An Assortment of Peptide Related Topics: Encapsulation of Transmembrane Helices, Heterochiral Quaternary Contacts, and Thioamide Containing Antigens

by

John Alphonse Mannone

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2025

Date of final oral examination: May 20th, 2025

The dissertation is approved by the following members of the Final Oral Committee:

Samuel Gellman, Professor, Chemistry

Jeffrey Martell, Assistant Professor, Chemistry

Katrina Forest, Professor, Bacteriology

Ive Hermans, Professor, Chemistry

Dedicated to my friends and family

Acknowledgments:

Before continuing into discussing the topics of dissertation, I have a plethora of people who have provided guidance and support to me during the grueling journey to complete this PhD that I cannot go without thanking:

Sam Gellman – You've been an exceptional mentor. This journey has been immensely difficult, and it's always meant a lot to me that no matter what was going on you've had my back. You've pushed me to grow and get out of my comfort zone. I truly would not be the scientist or even the person I am if I hadn't met you. As I said last we spoke "thank you for everything".

Mom and Dad – Thank you both for always supporting me. Not just in my academics but in everything I do. I don't say it enough but I'm really thankful to have both you.

Sophie – You are a hallmark definition of an annoying little sister, but I'm always happy when you call and we have our chats.

Beau, Jairo, Eric, Pat, Meg, Connor, Ray – my homies who have commiserated with me through this PhD ordeal. I could not have made it through this without you all. Thank you for all the late nights hangouts and giving me a reason to leave lab.

Brad, Mitch, and Chris – I love you goons. I will see you in Munich for Oktoberfest (this is a threat).

Dom, Rachel, Will, Dan, and Porti - You guys are some of my oldest friends and I love coming back to you in NY. Till next time, hopefully somewhere else than Bushwick…

Mathilde – L’amour de ma vie. Thank you for being one of the best things that has ever stumbled into my life. You’ve helped me learn not to take everything so seriously, and when I inevitably still take it too seriously you always are there for me. I love you so much and I can’t wait for our German chapter. I’ll be eagerly awaiting our reunion in Munich, ma chérie.

Table of Contents:

Chapter I

Development of a Protein Scaffold to Encapsulate and Study Helical Transmembrane Domains within a Membrane-Mimetic Environment

1.1 – Abstract

Proteins embedded in and spanning across the cellular membrane play critical roles in regulating biological processes and transferring signals across the membrane. For many membrane proteins oligomerization, via their transmembrane domains (TMDs), is required to access their biologically active conformation. The ability to study the structural features that promote oligomerization within cellular membranes would grant insight into membrane protein function. However, the development of methods to allow for high-resolution structural determination of membrane protein interactions, particularly the quaternary contacts formed within the cellular membrane, has been a long-standing challenge. The hydrophobic nature of the cellular membrane complicates the expression and isolation of membrane proteins, as many most methods of protein expression and isolation have been designed for aqueous environments.

In this study, the Saposin peptides (Saposin A and Saposin B) which have been adapted for the encapsulation of helical TMDs within a membrane-mimetic environment are explored. Our group had previously developed a Maltose binding protein – Saposin A Hybrid (MSH_A) platform which covalently binds a TMD of interest to the picodisc region used for encapsulation. The MSH platform was used to interrogate TMD interactions within the receptor tyrosine kinase and B-Cell Lymphoma 2 (BCL-2) transmembrane interactomes. Additionally, a second generation of the MSH_A platform containing Saposin B (MSH_B) was developed. Using the MSH_B platform the BCL-2 TMD interactions are further explored, as well as the trimeric TMD complex of the severe acute respiratory syndrome 2 spike protein. Non-covalent encapsulation approaches of TMDs are also explored using Saposin A.

1.2 – Introduction

Proteins embedded in and spanning across the cellular membrane (integral membrane proteins) play critical roles in biological processes involving signal transduction across the membrane[1,2]. Many of these processes are dependent upon the association of multiple membrane proteins to access a biologically active quaternary structure[3,4]. Many cases of membrane protein oligomerization are facilitated by single-pass transmembrane α-helices[5-7]. These are protein domains that contain a central region of ~30 residues that are comprised predominantly of hydrophobic residues. The domains fold into an α-helical secondary structure within the lipid environment and span across the membrane connecting the intra- and extra-cellular portions of the protein[8,9]. During membrane protein oligomerization, the transmembrane helices will engage with each other to stabilize the protein assembly. A well-studied class of proteins known to engage in this manner are receptor tyrosine kinases (RTKs), which require dimerization via their transmembrane (TM) helices to form the bioactive state of the receptor. Once assembled, these receptors can then bind their cognate ligands on the extracellular face of the membrane, inducing a conformational change through the transmembrane domain (TMD) to the intracellular region and induce phosphorylation. This increase in phosphorylation is communicated to the cell as receptor activation (**Figure 1.1**)[6,10].

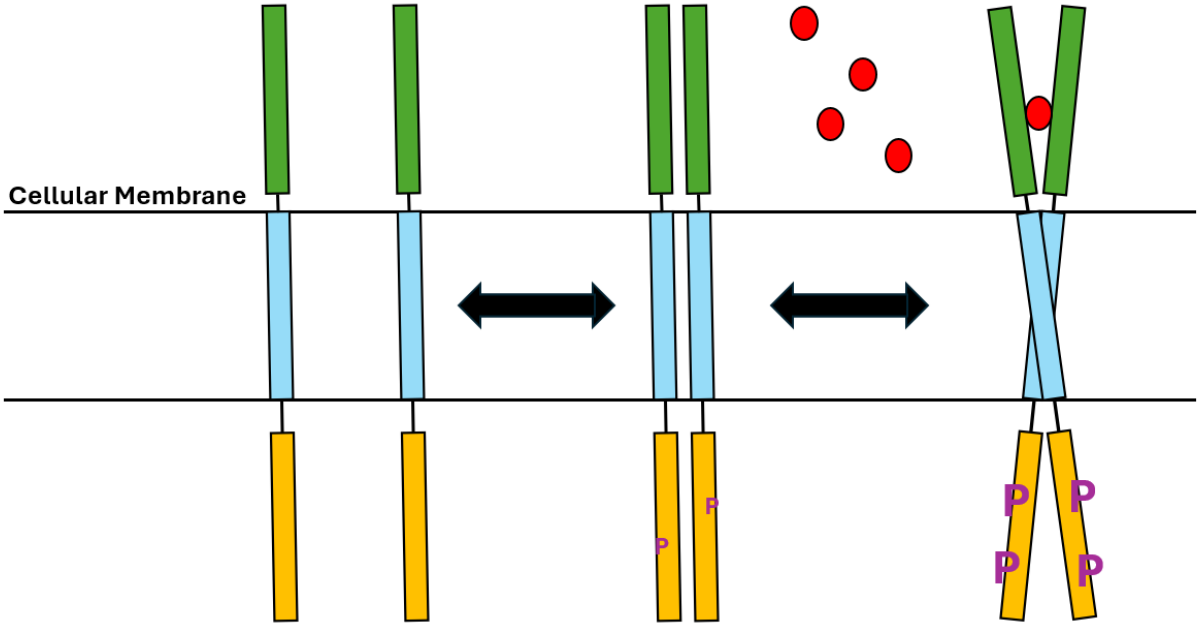


Figure 1.1: General mechanism for the activation of RTKs. Monomeric SPTMs dimerize which can then bind their cognate ligand on the extracellular face of the membrane. Once bound a ligand is bound in the extracellular domain, a conformational change is induced through the TMD to the intracellular face which manifests as increased levels of phosphorylation.

There are multiple examples of disease-associated mutations within the TMD of membrane proteins that result in altered levels of receptor activation and overall function such as: Fibroblast Growth Factor Receptor 3 (FGFR3) G380R – *Achondroplasia*[11], FGFR3 A391E – *Crouzon Syndrome*[12], Human Epidermal Growth Factor 2 (HER2) G660D and V659E – *non-small cell lung cancer*[13]. As these mutations are located within the TMD rather than the ligand binding domain or the intracellular phosphorylation domain, for the mutations to induce a change in communication across the membrane there must be conformational changes within the membrane region. This feature of being able to pass information through the membrane has made membrane proteins a popular topic in human health. Over 60% of FDA-approved drugs currently target some membrane protein[14]. Understanding the fundamental interactions that drive transmembrane (TM) association would provide

invaluable insight into how endogenous membrane proteins associate, function, and facilitate communication between the extra- and intra-cellular regions.

Techniques to study membrane proteins are underdeveloped when compared to their water-soluble counterparts. The challenges associated with studying membrane proteins are largely due to two intrinsic features of membrane proteins: low abundance of the membrane proteins[15,16] and the intrinsic duality of the native environments that these proteins are localized in (some portions within the hydrophobic lipid membrane and other regions exposed to aqueous solution)[17-19]. Both features have led to historic difficulties in obtaining enough protein for characterization and subsequent studies. To address the issue of abundance, recombinant protein expression is often employed. However, extraction and isolation of membrane proteins can often be tedious or impossible. Typical methods to isolate proteins from expression systems involve separations based on nickel affinity, maltose binding affinity, ion exchange, or size exclusion chromatography. These approaches can be hindered with aggregation or denaturing that can occur due to the aqueous conditions required for standard purification techniques coupled with the membrane protein's hydrophobic nature. This issue is commonly addressed by utilizing detergents, as their amphipathic

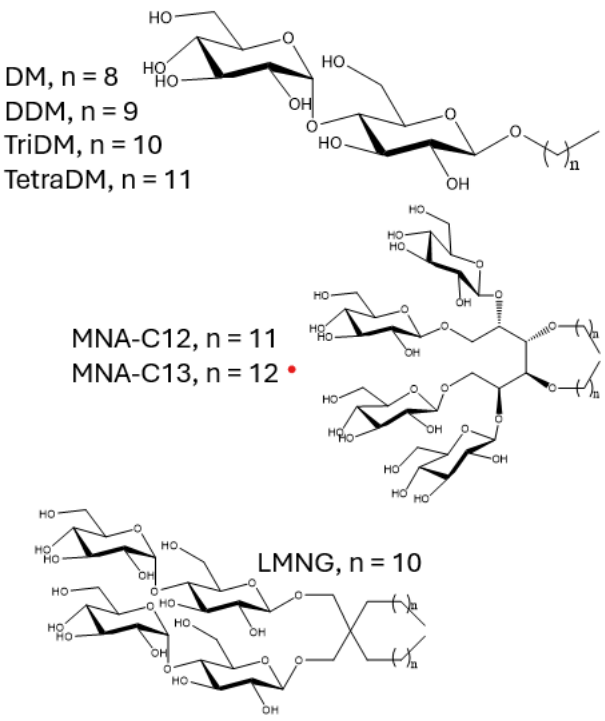


Figure 1.2: Common detergents shown to stabilize membrane proteins.

nature allows for stabilization of the hydrophobic TMDs in the aqueous milieu[20-21]. Detergents such as Lauryl Maltose Neopentyl Glycol (LMNG)[22,23], Mannitol Based Amphiphiles (MNA-C12 or MNA-C13)[24,25], and maltosides of varying length (DDM, TriDM, or TetraDM)[26-29] (**Figure 1.2**) have gained popularity as additives to prevent aggregation and stabilize membrane/hydrophobic proteins during isolation and purification.

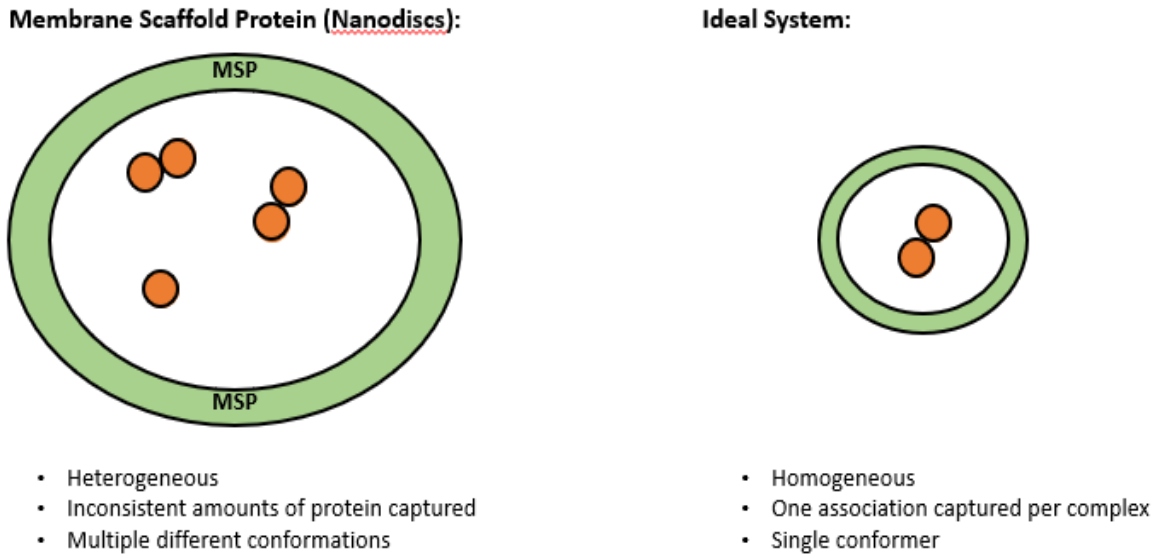


Figure 1.3: Depiction of attributes required to optimize membrane scaffold protein (MSP) system for high-resolution structural determination of membrane proteins. Green – disc/scaffold, Orange – instance of membrane protein

The use of detergents as stabilizers for the hydrophobic regions of membrane proteins comes at the cost of the loss of native lipophilic interactions. Detergents share structural features with lipids, both having a hydrophobic tail and polar headgroup. However, there are differences in headgroups, length of the hydrocarbon portion, branching within the hydrocarbon chain, and unsaturation. In addition, natural membranes contain multiple lipids, whereas detergents feature a single component.

To address some of the issues associated with replacing lipidic interactions with detergent interactions, Nanodiscs have been developed to encapsulate and study membrane proteins. Nanodiscs, which were pioneered by the Sligar group, exploit a protein construct known as a Membrane Scaffold Protein, which, upon dimerization, can encapsulate a modestly large bilayer of endogenous lipids to harbor a membrane protein of interest[30,31]. Upon reconstitution, a membrane protein of interest can then be studied in an isolatable and water-soluble model lipid bilayer[32,33]. This approach circumvents many problems associated with detergents by preserving the native lipid-protein interactions and keeping the membrane protein within a bilayer environment. Primarily due to their large size, Nanodiscs have been utilized to solve the structure of some entire transmembrane proteins via NMR[34,35] or Cryo-EM[36-45]. However, for studying transmembrane quaternary contacts isolated from the influence of the other regions of a membrane protein, techniques like Cryo-EM falter as the side of the system decreases. Since isolated transmembrane quaternary contacts are much smaller than whole membrane protein, X-ray crystallography is more suited to investing these smaller interactions. An ideal system with the membrane stabilizing capabilities of Nanodiscs would need to meet the following criteria to structurally elucidate isolated transmembrane quaternary contacts: 1) smaller overall size 2) homogenous 3) single instance of an encapsulated protein/protein per complex (**Figure 1.3)**.

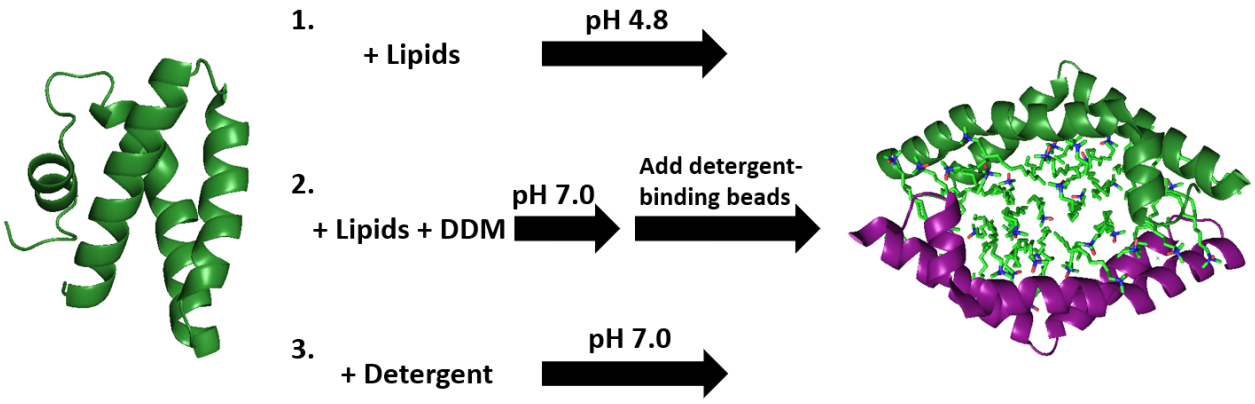


Figure 1.4: Approaches to facilitate the monomer to dimer transition of SapA. 1) Endogenous lipids at low pH. 2) Lipids + DDM detergent at neutral pH followed by detergent removal. 3) Detergents under neutral pH. Approaches 1 and 2 facilitate a lipid core while approach 3 forms a detergent core. Monomer (PDB: 2DOB) Dimer (PDB: 4DDJ)

We took inspiration from another class of proteins that has been used to encapsulate the TMD of membrane proteins, Saposin (Sap). Sap proteins are a class of short (~80 residues) proteins formed from the post-translational processing of pro-Sap to form SapA, SapB, SapC, and SapD[46]. These proteins have high sequence homology, are held together by three sets of conserved disulfide bonds and dimerize to form a disc-like secondary structure (**Figure 1.4**). The internal surface of the disc is lined with hydrophobic residues that function to encapsulate lipids and the TMDs of membrane proteins that are to be shuttled to the proteosome for degradation. Despite the native function of encapsulating and shuttling membrane proteins and lipids, Popovic et al. determined that SapA was able to encapsulate synthetic detergents as evidenced by the crystal structure of dimeric SapA encapsulating the nonionic detergent N,N-dimethyl-n-dodecylamine N-oxide[47]. Kurgan et al. subsequently studied SapA and modified versions of SapB, SapC and SapD containing a tryptophan (Trp) mutation within the hydrophobic pocket of the Sap dimers for the ability to encapsulate alternative non-ionic detergents: LMNG, MNA C12, and TetraDM[29].

We hypothesized that due to the small size, broad capability to encapsulate hydrophobic molecules such as lipids, TMDs, and detergents, and the capability of Sap proteins to facilitate high-resolution structural elucidation, that Sap proteins could serve as tools for structural determination of isolated TMD quaternary contacts.

1.3 – Non-Covalent Encapsulation of TMDs Inside of a SapA Dimer

1.3.1 – Introduction to SapA

A unique feature of SapA, compared to the other Sap homologs (SapB, SapC and SapD), is its ability to adopt varying oligomeric states[47]. Structures of SapA as a monomer (PDB: 2DOB) and dimer (PDB: 4DDJ) have been reported. While the other homologs have only been reported as dimeric. Controlling of the monomer to dimer transition of SapA is straightforward as SapA generally remains monomeric unless incubated with some form of surfactant, which, promotes the dimeric "picodisc" state (**Figure 1.4**). We believed that SapA could be useful for encapsulation of the TMDs of membrane proteins. To structurally elucidate the TM quaternary contacts of membrane proteins in isolation (without influence from the intra- or extracellular domains), TM peptides corresponding to the TMD of membrane proteins will be used in lieu of whole proteins. Additionally, the surfactants that promote dimerization of SapA and solubilization of the TM peptides may also be able to facilitate shuttling of the TM peptides into the dimeric picodisc.

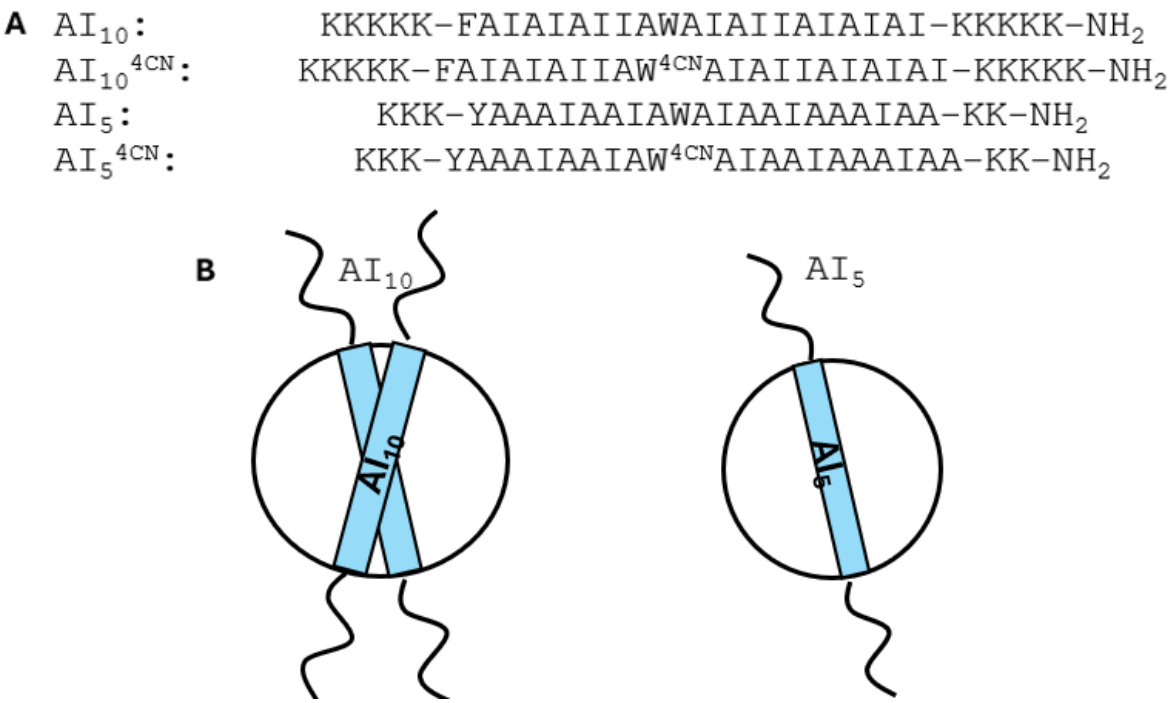


Figure 1.5: (A) Sequences of AI_{10} and AI_5 along with the corresponding Trp^{4CN} analog. (**B**) Depiction of the in membrane oligomeric state of AI_{10} (dimeric) and AI_5 (monomeric) within a detergent micelle.

With the capabilities of SapA protein to encapsulate detergent molecules already well-established[29], two *de novo* designed model peptides, AI_5 and AI_{10}, were chosen as representative TM peptides that within a membranal-mimetic environment (SDS micelles) exist in a monomeric or a dimeric oligomeric state, respectively[48]. Each peptide contains a hydrophobic segment rich in alanine and isoleucine with a centralized tryptophan residue. The N- and C- termini feature lysine residues to facilitate solubility in aqueous solvents

SapA was incubated with LMNG, MNA C12, or TetraDM along with a transmembrane peptide (AI_5 or AI_{10}), and the mixtures were analyzed via size exclusion chromatography (SEC). See Methods **Section 1.9.4** for experimental details. SEC separates molecules based on overall volume. Non-covalent assemblies can be analyzed without dissociation. SEC appeared to be an ideal method to observe the effects of TM peptide incorporation on the oligomeric state of SapA.

1.3.2 – SapA's Oligomeric State upon Encapsulation of Transmembrane Peptides

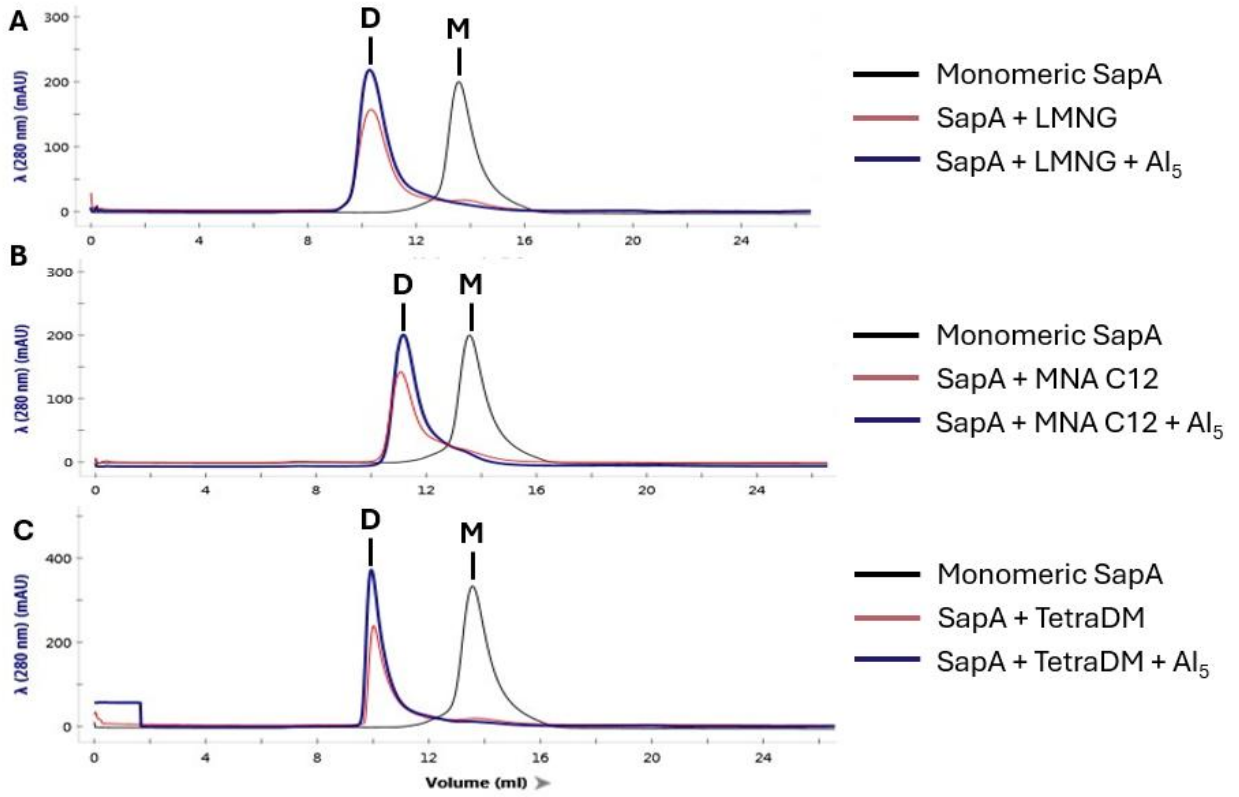


Figure 1.6: Size exclusion traces of SapA with transmembrane peptides and detergents. (**A**) SapA : LMNG : AI_5 (1:1:15), (**B**) SapA : MNA C12 : AI_5 (1:1:10), and (**C**) SapA : TetraDM : AI_5 (1:1:20). 280nm absorbance monitored. Black trace – monomeric SapA without detergent, red trace – SapA + detergent, blue trace – SapA + peptide + detergent. The peak which corresponds monomeric SapA is indicated with an "M". The peak which corresponds to the dimeric species formed from the presence detergent and/or TM peptide is indicated with a "D".

When incubated with AI_5, SapA formed a dimeric complex regardless of choice of detergent (**Figure 1.6**). The formation of a dimeric complex is evidenced by the diminished retention time relative to the monomeric SapA control (indicating a larger assembly) and comparable retention time of SapA dimerized with only detergent. The retention time of the SapA + peptide + detergent dimer varies slightly from the retention time of the SapA + detergent dimer. This slight difference is attributed to the necessity to displace detergent molecules within the hydrophobic pocket of the SapA dimer to make space for the TM peptide to occupy. This change in cargo within the hydrophobic pocket may lead to modest

changes in the overall volume of the dimer complex. An alternative explanation for this retention time change could be that encapsulation of the AI_5 peptide introduces contacts between the peptide and picodisc that distort the overall conformation of the picodisc, leading to an altered volume.

Dimeric assemblies formed from mixtures of SapA + detergent + TM peptide appeared to show higher 280nm absorbance values relative to the dimeric assemblies formed with only SapA + detergent (**Figure 1.6**). Absorbance at 280 arises largely from Trp side chains. SapA contains one Trp while the peptide contains one as well. Therefore, a dimer containing a peptide should display a stronger 280nm absorbance than a dimer lacking a peptide. Thus, both SEC retention times and the increased 280nm absorbance support the conclusion that the TM peptide is incorporated into the SapA + detergent picodisc.

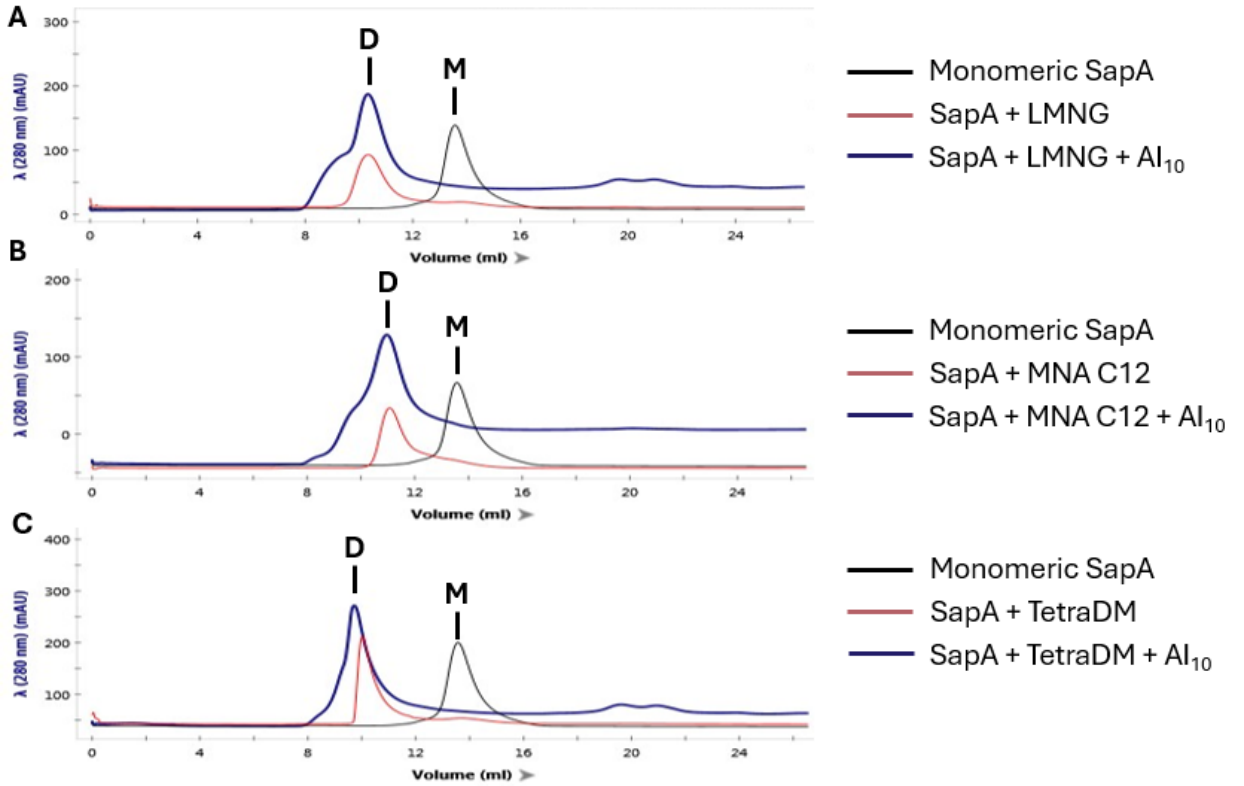


Figure 1.7: Size exclusion traces of SapA with transmembrane peptides and detergents. (**A**) SapA : LMNG : AI_{10} (1:1:15), (**B**) SapA : MNA C12 : AI_{10} (1:1:10), and (**C**) SapA: TetraDM : AI_{10} (1:1:20). 280nm absorbance monitored. Black trace – monomeric SapA without detergent, red trace – SapA + detergent, blue trace – SapA + peptide + detergent. Black trace – monomeric SapA without detergent, red trace – SapA + detergent, blue trace – SapA + peptide + detergent. The peak which corresponds monomeric SapA is indicated with an "M". The peak which corresponds to the dimeric species formed from the presence detergent and/or TM peptide is indicated with a "D".

When incubated with the dimer-forming TM peptide AI_{10} (**Figure 1.7**), the SapA + detergent mixture forms a predominantly dimeric assembly. However, unlike the mixture containing AI_5, higher-order oligomers (most likely trimeric), are also observed. I believe that this observation is due to two features of our system: 1) SapA's capability to adopt different oligomeric states and 2) the larger size of the AI_{10} dimer compared to the AI_5 monomer. SapA has been observed to form assemblies greater than dimers around the TMDs of much larger protein complexes[49-53].

After observing the mixture of oligomers when encapsulating AI_{10}, I hypothesized that the dimer-trimer equilibrium could be influenced by varying concentrations of detergent or peptide. The SEC experiment was repeated with either doubled detergent concentration or doubled peptide concentration. We also conducted a control experiment in which AI_{10} and detergent were incubated without SapA to determine whether we could visualize non-encapsulated peptide contained within detergent micelles.

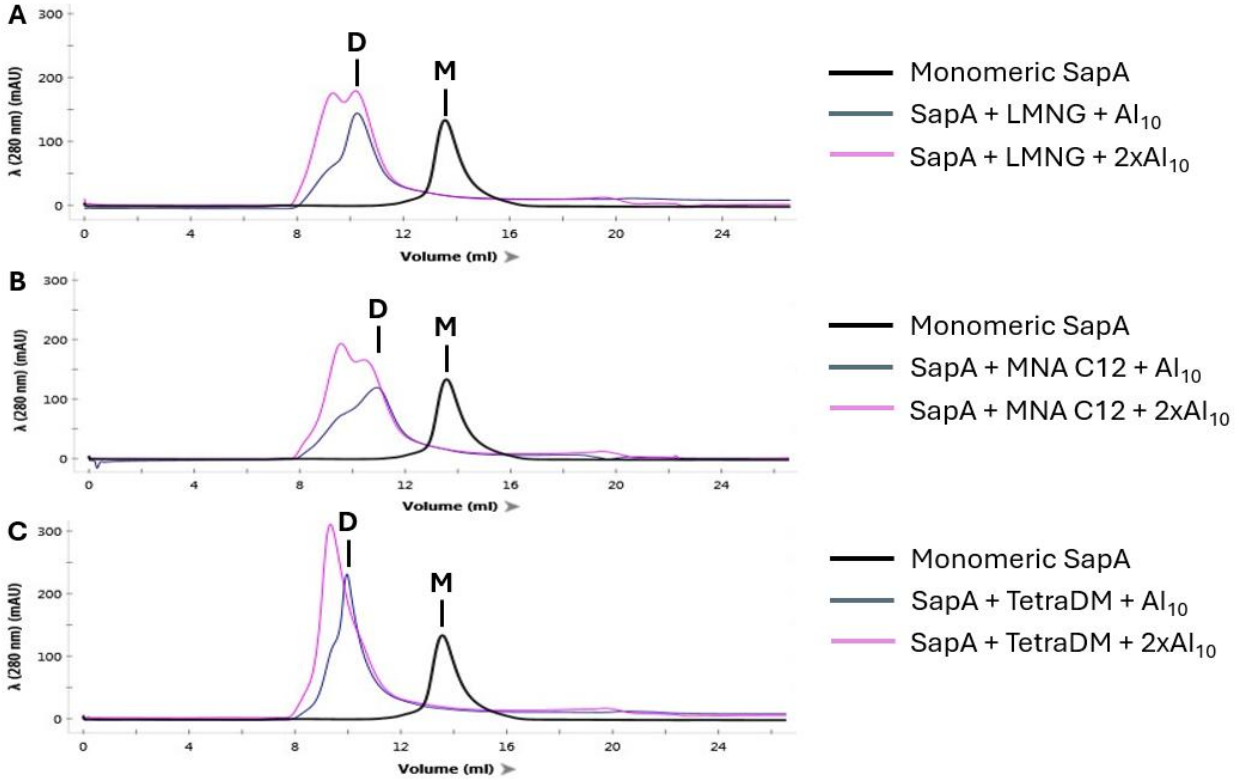


Figure 1.8: Size exclusion traces of SapA with transmembrane peptides and detergents. (**A**) SapA : LMNG : AI_{10} (1:2:15), (**B**) SapA : MNA C12 : AI_{10} (1:2:10), and (**C**) SapA : TetraDM : AI_{10} (1:2:20). 280nm absorbance monitored. Black trace – monomeric SapA without detergent, blue trace – SapA + peptide + detergent pink trace – SapA + 2x peptide + detergent. The peak which corresponds monomeric SapA is indicated with an "M". The peak which corresponds to the dimeric species formed from the presence detergent and/or TM peptide is indicated with a "D".

Doubling the concentration of peptide led to an increase in larger molecular weight species in the SEC trace, suggesting that increased peptide concentration may favor a trimeric picodisc (**Figure 1.8**). This may be attributed to the AI_{10} adopting different oligomeric states. Under the higher concentration of peptide used in **Figure 1.8** compared to **Figure 1.7**,

AI_{10} may be adopting a trimeric state that requires a trimeric SapA to fully encapsulate. AI_{10} adopting a trimeric state is supported by the effects observed upon increasing detergent concentration (**Figure 1.9**). In both cases where a double-tailed detergent (LMNG or MNA C12) was used there was no noticeable effect on the picodisc equilibrium. Collectively, these SEC data suggest that the oligomeric state of SapA picodiscs is determined by the oligomeric state of the TM peptide and not with the concentration of detergent. However, when using a single tailed detergent (TetraDM) there was a shift towards the larger molecular weight species. This may be due to the smaller size of TetraDM's hydrophobic tail, being a single-tailed detergent, which could facilitate more facile encapsulation within the SapA picodisc (**Figure 1.8C**).

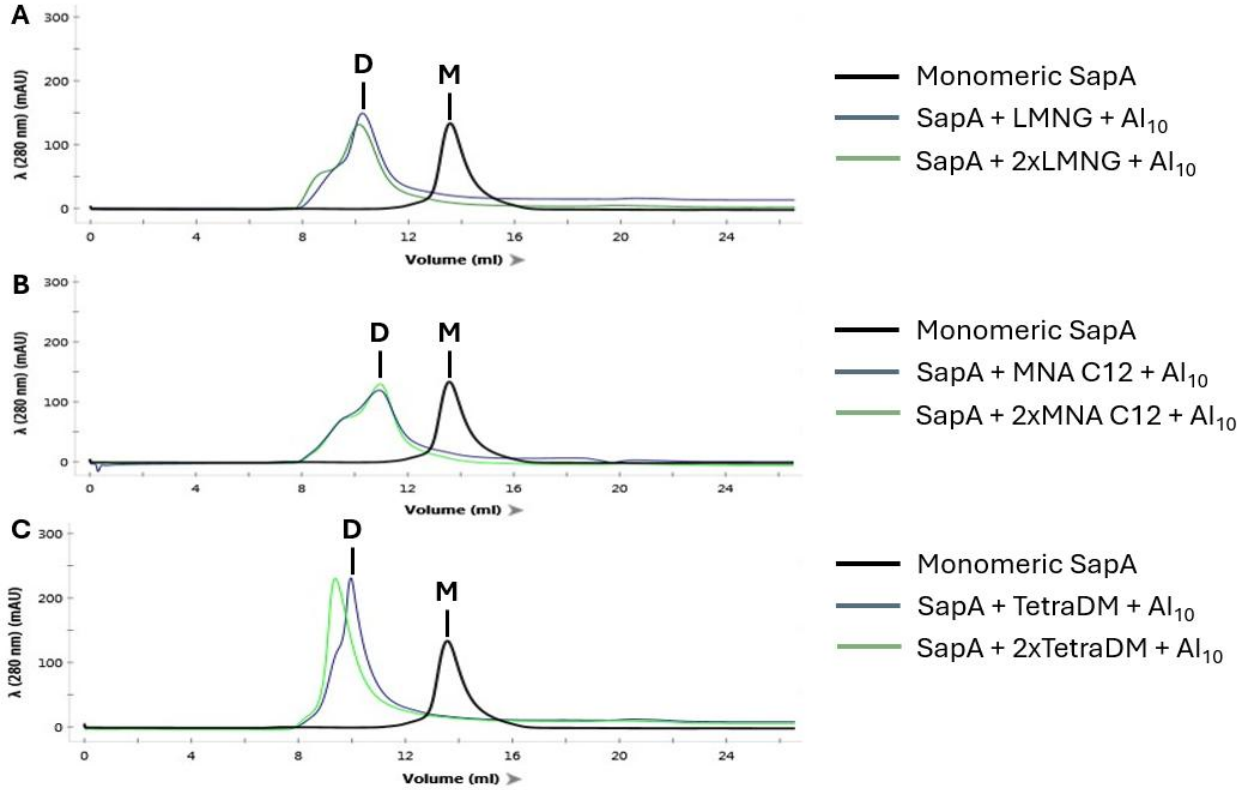


Figure 1.9: Size exclusion traces of SapA with transmembrane peptides and detergents. (**A**) SapA : LMNG : AI_{10} (1:1:30), (**B**) SapA : MNA C12 : AI_{10} (1:1:20), and (**C**) SapA : TetraDM : AI_{10} (1:1:40). 280 nm absorbance monitored. Black trace – monomeric SapA without detergent, blue trace – SapA + peptide + detergent pink trace – SapA + peptide + 2xdetergent. The peak which corresponds monomeric SapA is indicated with an "M". The

peak which corresponds to the dimeric species formed from the presence detergent and/or TM peptide is indicated with a “D”.

After observing the varying oligomeric states of SapA picodiscs upon encapsulation of AI_{10}, I wondered if full encapsulation of all AI_{10} peptides was occurring or if there was a substantial amount of non-encapsulated AI_{10}, presumably solubilized by detergent alone. Another SEC experiment where AI_{10} was incubated with just detergent was conducted (**Figure 1.10**). It was observed that the detergent solubilized peptide forms a micelle that is similar in size to the proposed picodisc encapsulated TM peptide. For the double-tailed detergents, the peptide micelles are typically closer in size to the dimeric SapA picodiscs than the trimeric form. The single-tailed TetraDM when incubated with AI_{10} has an almost identical retention time as when either detergent or peptide concentration is doubled. This further indicates that the smaller size of TetraDM tends to facilitate formation of larger complexes or higher order assembles.

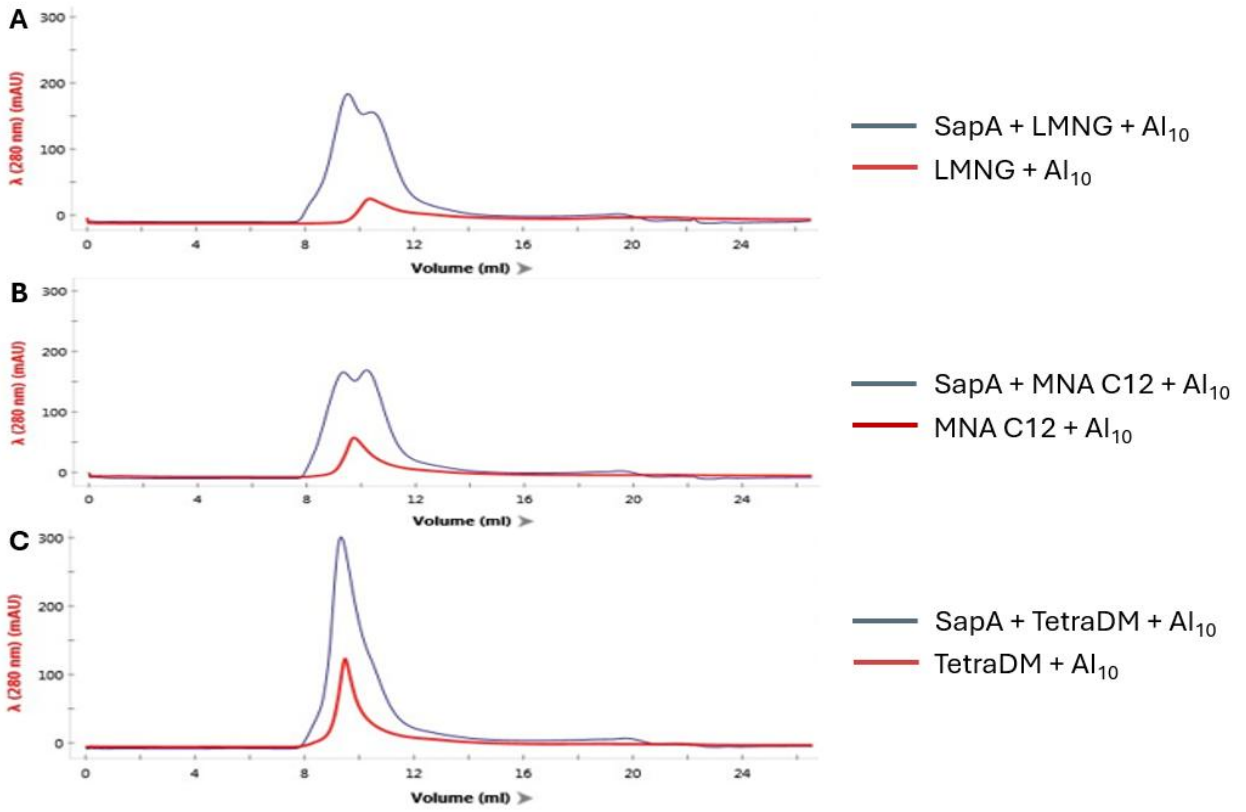


Figure 1.10: Size exclusion traces of SapA with transmembrane peptides and detergents. (**A**) LMNG : AI_{10} (1:15), (**B**) MNA C12 : AI_{10} (1:10), and (**C**) TetraDM : AI_{10} (1:20). 280nm absorbance monitored. Blue trace – SapA + peptide + detergent, red trace – peptide + detergent. The peak which corresponds monomeric SapA is indicated with an "M". The peak which corresponds to the dimeric species formed from the presence detergent and/or TM peptide is indicated with a "D".

1.3.3 – Fluorescence Energy Resonance Transfer Studies

The previous SEC experiments suggested that TM peptides can associate with picodiscs, evidenced by the change in SEC retention times when SapA was incubated with TM peptide and detergent. However, observing dimeric or higher oligomeric states of the SapA picodisc does not provide insight into how or where the peptides are interacting within the picodisc complex. Due to the hydrophobic composition of the inner surface and native

function of SapA, we can reasonably assume that the peptide will be located within the hydrophobic pocket, but to test this hypothesis this another method is needed.

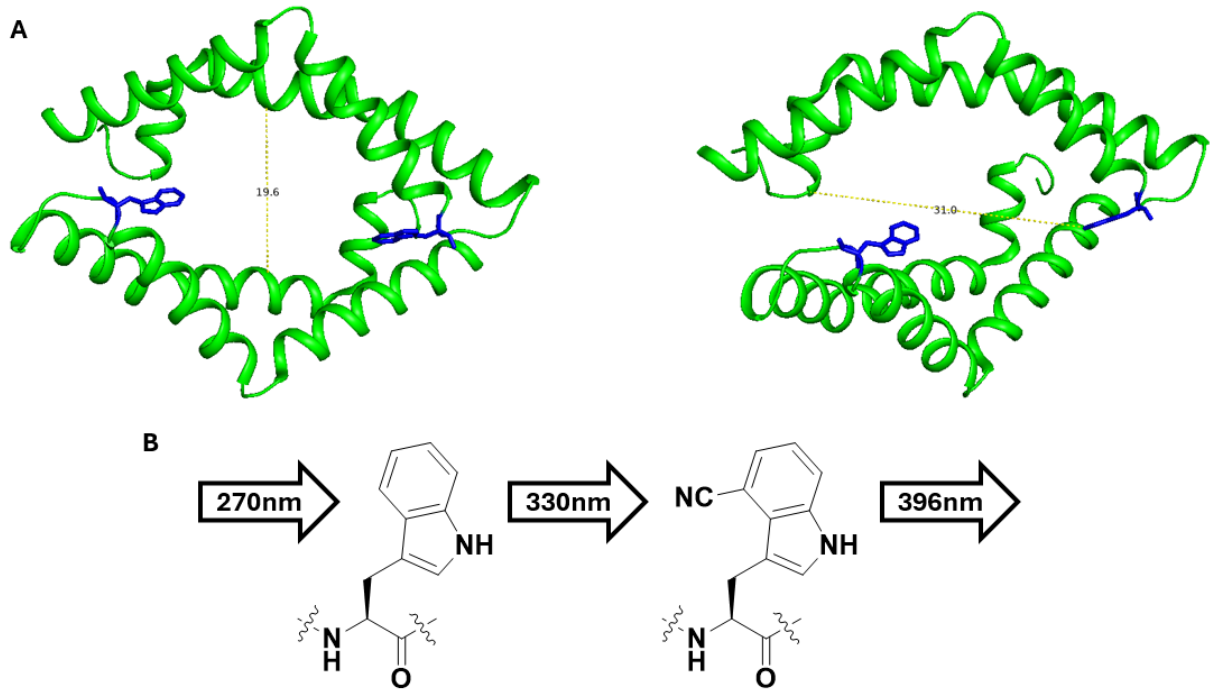


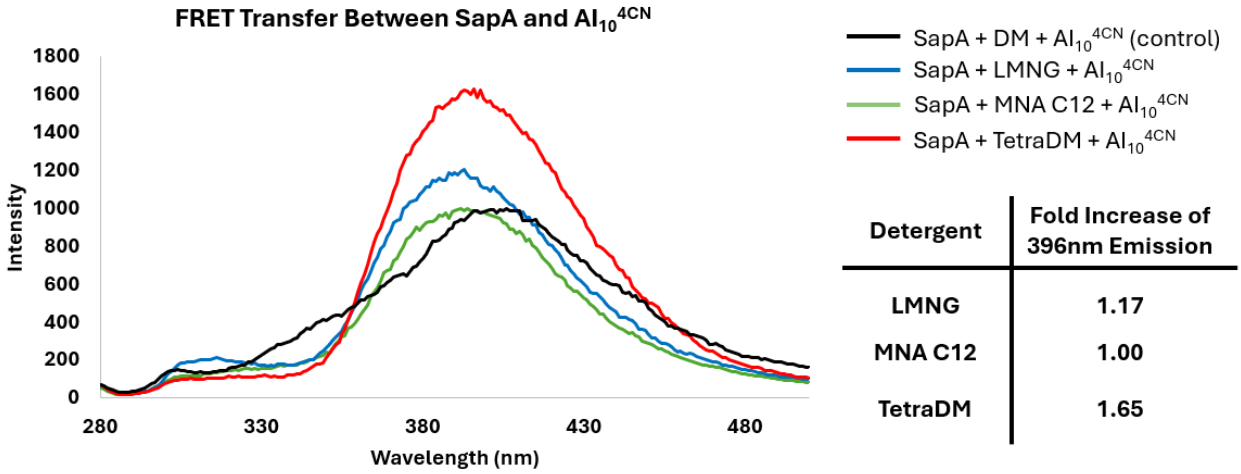
Figure 1.11: (**A**) Crystal structure of the SapA dimer (4DDJ) with Trp residues indicated in blue and the shortest and longest radii from either end of the hydrophobic pocket indicated (19.6 or 31.0Å).(**B**) Depiction of Trp-Trp^{4CN} FRET interaction with excitation and emission wavelengths shown.

Fluorescence Energy Resonance Transfer (FRET) is a popular and sensitive method to assess the proximity between two molecules. This technique utilizes the photophysical properties of two separate fluorescent molecules with compatible excitation and emission profiles. One fluorophore, the FRET doner, is excited with a light source that does not overlap with the excitation region of the second fluorophore, the FRET acceptor. Upon excitation, the FRET doner emits light at a wavelength that overlaps with the excitation region of the FRET acceptor. Once the FRET acceptor is excited, it will emit light at a third wavelength (**Figure 1.11B**). This phenomenon is distance-dependent; the two molecules should be within the

Förester distance (distance over which half of the fluorescence energy is transferred between fluorophores) for an efficient FRET interaction.

Many FRET pairs have been reported that range in size from small molecules to proteins housing fluorescent moieties; examples include FITC-TAMRA[54] and mScarletI-mNeongreen[55]. Since the protein-protein interactions being interrogated are in a membrane-mimetic environment, the fluorophores used must be compatible with hydrophobic environments. Many popular fluorescent molecules used for FRET have drastically altered photophysical properties when the polarity of their environment is changed. Recently, 4-cyanotryptophan (Trp^{4CN}) has been characterized as a FRET acceptor compatible with endogenous Trp residues as a FRET donor within both membrane and aqueous environments[56]. This FRET pair would be optimal for our system as the relatively small size of the cyano group makes it ideal for modifying a protein or peptide while preserving many physical properties.

Both SapA and AI_{10} have Trp residues that could either be replaced with Trp^{4CN} to generate the FRET acceptor. AI_{10} was chosen to be the FRET acceptor molecule as incorporating Trp^{4CN} can be done far more expeditiously via solid-phase peptide synthesis (SPPS) than protein expression. The Förester distance of Trp^{4CN} is 24.6 Å[56], Trp^{4CN} peptide within a picodisc should lead to FRET with a SapA Trp residue (**Figure 1.11A**).



Detergent	Fold Increase of 396nm Emission
LMNG	1.17
MNA C12	1.00
TetraDM	1.65

Figure 1.12: FRET emission spectra after excitation with 270nm light. 280-500nm range was monitored for FRET acceptor emission. All SapA and peptide concentrations are at 1 μM. Conditions for each FRET experiment: SapA : AI_{10}^{4CN} (1:1) in 1% DM, SapA : MNA C12 : AI_{10}^{4CN}: (1:10:1), SapA : LMNG : AI_{10}^{4CN} (1:15:1), and SapA : TetraDM : AI_{10}^{4CN} (1:20:1). Fold increase in emission intensity shown on table.

See **Section 1.9.5** for sample preparation. SapA and AI_{10}^{4CN} were incubated with the detergent n-decyl-β-D-maltoside (DM) was used as a negative control. It is known that DM does not facilitate picodisc formation with SapA[29]. When encapsulating AI_{10}^{4CN} we observe changes in the emission profile that vary with the detergent used for picodisc formation. When TetraDM was used, we see a 1.65-fold increase of the emission intensity at 396 nm suggesting that the Trp^{4CN} residue within the peptide is located near a Trp residue within the SapA picodisc. However, with double-tailed detergents, smaller or no change in emission intensity is observed. These variations in FRET emission can be intrepretted in two ways: 1) differing effects from detergents on photophysical properties of Trp^{4CN} 2) poor encapsulation of the TM peptide within the picodisc. Considering the drastic impact of local environment on an indole's photophysical properties[57-58], it is reasonable that the environments of LMNG or MNA C12 could have altered the emission or excitation region of SapA's Trp residue or the labeled peptide's Trp^{4CN}residue, respectively, to be less overlapped with one another. On the

other hand, this lack of detected Trp^{4CN} emission could also indicate that inefficient encapsulation of the labeled peptide was occuring. When TetraDM or LMNG was used, there was a noticeable increase in FRET, suggesting some degree of AI_{10}^{4CN} being encapsulated within the picodisc. However, presumably the majority is not located proximally to the Trp residue within SapA or is not encapsulated.

1.3.4 – Concluding Remarks on Non-Covalent Encapsulation Approach

The SEC and FRET data suggest some degree of encapsulation of the TM peptide within the picodisc. However, there are multiple other species besides the SapA + peptide + detergent species formed that would hamper the viability of this approach for crystallization trials. When encapsulating AI_5 the picodisc adopts exclusively the dimeric state, but when AI_{10} was encapsulated, a mixture of dimer and larger species emerged. This is presumably due to SapA's propensity to oligomerize into higher-order assemblies to encapsulate larger TMDs. This resulting heterogeneity might discourage crystallization.

Besides the inherent mixture of oligomeric states present when the dimer forming TM peptide is used for encapsulation, there are also micelles containing TM peptide that are of similar volume. This makes removing the non-encapsulated TM peptides in detergent micelles from the sample quite challenging.

