
% Lipid mixing
Time (min)


CTB5 conc = 0.5μM
% Lipid mixing
Time (min)

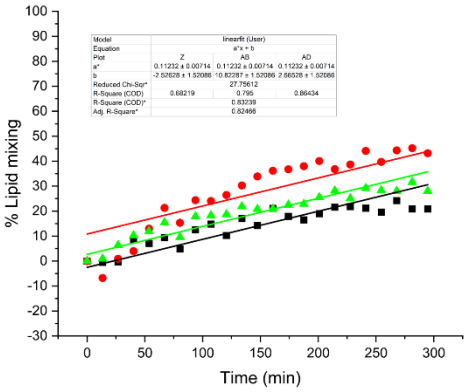

CTB5 conc = 0μM
% Lipid mixing
Time (min)

AP3

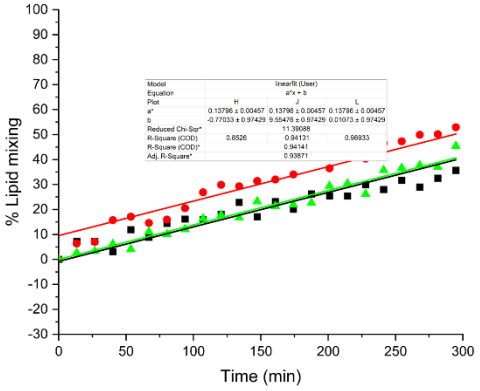
CTB_5 conc = 4μM



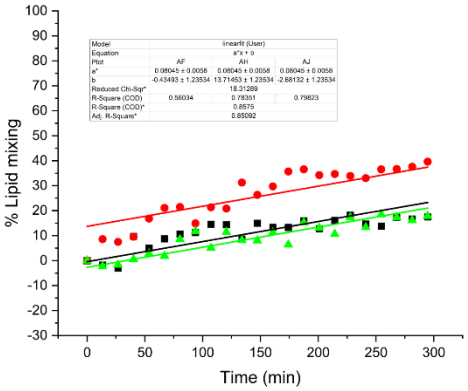
CTB_5 conc = 0.25μM



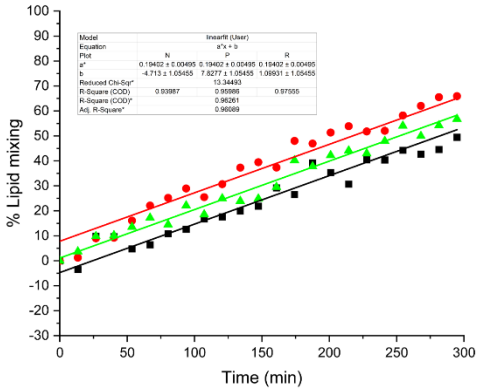
CTB_5 conc = 2μM



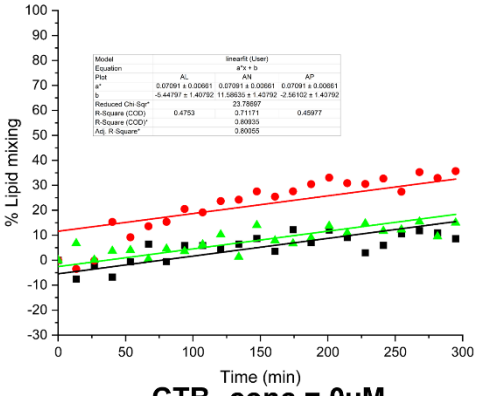
CTB_5 conc = 0.125μM



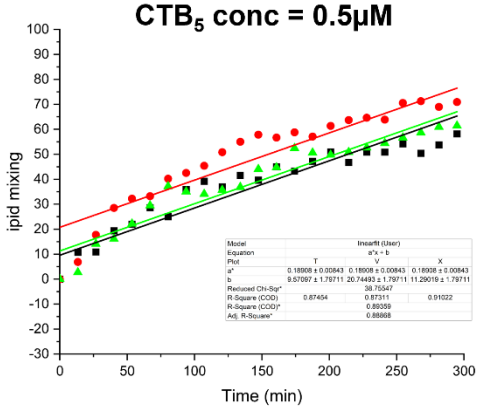
CTB_5 conc = 1μM



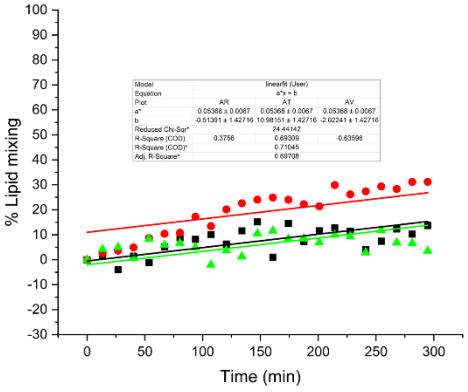
CTB_5 conc = 0.0625μM



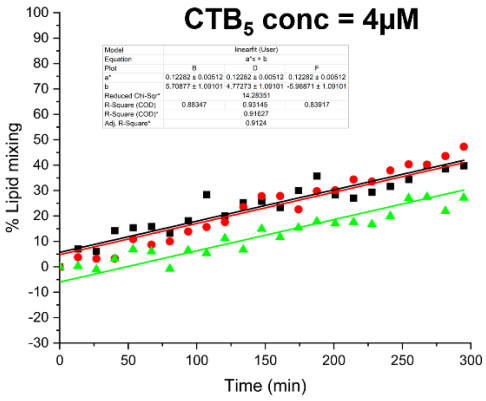
CTB_5 conc = 0.5μM

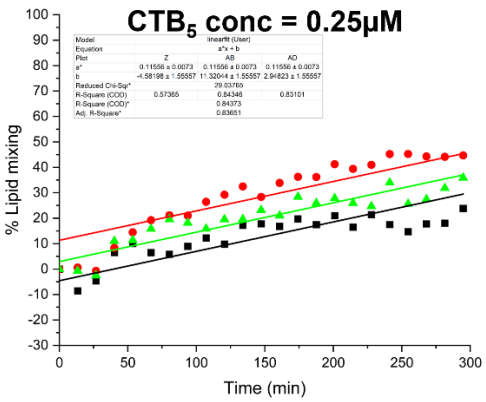


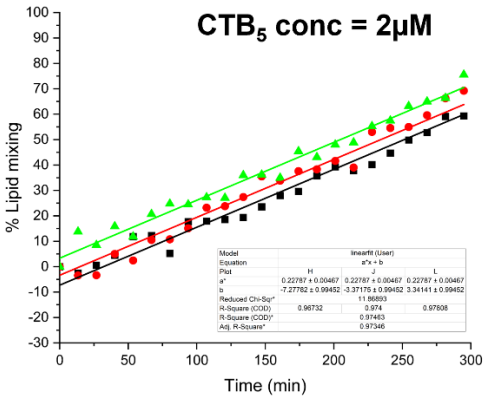
CTB_5 conc = 0μM

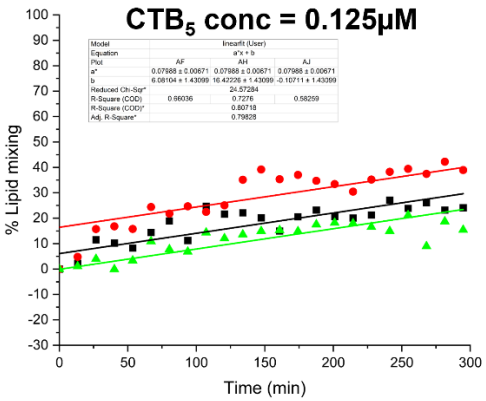


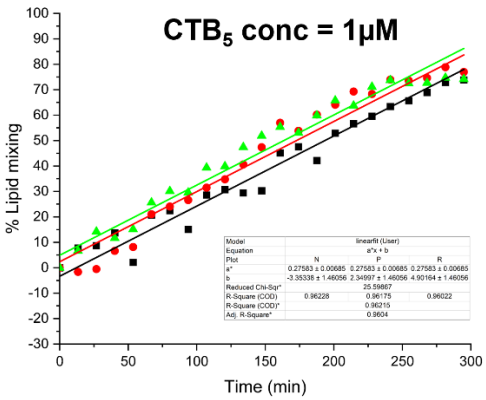
P-PP2.x

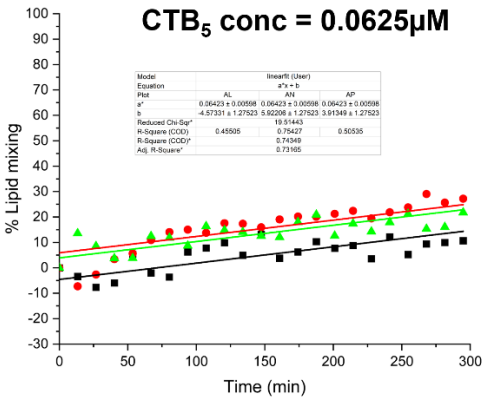

CTB5 conc = 4μM
% Lipid mixing
Time (min)

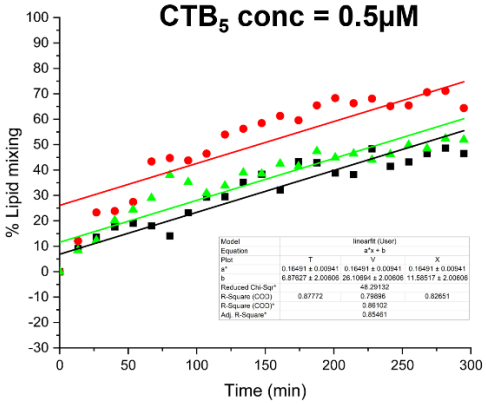

CTB5 conc = 0.25μM
% Lipid mixing
Time (min)

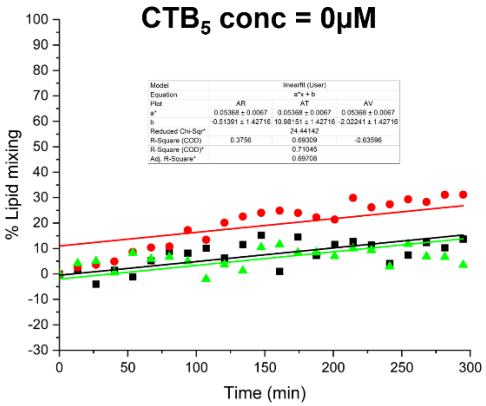

CTB5 conc = 2μM
% Lipid mixing
Time (min)


CTB5 conc = 0.125μM
% Lipid mixing
Time (min)


CTB5 conc = 1μM
% Lipid mixing
Time (min)

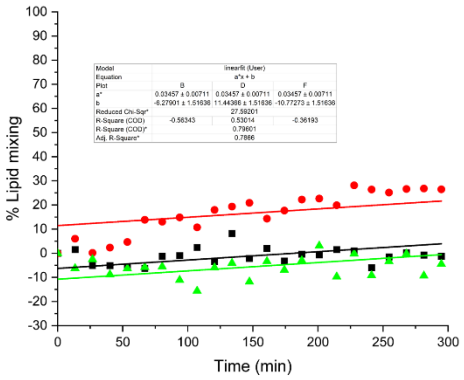

CTB5 conc = 0.0625μM
% Lipid mixing
Time (min)


CTB5 conc = 0.5μM
% Lipid mixing
Time (min)

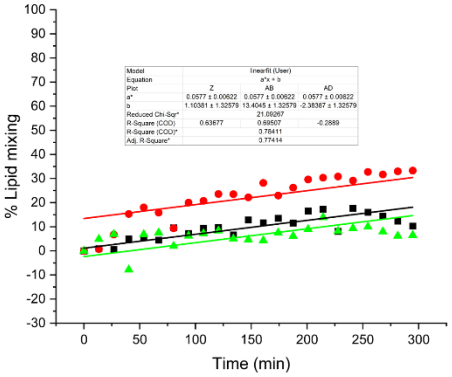

CTB5 conc = 0μM
% Lipid mixing
Time (min)