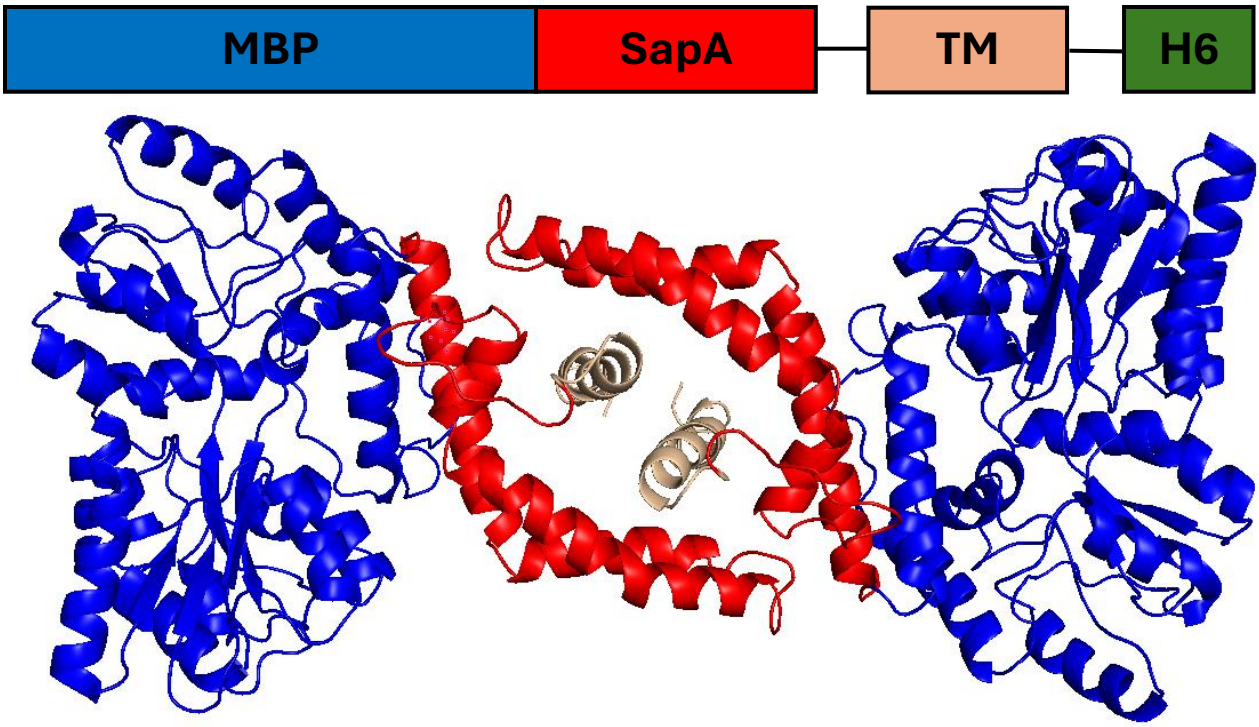
1.4 – Covalent Encapsulation RTK TMDs using a SapA Fusion Construct

1.4.1 – Introduction to Maltose Binding Protein-SapA Hybrid (MSH_A)

To address the issues documented in **Section 1.3.3**, regarding the non-covalent encapsulation of TM peptides with SapA, Dr. Zhen Yu (Gellman Group) set about generating a chimera protein of SapA fused to the TMD of a membrane protein of interest. We hoped that this would directly address the incomplete encapsulation issue of the TM peptides by covalently linking the TMD of interest to the picodisc with a short flexible linker. However, preliminary characterization of this bipartite SapA-TMD fusion protein construct for TMD encapsulation and crystallization was discouraging (unpublished). To further optimize the fusion construct for encapsulation and crystallization of TMD quaternary contacts, a tripartite fusion construct incorporating an N-terminal Maltose Binding Protein (MBP) domain was designed. We call this platform the "Maltose Saposin Hybrid" or MSH (**Figure 1.13**).

The incorporation of MBP onto the N-terminus of a protein has been shown to increase solubility[59,60], expression yields[61-63], and crystallization capacity/structural elucidation[64-68]. The latter is attributed primarily to two features of MBP: 1) MBP itself crystallizes quite readily under a wide range of conditions and 2) the C-terminus of MBP is α-helical, which when fused to a protein with an α-helical N-terminus, facilitates rigidification of the protein construct. Both these features are critical aspects of increasing the capacity of a protein to crystallize and have resulted in many reported MBP/MBP fusion structures.

Figure 1.13: Primary sequence by protein region of an MSH_A-TM protein. MBP-SapA hybridized domain followed by a short linker to a TMD of interest. Subsequentially followed by another a short linker to a His Tag



(H6). A crystal structure of MSH_A-FGFR3 is shown below. Note there is no defined electron density inside the hydrophobic pocket. The TM segments are modeled without electron density verification. The design MSH_A-FGFR3 construct, crystallization, and structure solution/refinement conducted by Dr. Zhen Yu.

Dr. Zhen Yu in collaboration with Dr. Dale Kreitler had successfully obtained preliminary crystallographic determination of an MSH_A (MSH construct where the saposin portion is based on the sequence of SapA) dimer with a TMD corresponding to the RTK FGFR3 (**Figure 1.13**). Unfortunately, the electron density was insufficient for most of the TMD to be assigned, the helices shown within the hydrophobic pocket are modeled based on the best approximations of the resolvable electron density. Attempts at addressing this issue, although not explicitly for the FGFR3 TMD, will be discussed **Sections 1.5-1.6**.

The fusion between the MBP and SapA regions within the MSH_A construct is comprised of a hybridized sequence of the C-terminal region of the MBP and the N-terminal region of SapA. This design allows the MSH_A construct to maintain the picodisc dimerization capabilities observed in native SapA without introducing any profound structural distortions.

Upon MSH_A dimerization, a hydrophobic pocket forms at the dimeric interface where the covalently bound TMD linked by a short flexible linker can reside (along with detergents if present). In principle any TMD could be incorporated within the MSH scaffold. We utilize standard molecular biology techniques to edit the MSH_A to contain whichever TMD sequence we desire. This feature has allowed us to generate multiple MSH_A constructs containing TMDs, the self-association of which can be studied. In addition, we have attempted crystallization of some MSH constructs.

1.4.2 – FGFR3 and G380R Homodimer Affinity within an MSH_A Construct

With the MSH_A platform we began to investigate a well-known RTK homodimer TM interaction, FGFR3. Multiple diseases associated with singular point mutations are located within the FGFR3 TMD. We focused primarily on the G380R mutation as it is known cause of Achondroplasia[11], the most common form of human dwarfism. Many disease-associated mutations in proteins are typically thought to affect a protein's biological function by blocking it from conducting its native function. However, when mutations are located inside of single pass transmembrane helices, they can impede by perturbing the oligomeric interface within the membrane. A computational model of the FGFR3 G380R mutation predicts that the membrane-embedded homodimer associates via a shifted dimeric interface about the coiled-coil axis[11]. The mutant dimer is predicted to have similar stability as the native homodimer, but the basal activity of the receptor is upregulated leading to the disease's phenotype. Another FGFR3 disease-associated mutation, A391E, adopts a similar oligomeric interface as the native, but the additional carboxylic acid functional group is

engaged in a non-native hydrogen bonding contact stabilizing the homodimer and upregulating the receptor's basal activity[12].

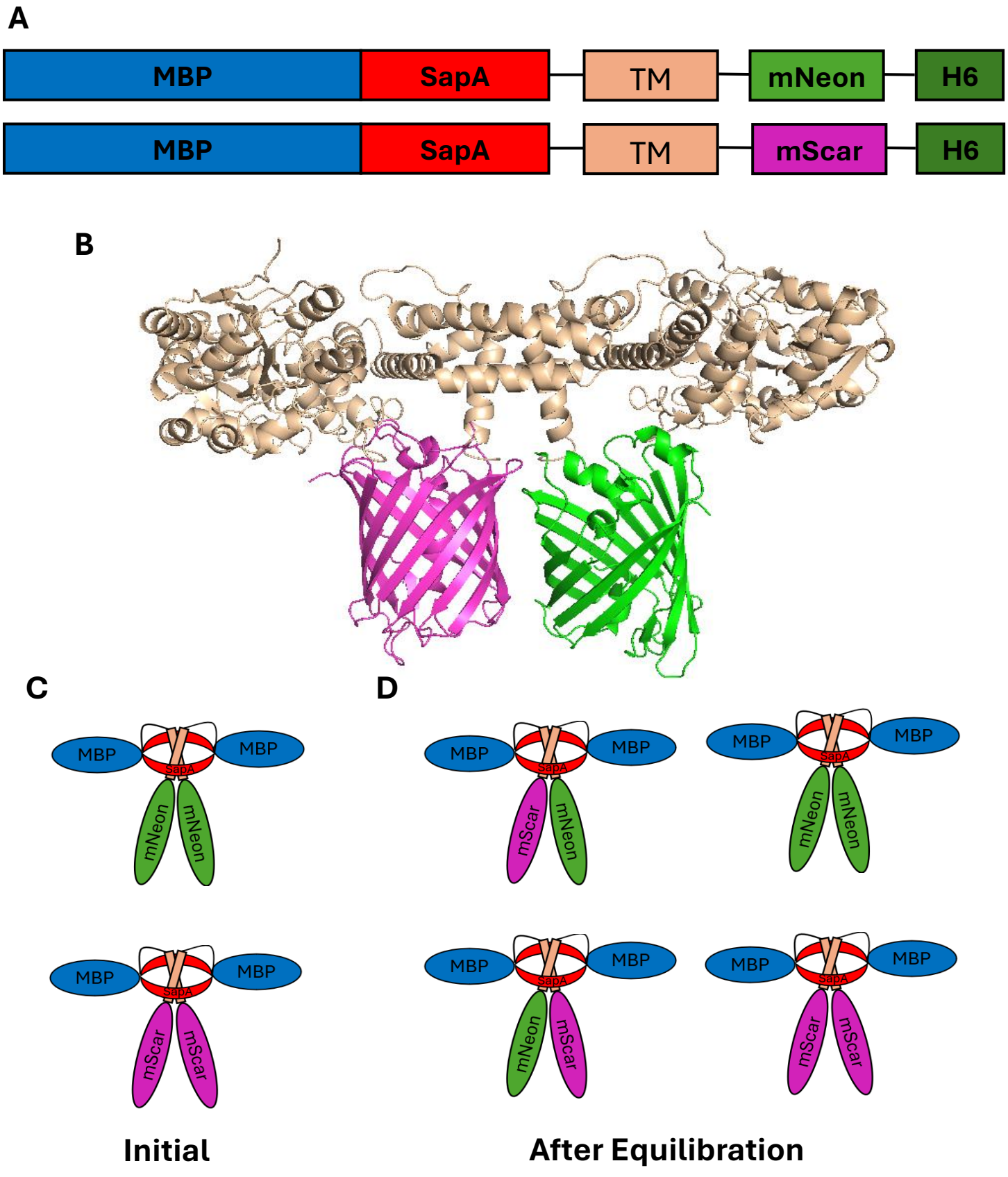

A
MBP
SapA
TM
mNeon
H6
MBP
SapA
TM
mScar
H6
B
C
D
MBP
SapA
mNeon
mScar
Initial
After Equilibration

Figure 1.14:)A) Primary sequence by protein region of an MSH_A-TM-Fluorescent protein MBP/SapA hybridized domain followed by a short linker to a TMD of interest. Subsequentially followed by a single alanine linker to a fluorescent protein connected by a short linker to a His Tag (H6). (**B**) A crystal structure of MSH_A-FGFR3 with mScarletI (PDB: 5LK4) and mNeongreen (5LTP) modeled after the TMD. (**C**) Solution of FRET pair proteins at the beginning of equilibration. (**D**) Solution of FRET pair proteins after equilibration.

We aimed to use the MSH_A construct to isolate the interactions between the TMDs of FGFR3 and FGFR3 G380R homodimers to determine contributions on overall stability and structure without influence from the solvent accessible regions of the endogenous protein. Our primary method to determine the strength of a TM homodimer interactions was a detergent titration assay. In the non-covalent encapsulation studies (**Section 1.3**) we used detergent primarily as a vehicle to solubilize detergents and facilitate SapA oligomerization. The stoichiometries of detergents used were reported within the literature to have a stabilizing effect on the dimeric form of SapA[30]. However, excess detergent can cause dimer dissociation. Taking advantage of this dissociative effect, we designed two variants of the MSH_A construct appended with a FRET Donor (mNeongreen) or FRET Acceptor (mScarletI). After purification of each construct, the fluorescently labeled MSH_A constructs were mixed along with the detergent used for titration and allowed to equilibrate at 4°C for 72 hours. The addition of detergent during equilibration is critical to facilitate the exchange of homodimers to heterodimers (**Figure 1.14C-D)**.

Once equilibrium is reached, detergent is titrated into the solution and the change in FRET efficiency (FE), defined as I_{590}/I_{520}, is interpreted as dimer dissociation. 100% dimer is defined as the initial FE value. The FE value must be constant over the beginning additions of detergent to assure that the sample is not partially dissociated under the concentrations of detergent used for equilibration. 0% is defined as the FE value once further additions of

detergent do not decrease the FE value. Plotting the percentage of dimer versus the log of detergent concentration (log[M]) added and then applying an exponential decay fitting, IC_{50} values were calculated that correspond to the amount of detergent required for 50% dissociation of a dimer. Since IC_{50} values in these experiments indicate the amount of detergent required to dissociate an MSH dimer, the relative strengths of each TMD interaction can be deduced where larger IC_{50} values indicate stronger interactions.

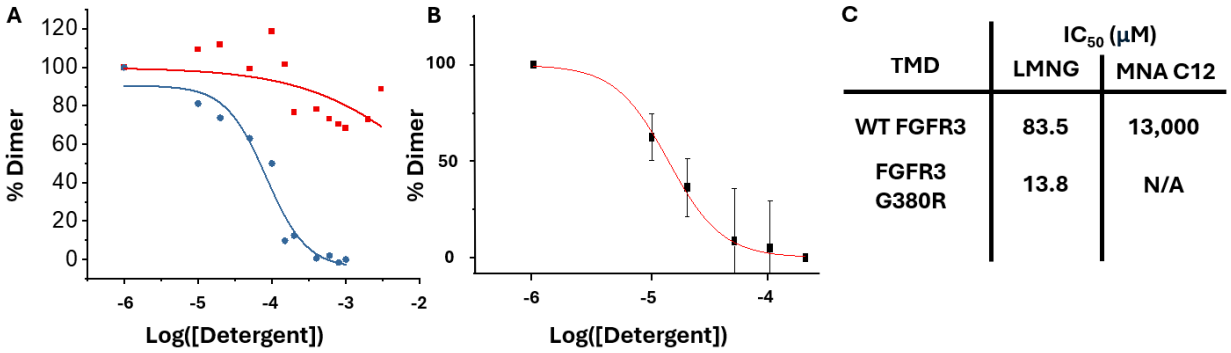


Figure 1.15: (**A**) Dissociation of MSH_A-FGFR3 with LMNG (blue) and MNA C12 (red). (**B**) Dissociation of MSH_A-FGFR3 G380R with LMNG. (**C**) Measured IC_{50} values in μM of each titration.

When titrating MSH_A-FGFR3 with LMNG or MNA C12 we calculated IC_{50} values of 83.5 μM and 13,000 μM, respectively (**Figure 1.15**). The complete dissociation of MSH_A-FGFR3 using MNA C12 was never realized as the complex did not fully dissociate as we approached the solubility limit of MNA C12. The IC_{50} value presented for MNA C12 titration represents the mathematical extrapolation from the exponential decay fitting. LMNG, on the other hand, showed complete dissociation and an IC_{50} of 83.5 μM. This indicates that LMNG can more favorably interact with the MSH_A construct and facilitate dissociation than MNA C12. Since LMNG has a higher propensity to engage with MSH_A constructs, it was used as the primary detergent for titration experiments going forward.

When conducting the LMNG titration with the disease associated construct, MSH_A-FGFR3 G380R, we measured an IC_{50} value of 13.8 μM. This indicates that G380R TM quaternary contacts are weaker than those of its native counterpart. This result is counter to previous reports on the strength of two homodimer associations as the G380R association is stated to be of equal or slightly stronger affinity[12]. These results can be interpreted through the following lens. 1) The strength of homodimerization for these TM peptides is dependent, at least in part, upon the local environment (detergent vs lipid in this case) of the TMD association. 2) The stability of FGFR3 G380R was induced by stabilizing contacts outside of the TMD. It is known that membrane protein structure and function depend upon their local environment, for instance lipid composition[69]. This feature cannot be ignored when discussing homodimer affinity inside membrane-mimetic environments. However, considering the computationally discovered shift in orientation of the oligomeric interface between FGFR3 and FGFR3 G380R[11], the increase in dimeric stability may have arisen from quaternary contacts outside of the membrane. For instance, rotation of the membrane-embedded oligomeric interface may lead to additional salt-bridging contacts in solvent accessible regions between dimeric partners that were previously inaccessible to the WT FGFR3. This increased stability from a shift in the membranal interface may manifest in the enhanced basal rate of phosphorylation of G380R compared to the WT. It is difficult to make conclusive claims on how this mutation affects the basal rate of phosphorylation without further interrogation of the entire protein. However, this observation from our detergent titration experiment highlights an ever-growing need to investigate TMD interactions on

overall protein structure and stability, specifically when involved in quaternary structure and disease-associated mutations.

1.4.3 – Native Mass Spectrometry Analysis of MSH_A-FGFR3 G380R

Alongside the detergent titration experiments, we sought additional approaches to evaluate the formation of MSH_A homodimers. To this end we turned to native mass spectrometry (NMS). These experiments were conducted by Dr. Kyle Brown (Gellman Group). This is a powerful technique that allows for the analysis of entire protein complexes without dissociation. We hoped that this technique would also allow the detection of detergent molecules associated within the MSH_A dimer's hydrophobic pocket.

Preparation of the protein sample for NMS is contingent upon buffer exchanging into 200 mM Ammonium Acetate (AmOAc). Two approaches were attempted for the buffer exchange, both using a purified homodimer of MSH_A-FGFR3 G380R in a buffer of 150 mM NaCl, 25 mM Tris at pH 7.5: 1) an overnight dialysis of the sample in 200 mM AmOAc buffer and 2) diluting the sample directly into 200 mM AmOAc and then passing through an SEC column using the same 200 mM AmOAc mobile phase. Surprisingly, the gentler overnight dialysis led to no observed ions when NMS was conducted. However, the cruder direct dilution and SEC purification approach afforded a faster general method to generate MSH_A samples for NMS.

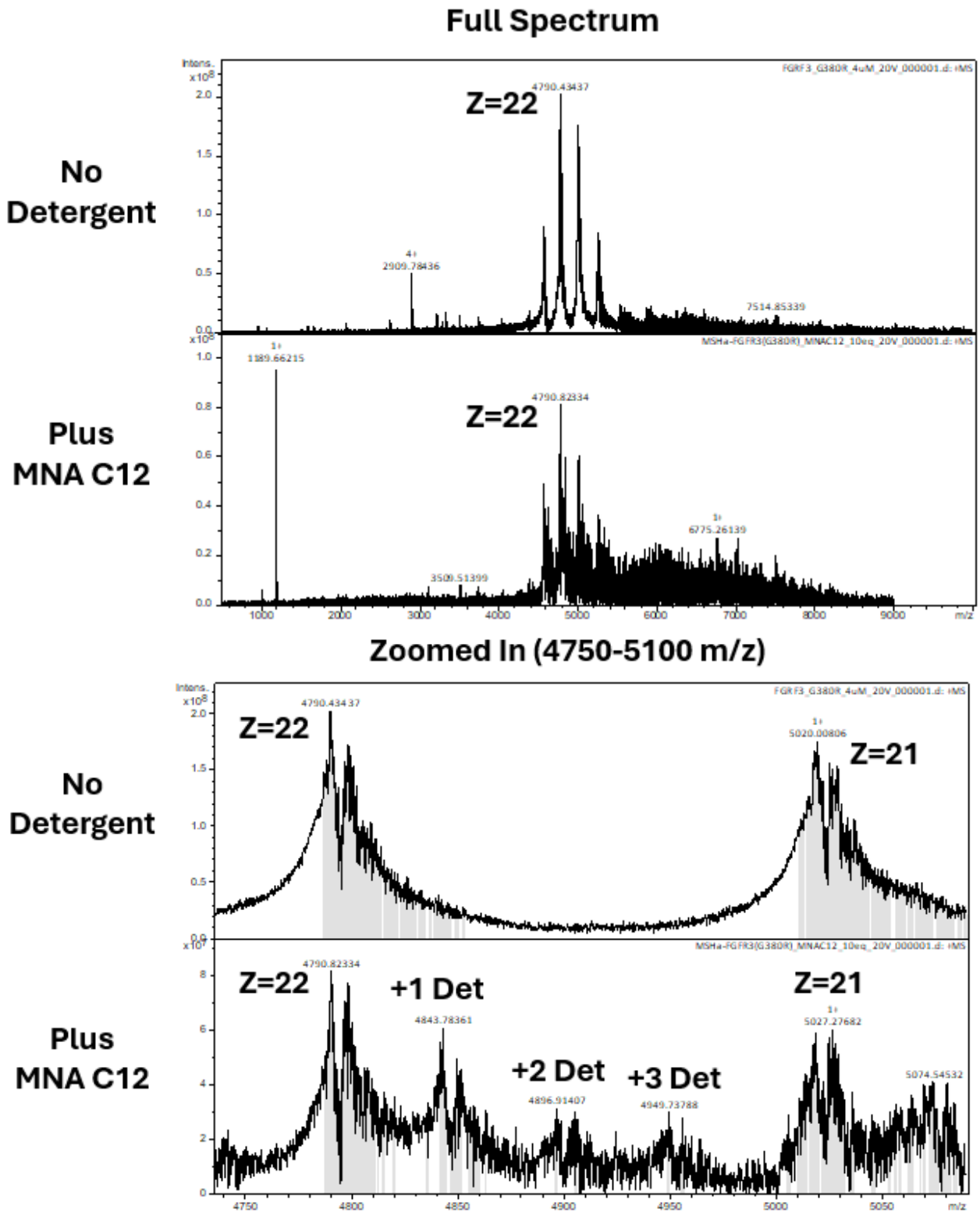


Figure 1.16: Mass spectrum of MSH_A-FGFR3 G380R with no detergent or MNA C12. For the Full Spectrum the Z=22 charge state is indicated. For the Zoomed in spectra, the Z=22 and 21 charge states are labeled. The +1, +2, and +3 detergent masses for the Z=22 charge state are indicated when applicable. Data acquisition and processing completed by Dr. Kyle Brown.

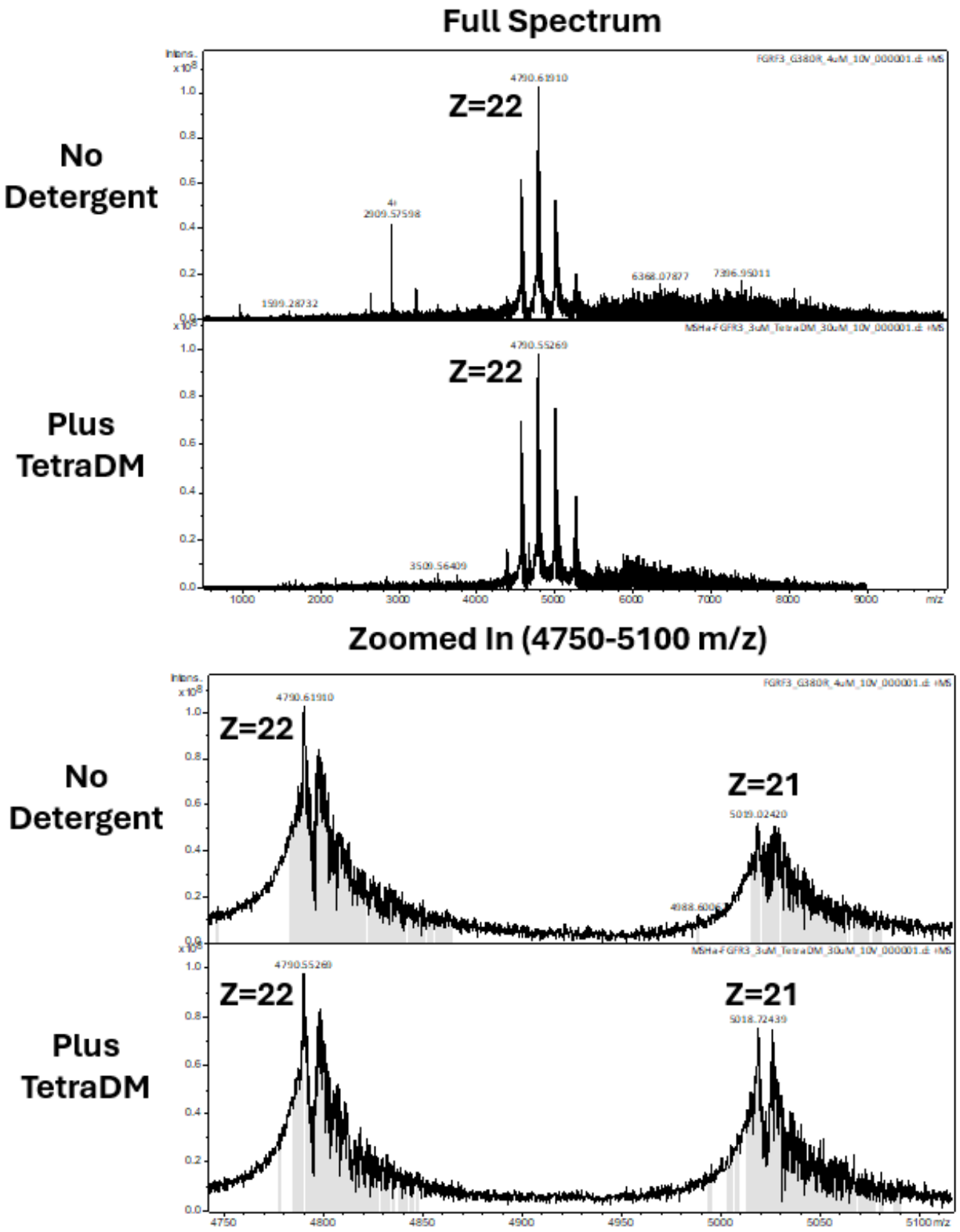


Figure 1.17: Mass spectrum of MSH_A-FGFR3 G380R with no detergent or TetraDM. For the Full Spectrum, the Z=22 charge state is indicated. For the Zoomed in spectra, the Z=22 and 21 charge states are labeled. Data acquisition and processing completed by Dr. Kyle Brown.

When ionized without the presence of any detergent we observe primarily the Z=20 to 23 charge states for the MSH_A-FGFR3 G380R homodimer (MW = 105521.20) (**Figure 1.16 and 1.17**). This verifies that even under mild ionization conditions the protein construct forms a non-covalently bound homodimer. When the detergents MNA C12 or TetraDM were added in a 10:1 ratio to the protein complex, only incorporation of the double tailed detergent, MNA C12, was observed within the protein complex. The number of MNA C12 molecules incorporated within the protein complex varied from one to three copies. The spread in number of encapsulated MNA C12 molecules may represent a heterogeneity amongst the population of MSH_A-FGFR3 G380R homodimers. This heterogeneity may arise from several factors. 1) Conformational dynamics – the picodisc or TMD inside it may sample multiple conformational states that each have a different volume of free space inside the hydrophobic pocket for the detergent molecules to occupy. 2) Detergent association/dissociation – detergent molecules may be freely associating/dissociating with the complex. 3) Ionization induced dissociation – despite NMS being a soft ionization technique, upon ionization the detergent molecules that are weakly associated with the complex may be dissociated. Regardless of whether one or all these potential explanations is correct, this experiment validates that double-tailed detergents can associate with the MSH_A protein construct. Additionally, as MNA C12 can associate with the protein complex this supports that our IC_{50} values obtained from the detergent titration experiments are likely from detergent engagement with the homodimer.

With TetraDM, there was no observation of any detergent association with the protein complex. The resulting NMS spectrum was nearly identical to the spectrum with no

detergent added. I concluded that the MSH_A constructs prefer double-tailed detergents for encapsulation or that single-tailed detergents dissociate readily upon mild ionization.

1.5 – Utilizing the MSH_A Platform to Study the BCL-2 Family Interactome

1.5.1 – The BCL-2 Family Overview

With the investigations of the WT FGFR3 and FGFR3 G380R homodimers within our MSH_A platform completed, we aimed to apply this platform to other classes of proteins with more complex interactomes. To achieve this goal, I focused next on the B-cell Leukemia/Lymphoma 2 (BCL-2) family of proteins.

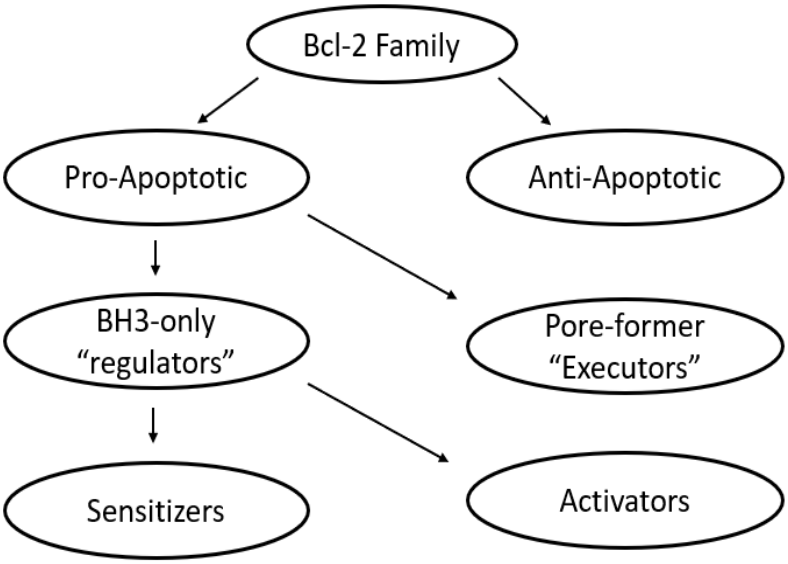


Figure 1.18: Cartoon depiction of the classes within the BCL-2 family of apoptosis regulating proteins.

The members of the BCL-2 family are each involved in regulating multiple biological processes. However, they are perhaps most famously known for their regulation of cellular apoptosis[70-72]. The BCL-2 family is divided into two classes: pro- and anti-apoptotic, with the pro-apoptotic members divided further into sensitizer BCL-2 Homology domain 3 (BH3), activator BH3, and pore-former subcategories. The oligomerization of the pore-former class (BAX and BAK) initiates apoptosis. When assembled into the bio-active oligomer, the pore-former complex can induce Mitochondrial Outer Membrane Permeabilization (MOMP),

which releases Capcases and begins the apoptotic bio-cascade[70]. Anti-apopotic members (BCL-2, BCL-X_L, and MCL-1) prevent apoptosis by binding pro-apoptotic pore-formers to prevent their oligomerization. The sensitizer and activator BH3 proteins play secondary roles in apoptosis regulation either in sequestering anti-apoptotic members to prevent pore-former binding (sensitizers and activators) or binding pore-formers to promote their oligomerization (activators only)[70].

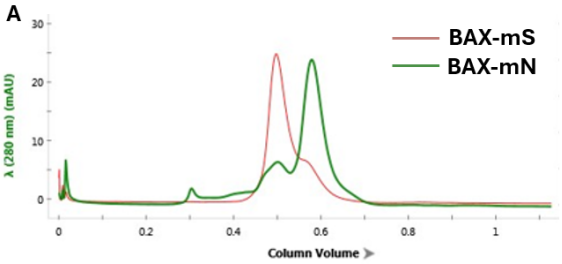
Since the BCL-2 family regulates the survival of a cell, multiple cancers, such as lymphoma[73], leukemia[74], prostate cancer[75], and lung cancer[76] to name a few, have been found to upregulate the endogenous expression of anti-apoptotic members of this family inside cancerous cells. This upregulation ultimately increases the amount of anti-apoptotic proteins available to sequester the pore-former proteins from oligomerizing, thus preventing MOMP and ultimately apoptosis. Due to the importance of this protein interactome in human oncology, understanding the fundamental protein-protein interactions that facilitate binding within the BCL-2 family is critical for further developing therapeutic approaches.

Until recently the predominant mode of association between BCL-2 family members was thought to be through the BH3 region (a region of high sequence homology amongst family members)[70]. This region is solvent accessible and allows for facile association between the membrane bound (anti-apoptotic and pore-former proteins) or the fully solubilized regulator proteins (sensitizer and activator BH3 proteins). However, Andreu-Fernandez et al. reported that chimera proteins comprised of the TMDs of BCL-X_L, BCL-2, or BAX with either the C- or N-terminal half of the Venus fluorescent protein were capable of TMD-mediated association[77]. Formation of each homodimer was detected as well as

heterodimers between anti- and proapoptotic TMDs (BCL-2/BAX and BCL-X_L/BAX), but not between antiapoptotic members. This study suggests that the BCL-2 interactome contained previously unexplored TMD interactions that may be involved in regulating cellular apoptotis. For these reasons, we chose to incorporate the TMDs of BCL-X_L and BAX into our MSH_A construct to investigate their homo- and heterodimer interactions.

1.5.2 – FRET Studies of BCL-X_L and BAX Family Homo- and Heterodimers

With the goal of assessing the affinities of the BCL-2 TMD dimers within the MSH_A construct, I prepared plasmids corresponding to FRET pairs of MSH_A constructs containing the TMDS of BCL-X_L and BAX for detergent titration experiments. As discussed in **Section 1.4.1**, the MSH_A scaffold is modular and exchanging the TMD is accessible by standard molecular biology techniques, see **Section 1.9.6**.



B

TMD Dimer	IC_{50} (uM)	FE (Start)	FE (End)	ΔFE
BAX-mS/BAX-mN	39	6.87%	6.33%	0.54%
BCL-X_L-mS/BCL-X_L-mN	4.2	7.45%	6.39%	1.06%
BAX-mS/BCL-X_L-mN	6.4	8.55%	7.15%	1.40%

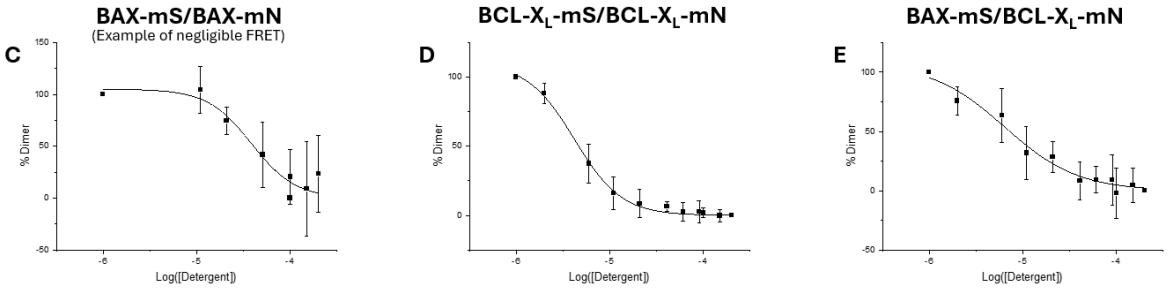


Figure 1.19: A) SEC comparison of MSH_A-BAX-mNeongreen and MSH_A-BAX-mScarletI at 1 μM. B) IC_{50} and FRET efficiency values for each pair of BAX/BCL-X_L homo- and heterodimers. Detergent titration of C) MSH_A-BAX homodimer D) MSH_A-BCL-X_L homodimer and E) MSH_A-BAX-mScarletI/MSH_A-BCL-X_L-mNeongreen heterodimer with LMNG. 280 nm absorbance monitored.

Before continuing into the results of these FRET experiments, it must be stated that as I went about interrogating the BCL-2 interactome it became clear that the BCL-2 TMD interactions rather weak interactions. Particularly, when the FRET proteins (mNeongreen or mScarletI) at the end of the TMD, isolating the MSH_A constructs in their dimeric state was challenging, if possible in certain instances. This is most evident upon examination of the MSH_A-BAX FRET proteins (**Figure 1.18A**). The MSH_A-BAX-mNeongreen construct resides primarily in a monomeric state with a minor dimeric population. MSH_A-BAX-mScarletI exhibited predominately the dimeric state with a minor monomeric population. I attribute this to weaker TMD interactions between BAX homodimers coupled with steric repulsion between the mNeongreen regions of the MSH_A construct. Since the MSH_A-BAX-mNeongreen homodimer is not capable of FRET by itself (FRET acceptor not present), I equilibrated the protein along with MSH_A-BAX-mScarletI and MSH_A-BCL-X_L-mScarletI in the hopes that fluorescent heterodimers would form.

The FRET detergent titration assay for the BCL-2 FRET proteins was conducted using the procedure indicated in **Section 1.4.2**. The LMNG titration experiments with BCL-X_L and BAX TMDs exhibited a low change in FRET efficiency (ΔFE), especially for dimers incorporating MSH_A-BAX-mScarletI as the FRET acceptor (**Figure 1.19B-E**). To exemplify the poor ΔFE observed and its correlation with measured IC_{50}, FE at the beginning and end of titration is shown for these weakly-associated TM interactions (**Figure 1.19B**). It is worth mentioning that no experiment displayed 0% FE, due to either background or a small population of FRET acceptors emitting at the FRET doner's excitation wavelength.

The monomeric nature of MSH_A-BAX-mNeongreen is displayed prominently in the MSH_A-BAX homodimer titration experiment. The dynamic range in ΔFE was remarkably low, a 0.54% reduction in FE over the course of the experiment. This led to an artificially high IC_{50} value that was not considered reliable (**Figure 1.19C**). The heterodimer of MSH_A-BAX-mScarletI and MSH_A-BCL-X_L-mNeongreen displayed the same behavior as the BAX homodimer (data not shown).

The remaining combinations of BAX and BCL-X_L constructs that did not incorporate MSH_A-BAX-mNeongreen displayed greater than double the dynamic range in ΔFE, albeit in both cases only between 1.0-1.5%. However, I believe this low dynamic range is enough to extract meaningful values in IC_{50}, as the error bars at the beginning and end of the titration do not overlap as in the case of the BAX homodimer titration.

As compared to the MSH_A-FGFR3 and MSH_A-FGFR3 G380R TMD homodimer detergent titration experiments (**Figure 1.15**), the IC_{50} values for the MSH_A-BCL-X_L-

mScarletI/MSH_A-BCL-X_L-mNeongreen TMD homodimer and MSH_A-BAX-mScarletI/MSH_A-BCL-X_L-mNeongreen TMD heterodimer are noticeably weaker. Although these interactions are weaker, BCL-2 TMD interactions within the MSH_A scaffold are still detectable using the detergent titration assay. However, I was unable to determine a plateau of 100% dimer in any of our experiments with this class of TMDs (See **Section 1.4.2** for details on FRET assay design). The lack of consistent dimeric species at the beginning of the titration experiment may be due to SapA's propensity to adopt varying oligomeric states and the weak interactions between BCL-2 family TMDs. The FRET dimers may be in a partially dissociated state at the beginning of the titration. Despite this limitation, the measured IC_{50} values, relative to one another, should still be correct (i.e. the MSH_A-BAX-mScarletI/MSH_A-BCL-X_L-mNeongreen heterodimer is a stronger interaction than MSH_A-BCL-X_L-mScarletI/MSH_ABCL-X_L-mNeongreen homodimer), but the actual IC_{50} values are lower than I have been able to measure (<1 μM).

NMS experiments of the BCL-2 family homo- and heterodimers were conducted but dimers were not detected (data not shown).

In conclusion, I have expanded the known capabilities of the MSH_A scaffold by detecting weak/transient and heterodimeric TMD quaternary contacts. These findings conclude suggest that quantitative analysis requires interactions of sufficient strength.

1.6 – Next Generation of MSH Scaffolds: MSH_B

1.6.1 – Comparing the Geometry of Sap Picodisc Homologs

At this point in development of the MSH platform, as highlighted in **Section 1.5.2**, our attempts at covalent encapsulation had been limited primarily to well-known strongly associating TMD quaternary contacts. I hypothesized this limitation to be derived from the propensity of SapA to adopt multiple oligomeric states coupled with encapsulating weaker associating TMDs. If the propensity of a TMD quaternary contact for dimerization is not adequate, the SapA portion of the MSH_A scaffold (forming the hydrophobic pocket) will dictate the oligomeric state. For BCL-2 interactions, and presumably other weak/transient TMD associations, the MSH scaffold tends to be monomeric when under any dissociation stress. As we briefly touched upon in **Section 1.4.1**, the MSH_A platform has allowed crystallographic determination of the MBP-SapA hybridized portion of the protein, but assignment of electron density within the hydrophobic pocket (i.e. the TMD quaternary contacts of interest) has been elusive thus far.

To address these issues, the homologs of SapA were considered as alternatives for the picodisc moiety within the MSH construct. SapA's ability to adopt several oligomeric states is unique amongst the Sap homologs. SapB-D[78,79] have been observed only as dimers, which may make them more optimal for the encapsulation of transient TMD dimers. We hoped that a reduction in oligomeric states, i.e. an increased dimerization propensity, would address the issues hampering the MSH_A scaffold. Additionally, we hypothesized that other Sap picodiscs may form hydrophobic pockets with decreased volume. The reduction in free

space could result in single conformation of the encapsulated TMD, increasing the probability of assigning electron density of the TMD within the hydrophobic pocket during an X-ray crystallography experiment.

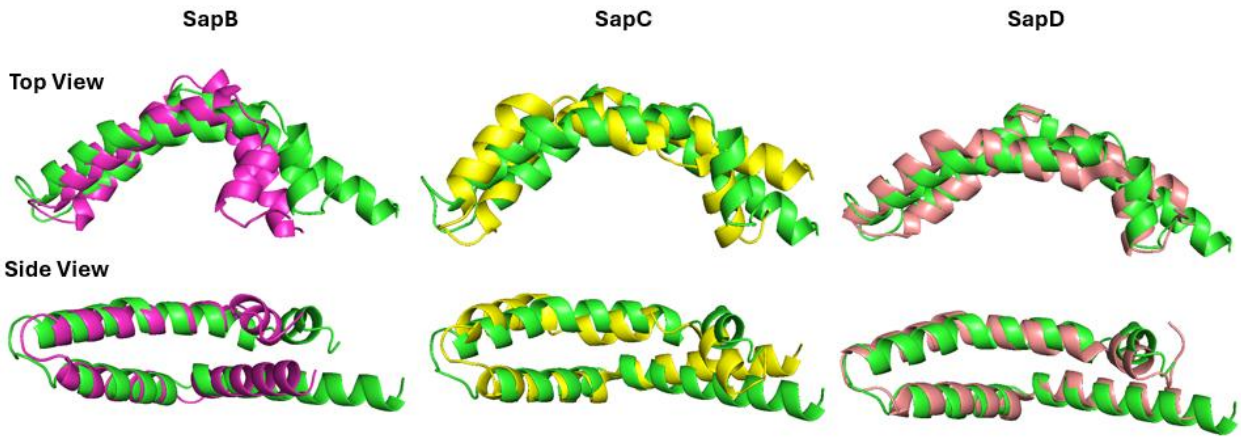


Figure 1.20: Alignment comparison of SapB (PDB: 1N69) in pink, SapC (2Z9A) in yellow, and SapD (5U85) in salmon with the picodisc portion of the MSH_A-FGFR3 structure.

A single copy of the dimeric peptide was overlayed with a single copy of the SapA domain within Dr. Zhen Yu's crystal structure of MSH_A-FGFR3. It is important to note that the reported structures of SapB-D are dimeric but have no molecules encapsulated within the hydrophobic pocket (apo-form). Additionally, there are no structures of SapA in the apo-form, only monomeric or housing detergent molecules (holo-form), complicating comparison between homologs. The crystal structure of SapB adopted a conformation that displayed a smaller internal radius than the other Sap homologs (**Figure 1.20**). Additionally, it has been reported that $SapB^W$ (a Trp point mutant of SapB) forms a smaller dimer than SapA when incubated with double-tailed detergents[30].

1.6.2 – Computational Design of an MSH_B Scaffold

With SapB having a higher dimeric propensity and a smaller volume than SapA, I hypothesized that an MSH_B scaffold (replacing MSH_A's SapA region with SapB) may address the SapA's issue of adopting multiple oligomeric states. Following the design of MSH_A, computational methods were employed to optimize the hydrophobic packing interactions between the MBP C-terminus and the SapB N-terminus.

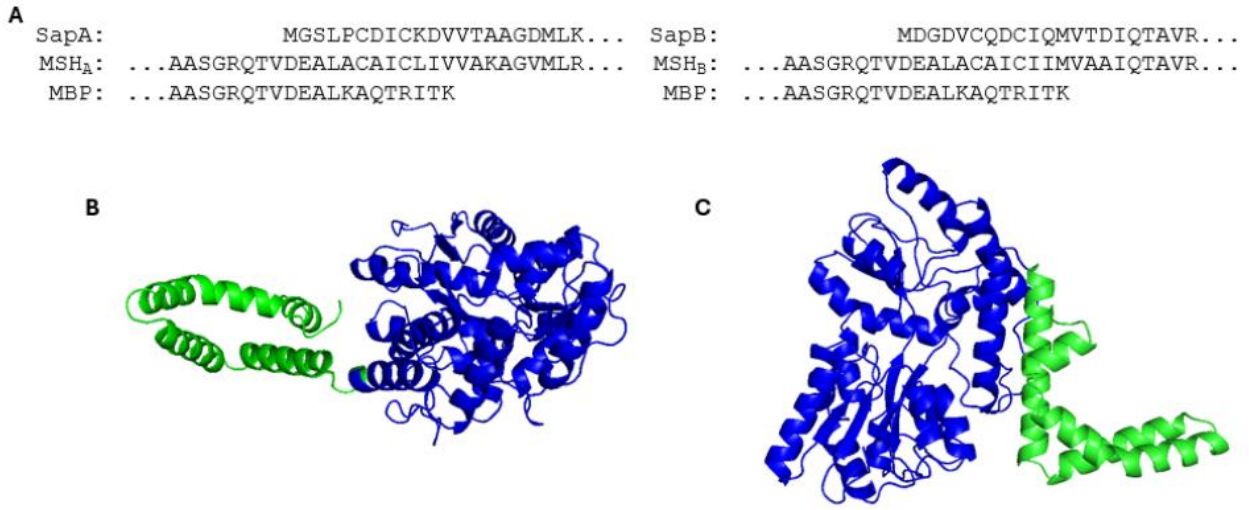


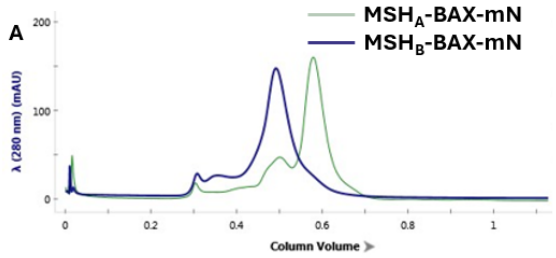
Figure 1.21: A) Sequence alignment of the MSH_A hybridized/fusion region of the N-terminal region of SapA and the C-terminal region of MBP and of the MSH_B hybridized region of the N-terminal region of SapB and the C-terminal region of MBP. B) ColabFold structure of direct fusion of MBP and SapB. C) ColabFold Structure of hybridized fusion of MBP and SapB.

For computational modeling I used ColabFold[80]. Direct fusion of the MBP C-terminal region and the SapB N-terminal region resulted in a computational model displaying no packing between the two domains (**Figure 1.21B**). Instead, the fusion construct resembled closer to the previously reported structures of both proteins with a short linker in between. With the simpler direct fusion approach unable to generate a computational model containing MBP/SapB packing interactions, the existing MSH_A-FGFR3 was used as a reference model. **Figure 1.21A** depicts a sequence alignment of the C-terminal region of MBP with the N-terminal region of SapA or SapB along with the corresponding MSH

sequence. Similar to the approach taken to generate the MSH_A sequence (unpublished), the region of overlap between the MBP and SapB domains is comprised mostly of residues from the MBP sequence while maintaining the cysteine residues used for disulfide bonds within the SapB region. The SapB domain is left largely intact with a few changes intended to facilitate van der Waals packing between the SapB domain and the C-terminal domain of the MBP.

1.6.3 – Revisiting the BCl-2 Interactome

I revisited the encapsulation and solution-phase interrogation of the weak/transient TMD interactions within the BCL-2 family with the MSH_B system. This began by expressing the mNeongreen and mScarletI analogous of BAX and BCL-X_L within the MSH_B scaffold, followed by reconducting the detergent titration experiments with the new scaffold.



B TMD Dimer	IC_{50} (uM)	FE (Start)	FE (End)	ΔFE
BAX-mS/BAX-mN	4	11.6%	6.7%	4.9%
BCL-X_L-mS/BCL-X_L-mN	3	12.2%	7.0%	5.2%
BAX-mS/BCL-X_L-mN	5	9.3%	6.3%	3.0%
BCL-X_L-mS/BAX-mN	3	8.5%	5.4%	3.1%

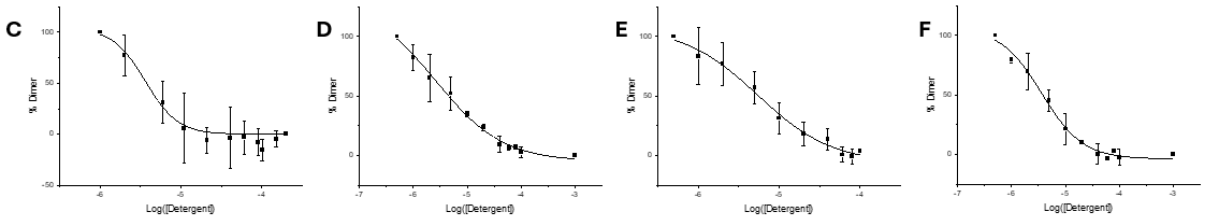


Figure 1.22: (**A**) SEC comparison of MSH_A-BAX-mNeongreen and MSH_B-BAX-mNeongreen at 1 and 5 μM, respectively. (**B**) IC_{50} and FRET efficiency values for each pair of BAX/BCL-X_L homo- and heterodimers. Detergent titration of the (**C**) MSH_B-BAX homodimer, (**D**) MSH_B-BCL-X_L homodimer, (**E**) MSH_B-BAX-mScarletI/MSH_B-BCL-X_L-mNeongreen heterodimer, and (**F**) MSH_B-BCL-X_L-mScarletI/MSH_B-BCL-X_L-mNeongreen heterodimer with LMNG. 280nm absorbance monitored.

When the MSH_B-BAX-mNeongreen was expressed, it existed primarily as a dimer, as opposed to its monomeric MSH_A counterpart (**Figure 1.19A**). This result verified that exchanging the picodisc portion from SapA to SapB did not compromise the ability of MSH_B to dimerize, instead promoting dimerization. With the promising result that SapB replacement is tolerated within the MSH platform, I was delighted to observe that there was a significant increase in FE, compared to MSH_A. Considering that the TMDs are constant from the MSH_A series (**Figure 1.19B-E**), I expected that the orientation of mNeongreen to mScarletI within each MSH construct would be comparable. With FRET being a distance-dependent measurement, the increase in FE should correlate to an increased dimer population. I interpret this as support for my hypothesis that the incorporation of SapB into the MSH platform promotes dimerization relative to SapA. Interestingly, the FE for both versions of the heterodimer was noticeably decreased compared to the BAX or BCL-X_L homodimers. This effect is also seen in the ΔFE over the course of titration for the homo- vs heterodimers, with smaller ΔFE values for heterodimers. This reduction in the ΔFE may be arising from competing equilibria between TMD homo- and heterodimers. As opposed to the TMD homodimer titration experiments which contain only one TMD interaction, i.e. BAX-BAX or BCL-X_L-BCL-X_L, the heterodimer titrations experiments contain three, i.e. BAX-BAX, BCL-X_L-BCL-X_L, and BAX-BCL-X_L. The formation of homodimers within the heterodimer titration experiments allowed for the formation of less FRET capable heterodimers to form. However, since only the heterodimer is capable of FRET, the IC_{50} value should be unaffected.

Although the ability of MSH_B to promote BCL-2 TMD dimerization is increased relative to MSH_A, BCL-2 TMD interactions are still fundamentally weak/transient. When titrating in

detergent, IC_{50} values less than 10 μM for each BCL-2 pairing were observed. The BCL-2 IC_{50} values are weaker than the homodimer detergent titrations of MSH_A-FGFR3 G380R or MSH_A-FGFR3, respectively (**Figure 1.15**). Additionally, plateaus of 100% dimer population at the beginning of a titration experiment were not observed, indicating the concentration of detergent used for equilibration facilitates partial dissociation of BCL-2 TMD dimers. An alternative interpretation could be that since each of the BCL-2 IC_{50} values seem nearly identical, the observed IC_{50} values may correlate to the dissociation of the SapB picodisc within the MSH_B scaffold, not a TMD dimer.

The propensity of the MSH_B platform to adopt a single dimeric state offers a substantial advantage over the first-generation MSH_A platform. As demonstrated in **Figure 1.22**, the MSH_B scaffold regardless of the encapsulated TMD adopts a dimeric state, even at the low concentrations required for FRET (50 nM). Due to this feature, we hypothesis that MSH_B could become an ideal platform for encapsulating TMD dimers of any affinity. However, as highlighted in **Section 1.4.1**, although the MSH_B platform has a substantially higher dimerization affinity and specificity than MSH_A, the picodisc portion of these constructs (SapA and SapB being homologs of one another) should contain a similar sized hydrophobic pocket. As seen in **Figure 1.13**, there should be still quite a large amount of free space within the hydrophobic pocket of MSH_B that could be filled with either detergent or alternative conformations/orientations of the encapsulated TMD. An approach to address this will be described in the final section of this chapter, **Section 1.7**.

1.7 – The Encapsulation of a Larger Oligomer by a Smaller Oligomer

1.7.1 – Combining Non-Covalent and Covalent Encapsulation Approaches

As discussed in **Section 1.6.1**, a large driving motive for designing the MSH_B scaffold was to enhance the crystallization capacity of the MSH platform. To increase the crystallization and structural elucidation potential of a protein, often two features are desired in a protein/protein construct: increased rigidity and conformational homogeneity. The replacement of SapA for SapB within the MSH construct was utilized to increase rigidity with the scaffold. However, the large degree of potential conformational homogeneity within the hydrophobic pocket still raised an issue. The large cavity within the hydrophobic pocket of the Sap homologs offers space for multiple conformations and orientations for the encapsulated TMD to adopt. This conformational heterogeneity could lead to challenges in the meaningful assignment of electron density, if possible. I hypothesized that if the size TMD quaternary structure to be encapsulated was increased the corresponding decrease in free space within the hydrophobic pocket of the MSH_B scaffold would lead to an increase in conformational homogeneity. To this end, I aimed to shift directions from the encapsulation of TMD dimers to trimers.

There is a fundamental limitation of the MSH platform when the TMD of interest's oligomeric state is not dimeric (**Figure 1.23**). The oligomeric state of the TMD of interest must match the oligomeric state of the MSH construct, i.e. TMD dimer must be encapsulated within an MSH dimer. If the encapsulated TMD favored a trimeric state while the MSH_B favored a dimeric state a competing equilibrium would be introduced, where either the

scaffold or TMD must adopt an unfavorable oligomeric state. However, if an additional copy of the TMD was introduced non-covalently from the scaffold via association within the hydrophobic pocket, then both the MSH_B and TMD would be able to adopt their favored oligomeric state simultaneously. To test this theory, I chose the TMD of the Severe acute respiratory syndrome Coronavirus 2 (SC2) spike protein as a model trimer for MSH_B encapsulation.