$CTA2B_5$

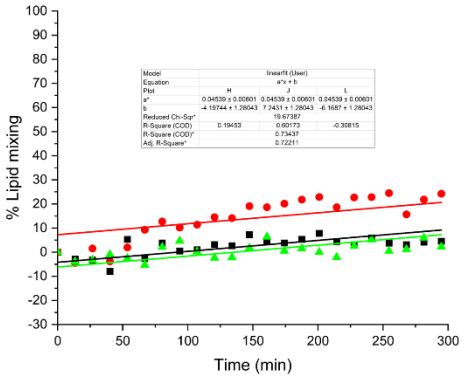
CTB_5 conc = 4 µM



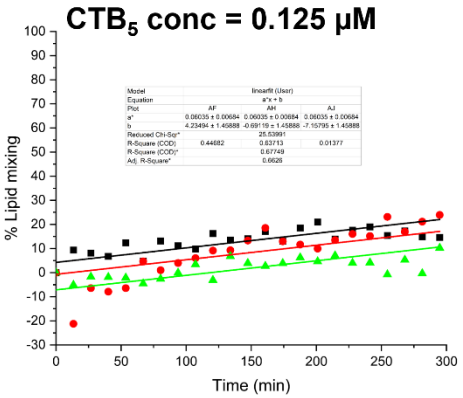
CTB_5 conc = 0.25 µM



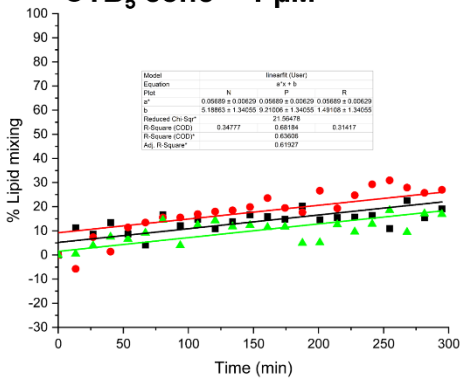
CTB_5 conc = 2 µM



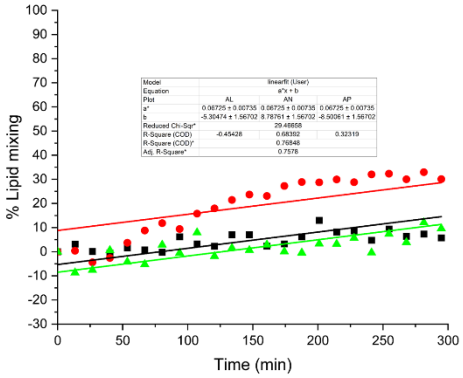
CTB_5 conc = 0.125 µM



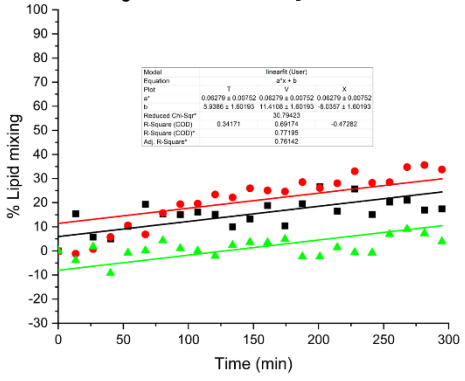
CTB_5 conc = 1 µM



CTB_5 conc = 0.0625 µM



CTB_5 conc = 0.5 µM



CTB_5 conc = 0 µM

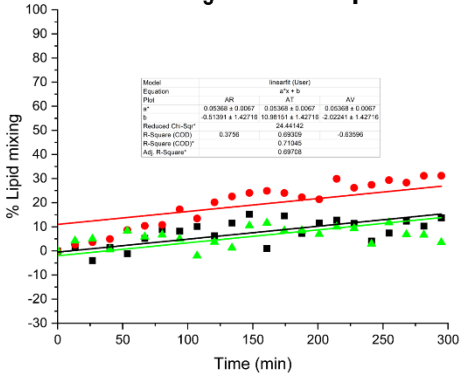


Figure VIII-16. Global fit of three repeats from lipid mixing results. The slope for each plot represents the percentage of lipid mixing (%) per minute, and the value is used to indicate the speed of lipid mixing rate.

Concentration of CTB_5 (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.24804	0.00668
2	0.24208	0.00549
1	0.17885	0.00587
0.5	0.11908	0.00663
0.25	0.08643	0.00748
0.125	0.05897	0.00564
0.0625	0.06521	0.00584
0	0.05368	0.0067

Table VIII-16. Lipid mixing rate for different concentration of AP2.1.

Concentration of CTB_5 (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.09545	0.00605
2	0.13796	0.00457
1	0.19402	0.00495
0.5	0.18908	0.00843
0.25	0.11232	0.00714
0.125	0.08045	0.0058
0.0625	0.07091	0.00661
0	0.05368	0.0067

Table VIII-17. Lipid mixing rate for different concentration of AP3.

Concentration of CTB_5 (μM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.12282	0.00512
2	0.22787	0.00467
1	0.27583	0.00685
0.5	0.16491	0.00941
0.25	0.11556	0.0073
0.125	0.07988	0.00671
0.0625	0.06423	0.00598
0	0.05368	0.0067

Table VIII-18. Lipid mixing rate for different concentration of P-PP2.x.

Concentration of CTB_5 (µM)	Lipid mixing rate (%·min^{-1})	Standard error
4	0.03457	0.00711
2	0.04539	0.00601
1	0.05689	0.00629
0.5	0.06279	0.00752
0.25	0.0577	0.00622
0.125	0.06035	0.00684
0.0625	0.06725	0.00735
0	0.05368	0.0067

Table VIII-19. Lipid mixing rate for different concentration of $CTA2B_5$.

Chapter IX References

1. Mitchell, M.J., M.M. Billingsley, R.M. Haley, M.E. Wechsler, N.A. Peppas, and R. Langer, *Engineering precision nanoparticles for drug delivery.* Nat Rev Drug Discov, 2021. **20**(2): p. 101-124.
2. Anselmo, A.C. and S. Mitragotri, *Nanoparticles in the clinic: An update post COVID-19 vaccines.* Bioeng Transl Med, 2021. **6**(3): p. e10246.
3. Varki, A., *Biological roles of glycans.* Glycobiology, 2017. **27**(1): p. 3-49.
4. Varki, A., *Biological roles of oligosaccharides: all of the theories are correct.* Glycobiology, 1993. **3**(2): p. 97-130.
5. Nguyen, L., K.A. McCord, D.T. Bui, K.M. Bouwman, E.N. Kitova, M. Elaish, D. Kumawat, G.C. Daskhan, I. Tomris, L. Han, P. Chopra, T.J. Yang, S.D. Willows, A.L. Mason, L.K. Mahal, T.L. Lowary, L.J. West, S.D. Hsu, T. Hobman, S.M. Tompkins, G.J. Boons, R.P. de Vries, M.S. Macauley, and J.S. Klassen, *Sialic acid-containing glycolipids mediate binding and viral entry of SARS-CoV-2.* Nat Chem Biol, 2022. **18**(1): p. 81-90.
6. Berois, N., A. Pittini, and E. Osinaga, *Targeting Tumor Glycans for Cancer Therapy: Successes, Limitations, and Perspectives.* Cancers (Basel), 2022. **14**(3).
7. Chiricozzi, E., G. Lunghi, E. Di Biase, M. Fazzari, S. Sonnino, and L. Mauri, *GM1 Ganglioside Is A Key Factor in Maintaining the Mammalian Neuronal Functions Avoiding Neurodegeneration.* International Journal of Molecular Sciences, 2020. **21**(3): p. 868.
8. Nilsson, O., F.T. Brezicka, J. Holmgren, S. Sorenson, L. Svennerholm, F. Yngvason, and L. Lindholm, *Detection of a ganglioside antigen associated with small cell lung carcinomas using monoclonal antibodies directed against fucosyl-GM1.* Cancer Res, 1986. **46**(3): p. 1403-7.
9. Wassarman, P.M. and E.S. Litscher, *Mammalian Fertilization Is Dependent on Multiple Membrane Fusion Events**, in *Cell Fusion: Overviews and Methods*, E.H. Chen, Editor. 2008, Humana Press: Totowa, NJ. p. 99-113.
10. Jahn, R., *Synaptic transmission: two players team up for a new tune.* Curr Biol, 1998. **8**(23): p. R856-8.
11. Ni, F., X. Chen, J. Shen, and Q. Wang, *Structural insights into the membrane fusion mechanism mediated by influenza virus hemagglutinin.* Biochemistry, 2014. **53**(5): p. 846-54.
12. Martens, S. and H.T. McMahon, *Mechanisms of membrane fusion: disparate players and common principles.* Nat Rev Mol Cell Biol, 2008. **9**(7): p. 543-56.

13. Chernomordik, L.V. and M.M. Kozlov, *Protein-lipid interplay in fusion and fission of biological membranes.* Annu Rev Biochem, 2003. **72**: p. 175-207.
14. Grill, D., A.L.L. Matos, W.C. de Vries, S. Kudruk, M. Heflik, W. Dorner, H.D. Mootz, B. Jan Ravoo, H.J. Galla, and V. Gerke, *Bridging of membrane surfaces by annexin A2.* Sci Rep, 2018. **8**(1): p. 14662.
15. Gong, Y., Y. Luo, and D. Bong, *Membrane activation: selective vesicle fusion via small molecule recognition.* J Am Chem Soc, 2006. **128**(45): p. 14430-1.
16. Ma, M., Y. Gong, and D. Bong, *Lipid membrane adhesion and fusion driven by designed, minimally multivalent hydrogen-bonding lipids.* J Am Chem Soc, 2009. **131**(46): p. 16919-26.
17. Ma, M., A. Paredes, and D. Bong, *Intra- and intermembrane pairwise molecular recognition between synthetic hydrogen-bonding phospholipids.* J Am Chem Soc, 2008. **130**(44): p. 14456-8.
18. Gong, Y., M. Ma, Y. Luo, and D. Bong, *Functional determinants of a synthetic vesicle fusion system.* J Am Chem Soc, 2008. **130**(19): p. 6196-205.
19. Joardar, A., G.P. Pattnaik, and H. Chakraborty, *Mechanism of Membrane Fusion: Interplay of Lipid and Peptide.* J Membr Biol, 2022. **255**(2-3): p. 211-224.
20. Kozlovsky, Y. and M.M. Kozlov, *Stalk model of membrane fusion: solution of energy crisis.* Biophys J, 2002. **82**(2): p. 882-95.
21. Brunger, A.T., *Structure and function of SNARE and SNARE-interacting proteins.* Q Rev Biophys, 2005. **38**(1): p. 1-47.
22. Bock, J.B., H.T. Matern, A.A. Peden, and R.H. Scheller, *A genomic perspective on membrane compartment organization.* Nature, 2001. **409**(6822): p. 839-41.
23. Lipka, V., C. Kwon, and R. Panstruga, *SNARE-ware: the role of SNARE-domain proteins in plant biology.* Annu Rev Cell Dev Biol, 2007. **23**: p. 147-74.
24. Valdez-Taubas, J. and H. Pelham, *Swf1-dependent palmitoylation of the SNARE Tlg1 prevents its ubiquitination and degradation.* EMBO J, 2005. **24**(14): p. 2524-32.
25. Jahn, R. and R.H. Scheller, *SNAREs--engines for membrane fusion.* Nat Rev Mol Cell Biol, 2006. **7**(9): p. 631-43.
26. Fasshauer, D., H. Otto, W.K. Eliason, R. Jahn, and A.T. Brunger, *Structural changes are associated with soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor complex formation.* J Biol Chem, 1997. **272**(44): p. 28036-41.

27. Robson Marsden, H., N.A. Elbers, P.H. Bomans, N.A. Sommerdijk, and A. Kros, *A reduced SNARE model for membrane fusion.* Angew Chem Int Ed Engl, 2009. **48**(13): p. 2330-3.
28. Weber, T., B.V. Zemelman, J.A. McNew, B. Westermann, M. Gmachl, F. Parlati, T.H. Sollner, and J.E. Rothman, *SNAREpins: minimal machinery for membrane fusion.* Cell, 1998. **92**(6): p. 759-72.
29. Kashiwada, A., M. Tsuboi, and K. Matsuda, *Target-selective one-way membrane fusion system based on a pH-responsive coiled coil assembly at the interface of liposomal vesicles.* Langmuir, 2011. **27**(4): p. 1403-8.
30. Kashiwada, A., M. Tsuboi, N. Takamura, E. Brandenburg, K. Matsuda, and B. Koksch, *Design and characterization of endosomal-pH-responsive coiled coils for constructing an artificial membrane fusion system.* Chemistry, 2011. **17**(22): p. 6179-86.
31. Davis, J.H., A.J. Rubin, and R.T. Sauer, *Design, construction and characterization of a set of insulated bacterial promoters.* Nucleic Acids Res, 2011. **39**(3): p. 1131-41.
32. Lopez, A. and J. Liu, *DNA Oligonucleotide-Functionalized Liposomes: Bioconjugate Chemistry, Biointerfaces, and Applications.* Langmuir, 2018. **34**(49): p. 15000-15013.
33. Whitehouse, W.L., J.E. Noble, M.G. Ryadnov, and S. Howorka, *Cholesterol Anchors Enable Efficient Binding and Intracellular Uptake of DNA Nanostructures.* Bioconjug Chem, 2019. **30**(7): p. 1836-1844.
34. Ries, O., P.M.G. Loffler, A. Rabe, J.J. Malavan, and S. Vogel, *Efficient liposome fusion mediated by lipid-nucleic acid conjugates.* Org Biomol Chem, 2017. **15**(42): p. 8936-8945.
35. Jumeaux, C., O. Wahlsten, S. Block, E. Kim, R. Chandrawati, P.D. Howes, F. Hook, and M.M. Stevens, *MicroRNA Detection by DNA-Mediated Liposome Fusion.* Chembiochem, 2018. **19**(5): p. 434-438.
36. Stengel, G., R. Zahn, and F. Hook, *DNA-induced programmable fusion of phospholipid vesicles.* J Am Chem Soc, 2007. **129**(31): p. 9584-5.
37. Stengel, G., L. Simonsson, R.A. Campbell, and F. Hook, *Determinants for membrane fusion induced by cholesterol-modified DNA zippers.* J Phys Chem B, 2008. **112**(28): p. 8264-74.
38. Simonsson, L., P. Jonsson, G. Stengel, and F. Hook, *Site-specific DNA-controlled fusion of single lipid vesicles to supported lipid bilayers.* Chemphyschem, 2010. **11**(5): p. 1011-7.
39. Yang, J., A. Bahreman, G. Daudey, J. Bussmann, R.C. Olsthoorn, and A. Kros, *Drug Delivery via Cell Membrane Fusion Using Lipopeptide Modified Liposomes.* ACS Cent Sci, 2016. **2**(9): p. 621-630.