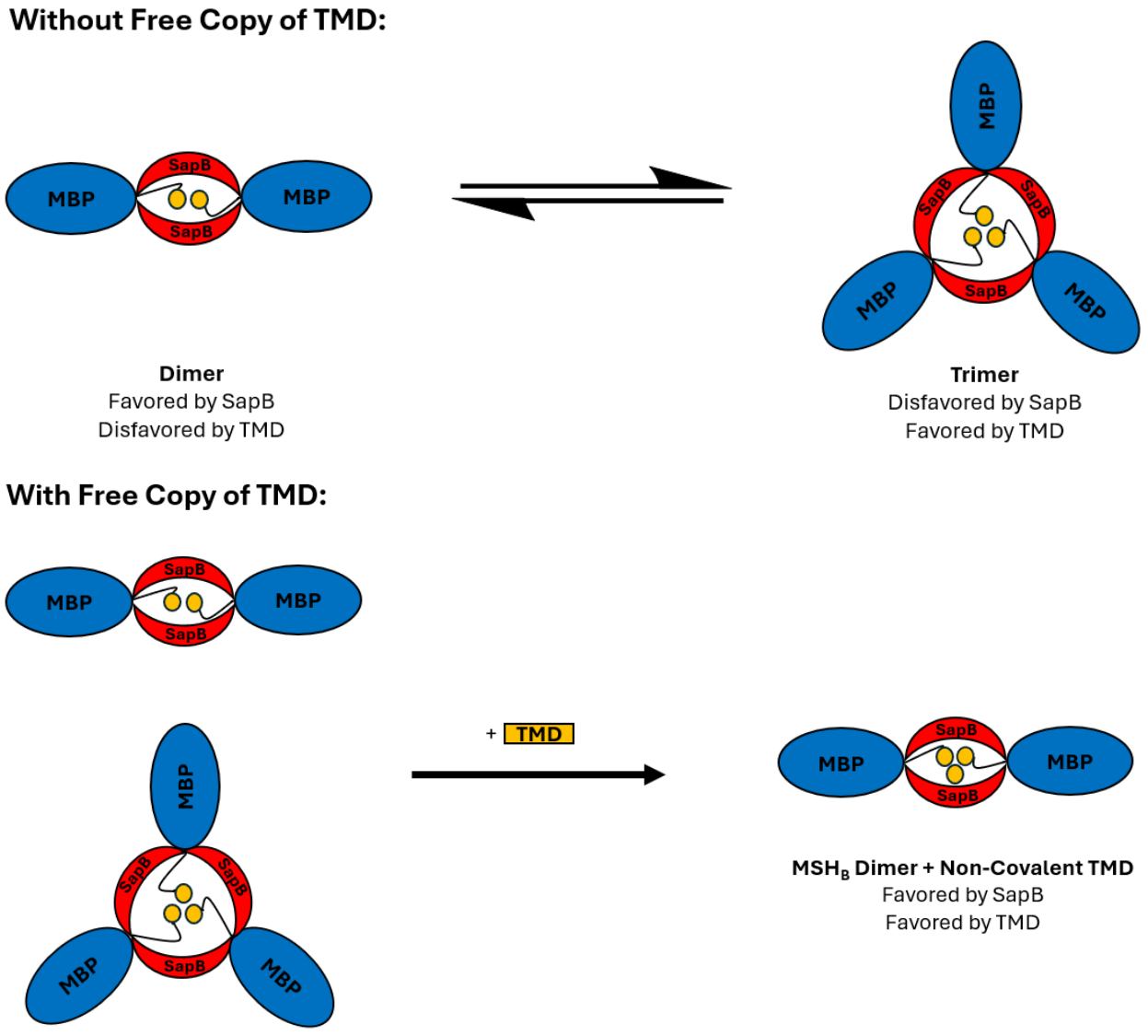


Figure 1.23: Model equilibria of an MSH_B scaffold containing a TMD that prefers the trimeric state without and with a non-covalent copy of the TMD present.

There are currently no crystal structures available of the spike TMD[81]. There is a single NMR structure in detergent micelles reported (PDB: 7LC8). However, the overall chemical shift assignment is relatively poor, at 9%. As the TMD of the protein that facilitates the entry of the SC2 virus through the cellular membrane, the understanding of the quaternary contacts that facilitate the bio-active trimeric state within the membrane would prove invaluable as insight into designing therapeutic approaches for the treatment of SC2[82]. Additionally, successful crystallization and structural elucidation of a TM trimer within the MSH_B scaffold could serve as a general method for the structural elucidation of other TM trimeric associations.

1.7.2 – Characterizing the Oligomeric Flexibility of an MSH_B-SC2 Construct

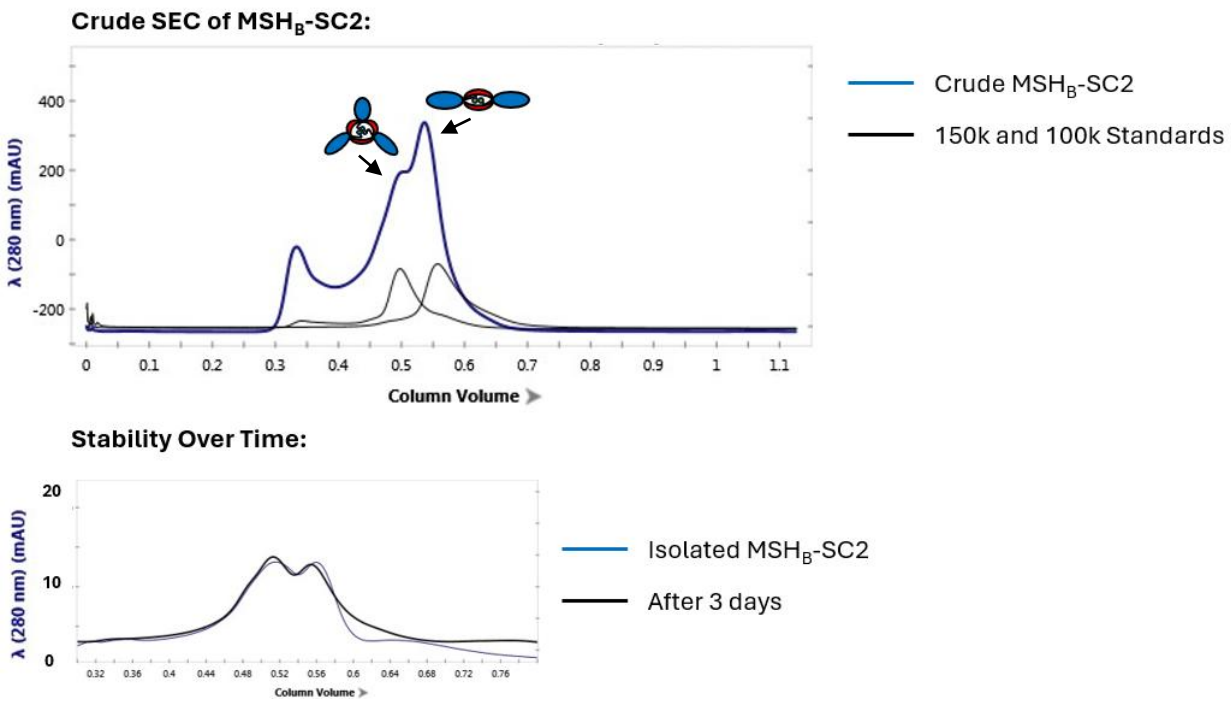


Figure 1.24: (**A**) First SEC trace of MSH_B-SC2 after Ni and amylose affinity purification. (**B**) Analytical SEC to assess the oligomer equilibria stability of MSH_B-SC2 after SEC and after resting at 4°C for 72 hours. 280nm absorbance monitored. Dimer and trimer oligomeric states are indicated by a cartoon of the dimeric or trimeric assembly.

SEC was used to determine the oligomeric state of the MSH_B-SC2 construct without the presence of a non-covalent copy of the SC2 TMD. Most of the material in the protein sample displayed an SEC retention time consistent with species of molecular weights of 100k or 150k (**Figure 1.24**). A single copy of MSH_B-SC2 has a molecular weight of 50k which indicates that the protein predominantly adopts dimeric and trimeric oligomerization states. The displayed mixture of dimeric and trimeric oligomerization states supports the hypothesis of a competing equilibrium between the MSH_B and TMD regions (**Figure 1.23**). Another population of higher-order (greater than trimer) oligomers was observed at 0.34 column volume retention time. These higher-order oligomers are typical byproducts when expressing MSH constructs and are easily removed during the SEC purification. To determine whether the dimer-trimer equilibrium observed was an artifact of the purification process, MSH_B-SC2 was analyzed via SEC immediately after isolation and after 3 days at 4°C. No detectable changes in dimer-trimer ratio (~50:50) were observed, indicating that during the purification process the MSH_B-SC2 construct was at equilibrium.

RRR-SC2: Ac-RRR-WLGFIAGLIAIVLVTILL-NH_2

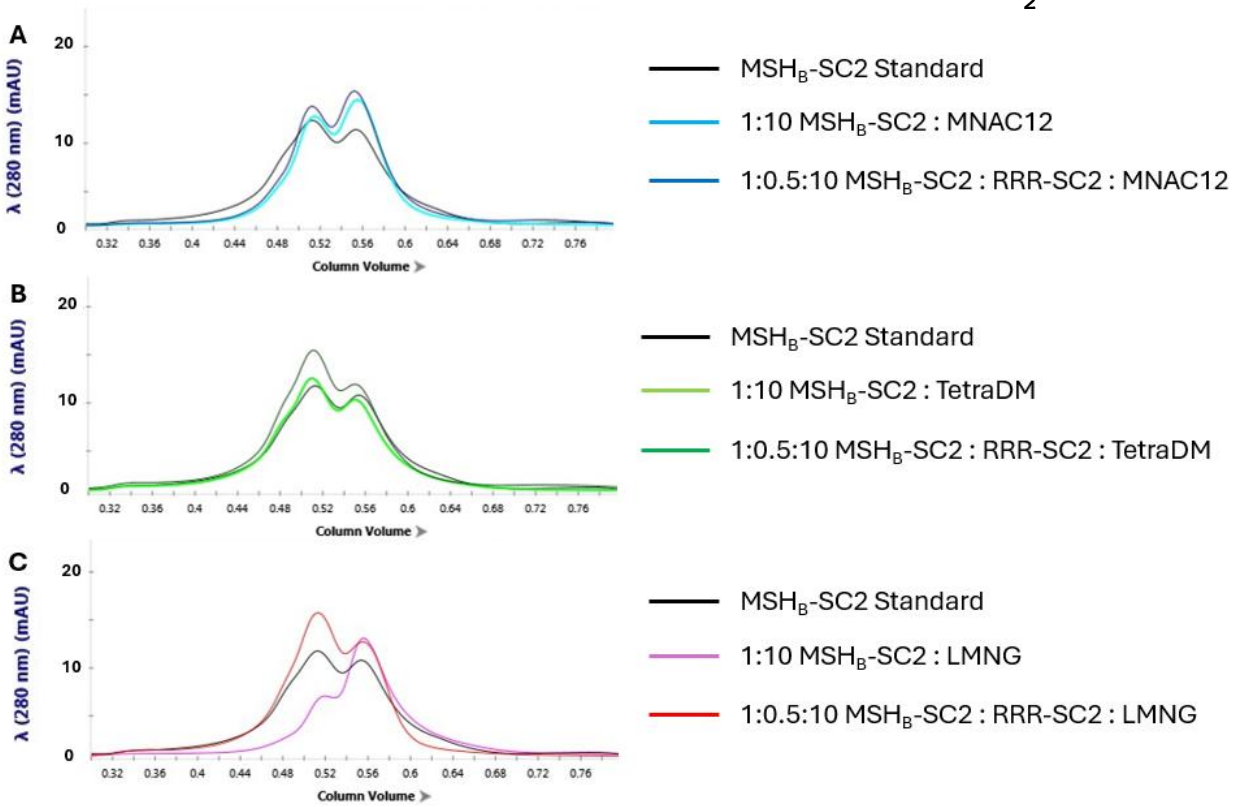


Figure 1.25: Analytical SEC traces of MSH_B-SC2 (1 μM) incubated with (**A**) MNA C12 or MNA C12 with RRR-SC2, (**B**) TetraDM or TetraDM with RRR-SC2, and (**C**) LMNG or LMNG with RRR-SC2 after equilibrating at 4°C for 72 hours. 280nm absorbance monitored.

The dimer-trimer mixture of MSH_B-SC2 was incubated with detergent and non-covalently bound peptides corresponding to the TMD of the SC2 spike protein (RRR-SC2). The peptide included the TMD of the SC2 spike protein along with the endogenous N-terminal juxta-membrane Trp residue and a non-native tri-arginine motif. The tri-arginine motif was added to RRR-SC2 to aid in purification by imparting charged character to the peptide. When MSH_B-SC2 was incubated with either TetraDM or MNA C12, only modest changes in oligomeric state were observed regardless of the presence of RRR-SC2 (**Figure 1.25**). When TetraDM was used, there was very minimal change besides a modest preference for the trimeric state. When incubated with MNA C12 a similar change was observed, but with a slight preference for the dimeric state. Both detergents seemed to have

a minimal impact on the oligomeric state of the MSH_B-SC2 construct, regardless of the presence of RRR-SC2 (**Figure 1.25A-B**).

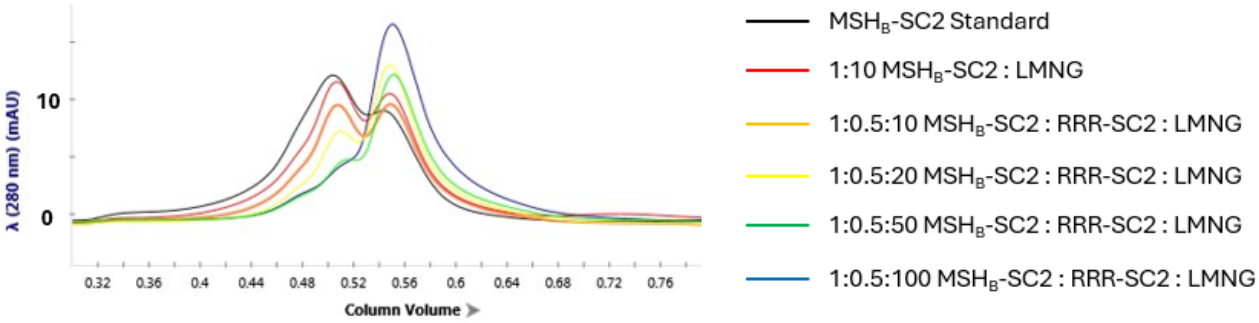


Figure 1.26: Analytical SEC traces of MSH_B-SC2 (1 μM) incubated with increasing ratios of LMNG after 24 hours of equilibration at 4°C. 280nm absorbance monitored.

LMNG, on the other hand, provided an initially counter intuitive result (**Figure 1.25C**). When MSH_B-SC2 was incubated only in the presence of detergent there was a strong shift towards the dimeric state. This is less surprising as we have previously characterized in this chapter that LMNG can associate strongly with Saposin moieties (See **Section 1.3-1.5**). However, when RRR-SC2 was incubated with LMNG, the dimeric shift was abolished. I hypothesized that since the detergent in these experiments plays two roles, facilitating oligomeric transitions of MSH_B-SC2 and solubilization of RRR-SC2, that the detergent facilitating solubilization was sequestering detergent from preforming the oligomeric transition of MSH_B-SC2. A second rendition of this experiment was conducted where MSH_B-SC2 and RRR-SC2 were incubated with varying concentrations of LMNG (**Figure 1.26**). As the concentration of LMNG was increased MSH_B-SC2 was observed more prominently in the dimeric form. The SEC data indicates that higher concentrations of LMNG are required to facilitate the dimer to trimer transition when RRR-SC2 solubilization is happening concurrently.

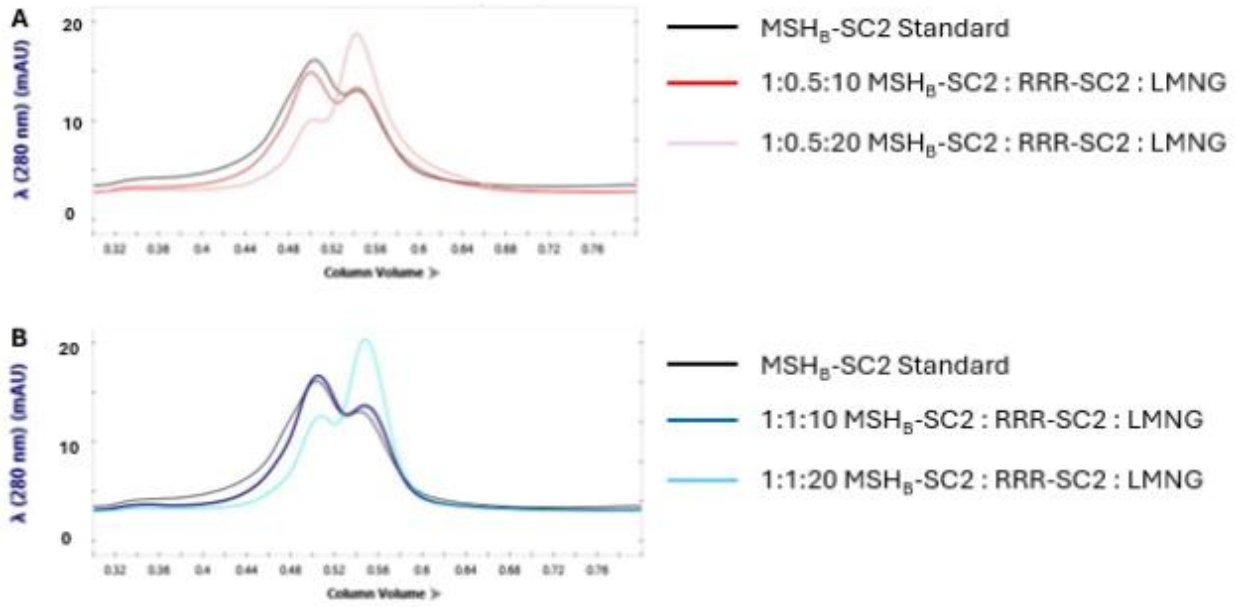


Figure 1.27: Analytical SEC traces of MSH_B-SC2 (1 μM) incubated with a (**A**) 2:1 or (**B**) 1:1 ratio with RRR-SC2. SEC analysis was conducted after equilibrating for 72 hours at 4°C. 280nm absorbance monitored.

With conditions established to promote the dimeric form of MSH_B-SC2 in the presence of RRR-SC2, we next needed to determine whether RRR-SC2 was incorporated with MSH_B-SC2 or remaining solubilized in detergent micelles. To interrogate this, a series of SEC experiments varying the concentration of RRR-SC2 were performed. I hypothesized that if enough detergent was present to facilitate dimer-trimer exchange and solubilization of RRR-SC2, that upon increasing the concentration of RRR-SC2 an increase in the dimeric population of MSH_B-SC2 would be observed. An increase in the dimeric population of MSH_B-SC2 as a function of increasing concentration of RRR-SC2 would indicate the formation of the trimeric RRR-SC2 inside of an MSH_B-SC2 dimer (**Figure 1.26**). Regrettably, I observed no change in oligomeric state when increasing the concentration of RRR-SC2. At both concentrations of RRR-SC2 tested, it was evident that LMNG was the predominant factor in dictating oligomeric state. Although there was no shift in dimeric population as a function of RRR-SC2 added, there was no evidence to suggest that RRR-SC2 was not incorporated

within the MSH_B-SC2 scaffold. The primary conclusion from these SEC experiments is that the oligomeric state of the MSH_B-SC2 scaffold is controlled primarily by protein-detergent interactions.

1.7.3 – Labelled RRR-SC2 Association within the MSH_B-SC2 Scaffold

To determine whether RRR-SC2 was associated with the MSH_B-SC2 scaffold, a labeled form of the RRR-SC2 peptide containing a Trp^{4CN} (RRR-$SC2^{4CN}$) to facilitate FRET transfer was synthesized. The TMD of the MSH_B-SC2 construct contains a native Trp residue that can function as a FRET Donor. The labeled peptide was incubated with the MSH_B-SC2 protein, and LMNG to facilitate equilibration. Once equilibration was completed, the sample was exposed to 270 nm light and the emission spectrum from 300-800 nm was monitored.

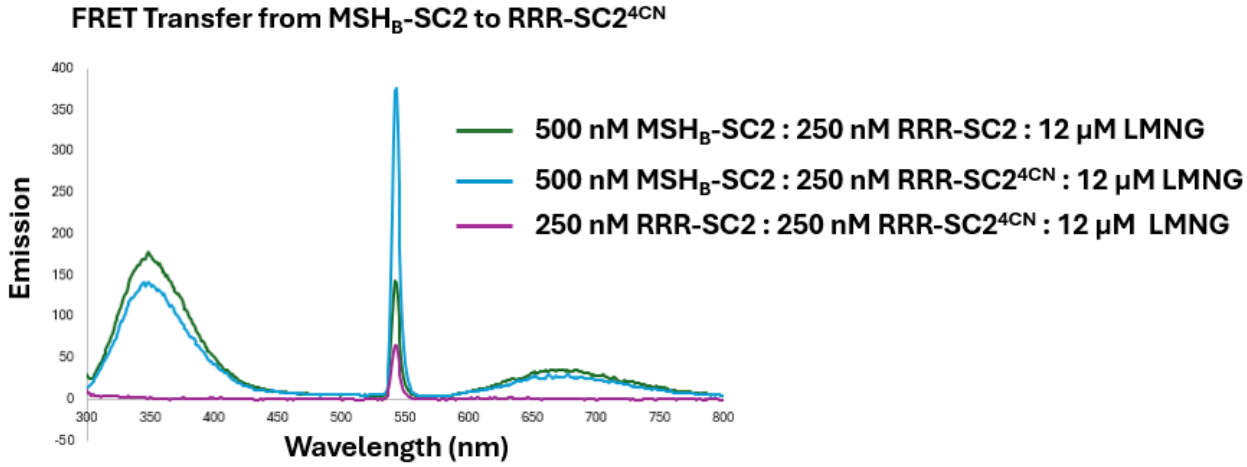


Figure 1.28: FRET emission spectra after excitation with 270 nm light. 280-800 nm range was monitored for FRET acceptor emission. Traces shown are after a 72-hour equilibration period at 4°C for MSH_B-SC2 : RRR-SC2 : LMNG (2:1:80) (green), MSH_B-SC2 : RRR-$SC2^{4CN}$: LMNG (2:1:80) (blue), and RRR-SC2:RRR-$SC2^{4CN}$: LMNG (1:1:80) (purple).

The FRET experiment was conducted as described in **Section 1.3.2**. A large, broad emission with a maximum at 350 nm was observed which correlates with the emission of the native Trp residues (nine residues in each MSH_B-SC2 protein) (**Figure 1.28**). Another emission maximum at 542 nm was observed that correlates with an emission maximum of

$Trp^{4CN\ 56}$. The increase in 542 nm emission when MSH_B-SC2 and RRR-$SC2^{4CN}$ are incubated in the presence of LMNG suggests that the two are associated with one another.

1.8 – Conclusions and Future Directions

In this chapter, I report our attempts at the encapsulation of TMD peptides within a Sap inspired picodisc through non-covalent and covalent approaches. In summary, I discovered that non-covalent encapsulation of a TMD inside of SapA is possible, although with the caveat of high degrees of heterogeneity within the sample. I observed that a variety of assemblies formed when incubating peptide, picodisc, and detergent together. Thus, making complex purification exceedingly difficult and drastically reducing the viability of a non-covalent encapsulation approach for the high-resolution structural determination of TMDs.

After attempting non-covalent encapsulation, I explored covalent encapsulation with an MSH_A scaffold. I found that this approach introduced significant advantages over the former: localization of the picodisc to the TMD, enhanced solubility, and capability of crystallization. The homo-interactions of FGFR3 and FGFR3 G380R were investigated by determining their relative strengths upon detergent dissociation and NMS. I discovered an inverse trend from previous reports[12] which had previously proposed that the WT homodimer was a weaker interaction, where under my conditions the WT proved to have a substantially higher affinity than the mutant. With an approach to interrogate the homodimers established, I moved into two different areas: weak TMD interactions and heterodimers.

Interactions between the TMDs of the BCL-2 family protein, BAX and BCL-X_L, were interrogated for homo- and heterodimer formation utilizing our MSH_A scaffold, but the interactions were too easily dissociated to measure. To address this, a second generation MSH_B scaffold was developed to improve dimer affinity and allow for encapsulation of weak TMD interactions. Using this scaffold we detected all homo- and heterodimeric interactions between the BAX and BCL-X_L TMDs. Although detectable by FRET, the interactions were too transient to determine IC_{50} values during detergent titration experiments.

With the MSH_B scaffold providing a greater population of dimers compared to the MSH_A scaffold, I aimed to utilize this scaffold to encapsulate trimeric TMD interactions utilizing a combined non-covalent/covalent encapsulation approach. I characterized the detergent and peptide effects on the oligomeric state of an MSH_B-SC2 construct. I established methods to promote the dimeric state over the trimeric state and utilized FRET to verify the encapsulation of a labeled peptide within the dimeric picodisc.

The MSH_B scaffold represents an alternative version of the MSH_A scaffold with an improved ability to adopt a dimeric state. I hypothesis that SapB's selectivity for the dimeric picodisc state drives the greater population of MSH_B dimers observed, compared to MSH_A . This selectivity provides a more reliable route to form stable picodiscs containing TMD quaternary contacts between homo- and heterodimers and larger trimeric TMD assemblies. Looking forward, the next major goal to be addressed will be the crystallization potential of the MSH_B scaffold. Preliminary crystal trials have provided micro crystal formation using commercial crystal screening kits of the MSH_B-SC2 scaffold with RRR-SC2, but high-resolution structural determination of TMD's within an MSH_B scaffold is yet to be completed.

1.9 – Methods and Characterization

1.9.1 – SapA Expression and Purification

The DNA Sequence encoding SapA (*H. sapiens*, Uniprot entry P07602) was codon-optimized by Genscript for expression in *E. coli*. The DNA sequence of SapA is shown below with the 6x-His tag, thrombin cleavage site, and enzymatic degradation sites are highlighted in blue, yellow, and green, respectively:

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCA

TATGCTCGAGAGCCTGCCGTGCGACATCTGCAAGGATGTGGTTACCGCGGCGGGTGACAT

GCTGAAAGATAACGCGACCGAGGAAGAGATTCTGGTGTACCTGGAAAAGACCTGCGACTG

GCTGCCGAAACCGAACATGAGCGCGAGCTGCAAGGAGATCGTGGACAGCTATCTGCCGG

TTATCCTGGATATCATCAAGGGTGAAATGAGCCGTCCGGGCGAGGTTTGCAGCGCGCTGA

ACCTGTGCGAAAGCTAAGGATCCGGCTGCTAACAAAGCCCGA

The resulting plasmid was transformed into *E. coli* SHuffle cells (New England Biolabs Inc.). SHuffle cells were used as they are an engineered *E. coli* strain with a non-reducing cytoplasm. This environment is critical for forming the three sets of disulfide bonds that Sap proteins require for folding.

Starter cultures for protein expression were grown by adding 10 μL of a thawed glycerol stock of SHuffle cells containing the SapA plasmid (stored at -80°C) into 40 mL of sterilized Terrific broth media (47.6 g/L Terrific broth, 4 mL/L glycerol, 0.1 mg/mL ampicillin). The starter culture was lightly shaken at 37°C overnight (~14 hours). 4 mL/L of the starter

culture was added to 2 L sterilized baffled flasks containing 1 L of sterilized Terrific broth media. The cultures were then shaken at 37°C until an optical density at 600 nm (OD_{600}) value between 0.4-0.8 was measured (~4 hours). The cell media was then left at 4°C for one hour to halt cell growth. Induction was then performed by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The induced cell cultures were then allowed to shake overnight (~14 hours) at 18°C. Once the induction period had concluded, the cells were harvested by centrifugation at 5k rpm for 10 minutes.

The cells were resuspended in lysis buffer (150 mM NaCl, 25 mM Tris-HCl, pH 7.5, 0.2 mg/mL lysosome, 2 mM $MgCl_2$, and 20 µg/mL DNase) and subjected to lysis by sonication at 4°C. The lysates were centrifuged at 10k rpm for 10 minutes and the supernatant was collected. The isolated supernatants were then heated at 80°C for 10 minutes. Since SapA remains soluble at high temperatures, this allows for the precipitation of proteins that aggregate upon heating. The solution was then cooled to 4°C and subjected to another round of centrifugation at 10k rpm for 45 minutes. The supernatant was collected for affinity purification.

The collected supernatant was loaded onto a 5 mL His-Trap column (Cytiva). Once all the supernatant was loaded onto the Ni column, SapA was eluted using a linear gradient of increasing elution buffer (150 mM NaCl, 25 mM Tris-HCl, 500 mM Imidazole, pH 8; 3-50% elution buffer) performed on an NGC Quest FPLC system (Bio-Rad). Following the Ni affinity purification, the SapA fractions were incubated with thrombin (1.5kU/500nmol SapA) overnight at room temperature with gentle mixing to remove the N-terminal His tag.

After thrombin removal, the samples were heated at 80°C for 1 hour to denature and precipitate the thrombin. The solution was then centrifuged at 10k rpm for 30 minutes, and the supernatant was collected. The cleaved protein was first purified through affinity chromatography by passing the supernatant back through the Ni column and collecting the flow through. The cleaved protein was then further purified by SEC on a Superdex 75 16/600 120 mL (Cytiva). The elution buffer used for this purification step was 150 mM NaCl, 25 mM Tris-HCl, pH 7.5. The SapA was purified as its monomeric form and concentrated using a 5k molecular weight cutoff (MWCO) Amicon size exclusion filter. The concentrated protein was then stored at 4°C until use.

1.9.2 – Synthesis of 4-Cyanotryptophan

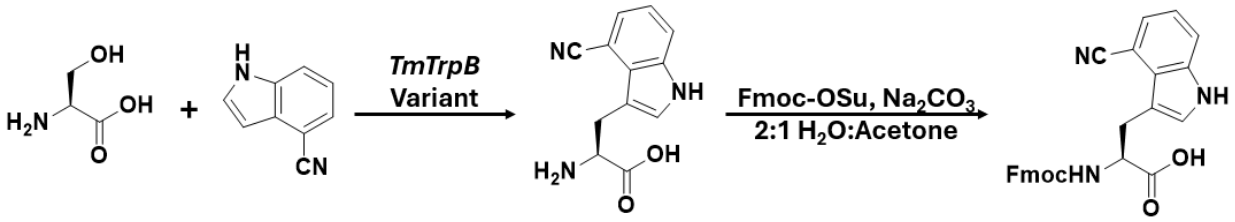


Figure 1.29: Synthetic scheme for the synthesis of Fmoc-Trp^{4CN}-OH

H-Trp^{4CN}-OH synthesis from serine and 4-cyanoindole was conducted using a previously reported literature synthesis[84].

Fmoc-Trp^{4CN}-OH was synthesized by dissolving crude H-Trp^{4CN}-OH (300 mg, 1.31 mmol) in a 2:1 mixture of 0.1 M $Na_2CO_{3(aq)}$ and acetone (0.1 M). Fmoc-OSu (660 mg, 1.96 mmol) was added to the reaction mixture, which was allowed to stir at room temperature overnight. The reaction mixture was diluted with 50 mL of water, and the acetone was removed under vacuum. The aqueous layer was then washed with diethyl ether (2x100 mL)

to remove unreacted Fmoc-OSu. The pH of the aqueous layer was adjusted to 1 with 1 M $HCl_{(aq)}$ to induce precipitation of the product. The product was extracted with ethyl acetate (2x100 mL), the organic layer was dried over Na_2SO_4 and concentrated *in vacuo* to afford the crude product. The product was then redissolved in the minimal amount of ethyl acetate and then diluted with heptane to induce recrystallization. The sample was then left at 4°C overnight. The following day the solution was filtered and the solid was washed with heptane and allowed to dry under aspiration. 616 mg of the resulting off-white solid was recovered and used without further purification for SPPS.

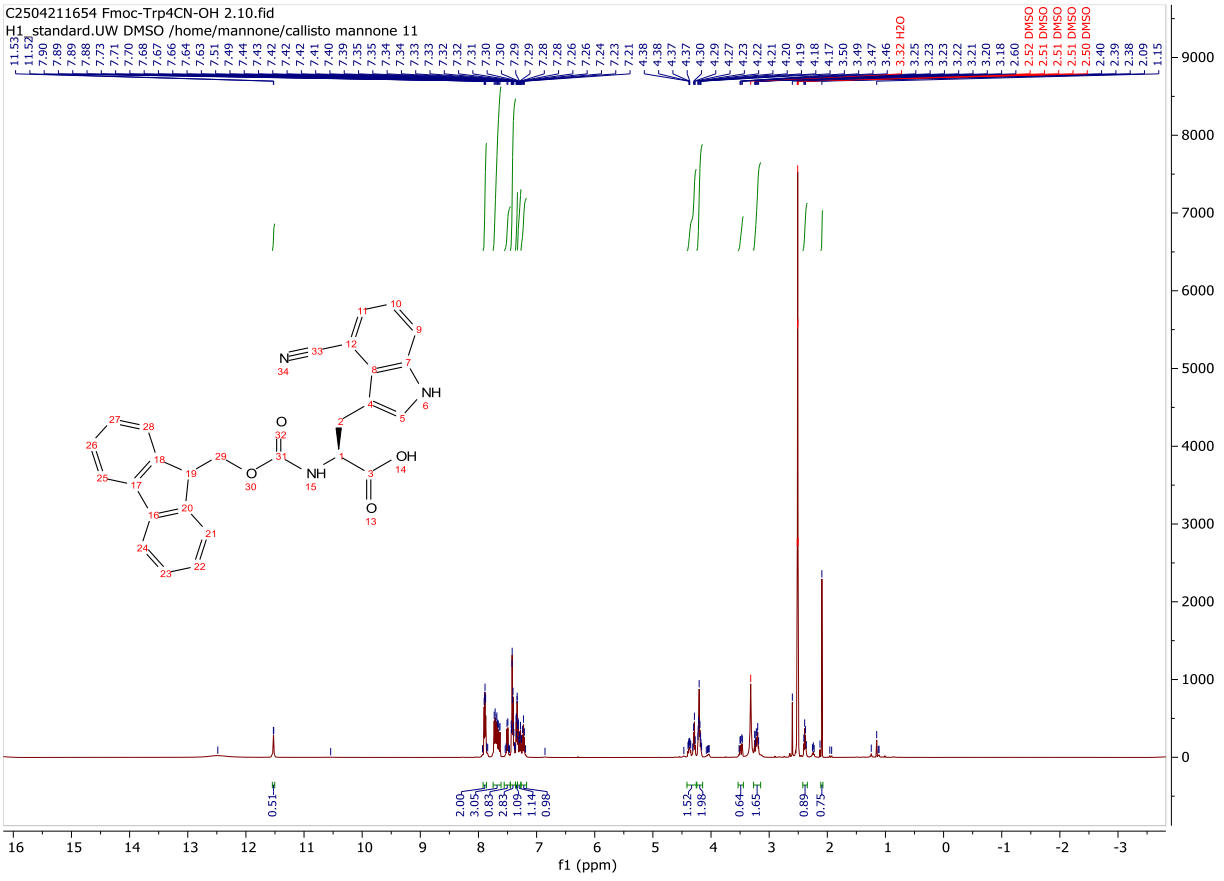


Figure 1.30: ^{1}H-NMR of Fmoc-Trp^{4CN}-OH in *d6*-DMSO. 500mHz

1.9.3 – Peptide Synthesis, Purification and Characterization

All TMD peptides were synthesized via automated microwave-assisted reactions on the Liberty Blue Peptide Synthesizer (CEM) on Low Load (LL) Fmoc-Rink Amide Resin at 50 or 100 µmol scale. For each coupling reaction, Fmoc-protected amino acid (5 eq), Oxyma (5 e, and DIC (10eq) in biotech-grade DMF were mixed with resin, and the mixture was heated to 75°C for 10 minutes while stirring. After the coupling period, the reaction vessel was drained and washed with ACS-grade DMF (3x10 mL), and the coupling reaction was repeated (double coupling). After the second coupling reaction, the resin was washed with ACS-grade DMF (3x10 mL). For Fmoc-deprotection, 20% (v/v) piperidine in ACS-grade DMF (5 mL) was added to the resin, and the mixture was heated at 80°C for 2 minutes while stirring. After deprotection, the resin was washed with ACS-grade DMF (3x10 mL). Once peptide elongation was completed, a final deprotection reaction was completed, the resin was washed with ACS-grade DMF (3x10 mL), and then 4 mL of an acetylation cocktail (8:2:1 BioTech-grade DMF : Acetic Anhydride : DIEA) was added. The mixture was allowed to stir for 30 minutes at room temperature. The resin was then washed with ACS-grade DMF (3x10mL), followed by DCM (3x1 0mL), transferred to an SPPS vessel, and placed under reduced pressure on an aspirator until dry. Peptides were then cleaved from the resin, with concomitant side chain deprotection, by combining the resin with 2 mL per 25 umol resin of Reagent K cleavage cocktail (82.5% TFA, 5% water, 5% thioanisole, 5% phenol, and 2.5% 1,2-ethanedithiol), and gently rocking the vessel for 3 hours (AI series) or 5 hours (RRR-SC2 series). The reaction solution was then drained into a 45mL falcon tube, and the resin was washed twice with 2 mL TFA. The combined peptide cleavage solution was then

concentrated under a stream of nitrogen gas until less than 1 mL remained or precipitation of the peptide was observed. The crude peptide was then precipitated with 40 mL of cold diethyl ether. This mixture was centrifuged at 4k rpm for 10 minutes to pellet the suspended solid, and then the ether was decanted. A second volume of 40 mL cold diethyl ether was added, and the process of centrifugation and decantation was repeated. The resulting solid was dried under a stream of nitrogen gas to afford the crude product.

To purify the crude material, the solid peptide was dissolved in the minimal volume of ACN/H_2O (AI series) or DMSO (RRR-SC2 series) required to fully dissolve the material (1-4 mL depending on peptide solubility) and passed through a 0.22 µm filter. Purification was conducted via reverse-phase high performance liquid chromatography (RP-HPLC). For the AI_5 and AI_{10} peptides, and their Trp^{4CN} derivates, the following conditions were used: CSH C18 stationary phase, 5 µm, 19 mm x 250 mm, the solvent gradient was 10-60% acetonitrile over 20 minutes. For RRR-SC2 and its Trp^{4CN} derivative the following conditions were used: CSH C4 stationary phase, 5 µm, 19 mm x 250 mm, the solvent gradient was 10-80% acetonitrile over 20 minutes. The purity of each AI peptide was assessed via ultra-high performance liquid chromatography (UPLC; CSH C18 stationary phase, 1.7 µm, 2.1 mm x 100 mm, the solvent gradient was 10-60% acetonitrile over 10 minutes). The purity of each RRR-SC2 peptide was assessed via ultra-high performance liquid chromatography (UPLC; BEH C4 stationary phase, 1.7 µm, 2.1 mm x 100 mm, the solvent gradient was 10-80% acetonitrile over 10 minutes). Peptide purity was assessed by integration of area under the curve while monitoring at 220 nm. Peptide concentration was determined by absorbance at 280 nm. Each peptides molar extinction coefficient was determined using previously reported

literature values of proteogenic amino acids at 280nm[85]. The peptide solution was then dissolved in water (if needed, with a minimal amount of acetonitrile), frozen, and lyophilized to provide a white powder. The material was then stored at 4 °C until use.

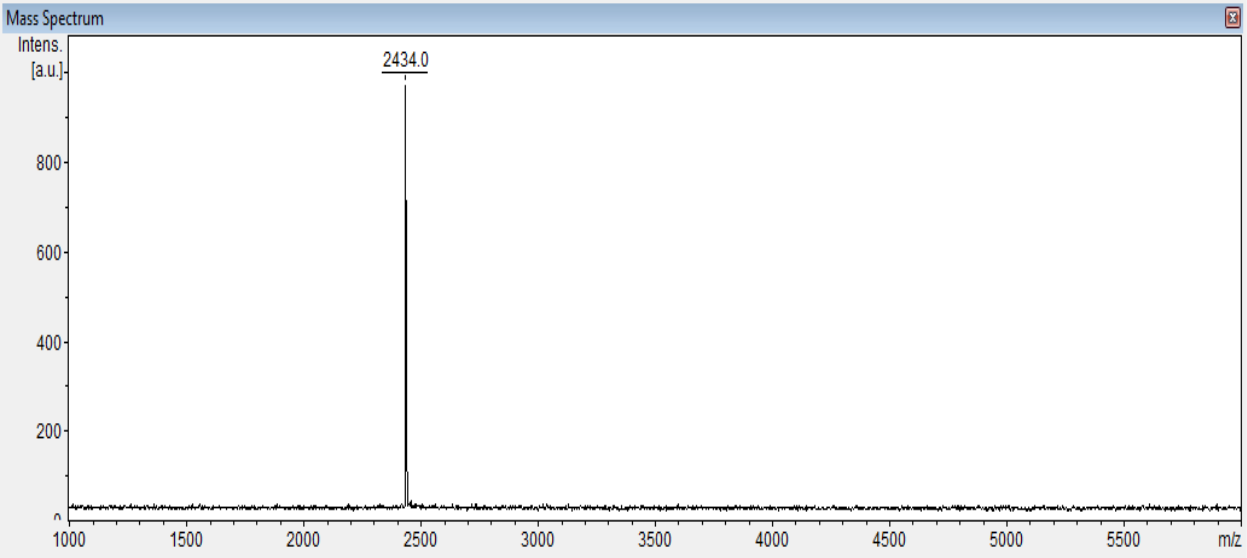


Figure 1.31: MALDI-TOF characterization of RRR-SC2. Observed: $[M+H]^+$ = 2434.0, Calculated: $[M+H]^+$ = 2433.55

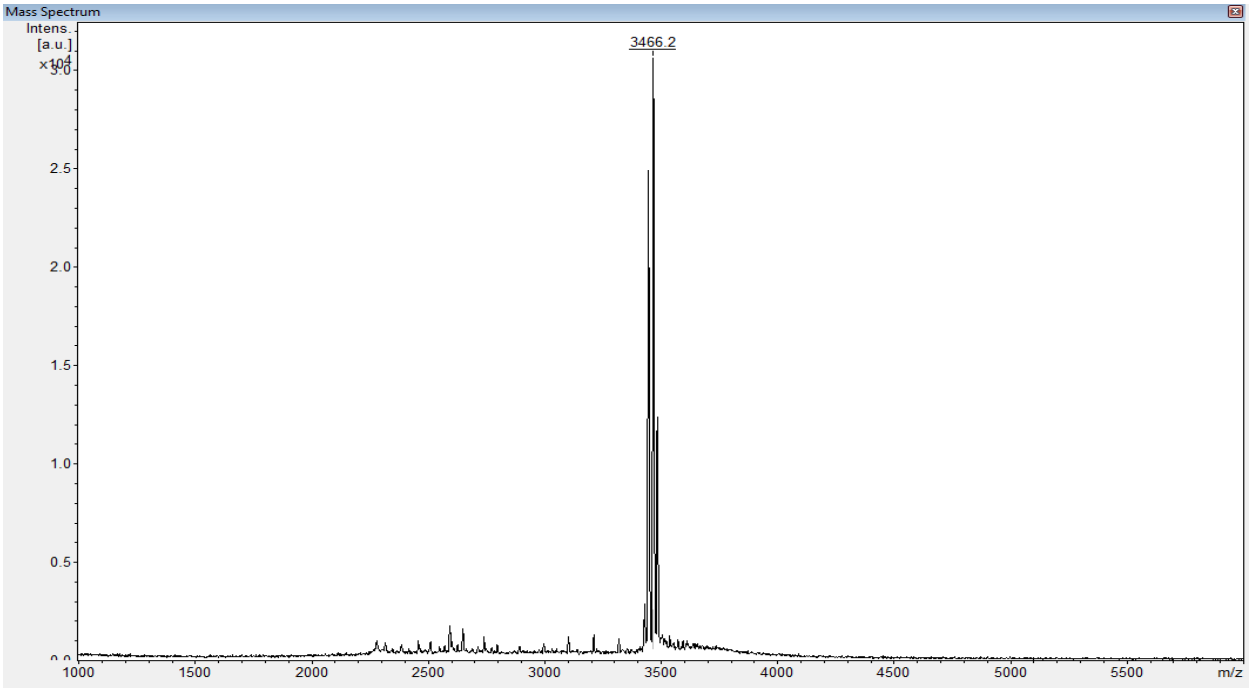


Figure 1.32: MALDI-TOF characterization of Al_{10}. Observed: $[M+Na]^+$ = 3466.2, Calculated: $[M+Na]^+$ = 3467.5

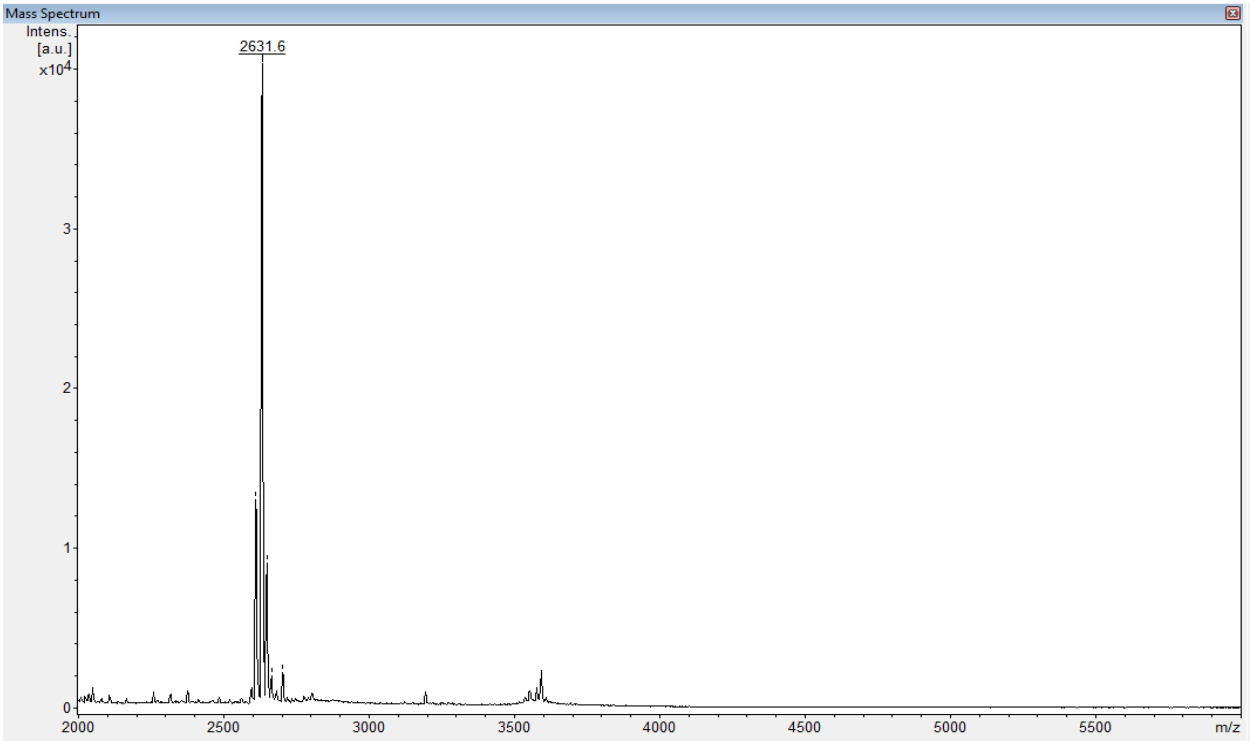


Figure 1.33: MALDI-TOF characterization of Al_5 Observed: $[M+Na]^+$ = 2631.6, Calculated: $[M+Na]^+$ = 2632.2

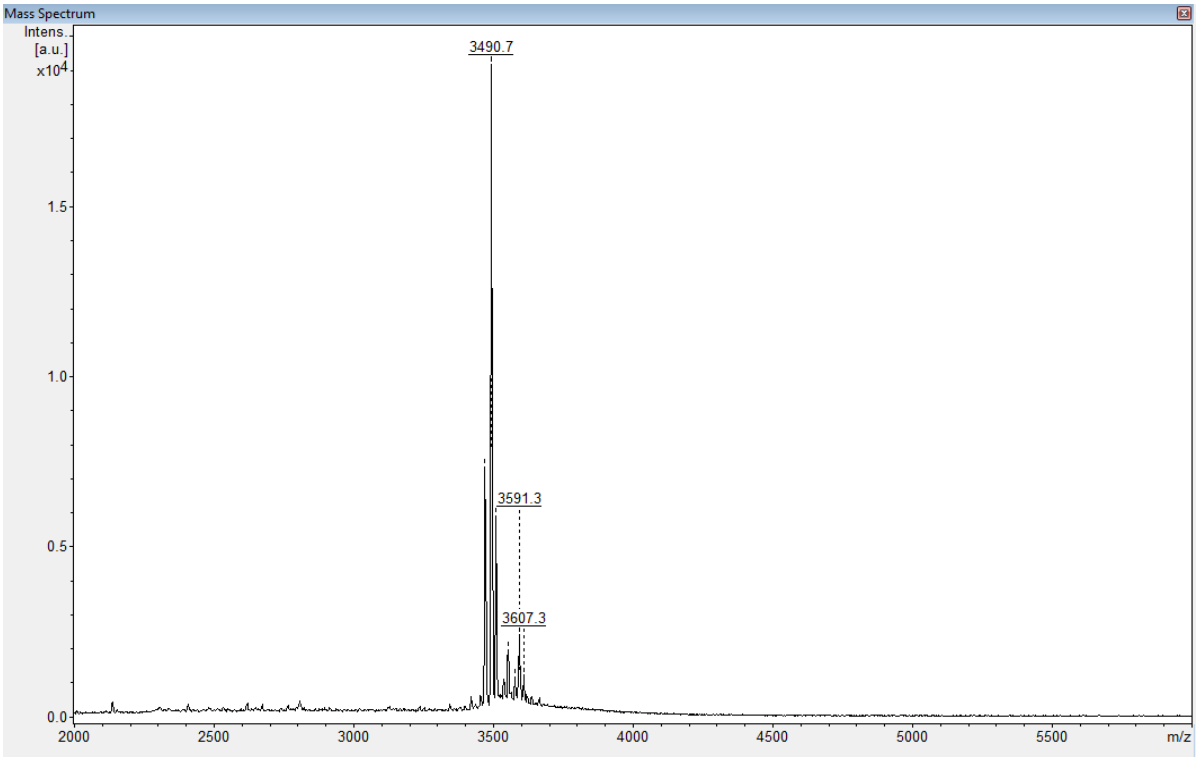


Figure 1.34: MALDI-TOF characterization of Al_{10}^{4CN}. Observed: $[M+Na]^+ = 3490.7$, Calculated: $[M+Na]^+ = 2491.3$

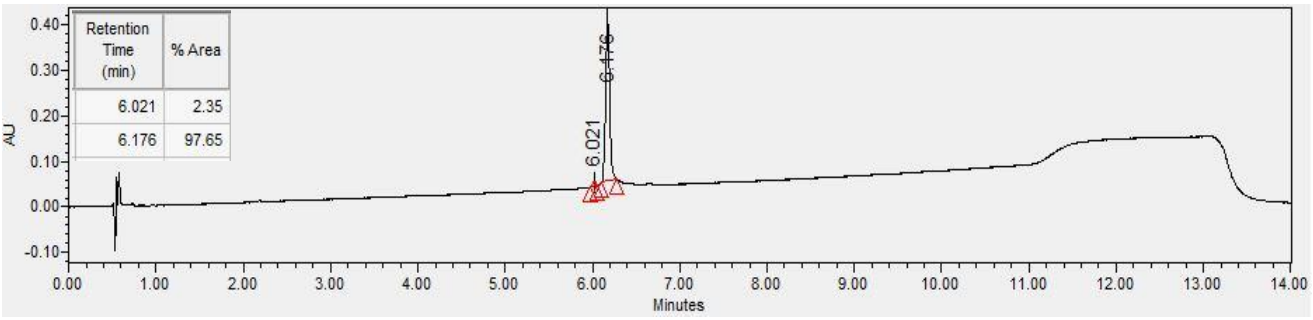


Figure 1.35: UPLC chromatogram of RRR-SC2 monitored at 220 nm. Purity >97%.

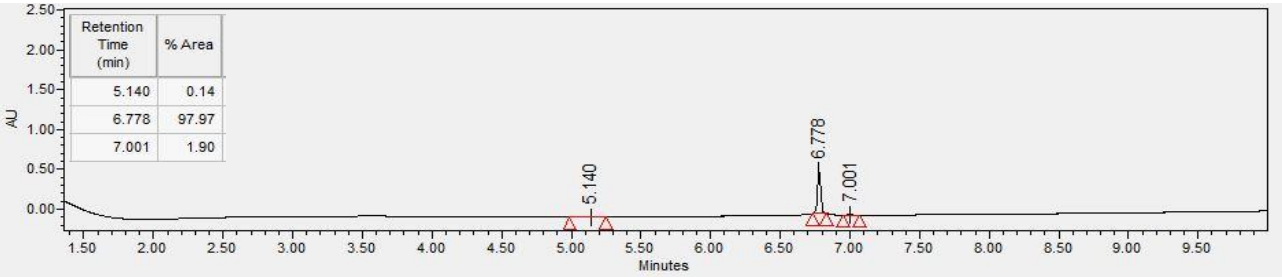


Figure 1.36: UPLC chromatogram of AI_5 monitored at 220 nm. Purity >97%

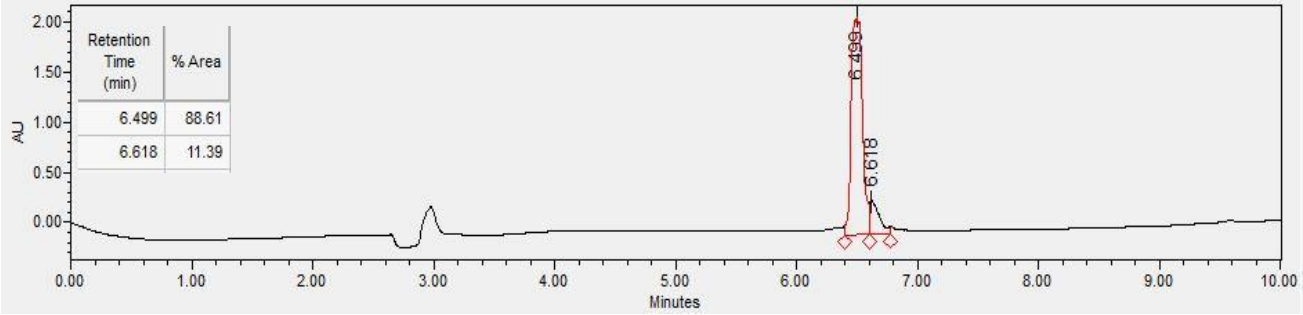


Figure 1.37: UPLC chromatogram of Al_{10}^{4CN} monitored at 220 nm. Purity >88%

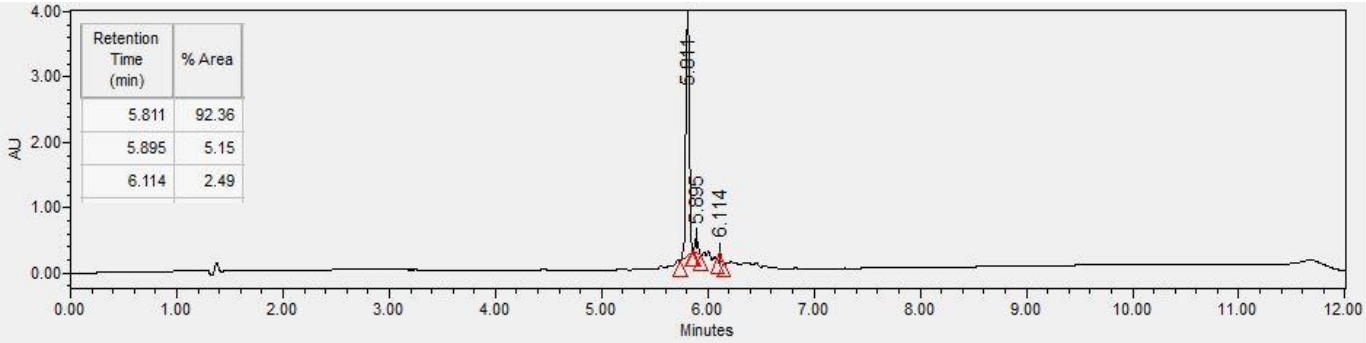


Figure 1.38: UPLC chromatogram of AI_{10} monitored at 220 nm. Purity >92%

1.9.4 – Analytical SEC of SapA Picodiscs

AI_5 and AI_{10} peptides were dissolved in a buffer solution of 150 mM NaCl, 25 mM Tris-HCl, and 1% DM at pH 7.5. The purified monomeric SapA and solubilized AI peptide (50 nmol each) were then combined and mixed with MNA C12 (10 eq), LMNG (15 eq), or TetraDM (20 eq), followed by a dilution to 1 mL total volume (50 μM of SapA and AI peptide). The equivalents of detergent are based on previously reported literature values[29]. The solutions were then incubated at 37 °C for 45 minutes to facilitate picodisc formation. The samples were then cooled to 4 °C and then loaded onto a Superdex 75 10/300 GL column (Cytiva) and eluted with a buffer of 150mM NaCl and 25mM Tris-HCl at pH 7.5.

1.9.5 – FRET Assay for SapA Picodisc

AI_{10}^{4CN} was dissolved in a buffer solution of 150 mM NaCl, 2 5mM Tris-HCl, and 1% DM at pH 7.5. The purified monomeric SapA and solubilized AI_{10}^{4CN} peptide (1:2) were then combined and mixed with MNA C12 (20 eq), LMNG (30 eq), or TetraDM (40 eq) and diluted with the 150 mM NaCl, 25 mM Tris-HCl at pH 7.5 to a concentration of SapA at 12.5 μM. The solution was then incubated at 37°C for 45 mins to facilitate picodisc formation. The solution was then diluted down to 1 μM and excited with 270 nm light. The fluorescent emission was monitored from 280-500 nm.

1.9.6 – Plasmid Construction and General Method for MSH Protein Expression and Purification

Our group has previously developed a modified pETDuetI vector containing an encoding domain for MSH and analogues with a C-terminal His tag. DNA fragments used for

exchanging MSH_A for MSH_B and exchanging TMDs were purchased from Twist Bioscience, and inserted between HindIII and NotI, and EcoRI and NotI sites of a pETDuet1 vector, respectively. The resulting plasmids that overexpress the MSH analogues were transformed into *E. coli* SHuffle cells, and the proteins were isolated using the following procedure:

Starter cultures for protein expression were grown by adding 10 µL of a thawed glycerol stock of SHuffle cells containing the MSH plasmid (stored at -80°C) into 40 mL of sterilized Terrific broth media (47.6 g/L Terrific broth, 4 mL/L glycerol, 0.1 mg/mL ampicillin). The starter culture was lightly shaken at 37°C overnight (~14 hours). 4 mL/L of the starter culture was added to 2 L sterilized baffled flasks containing 1 L of sterilized Terrific broth media. The cultures were then shaken at 37°C until an optical density at 600 nm (OD_{600}) value of 0.8 was measured (~4 hours). The cell medium was then left at 4°C for one hour to halt cell growth. Induction was then performed by the addition of IPTG to a final concentration of 1 mM. The induced cell cultures were then allowed to shake overnight (~14 hours) at 18°C. Once the induction period had concluded, the cells were harvested by centrifugation at 5k rpm for 10 minutes.

The cells were resuspended in lysis buffer (300 mM NaCl, 50 mM Tris-HCl, pH 7.5, 0.1% v/v Triton X-100, 2 mM $MgCl_2$, and 20 µg/mL DNase) and allowed to stir at 4°C overnight (~14 hours). The lysates were centrifuged at 10k rpm for 10 minutes and the supernatant was collected. The collected supernatant was loaded onto a 20 mL HisPrep FF column (Cytiva). The loaded Ni column was then washed with 3 column volumes of wash buffer (300 mM NaCl, 5 0mM Tris-HCl, pH 7.5). The MSH protein was eluted with 300 mM NaCl, 50 mM Tris-HCl, and 500 mM Imidazole at pH 8.0 onto an amylose column. The amylose column was

then washed with 2 column volumes of wash buffer and then eluted with 300 mM NaCl, 50 mM Tris-HCl and 1 mM maltose at pH 7.5. The protein was subsequently concentrated using a 50k Da Molecular Weight Cutoff (MWCO) Amicon filter. The concentrated protein was then loaded onto a Superdex 200 10/300 GL column (Cytiva) for a final purification with wash buffer. The protein was then concentrated using 50k Da MWCO Amicon filter and stored at 4°C until use.

1.9.7 – Analytical SEC of MSH Constructs

Purified MSH constructs were diluted in a buffer solution of 300 mM NaCl and 50 mM Tris-HCl at pH 7.5 to an indicated final concentration (typically 500 nM, 1 µM, or 5 µM). Once diluted the samples were loaded onto a Superdex 200 10/300 GL column (Cytiva) and eluted with a buffer of 300mM NaCl and 50mM Tris-HCl at pH 7.5. Absorbance monitored at 280nm.

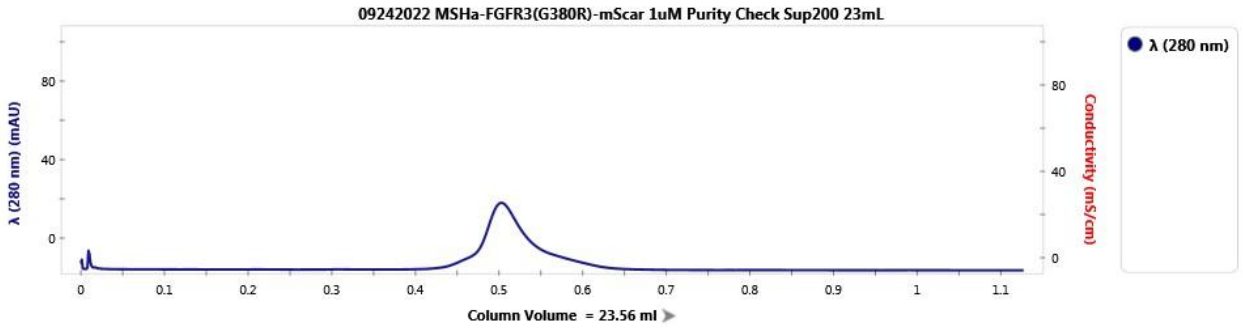


Figure 1.39: 1 μM SEC trace of purified MSH_A-FGFR3 G380R-mNeongreen.

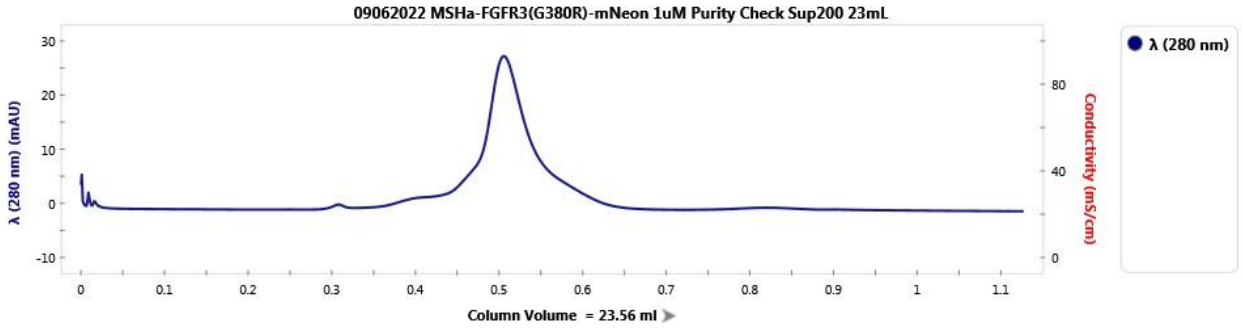


Figure 1.40: 1 μM SEC trace of purified MSH_A-FGFR3 G380R-mScarletI

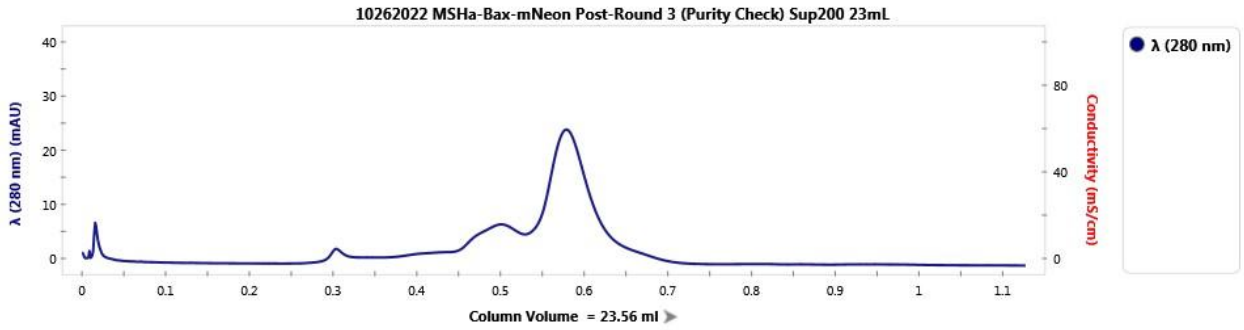


Figure 1.41: 1 μM SEC trace of purified MSH_A-BAX-mNeongreen.

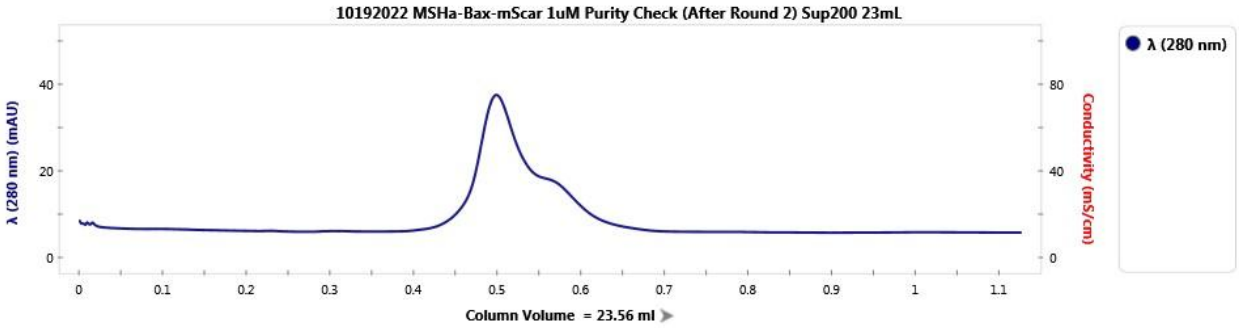


Figure 1.42: 1 µM SEC trace of purified MSH_A-BAX-mScarletI

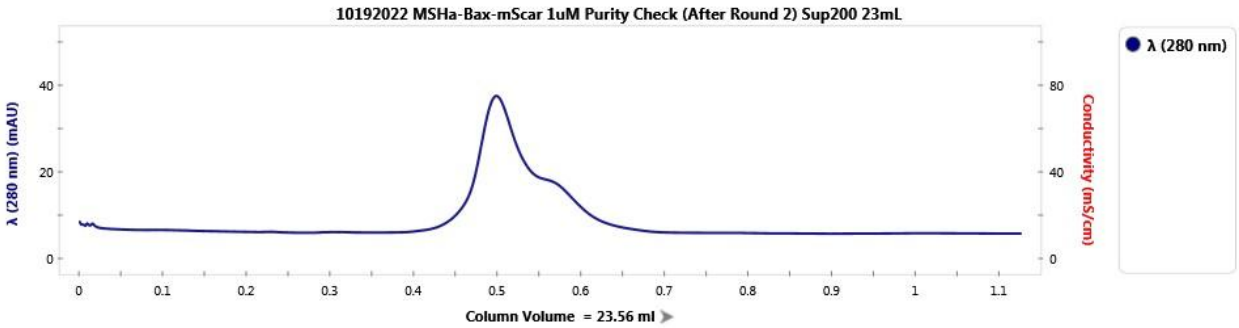


Figure 1.43: 1 μM SEC trace of purified MSH_A-BCL-X_L-mNeongreen

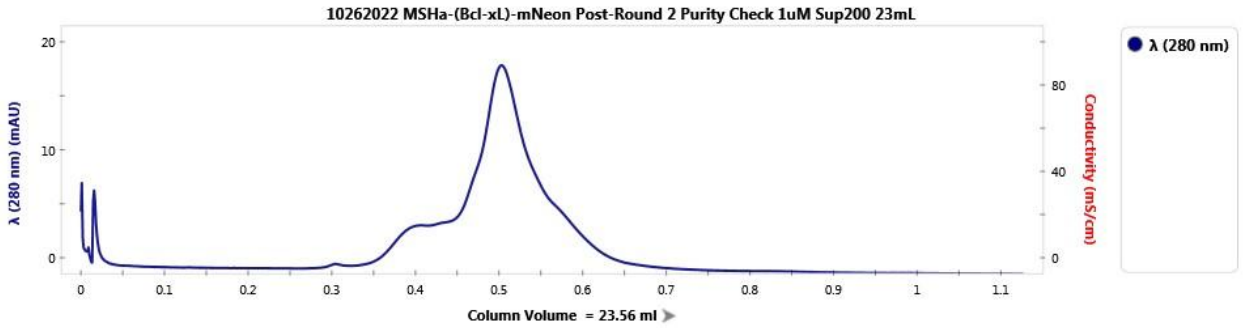


Figure 1.44: 1 μM SEC trace of purified MSH_A-BCL-X_L-mNeongreen

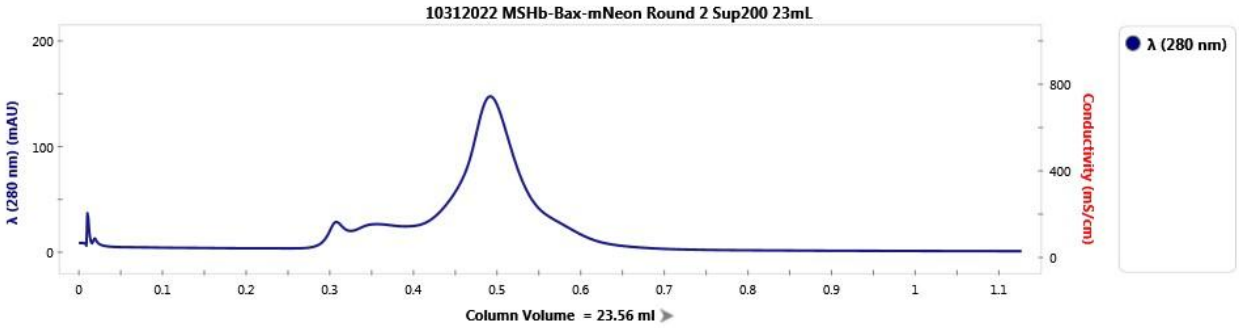


Figure 1.45: 5 µM SEC trace of purified MSH_B-BAX-mNeongreen

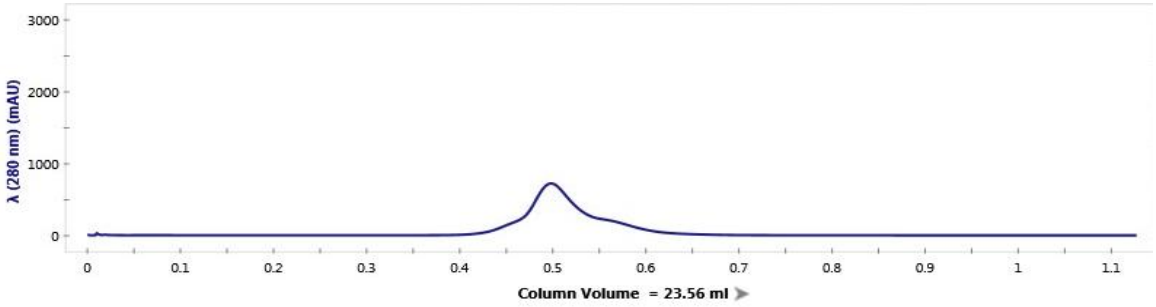


Figure 1.46: 50 µM SEC trace of purified MSH_B-BAX-mScarletI

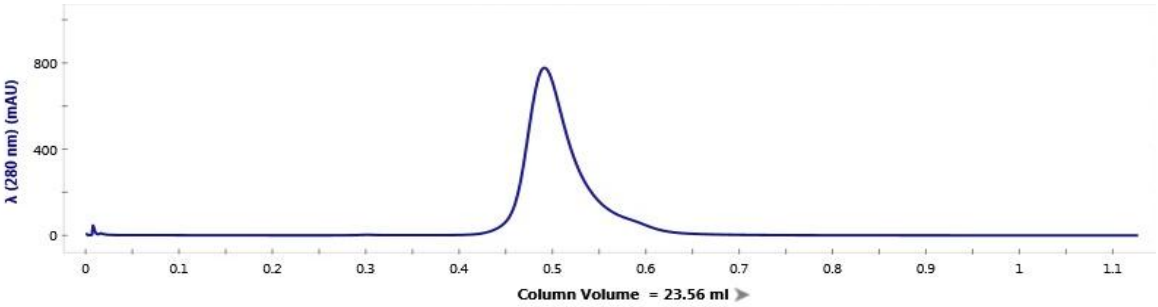


Figure 1.47: 50 μM SEC trace of purified MSH_B-BCL-X_L-mNeongreen

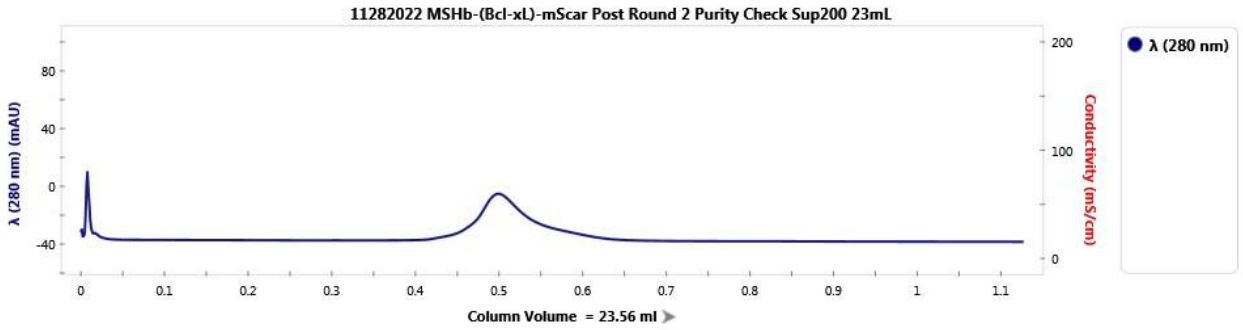


Figure 1.48: 1 µM SEC trace of purified MSH_B-BCL-X_L-mScarletI

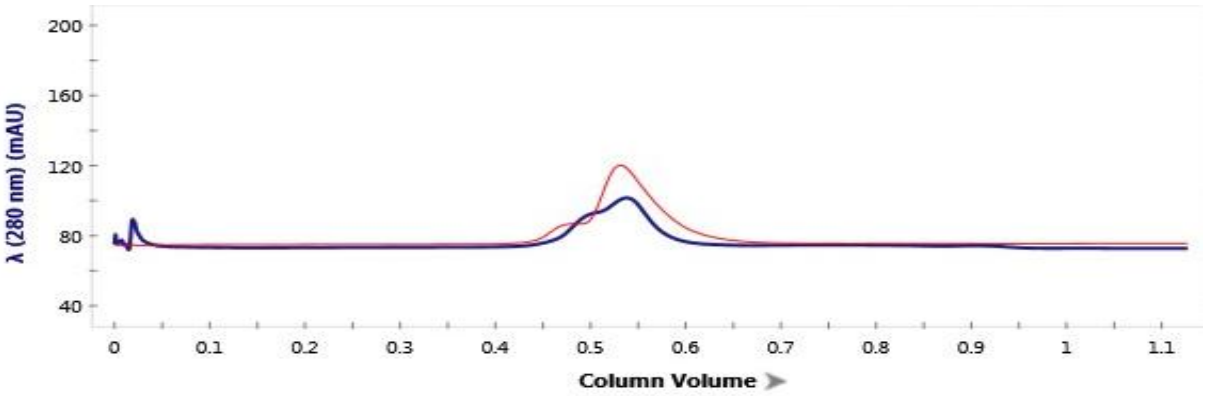


Figure 1.49: 50 µM SEC trace of purified MSH_B-SC2 before (Red) and after (Blue) buffer exchange into 200 mM AmOAc

1.9.8 – FRET Assay for MSH Proteins

Pairs of MSH-TM-mNeongreen and MSH-TM-mScarletI were mixed in 1:1 ratio (10 μM each) with LMNG (10 eq) or MNAC12 (10 eq) to allow FRET couple equilibration. This solution was diluted to 100 nM MSH-TM-mNeongreen and 100 nM MSH-TM-mScarletI with 25 mM Tris-HCl (pH 7.4) and 150 mM NaCl. The emission intensity ratio $I_{590\ nm}/I_{520\ nm}$ (excitation: 480 nm) of this solution was normalized as 100% FRET. Detergent micelles were prepared by dissolving detergent in 25 mM Tris-HCl (pH 7.4) and 150 mM NaCl. Various concentrations of the detergent micelles were titrated into the MSH-TM-FP solution, with emission ratio $I_{590\ nm}/I_{520\ nm}$ measured to monitor FRET between fluorescent protein couple. 0% FRET is normalized to $I_{590\ nm}/I_{520\ nm}$ at the average FE once a plateau has been established at the end of the FRET experiment.

1.9.9 – Native Mass Spectrometry Experiments of MSH Dimers

50 μL of purified MSH constructs (between 10-20 μM) in a buffer solution of 300 mM NaCl and 50 mM Tris-HCl at pH 7.5 were diluted in 950μL of 200 mM AmOAc and mixed gently. The sample was then loaded onto a Superdex 200 10/300 GL column (Cytiva) and eluted with a buffer of 200mM AmOAc. The solution was then concentrated with a 50k Da MWCO Amicon till the concentration of MSH was between 4-10 μM.

Samples were diluted to 10 μM in 150 mM ammonium acetate then infused with a nano-ESI sprayer (TriVersa NanoMate; Advion Bioscience) using a voltage of 1.4-1.6 kV versus the inlet and 0.3-1.0 psi drying gas into a Bruker solariX 12T FTICR-MS (Bruker Daltonics). Data were collected with 5 x 10^5 data points from 200-10,000 m/z. The following

instrument additional parameters were used: Funnel 1 150 V; Skimmer 1 125V; Funnel RF amplitude 300 Vpp; ion accumulation 700 ms; Collision cell frequency 1.4 MHz; Collision RF amplitude 600 Vpp; Transfer Optics time of flight 3 ms; Transfer Optics frequency 2 MHz; Transfer Optics RF Amplitude 450 Vpp; Sweep excitation power 40%; Gas control 40%; collision energy of 10-30 V.

1.9.10 – Assessing RRR-SC2 Oligomeric State with Analytical SEC

MSH_B-SC2 purified in a dimer-trimer equilibrium was mixed with a set ratio of detergent and RRR-SC2 (see experiment for specific detergent type and ratio) and diluted to 1 µM with a buffer solution of 300 mM NaCl and 50 mM Tris-HCl at pH 7.5. The samples were allowed to equilibrate for 1 or 3 days (see experiment for specific equilibration time) at 4°C. Once the equilibration period had concluded, the samples were loaded onto a Superdex 200 10/300 GL column (Cytiva) and eluted with a buffer of 300 mM NaCl and 50 mM Tris-HCl at pH 7.5.

1.9.11 – MSH_B-SC2 to RRR-SC2 FRET Transfer

RRR-SC2 and RRR-$SC2^{4CN}$ (500 µM) were both dissolved individually with a solution of 10 mM LMNG micelles in a buffer of 300 mM NaCl and 50 mM Tris-HCl at pH 7.5. The peptide/detergent solutions were added to a MSH_B-SC2 solution in a dimer-trimer equilibrium of the same buffer composition and diluted till MSH_B-SC2 final concentration was 10uM. The final ratio of MSH_B-SC2 : Peptide : LMNG was 2:1:80. The sample was allowed to equilibrate at 4°C for a minimum of 72 hours. A control solution of RRR-SC2:RRR-$SC2^{4CN}$:LMNG 1:1:80 was made as a control. Once equilibrated, the solution was diluted

with the aforementioned buffer until a final concentration of 300 nM MSH_B-SC2, or 150 nM peptide if no protein was present. The solutions were then excited with light at 270 nm, and the emission profile from 300-800 nm was recorded.

1.9.12 – Preliminary Crystal Screening of MSH_B-SC2/RRR-SC2 Non-Covalent Assembly

RRR-SC2 (500uM) was dissolved in a solution of 10 mM LMNG micelles in a buffer of 300 mM NaCl and 50 mM Tris-HCl at pH 7.5. The peptide/detergent solutions were added to a MSH_B-SC2 solution in a dimer-trimer equilibrium of the same buffer composition and diluted until the MSH_B-SC2 final concentration was 30 mg/mL. The final ratio of MSH_B-SC2 : RRR-SC2 : LMNG was 2:1:80. The sample was allowed to equilibrate at 4°C for a minimum of 72 hours. Once equilibrated, a sitting drop crystal screen was conducted using the HR2-144 crystal screening kit (Hampton Research). 1uL of sample was mixed with 1 µL of precipitant over a well of 150 µL of precipitant. Crystal trays were left at room temperature and monitored for crystal growth.

1.9.13 – MSH DNA Sequences and Translated Sequences

Restriction enzyme cleavage site

NcoI SalI HindIII EcoRI NotI

MSH_A-(BCL-X_L)-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC

GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCG TGT GCG ATT TGT CTT ATT GTT GTT GCC AAA GCG GGC GTG ATG TTA CGC GAT AAT GCG ACG GAA GAA GAG ATT TTA GTT TAT TTG GAG AAG ACC TGT GAT TGG CTT CCT AAA CCT AAT ATG TCG GCA TCG TGC AAA GAA ATT GTG GAT AGT TAT TTG CCA

GTG ATC CTG GAT ATT ATC AAA GGC GAG ATG TCA CGC CCA GGG GAG GTG TGC TCG GCC TTG AAT CTT TGC GAA GGC GGT TCG AAT TCA

TCG CGT AAG GGT CAG GAG CGT TTT AAT CGC TGG TTT CTG ACG GGA ATG ACC GTC GCA GGC GTA GTA CTG CTG GGG TCA CTT TTC TCT CGT AAG

CAC CAT CAT CAC CAC CAT TAA GCG GCC GC

Translated sequence

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA MSYFWYAVRTAVINAASGRQTVDEALACAICLIVVAKAGVMLRDNATEEEILVYLEKTCDWLPKPNM SASCKEIVDSYLPVILDIIKGEMSRPGEVCSALNLCEGGSNSSRKGQERFNRWFLTGMTVAGVVLL GSLFSRKHHHHHH-AA

MSH$_A$-BAX-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC

GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCG TGT GCG ATT TGT CTT ATT GTT GTT GCC AAA GCG GGC GTG ATG TTA CGC GAT AAT GCG ACG GAA GAA GAG ATT TTA GTT TAT TTG GAG AAG ACC TGT GAT TGG CTT CCT AAA CCT AAT ATG TCG GCA TCG TGC AAA GAA ATT GTG GAT AGT TAT TTG CCA

GTG ATC CTG GAT ATT ATC AAA GGC GAG ATG TCA CGC CCA GGG GAG GTG TGC TCG GCC TTG AAT CTT TGC GAA GGC GGT TCG AAT TCA

ACT TGG CAG ACT GTG ACC ATC TTT GTC GCA GGC GTA CTG ACA GCT TCC TTA ACA ATC T GG AAG AAA ATG GGT

CAC CAT CAT CAC CAC CAT TAA GCG GCC GC

Translated Sequence:

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA MSYFWYAVRTAVINAASGRQTVDEALACAICLIVVAKAGVMLRDNATEEEILVYLEKTCDWLPKPNM SASCKEIVDSYLPVILDIIKGEMSRPGEVCSALNLCEGGSNSTWQTVTIFVAGVLTASLTIWKKMGHH HHHH-AA

MSH$_A$-(BCL-X$_L$)-mNeongreen-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC

GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCG TGT GCG ATT TGT CTT ATT GTT GTT GCC AAA GCG GGC GTG ATG TTA CGC GAT AAT GCG ACG GAA GAA GAG ATT TTA GTT TAT TTG GAG AAG ACC TGT GAT TGG CTT CCT AAA CCT AAT ATG TCG GCA TCG TGC AAA GAA ATT GTG GAT AGT TAT TTG CCA

GTG ATC CTG GAT ATT ATC AAA GGC GAG ATG TCA CGC CCA GGG GAG GTG TGC TCG GCC TTG AAT CTT TGC GAA GGC GGT TCG AAT TCA

TCG CGT AAG GGT CAG GAG CGT TTT AAT CGC TGG TTT CTG ACG GGA ATG ACC GTC GCA GGC GTA GTA CTG CTG GGG TCA CTT TTC TCT CGT AAG

GAC GTC CTG CCG GCT ACC CAT GAA CTG CAC ATC TTC GGG AGC ATC AAC GGC GTG GAC TTT GAT ATG GTG GGT CAG GGC ACC GGA AAC CCG AAC GAT GGC TAT GAG GAG CTT AAT CTG AAG AGC ACC AAA GGC GAC CTT CAG TTT AGT CCG TGG ATT CTT GTG CCG CAT ATT GGG TAT GGC TTC CAT CAG TAT TTA CCA TAC CCA GAT GGT ATG TCG CCG TTC CAA GCC GCG ATG GTC GAT GGG TCC GGC TAT CAG GTT CAT CGT ACC ATG CAG TTT GAG GAT GGG GCC TCA TTA ACG GTG AAC TAC CGC TAT ACC TAT GAA GGG AGT CAC ATT AAA GGT GAA GCT CAA GTC AAA GGC ACC GGC TTT CCT GCC GAC GGA CCT GTG ATG ACC AAT TCT TTA ACT GCG GCG GAC TGG GCG CGC AGC AAG AAA ACG TAT CCG AAT GAT AAA ACG ATC ATT TCC ACG TTC AAA TGG AGC TAT ACT ACC GGC AAT GGT AAG CGC TAT CGT TCA ACC GCC CGT ACG ACT TAC ACG TTT GCG AAA CCG ATG GCG GCG AAC TAT TTA AAG AAT CAA CCG ATG TAC GTT TTT CGC AAA ACG GAG TTG AAA CAC TCC AAG ACT GAG CTG AAT TTT AAA GAG TGG CAG AAA GCG TTC ACT GAT GTG ATG GGT ATG GAT GAG TTG TAT AAA

CAC CAT CAT CAC CAC CAT TAA GCG GCC GC

Translated sequence

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD
RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA
LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN
KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN
AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA
MSYFWYAVRTAVINAASGRQTVDEALACAICLIVVAKAGVMLRDNATEEEILVYLEKTCDWLPKPNM
SASCKEIVDSYLPVILDIIKGEMSRPGEVCSALNLCEGGSNSSRKGQERFNRWFLTGMTVAGVVLL
GSLFSRKDVLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGY
GFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVKGTGFP
ADGPVMTNSLTAADWARSKKTYPNDKTIISTFKWSYTTGNGKRYRSTARTTYTFAKPMAANYLKNQP
MYVFRKTELKHSKTELNFKEWQKAFTDVMGMDELYKHHHHHH-AA

MSH$_A$-(BCL-X$_L$)-mScarletI-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT
AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC
GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT
GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC
CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC
TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG
TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG
ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC

CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCG TGT GCG ATT TGT CTT ATT GTT GTT GCC AAA GCG GGC GTG ATG TTA CGC GAT AAT GCG ACG GAA GAA GAG ATT TTA GTT TAT TTG GAG AAG ACC TGT GAT TGG CTT CCT AAA CCT AAT ATG TCG GCA TCG TGC AAA GAA ATT GTG GAT AGT TAT TTG CCA GTG ATC CTG GAT ATT ATC AAA GGC GAG ATG TCA CGC CCA GGG GAG GTG TGC TCG GCC TTG AAT CTT TGC GAA GGC GGT TCG AAT TCA

TCG CGT AAG GGT CAG GAG CGT TTT AAT CGC TGG TTT CTG ACG GGA ATG ACC GTC GCA GGC GTA GTA CTG CTG GGG TCA CTT TTC TCT CGT AAG

GAC GTC ATT AAA GAA TTT ATG CGC TTT AAG GTA CAC ATG GAA GGT TCG ATG AAC GGG CAC GAG TTC GAG ATT GAG GGA GAG GGA GAG GGG CGC CCG TAT GAG GGG ACC CAA ACT GCC AAA TTG AAG GTC ACG AAA GGC GGC CCG TTA CCA TTT TCT TGG GAC ATC CTT TCG CCT CAA TTC ATG TAC GGT AGC CGC GCA TTT ATT AAG CAT CCA GCG GAT ATC CCT GAC TAC TAT AAG CAG TCA TTC CCT GAA GGA TTT AAG TGG GAA CGT GTA ATG AAT TTC GAG GAC GGA GGA GCG GTT ACT GTC ACT CAG GAT ACC AGT CTT GAA GAC GGC ACA TTA ATT TAT AAG GTT AAA TTG CGC GGG ACA AAC TTC CCT CCT GAT GGT CCA GTT ATG CAG AAA AAG ACT ATG GGT TGG GAA GCC TCC ACC GAA CGC CTT TAC CCC GAA GAC GGT GTC TTA AAA GGG GAT ATC AAG ATG GCA TTG CGT CTG AAA GAC GGT GGG CGT TAC TTG GCC GAT TTC AAG ACC ACC TAC AAG GCT AAG AAG CCT GTG CAA ATG CCA GGC GCG TAT AAC GTA GAT CGC AAG TTG GAC ATC ACA TCA CAT AAT GAA GAT TAC ACC GTG GTC GAA CAG TAC GAG CGC TCT GAA GGG CGC CAC TCA ACA GGT GGT ATG GAT GAA CTT TAT AAG CAC CAT CAT CAC CAC CAT TAA GCG GCC GC

Translated sequence

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA MSYFWYAVRTAVINAASGRQTVDEALACAICLIVVAKAGVMLRDNATEEEILVYLEKTCDWLPKPNM

SASCKEIVDSYLPVILDIIKGEMSRPGEVCSALNLCEGGSNSSRKGQERFNRWFLTGMTVAGVVLL

GSLFSRKDVIKEFMRFKVHMEGSMNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWDILSPQ

FMYGSRAFIKHPADIPDYYKQSFPEGFKWERVMNFEDGGAVTVTQDTSLEDGTLIYKVKLRGTNFPP

DGPVMQKKTMGWEASTERLYPEDGVLKGDIKMALRLKDGGRYLADFKTTYKAKKPVQMPGAYNV

DRKLDITSHNEDYTVVEQYERSEGRHSTGGMDELYKHHHHHH-AA

MSH$_A$-BAX-mNeonGreen-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT

AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC

GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT

GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC

CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC

TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG

TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG

ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC

CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT

AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG

AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC

ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC

AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG

GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA

GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCG TGT GCG ATT TGT CTT ATT GTT GTT GCC AAA GCG GGC GTG ATG TTA CGC GAT AAT GCG ACG GAA GAA GAG ATT TTA GTT TAT TTG GAG AAG ACC TGT GAT TGG CTT CCT AAA CCT AAT ATG TCG GCA TCG TGC AAA GAA ATT GTG GAT AGT TAT TTG CCA GTG ATC CTG GAT ATT ATC AAA GGC GAG ATG TCA CGC CCA GGG GAG GTG TGC TCG GCC TTG AAT CTT TGC GAA GGC GGT TCG AAT TCA

ACT TGG CAG ACT GTG ACC ATC TTT GTC GCA GGC GTA CTG ACA GCT TCC TTA ACA ATC T GG AAG AAA ATG GGT

GAC GTC CTG CCG GCT ACC CAT GAA CTG CAC ATC TTC GGG AGC ATC AAC GGC GTG GAC TTT GAT ATG GTG GGT CAG GGC ACC GGA AAC CCG AAC GAT GGC TAT GAG GAG CTT AAT CTG AAG AGC ACC AAA GGC GAC CTT CAG TTT AGT CCG TGG ATT CTT GTG CCG CAT ATT GGG TAT GGC TTC CAT CAG TAT TTA CCA TAC CCA GAT GGT ATG TCG CCG TTC CAA GCC GCG ATG GTC GAT GGG TCC GGC TAT CAG GTT CAT CGT ACC ATG CAG TTT GAG GAT GGG GCC TCA TTA ACG GTG AAC TAC CGC TAT ACC TAT GAA GGG AGT CAC ATT AAA GGT GAA GCT CAA GTC AAA GGC ACC GGC TTT CCT GCC GAC GGA CCT GTG ATG ACC AAT TCT

TTA ACT GCG GCG GAC TGG GCG CGC AGC AAG AAA ACG TAT CCG AAT GAT AAA ACG ATC ATT TCC ACG TTC AAA TGG AGC TAT ACT ACC GGC AAT GGT AAG CGC TAT CGT TCA ACC GCC CGT ACG ACT TAC ACG TTT GCG AAA CCG ATG GCG GCG AAC TAT TTA AAG AAT CAA CCG ATG TAC GTT TTT CGC AAA ACG GAG TTG AAA CAC TCC AAG ACT GAG CTG AAT TTT AAA GAG TGG CAG AAA GCG TTC ACT GAT GTG ATG GGT ATG GAT GAG TTG TAT AAA

CAC CAT CAT CAC CAC CAT TAA GCG GCC GC

Translated Sequence:

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA MSYFWYAVRTAVINAASGRQTVDEALACAICLIVVAKAGVMLRDNATEEEILVYLEKTCDWLPKPNM SASCKEIVDSYLPVILDIIKGEMSRPGEVCSALNLCEGGSNSTWQTVTIFVAGVLTASLTIWKKMGDV LPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYP DGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVKGTGFPADGPVMTNS LTAADWARSKKTYPNDKTIISTFKWSYTTGNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELK HSKTELNFKEWQKAFTDVMGMDELYKHHHHHH-AA

MSH$_A$-BAX-mScarletI-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCG TGT GCG ATT TGT CTT ATT GTT GTT GCC AAA GCG GGC GTG ATG TTA CGC GAT AAT GCG ACG GAA GAA GAG ATT TTA GTT TAT TTG GAG AAG ACC TGT GAT TGG CTT CCT AAA CCT AAT ATG TCG GCA TCG TGC AAA GAA ATT GTG GAT AGT TAT TTG CCA GTG ATC CTG GAT ATT ATC AAA GGC GAG ATG TCA CGC CCA GGG GAG GTG TGC TCG GCC TTG AAT CTT TGC GAA GGC GGT TCG AAT TCA

ACT TGG CAG ACT GTG ACC ATC TTT GTC GCA GGC GTA CTG ACA GCT TCC TTA ACA ATC T GG AAG AAA ATG GGT

GAC GTC ATT AAA GAA TTT ATG CGC TTT AAG GTA CAC ATG GAA GGT TCG ATG AAC GGG CAC GAG TTC GAG ATT GAG GGA GAG GGA GAG GGG CGC CCG TAT GAG GGG ACC CAA ACT GCC AAA TTG AAG GTC ACG AAA GGC GGC CCG TTA CCA TTT TCT TGG GAC ATC CTT TCG CCT CAA TTC ATG TAC GGT AGC CGC GCA TTT ATT AAG CAT CCA GCG GAT ATC CCT GAC TAC TAT AAG CAG TCA TTC CCT GAA GGA TTT AAG TGG GAA CGT GTA ATG AAT TTC GAG GAC GGA GGA GCG GTT ACT GTC ACT CAG GAT ACC AGT CTT GAA GAC GGC ACA TTA ATT TAT AAG GTT AAA TTG CGC GGG ACA AAC TTC CCT CCT GAT GGT CCA GTT ATG CAG AAA AAG ACT ATG GGT TGG GAA GCC TCC ACC GAA CGC CTT TAC CCC GAA GAC GGT GTC TTA AAA GGG GAT ATC AAG ATG GCA TTG CGT CTG AAA GAC GGT GGG CGT TAC TTG GCC GAT TTC AAG ACC ACC TAC AAG GCT AAG AAG CCT GTG CAA ATG CCA GGC GCG TAT AAC GTA GAT CGC AAG TTG GAC ATC ACA TCA CAT AAT GAA GAT TAC ACC GTG GTC GAA CAG TAC GAG CGC TCT GAA GGG CGC CAC TCA ACA GGT GGT ATG GAT GAA CTT TAT AAG

CAC CAC CAC CAT CAC CAC TAA GCG GCC GC

Translated Sequence:

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD
RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA
LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN
KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN
AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA
MSYFWYAVRTAVINAASGRQTVDEALACAICIIMVAAIQTAVRTNSTFVQALVEHVKEECDRLGPGM
ADICKNYISQYSEIAIQMMMHMQPKEICALVGFCDEGGSNSTWQTVTIFVAGVLTASLTIWKKMGD
VIKEFMRFKVHMEGSMNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWDILSPQFMYGSRAF
IKHPADIPDYYKQSFPEGFKWERVMNFEDGGAVTVTQDTSLEDGTLIYKVKLRGTNFPPDGPVMQK
KTMGWEASTERLYPEDGVLKGDIKMALRLKDGGRYLADFKTTYKAKKPVQMPGAYNVDRKLDITSH
NEDYTVVEQYERSEGRHSTGGMDELYKHHHHHH-AA

MSH$_B$-(BCL-X$_L$)-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT
AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC
GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT
GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC
CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC

TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCC TGC GCT ATT TGT ATT ATT ATG GTA GCG GCA ATT CAA ACC GCT GTC CGC ACC AAT AGC ACG TTT GTG CAG GCG CTG GTA GAA CAC GTC AAG GAA GAA TGC GAT CGT TTG GGA CCT GGG ATG GCA GAT ATC TGC AAG AAC TAT ATT TCG CAG TAC AGC GAG ATC GCC ATT CAA ATG ATG ATG CAT ATG CAG CCA AAA GAG ATT TGT GCG CTG GTC GGA TTT TGC GAC GAG GGC GGT TCG AAT TCA

TCG CGT AAG GGT CAG GAG CGT TTT AAT CGC TGG TTT CTG ACG GGA ATG ACC GTC GCA GGC GTA GTA CTG CTG GGG TCA CTT TTC TCT CGT AAG

CAC CAC CAC CAT CAC CAC TAA GCG GCC GC

Translated Sequence:

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA MSYFWYAVRTAVINAASGRQTVDEALACAICIIMVAAIQTAVRTNSTFVQALVEHVKEECDRLGPGM ADICKNYISQYSEIAIQMMMHMQPKEICALVGFCDEGGSNSSRKGQERFNRWFLTGMTVAGVVLL GSLFSRKHHHHHH

MSH$_B$-BAX-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC

TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCC TGC GCT ATT TGT ATT ATT ATG GTA GCG GCA ATT CAA ACC GCT GTC CGC ACC AAT AGC ACG TTT GTG CAG GCG CTG GTA GAA CAC GTC AAG GAA GAA TGC GAT CGT TTG GGA CCT GGG ATG GCA GAT ATC TGC AAG AAC TAT ATT TCG CAG TAC AGC GAG ATC GCC ATT CAA ATG ATG ATG CAT ATG CAG CCA AAA GAG ATT TGT GCG CTG GTC GGA TTT TGC GAC GAG GGC GGT TCG AAT TCA

ACT TGG CAG ACT GTG ACC ATC TTT GTC GCA GGC GTA CTG ACA GCT TCC TTA ACA ATC T

GG AAG AAA ATG GGT

CAC CAC CAC CAT CAC CAC TAA GCG GCC GC

Translated Sequence:

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD

RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA

LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN

KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN

AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA

MSYFWYAVRTAVINAASGRQTVDEALACAICIIMVAAIQTAVRTNSTFVQALVEHVKEECDRLGPGM

ADICKNYISQYSEIAIQMMMHMQPKEICALVGFCDEGGSNSTWQTVTIFVAGVLTASLTIWKKMGH

HHHHH-AA

MSH_B-(BCL-X_L)-mNeongreen-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT

AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC

GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT

GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC

CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCC TGC GCT ATT TGT ATT ATT ATG GTA GCG GCA ATT CAA ACC GCT GTC CGC ACC AAT AGC ACG TTT GTG CAG GCG CTG GTA GAA CAC GTC AAG GAA GAA TGC GAT CGT TTG GGA CCT GGG ATG GCA GAT ATC TGC AAG AAC TAT ATT TCG CAG TAC AGC GAG ATC GCC ATT CAA ATG ATG ATG CAT ATG CAG CCA AAA GAG ATT TGT GCG CTG GTC GGA TTT TGC GAC GAG GGC GGT TCG AAT TCA

TCG CGT AAG GGT CAG GAG CGT TTT AAT CGC TGG TTT CTG ACG GGA ATG ACC GTC GCA GGC GTA GTA CTG CTG GGG TCA CTT TTC TCT CGT AAG

GAC GTC CTG CCG GCT ACC CAT GAA CTG CAC ATC TTC GGG AGC ATC AAC GGC GTG GAC TTT GAT ATG GTG GGT CAG GGC ACC GGA AAC CCG AAC GAT GGC TAT GAG GAG CTT AAT CTG AAG AGC ACC AAA GGC GAC CTT CAG TTT AGT CCG TGG ATT CTT GTG CCG CAT ATT GGG TAT GGC TTC CAT CAG TAT TTA CCA TAC CCA GAT GGT ATG TCG CCG TTC CAA GCC GCG ATG GTC GAT GGG TCC GGC TAT CAG GTT CAT CGT ACC ATG CAG TTT GAG GAT GGG GCC TCA TTA ACG GTG AAC TAC CGC TAT ACC TAT GAA GGG AGT CAC ATT AAA GGT GAA GCT CAA GTC AAA GGC ACC GGC TTT CCT GCC GAC GGA CCT GTG ATG ACC AAT TCT TTA ACT GCG GCG GAC TGG GCG CGC AGC AAG AAA ACG TAT CCG AAT GAT AAA ACG ATC ATT TCC ACG TTC AAA TGG AGC TAT ACT ACC GGC AAT GGT AAG CGC TAT CGT TCA ACC GCC CGT ACG ACT TAC ACG TTT GCG AAA CCG ATG GCG GCG AAC TAT TTA AAG AAT CAA CCG ATG TAC GTT TTT CGC AAA ACG GAG TTG AAA CAC TCC AAG ACT GAG CTG AAT TTT AAA GAG TGG CAG AAA GCG TTC ACT GAT GTG ATG GGT ATG GAT GAG TTG TAT AAA

CAC CAC CAC CAT CAC CAC TAA GCG GCC GC

Translated Sequence:

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN

KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN
AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA
MSYFWYAVRTAVINAASGRQTVDEALACAICIIMVAAIQTAVRTNSTFVQALVEHVKEECDRLGPGM
ADICKNYISQYSEIAIQMMMHMQPKEICALVGFCDEGGSNSSRKGQERFNRWFLTGMTVAGVVLL
GSLFSRKDVLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGY
GFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVKGTGFP
ADGPVMTNSLTAADWARSKKTYPNDKTIISTFKWSYTTGNGKRYRSTARTTYTFAKPMAANYLKNQP
MYVFRKTELKHSKTELNFKEWQKAFTDVMGMDELYKHHHHHH-AA

MSH_B-(BCL-X_L)-mScarletI-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT
AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC
GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT
GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC
CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC
TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG
TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG
ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC
CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT
AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG
AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC

ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCC TGC GCT ATT TGT ATT ATT ATG GTA GCG GCA ATT CAA ACC GCT GTC CGC ACC AAT AGC ACG TTT GTG CAG GCG CTG GTA GAA CAC GTC AAG GAA GAA TGC GAT CGT TTG GGA CCT GGG ATG GCA GAT ATC TGC AAG AAC TAT ATT TCG CAG TAC AGC GAG ATC GCC ATT CAA ATG ATG ATG CAT ATG CAG CCA AAA GAG ATT TGT GCG CTG GTC GGA TTT TGC GAC GAG GGC GGT TCG AAT TCA

TCG CGT AAG GGT CAG GAG CGT TTT AAT CGC TGG TTT CTG ACG GGA ATG ACC GTC GCA GGC GTA GTA CTG CTG GGG TCA CTT TTC TCT CGT AAG

GAC GTC ATT AAA GAA TTT ATG CGC TTT AAG GTA CAC ATG GAA GGT TCG ATG AAC GGG CAC GAG TTC GAG ATT GAG GGA GAG GGA GAG GGG CGC CCG TAT GAG GGG ACC CAA ACT GCC AAA TTG AAG GTC ACG AAA GGC GGC CCG TTA CCA TTT TCT TGG GAC ATC CTT TCG CCT CAA TTC ATG TAC GGT AGC CGC GCA TTT ATT AAG CAT CCA GCG GAT ATC CCT

GAC TAC TAT AAG CAG TCA TTC CCT GAA GGA TTT AAG TGG GAA CGT GTA ATG AAT TTC GAG GAC GGA GGA GCG GTT ACT GTC ACT CAG GAT ACC AGT CTT GAA GAC GGC ACA TTA ATT TAT AAG GTT AAA TTG CGC GGG ACA AAC TTC CCT CCT GAT GGT CCA GTT ATG CAG AAA AAG ACT ATG GGT TGG GAA GCC TCC ACC GAA CGC CTT TAC CCC GAA GAC GGT GTC TTA AAA GGG GAT ATC AAG ATG GCA TTG CGT CTG AAA GAC GGT GGG CGT TAC TTG GCC GAT TTC AAG ACC ACC TAC AAG GCT AAG AAG CCT GTG CAA ATG CCA GGC GCG TAT AAC GTA GAT CGC AAG TTG GAC ATC ACA TCA CAT AAT GAA GAT TAC ACC GTG GTC GAA CAG TAC GAG CGC TCT GAA GGG CGC CAC TCA ACA GGT GGT ATG GAT GAA CTT TAT AAG

CAC CAC CAC CAT CAC CAC TAA GCG GCC GC

Translated Sequence:

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA MSYFWYAVRTAVINAASGRQTVDEALACAICIIMVAAIQTAVRTNSTFVQALVEHVKEECDRLGPGM ADICKNYISQYSEIAIQMMMHMQPKEICALVGFCDEGGSNSSRKGQERFNRWFLTGMTVAGVVLL GSLFSRKDVIKEFMRFKVHMEGSMNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWDILSPQ FMYGSRAFIKHPADIPDYYKQSFPEGFKWERVMNFEDGGAVTVTQDTSLEDGTLIYKVKLRGTNFPP DGPVMQKKTMGWEASTERLYPEDGVLKGDIKMALRLKDGGRYLADFKTTYKAKKPVQMPGAYNV DRKLDITSHNEDYTVVEQYERSEGRHSTGGMDELYKHHHHHH-AA

MSH_8-BAX-mNeongreen-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCC TGC GCT ATT TGT ATT ATT ATG GTA GCG GCA ATT CAA ACC GCT GTC CGC ACC AAT AGC ACG TTT GTG CAG GCG CTG GTA GAA CAC GTC AAG GAA GAA TGC GAT CGT TTG GGA CCT GGG ATG GCA GAT ATC TGC AAG AAC TAT ATT TCG CAG TAC AGC GAG ATC GCC ATT CAA ATG ATG ATG CAT ATG CAG CCA AAA GAG ATT TGT GCG CTG GTC GGA TTT TGC GAC GAG GGC GGT TCG AAT TCA

ACT TGG CAG ACT GTG ACC ATC TTT GTC GCA GGC GTA CTG ACA GCT TCC TTA ACA ATC T GG AAG AAA ATG GGT

GAC GTC CTG CCG GCT ACC CAT GAA CTG CAC ATC TTC GGG AGC ATC AAC GGC GTG GAC TTT GAT ATG GTG GGT CAG GGC ACC GGA AAC CCG AAC GAT GGC TAT GAG GAG CTT AAT CTG AAG AGC ACC AAA GGC GAC CTT CAG TTT AGT CCG TGG ATT CTT GTG CCG CAT ATT GGG TAT GGC TTC CAT CAG TAT TTA CCA TAC CCA GAT GGT ATG TCG CCG TTC CAA GCC GCG ATG GTC GAT GGG TCC GGC TAT CAG GTT CAT CGT ACC ATG CAG TTT GAG GAT GGG GCC TCA TTA ACG GTG AAC TAC CGC TAT ACC TAT GAA GGG AGT CAC ATT AAA GGT GAA GCT CAA GTC AAA GGC ACC GGC TTT CCT GCC GAC GGA CCT GTG ATG ACC AAT TCT TTA ACT GCG GCG GAC TGG GCG CGC AGC AAG AAA ACG TAT CCG AAT GAT AAA ACG ATC ATT TCC ACG TTC AAA TGG AGC TAT ACT ACC GGC AAT GGT AAG CGC TAT CGT TCA ACC GCC CGT ACG ACT TAC ACG TTT GCG AAA CCG ATG GCG GCG AAC TAT TTA AAG AAT CAA CCG ATG TAC GTT TTT CGC AAA ACG GAG TTG AAA CAC TCC AAG ACT GAG CTG AAT TTT AAA GAG TGG CAG AAA GCG TTC ACT GAT GTG ATG GGT ATG GAT GAG TTG TAT AAA

CAC CAC CAC CAT CAC CAC TAA GCG GCC GC

Translated Sequence:

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD
RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA
LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN
KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN
AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA
MSYFWYAVRTAVINAASGRQTVDEALACAICIIMVAAIQTAVRTNSTFVQALVEHVKEECDRLGPGM
ADICKNYISQYSEIAIQMMMHMQPKEICALVGFCDEGGSNSTWQTVTIFVAGVLTASLTIWKKMGD
VLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPY
PDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVKGTGFPADGPVMTN
SLTAADWARSKKTYPNDKTIISTFKWSYTTGNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTEL
KHSKTELNFKEWQKAFTDVMGMDELYKHHHHHH-AA

MSH$_B$-BAX-mScarletI-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT
AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC
GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT
GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC
CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC

TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCC TGC GCT ATT TGT ATT ATT ATG GTA GCG GCA ATT CAA ACC GCT GTC CGC ACC AAT AGC ACG TTT GTG CAG GCG CTG GTA GAA CAC GTC AAG GAA GAA TGC GAT CGT TTG GGA CCT GGG ATG GCA GAT ATC TGC AAG AAC TAT ATT TCG CAG TAC AGC GAG ATC GCC ATT CAA ATG ATG ATG CAT ATG CAG CCA AAA GAG ATT TGT GCG CTG GTC GGA TTT TGC GAC GAG GGC GGT TCG AAT TCA

ACT TGG CAG ACT GTG ACC ATC TTT GTC GCA GGC GTA CTG ACA GCT TCC TTA ACA ATC T

GG AAG AAA ATG GGT

GAC GTC ATT AAA GAA TTT ATG CGC TTT AAG GTA CAC ATG GAA GGT TCG ATG AAC GGG

CAC GAG TTC GAG ATT GAG GGA GAG GGA GAG GGG CGC CCG TAT GAG GGG ACC CAA

ACT GCC AAA TTG AAG GTC ACG AAA GGC GGC CCG TTA CCA TTT TCT TGG GAC ATC CTT

TCG CCT CAA TTC ATG TAC GGT AGC CGC GCA TTT ATT AAG CAT CCA GCG GAT ATC CCT

GAC TAC TAT AAG CAG TCA TTC CCT GAA GGA TTT AAG TGG GAA CGT GTA ATG AAT TTC

GAG GAC GGA GGA GCG GTT ACT GTC ACT CAG GAT ACC AGT CTT GAA GAC GGC ACA TTA

ATT TAT AAG GTT AAA TTG CGC GGG ACA AAC TTC CCT CCT GAT GGT CCA GTT ATG CAG

AAA AAG ACT ATG GGT TGG GAA GCC TCC ACC GAA CGC CTT TAC CCC GAA GAC GGT

GTC TTA AAA GGG GAT ATC AAG ATG GCA TTG CGT CTG AAA GAC GGT GGG CGT TAC TTG

GCC GAT TTC AAG ACC ACC TAC AAG GCT AAG AAG CCT GTG CAA ATG CCA GGC GCG TAT

AAC GTA GAT CGC AAG TTG GAC ATC ACA TCA CAT AAT GAA GAT TAC ACC GTG GTC GAA

CAG TAC GAG CGC TCT GAA GGG CGC CAC TCA ACA GGT GGT ATG GAT GAA CTT TAT AAG

CAC CAC CAC CAT CAC CAC TAA GCG GCC GC

Translated Sequence:

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD

RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA

LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN

KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN

AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA
MSYFWYAVRTAVINAASGRQTVDEALACAICIIMVAAIQTAVRTNSTFVQALVEHVKEECDRLGPGM
ADICKNYISQYSEIAIQMMMHMQPKEICALVGFCDEGGSNSTWQTVTIFVAGVLTASLTIWKKMGD
VIKEFMRFKVHMEGSMNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWDILSPQFMYGSRAF
IKHPADIPDYYKQSFPEGFKWERVMNFEDGGAVTVTQDTSLEDGTLIYKVKLRGTNFPPDGPVMQK
KTMGWEASTERLYPEDGVLKGDIKMALRLKDGGRYLADFKTTYKAKKPVQMPGAYNVDRKLDITSH
NEDYTVVEQYERSEGRHSTGGMDELYKHHHHHH-AA

MSH$_B$-SC2-H6:

CC

ATG GGG AAA ATC GAA GAA GGC AAA CTG GTA ATC TGG ATT AAT GGA GAT AAA GGC TAT
AAT GGT CTC GCG GAA GTC GGA AAG AAA TTC GAG AAA GAT ACC GGC ATT AAA GTC ACC
GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA TTC CCA CAG GTT GCG GCA ACT GGC GAT
GGC CCT GAC ATT ATC TTC TGG GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC
CTG TTG GCT GAA ATC ACC CCG GAC AAA GCG TTT CAG GAT AAG CTG TAT CCG TTT ACC
TGG GAC GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT GAA GCG
TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA AAA ACC TGG GAA GAG
ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA GGT AAG AGC GCG CTG ATG TTT AAC
CTG CAA GAA CCG TAC TTC ACC TGG CCG CTG ATT GCT GCC GAC GGG GGT TAT GCG TTT
AAG TAT GAT AAC GGC AAG TAC GAC ATT AAA GCT GTG GGC GTG GCA AAC GCT GGC GCG
AAA GCG GGT CTG ACC TTT CTG GTC GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA GAC
ACC GAT TAC TCC ATC GCA GAA GCA GCC TTT AAT AAA GGC GAA ACA GCG ATG ACC ATC

AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA TCC AAA CCG TTT GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT CCG AAC AAA GAG CTG GCA AAA GAG TTT CTC GAA AAT TAT CTG CTG ACT GAT GAA GGT CTG GAA GCG GTT AAT AAA GAT AAA CCG CTG GGT GCC GTG GCG CTG AAG TCT TAT GAG GAA GAG TTG GTG AAA GAT CCG CGT ATT GCC GCC ACG ATG GAG AAT GCC CAG AAA GGT GAA ATC ATG CCG AAT ATC CCG GCA ATG AGC TAT TTC TGG TAT GCC GTG CGT ACG GCG GTG ATC AAT GCC GCC TCT GGG CGT CAG ACG GTC GAT

GAA GCT TTA GCC TGC GCT ATT TGT ATT ATT ATG GTA GCG GCA ATT CAA ACC GCT GTC CGC ACC AAT AGC ACG TTT GTG CAG GCG CTG GTA GAA CAC GTC AAG GAA GAA TGC GAT CGT TTG GGA CCT GGG ATG GCA GAT ATC TGC AAG AAC TAT ATT TCG CAG TAC AGC GAG ATC GCC ATT CAA ATG ATG ATG CAT ATG CAG CCA AAA GAG ATT TGT GCG CTG GTC GGA TTT TGC GAC GAG GGC GGT TCG AAT TCA

TGG TTA GGC TTT ATT GCG GGA TTA ATT GCC ATT GTA CTT GTC ACG ATC CTT TTG

CAC CAC CAC CAT CAC CAC TAA GCG GCC GC

Translated Sequence:

MGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYDNGKYDIKAVGVANAGAKAGLTFLVDLIKN

KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGIN

AASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPA

MSYFWYAVRTAVINAASGRQTVDEALACAICIIMVAAIQTAVRTNSTFVQALVEHVKEECDRLGPGM

ADICKNYISQYSEIAIQMMMHMQPKEICALVGFCDEGGSNSWLGFIAGLIAIVLVTILLHHHHHH-

AA

1.10 – References

1. Jelokhani-Niaraki, M., Membrane Proteins: Structure, Function and Motion, *Int. J. Mol. Sci.*, **2023**, *24*, 468
2. Sachs, J.; Engelman, D., Introduction to the Membrane Protein Reviews: the Interplay of Structure, Dynamics, and Environmental in Membrane Protein Function, *Annual Review of Biochemistry*, **2006**, *53*, 707-712
3. Gupta, K.; Donlan, J.; Hopper, J.; Uzdavinys, P.; Landreh, M.; Struwe, W.; Drew, D.; Baldwin, A.; Stansfeld, P.; Robinson, C., The Role of Interfacial Lipids in Stabilizing Membrane Protein Oligomers, *Nature*, **2017**, *541*, 421-424
4. Klingenberg, M., Membrane Protein Oligomeric Structure and Transport Function, *Nature*, **1981**, *290*, 449-454
5. Adams, P.; Engelman, D.; Brunger, A., Improved Prediction for the Structure of the Dimeric Transmembrane Domain of Glycophorin A Obtained through Global Searching, *Proteins*, **1996**, *26*, 257-261
6. Paul M.; Hristova, K., The RTK Interactome: Overview and Perspective on RTK Heterointeractions, *Chemical Reviews*, **2018**, *119*, 5881-5921
7. Call, M.; Schnell, J.; Xu, C.; Lutz, R.; Chou, J.; Wucherpfennig, K,, The Structure of the zz Transmembrane Dimer Reveals Features Essential for Its Assembly with the T Cell Receptor, *Cell*, **2006**, *127*, 355-368
8. Pogozheva, I.; Lomize, A., Evolution and Adaptation of Single-Pass Transmembrane Proteins, *Biochimica et Biophysica Acta – Biomembranes*, **2018**, *1860*, 364-377

9. Liu, S.; Payne, A.; Wang, J.; Zhu, L.; Paknejad, N.; Eng, E.; Liu, W.; Maio, Y.; Hite, R.; Huang, X., Architecture and Activation of Single-Pass Transmembrane Receptor Guanylyl Cyclase, *Nature Structural and Molecular Biology*, **2025**, *32*, 469-478
10. Li, E.; Hristova, K., Receptor Tyrosine Kinase Transmembrane Domains – Function, Dimer Structure, and Dimerization Energetics, *Cell Adhesion and Migration*, **2009**, *4*, 249-254
11. He, L.; Horton, W.; Hristova, K., Physical Basis Behind Achondroplasia, the Most Common Form of Human Dwarfism, *Journal of Biological Chemistry*, **2010**, *285*, 30103-30114
12. Li, E.; You, M.; Hristova, K., FGFR3 Dimer Stabilization Due to a Single Amino Acid Pathogenic Mutation, *Journal of Molecular Biology*, **2006**, *356*, 600-612
13. Ou, S.; Schrock, A.; Bocharov, E.; Klempner, S.; Haddad, C.; Steinecker, G.; Johnson, J.; Gitlitz, B.; Chung, J.; Campregher, P.; Ross, J.; Stephens, P.; Miller, V.; Suh, J.; Ali, S.; Velcheti, V., HER2 Transmembrane Domain (TMD) Mutations (V659/G660) that Stabilize Homo- and Heterodimerization are Rare Oncogenic Drivers in Lung Adenocarcinoma that Respond to Afatinib, *Journal of Thoracic Oncology,* **2017**,*12*, 446–57
14. Errey, J.; Fiez-Vandal, C., Production of Membrane Proteins in Industry: the Example of GPCRs, *Protein Expression and Purification*, **2020**, *169*, 105569
15. Braun, R. J.; Kinkl, N.; Beer, M.; Ueffing, M., Two-Dimensional Electrophoresis of Membrane Proteins, *Anal. Bioanal. Chem.*, **2007**, *389*, 1033–45.
16. Macher, B.; Yen, T., Proteins at Membrane Surfaces – a Review of Approaches, *Molecular BioSystems*, **2007**, *3*, 705-713

17. Carpenter, E.P.; Beis, K.; Cameron, A.D.; Iwata, S., Overcoming the Challenges of Membrane Protein Crystallography, *Current Opinion in Structural Biology*, **2008**, *18*, 581-586
18. Lacapere, J.J.; Pebay-Peyroula, E.; Neumann, J.M.; and Etchebest, C., Determining Membrane Protein Structures: Still a Challenge!, *Trends in Biochemical Sciences*, **2007**, *32*, 259-270
19. Palomar-Alonso, N.; Lee, M.; Kim, M., Exosomes: Membrane-Associated proteins, Challenges and Perspectives, *Biochemistry and Biophysics Reports*, **2024**, *37*, 101599
20. Privé, G.G., Detergents for the stabilization and crystallization of membrane proteins. *Methods*, **2007**, *41*, 388-397
21. Champeil, P.; Orlowski, S.; Babin, S.; Lund, S.; Le Maire, M.; Møller, J.; Lenoir, G.; Montigny, C., A Robust Method to Screen Detergents for Membrane Protein Stabilization, Revisited. *Analytical Biochemistry*, **2016**, *511*, 31-35
22. Breyton, C.; Javed, W.; Vermot, A.; Arnaud, C.A.; Hajjar, C.; Dupuy, J.; Petit-Hartlein; I., Le Roy, A.; Martel, A.; Thépaut, M.; Orelle, C., 2019. Assemblies of Lauryl Maltose Neopentyl Glycol (LMNG) and LMNG-Solubilized Membrane Proteins, *Biochimica et Biophysica Acta (BBA)-Biomembranes*, **2019**, *1861*, 939-957
23. Aduri, N.G.; Ernst, H.A.; Prabhala, B.K.; Bhatt, S.; Boesen, T.; Gajhede, M.; Mirza, O., Human Proton Coupled Folic Acid Transporter is a Monodisperse Oligomer in the Lauryl Maltose Neopentyl Glycol Solubilized State. *Biochemical and Biophysical Research Communications*, **2018**, *495*, 1738-1743

24. Hussain, H.; Du, Y.; Scull, N.J.; Mortensen, J.S.; Tarrasch, J.; Bae, H.E.; Loland, C.J.; Byrne, B.; Kobilka, B.K.; Chae, P.S., Accessible Mannitol-Based Amphiphiles (MNAs) for Membrane Protein Solubilisation and Stabilisation, *Chemistry–A European Journal*, **2016**, *22*, 7068-7073

25. Hussain, H.; Helton, T.; Du, Y.; Mortensen, J.S.; Hariharan, P.; Ehsan, M.; Byrne, B., Loland, C.J.; Kobilka, B.K.; Guan, L.; Chae, P.S., A Comparative Study of Branched and Linear Mannitol-Based Amphiphiles on Membrane Protein Stability. *Analyst*, **2018**, *143*, 5702-5710

26. Døvling Kaspersen, J.; Moestrup Jessen, C.; Stougaard Vad, B.; Skipper Sørensen, E.; Kleiner Andersen, K.; Glasius, M.; Pinto Oliveira, C.L.; Otzen, D.E.; Pedersen, J.S., Low-Resolution Structures of OmpA· DDM Protein–Detergent Complexes. *ChemBioChem*, **2014,** *15*(14), 2113-2124

27. Hutchison, J.M.; Lu, Z.; Li, G.C.; Travis, B.; Mittal, R.; Deatherage, C.L.; Sanders, C.R., Dodecyl-β-melibioside detergent micelles as a medium for membrane proteins. *Biochemistry*, **2017**, *56*, 5481-5484

28. Baradaran, R., Berrisford, J.M., Minhas, G.S. and Sazanov, L.A., 2013. Crystal structure of the entire respiratory complex I. *Nature*, *494*(7438), pp.443-448.

29. Kurgan, K.; Chen, B.; Brown, K.; Cobra, P.; Ye, X.; Ge, Y.; Gellman, S., Stable Picodisc Assemblies from Saposin Proteins and Branched Detergents, *Biochemistry*, **2021**, *60*, 1108-1119

30. Denisov, I.G.; Grinkova, Y.V.; Lazarides, A.A.; Sligar, S.G., Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs With Controlled Size, *JACS*, **2004**, *126*, 3477-3487

31. Nath, A.; Atkins, W. M,; Sligar, S. G., Applications of Phospholipid Bilayer Nanodiscs in the Study of Membranes and Membrane Proteins, *Biochemistry*, **2004**, *46*, 2059-2069

32. Bayburt, T. H.; Grinkova, Y. V.; Sligar, S. G., Assembly of Single Bacterorhodopsin Trimers in Bilayer Nanodiscs *Archives Biochemistry and Biophysics*, **2006**, *450*, 215-222

33. Nath, A. Grinkova, Y.V.; Sligar, S. G.; Atkins, W. M., Ligand Binding to Cytochrome P450 3A4 in Phospholipid Bilayer Nanodiscs: The Effect of Model Membranes." *J Biol. Chem.*, **2007**, *282*, 28309-28320

34. Hagn, F.; Etzkorn, M.; Raschle, T.; Wagner, G. Optimized Phospholipid Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins, *JACS*, **2013**, *135,*1919-1925.

35. Dutta, S. K.; Yao, Y.; Marassi. F. M. Structural Insights into the Yersinia pestis Outer Membrane Protein A in Lipid Bilayers, *J. Phys. Chem. B.,* **2017**, *121*, 7561-7570.

36. Gao, Y.; Cao, E.; Julius, D.; Cheng, Y. TRPV1 Structures in Nanodiscs Reveal Mechanisms of Ligand and Lipid Action, *Nature,* **2016**, *534*, 347-351

37. Gatsogiannis, C.; Merino, F.; Prumbaum, D.; Roderer, D.; Leidreiter, F.; Meusch, D.; Raunser, S. Membrane insertion of a Tc Toxin in Near-Atomic Detail. *Nat. Struc. Biol.* **2016**, *23*, 884-890.

38. Taylor, N. M. I.; Manolaridis, I.; Jackson, S. M.; Kowal, J.; Stahlberg, H.; Locher, K. P. Structure of the Human Multidrug Transporter ABCG2, *Nature*, **2017**, *546*, 504-509.

39. Mi, W.; Li, Y.; Yoon, S. H.; Ernst, R. K., Walz, T.; Liao, M. Structural Basis of MsbA-mediated Liposaccharide Transport, *Nature*, **2017**, *549*, 223-237.

40. Chen, Q.; She, J.; Zeng, W.; Guo, J.; Xu, H.; Bai, X. C.; Jiang, Y. Structure of Mammalian Endolysosomal TRPML1 Channel in Nanodiscs, *Nature*, **2017**, *550*, 415-418.

41. McGoldrick, L. L.; Singh, A. K.; Saotome, K. Yelshanskaya, M. V.; Twomey, E. C.; Grassucci, R. A.; Sobolevsky, A. I. Opening of the Human Epithelial Calcium TRPV6, *Nature*, **2018**, *553*, 233-237.

42. Autzen, H. E.; Myasnikov, A. G.; Cambell, M. G.; Asamow, D.; Julius, D.; Cheng, Y. Structure of the Human TRPM4 Ion Channel in a Lipid Nanodisc, *Science*, **2018**, *359*, 228-232.

43. Srivastava, A. P.; Luo, M.; Zhou, W.; Symersky, J.; Bai, D.; Chambers, M. G.; Feraldo-Gómez, J. D.; Liao, M.; Mueller, D. M. High-Resolution Cryo-EM Analysis of the Yeast ATP Synthase in a Lipid Membrane, *Science*, **2018**, *360*, eaas9699.

44. Zhao, D. Y.; Pöge, M.; Morizumi, T.; Gulati, S.; Van Eps, N.; Zhang, J.; Miszta, P.; Filipek, S.; Mahamid, J.; Plitzko, J. M.; et al. Cryo-EM Structure of the Native Rhodopsin Dimer in Nanodiscs, *J. Biol. Chem.*, **2019**, *294*, 14215-14230.

45. Dang, S.; van Goor, M. K.; Asarnow, D.; Wang, Y.; Julius, D.; Cheng, Y.; van der Wijst, . Structural Insight into TRPV5 Channel Function and Modulation., *Proc Natl Acad Sci USA* **2019**, *116 (18)*, 8869-8878.

46. O'Brien, J.S.; Kishimoto, Y., Saposin Proteins: Structure, Function, and Role in Human Lysosomal Storage Disorders. *The FASEB journal*, **1991**, *5*, 301-308

47. Popovic, K.; Holyoake, J.; Pomès, R.; Privé, G.G., Structure of Saposin A Lipoprotein Discs. *Proceedings of the National Academy of Sciences*, **2012**, *109*, 2908-2912

48. Tulumello, D.V.; Deber, C.M., SDS Micelles as a Membrane-Mimetic Environment for Transmembrane Segments. *Biochemistry*, **2009**, *48*, 12096-12103

49. Chien, C. H.; Helfinger, L. R.; Bostock, M. J.; Solt, A.; Tan, Y. L.; Nietlispach, D. An Adaptable Phospholipid Membrane Mimetic System for Solution NMR Studies of Membrane Proteins, *J. Am. Chem. Soc.*, **2017**, *139*, 14829-14832.

50. Frauenfeld, J.; Löving, R.; Armache, J.; Sonnen, A. F.; Guettou, F.; Moberg, P.; Zhu, L.; Jegerschöld, C.; Flayhan, A.; Briggs, J. A. G.; Garoff, H.; Löw, C.; Cheng, Y.; Norlund, P. A Saposin-Lipoprotein Nanoparticle System for Membrane Proteins, *Nat. Methods* **2016**, *13*, 345-351.

51. Kintzer, A. F.; Green, E. M.; Dominik, P. K.; Bridges, M.; Armache, J. P.; Deneka, D.; Kim, S. S.; Hubbell, W.; Kossiakoff, A. A.; Cheng, Y.; Stroud, R. M. Structural Basis for Activation of Voltage Sensor Domains in an Ion Channel TPC1, *PNAS*, **2018**, *115*, E9095-E9104.

52. Nguyen, N. X.; Armache, J. P.; Lee, C.; Yang, Y.; Zeng, W.; Mootha, V. K.; Cheng, Y.; Bai, X.; Jiang, Y. Cryo-EM Structure of a Fungal Mitochondrial Calcium Uniporter, *Nature* **2018**, *559*, 570-574.

53. Du, D.; Neuberger, A.; Orr, M. W.; Newman, C. E.; Hsu, P.; Samsudin, F.; Szewczak-Harris, A.; Ramos, L. M,; Debela, M.; Khalid, S.; Storz, G.; Luisi, B. F. Interactions of a Bacterial RND Transporter with a Transmembrane Small Protein in a Lipid Environment, *Structure*, **2020**, *28*, 625-634.e6.

54. Vaasa, A.; Lust, M.; Terrin, A.; Uri, A.; Zaccolo, M., Small-Molecule FRET Probes for Protein Kinase Activity Monitoring in Living Cells. *Biochemical and Biophysical Research Communications*, **2010**, *397*, 750-755

55. McCullock, T.W.; MacLean, D.M; Kammermeier, P.J., Comparing the Performance of mScarlet-I, mRuby3, and mCherry as FRET Acceptors for mNeonGreen. *PloS one*, **2020**, *15*, p.e0219886.

56. Zhang, K.; Ahmed, I.A.; Kratochvil, H.T.; DeGrado, W.F.; Gai, F.; Jo, H., Synthesis and Application of the Blue Fluorescent Amino Acid L-4-Cyanotryptophan to Assess Peptide–Membrane Interactions. *Chemical Communications*, **2019**, *55*, 5095-5098

57. Sun, F., Zong, W., Liu, R., Chai, J. and Liu, Y., 2010. Micro-Environmental Influences on the Fluorescence of Tryptophan, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **2010**, *76*, 142-145

58. Feitelson, J., Environmental Effects on the Fluorescence of Tryptophan and Other Indole Derivatives *Israel Journal of Chemistry*, **1970**, *8*, 241-252

59. Fox, J.D.; Waugh, D.S., Maltose-binding protein as a solubility enhancer. *E. coli Gene Expression Protocols*, **2003**, 99-117.

60. Sun, P.; Tropea, J.E.; Waugh, D.S., Enhancing the Solubility of Recombinant Proteins in Escherichia Coli by using Hexahistidine-Tagged Maltose-Binding Protein as a Fusion Partner. *Heterologous gene expression in E. coli: methods and protocols*, **2010**, 259-274

61. Riggs, P., Expression and Purification of Recombinant Proteins by Fusion to Maltose-Binding Protein. *Molecular biotechnology*, **2000**, *15*, 51-63

62. Lebendiker, M; Danieli, T., Purification of proteins fused to maltose-binding protein. In *Protein Chromatography: Methods and Protocols*, **2010**, 281-293

63. Pryor, K.D.; Leiting, B., High-level expression of soluble protein in Escherichia coli Using a his6-tag and maltose-binding-protein double-affinity fusion system. *Protein expression and purification*,**2010**, *10*, 309-319

64. Bokhove, M.; Al Hosseini, H.S.; Saito, T.; Dioguardi, E.; Gegenschatz-Schmid, K.; Nishimura, K.; Raj, I.; de Sanctis, D.; Han, L.; Jovine, L., Easy mammalian expression and crystallography of maltose-binding protein-fused human proteins. *Journal of structural biology*, **2016** *194*, 1-7

65. Sachdev, D. and Chirgwin, J.M., Fusions to maltose-binding protein: control of folding and solubility in protein purification *Methods in enzymology*, **2000**, *326*, 312-321

66. Clifton, M.C.; Dranow, D.M.; Leed, A.; Fulroth, B.; Fairman, J.W.; Abendroth, J.; Atkins, K.A.; Wallace, E.; Fan, D.; Xu, G.; Ni, Z.J., A maltose-binding protein fusion construct yields a robust crystallography platform for MCL1. *PLoS One*, **2015**, *10*, p.e0125010.

67. Center, R.J.; Kobe, B.; Wilson, K.A.; Teh, T.; Kemp, B.E.; Poumbourios, P.; Howlett, G.J., Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein. *Protein science*, **1998**, *7*, 1612-1619

68. Waugh, D.S., Crystal structures of MBP fusion proteins. *Protein Science*, **2016**, *25*, 559-571

69. Sanders, C.R.; Mittendorf, K.F., Tolerance to changes in membrane lipid composition as a selected trait of membrane proteins. *Biochemistry*, **2011**, *50*, 7858-7867

70. Kale, J.; Osterlund E.; Andrews, D. BCL-2 family of proteins: changing partners in the dance towards death, *Cell Death and Differentiation*, **2018**, *25*, 65-80

71. Coultas, L.; Strasser, A., The role of the Bcl-2 protein family in cancer, *Seminars in cancer biology*, **2003**, *13*,115-123

72. Adams, J.M.; Cory, S., The Bcl-2 protein family: arbiters of cell survival. *Science*, **1998**, *281*, 1322-1326

73. Gascoyne, R.D.; Adomat, S.A.; Krajewski, S.; Krajewska, M.; Horsman, D.E.; Tolcher, A.W.; O'Reilly, S.E.; Hoskins, P.; Coldman, A.J.; Reed, J.C.; Connors, J.M., Prognostic significance of Bcl-2 protein expression and Bcl-2 gene rearrangement in diffuse aggressive non-Hodgkin's lymphoma. *Blood, The Journal of the American Society of Hematology*, **1997**, *90*, 244-251

74. Buggins, A.G.; Pepper, C.J., The role of Bcl-2 family proteins in chronic lymphocytic leukaemia. *Leukemia research*, **2010** *34*, 837-842

75. Karnak, D.; Xu, L., Chemosensitization of prostate cancer by modulating Bcl-2 family proteins. *Current drug targets*, **2010**, *11*, 699-707

76. Ikegaki, N.; Katsumata, M.; Minna, J.; Tsujimoto, Y., Expression of bcl-2 in small cell lung carcinoma cells. *Cancer research*, **1994**, *54*, 6-8

77. Andreu-Fernández, V.; Sancho, M.; Genovés, A.; Lucendo, E.; Todt, F.; Lauterwasser, J.; Funk, K.; Jahreis, G.; Pérez-Payá, E.; Mingarro, I.; Edlich, F., Bax transmembrane domain interacts with prosurvival Bcl-2 proteins in biological membranes. *Proceedings of the National Academy of Sciences*, **2017**, *114*, 310-315

78. Ahn, V.E.; Faull, K.F.; Whitelegge, J.P.; Fluharty, A.L.; Privé, G.G., Crystal structure of saposin B reveals a dimeric shell for lipid binding. *Proceedings of the National Academy of Sciences*, **2003**, *100*, 38-43

79. Rossmann, M.; Schultz-Heienbrok, R.; Behlke, J.; Remmel, N.; Alings, C.; Sandhoff, K., Saenger, W.; Maier, T., Crystal structures of human saposins C and D: implications for lipid recognition and membrane interactions. *Structure*, **2008**, *16*, 809-817

80. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M., ColabFold: making protein folding accessible to all. *Nature methods*, **2022**, *19*, 679-682

81. Fu, Q.; Chou, J., A Trimeric Hydrophobic Zipper Mediates the Intramembrane Assembly of the SARA-CoV-2 Spike, *J. Am. Chem. Soc*, **2021**, *143*, 8543-8546

82. Zhang, J.; Xiao, T.; Cai, Y.; Chen, B., Structure of SARS-CoV-2 spike protein. *Current opinion in virology*, **2021**, *50*, 173-182

83. Hong, H.; Park, S.; Jimenez, R.; Rinehart, D.; Tamm, L.; Role of aromatic side chains in the folding and thermodynamic stability of integral membrane proteins, *Journal of the American Chemical Society*, **2007**, *129*, 8320-8327

84. Boville, C.; Romney, D.; Almhjell, P.; Sieben, M.; Arnold, F., Improved Synthesis of 4-Cyanotryptophan and Other Tryptophan Analogues in Aqueous Solvent Using Variants of TrpB from *Thermotoga maritima*, *J. Org. Chem.*, **2018**, *83*, 7447-7452

85. Anthis, N.; Clore, G., Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm, *Protein Science*, **2013**, *22*, 851-8.

Chapter II

Promotion of Novel Heterochiral Quaternary Contacts by Substituting Proposed Hendecad Positions of the GCN4 Dimerization Domain

2.1 – Abstract

Racemic crystallography has risen as a powerful tool to facilitate the process of acquiring structural data for many protein/peptide monomers and complexes. Since biology utilizes a single stereochemistry for its protein building blocks (L-amino acids), this can hinder crystal growth by limiting proteins to the 65 chiral space groups, out of the total 230. Racemic crystallography provides access to these previously forbidden space groups by mixing peptide racemates (L- and D-chirality) during crystallization. For the structural data obtained from racemic crystals to depict the biologically relevant quaternary structure, the association should be between molecules of the same chirality. Crystallization of a racemic peptide corresponding to the transmembrane domain of the Influenza A M2 protein revealed novel quaternary contacts between enantiomeric peptides to form a heterochiral α-helical dimer, rather than the native homochiral coiled-coil tetramer. The heterochiral dimer displays the less commonly observed hendecad sequence repeat pattern rather than the more commonly observed heptad sequence repeat. The heterochiral dimer displays no supercoiling and adopts a third position interfacial position within the oligomeric interface.

In this study I examined a series of hydrophobic residue substitutions within the dimerization domain of the *Saccharomyces Cerevisiae* transcriptional activator protein GCN4. The substitutions were aimed at promoting heterochiral associations within a water-soluble Coiled-coil forming system. Crystallization and X-ray diffraction experiments afforded seven unique data sets diffracting between 1.18-1.68Å for two potential hendecad design schemes. One structure has been solved at 1.42Å resolution.

2.2 – Introduction

2.2.1 – Coiled Coils

Coiled coils (CCs) are amongst one of the most common protein quaternary structures in biology. This motif is found in about 10% of all protein sequences and is a driving force for many protein assemblies[1]. This motif plays critical roles in the tertiary and quaternary structure within a wide range of protein structures and assemblies. For example, the motif is utilized as the locus of oligomerization in the following systems: the homodimerization of GCN4[2-,4] and the homotetrameric assembly of the Influenza A M2 protein (M2-TM)[5-7].

The relative simplicity of this structural motif has led to multiple investigations to design CC forming sequences[8-12]. Much work of this has gone towards elucidation of the fundamental contacts that drive oligomerization between α-helices and how to apply this knowledge to generate new *de novo* CC-forming peptide sequences. The rational design of CC peptides not observed in nature has led to α-helical barrels of tunable oligomeric size[10], switchable states of α-helical barrels[11], and membrane embeddable CCs systems[12].

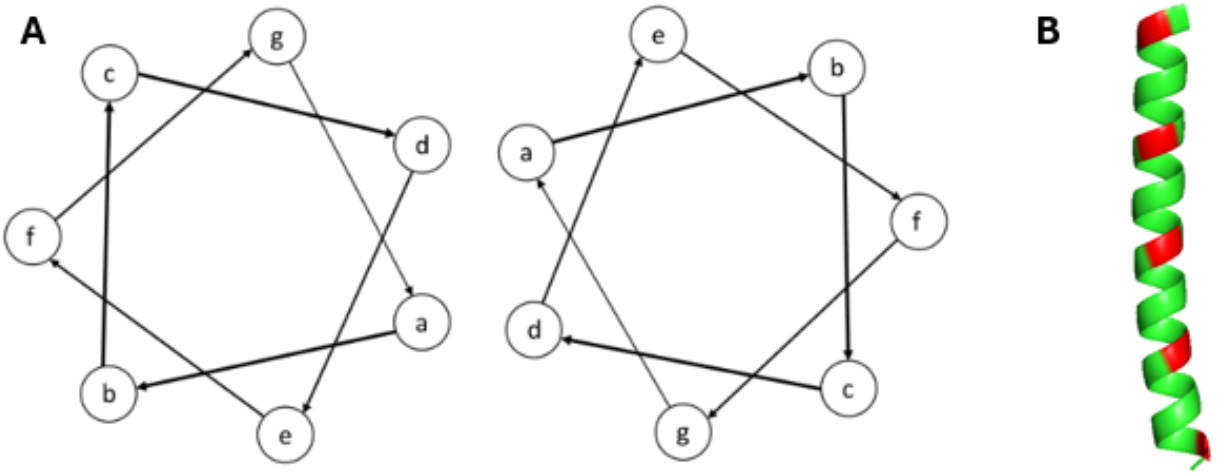


Figure 2.1: (**A**) Helical wheel diagram of an antiparallel homochiral heptad CC dimer assembly. ***a*** and ***d*** residues occupy the positions forming the oligomeric interface. ***e*** and ***g*** residues occupy flanking positions next to the oligomeric interface. Positions on the opposite helical face of the oligomeric interface are comprised of ***c***, ***f***, and ***b*** residues. (**B**) Single copy of a monomer involved in the GCN4 homochiral dimer (PDB: 2ZTA). The ***a*** residues are shown in red. ***a*** residues are displayed on one face of the α-helix with a drift arising from supercoiling.

The heptad repeat positions designated ***abcdefg*** are commonly used to denote the relative positions of residues within an α-helix. The typical α-helix within an CC has been observed to exhibit a periodicity of 3.51 residues/helical turn[13]. Residues of the same designation will be aligned on top of each other along with helical axis and displayed along one face of the helix (**Figure 2.1B**). Observed CC-forming α-helices vary in periodicity from an ideal α-helix (3.65 residues/helical turn), which results in a drift in display of the side chains at a given heptad position along the helical face.

When considering water-soluble CCs, the ***a*** and ***d*** positions are typically occupied by hydrophobic side chains. The ***a***/***d*** "stripes" of these α-helices pack against one another upon association. At the helical interface, the ***a*** and ***d*** side chains interdigitate in a "Knobs-in-Holes" fashion. In some cases, polar residues are observed at these positions. These residues can form interhelical hydrogen bonds to promote partner specificity[2]. Positions ***e*** and ***g*** flank the hydrophobic residues and commonly display charged/polar residues, such as lysine, glutamic acid, or glutamine. These residues stabilize helix associations through salt bridging or hydrogen bonding interactions. Positions ***b***, ***c***, and ***f*** are located on the opposite face of the α-helix relative to the hydrophobic interface and are primarily composed of polar/charged residues that interact with the solvent. The relative placement of these hydrophobic and hydrophilic residues in peptide sequences that form CCs gives rise to their characteristic amphipathic nature, which leads to a driving force for oligomerization.

2.2.2 – Stereochemical Preference within Biological Systems and Racemic Crystallography

Most biomolecules are comprised of building blocks that display a single stereochemistry. Over the previous decades, noncanonical building blocks for biomolecules molecules has arisen as a promising method to develop certain therapeutics with higher *in vivo* lifetimes and overall efficacy[15-17]. Peptidomimetic molecules that incorporate, or are fully composed of, noncanonical building blocks, such as D-amino acids[18-20] and β-amino acids[21,22] have been reported to bind with native biomolecules and exhibit resistance to proteolysis. Additionally, in the case of biomimetic molecules composed of D-amino acids, there have been multiple studies utilizing mirror-image phage display to discover L-peptides that bind with high affinity to D-proteins[23-28] (heterochiral association). The capability of a D-peptide to be sequence optimized for target specificity, coupled with high resistance towards proteolytic degradation makes this an attractive approach to develop peptidomimetic therapeutics. However, there is a lack of structural analysis of these heterochiral assemblies to understand the fundamental interactions that promote their association with native biomolecules. This is in stark contrast to their well-reported homochiral associating counterparts. To understand the principles that drive these non-canonical quaternary contacts, we must be able to obtain high-resolution structures of these assemblies and analyze them for general features that can elucidate fundamental interactions that promote their formation.

Crystallography is one of the most common approaches to generate high-resolution structures of proteins/peptides. Protein crystals, which are often composed of roughly 50%

water by volume, are formed from different driving forces than their small molecule counterparts. Small molecule crystals are typically formed by maximizing close packing of the molecules of interest within the crystal and excluding as many solvent molecules as possible[29]. Due to the high-water content of protein crystals necessary to generate a crystal that maintains a native-like conformation for the protein of interest, maximizing close packing contacts of the proteins within the crystal is not enough to facilitate protein crystal growth. Instead, protein crystals formation is typically thought to be driven by increasing the number of rigid-body degrees of freedom for a given protein within a crystal's unit cell. This model was initially proposed by Wukovitz et al.[30] and summarized by **Equation 2.1**:

(Equation 2.1) $$D = S + L - C$$

In the above equation D = number of rigid-body degrees of freedom, S = number of meaningful degrees of freedom within the unit cell, L = number of independent parameters to define the unit cell, and C = number of unique crystal contacts. The larger the D value (maximum of 8) the greater the propensity for a protein to crystallize into a given space group. The D value may also be interpreted as the dimensionality of the rigid-body space that gives rise to a connected crystal, where again the greater D value is preferred.

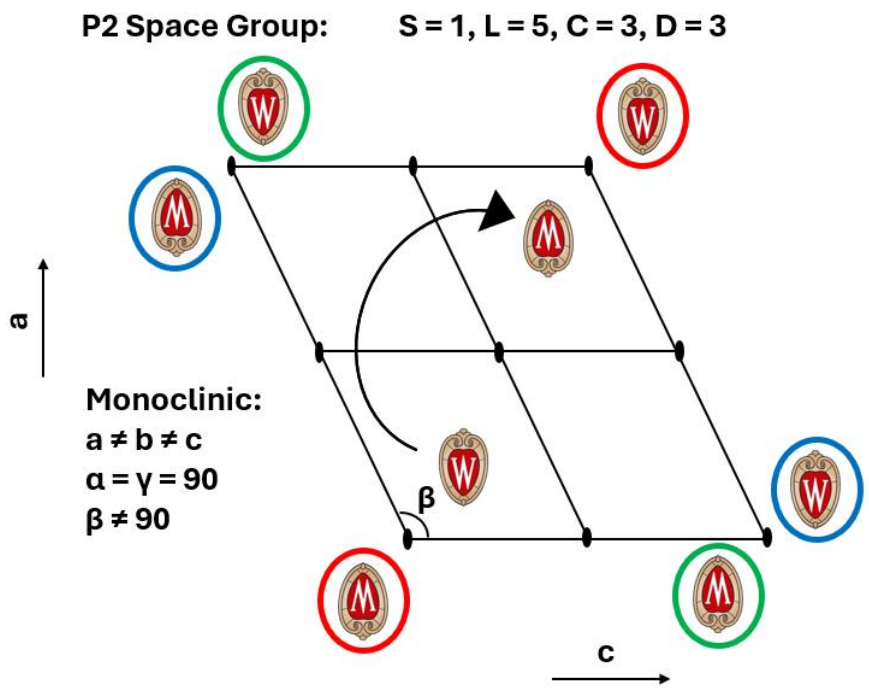


Figure 2.2: D value calculation of the P2 space group. S, L, C, and D values are displayed as integer values. 180° rotation about the unit cell's central axis (S value) shown as curved arrow. Crystal lattice and corresponding unit cell parameters indicated (L value). Unique crystal contacts (C value) circled in red, blue, and green. Circles of the same color indicate redundant crystal contacts.

As an example for applying this formula, I will assign a D value for the P2 space group (**Figure 2.2**). In the P2 space group there are two unique copies of the molecule inside the unit cell that are related to each other through a 180° rotation about the central axis of the unit cell. This affords the P2 space group an S value of 1. The L value is calculated by the unique number of parameters required to define the unit cell's crystal lattice. The P2 space group resides in a monoclinic space group ($a \neq b \neq c$, $\alpha = \gamma = 90$, $\beta \neq 90$) which results in an L value of 5. The C value is calculated by determining the number of unique contacts from adjacent unit cells. This affords a C value of 3 for the P2 space group. When S, L, and C values are inputted in **Equation 2.1** the D value is calculated to be 3.

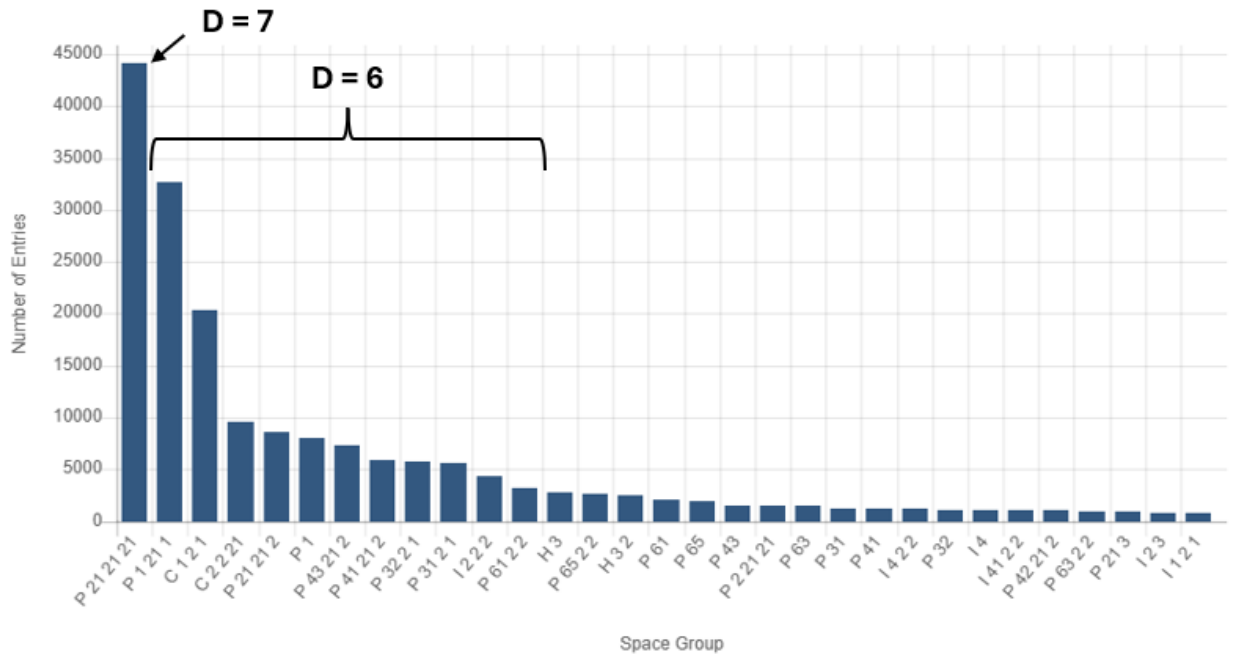


Figure 2.3: Number of protein structure entries reported in the PDB as of May 2025, sorted by reported space group. Space groups with a D value of 6 or higher are indicated. (Screenshot from rscb.org)

Figure 2.3 displays the prevalence of different space groups reported within the protein structure database. The most common space groups observed have a D value of 6 or higher. This trend supports the theory that proteins have a higher probability of crystallizing in a space group with a larger D value. However, not all space groups are accessible for a single enantiomer of a chiral molecule. Traditional (all-L) protein crystallization is fundamentally limited to the 65 chiral space groups, out of a total 230[30,31].

There are multiple achiral space groups that have D values of 7 or higher. Racemic mixtures can access these previously forbidden achiral space groups. Crystallization of racemates grants access to the high D value space groups: $P\bar{1}$ (D = 8), $P2_1/c$ (D=7), and C2/c (D=7). Recent literature examples of the application of racemic crystallography; the heterochiral assembly of the M2-TM peptide dimer[32] (**Fig 2.5 A,C**), the quasi-racemic

crystallization of Ubiquitin[33], and racemic plectasin[34]. These crystal structures were solved in the space groups $P\bar{1}$, P1 (pseudo-$P\bar{1}$), and $P\bar{1}$, respectively.

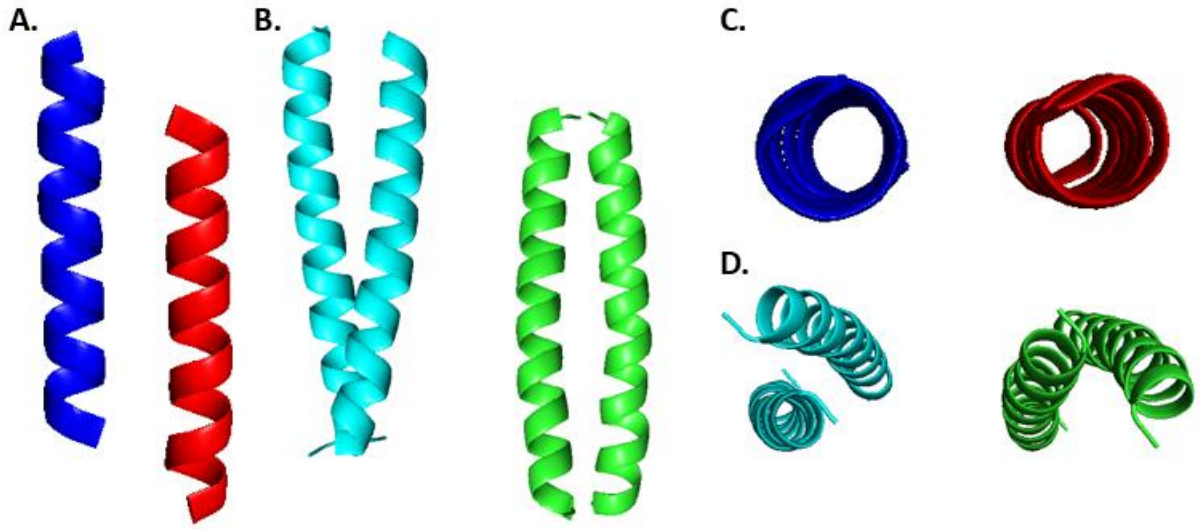


Figure 2.5: Crystal structures of racemic M2-TM (L-peptide in blue, D-peptide in red) and GCN4 (L-peptide in green, D-peptide in cyan). (**A**) Sideview of heterochiral antiparallel M2-TM Dimer. (**B**) Sideview of homochiral parallel GCN4 Dimer. View down the interhelical axis of (**C**) M2-TM (4RWC) and (**D**) GCN4 (2ZTA).

Racemic crystallography offers the possibility of observing quaternary contacts where peptides of opposite stereochemistry assemble (heterochiral)[32,35], instead of peptides of the same stereochemistry[36]. Observing a heterochiral assembly in a crystal structure would represent a quaternary structure that would not represent the biologically relevant assembly. The formation of a heterochiral assembly would then suggest that the non-native protein-protein interaction would be more favorable than the anticipated homochiral counterpart, at least under a given crystal growth condition.

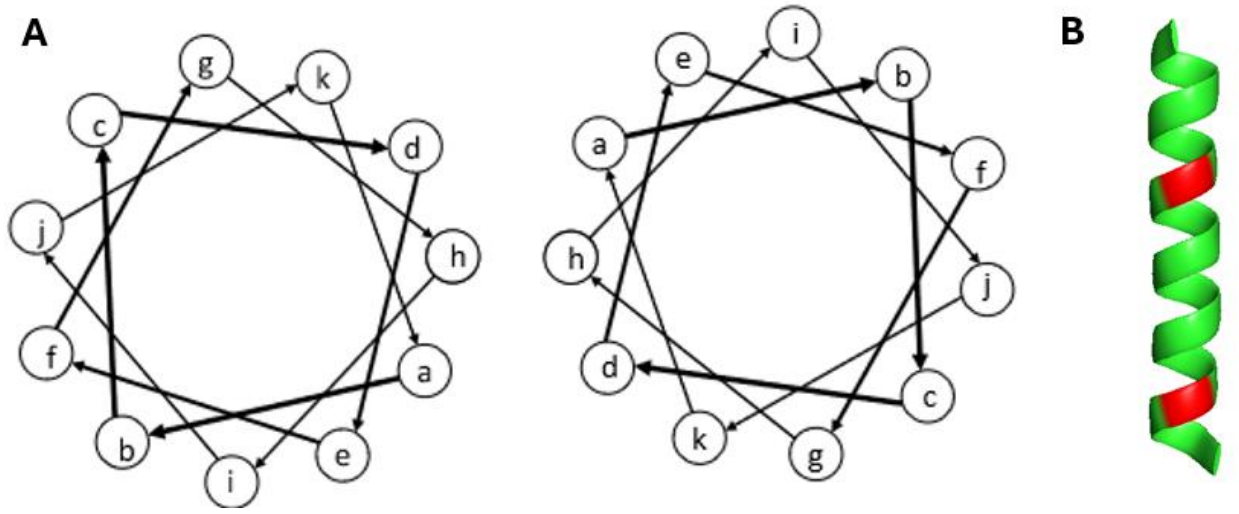


Figure 2.6: (**A**) Helical wheel diagram of an antiparallel homochiral hendecad CC dimer assembly. ***h*** residues occupy the central hydrophobic position and face directly towards the dimerizing partner helix. ***a*** and ***d*** residues occupy the terminal hydrophobic positions within the oligomeric interface. ***e*** and ***k*** residues occupy positions that flank the oligomeric interface. Positions on the opposite helical face of the oligomeric interface are comprised of ***b***, ***f***, ***j***, ***c***, and ***g*** residues. (**B**) Single copy of a monomer involved in the M2-TM heterochiral dimer (PDB: 6MPL). The ***a*** residues are shown in red. ***a*** residues are displayed on one face of the α-helix with near perfect alignment from the loss of supercoiling.

Previous work from our group by Dr. Dale Kreitler determined that the racemic M2-TM crystallizes as an antiparallel heterochiral dimer[32], as opposed to the native homochiral tetrameric structure[5]. The individual α-helices within the heterochiral assembly display a hendecad repeat pattern. As illustrated in **Section 2.2.2**, in homochiral CC associations the participating α-helices typically adopt a heptad repeat; ***abcdefg*** (**Figure 2.1A**). The heptad pattern places hydrophobic residues at the ***a*** and ***d*** positions forming a diagonal stripe of hydrophobic positions along one side of the helix with a periodicity of 3.51 residues/helical turn (**Figure 2.1B**)[13]. The ideal α-helix would require a periodicity of 3.65 residues/helical turn, but with heptad-adopting α-helices, the lower periodicity results in left-handed supercoiling of the CC dimer. The racemic structure of the M2-TM forms a heterochiral dimer displaying a hendecad repeat, ***abcdefghijk*** (**Figure 2.5B,D**), where ***a***, ***d***, and ***h*** make up the hydrophobic residues in a *3,4,4* spacing (*i, i+3, i+7, i+11*). This results in a periodicity of 3.67

residues/turn and a near perfect alignment of a given hendecad position along the helix (**Figure 2.6B**)[32]. For a homochiral hendecad a slight right-handed supercoiling would be exhibited about the dimerizing peptide interface. However, supercoiling in the heterochiral hendecad case is impossible due to the opposite-handedness of the interlocked racemic peptides.

The new hydrophobic position, ***h***, of the hendecad repeat pattern projects directly into the hydrophobic interface (**Figure 2.6A**). Due to the close proximity between ***h*** positions of dimerizing peptides larger hydrophobic residues such as leucine or isoleucine are not tolerated at these positions[31]. In the M2-TM I39A (PDB: 6MPL) structure, valine and alanine, occur as the corresponding ***h*** positions. Then the native sequence of M2-TM (I39 in an ***h*** position) was attempted to be crystallized with its racemate, however, no crystal formation was detected[32]. The packing of Valine28 against Alanine39 is shown via a space filling model to exemplify this point (**Figure 2.7**). The isopropyl sidechain of Valine28 reaches almost to the backbone of its partner helix, while also making van der Waals contacts with Alanine39. If larger hydrophobic residues were in the 29th position, steric repulsion would occur.

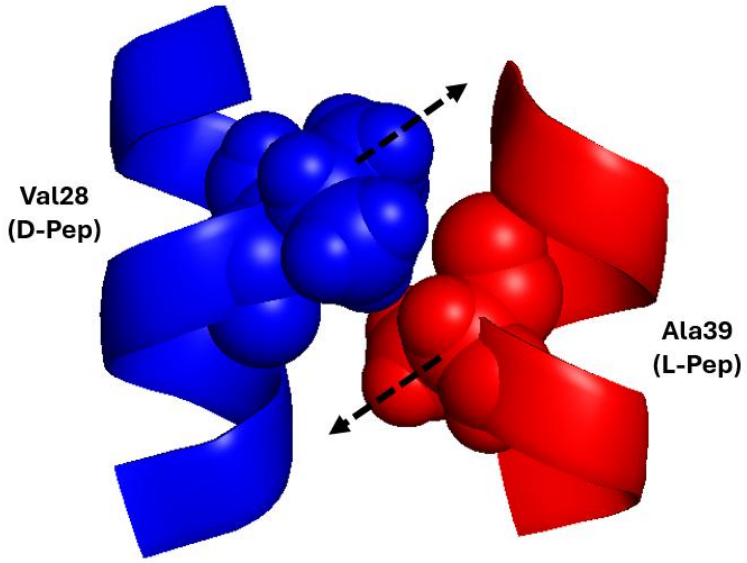


Figure 2.7: Packing of ***h*** position residues in a section of the racemic I39A M2-TM peptide (antiparallel heterochiral dimer)[32]. A39 and V27 are shown in space-filling fashion. Dotted arrows are used to signify the C_α-C_β vector.

2.2.3 – Rationally Designing a Water-Soluble Peptide to Form Heterochiral Dimers

The reported racemic crystal structure of M2-TM introduced a novel mode of heterochiral helical association[32,35]. The high percentage of sequence composition being hydrophobic residues most likely played a substantial role in the heterochiral association being observed. Hendecad repeat patterns require hydrophobic residues at the ***a***, ***d***, and ***h*** positions. M2-TM displays predominately hydrophobic residues within its sequence which allows for more opportunities to satisfy the requirement of hydrophobic residues at the ***a***, ***d***, and ***h*** positions of the hendecad repeat pattern.

I wondered whether observing heterochiral quaternary structures was a phenomena reserved for peptide sequences containing predominately hydrophobic residues or if the requirement of the ***a***, ***d***, and ***h*** positions containing hydrophobic residues would be enough to promote heterochiral assemblies. I hypothesized that if a water-soluble peptide (rich in polar residues) that naturally forms a homochiral CC was modified with the minimal

hydrophobic residue substitutions required to satisfy a hendecad interface, then a heterochiral association would be observed for the racemic form. As mentioned in **Section 2.2.2**, there have been multiple reports via mirror-phage display of D-peptides (derived from an L-peptide sequence) that bind L-proteins[23-28]. Uncovering motifs that induce heterochiral association may become a powerful tool to enabling next-generation peptidomimetic therapeutics.

A peptide corresponding to the dimerization domain of GCN4[3,4], was chosen as a model system (referred to as GCN4 henceforth). As opposed to M2-TM, GCN4, and variants of the GCN4 sequence, are water-soluble peptides that can adopt a dimeric[2], trimeric[37], or tetrameric[38] oligomerization state depending on the composition of the ***a*** and ***d*** residues within its heptad repeat pattern. The native oligomerization state of GCN4, and its parent protein, are both homodimeric. A previous member of our group, Dr. David Mortenson, attempted the crystallization of racemic GCN4 and obtained a preliminary crystal structure (unpublished). When GCN4 was crystallized with both racemates present, the observed GCN4 structure remains assembled as a homochiral dimer (**Figure 2.3B,D**). Since the GCN4 sequence is rich in polar and charged residues, I hypothesized that the racemates were unable to associate with one another as one or more polar or charged residues would be located at an ***a***, ***d***, or ***h*** position when adopting an α-helix conformation that displays a hendecad repeat pattern.

Native: Ac–RMKQLEDKVEELLSKNYHLENEVARLKKLVG–NH_2
...gabcdefgabcdefgabcdefgabcdefgab...

344: Ac–RMKQLEDKVEELLSKLYHLANEVARLLKLVG–NH_2
...kabcdefghijkabcdefghijkabcdefgh...

434: Ac–RMKQLEDLVEEALSKLYHLENEVARLLKLVG–NH_2
...hijkabcdefghijkabcdefghijkabcde...

443: Ac–RMKQAEDKVEELLSKAYHLLNEVARLAKLVG–NH_2
...defghijkabcdefghijkabcdefghijka...

Figure 2.8: Shown are the sequences of GCN4 (Native) and three variants I designed that I hypothesized to adopt a heterochiral quaternary structure. Underneath the GCN4 sequence, the heptad repeat pattern that it adopts is shown. ***a*** and ***d*** residues are highlighted in red. Underneath the GCN4, the three variant GCN4 sequences are shown: 344, 434, and 434. Any substitutions there were made to the parent sequence are highlighted in orange within the 344, 434, and 443 sequences. The postulated hendecad repeat pattern is shown aligned underneath each sequence. The ***h*** position is highlighted in green.

To induce the GCN4 peptide to adopt a heterochiral assembly, I designed three variants of the GCN4 that introduce hydrophobic residue substitutions that were hypothesized to stabilize a hendecad repeat pattern adopting quaternary structure. These variants were designed by first overlaying the heptad repeat pattern observed within the GCN4 homodimer and then best aligning a hendecad repeat pattern with the heptad repeat pattern. In all attempts to overlay the two repeat patterns at least three ***a***, ***d***, or ***h*** residues within any of the hendecad repeat patterns would contain charged or polar residues. Placing charged or polar residues at a hydrophobic position would be inherently destabilizing to forming a heterochiral dimer featuring a hendecad repeat pattern (**Figure 2.8**). These hendecad-destabilizing features were a charged or polar residue at an ***a*** or ***d*** position, or a charged, polar or large hydrophobic residue at an ***h*** position. I named each variant sequence based on the number of residues between each ***a***, ***d***, and ***h*** when reading from the N- to C-

terminus. For example, reading the 344-sequence left to right shows an *i*(***a***), *i+3*(***d***), *i+7*(***h***), *i+11*(***a***) spacing between hydrophobic positions. The variants are related to one another by translation of the hendecad repeat pattern three or four residues along the peptide sequence.

To remedy the presence of charged or polar hendecad-destabilizing positions, leucine residues were substituted into the sequence at any ***a*** or ***d*** position that was not a hydrophobic residue (***a/d***-type substitution). Any ***h*** positions that were not alanine or valine were substituted to an alanine (***h-***type substitution). Variants 344 and 434 have similar sets of substitutions to form their hendecad repeat pattern sequences: two or three ***a/d***-type, respectively, and one ***h***-type substitution. Variant 443 had one ***a/d***-type and three ***h***-type. These substitutions allow the variant sequences to achieve a theoretically ideal hydrophobic residue spacing for adopting a hendecad repeat pattern.

In the following sections, the results of the racemic crystallization of the native GCN4 peptide and each of the variants, followed by X-ray diffraction of the best-growing crystals, and crystal structures deduced from our X-ray diffraction datasets will be discussed.

2.3 – Results of Racemic Crystallization

All X-ray diffraction data was collected by Dr. Craig Bingman. Crystal structure solution and refinement of 344-variant conducted by Christopher Dade.

2.3.1 – Racemic Crystallization of GCN4

As mentioned in **Section2.2.3**, our group had determined that when crystallized with its racemate GCN4 adopts a homochiral dimer (unpublished). However, the preliminary

crystal structure obtained was unable to be refined to sufficient quality to deposit into the protein structure database. To establish a standard for comparison of the GCN4 variants and generate a deposition-quality structure, I began by determining the crystal structure of the racemic form of the native GCN4 dimerization domain.