40. Wagle, S., V.N. Georgiev, T. Robinson, R. Dimova, R. Lipowsky, and A. Grafmuller, *Interaction of SNARE Mimetic Peptides with Lipid bilayers: Effects of Secondary Structure, Bilayer Composition and Lipid Anchoring.* Sci Rep, 2019. **9**(1): p. 7708.
41. Kumar, P., S. Guha, and U. Diederichsen, *SNARE protein analog-mediated membrane fusion.* J Pept Sci, 2015. **21**(8): p. 621-9.
42. Daudey, G.A., C. Schwieger, M. Rabe, and A. Kros, *Influence of Membrane-Fusogen Distance on the Secondary Structure of Fusogenic Coiled Coil Peptides.* Langmuir, 2019. **35**(16): p. 5501-5508.
43. Zheng, T., J. Voskuhl, F. Versluis, H.R. Zope, I. Tomatsu, H.R. Marsden, and A. Kros, *Controlling the rate of coiled coil driven membrane fusion.* Chem Commun (Camb), 2013. **49**(35): p. 3649-51.
44. Koukalova, A., S. Pokorna, A.L. Boyle, N. Lopez Mora, A. Kros, M. Hof, and R. Sachl, *Distinct roles of SNARE-mimicking lipopeptides during initial steps of membrane fusion.* Nanoscale, 2018. **10**(40): p. 19064-19073.
45. Crone, N.S.A., D. Minnee, A. Kros, and A.L. Boyle, *Peptide-Mediated Liposome Fusion: The Effect of Anchor Positioning.* Int J Mol Sci, 2018. **19**(1).
46. Pahler, G., C. Panse, U. Diederichsen, and A. Janshoff, *Coiled-coil formation on lipid bilayers--implications for docking and fusion efficiency.* Biophys J, 2012. **103**(11): p. 2295-303.
47. Versluis, F., J. Dominguez, J. Voskuhl, and A. Kros, *Coiled-coil driven membrane fusion: zipper-like vs. non-zipper-like peptide orientation.* Faraday Discuss, 2013. **166**: p. 349-59.
48. Sadek, M., D. Berndt, D. Milovanovic, R. Jahn, and U. Diederichsen, *Distance Regulated Vesicle Fusion and Docking Mediated by beta-Peptide Nucleic Acid SNARE Protein Analogues.* Chembiochem, 2016. **17**(6): p. 479-85.
49. Meyenberg, K., A.S. Lygina, G. van den Bogaart, R. Jahn, and U. Diederichsen, *SNARE derived peptide mimic inducing membrane fusion.* Chem Commun (Camb), 2011. **47**(33): p. 9405-7.
50. Kahne, D., C. Leimkuhler, W. Lu, and C. Walsh, *Glycopeptide and lipoglycopeptide antibiotics.* Chem Rev, 2005. **105**(2): p. 425-48.
51. Beddoe, T., A.W. Paton, J. Le Nours, J. Rossjohn, and J.C. Paton, *Structure, biological functions and applications of the AB5 toxins.* Trends Biochem Sci, 2010. **35**(7): p. 411-8.
52. Paton, J.C. and A.W. Paton, *Pathogenesis and diagnosis of Shiga toxin-producing Escherichia coli infections.* Clin Microbiol Rev, 1998. **11**(3): p. 450-79.

53. Fan, E., E.A. Merritt, C.L. Verlinde, and W.G. Hol, *AB(5) toxins: structures and inhibitor design.* Curr Opin Struct Biol, 2000. **10**(6): p. 680-6.
54. Paton, A.W. and J.C. Paton, *Direct detection and characterization of Shiga toxigenic Escherichia coli by multiplex PCR for stx1, stx2, eae, ehxA, and saa.* J Clin Microbiol, 2002. **40**(1): p. 271-4.
55. Goins, B. and E. Freire, *Thermal stability and intersubunit interactions of cholera toxin in solution and in association with its cell-surface receptor ganglioside GM1.* Biochemistry, 1988. **27**(6): p. 2046-52.
56. Tinker, J.K., J.L. Erbe, W.G. Hol, and R.K. Holmes, *Cholera holotoxin assembly requires a hydrophobic domain at the A-B5 interface: mutational analysis and development of an in vitro assembly system.* Infect Immun, 2003. **71**(7): p. 4093-101.
57. Merritt, E.A., S. Sarfaty, F. van den Akker, C. L'Hoir, J.A. Martial, and W.G. Hol, *Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide.* Protein Sci, 1994. **3**(2): p. 166-75.
58. Heim, J.B., V. Hodnik, J.E. Heggelund, G. Anderluh, and U. Krengel, *Crystal structures of cholera toxin in complex with fucosylated receptors point to importance of secondary binding site.* Sci Rep, 2019. **9**(1): p. 12243.
59. Pezeshkian, W., H. Gao, S. Arumugam, U. Becken, P. Bassereau, J.C. Florent, J.H. Ipsen, L. Johannes, and J.C. Shillcock, *Mechanism of Shiga Toxin Clustering on Membranes.* ACS Nano, 2017. **11**(1): p. 314-324.
60. Pezeshkian, W., L.J. Nabo, and J.H. Ipsen, *Cholera toxin B subunit induces local curvature on lipid bilayers.* FEBS Open Bio, 2017. **7**(11): p. 1638-1645.
61. Turnbull, W.B., B.L. Precious, and S.W. Homans, *Dissecting the cholera toxin-ganglioside GM1 interaction by isothermal titration calorimetry.* J Am Chem Soc, 2004. **126**(4): p. 1047-54.
62. St Hilaire, P.M., M.K. Boyd, and E.J. Toone, *Interaction of the Shiga-like toxin type 1 B-subunit with its carbohydrate receptor.* Biochemistry, 1994. **33**(48): p. 14452-63.
63. Romer, W., L. Berland, V. Chambon, K. Gaus, B. Windschiegl, D. Tenza, M.R. Aly, V. Fraisier, J.C. Florent, D. Perrais, C. Lamaze, G. Raposo, C. Steinem, P. Sens, P. Bassereau, and L. Johannes, *Shiga toxin induces tubular membrane invaginations for its uptake into cells.* Nature, 2007. **450**(7170): p. 670-5.
64. Windschiegl, B., A. Orth, W. Romer, L. Berland, B. Stechmann, P. Bassereau, L. Johannes, and C. Steinem, *Lipid reorganization induced by Shiga toxin clustering on planar membranes.* PLoS One, 2009. **4**(7):

p. e6238.
65. Ling, H., A. Boodhoo, B. Hazes, M.D. Cummings, G.D. Armstrong, J.L. Brunton, and R.J. Read, *Structure of the shiga-like toxin I B-pentamer complexed with an analogue of its receptor Gb3.* Biochemistry, 1998. **37**(7): p. 1777-88.
66. Jobling, M.G., Z. Yang, W.R. Kam, W.I. Lencer, and R.K. Holmes, *A Single Native Ganglioside GM₁-Binding Site Is Sufficient for Cholera Toxin To Bind to Cells and Complete the Intoxication Pathway.* mBio, 2012. **3**(6): p. e00401-12.
67. Krishnan, P., A. Singla, C.A. Lee, J.D. Weatherston, N.C. Worstell, and H.J. Wu, *Hetero-multivalent binding of cholera toxin subunit B with glycolipid mixtures.* Colloids Surf B Biointerfaces, 2017. **160**: p. 281-288.
68. Brown, D.A. and E. London, *Functions of lipid rafts in biological membranes.* Annu Rev Cell Dev Biol, 1998. **14**: p. 111-36.
69. Pezeshkian, W., H. Gao, S. Arumugam, U. Becken, P. Bassereau, J.C. Florent, J.H. Ipsen, L. Johannes, and J.C. Shillcock, *Correction to Mechanism of Shiga Toxin Clustering on Membranes.* ACS Nano, 2018. **12**(2): p. 2079.
70. Schutte, O.M., L.J. Patalag, L.M. Weber, A. Ries, W. Romer, D.B. Werz, and C. Steinem, *2-Hydroxy Fatty Acid Enantiomers of Gb3 Impact Shiga Toxin Binding and Membrane Organization.* Biophys J, 2015. **108**(12): p. 2775-8.
71. Ewers, H., W. Romer, A.E. Smith, K. Bacia, S. Dmitrieff, W. Chai, R. Mancini, J. Kartenbeck, V. Chambon, L. Berland, A. Oppenheim, G. Schwarzmann, T. Feizi, P. Schwille, P. Sens, A. Helenius, and L. Johannes, *GM1 structure determines SV40-induced membrane invagination and infection.* Nat Cell Biol, 2010. **12**(1): p. 11-8; sup pp 1-12.
72. Johannes, L. and W. Romer, *Shiga toxins--from cell biology to biomedical applications.* Nat Rev Microbiol, 2010. **8**(2): p. 105-16.
73. Saslowsky, D.E. and W.I. Lencer, *Conversion of apical plasma membrane sphingomyelin to ceramide attenuates the intoxication of host cells by cholera toxin.* Cell Microbiol, 2008. **10**(1): p. 67-80.
74. Arnaud, J., J. Claudinon, K. Trondle, M. Trovaslet, G. Larson, A. Thomas, A. Varrot, W. Romer, A. Imberty, and A. Audfray, *Reduction of lectin valency drastically changes glycolipid dynamics in membranes but not surface avidity.* ACS Chem Biol, 2013. **8**(9): p. 1918-24.
75. Arnaud, J., K. Trondle, J. Claudinon, A. Audfray, A. Varrot, W. Romer, and A. Imberty, *Membrane deformation by neolectins with engineered glycolipid binding sites.* Angew Chem Int Ed Engl, 2014. **53**(35): p.