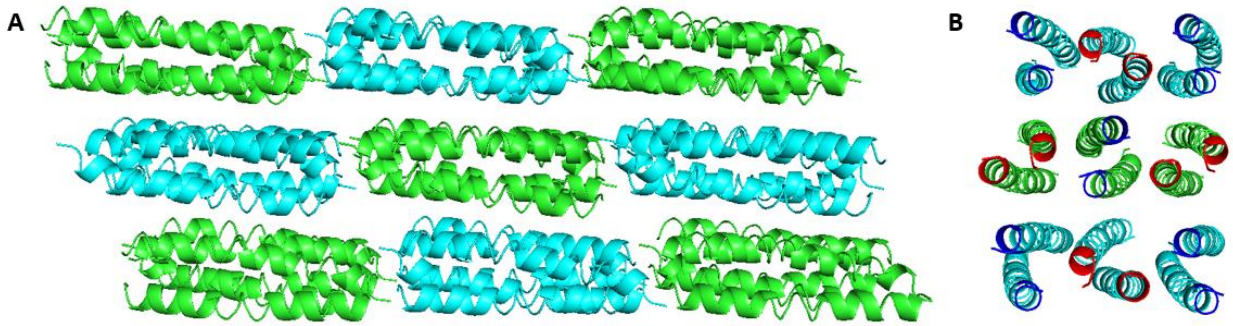


Figure 2.9: (**A**) Crystal packing side view of a 3x3x3 Rac-GCN4 CC array. (**B**) Aerial view of the middle segment in **A**. L-peptides are colored green and D-peptides are colored cyan in both images. In **B**, the nearest terminus of each peptide, C-term or N-term, has been colored in red or blue, respectively.

I was able to grow a GCN4 peptide crystal that produced a complete X-ray diffraction dataset, diffracting to 1.53Å. The resulting data set revealed that the crystal system resided in the C*2/c* space group. Notably, C*2/c* is an achiral space group with a D value of 7, as opposed to C2 which is a chiral space group with a D value of 6 that the all-L GCN4 crystallized. The crystal structure of racemic GCN4 reveals a parallel homochiral CC dimer which displays a heptad repeat pattern (**Figure 2.9**). The characteristic structural features of the all-L GCN4 dimer, such as its heptad repeat pattern and N16-N16' hydrogen bond buried within the hydrophobic interface are evident[2]. Layers of L-dimers alternate with layers of D-dimers within the crystal lattice. Each dimer makes six crystal contacts to neighboring CCs (**Figure 2.9B**).

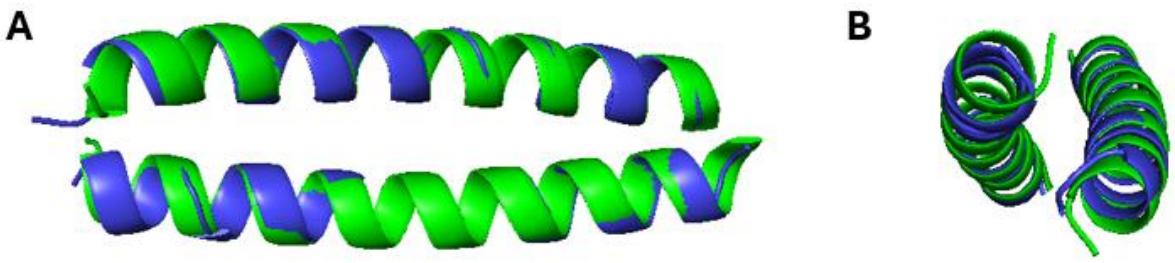


Figure 2.10: Alignment of the racemic GCN4 with optically active GCN4 (2ZTA) from the (**A**) side profile and (**B**) top view. A backbone RMSD of 0.452Å was calculated. Racemic GCN4 (blue) and all-L GCN4 (Green).

When an L-dimer of our racemic GCN4 structure was aligned with the previously reported GCN4 (PDB: 2ZTA), we calculated a backbone RMSD = 0.452Å. Thus, crystallization of GCN4 in presence of its racemate does not influence the homochiral coiled-coil quaternary structure.

2.3.2 – Summary of Crystallization and Diffraction of Racemic GCN4 Variants

Synthesis, crystallization trials, and X-ray diffraction experiments for each of the GCN4 variants as a racemic mixture were conducted. Crystallization of racemic 344-variant resulted in five crystals of X-ray diffraction quality. X-ray diffraction experiments of each of the crystals afford five datasets ranging from 1.28-1.68 Å. From those datasets, we solved a single structure with a resolution of 1.42 Å (see **Section 2.3.3**). Crystallization of 434-variant resulted in two crystals of X-ray diffraction quality. X-ray diffraction experiments with the 434-variant crystals resulted in two datasets at a resolutions of 1.18 and 1.29 Å. From those two datasets we solved a preliminary structure solution of the 1.29 Å dataset (see **Section 2.3.4**). Two crystals from a racemic mixture of the 443-variant were grown and subjected to X-ray diffraction. However, both 443-variant crystals were unable to produce complete datasets (data not shown). Crystallization conditions, resolutions observed, completeness of the X-

ray diffraction datasets, and the observed space groups observed for the crystals grown of racemic mixtures of GCN4, 344-variant, and 434-variants are summarized in **Table 2.1**.

Sample	Crystallization Condition	Resolution	I/σ	Completeness Inner/Outer	Space Group
Rac-GCN4	0.1M Potassium thiocyanate, 30% w/v Polyethylene glycol monoethyl ether 2,000	1.53Å	21.2	100/99	C2/c
Rac-344	0.1M HEPES pH 7.5, 0.5M Magnesium formate dihydrate	1.44Å	50.3	97.5/91.7	P222
	0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350	1.42Å	63.7	100/99.9	$P2_1$
	0.2 M Trimethylamine N-oxide dihydrate, 0.1 M Tris pH 8.5, 20% w/v Polyethylene glycol monomethyl ether 2,000	1.68Å	38.5	99.8/100	$P2_1$
	0.2 M Sodium chloride, 0.1 M Tris pH 8.5, 25% w/v Polyethylene glycol 3,350	1.28Å	28.5	91.6/18.5	P2
	0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350	1.57Å	36.4	99.5/95.1	$P2_12_12_1$
Rac-434	0.1 M BIS-TRIS pH 6.5, 3.0 M Sodium chloride	1.29Å	49.6	100/100	$P\bar{4}2_1m$
	1.0 M Ammonium sulfate, 0.1 M BIS-TRIS pH 5.5, 1% w/v Polyethylene glycol 3,350	1.18Å	63.3	100/54	$P\bar{4}2_1m$

Table 2.1: Crystallization conditions for each diffraction data set collected along with the resolution, I/σ, the completeness of the inner and outer sphere, and the observed space group. All data shown are from isotropic processing.

The racemic 344-variant, like racemic GCN4, crystallized in monoclinic crystal systems (P2, $P2_1$, and $P2_12_12_1$). Unlike racemic GCN4, the racemic 344-variant crystallized exclusively in chiral space groups. A possible explanation for observing chiral space groups

is the spontaneous resolution of enantiomers[39,40]. This is the spontaneous segregation of enantiomers in a racemic mixture to then form enantiomeric crystals. Racemic 434-variant crystallized into an achiral space group, $P\bar{4}2_1m$, suggesting incorporation of both enantiomers within the crystallographic unit cell.

2.3.3 – GCN4 344-Variant Structure

The structure solution and refinement of the 344-variant were completed by Christopher Dade.

As mentioned in **Section 2.3.2**, one structure has been solved from the 344-variant diffraction data sets. This structure was solved via molecular replacement using a theoretical model of an L-344-variant monomer search model. The model was designed by modifying an L-GCN4 monomer from within the racemic GCN4 structure (**Figure 2.9**) and "mutating" three residues (N16L, E20A, and K27L) to match the 344-variant sequence. This approach to design a theoretical L-344-variant allows the search model to retain the conformation that the GCN4 monomer would adopt when forming a homochiral coiled coil while displaying the same residues within the 344-variant sequence.

The racemic 344-variant structure displayed a sparsely reported mode of tetrameric coiled coil assembly, an antiparallel-dimer of parallel dimers (AD-PD). To the best of my knowledge, the GCN4 variant GCN4-pAe is only one other example of a GCN4 variant adopting an AD-PD quaternary structure[41]**.** The differences between the 344-variant and GCN4-pAe will be discussed later in this section. The 344-variant structure was solved in the space group $P2_1$ with a resolution of 1.42Å and an R_{free} of 0.19. Since this structure is

comprised exclusively of L-344 peptides in a chiral space group, crystallization of racemic 344 resulted in spontaneous resolution of enantiomers.

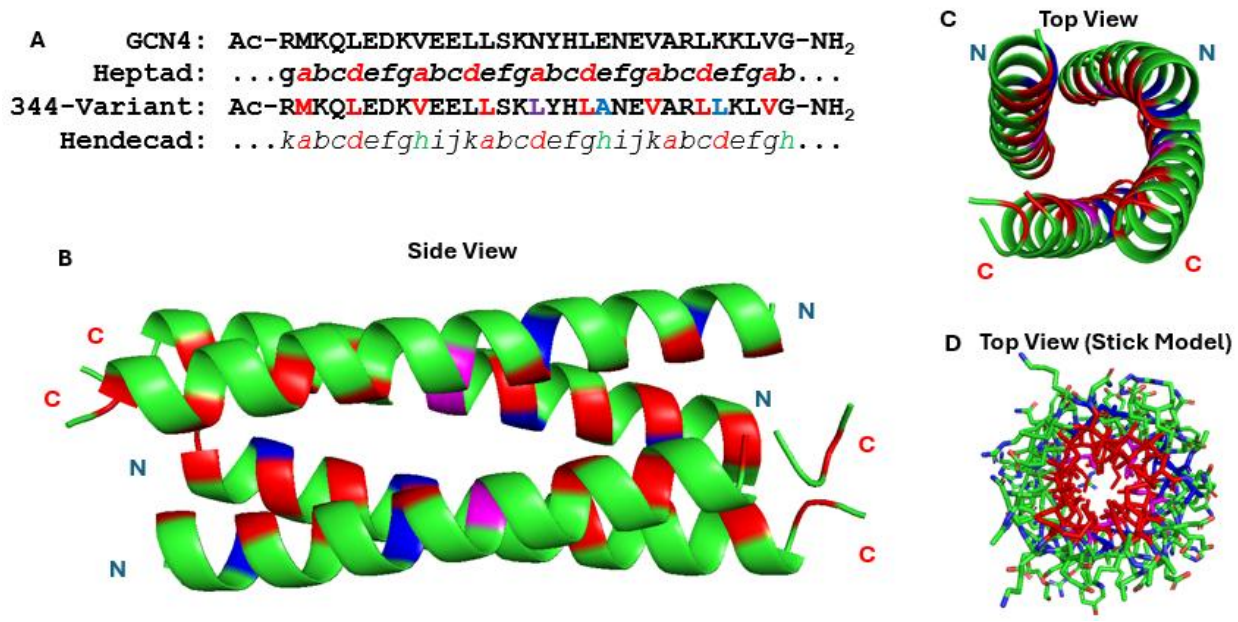


Figure 2.11: Structure and sequence alignment of the L-344 tetramer. (**A**) Native GCN4 sequence aligned with its observed heptad repeat and 344-variant aligned with its HRP. For Heptad and Hendecad repeat pattern overlays, ***a*** and ***d*** positions and ***h*** positions are shown in red and green, respectively. (**B**) Side and (**C**) top view of the tetrameric CC assembly of 344-variant. (**D**) Alternative top view where the cartoon has been replaced with sticks. For all structure and sequence of 344-variant, residues that appear in GCN4's heptad repeat are shown in red. Residues substituted to promote a heterochiral assembly that are shared by the heptad and hendecad repeat patterns are shown in purple. Residue substitutions unique to the hendecad repeat pattern are shown in blue. C- and N-terminus in panels **B** and **C** labeled with a red "C" or blue "N", respectively.

The residues of each 344-variant monomer that form the hydrophobic interface include the native GCN4 heptad repeat pattern residues and the hendecad-promoting substitutions (**Figure 2.11B**). When looking down the tetramer axis, it becomes apparent that the oligomeric state is driven by maximizing the hydrophobic burial of the heptad and hendecad residues. In both the cartoon and stick model, the hydrophobic residues of the heptad and hendecad repeat patterns face towards the center of the tetramer to facilitate van der Waals packing interactions between adjacent helices. I suspect that the inclusion of additional hydrophobic residues within the GCN4 sequence increased the size of the

hydrophobic face, which allowed for additional partner helices to oligomerize relative to the GCN4 homodimer.

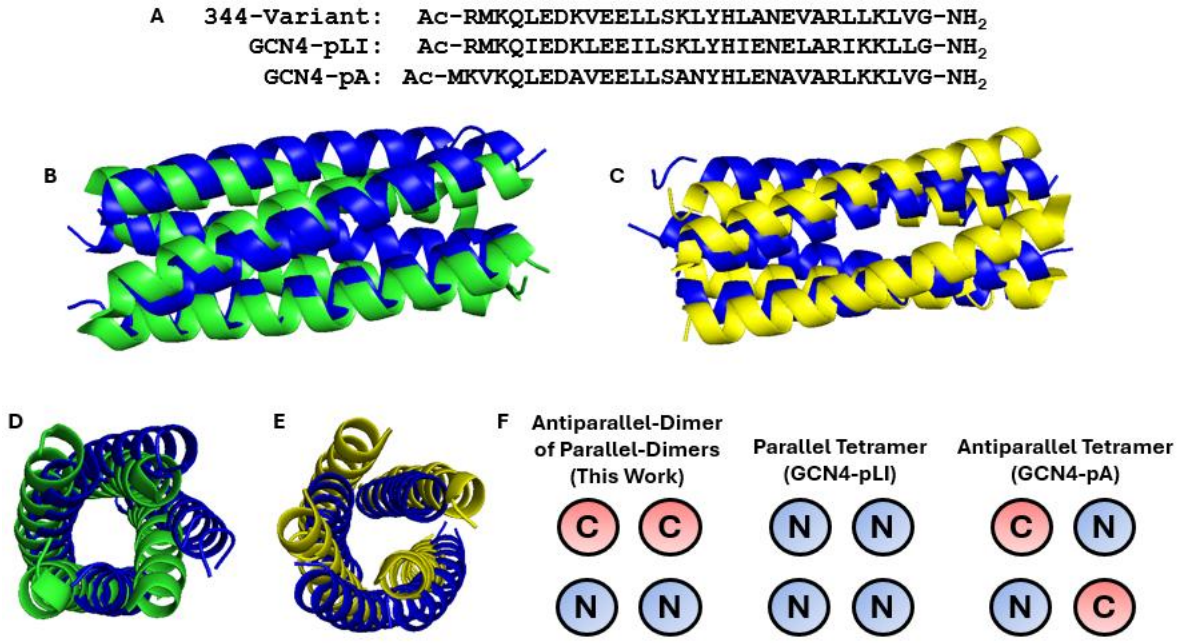


Figure 2.12: Comparison of 344-variant tetramer with the GCN4-pLI parallel tetramer (1GCL) and the GCN4-pA antiparallel tetramer (2B1F). (**A**) 344-variant sequence aligned with GCN4-pLI and GCN4-pA. (**B**) Side view and (**D**) top view alignment of 344-variant (blue) with GCN4-pA (green). (**C**) Side view and (**E**) top view alignment of 344-variant (blue) with GCN4-pLI (yellow). (**F**) Directionality of monomers inside of the each GCN4 tetramer. C-termini labeled with a "C" and N-termini labeled with an "N".

To verify that our structure is in fact a CC, we input our structure into the CC-detecting software, Socket2, developed by the Woolfson Lab[42,43]. Socket2 evaluates protein assemblies for the characteristic "Knobs in Holes" interactions that define CC assemblies. Using Socket2, I verified that the 344-variant tetramer satisfies the requirements for a CC for residues 5-30 for each monomer. Each monomer of the 344-variant must make two distinct sets of contacts with a neighboring helix, one parallel and one antiparallel.

Next the 344-variant structure will be compared to the GCN4-pLI and GCN4-pA structures as the representatives of a parallel and antiparallel tetrameric CC, respectively. GCN4-pLI, a GCN4 variant with all ***a*** and ***d*** positions substituted to leucine and isoleucine

respectively, crystallizes as a parallel homotetramer[38]. Each monomer within the GCN4-pLI tetramer contacts two copies of itself both in a parallel orientation. When aligned with the 344-structure (**Figure 2.12C,E**), an RMSD of 9.565 Å is observed. This indicates substantial structural deviations between the two structures. GCN4-pA is a GCN4 variant with five substitutions that crystallizes as an antiparallel homotetramer[44]. Each monomer within the GCN4-pA tetramer makes contacts with two adjacent α-helices in an antiparallel orientation. When aligned with the 344-structure (**Figure 2.12B,D**), an RMSD of 6.553 Å is observed.

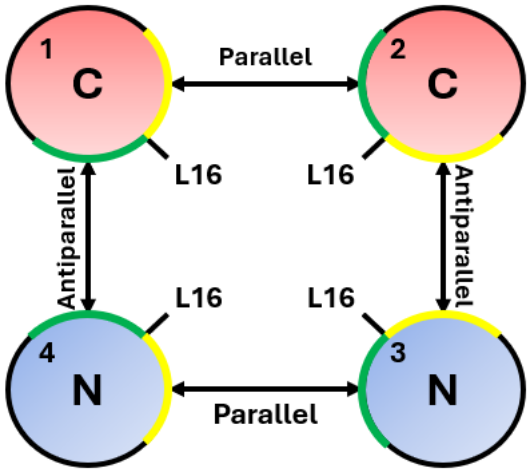


Figure 2.13: Directionality of monomers inside of the 344-variant antiparallel-dimer of parallel-dimers. The helices are numbered 1-4 clockwise starting from the top-left monomer. Leucine16 is depicted pointing towards the center of the helical bundle to indicate the orientation of each monomer within the bundle. Parallel and antiparallel relationships between adjacent helices are indicated. C-termini labeled with a "C" and N-termini labeled with an "N", and colored red and blue, respectively. The oligomeric interface of each monomer is divided into two regions, green and yellow, with the division being marked by the central Leucine16 residue.

The largest departure that an AD-PD tetramer displays from traditional CC tetramers is that a given hydrophobic interface of a monomer is selective for a single face of a partner helix depending on whether the partner helix is orientated parallel or antiparallel. This feature is highlighted in **Figure 2.13**. Interactions between parallel helices are facilitated through different regions of the hydrophobic interface (green-yellow), while interactions between antiparallel helices are facilitated through interactions between the same regions (green-green or yellow-yellow).

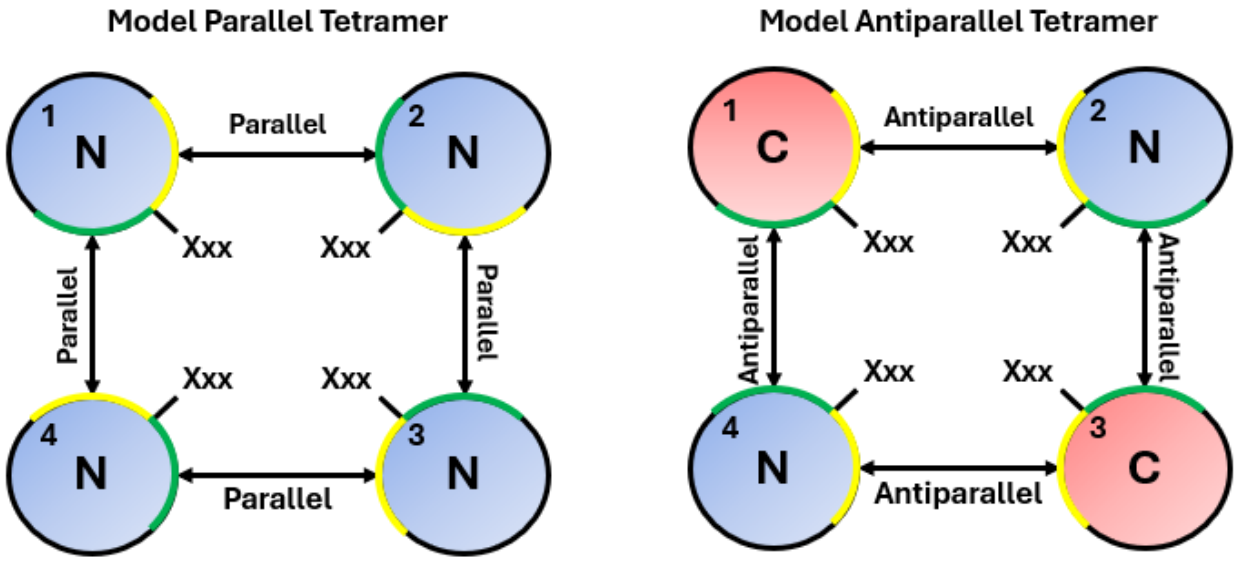


Figure 2.14: Directionality of monomers inside of the 344-variant antiparallel-dimer of parallel-dimers. The helices are numbered 1-4 clockwise starting from the top-left monomer. Xxx depicts a central residue that points towards the center of the helical bundle. Parallel and antiparallel relationships between adjacent helices are indicated. C-termini labeled with a "C" and N-termini labeled with an "N", and colored red and blue, respectively. The oligomeric interface of each monomer is divided into two regions, green and yellow, with the division being marked by the central Xxx residue.

A single peptide having multiple orientations for homo-oligomerization is an atypical feature in CCs. For instance, in a parallel tetrameric CC, each monomer will make one set of parallel contacts with the adjacent monomer's oligomeric interface (green-yellow). While antiparallel tetrameric CCs will make two sets of antiparallel contacts between the adjacent monomers (green-green and yellow-yellow), one on each side of a participating α-helix (**Figure 2.14**).

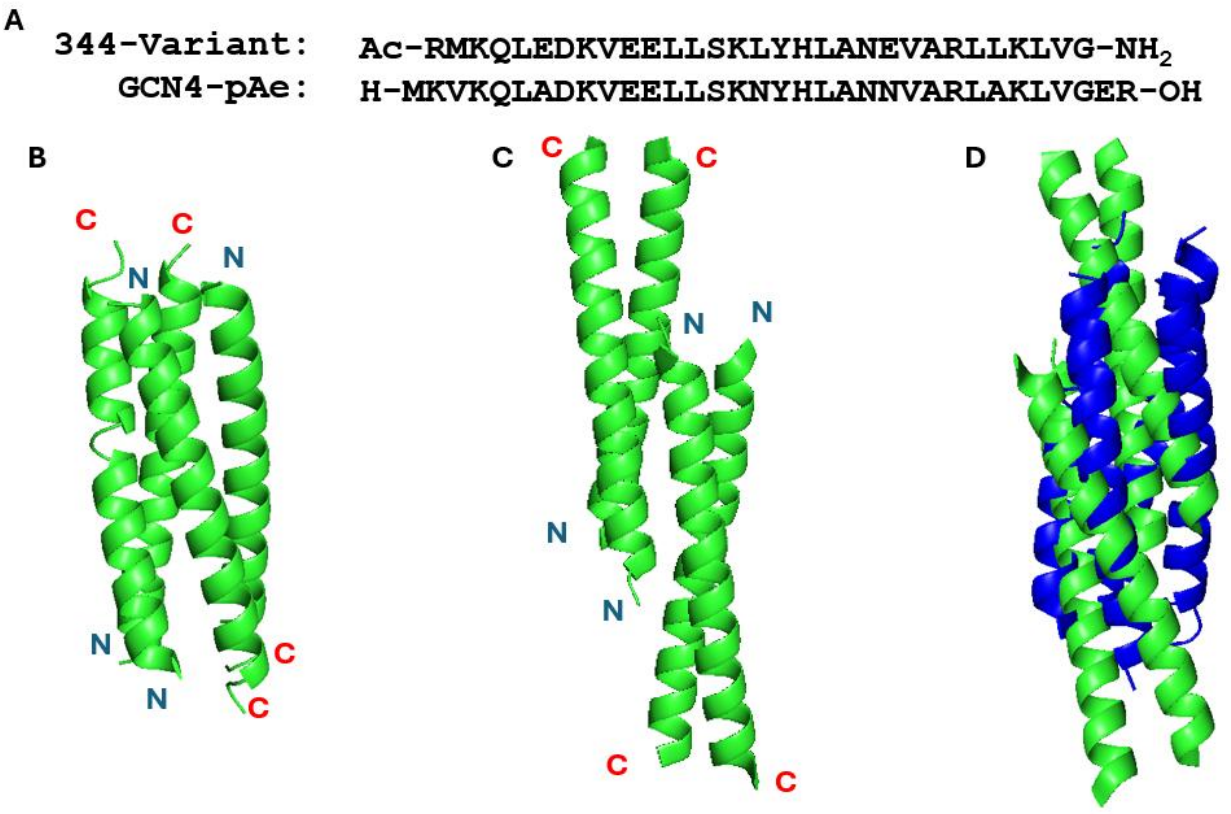


Figure 2.15: (**A**) Sequences of 344-variant and GCN4-pAe. (**B**) Structure of 344-variant and (**C**) Structure of GCN4-pAe (PDB: 2NRN). The C- and N-termini are labelled with a red "C" and blue "N", respectively. (**D**) Alignment of 344-variant and GCN4-pAe. The 344-variant is colored in blue and GCN4-pAe is colored in green.

As mentioned earlier in this section, to the best of my knowledge, GCN4-pAe is the only other GCN4 variant that displays an AD-PD quaternary structure[41]. Despite both assemblies displaying AD-PD quaternary structure and based on the GCN4 sequence, there are substantial differences between each structure. The AD-PD domain observed within the GCN4-pAe structure is formed between the first two heptad repeats from the N-terminus of each monomer. The C-terminal region of each monomer of GCN4-pAe exhibits a parallel homodimer quaternary structure. The 344-variant adopts an AD-PD quaternary structure throughout the entire length of the tetramer . When the two assemblies are aligned with one another, an RMSD value of 14.758 Å was calculated (**Figure 2.15**). The calculated RMSD value of the 344-variant aligned with GCN4-pAe is greater than when aligned with GCN4-pLI

(9.565 Å) or GCN4-pA (6.553 Å). Despite the 344-variant and GCN4-pAe both adopting an AD-PD quaternary structure, the two structures are less similar than the 344-variant compared to a model parallel or antiparallel GCN4 variant tetramer.

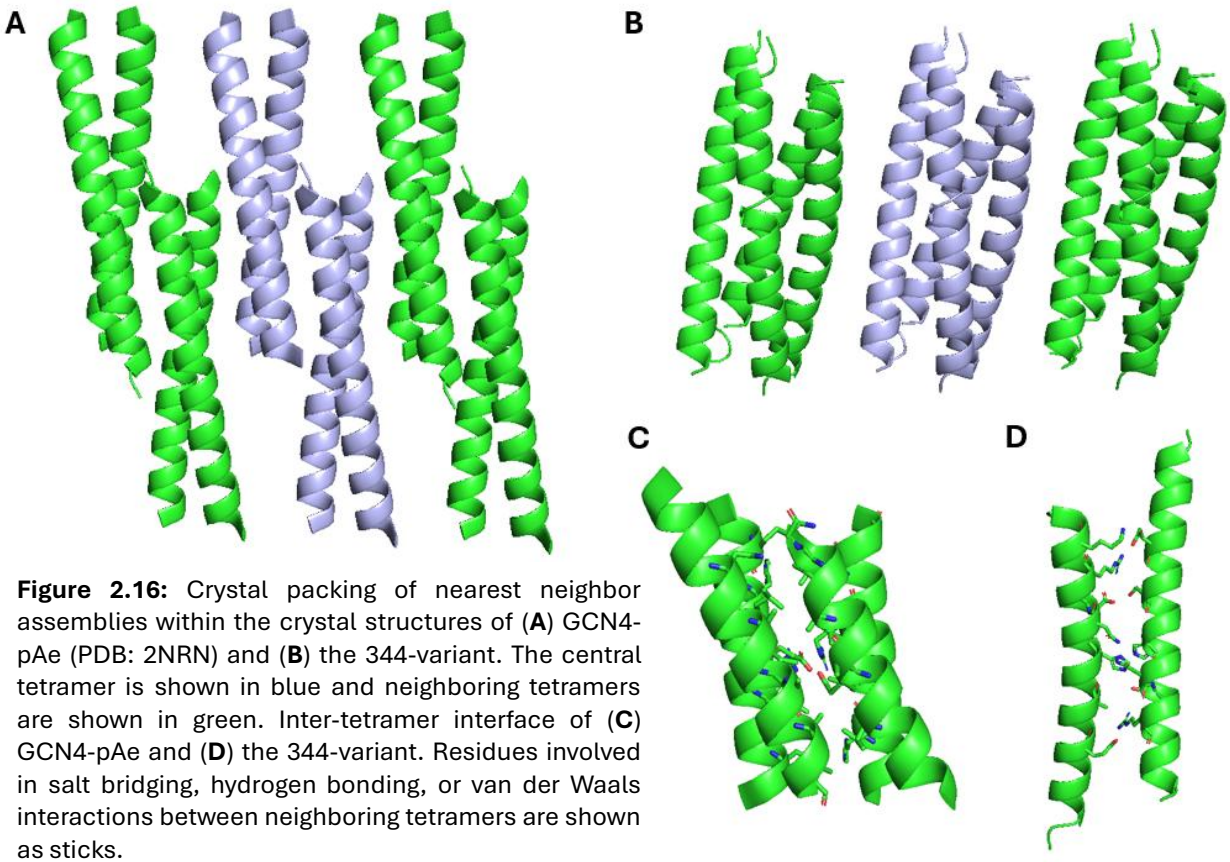


Figure 2.16: Crystal packing of nearest neighbor assemblies within the crystal structures of (**A**) GCN4-pAe (PDB: 2NRN) and (**B**) the 344-variant. The central tetramer is shown in blue and neighboring tetramers are shown in green. Inter-tetramer interface of (**C**) GCN4-pAe and (**D**) the 344-variant. Residues involved in salt bridging, hydrogen bonding, or van der Waals interactions between neighboring tetramers are shown as sticks.

In solution the tetrameric AD-PD crystal structure of GCN4-pAe was observed to sediment as a dimer between 50–800 μM after analytical ultracentrifugation[41]. Indicating that the observed AD-PD may be an artifact of crystallization and not the in solution oligomeric state. The crystal packing of GCN4-pAe also supports that the observed tetrameric AD-PD quaternary structure may be a crystallographic artifact. The GCN4-pAe tetramers are located close to the neighboring assembly allowing for inter-tetramer interactions (**Figure 2.16A,C**). Adjacent tetramers are associated via salt bridging, hydrogen

bonding, and van der Waals interactions. These inter-tetramer interactions result in lateral association and then fibrillation of the GCN4-pAe tetramers.

The 344-variant structure displays less inter-tetramer interactions than the GCN4-pAe structure. The inter-tetramer contacts of the 344-variant structure are facilitated by between two monomers within each tetramer, as opposed to all monomers in the GCN4-pAe structure (**Figure 2.16B,D)** . The 344-variant's inter-tetramer contacts are formed via salt bridging and hydrogen bonding interactions between adjacent tetramers. The reduction in inter-tetramer contacts shown in the 344-variant structure compared to the GCN4-pAe supports that the AD-PD quaternary structure of the 344-variant is not a crystallographic artifact, but further studies are required to verify the in-solution oligomeric state.

2.3.4 – Racemic GCN4 434-Variant Structure

The structure solution and refinement of the 434-variant were completed by Christopher Dade.

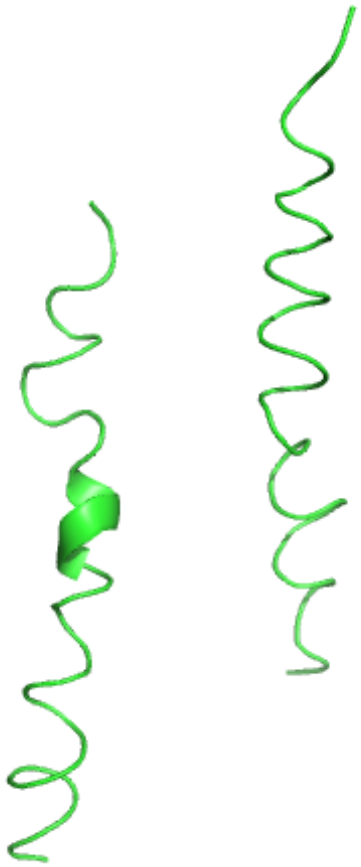


Figure 2.17: Preliminary structure of 434-variant heterochiral dimer. An L- and D-peptide are both detectable within the unit cell. This structure is not ready for deposition as of yet!

We have been able to deduce a unreported mode of assembly for tetrameric CCs with our 344-variant structure. However, we have not yet been able to generate a well-refined model of a heterochiral GCN4 variant. We currently have a been able to solve a X-ray diffraction set of the 434-variant in the achiral space group of $P\bar{4}2_1m$. The current R_{free} for this structure is greater than 0.4, which is not adequate for PDB deposition or preliminary deduction of quaternary contacts. We deduce from this dataset that the 434-variant crystallizes as a racemate, as with one copy of each enantiomer is present within the unit cell (**Figure 2.17**).

We will continue to attempt refinement of this structure.

2.4 – Conclusions

In this chapter, I report on the utilization of racemic crystallization to investigate the quaternary contacts of GCN4 and GCN4 sequence variants. We discovered that the crystal structure of homochiral GCN4 as a dimeric CC reported by O'Shea et al.[2] is consistent when GCN4 is crystallized as a racemic mixture. However, the space group in which the peptide crystallizes is achiral, C2/c.

I designed three variants of GCN4, each aimed at promoting a dimer based on a heterochiral hendecad repeat pattern. Two of the variants were able to produce diffraction datasets that we solved, to varying degrees of success. The 344-variant underwent spontaneous resolution of enantiomers, and the resulting tetramer structure was comprised entirely of L-peptides. The CC tetramer structure solved for the 344-variant revealed an AD-PD quaternary structure that deviated significantly from the GCN4-pAe AD-PD structure[41].

We also managed to find a structure solution for a 434-variant crystal. The 434-variant crystallized in an achiral space group, $P\bar{4}2_1m$, and within its unit cell one copy of each the L- and D-peptide were detected. While more structure refinement is required to deduce the quaternary contacts within the 434-variant crystal structure, the current model suggests a heterochiral dimeric association.

2.5 – Methods and Characterization

2.5.1 – Peptide Synthesis and Purification

All peptides were synthesized by manual microwave-assisted solid-phase peptide synthesis at 100 µmol scale. Low-Load Rink Amide resin was used as solid support. For each coupling reaction, Fmoc-protected amino acid (4 eq), DIEA (4 eq), and HATU (eq) in biotech-grade DMF were mixed with resin, and the mixture was heated to 75°C for 5 minutes while stirring. After the coupling period, the reaction vessel was drained and washed with ACS-grade DMF (3x10 mL). For Fmoc-deprotection, 20 % (v/v) piperidine in ACS-grade DMF (5 mL) was added to the resin, and the mixture was heated at 80°C for 2 minutes while stirring. After deprotection, the resin was washed with ACS-grade DMF (3x10 mL). Once peptide elongation was completed, a final deprotection reaction was completed, the resin was washed with ACS-grade DMF (3x10 mL) and 4 mL of an acetylation cocktail (8:2:1 BioTech-grade DMF:acetic anhydride:DIEA) was added and allowed to stir for 30 minutes at room temperature. The resin was then washed with ACS-grade DMF (3x10 mL), followed by DCM (3x10 mL) and placed under reduced pressure on an aspirator until dry. Peptides were then cleaved from the resin, with concomitant side chain deprotection, by combining the resin with 4 mL per 25 umol resin of Reagent K cleavage cocktail (82.5% TFA, 5% water, 5% thioanisole, 5% phenol, and 2.5% 1,2-ethanedithiol), and gently rocking the vessel for 3 hours. The reaction solution was then drained into a 45mL falcon tube, and the resin was washed twice with 2 mL TFA. The combined peptide cleavage solution was then concentrated under a stream of nitrogen gas until less than 1.0 mL remained or precipitation of the peptide was observed. The crude peptide was then precipitated with 40mL of cold

diethyl ether. This mixture was centrifuged at 4k rpm for 10 minutes to pellet the suspended solid, and then the ether was decanted. A second volume of 40mL cold diethyl ether was added, and the process of centrifugation and decanting was repeated. The resulting solid was dried under a stream of nitrogen gas to afford the crude product.

To purify the crude material, the solid peptide was dissolved in the minimal volume of ACN/H_2O required to fully dissolve the material (1-4mL depending on the peptide solubility) and passed through a 0.22µm filter. Purification was conducted via reverse-phase high performance liquid chromatography (RP-HPLC). The following conditions were used - CSH C18 stationary phase, 5 µm, 19mm x 250mm, the solvent gradient was 10-60% acetonitrile over 20 minutes. The purity of each peptide was assessed via ultra-high performance liquid chromatography (UPLC; CSH C18 stationary phase, 1.7 µm, 2.1 mm x 100 mm, the solvent gradient was 10-60% acetonitrile over 10 minutes). Peptide purity was assessed by integration of area under the curve while monitoring at 220 nm. Peptide concentration was determined by absorbance at 280 nm. This value was determined using previously reported literature values of proteogenic amino acids at 280 nm. The peptide solution was then dissolved in water, frozen, and lyophilized to provide a white powder. The material was then stored at 4 °C until use.

2.5.2 – Crystal Screening and Crystal Harvesting

Crystal harvesting was conducted by Dr. Craig Bingman

A stock solution of 5.0 mg/mL racemic peptide (2.5 mg/mL each enantiomer) was prepared in nanopure water. Crystal screens were conducted using the HR2-144 Crystal

Screening kit (Hampton Research) under hanging drop conditions. 1 µL of racemic peptide solution was added to the glass coverslip, and 1 µL of precipitant was added and the solution was mixed via pipetting up and down ten times. The glass slip was sealed over a well of 100 µL of precipitant, and the crystal tray was left at room temperature while crystal growth was monitored. Once a crystal had grown to X-ray diffraction quality size and dimensions, the crystal was harvested.

2.5.3 – X-Ray Diffraction

X-ray diffraction experiments were conducted by Dr. Craig Bingman

Harvested crystals had X-ray diffraction experiments conducted at the advanced photon source located at Argonne National Laboratory.

2.5.4 – Structure Solution and Refinement

344-variant structure solution and refinement was completed by Christopher Dade

All structures were solved using a single monomer of L-GCN4 which had been “mutated” to include the residues of the 344- or 434-variant as a search model using Phenix. Once a reasonable molecular replacement structure was solved, the resulting solution was imported to Coot and refined against the X-ray diffraction data. The refined structure was then scored in Phenix. The process of refining in Coot and scoring in Phenix was repeated until the final structure was solved.

2.5.5 – Refinement Statistics for 344- GCN4 Structure

Wavelength	
Resolution range	**24.46 - 1.42 (1.49 - 1.42)**
Space group	**P 1 21 1**
Unit cell	**24.897 46.27 46.997 90 100.786 90**
Total reflections	
Unique reflections	**19642 (2825)**
Multiplicity	
Completeness (%)	**98.56 (99.72)**
Mean I/sigma(I)	
Wilson B-factor	**17.07**
R-merge	
R-meas	

R-pim	
CC1/2	
Reflections used in refinement	**19642 (2825)**
Reflections used for R-free	**1044 (152)**
R-work	**0.1739 (0.2218)**
R-free	**0.1988 (0.2275)**
Number of non-hydrogen atoms	**1139**
macromolecules	**1024**
ligands	**12**
solvent	**103**
Protein residues	**123**
RMS(bonds)	**0.011**
RMS(angles)	**1.20**

Ramachandran favored (%)	**100.00**
Ramachandran allowed (%)	**0.00**
Ramachandran outliers (%)	**0.00**
Rotamer outliers (%)	**0.00**
Clashscore	**3.27**
Average B-factor	**24.62**
macromolecules	**23.55**
ligands	**26.95**
solvent	**34.97**

Statistics for the highest-resolution shell are shown in parentheses.

2.5.6 – Peptide Characterization

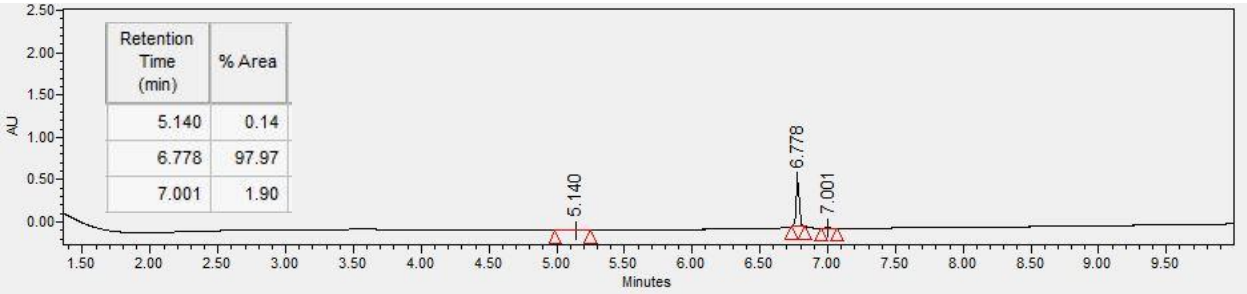


Figure 2.18: UPLC chromatogram of D-344-Variant monitored at 220nm. Purity >97%.

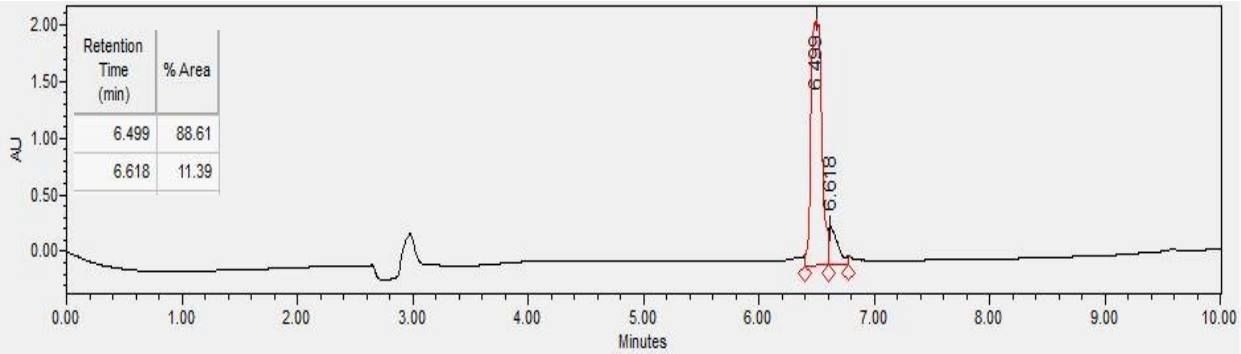


Figure 2.19: UPLC chromatogram of L-344-Variant monitored at 220nm. Purity >88%.

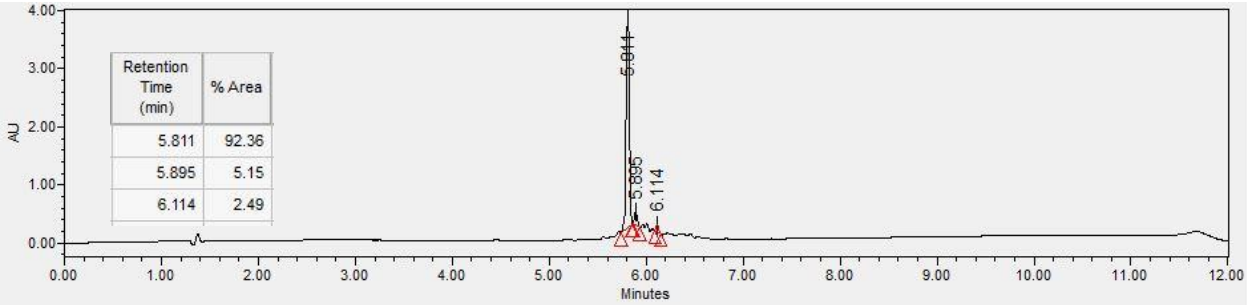


Figure 2.20: UPLC chromatogram of L-434-Variant monitored at 220nm. Purity >92%.

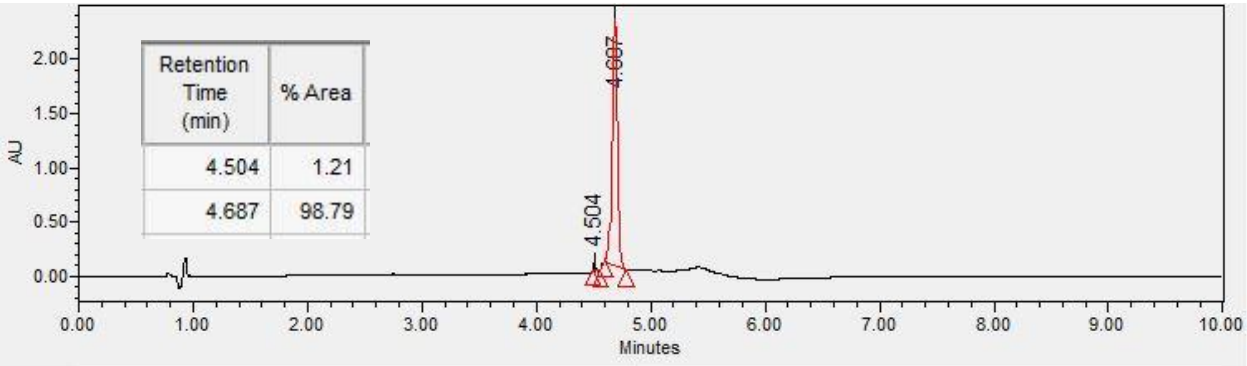


Figure 2.21: UPLC chromatogram of D-443-Variant monitored at 220nm. Purity >98%.

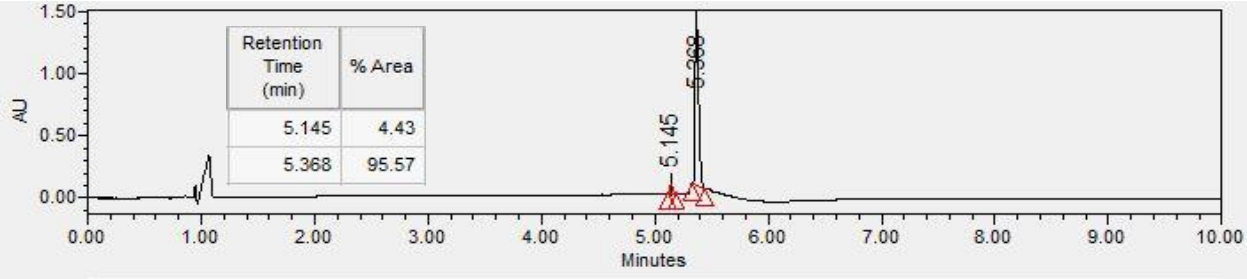


Figure 2.22: UPLC chromatogram of L-443-Variant monitored at 220nm. Purity >95%.

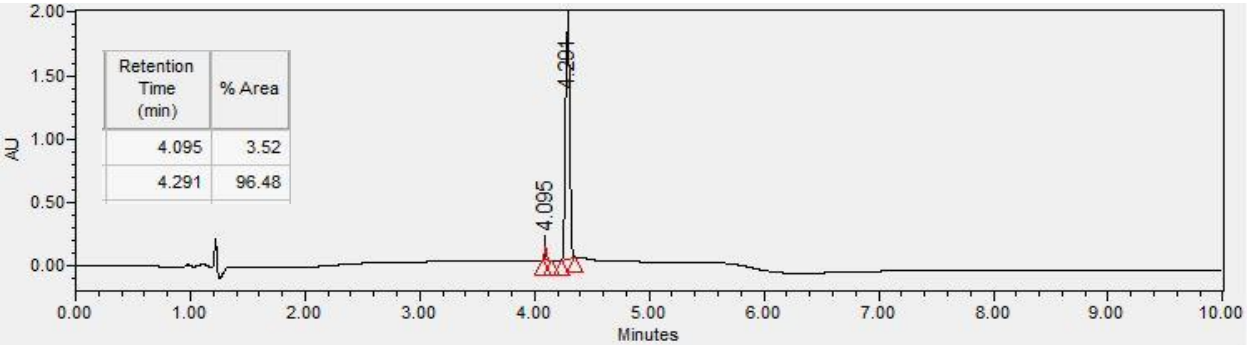


Figure 2.23: UPLC chromatogram of D-GCN4 monitored at 220nm. Purity >96%.

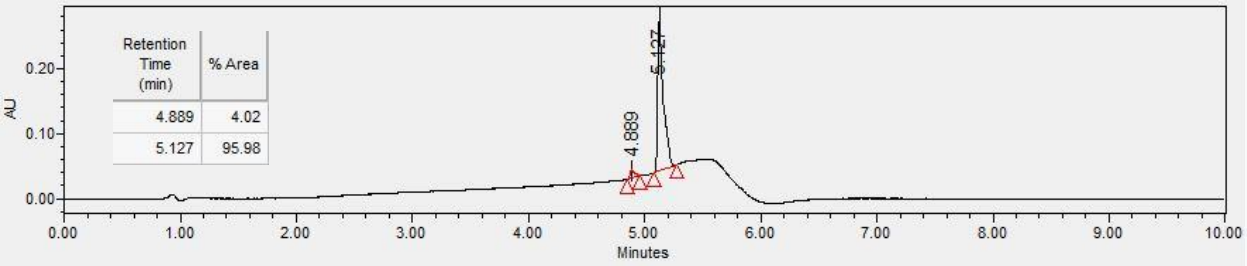


Figure 2.24: UPLC chromatogram of D-434-Variant monitored at 220nm. Purity >95%.

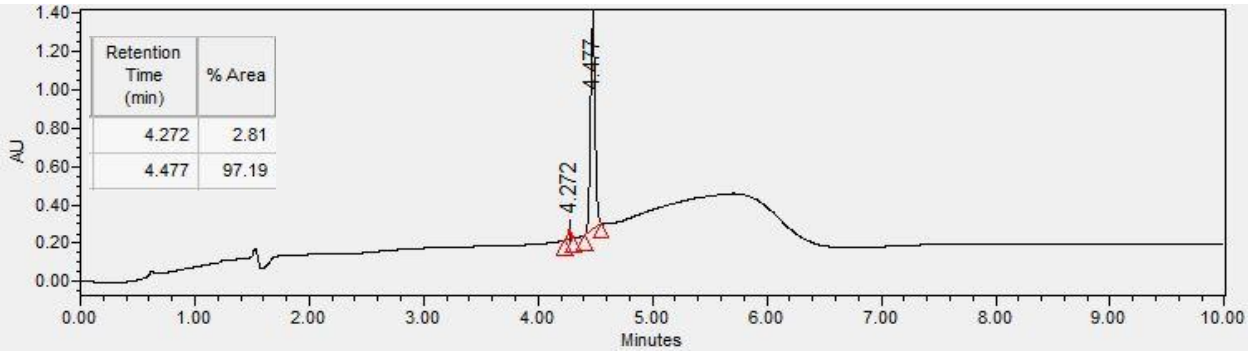


Figure 2.25: UPLC chromatogram of L-GCN4 monitored at 220nm. Purity >97%.