9267-70.
76. Villringer, S., J. Madl, T. Sych, C. Manner, A. Imberty, and W. Romer, *Lectin-mediated protocell crosslinking to mimic cell-cell junctions and adhesion.* Sci Rep, 2018. **8**(1): p. 1932.
77. Ribeiro, J.P., S. Villringer, D. Goyard, L. Coche-Guerente, M. Hoferlin, O. Renaudet, W. Romer, and A. Imberty, *Tailor-made Janus lectin with dual avidity assembles glycoconjugate multilayers and crosslinks protocells.* Chem Sci, 2018. **9**(39): p. 7634-7641.
78. Notova, S., L. Siukstaite, F. Rosato, F. Vena, A. Audfray, N. Bovin, L. Landemarre, W. Römer, and A. Imberty, *Extending Janus lectins architecture: Characterization and application to protocells.* Computational and Structural Biotechnology Journal, 2022. **20**: p. 6108-6119.
79. Siukstaite, L., A. Imberty, and W. Romer, *Structural Diversities of Lectins Binding to the Glycosphingolipid Gb3.* Front Mol Biosci, 2021. **8**: p. 704685.
80. Wehrum, S., L. Siukstaite, D.J. Williamson, T.R. Branson, T. Sych, J. Madl, G.C. Wildsmith, W. Dai, E. Kempmann, J.F. Ross, M. Thomsen, M.E. Webb, W. Römer, and W.B. Turnbull, *Membrane Fusion Mediated by Non-covalent Binding of Re-engineered Cholera Toxin Assemblies to Glycolipids.* ACS Synthetic Biology, 2022.
81. Marsden, H.R., I. Tomatsu, and A. Kros, *Model systems for membrane fusion.* Chem Soc Rev, 2011. **40**(3): p. 1572-85.
82. McMahon, H.T., M.M. Kozlov, and S. Martens, *Membrane curvature in synaptic vesicle fusion and beyond.* Cell, 2010. **140**(5): p. 601-5.
83. Grothe, T. and P.J. Walla, *Fluorescence Lifetime and Cross-correlation Spectroscopy for Observing Membrane Fusion of Liposome Models Containing Synaptic Proteins.* Methods Mol Biol, 2022. **2417**: p. 167-180.
84. Mora, N.L., A.L. Boyle, B.J.V. Kolck, A. Rossen, S. Pokorna, A. Koukalova, R. Sachl, M. Hof, and A. Kros, *Controlled Peptide-Mediated Vesicle Fusion Assessed by Simultaneous Dual-Colour Time-Lapsed Fluorescence Microscopy.* Sci Rep, 2020. **10**(1): p. 3087.
85. Arribas Perez, M. and P.A. Beales, *Biomimetic Curvature and Tension-Driven Membrane Fusion Induced by Silica Nanoparticles.* Langmuir, 2021. **37**(47): p. 13917-13931.
86. Witt, H., F. Savic, S. Verbeek, J. Dietz, G. Tarantola, M. Oelkers, B. Geil, and A. Janshoff, *Membrane fusion studied by colloidal probes.* Eur Biophys J, 2021. **50**(2): p. 223-237.
87. Tabaei, S., J. Jackman, M. Kim, S. Yorulmaz Avsar, S. Vafaei, and N. Cho,

Biomembrane Fabrication by the Solvent-assisted Lipid Bilayer (SALB) Method. Journal of Visualized Experiments, 2015. 2015.
88. Khan, M.S., N.S. Dosoky, D. Patel, J. Weimer, and J.D. Williams, *Lipid Bilayer Membrane in a Silicon Based Micron Sized Cavity Accessed by Atomic Force Microscopy and Electrochemical Impedance Spectroscopy.* Biosensors (Basel), 2017. 7(3).
89. Shi, L., K. Howan, Q.T. Shen, Y.J. Wang, J.E. Rothman, and F. Pincet, *Preparation and characterization of SNARE-containing nanodiscs and direct study of cargo release through fusion pores.* Nat Protoc, 2013. 8(5): p. 935-48.
90. François-Martin, C. and F. Pincet, *Actual fusion efficiency in the lipid mixing assay - Comparison between nanodiscs and liposomes.* Scientific Reports, 2017. 7(1): p. 43860.
91. Denisov, I.G. and S.G. Sligar, *Nanodiscs for structural and functional studies of membrane proteins.* Nat Struct Mol Biol, 2016. 23(6): p. 481-6.
92. Mattheyses, A.L., S.M. Simon, and J.Z. Rappoport, *Imaging with total internal reflection fluorescence microscopy for the cell biologist.* J Cell Sci, 2010. 123(Pt 21): p. 3621-8.
93. Fairhead, M., D. Krndija, E.D. Lowe, and M. Howarth, *Plug-and-play pairing via defined divalent streptavidins.* J Mol Biol, 2014. 426(1): p. 199-214.
94. Tiede, C., R. Bedford, S.J. Heseltine, G. Smith, I. Wijetunga, R. Ross, D. AlQallaf, A.P. Roberts, A. Balls, A. Curd, R.E. Hughes, H. Martin, S.R. Needham, L.C. Zanetti-Domingues, Y. Sadigh, T.P. Peacock, A.A. Tang, N. Gibson, H. Kyle, G.W. Platt, N. Ingram, T. Taylor, L.P. Coletta, I. Manfield, M. Knowles, S. Bell, F. Esteves, A. Maqbool, R.K. Prasad, M. Drinkhill, R.S. Bon, V. Patel, S.A. Goodchild, M. Martin-Fernandez, R.J. Owens, J.E. Nettleship, M.E. Webb, M. Harrison, J.D. Lippiat, S. Ponnambalam, M. Peckham, A. Smith, P.K. Ferrigno, M. Johnson, M.J. McPherson, and D.C. Tomlinson, *Affimer proteins are versatile and renewable affinity reagents.* Elife, 2017. 6.
95. Balmforth, M.R., J. Haigh, V. Kumar, W. Dai, C. Tiede, D.C. Tomlinson, J. Deuchars, M.E. Webb, and W.B. Turnbull, *Piggybacking on the Cholera Toxin: Identification of a CTB-Binding Protein as an Approach for Targeted Delivery of Proteins to Motor Neurons.* Bioconjug Chem, 2021. 32(10): p. 2205-2212.
96. Danielewicz, N., W. Dai, F. Rosato, M.E. Webb, G. Striedner, W. Romer, W.B. Turnbull, and J. Mairhofer, *In-Depth Characterization of a Re-Engineered Cholera Toxin Manufacturing Process Using Growth-*

Decoupled Production in Escherichia coli. Toxins (Basel), 2022. **14**(6).
97. Stargardt, P., L. Feuchtenhofer, M. Cserjan-Puschmann, G. Striedner, and J. Mairhofer, *Bacteriophage Inspired Growth-Decoupled Recombinant Protein Production in Escherichia coli.* ACS Synth Biol, 2020. **9**(6): p. 1336-1348.
98. Nojima, Y., K. Iguchi, Y. Suzuki, and A. Sato, *The pH-dependent formation of PEGylated bovine lactoferrin by branched polyethylene glycol (PEG)-N-hydroxysuccinimide (NHS) active esters.* Biol Pharm Bull, 2009. **32**(3): p. 523-6.
99. Burger, K.N.J., *Greasing Membrane Fusion and Fission Machineries.* Traffic, 2000. **1**(8): p. 605-613.
100. Truebestein, L. and T.A. Leonard, *Coiled-coils: The long and short of it.* Bioessays, 2016. **38**(9): p. 903-16.
101. Fraser, R.D., T.P. Macrae, and A. Miller, *The Fourier Transform of the Coiled-Coil Model for Alpha-Keratin. A Correction.* Acta Crystallogr, 1965. **18**: p. 1087.
102. Chothia, C., M. Levitt, and D. Richardson, *Structure of proteins: packing of alpha-helices and pleated sheets.* Proc Natl Acad Sci U S A, 1977. **74**(10): p. 4130-4.
103. Walshaw, J. and D.N. Woolfson, *Socket: a program for identifying and analysing coiled-coil motifs within protein structures.* J Mol Biol, 2001. **307**(5): p. 1427-50.
104. Acharya, A., V. Rishi, J. Moll, and C. Vinson, *Experimental identification of homodimerizing B-ZIP families in Homo sapiens.* J Struct Biol, 2006. **155**(2): p. 130-9.
105. Grigoryan, G. and A.E. Keating, *Structural specificity in coiled-coil interactions.* Curr Opin Struct Biol, 2008. **18**(4): p. 477-83.
106. Lumb, K.J., C.M. Carr, and P.S. Kim, *Subdomain folding of the coiled coil leucine zipper from the bZIP transcriptional activator GCN4.* Biochemistry, 1994. **33**(23): p. 7361-7.
107. Woolfson, D.N., *The design of coiled-coil structures and assemblies.* Adv Protein Chem, 2005. **70**: p. 79-112.
108. Fletcher, J.M., A.L. Boyle, M. Bruning, G.J. Bartlett, T.L. Vincent, N.R. Zaccai, C.T. Armstrong, E.H. Bromley, P.J. Booth, R.L. Brady, A.R. Thomson, and D.N. Woolfson, *A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology.* ACS Synth Biol, 2012. **1**(6): p. 240-50.
109. Testa, O.D., E. Moutevelis, and D.N. Woolfson, *CC+: a relational database of coiled-coil structures.* Nucleic Acids Res, 2009. **37**(Database issue): p. D315-22.

110. Cristie-David, A.S., A. Sciore, S. Badieyan, J.D. Escheweiler, P. Koldewey, J.C. Bardwell, B.T. Ruotolo, and E.N.G. Marsh, *Evaluation of de novo-designed coiled coils as off-the-shelf components for protein assembly.* Molecular Systems Design & Engineering, 2017. **2**(2): p. 140-148.
111. Padilla, J.E., C. Colovos, and T.O. Yeates, *Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments.* Proceedings of the National Academy of Sciences, 2001. **98**(5): p. 2217.
112. Ross, J.F., G.C. Wildsmith, M. Johnson, D.L. Hurdiss, K. Hollingsworth, R.F. Thompson, M. Mosayebi, C.H. Trinh, E. Paci, A.R. Pearson, M.E. Webb, and W.B. Turnbull, *Directed Assembly of Homopentameric Cholera Toxin B-Subunit Proteins into Higher-Order Structures Using Coiled-Coil Appendages.* J Am Chem Soc, 2019. **141**(13): p. 5211-5219.
113. Cristie-David, A.S., P. Koldewey, B.A. Meinen, J.C.A. Bardwell, and E.N.G. Marsh, *Elaborating a coiled-coil-assembled octahedral protein cage with additional protein domains.* Protein Sci, 2018. **27**(11): p. 1893-1900.
114. Cristie-David, A.S., J. Chen, D.B. Nowak, A.L. Bondy, K. Sun, S.I. Park, M.M. Banaszak Holl, M. Su, and E.N.G. Marsh, *Coiled-Coil-Mediated Assembly of an Icosahedral Protein Cage with Extremely High Thermal and Chemical Stability.* J Am Chem Soc, 2019. **141**(23): p. 9207-9216.
115. Gradisar, H. and R. Jerala, *De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers.* J Pept Sci, 2011. **17**(2): p. 100-6.
116. Thomas, F., A.L. Boyle, A.J. Burton, and D.N. Woolfson, *A set of de novo designed parallel heterodimeric coiled coils with quantified dissociation constants in the micromolar to sub-nanomolar regime.* J Am Chem Soc, 2013. **135**(13): p. 5161-6.
117. Fletcher, J.M., G.J. Bartlett, A.L. Boyle, J.J. Danon, L.E. Rush, A.N. Lupas, and D.N. Woolfson, *N@a and N@d: Oligomer and Partner Specification by Asparagine in Coiled-Coil Interfaces.* ACS Chem Biol, 2017. **12**(2): p. 528-538.
118. McClain, D.L., H.L. Woods, and M.G. Oakley, *Design and characterization of a heterodimeric coiled coil that forms exclusively with an antiparallel relative helix orientation.* J Am Chem Soc, 2001. **123**(13): p. 3151-2.
119. O'Neil, K.T. and W.F. DeGrado, *A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids.* Science, 1990. **250**(4981): p. 646-51.
120. Thompson, K.E., C.J. Bashor, W.A. Lim, and A.E. Keating, *SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific*