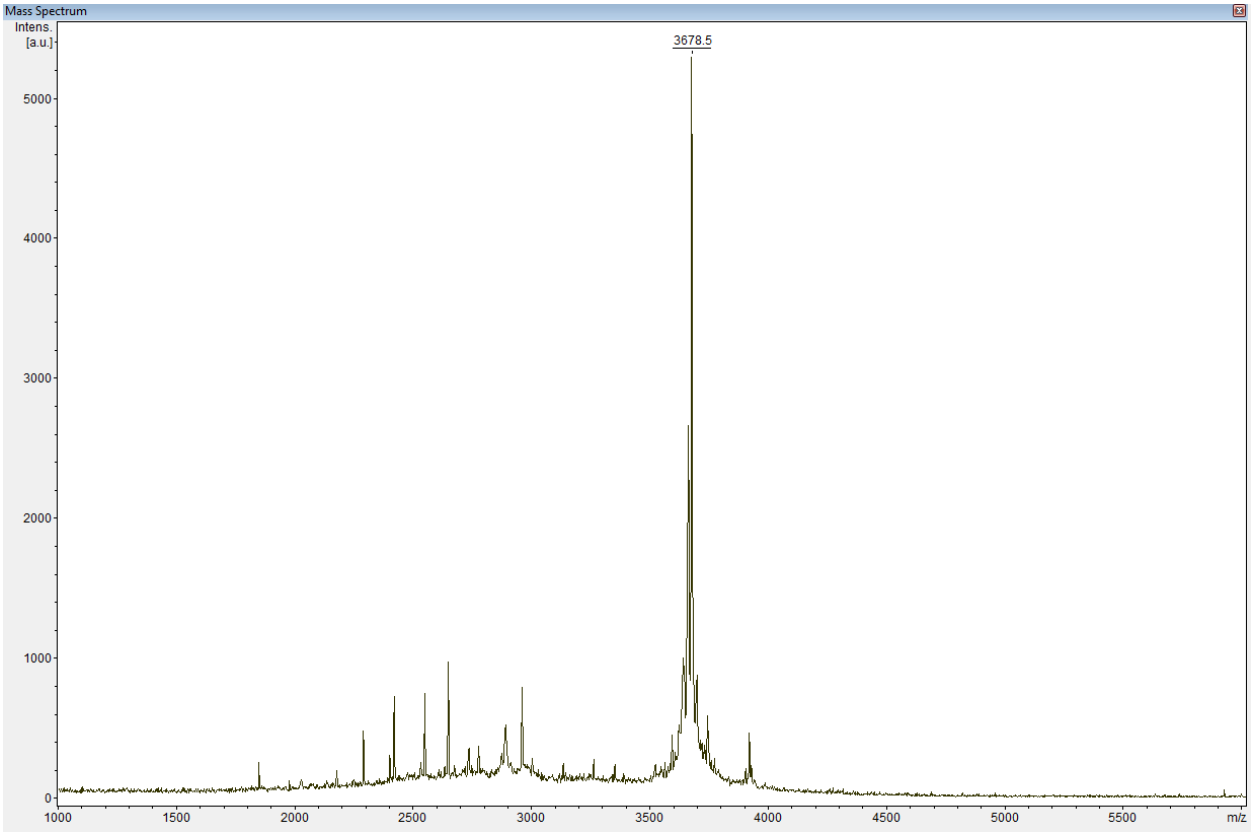


Figure 2.26: MALDI-TOF-MS of D-344-Variant. Observed mass: $[M+H]^+$ = 3678.5. Calculated mass: $[M+H]^+$ = 3677.1

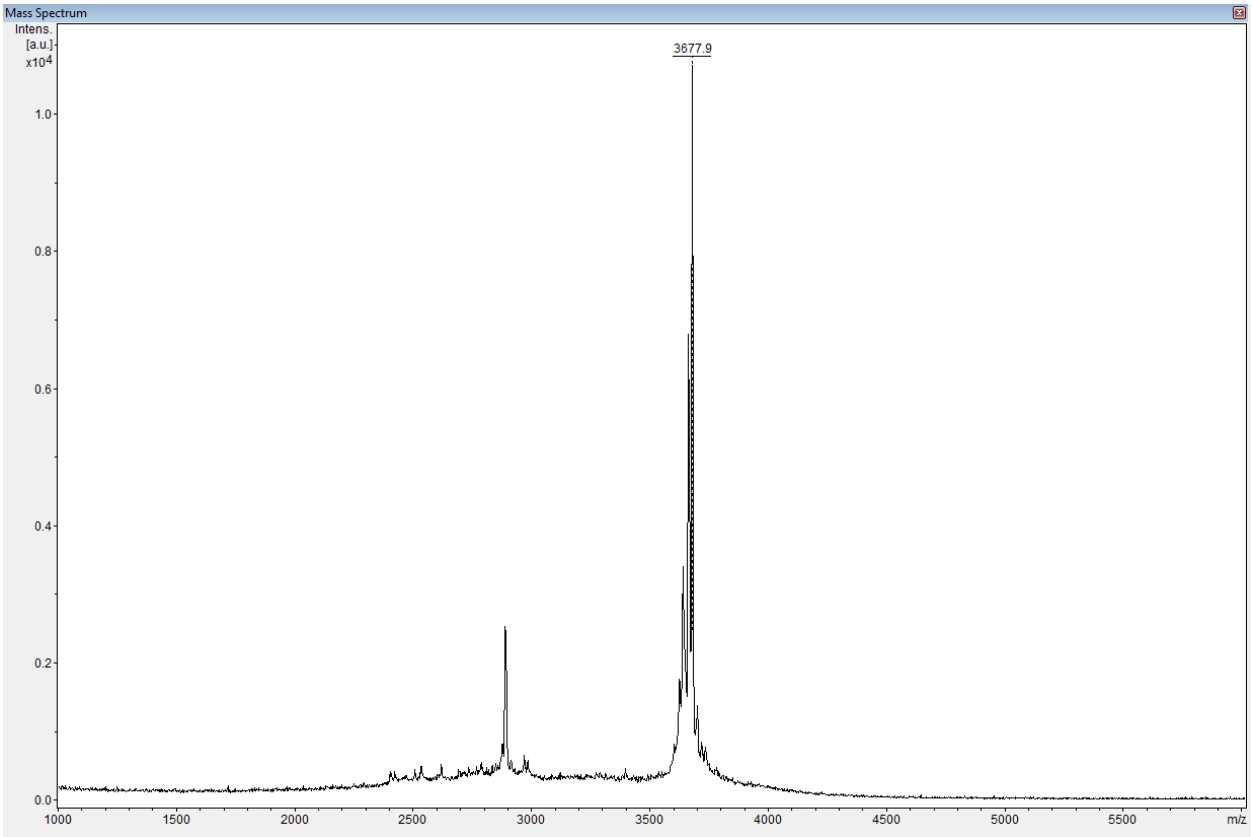


Figure 2.27: MALDI-TOF-MS of L-344-Variant. Observed mass: $[M+H]^+$ = 3677.9. Calculated mass: $[M+H]^+$ = 3677.1

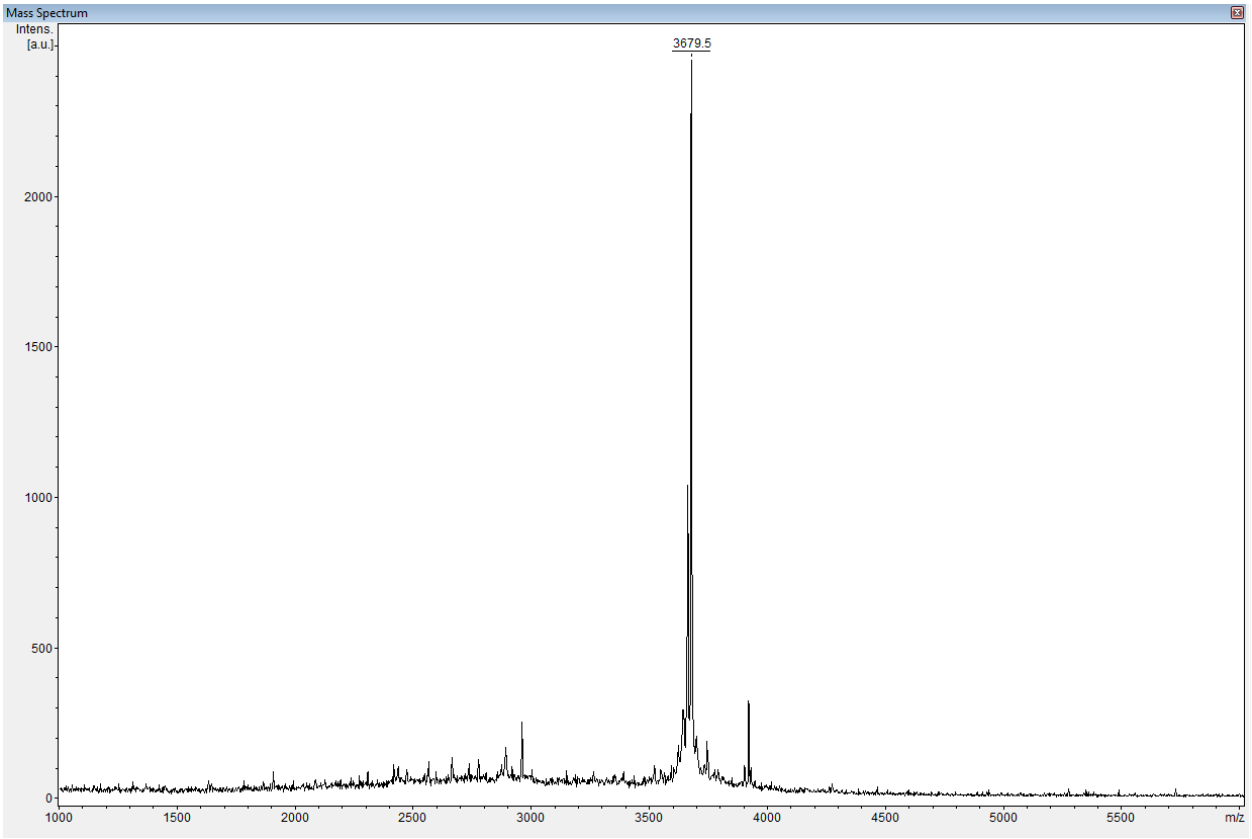


Figure 2.28: MALDI-TOF-MS of D-434-Variant. Observed mass: $[M+H]^+$ = 3679.5. Calculated mass: $[M+H]^+$ = 3677.1

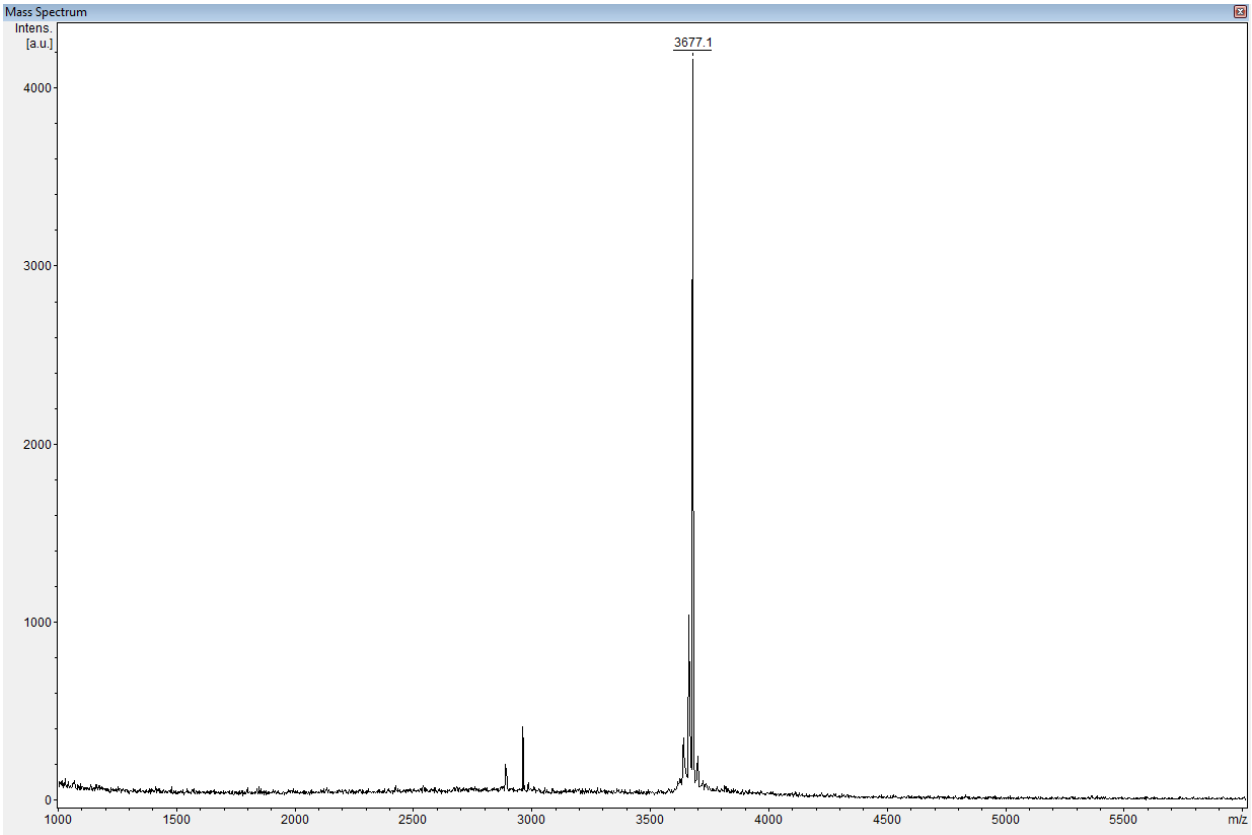


Figure 2.29: MALDI-TOF-MS of L-434-Variant. Observed mass: $[M+H]^+$ = 3677.1. Calculated mass: $[M+H]^+$ = 3677.1

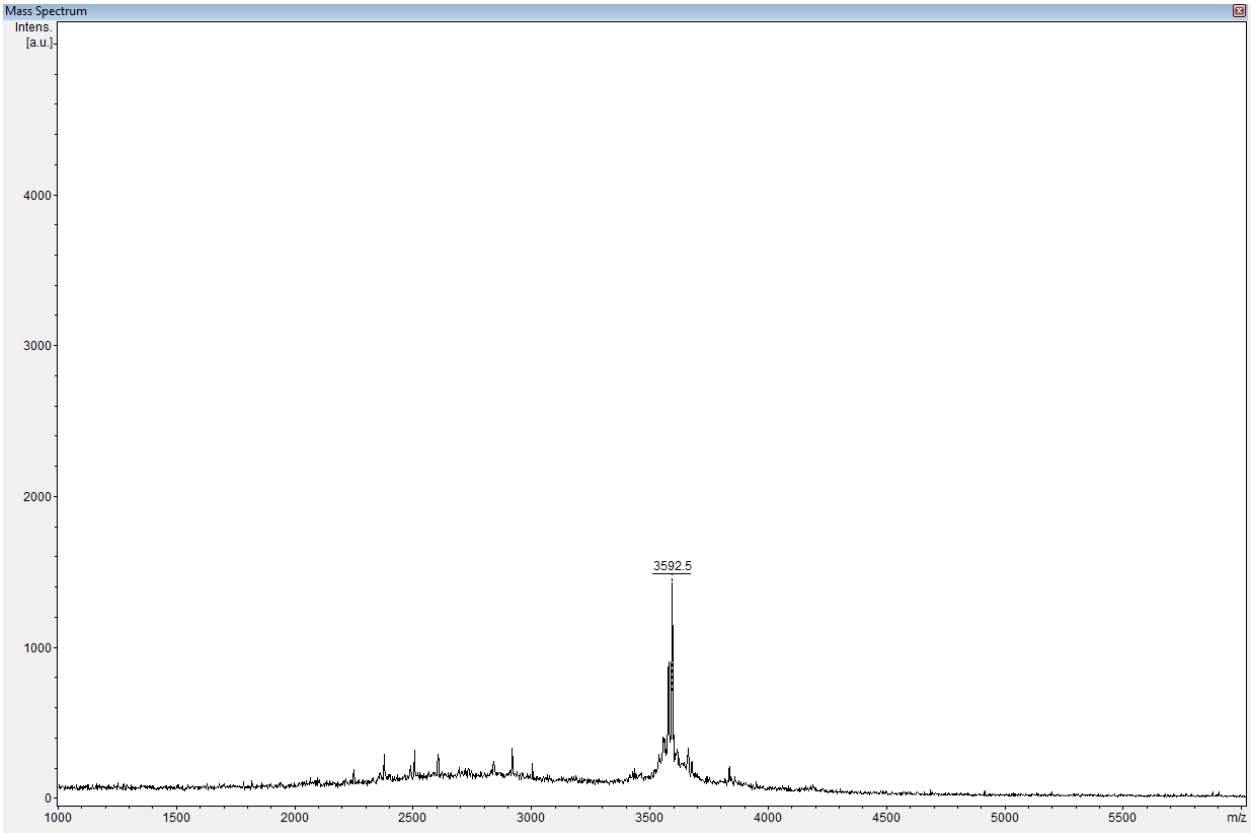


Figure 2.30: MALDI-TOF-MS of D-443-Variant. Observed mass: $[M+H]^+$ = 3592.5. Calculated mass: $[M+H]^+$ = 3593.0

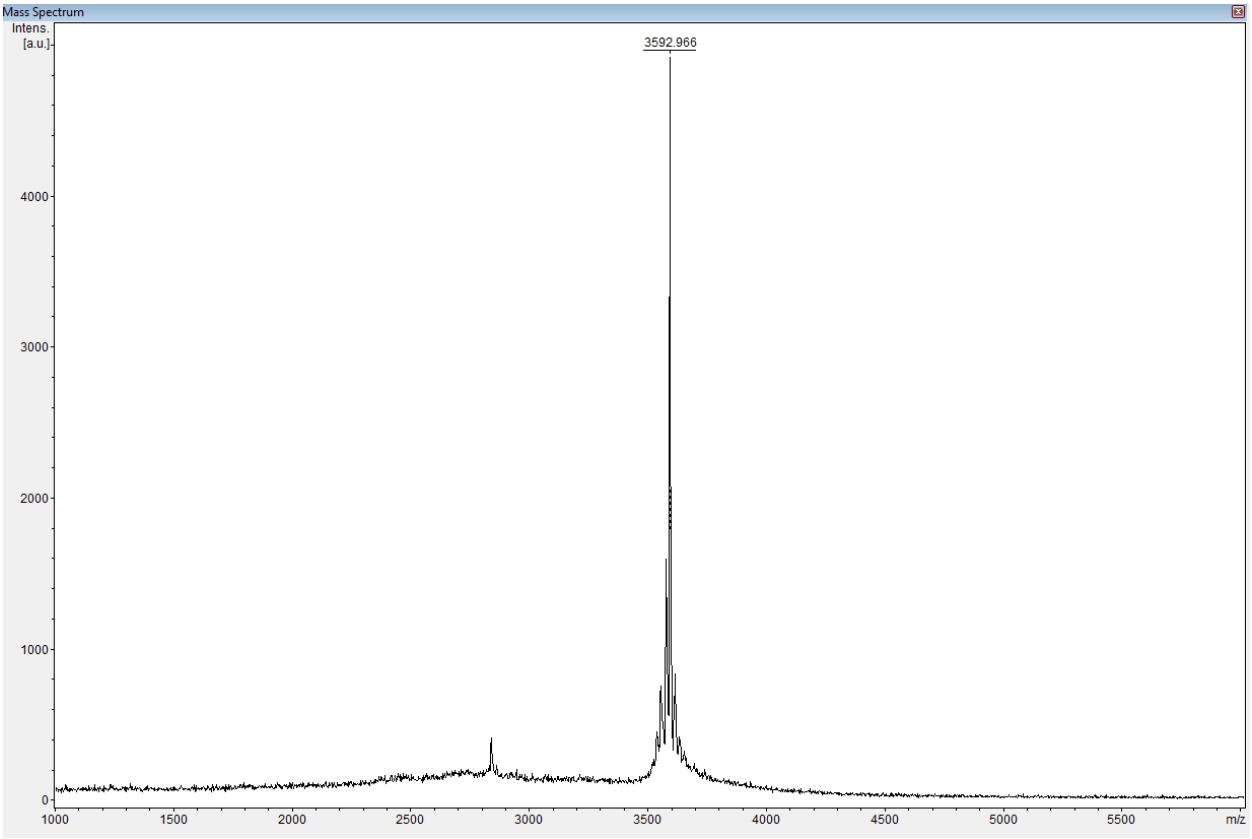


Figure 2.31: MALDI-TOF-MS of L-443-Variant. Observed mass: $[M+H]^+$ = 3593.0. Calculated mass: $[M+H]^+$ = 3593.0

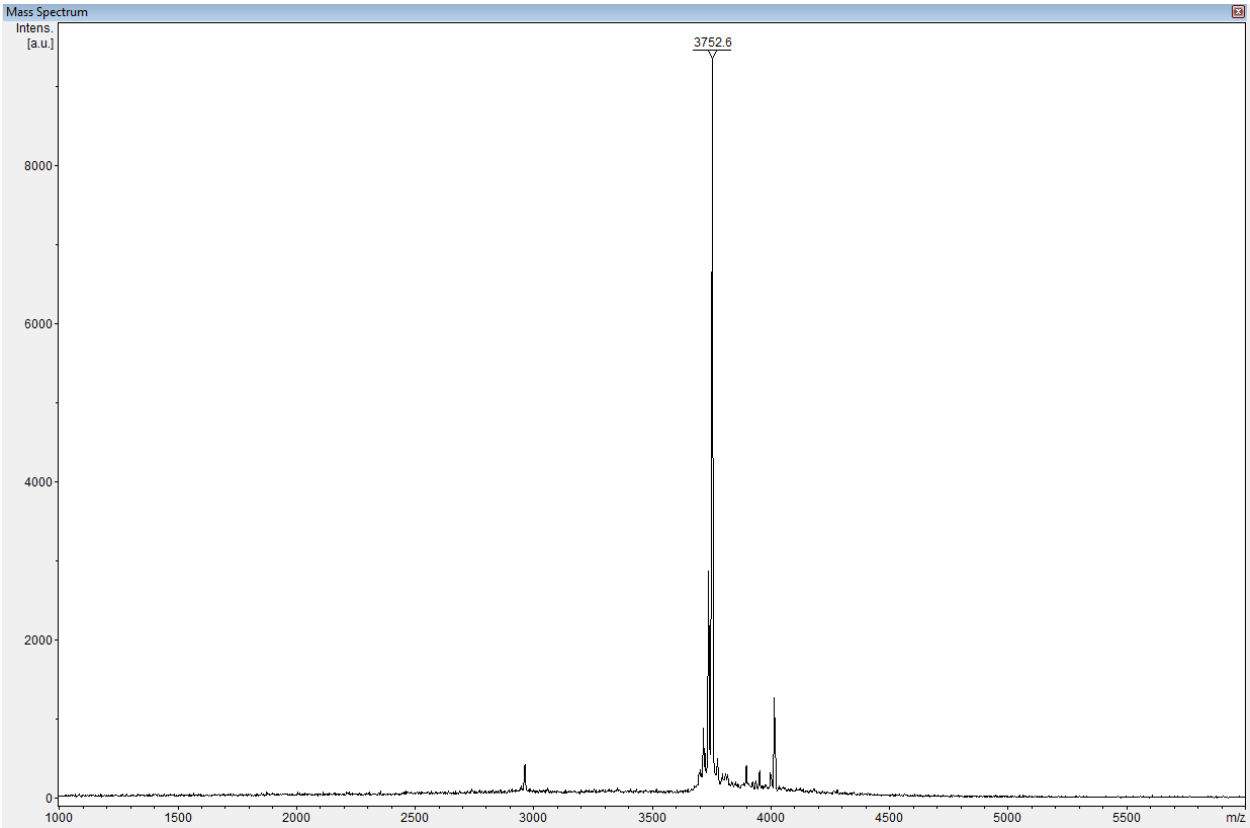


Figure 2.32: MALDI-TOF-MS of D-GCN4. Observed mass: $[M+H]^+$ = 3752.6. Calculated mass: $[M+H]^+$ = 3751.0

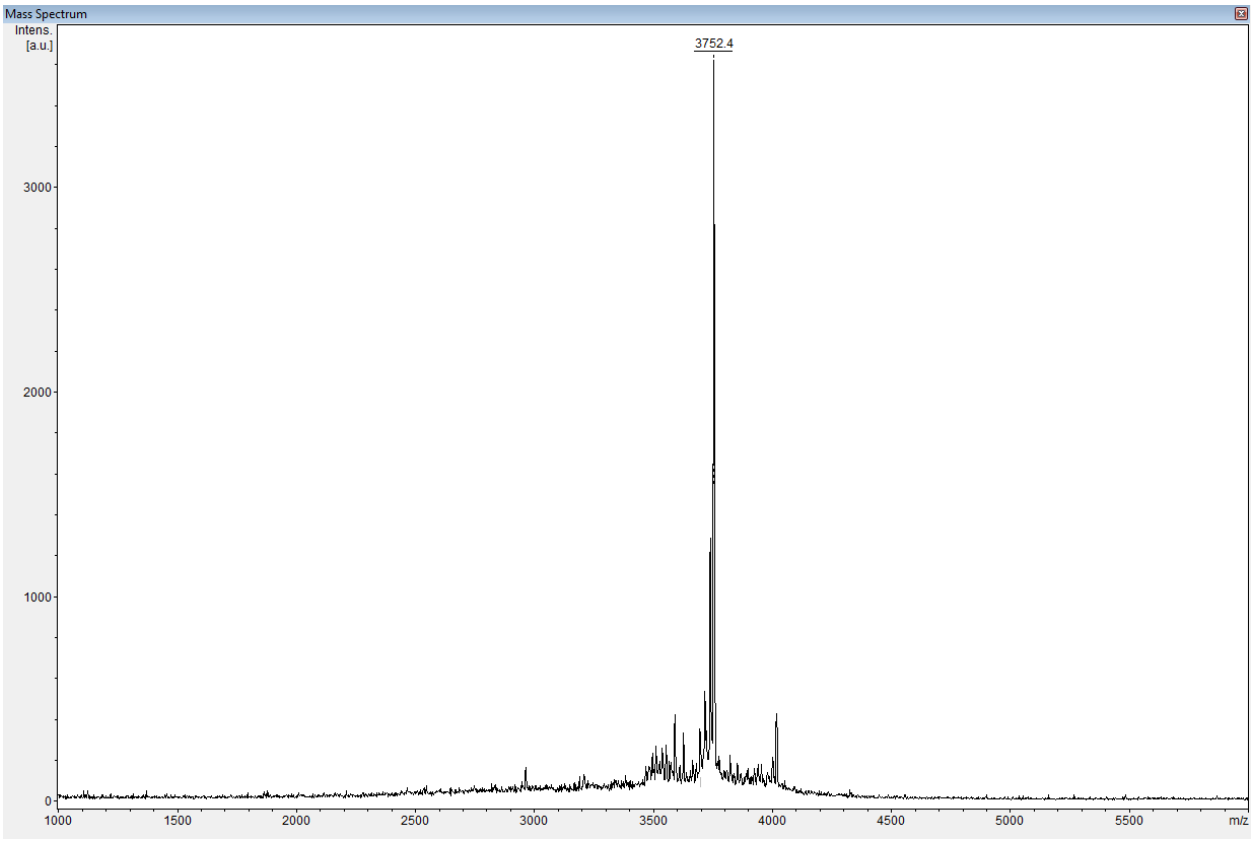


Figure 2.33: MALDI-TOF-MS of L-GCN4. Observed mass: $[M+H]^+$ = 3752.4. Calculated mass: $[M+H]^+$ = 3751.0

2.6 – References

1. McFarlane, A.; Orriss, G.; Stetefeld, J., The Use of Coiled-Coil Proteins in Drug Delivery, *European Journal of Pharmacology*, **2009**, *625*, 101-107
2. O'Shea, E.; Klemm, J.; Alber, T., X-Ray Crystal Structure of the GCN4 Leucine Zipper, a Two-Stranded, Parallel Coiled Coil, *Science*, **1991**, *254*, 539-544
3. Ellenberger, T.; Brandi, C.; Struhl, K.; Harrison, S., The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted α-Helices: Crystal Structure of Protein-DNA Complex, *Cell*, **1992**, *71*, 1223-1237
4. Hope, I.; Struhl, K., GCN4, a Eukaryotic Transcriptional Activator Protein, Binds as a Dimer to Target DNA, *The EMBO Journal*, **1987**, *6*, 2781-2784
5. Schnell, J.; Chou, J., Structure and Mechanism of the M2 Proton Channel of Influenza A Virus, *Nature*, **2008**, *451*, 591-596
6. Wang, J., Qiu, J.X., Soto, C. and DeGrado, W.F., 2011. Structural and dynamic mechanisms for the function and inhibition of the M2 proton channel from influenza A virus. *Current opinion in structural biology*, *21*(1), pp.68-80.
7. Hong, M. and DeGrado, W.F., 2012. Structural basis for proton conduction and inhibition by the influenza M2 protein. *Protein Science*, *21*(11), pp.1620-1633.
8. Moutevelis, E.; Wooflson, D., A Periodic Table of Coiled-Coil Protein Structures, *J. Mol. Biol.*, **2009**, *385*, 726-732
9. Boyle, A.L. and Woolfson, D.N., 2011. De novo designed peptides for biological applications. *Chemical Society Reviews*, *40*(8), pp.4295-4306.

10. Woolfson, D.N., 2017. Coiled-coil design: updated and upgraded. *Fibrous Proteins: Structures and Mechanisms*, pp.35-61.
11. Dawson, W.; Martin, F.; Rhys, G.; Shelley, K.; Brady, L.; Woolfson, D., Coiled Coils 9-to-5 *De Novo* Design of α-Helical Barrels with Tunable Oligomeric States, *Chemical Science*, **2021**, *12*, 6923-6928
12. Dawson, W.; Lang, E.; Rhys, G.; Shelley, K.; Williams, C.; Brady, R.; Crump, M.; Mulholland, A.; Woolfson, D., Structural Resolution of Switchable States of a *De Novo* Peptide Assembly, *Nature Communications*, **2021**, *12*, 1530
13. McFarlane, A.A., Orriss, G.L. and Stetefeld, J., 2009. The use of coiled-coil proteins in drug delivery systems. *European journal of pharmacology*, *625*(1-3), pp.101-107.
14. Harrington, L.; Fletcher, J.; Heermann, T.; Woolfson, D.; and Schwille, P., *De Novo* Design of Reversible Phosphorylation-Dependent Switch for Membrane Targeting, *Nature Communications*, **2021**, *12*, 1472
15. Vagner, J.; Qu, H.; Hruby, V., Peptidomimetics, a Synthetic Tool of Drug Discovery, *Current Opinion in Chemical Biology*, **2008**, *12*, 292-296
16. Hossain, M.; Bathgate, R., Challenges in the Design of Insulin and Relaxin/Insulin-Like Peptide Mimetics, *Bioorganic and Medicinal Chemistry*, ***2018***, *26*, 2827-2841
17. Mabonga, L.; Kappo, A., Peptidomimetics: A Synthetic Tool for Inhibiting Protein-Protein Interactions in Cancer, *International Journal of Peptide Research and Therapeutics*, **2020**, *26*, 225-241
18. Feng, Z.; Xu, B., Inspiration from the Mirror: D-Amino Acid Containing Peptides in Biomedical Approaches, *Biomol Concepts*, **2016**, *7(3)*, 179-187

19. Hamamoto, K.; Kida, Y.; Zhang, Y., Shimizu, T.; Kuwano, K.; Antimicrobial Activity and Stability to Proteolysis of Small Linear Cationic Peptides with D-Amino Acid Substitutions, *Microbiol. Immunol.*, **2002**, *46(11)*, 741-749

20. Lu, J.; Xu, H.; Xia, J.; Ma, J.; Xu, J.; Li, Y.; Feng, J., D- and Unnatural Amino Acid Substituted Antimicrobial Peptides with Improved Proteolytic Resistance and Their Proteolytic Degradation Characteristics, *Frontiers in Microbiology*, **2020**, *11*, 1-17

21. Checco, J.; Kreitler, D.; Thomas, N.; Belair, D.; Rettko, N.; Murphy, W.; Forest, K.; Gellman, S., Targeting Diverse Protein-Protein Interaction Interfaces with α/β-Peptides Derived from the Z-Domain Scaffold, *PNAS*, **2015**, *122*, 4552-4557

22. Cabrele, C.; Martinek, T.; Reiser, O.; Berlicki, L., Peptides Containing β-Amino Acid Patterns: Challenges and Successes in Medicinal Chemistry, *J. Med. Chem.*, **2014**, *57*, 9718-9739

23. B. Welch, B. D.; VanDemark, A. P.; Heroux, A.; Hill, C. P.; Kay, M. S. Potent D-Peptide Inhibitors of HIV-1 Entry. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104 (43)*, 16828−16833.

24. C. Liu, M.; Li, C.; Pazgier, M.; Li, C.; Mao, Y.; Lv, Y.; Gu, B.; Wei, G.; Yuan, W.; Zhan, C.; Lu, W.-Y.; Lu, W. D-Peptide Inhibitors of the p53-MDM2 Interaction for Targeted Molecular Therapy of Malignant Neoplasms, *Proc. Natl. Acad. Sci. U. S. A.*, **2010**, *107 (32)*, 14321−14326.

25. D. Van Groen, T.; Wiesehan, K.; Funke, S. A.; Kadish, I.; NagelSteger, L.; Willbold, D. Reduction of Alzheimer's Disease Amyloid Plaque Load in Transgenic Mice by D3, a D-Enantiomeric Peptide Identified by Mirror Image Phage Display, *ChemMedChem*, **2008**, *3 (12)*, 1848−1852.

26. E. Schumacher, T. N. M.; Mayr, L. M.; Minor, D. L., Jr.; Milhollen, M. A.; Burgess, M. W.; Kim, P. S. Identification of D-Peptide Ligands Through Mirror-Image Phage Display, *Science*, **1996**, *271*, 1854−1857.

27. F. Chang, H. N.; Liu, B. Y.; Qi, Y. K.; Zhou, Y.; Chen, Y. P.; Pan, K. M.; Li, W. W.; Zhou, X. M.; Ma, W. W.; Fu, C. Y.; Qi, Y. M.; Liu, L.; Gao, Y. F. Blocking of the PD-1/PD-L1 Interaction by a D-Peptide Antagonist for Cancer Immunotherapy, *Angew. Chem., Int. Ed.*, **2015**, **54 (40)**, 11760−11764.

28. G. Mandal, K.; Uppalapati, M.; Ault-Riche, D.; Kenney, J.; Lowitz, J.; Sidhu, S. S.; Kent, S. B. H., Chemical Synthesis and X-Ray Structure of a Heterochiral {D-Protein Antagonist plus Vascular Endothelial Growth Factor} Protein Complex by Racemic Crystallography, *Proc. Natl. Acad. Sci. U. S. A.*, **2012**, *109 (37)*, 14779−14784.

29. Kitaigorodskii, A.I. Organic Chemical Crystallography (Consultants Bureau: New York, 1961) (Originally published in Russian by Press of the Academy of Sciences of the USSR, Moscow, 1955).

30. Wukovitz. S.; Yeates, T., Why Proteins Favour Some Space Groups Over Others, *Nature Structural Biology*, **1995**, *2*, 1062-1067

31. Yeates, T., Kent. S.; Racemic Protein Crystallography, *Annu. Rev. Biophys.*, **2012**, *41*, 41-61

32. Kreitler, D.; Yao, Z.; Steinkruger, J.; Mortenson, D.; Huang, L. Mittal, R.; Travis, B.; Forest, K.; Gellman S., A Hendecad Motif is Preferred for Heterochiral Coiled Coils, *J. Am. Chem. Soc.*, **2019**, *141*, 1583−1592

33. Pan, M.; Gao, S.; Zheng, Y.; Tan, X.; Lan, H.; Tan, X.; Sun, D.; Lu, L.; Wang, T.; Zheng, Q.; Huang, Y.; Wang, J., and Liu L., Monomer/Oligomer Quasi-Racemic Protein Crystallography, *J. Am. Chem. Soc.*, **2016**, *138*, 7429−7435

34. Mandal, K.; Pentelude, B.; Tereshko, V.; Thammavongsa, V.; Schneewind, O.; Kossiakoff, A.; Kent, S., Racemic Crystallography of Synthetic Protein Enantiomers Used to Determine the X-Ray Structure of Plectasin by Direct Methods, *Protein Science*, **2009**, *18*, 1146-1154

35. Mortenson, D.E.; Steinkruger, J.D.; Kreitler, D.F.; Perroni, D.V.; Sorenson, G.P.; Huang, L.; Mittal, R.; Yun ,H.G.; Travis, B.R.; Mahanthappa, M.K.; Forest, K.T.; Gellman, S.H., High-resolution structures of a heterochiral coiled-coil. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 13144.

36. Kurgan, K.W.; Kleman, A.F.; Bingman C.A.; Kreitler, D.F.; Weisblum, B.; Forest, K.T.; Gellman, S.H., Retention of Native Quaternary Structure in Racemic Melittin Crystals. *J. Am. Chem. Soc.*, **2019**, *141*, 7704-7708.

37. Harbury, P.; Kim, P.; Alber, T., Crystal Structure of an Isoleucine-Zipper Trimer, *Nature*, **1994**, *371*, 80-83

38. Harbury, P.; Zhang, T.; Kim, P.; Alber, T., A Switch Between Two-, Three-, and Four-Stranded Coiled Coils in GCN4 Leucine Zipper Mutants, *Science*, **1993**, *262*, 1401-1407

39. D'Oria, E.; Karamertzanis, P.; Price. S, Spontaneous Resolution of Enantiomers by Crystallization: Insights from Computed Crystal Energy Landscapes, *Crystal Growth and Design*, **2010**, *10*, 1749-1756

40. Zlenko, D.; Zanin, A.; Skoblin, A.; Tverdislov, V.; Stovbum, S., Spontaneous Resolution in Racemic Solutions of N-Trifluoroacetylated α-aminoalchols, *Journal of Molecular Structure*, **2019**, *1183*, 8-13

41. Deng, Y., Zheng, Q., Liu, J., Cheng, C.S., Kallenbach, N.R. and Lu, M., Self-assembly of coiled-coil tetramers in the 1.40 Å structure of a leucine-zipper mutant. *Protein science*, **2007**, *16*(2), 323-328.

42. Kumar, P.; Woolfson, D., Socket2: a Program for Locating, Visualizing and Analyzing Coiled-Coil Interfaces in Protein Structures, *Bioinformatics*, **2021**, *37*, 4575-4577

43. Walshaw, J.; Woolfson, D., SOCKET: a Program for Identifying and Analysing Coiled-Coil Motifs within Protein Structures, *J. Mol. Biol.*, **2001**, *307*, 1427-1450

44. Deng, Y., Liu, J., Zheng, Q., Eliezer, D., Kallenbach, N.R. and Lu, M., 2006. Antiparallel four-stranded coiled coil specified by a 3-3-1 hydrophobic heptad repeat. *Structure*, **2006**, *14*(2), 247-255.

Chapter III

Backbone Thioamide Modifications Within Antigenic Peptides to Increase their Immunogenic Profile

3.1 – Abstract

Peptide vaccines offer the most direct route to initiate an immunogenic response to a defined antigen epitope compared to other common methods of vaccine, such as mRNA or live-attenuated vaccines. Peptide vaccines do not require any additional processing once delivered to a patient to access an antigen epitope. However, peptide vaccines, and other peptide drugs, are unstable *in vivo*, as they are subjected to rapid enzymatic proteolysis. In this study, I will replace the native backbone oxoamides within the murine major histocompatibility complex (MHC) 1-resitricted antigen SIINFEKL with an isosteric thioamide moiety. Thioamides represent a minimalist modification (carbonyl O to S) to instill protection from proteolytic degradation. We hypothesized that replacement of the native backbone oxoamides would result in modified antigens with increased stability and functional mimicry of the native antigen. We discovered that the modification is generally well-tolerated and able to stimulate an immune response *in vitro* and can offer substantial protection from serum proteases. Finally, we immunized live mice with the thio-SIINFEKL antigens. We determined that the immunized CD8+ T-cells from the mice were cross-reactive with the native antigen and in some cases effective at suppressing tumor growth in OVA-expressing tumor-bearing mice.

3.2 – Introduction

Peptide vaccination offers the most direct route to initiate an immunogenic response to a defined antigen epitope. Other popular methods of vaccination, such as mRNA vaccines[1-2] or live-attenuated vaccines[3], require cellular processing to generate the target antigenic species. Major Histocompatibility Complex Class 1 (MHC1) restricted antigens are typically derived from cytosolic proteins and are 8 to 10 residue peptides once fully processed by cellular machinery. Peptides of this size are, generally, easily accessible using solid phase peptide synthesis (SPPS) providing the peptide sequence of the antigen is known. However, peptides are unstable *in vivo* as they are subject to rapid enzymatic proteolysis[4-6]. This degradation issue has long hampered the implementation of peptides as any form of therapeutic. Although new peptide drugs, such as the Glucagon Like Peptide 1 analogs Semaglutide[7,8] and Tirzepatide[9-11], have recently seen widespread clinical application.

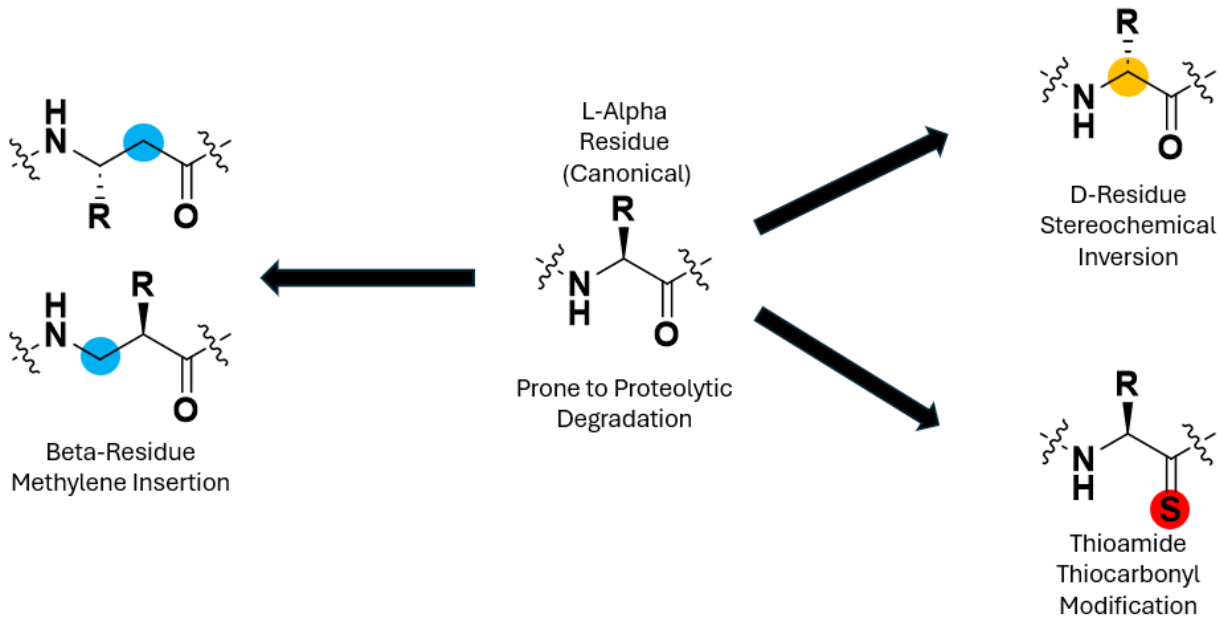


Figure 3.1: Structure of a canonical L-α-residue compared with the classic peptidomimetic modifications of the β-residue or D-α-residue. The newer thioamide modification is shown on the bottom right.

Over the last decades many forms of peptidomimetics have emerged as approaches to circumvent the degradation issue and increase therapeutic viability: such as β-amino acids[12-17] and D-amino acids[18-22] (**Figure 3.1**). Both the approaches indicated are known to increase proteolytic stability for the canonical amides located near the site of the peptidomimetic modification incorporation. However, subtle changes to an antigen's sequence can induce structural perturbations that can result in reduced MHC1 affinity[23-25]. These structural perturbations are present in any peptide with a peptidomimetic modification included. However, the perturbations are more prominent within MHC1 restricted antigens as the peptide sequences are only 8 to 10 residues. An individual modifications will occupy more of the peptide's sequence, making the modifications more pronounced than compared to longer peptides.

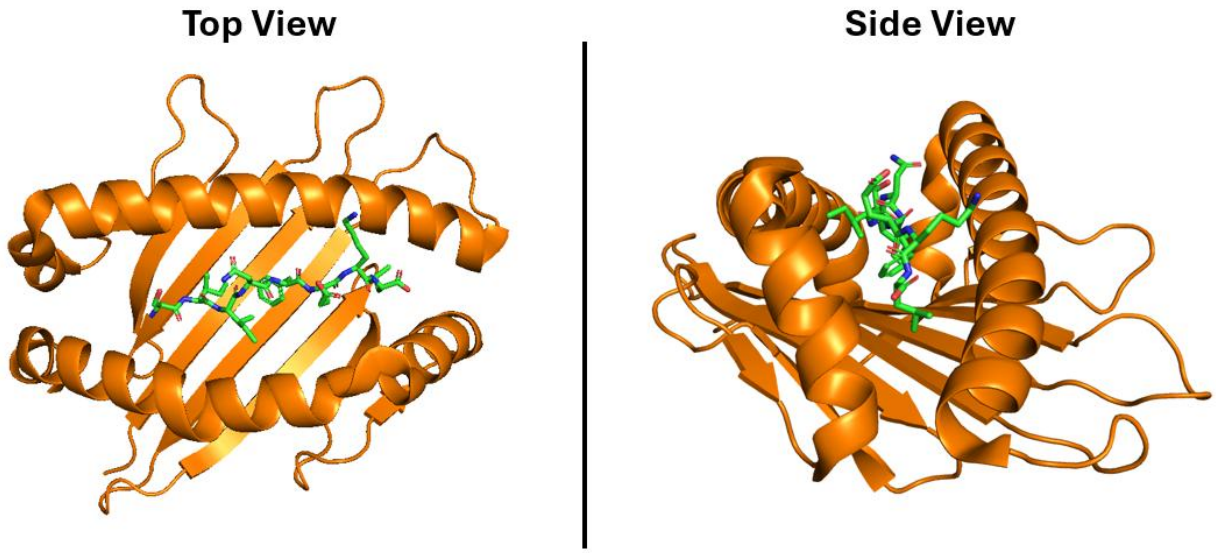


Figure 3.2: Structure of a SIINFEKL bound to a H-2K^b (PDB: 3P9L). The antigen is bound to the MHC binding site in an extended conformation. Peptide shown in green, MHC shown in orange.

To examine the impact of peptidomimetic approaches on antigen-MHC binding, I will discuss a structure of the H-2K^b bound to the SIINFEKL antigen (**Figure 3.2**). Due to the short size of Class 1-restricted antigens, there typically is no detectable secondary structure of the peptide in solution. Upon binding to the antigen's cognate MHC, the peptide adopts an extended conformation that determines the specificity of the binding interactions. Once the antigen is bound to an MHC within the endoplasmic reticulum, the MHC-antigen complex is shuttled to the cell surface where it can display the antigen to immunoregulatory cells[27,28]. When modifications such as replacing an α-amino acids to β-amino acids are implemented, the backbone becomes elongated which displaces the antigenic peptide's sidechains and hinders efficient binding of the antigen to the MHC[14,29].

For peptide vaccination to be a viable approach for immunization, there must be a peptidomimetic modification that minimally impacts the binding conformation of the backbone and orientation of the sidechains, to provide proteolytic resistance without loss of affinity or specificity for the antigen to the cognate receptor. The thioamide moiety is a minimalistic isosteric modification for the canonical backbone oxoamide that has been reported to increase proteolytic stability[30-37]. Since the thioamide modification is known to increase proteolytic stability and does not directly impact the backbone length or sidechain orientation, Dr. Rylie Morris (Gellman Lab) and Dr. Ruslan Gibadullin (Gellman Group) designed and synthesized a series of thioamide modified antigen (thio-antigens) derivates of two human leukocyte antigen-A2 antigens (HLA-A2): GILGFVFTL (GIL) and ELAGIGILTV (ELA)[38]. Both antigens had been previously evaluated *in vitro* or in a clinical setting for the potential of becoming peptide vaccines[39-43]. Most of the thio-antigens of GIL and ELA

displayed the capability to bind to HLA-A2 and stimulate T cells comparably to their native antigens *in vitro*. Additionally, prolonged stability for GIL and ELA thio-antigens was observed when exposed to an aggressive non-specific protease, Proteinase K[44]. The *in vitro* results of the GIL and ELA thio-antigens to mimic the function their native sequences and display enhanced proteolytic stability is encouraging for the application of thio-antigens in peptide vaccines.

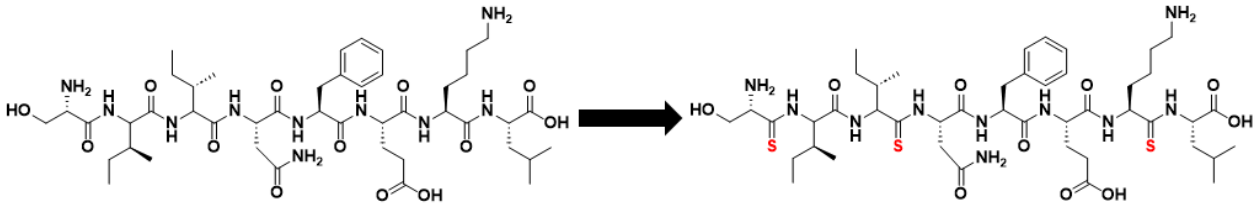


Figure 3.3: Implementation of thioamide modifications within a well-studied antigenic peptide, mouse OVA. The native backbone will be decorated with thioamide modifications to enhance the antigens proteolytic stability. Positions of sulfur modification are indicated with the thiocarbonyl highlighted in red.

To interrogate the viability of implementing thioamide modifications within live mice, the well-studied mouse MHC Class 1-restricted OVA antigen[29,45-49] ,SIINFEKL, was chosen as a "proof of principle" model system to implement the approach *in vivo*. Our group has previously attempted to incorporate Beta-3 residues within the SIINFEKL epitope, which will be mentioned in comparison to the thioamide work as we continue into this section[17].

This work is the result of an ongoing collaboration between the Gellman group (University of Wisconsin-Madison, Chemistry Department), the McNeel Group (University of Wisconsin-Madison, Carbone Cancer Center), and the Chatterjee Group (Indian Institute of Science, Molecular Biophysics Unit). Work that has been conducted outside of the Gellman Group well be indicated where applicable.

3.3 – Synthesis of Thio-SIINFEKL Derivatives

3.3.1 – Synthesis Thioamide Monomers for SPPS

Synthetic approaches to convert standard Fmoc-protected amino acids (Fmoc-AAs) used for SPPS into their corresponding thioamide producing counterparts have been introduced over the last decades[18,19]. I will summarize the approach below, see **Methods and Characterization** for specific synthetic details:

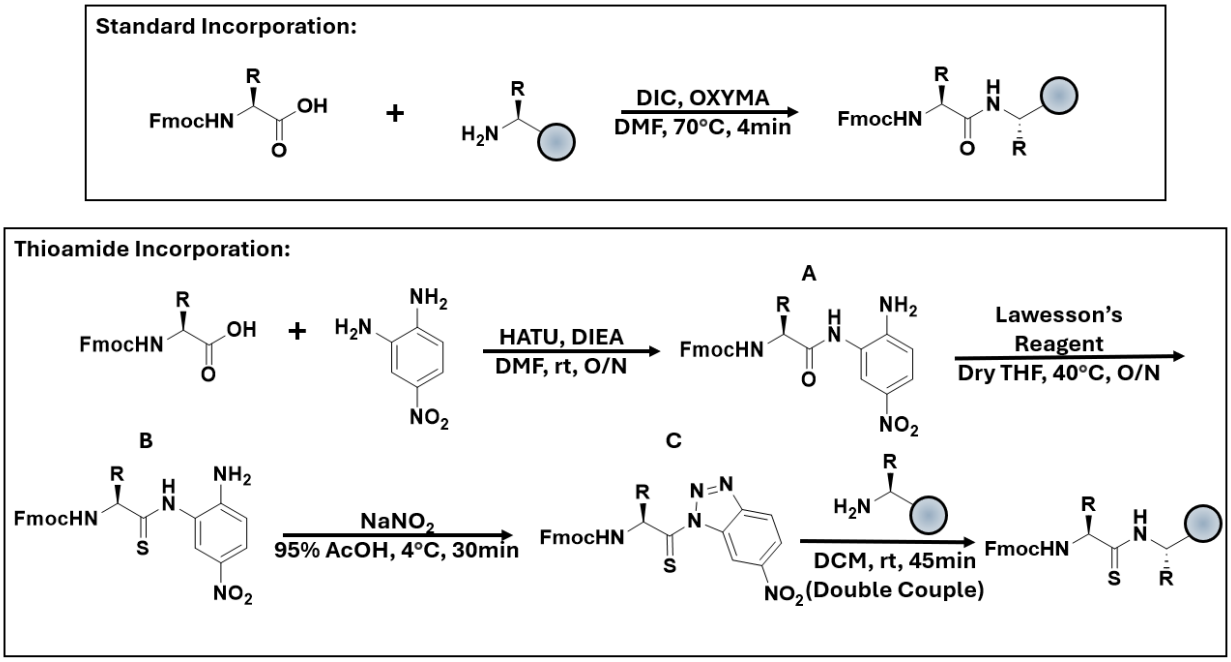


Figure 3.4: In the top panel a conventional coupling reaction for the incorporation of an Fmoc-protected amino acid onto solid support is depicted. DIC and OXYMA are used as the activator and activator base, respectively. In the bottom panel a general synthetic route to transform an Fmoc-protected amino acid into an Fmoc-protected thioacylating is depicted along with its incorporation into solid support.

To convert an Fmoc-AA into a corresponding Fmoc-protected thioacylating reagent (Compound **C**) a three-step chemical process is required (**Figure 3.4**). Firstly, the Fmoc-AA is coupled to 4-nitro-*o*-phenylenediamine (4NoPDA) by pre-activating the Fmoc-AA with O-(7-azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HATU) and diisopropylethylamine (DIEA) in N,N-dimethylformamide (DMF), and then 4NoPDA is added.

After the reaction has been completed, an aqueous workup will be conducted (see **Section 3.7.3**). To remove residual water Na_2SO_4 should be used as the drying agent and not $MgSO_4$. $MgSO_4$ will coordinate this thioacylating reagent (product of step three) and prevent efficient incorporation of the thioacylating reagent (compound **C**) onto solid support.

The crude product from the previous reaction (compound **A)** is then added to a reaction vessel with Lawesson's reagent (LR), the vessel is then sealed and purged of air under a stream of N_2. Dry tetrahydrofuran (THF) is added to dissolve crude compound **A** and LR. LR readily hydrolyzes so inert atmospheres are critical for oxoamide-to-thioamide transformations. Flash chromatography is implemented to purify compound **B** after this step. See **Section 3.3.2** for more information about LR specifications.

Compound **B** was activated by introducing $NaNO_2$ under acidic conditions to form a benzotriazole containing species. Once the reaction was completed, the thioacylating reagent (compound **C**) was removed from the acidic solution and prepared for solid-phase incorporation. Immediately prior to incorporation on to solid support the thioacylating reagent (compound **C**) was dissolved in DCM, and the solution was added to the solid support. Many literature reports of thioacylation on solid support incorporate 0.5eq of DIEA[50,51]. We and others have observed that for certain positions within peptide sequences, the incorporation of the thioamide on solid support can be completely thwarted by the presence of DIEA[52]. We choose to omit DIEA in all thioacylation reactions for this reason.

3.3.2 – Variability in Thionation Efficiency with Lawesson's Reagent

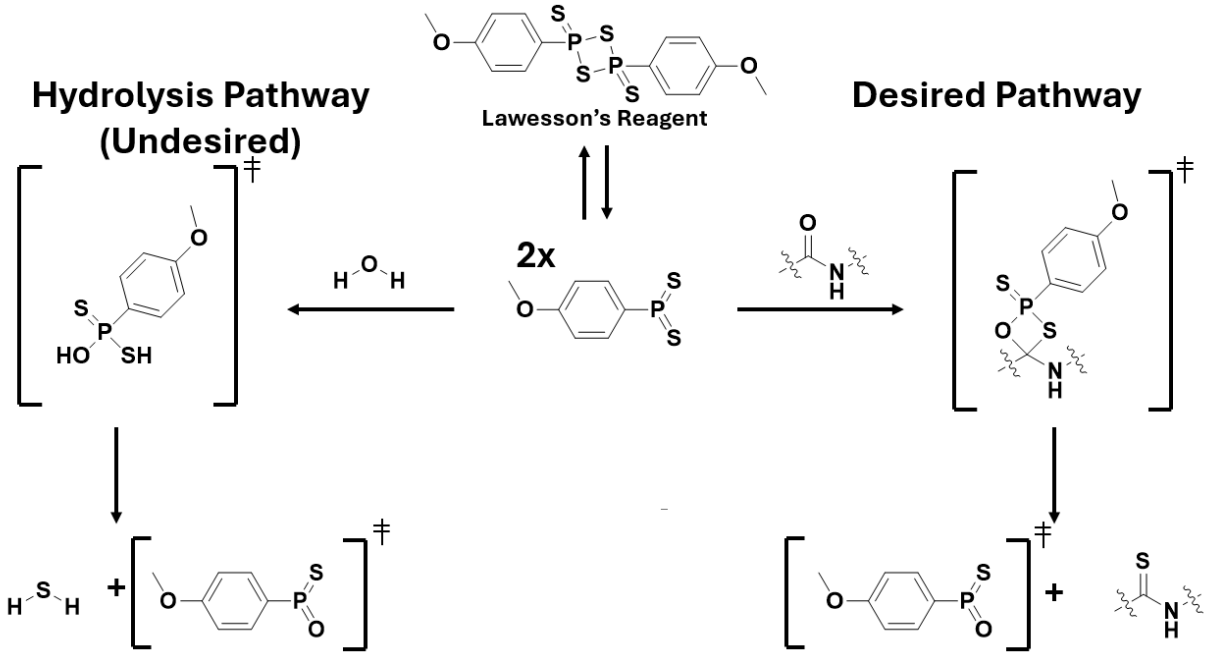


Figure 3.5: Competing pathways of thioamide formation and hydrolysis of Lawesson's reagent (LR). Exclusion of water is critical to facilitating high yielding thionation reactions.

Before continuing further into synthetic approaches, I will discuss the temperamental nature of oxoamide-to-thioamide conversions and our methods to tame them. LR is the most widely used thionating reagent and as LR readily facilities carbonyl O to S conversion with a wide range of carbonyl functional groups: amides, esters, ketones, and aldehydes[53-55]. Although very efficient for facilitating thionation reactions, particularly with amide to thioamide transformations, LR is highly prone to hydrolysis (**Figure 3.5**). The hydrolysis of LR leads to a competing pathway when conducting a thionation reaction and limits the stability of the thionating reagent when stored. Stability can be enhanced by storing LR under cold (4°C or -20°C) and dry (inside a desiccator or with moisture trapping packets) conditions, but after three or more months after opening the reagent bottle, a reduction in yield will begin to be observed. In some cases, I have observed no reactivity using LR from bottles less than six

months after purchasing. I also observed inconsistencies from suppliers; a batch of LR reagent from a given supplier could work well for months, while another batch degraded rapidly. The poor shelf stability and inconsistency from suppliers pushed us to find alternative routes to obtain high quality LR.

I began synthesizing LR in-house by coupling phosphorus pentasulfide with anisole via an electrophilic aromatic substitution reaction[56]. This resulted in the best yields and reproducibility between batches for amide to thioamide conversions.

3.3.3 – Liberation of Thio-Peptides from Solid Support

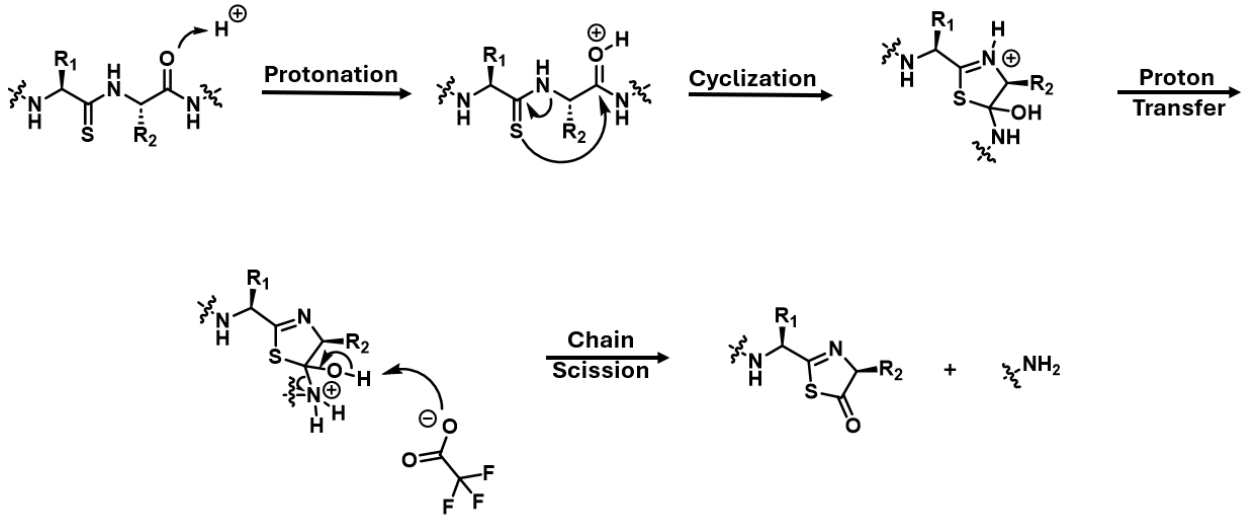


Figure 3.6: Mechanism of the cyclization that causes degradation of a thioamide containing peptide under acidic conditions.