coiled-coil interaction domains. ACS Synth Biol, 2012. **1**(4): p. 118-29.
121. Anderson, G.P., L.C. Shriver-Lake, J.L. Liu, and E.R. Goldman, *Orthogonal Synthetic Zippers as Protein Scaffolds.* ACS Omega, 2018. **3**(5): p. 4810-4815.
122. Patterson, D.P., M. Su, T.M. Franzmann, A. Sciore, G. Skiniotis, and E.N. Marsh, *Characterization of a highly flexible self-assembling protein system designed to form nanocages.* Protein Sci, 2014. **23**(2): p. 190-9.
123. Petrone, P.M., C.D. Snow, D. Lucent, and V.S. Pande, *Side-chain recognition and gating in the ribosome exit tunnel.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16549-54.
124. BioLabs, N.E. *Q5® Site-Directed Mutagenesis Kit Protocol (E0554)*. 2013 04/December/2022]; Available from: https://international.neb.com/protocols/2013/01/26/q5-site-directed-mutagenesis-kit-protocol-e0554.
125. Wood, C.W. and D.N. Woolfson, *CCBuilder 2.0: Powerful and accessible coiled-coil modeling.* Protein Sci, 2018. **27**(1): p. 103-111.
126. Gurnon, D.G., J.A. Whitaker, and M.G. Oakley, *Design and characterization of a homodimeric antiparallel coiled coil.* J Am Chem Soc, 2003. **125**(25): p. 7518-9.
127. Wood, C.W., J.W. Heal, A.R. Thomson, G.J. Bartlett, A.A. Ibarra, R.L. Brady, R.B. Sessions, and D.N. Woolfson, *ISAMBARD: an open-source computational environment for biomolecular analysis, modelling and design.* Bioinformatics, 2017. **33**(19): p. 3043-3050.
128. Plaper, T., J. Aupic, P. Dekleva, F. Lapenta, M.M. Keber, R. Jerala, and M. Bencina, *Coiled-coil heterodimers with increased stability for cellular regulation and sensing SARS-CoV-2 spike protein-mediated cell fusion.* Sci Rep, 2021. **11**(1): p. 9136.
129. Chatzopoulou, M. and N.M. Glykos, *Geometric modeling of coiled-coils using Isambard: the case of the RM6 variant of the Repressor of Primer protein.*
130. Rhys, G.G., C.W. Wood, E.J.M. Lang, A.J. Mulholland, R.L. Brady, A.R. Thomson, and D.N. Woolfson, *Maintaining and breaking symmetry in homomeric coiled-coil assemblies.* Nat Commun, 2018. **9**(1): p. 4132.
131. Maier, J.A., C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, and C. Simmerling, *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* J Chem Theory Comput, 2015. **11**(8): p. 3696-713.
132. MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen,

B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* J Phys Chem B, 1998. 102(18): p. 3586-616.
133. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N·log (N) method for Ewald sums in large systems.* The Journal of chemical physics, 1993. 98(12): p. 10089-10092.
134. Hammonds, K.D. and J.-P. Ryckaert, *On the convergence of the SHAKE algorithm.* Computer physics communications, 1991. 62(2-3): p. 336-351.
135. Guo, Z. and D. Cremer, *Methods for a Rapid and Automated Description of Proteins: Protein Structure, Protein Similarity, and Protein Folding.* Reviews in Computational Chemistry, 2016. 29: p. 369-438.
136. Tatulian, S.A., *Determination of helix orientations in proteins.* Comput Biol Chem, 2008. 32(5): p. 370-4.
137. Serrano, A., J.L. Guyette, J.B. Heim, M. Taylor, P. Cherubin, U. Krengel, K. Teter, and S.A. Tatulian, *Holotoxin disassembly by protein disulfide isomerase is less efficient for Escherichia coli heat-labile enterotoxin than cholera toxin.* Sci Rep, 2022. 12(1): p. 34.
138. Wang, T., E.A. Smith, E.R. Chapman, and J.C. Weisshaar, *Lipid mixing and content release in single-vesicle, SNARE-driven fusion assay with 1-5 ms resolution.* Biophys J, 2009. 96(10): p. 4122-31.
139. Oude Blenke, E.E., J. van den Dikkenberg, B. van Kolck, A. Kros, and E. Mastrobattista, *Coiled coil interactions for the targeting of liposomes for nucleic acid delivery.* Nanoscale, 2016. 8(16): p. 8955-65.
140. Daudey, G.A., M. Shen, A. Singhal, P. van der Est, G.J.A. Sevink, A.L. Boyle, and A. Kros, *Liposome fusion with orthogonal coiled coil peptides as fusogens: the efficacy of roleplaying peptides.* Chem Sci, 2021. 12(41): p. 13782-13792.
141. Gimenez-Andres, M., A. Copic, and B. Antonny, *The Many Faces of Amphipathic Helices.* Biomolecules, 2018. 8(3).
142. Tiede, C., A.A. Tang, S.E. Deacon, U. Mandal, J.E. Nettleship, R.L. Owen, S.E. George, D.J. Harrison, R.J. Owens, D.C. Tomlinson, and M.J. McPherson, *Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications.* Protein Eng Des Sel, 2014. 27(5): p. 145-55.
143. Yang, B. and I. Kwon, *Chemical Modification of Cysteine with 3-Arylpropriolonitrile Improves the In Vivo Stability of Albumin-Conjugated Urate Oxidase Therapeutic Protein.* Biomedicines, 2021.

9(10).
144. Wang, C.L., J.M. Chalovich, P. Graceffa, R.C. Lu, K. Mabuchi, and W.F. Stafford, *A long helix from the central region of smooth muscle caldesmon.* J Biol Chem, 1991. **266**(21): p. 13958-63.
145. Wall, A., A.G. Wills, N. Forte, C. Bahou, L. Bonin, K. Nicholls, M.T. Ma, V. Chudasama, and J.R. Baker, *One-pot thiol-amine bioconjugation to maleimides: simultaneous stabilisation and dual functionalisation.* Chem Sci, 2020. **11**(42): p. 11455-11460.
146. Wilschut, J. and D. Papahadjopoulos, *Ca2+-induced fusion of phospholipid vesicles monitored by mixing of aqueous contents.* Nature, 1979. **281**(5733): p. 690-2.
147. Chan, Y.H., B. van Lengerich, and S.G. Boxer, *Lipid-anchored DNA mediates vesicle fusion as observed by lipid and content mixing.* Biointerphases, 2008. **3**(2): p. FA17.
148. Meers, P., S. Ali, R. Erukulla, and A.S. Janoff, *Novel inner monolayer fusion assays reveal differential monolayer mixing associated with cation-dependent membrane fusion.* Biochim Biophys Acta, 2000. **1467**(1): p. 227-43.
149. Xu, Y., F. Zhang, Z. Su, J.A. McNew, and Y.K. Shin, *Hemifusion in SNARE-mediated membrane fusion.* Nat Struct Mol Biol, 2005. **12**(5): p. 417-22.
150. Polayes, D.A., T.D. Parks, S.A. Johnston, and W.G. Dougherty, *Application of TEV Protease in Protein Production.* Methods Mol Med, 1998. **13**: p. 169-83.
151. Wu, D. and G. Piszczek, *Rapid Determination of Antibody-Antigen Affinity by Mass Photometry.* J Vis Exp, 2021(168).
152. Madl, J., S. Villringer, and W. Römer, *Delving into Lipid-Driven Endocytic Mechanisms Using Biomimetic Membranes*, in *Chemical and Synthetic Approaches in Membrane Biology*, A.K. Shukla, Editor. 2017, Springer New York: New York, NY. p. 17-36.
153. Krivov, G.G., M.V. Shapovalov, and R.L. Dunbrack, Jr., *Improved prediction of protein side-chain conformations with SCWRL4.* Proteins, 2009. **77**(4): p. 778-95.
154. McIntosh-Smith, S., J. Price, R.B. Sessions, and A.A. Ibarra, *High performance in silico virtual drug screening on many-core processors.* Int J High Perform Comput Appl, 2015. **29**(2): p. 119-134.