Incorporation of thioamides onto solid support, when utilizing the procedure indicated in **Section 3.3.1** and described more fully in **Methods and Characterization**, is a relatively straightforward and efficient process. A long-standing problem in generating thioamide containing peptides (thio-peptides) is the degradation that occurs under acidic conditions. The degradation results from the nucleophilic nature of sulfur compared to

oxygen. The sulfur undergoes nucleophilic addition to the *i+1* carbonyl carbon under acidic conditions (**Figure 3.5**). Since many SPPS procedures rely on strong acid, typically trifluoroacetic acid (TFA), to liberate the peptide from the solid support and remove side chain protecting groups it was inevitable that our thioamide moiety would be exposed to strong acid and undergo degradation, leading to low yields. In the best cases, we would observe very low isolated yield (<1%) after purification. Often syntheses that afforded a small amount of product would not be reproducible when tried another day. I required an alternative approach to prevent the degradation of thio-peptides

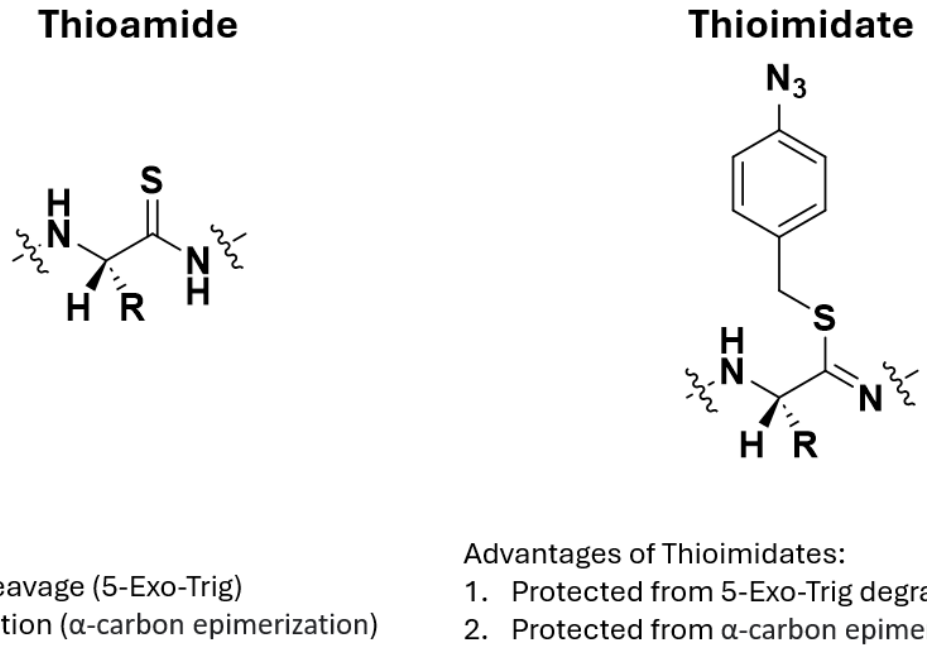


Figure 3.7: (left) Depiction of a thioamide incorporated within a peptide and the common routes of degradation observed during SPPS and resin cleavage. (right) Depiction of a para-azido-thioimidate protected thioamide and the stated chemical protections the thioimidate provides during SPPS and resin cleavage.

The thioamide's degradation propensity under acidic conditions necessitated modified cleavage conditions or a new protecting group to prevent 5-exo-trig degradation during peptide cleavage. Byerly-Duke et al. had reported that the conversion of a thioamide to a thioimidate in-solution and on solid support attenuates the nucleophilicity of the sulfur

under acidic conditions[57,58]. This attenuation of nucleophilicity blocks degradation. After cleavage, the thioimidate could be quantitatively removed under reducing conditions. Thioimidates thus offered a method to efficiently incorporate backbone thioamides during SPPS to make thio-peptides.

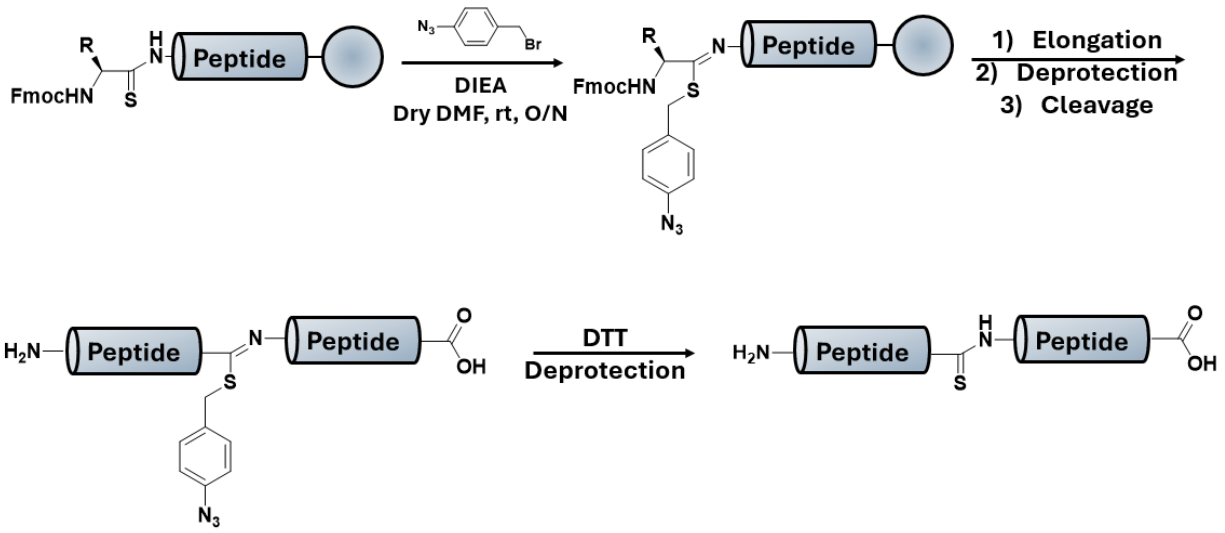


Figure 3.8: Depiction of thioimidate protection of a thioamide incorporated onto solid support and subsequent elongation, cleavage, and thioamide regeneration. Once a thioamide is incorporated onto solid support (see **Figure 3.4**) the thioamide is protected by coupling of a 4-azido-methylbromobenzene to the sulfur of the backbone thioamide. Once protected the peptide is elongated till the desired sequence is reached, followed by N-terminal Fmoc-deprotection. The peptide is then cleaved with a cocktail of 75%TFA, 23%TIPS, and 2% triisopropylsilane. Post-cleavage and removal of excess TFA the thioamide is regenerated by reduction of the thioimidate with dithiothreitol, furnishing the target thio-peptide.

Thioimidate protecting groups are easily incorporated into the general SPPS workflow (**Figure 3.8**), albeit lengthening the process by roughly 24 hours per thioimidate unit. 4-azido-methylbromobenzene is prepared in two steps from *p*-toluidine: amine to azide via a diazonium salt intermediate, followed by benzylic bromination (see **Methods and Characterization**)[26]. In addition to blocking degradation, the thioimidate prevents α-carbon epimerization during Fmoc-deprotection[59]. I suggest thioimidate protecting any thioamide

before continuing peptide synthesis to prevent epimerization during basic Fmoc-deprotection reactions during SPPS (**Figure 3.7**).

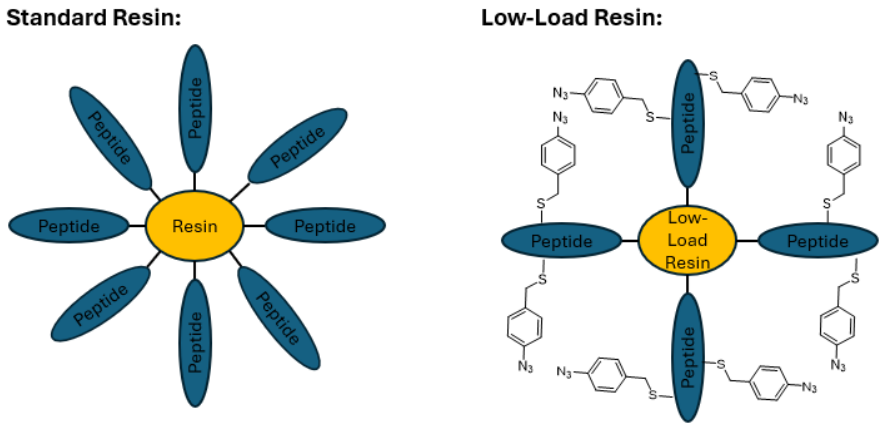


Figure 3.9: Cartoon illustration of standard resin versus Low-Load Wang resin when thioimidate protecting groups are installed onto a growing peptide chain. Thioimidate protecting group is depicted as a 4-azido- benzene group, which is attached via the thioamide sulfur, perpendicular to the growth axis of a given peptide chain.

Use of the thioimidate protecting group enhanced the reproducibility of the thio-peptide syntheses. However, the isolated yields of each thio-peptide synthesis were not conducive for live mouse model studies (see **Section 3.6** for results of the *in vivo* studies). I observed that yields of thio-peptides increased when using Low Load (LL) Fmoc-Leu Wang Resin (0.3-0.45 mmol/g) as opposed to standard loading Fmoc-Leu Wang Resin (0.5-0.7 mmol/g). LL resins are often employed for syntheses of long and/or hydrophobic peptides as those molecules tend to aggregate on the surface of the solid support, which prevents the N-terminus of growing peptide chains from accessing the later additions of peptide coupling solution. The additional spacing between growing peptides on LL resin helps to prevent aggregation. I hypothesized that instead of sequestering the N-terminus from elongation that standard loading resin hindered access of the 4-azido-methylbromo benzene to the thioamide. LL resin would provide additional space between growing peptide chains to allow the 4-azido-methylbromo benzene better access to the thioamide on solid support. After

incorporating LL resin in the syntheses, ~4-fold increase in isolated yield of thio-peptides was observed.

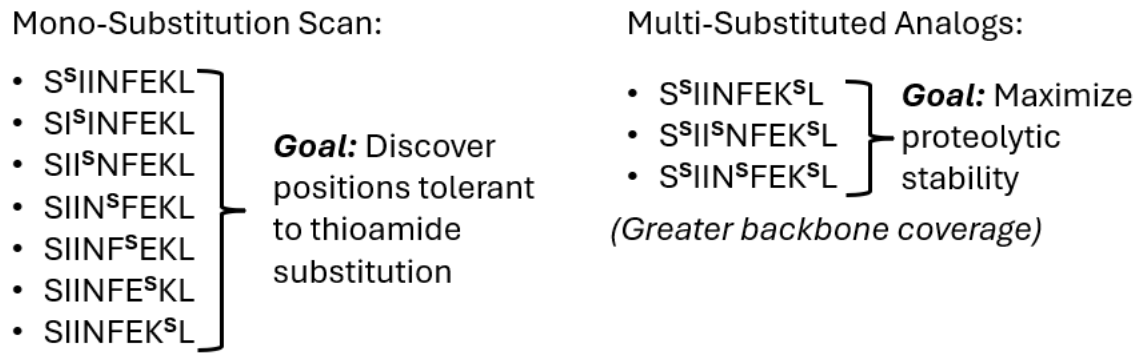


Figure 3.10: List of thio-SIINFEKL peptides examined for *in vitro* immunogenic response

With a general method established for generating thio-peptides, we synthesized a series of thio-SIINFEKL derivatives. I divided the derivatives into two groups: mono-substitution scan and multi-substituted analogs. The mono-substitution scan was designed to identify positions at which the SIINFEKL antigen could be modified without loss of immunological activity. The multi-substituted analogs were designed to provide the maximum backbone protection against proteolytic degradation.

3.4 – *In Vitro* Stimulation of an Immune Response with Thio-SIINFEKL Derivatives

3.4.1 – Binding of Thio-SIINFEKL Derivatives to the MHC

All experiments described in **Section 3.4.1** were performed by Dr. Ichwaku Rastogi (McNeel Group).

Thio-SIINFEKL derivatives were incubated with MHCs at room temperature to allow for binding to occur. Once the incubation period had concluded the MHCs were washed to

remove unbound peptide and anti-H-2K^b (an antibody selective for the H-2K^b-SIINFEKL complex) was added.

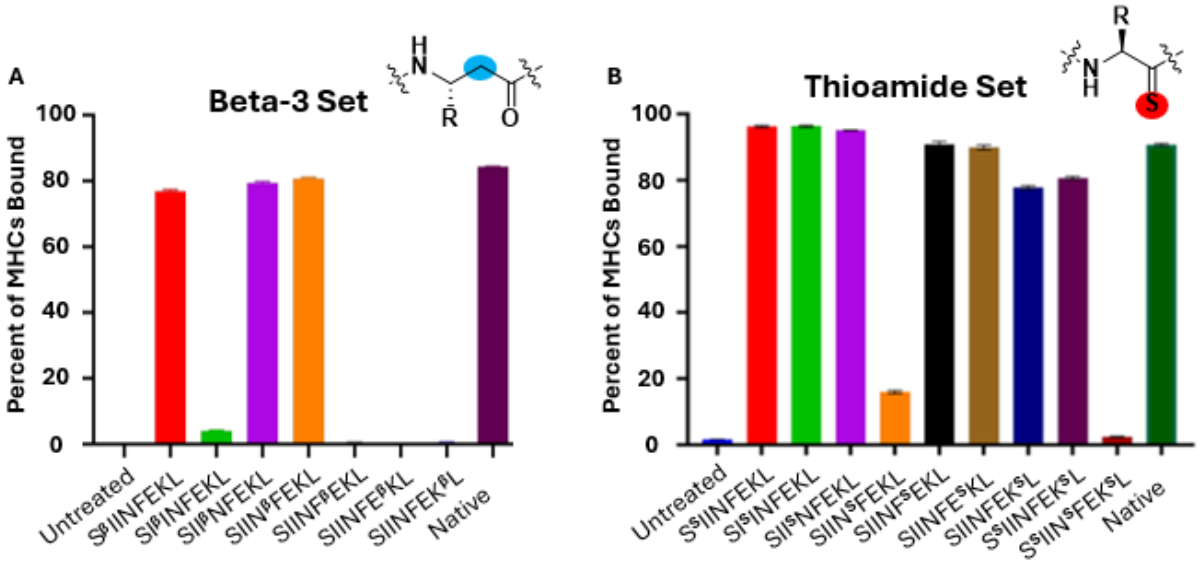


Figure 3.11: Percent of (**A**) beta-3 or (**B**) thioamide modified SIINFEKL antigens bound to mouse class I MHC molecule H-2k^b. SIINFEKL peptides were co-incubated with purified H-2K^b MHC. Quantification of percent MHC bound was detected by antibody detection using anti-H-2K^b. Superscript β or S within a peptide sequence indicates the residue where the Beta-3 or thioamide modification has been incorporated, respectively. Experiments and results shown in this figure were performed by Dr. Ichwaku Rastogi (McNeel Group).

To provide a context for the results obtained with the thio-antigens, I will first discuss the MHC binding of beta-3 modified SIINFEKL analogs to H-2k^b [29]. As discussed in **Section 3.2**, beta residue substitutions introduce an additional methylene unit into the peptide backbone. We expected that due to the short length and extended conformation adopted upon binding of the SIINFEKL antigen, beta residues would be poorly tolerated[14,29]. **Figure 3.11A** shows that four of the seven beta-3 substituted derivatives lose essentially all capability to bind to the MHC. The Ser1, Ile3, and Asn4 substitutions maintain MHC binding.

The thioamide series do not display the general loss of binding capability that is observed with the Beta-3 series. Most thioamide peptides display binding capacities that are

on par with the native peptide (**Figure 3.11B**). The oxoamide-to-thioamide substitutions within the thio-SIINFEKL derivatives are generally well-tolerated, except when placed at Asn4. Thioamides, compared to oxoamides, display stronger hydrogen bond donating and weaker hydrogen bond accepting capabilities[60]. In the crystal structure of SIINFEKL bound to H-2K^b (PDB: 3P9L), the backbone amide of Asn4 within the antigen accepts a hydrogen bond from the sidechain of Arg155 within the MHC binding groove. We suspect that the observed decrease in MHC binding for the thio-peptides containing a thioamide at position four comes from the thiocarbonyl's worse ability to accept hydrogen bonds compared to the native oxoamide.

The di-substituted thio-antigen retains affinity for the MHC, perhaps because each of the substitutions, at Ser1 and Lys7, are individually well tolerated. The tri-substituted thio-antigen does not bind well, a result that we attribute the substitution at Asn4, which was deleterious on its own.

3.4.2 – T Cell Stimulation with Presented Thio-SIINFEKL Antigens

All experiments described in **Section 3.4.2** were performed by Dr. Ichwaku Rastogi (McNeel Group).

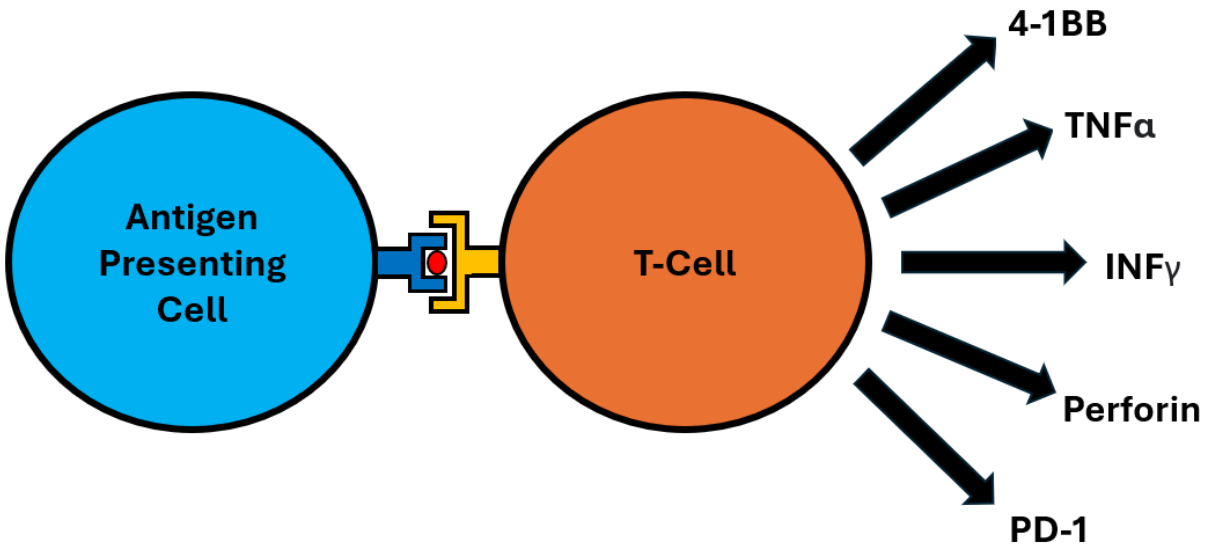


Figure 3.12: Cartoon model of an antigen presenting cell with an cell surface MHC (blue) presenting an antigen (red) to a T-cell which binds the MHC-peptide complex via the T-Cell Receptor (yellow). Upon stimulation of the T-cell causes expression and/or release of multiple protein, matching those shown.

The data in **Section 3.4.1** suggests that most sites of oxoamide-to-thioamide substitution in the SIINFEKL antigen are tolerated without loss of MHC1 binding. However, MHC binding is only part of the process of an antigen mediated immune response. We next asked whether MHC-thio-antigen complexes could be recognized by T-cell receptors (**Figure 3.12**).

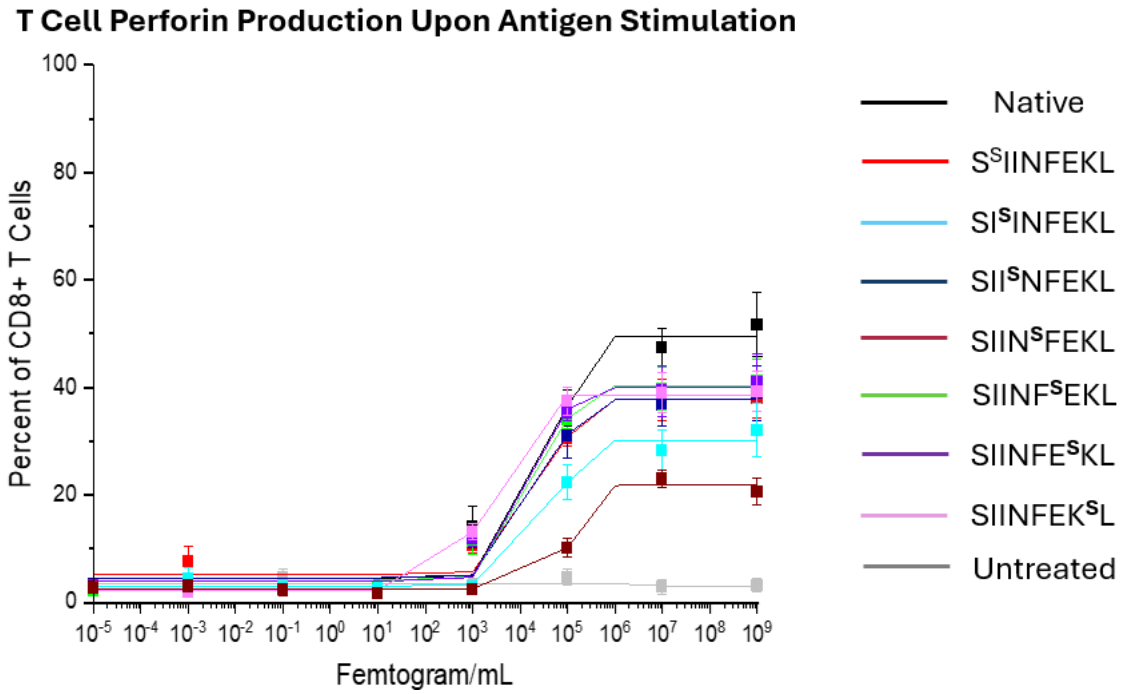


Figure 3.13: The percentage of immunized CD8+ T Cells out of total T-Cells that produce perforin, as a function of peptide concentration. In this series we report the mono-substituted peptides versus the native. Superscript S within a peptide sequence indicates the residue where a thioamide modification has been incorporated. Experiments and results shown in this figure were performed by Dr. Ichwaku Rastogi (McNeel Group).

Each of our mono-substituted antigens and the native peptide were used to stimulate OT-1 splenocytes (spleen cells that contain SIINFEKL-specific CD8+ T-Cells) at varying concentrations of peptide (10zg/ml to 1µg/mL). Utilizing flow cytometry, we quantified the number of T-Cells producing immune response associated proteins (4-1BB, TNFα, IFNγ, Perforin, GrzB, and PD-1). Results of the perforin secretion are shown in **Figure 3.13**. For the results of the other proteins monitored see **Methods and Characterization**.

Most mono-substituted SIINFEKL analogs stimulated the T-cells on par with the native. All thio-peptides displayed a slightly lower maximal response compared to the native. There were two outliers in the dataset, Ile2 and Asn4. The Asn4 thio-peptide displayed a lower maximal response and higher concentrations of peptide were required to observe perforin production by T cells. We attribute this to the weaker MHC binding of the Asn4 thio-

peptide, as shown in **Section 3.4.1**. The Ile2 thio-peptide, however, did not display reduced binding to the MHC. Ile2 has been reported as an anchoring residue (significant contacts with the MHC) along with Leu8. Anchoring residues have been reported as substantially intolerant of modification of any sort[61-62].

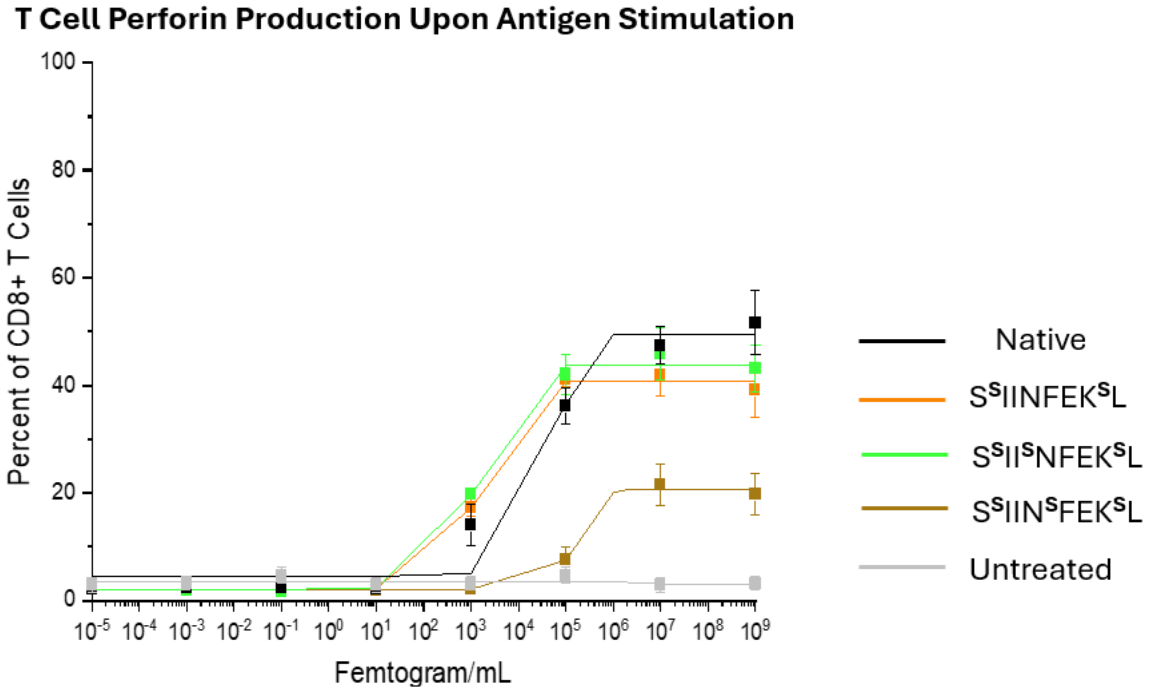


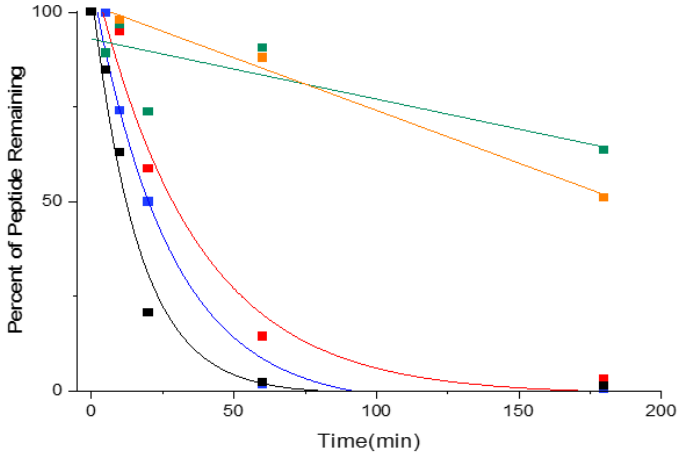
Figure 3.14: The percentage of immunized CD8+ T Cells out of total T-Cells that produce perforin, as a function of peptide concentration. In this series we report the multi-substituted peptides versus the native. Superscript S within a peptide sequence indicates the residue where a thioamide modification has been incorporated. Experiments and results shown in this figure were performed by Dr. Ichwaku Rastogi (McNeel Group).

Three multi-substituted SIINFEKL derivatives were evaluated for the ability to induce perforin secretion (**Figure 3.14**). The double-substituted analog, with modification at Ser1 and Lys7, and a triple-substituted thioamides at Ser1, Ile3, and Lys7 displayed maximal responses on par with most of the mono-substituted analogs of SIINFEKL (**Figure 3.13**). The potencies of these two thio-peptides were also greater than the native as less peptide was required to observe T-cell perforin production. The triple-substituted analog with thioamides placed at Ser1, Asn4, and Lys7, like the mono-substituted Asn4, displayed decreased

potency and lower maximal response. We again attribute this to the weaker MHC binding observed of SIINFEKL peptides containing an Asn4 modification.

3.4.3 – Proteolytic Stability of Thio-Peptides

With the *in vitro* activity of many of the SIINFEKL peptides observed to match that of the native epitope, I next sought to determine whether the thioamide incorporation imparted enhanced proteolytic stability relative to the native SIINFEKL antigen.



Peptide	$t_{1/2}$ (min)	$t_{1/2}$ (Rel)
Native	16	1
S^{S}IINFEKL	36	2.3
SII^{S}NFEKL	27	1.7
S^{S}IINFEK^{S}L	***	***
S^{S}II^{S}NFEK^{S}L	***	***

Figure 3.15: (left) plot of the remaining amount of each peptide at a given timepoint after incubation with 25% v/v mouse serum at 37°C. (right) Half-lives of each peptide tested in both minutes and relative to the native. *** indicates that the half-live of a peptide was not reached after 3 hours of incubation with mouse serum. This indicates a minimum 10-fold increase in stability towards the proteases present in mouse serum. Superscript "S" within a peptide sequence indicates the residue where a thioamide modification has been incorporated.

The half-life ($t_{1/2}$) of selected peptides was determined in the presence of mouse serum (**Figure 3.15**). The $t_{1/2}$ of the native SIINFEKL peptide was observed at 16 minutes in our assay. When a single thioamide was incorporated at either Ser1 or Ile3 the half-life of each peptide was roughly doubled. The $t_{1/2}$ value could not be determined with the di- or tri-substituted analogs, as after 3 hours over 50% of each peptide remained. This assay will need to be repeated with longer time points, but we can deduce that thioamides modifications when put at the N- and C-terminal amide increase the proteolytic stability of

the SIINFEKL peptide over 10-fold. This assay also supports that the predominant proteolytic susceptibility of the SIINFEKL peptide is likely from exopeptidases.

3.5 – *In Vivo* Stimulation of an Immune Response with Thio-SIINFEKL Derivates

All experiments described in **Section 3.5** were performed by Dr. Ichwaku Rastogi (McNeel Group). Synthesizing milligram quantities of the multi-substituted peptides has still been a challenge even with substantial synthetic protocol optimization. To assist with generating multi-milligram quantities of the multi-substituted thio-SIINFEKL analogs, Nishant Raj (Chatterjee Group) synthesized the Ser1-Lys7 and Ser1-Ile3-Lys7 thio-peptides. All other peptides used in **Section 3.5** were synthesized by John Mannone (Gellman Group).

Data in **Section 3.4** identify the thio-SIINFEKL analogs that can stimulate T-cells *in vitro* and resist degradation in mouse serum. Based on these results, five thio-SIINFEKL derivates were selected for animal model studies. Both multi-substituted peptides that elicited an immune response (Ser1-Lys7 and Ser1-Ile3-Lys7), two mono-substituted peptides that had modifications at one of the sites in the triple-substituted thio-peptide (Ser1 and Ile3), and a mono-substituted negative control (Ile2). Asn4 was initially chosen as the negative control, but that peptide proved too difficult to synthesize in quantities sufficient for *in vivo* studies. Instead, Ile2 was chosen for the negative control as the thioamide is at an anchoring residue and showed poor stimulation of perforin secretion in **Section 3.4.2**.

3.5.1 – *In Vivo* T Cell Immunization

C57BL/6 naïve male mice were immunized with the five chosen thio-peptides and the native peptide. The immunizations were conducted in the presence of Toll-Like Receptor (TLR) 3 and 9 agonists to stimulate the immune system for better response to the antigens. TLR3 and TLR9 agonists stimulate the immune system to allow for more efficient immunization[65]. Immunizations consisted of two deliveries of 100 µg of antigen spaced seven days apart. After fourteen days the animals were euthanized, their spleens were collected, and flow cytometry coupled with IFNγ ELISpot assays were conducted to determine the percent of CD8+ T cells that had been activated.

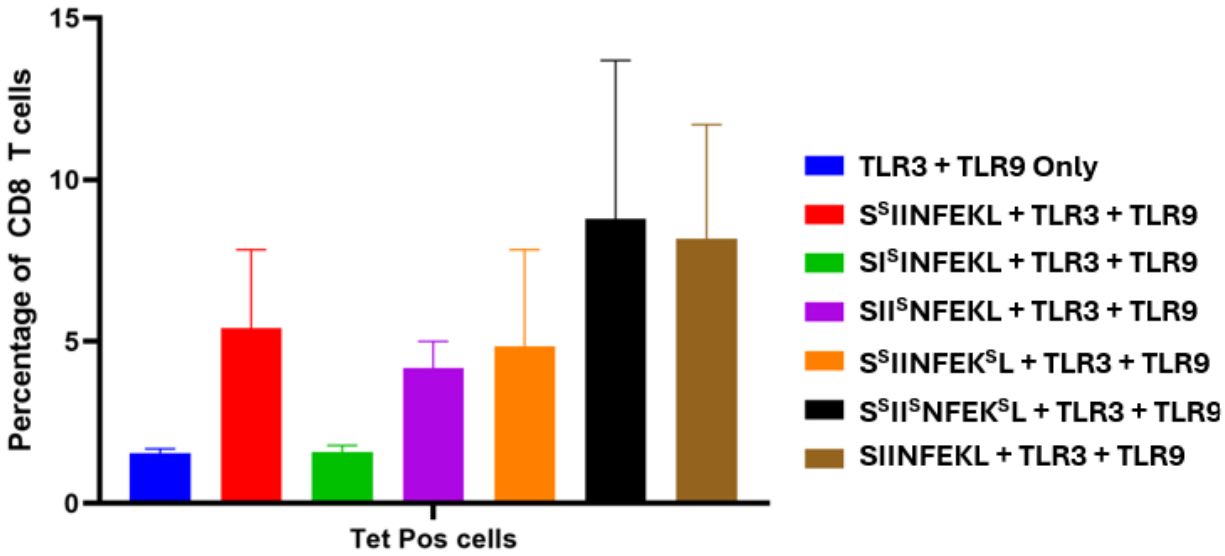


Figure 3.16: Bar graph indicating the percentage of CD8+ T cells detected after immunization with a given peptide antigen. For immunization, TLR3 + TLR9 without an antigen represents the negative control. SIINFEKL + TLR3 + TLR9 represents a benchmark for an antigen that elicits an immunogenic response. Superscript S within a peptide sequence indicates the residue where a thioamide modification has been incorporated. Experiments and results shown in this figure were performed by Dr. Ichwaku Rastogi (McNeel Group).

After measuring the degree of CD8+ T cell activation, we observed that the location of a thioamide modification can be indicative of the peptide's immunogenic profile *in vivo*. For example, all peptides that contain the Ser1 substitution elicit CD8+ T cell response on

par with the native SIINFEKL peptide. The Ile2 negative control was not able to stimulate a response which agrees with weaker stimulation of T cells observed in the perforin expression assay (**Figure 3.13**).

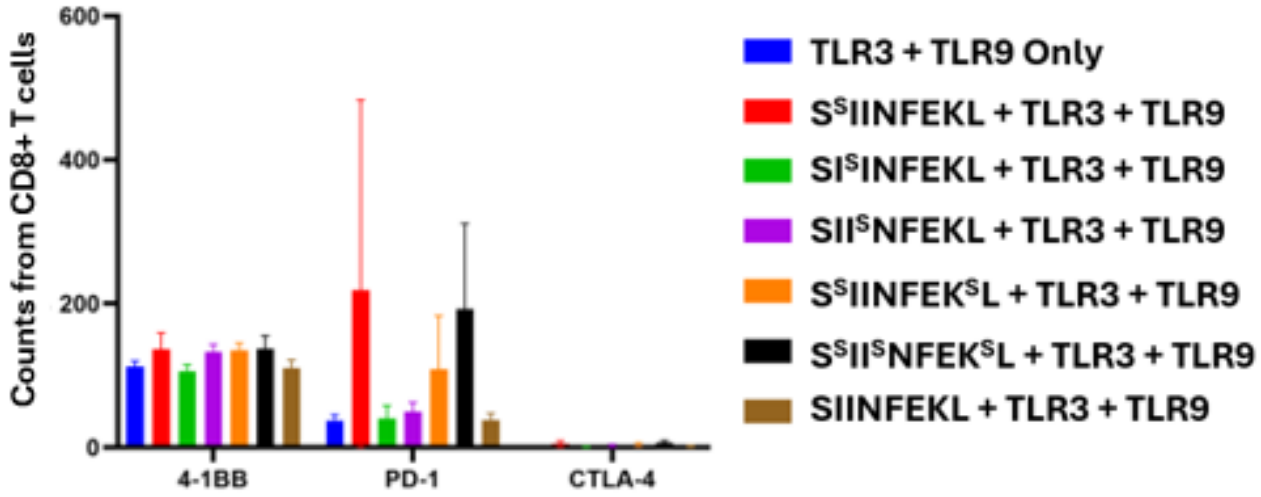


Figure 3.17: Bar graph indicating the counts of immune response associated proteins (4-1BB, PD-1, and CTLA-4) measured from T cells collected from mouse spleens. For immunization, TLR3 + TLR9 without an antigen represents the negative control. SIINFEKL + TLR3 + TLR9 represents a benchmark for an antigen that elicits an immunogenic response. Superscript S within a peptide sequence indicates the residue where a thioamide modification has been incorporated. Experiments and results shown in this figure were performed by Dr. Ichwaku Rastogi (McNeel Group).

In an analogous assay to the *in vitro* perforin expression monitoring assay (**Figure 3.13 and Figure 3.14**), we measured the expression of 4-1BB, PD-1, and CTLA-4 from T cells extracted from the spleens of immunized mice. We were unable to detect significant changes in expression levels of 4-1BB or CTLA-4 between the negative control or any of the antigens tested. PD-1, similar to the T cell activation assay (**Figure 3.16**), displayed drastically levels of expression for antigens that contained the Ser1 thioamide modification. Increased expression levels of PD-1 are typically associated with prolonged immune response signaling and continued activation of the T cell[63,64]. Although increased MHC

binding is a generally desirable trait for an mimetic antigen, increased expression of PD-1 is linked to T-Cell death and thus an undesirable trait for peptide vaccines.

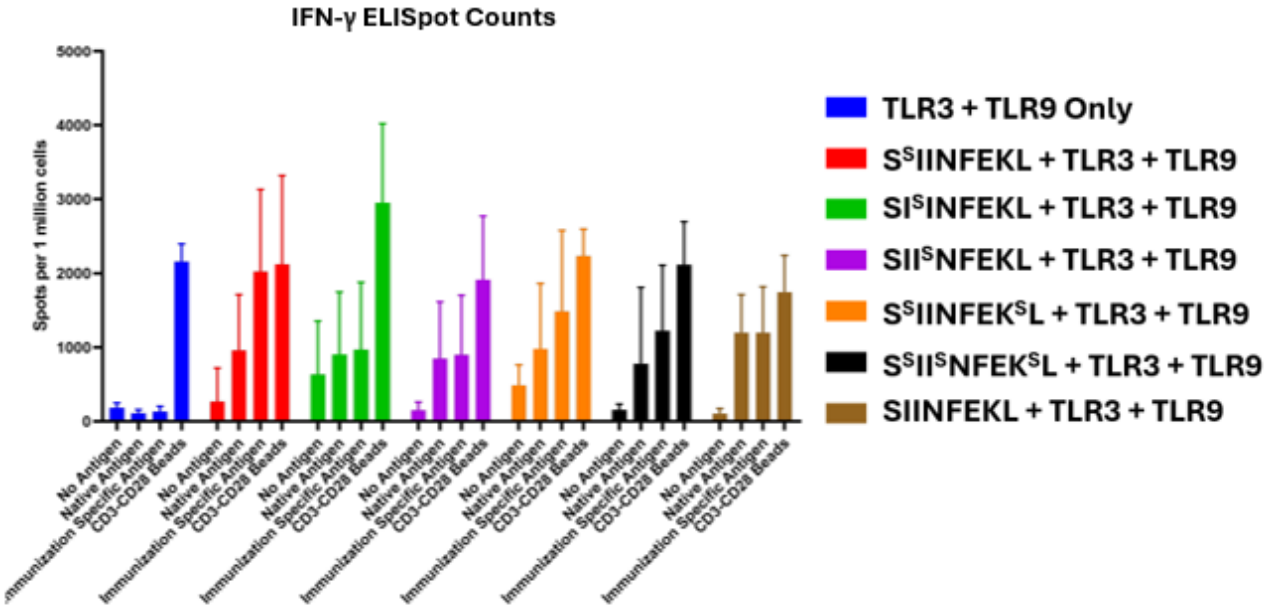


Figure 3.18: Bar graph indicating the cross-reactivity of antigens. T cells from the spleens of mice that had been immunized with an antigen were exposed to the native antigen to determine if they are cross reactive to one another. Each group indicates the immunization conditions that a mouse was given for the T cells used in the assay. T cells collected were given no antigen as a negative control, or CD3-CD28 beads were used as a positive control. The T cells were then exposed to the antigen of immunization or the native antigen to determine if the immunized T cells were response to the native antigen or just the immunization specific antigen. For immunization, TLR3 + TLR9 without an antigen represents the negative control. SIINFEKL + TLR3 + TLR9 represents a benchmark for an antigen that elicits an immunogenic response. Superscript S within a peptide sequence indicates the residue where a thioamide modification has been incorporated. Experiments and results shown in this figure were performed by Dr. Ichwaku Rastogi (McNeel Group).

Cross-reactivity, a T cell immunized with a thio-antigen recognizing the native antigen in this case, is a critical feature for the success of this vaccination approach. After immunization, T cells were exposed to the antigen used for immunization as a positive control or to the native SIINFEKL antigen to test for cross-reactivity. Any population of T cells that had been immunized with a thio-antigen was also cross-reactive with the native antigen (**Figure 3.18**). This result indicates that vaccination with thio-antigens immunizes T cells for recognition of the native antigen. In contrast, when the beta-3 series of SIINFEKL derivatives

were examined, it was discovered that they were able to elicit an immune response but displayed no cross-reactivity[29]. This cross-reactivity of thio-antigen immunized T-cells suggests that the thio-peptides are interacting with the MHC and T-cell receptor in a manner analogous to the native peptide.

3.5.2 – *In Vivo* Capability of Thio-SIINFEKL Antigens to Hinder Tumor Growth

Tumor growing cells (E.G7-OVA cells) were implanted into mice and allowed to grow for 9 days. On the 9th, day OT-1 cells (T cell population that, when immunized with the SIINFEKL antigen, stimulate an immune response against E.G7-OVA cells) and were adoptively transferred into the mice. The following day the mice were immunized with 100 µg of an antigen and Toll-Like receptor (TLR) 3 and TLR9 agonists. TLR3 and TLR9 agonists stimulate the immune system to allow for more efficient immunization[65]. Tumor growth was measured every other day, and when tumor volume exceeded 2 cm^3 the mice were sacrificed.

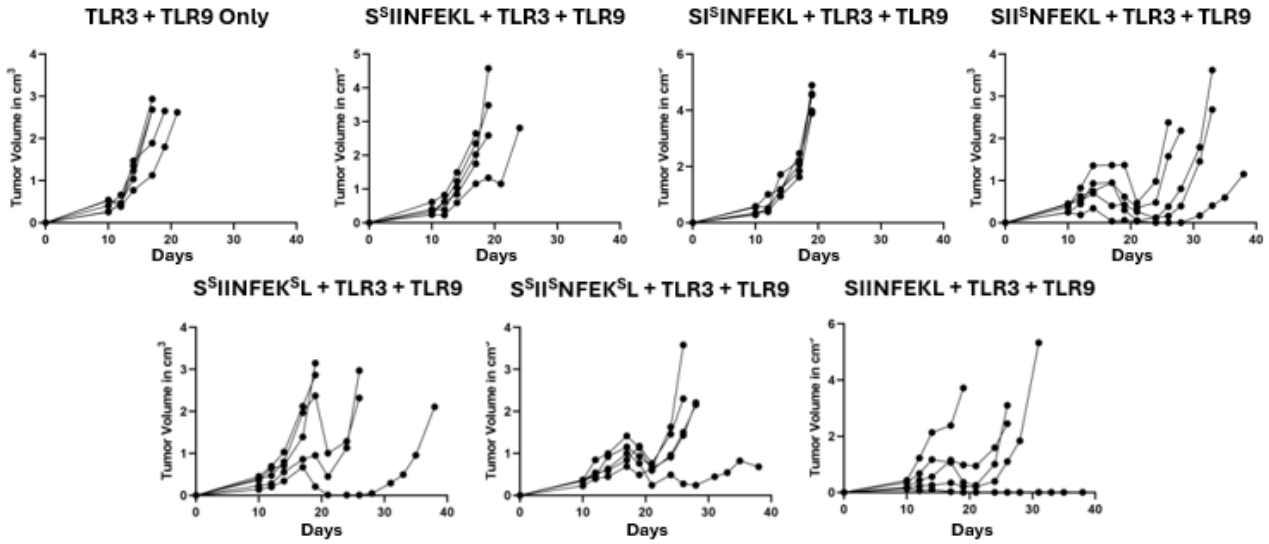


Figure 3.19: Plot of the survival rate of mice immunized with an antigen. Mice implanted with tumors were measured on alternating days to determine tumor growth. Once tumor volume exceeded 2 cm^3 the mice were sacrificed. For immunization, TLR3 + TLR9 without an antigen represents the negative control. SIINFEKL + TLR3

+ TLR9 represents a benchmark for an antigen that elicits an immunogenic response. Superscript S within a peptide sequence indicates the residue where a thioamide modification has been incorporated. Experiments and results shown in this figure were performed by Dr. Ichwaku Rastogi (McNeel Group).

Using TLR3 + TLR9 agonists only as negative control, it can be deduced that there was negligible reduction in tumor growth when either the Ser1 or Ile2 mono-substituted peptides were used for immunization. However, when any thio-antigen incorporating a thioamide at Ile3 was used for immunization, the reduction in tumor growth similar to the native antigen's ability to reduce tumor growth. The Ile3 thioamide containing peptides reduced tumor growth to prolong mouse survival past twenty days in all cases, this prolonged survival is present in most of the mice immunized with the native antigen. For immunization with the di-substituted antigen (Ser1-Lys7), about half of the mice displayed survival past twenty days while the other half displayed tumor growth at rates similar to the negative control. Similar to the Ser1 substitution being correlated to increased immunization rates (**Figure 3.16**) and expression of PD-1 (**Figure 3.17**), Ile3 substitution seems correlated to an increased capability to reduce the rate of tumor growth (**Figure 3.19**).

3.6 – Conclusions

In this study we report *in vitro* and *in vivo* evidence that thioamide modifications can be incorporated into antigenic peptides while maintaining immunogenicity. We have characterized *in vitro* that thioamide modified antigens can bind to MHC molecules on par with the native antigen. Additionally, the binding of thio-antigens to their cognate MHC can facilitate communication with CD8+ T cells in a manner analogous to the native peptide.

Peptide vaccines, and more broadly peptide drugs, have had their pharmaceutical utility hampered by poor stability against enzymatic proteases. I developed and conducted

a proteolysis assay that screens an antigen against all endogenous proteases present in a given serum. I discovered that when N- and C-terminal amides of the SIINFEKL epitope are replaced with thioamides a decreased susceptibility to degradation in serum is observed.

To determine if our *in vitro* data translated into an *in vivo* system, we immunized mice with thio-SIINFEKL derivatives and observed that thio-antigens containing a thioamide modification at Ser1 lead to the highest degree of T-cell immunization. We discovered that modified antigens with a thioamide at Ser1 did lead to increased expression of PD-1 which may facilitate cell termination and reduce the efficacy of certain thioamide containing peptides as vaccine candidates. Finally, when monitoring tumor growth after immunization, we observed that thio-antigens containing the Ile3 modification have a propensity to reduce tumor rate growth.

We believe that thioamide modifications within antigenic peptides present a promising route to addressing common issues for peptide drug application. We have observed multiple features that make thio-antigens a potentially viable route to increase the utility of peptide vaccines. With judicious selection of location for thioamide incorporation within an antigen's sequence, the *in vivo* properties of an antigen can potentially be tailored to the afford a desired outcome. Further investigation is required to determine whether this is a viable strategy for human vaccination.

3.7 – Methods and Characterization

3.7.1 – Resin Loading Procedure

Dry resin was swelled in a mixture of DCM and DMF (9:1, v/v; 10 mL/g) in a 45 mL falcon tube. In a separate container 4 eq of Fmoc-Leucine-OH and 4 eq of HOBt were dissolved in the minimal amount of DMF, and the solution was added to the resin solution. 4 eq of DIC was then added to the reaction mixture followed by 0.1 eq of DMAP. The mixture was agitated with a mechanical shaker overnight (14 hours). The next day, the resin was drained into an SPPS syringe vessel and washed with DCM (3x20 mL). The unreacted hydroxyl groups were then capped with 2 eq of acetic anhydride and 2 eq of DIEA in the minimal amount of DCM to fully submerge the resin. The resin mixture was then agitated for an additional 30 minutes on the mechanical shaker. The resin was then drained of the reaction solution and washed with DCM (3x20 mL), then dried for an hour on an aspirator. The loading was estimated by weight gained of the resin and then the resin was used for SPPS.

3.7.2 – General Method for Thioamide Peptide Synthesis

All thioamide monomers for peptide synthesis have been prepared as previously reported[50-51].

Synthesis of Fmoc-AA-4NoPDA

Fmoc-protected amino acid was dissolved in DMF along with HATU (1.5 eq) and DIEA (3.0 eq). The solution was stirred for 5-10 mins to allow for activation of the carboxylic acid. Once activation was completed, 4-nitrophenylene diamine (4NoPDA) (1.2 eq) was then added to

the reaction vessel, and the solution was allowed to stir overnight (14 hours). Once all starting material was consumed (monitored by TLC), the crude product was precipitated with cold water and filtered. The crude material was then washed several times with cold water (3x150 mL) and then redissolved in ethyl acetate (EA). The solution was transferred to a separatory funnel. The organic layer was washed with sat. NH_4Cl, sat. $NaHCO_3$, and brine (3x150 mL each), respectively. The organic layer was then dried over Na_2SO_4, filtered, concentrated *in vacuo* and used without further purification for the next step.

Synthesis of Lawesson's Reagent

Anisole (50 mL, 464 mmol) was added directly to P_4S_{10} (20.6 g, 46.5 mmol,) and the mixture was refluxed at 160°C for 2 hours. After 1 hour the P_4S_{10} began to dissolve, and after two hours precipitate formed. The mixture was allowed to cool to room temperature. The reaction mixture was filtered through a Buchner funnel, and the solid was washed with diethyl ether (3x50 mL) and then chloroform (3x50 mL). The container housing the LR was purged of air under a stream of nitrogen gas and stored at 4°C until use. 52 g of LR was recovered.

Synthesis of Fmoc-Thio(AA)-4NoPDA

Crude Fmoc-AA-4NoPDA was added to a reaction vessel with LR and the reaction vessel was purged under a stream of N_2 for 10 minutes. Dry THF was added to the flask, and the mixture was mildly heated (40°C) while stirring overnight (14 hours). The following day the reaction was stopped by removing the solvent under vacuum. The crude mixture was dissolved in EA and washed with sat. $NaHCO_3$ (3x150 mL). The organic layer was then dried over Na_2SO_4 and

concentrated *in vacuo*. The material was purified by flash column chromatography over silica using a mobile of pentanes and EA. The purified material was concentrated *in vacuo* and stored at 4°C until incorporation onto solid support. The yields for this step ranged from 15-80%, depending on the amino acid.

Synthesis of Fmoc-Thio(AAA)-Benzotriazole and Thioamide Formation on Resin

Fmoc-Thio(Xxx)-4NoPDA was dissolved in the minimal amount of 95 % acetic acid solution, and $NaNO_2$ (10eq) was added to the solution. The solution was quickly capped to prevent escape of $NO_{(g)}$ and then allowed to stir at 4°C until the full conversion by TLC. The crude product was then precipitated with cold water and then redissolved in DCM. The organic layer then washed with sat $NaHCO_3$ until the pH of the aqueous layer was neutral. The organic layer was then dried over Na_2SO_4 and concentrated *in vacuo*. The crude product was then dissolved in 1-2 mL DCM per 25 µmol of resin. 2 eq of activated thioacylating reagent was added to pre-swelled resin and allowed to mix under protection from light. After 45 minutes the solution was drained from the resin, and another 2 eq of activated thioacylating reagent was added. The reaction mixture was again allowed to mix under protection of light for 45 minutes. The solution was drained from the resin and the resin was washed with DCM 3x10 mL.

Synthesis of 4-azido-methylbenzene

p-Toluidine (10 g, 92.8 mmol) was dissolved in 46mL of H_2O:HCl (1:1) at 0°C. $NaNO_2$ (6.33 g, 92.8 mmol) was dissolved in 26 mL of water and added dropwise to the *p*-toluidine solution. Once all $NaNO_2$ was added to the flask, the mixture was allowed to stir for 10 minutes. NaN_3

(6.1 g, 92.8 mmol) was dissolved in 24 mL of water, and the solution was added slowly and dropwise to the reaction mixture. Once all NaN_3 was added, the reaction mixture was allowed to stir for 1 hour with an open top. The reaction was then extracted with 3x200mL of chloroform. The combined organic layers were then washed with H_2O (3x100mL), dried over Na_2SO_4 and concentrated in vacuo. The next step was conducted without further purification.

Synthesis of 4-azido-methylbromobenzene

4-Azido-methylbenzene was dissolved in 250 mL of benzene under a nitrogen atmosphere. *N*-bromosuccinimide (16.5 g, 92.8 mmol) and azobisisobutyronitrile (4.6 g, 27.8 mmol) were added to the reaction mixture, which was allowed to reflux overnight (14 hours). The next day, 200 mL of water was added to the reaction mixture. The reaction solution was then transferred to a separatory funnel. Diethyl ether (3x150 mL) was then used to extract the crude product. The combined organic layers were dried over Na_2SO_4 and concentrated *in vacuo*. The crude oil was purified via flash column chromatography with a pentane mobile phase. The product was isolated as a brown oil (14g, 66.5mmol) and stored at 4°C.

Thioimidate Protection on Resin

Immediately after formation of a thioamide on resin, 60 uL of 4-azido-methylbromobenzene and 15 uL of DIEA are mixed in 2 mL of Dry DMF and added to the SPPS reaction vessel. The reaction is allowed to mix overnight (14 hours) under protection from light.

Peptide Synthesis and Cleavage

Peptides were synthesized via microwave-assisted reactions on Fmoc-Leucine Wang Resin at 25 µmol scale. For each coupling reaction, 5 equivalents of Fmoc-protected amino acid, 5 equivalents of Oxyma, and 10 equivalents of DIC in biotech-grade DMF were mixed with resin, and the mixture was heated to 75°C for 4 minutes in an SPPS reaction vessel while stirring. After the coupling reaction, the resin was washed with ACS-grade DMF (3x5mL). For Fmoc-deprotection, 20% (v/v) piperidine in DMF (3 mL) was added to the resin, and the mixture was heated at 80°C for 2 minutes while stirring. After deprotection, the resin was washed with DMF (3x5mL). Once peptide elongation was completed, a final deprotection step was undertaken, and the resin was washed with ACS-grade DMF (3x5mL), then DCM (3x5mL) and placed under reduced pressure on an aspirator until dry. Peptides were then cleaved from the resin, with concomitant side chain deprotection, by combining the resin with a 2 mL per 50 µmol solution of 75% TFA, 23% phenol, and 2% triisopropylsilane, and gently rocking the vessel for 3 hours. The reaction solution was then drained into a 45 mL falcon tube, and the resin was washed twice with 2 mL DCM. The combined peptide cleavage solution was then concentrated under a stream of nitrogen gas until less than 0.5 mL remained. The crude peptide was then precipitated with 40 mL of cold diethyl ether. This mixture was centrifuged at 4k rpm for 10 minutes to pellet the suspended solid, and then the ether was decanted. A second volume of 40 mL cold diethyl ether was added, and the process of centrifugation and decanting was repeated. The resulting solid was dried under a stream of nitrogen gas.

Thioimidate Deprotection

Crude peptide powder after cleavage was mixed with 10 eq of DTT and 2e q of DIEA in 1 mL per 25 µmol Dry DMF. The solution was mixed overnight (14 hours). The following day the product was precipitated with cold diethyl ether that had been chilled in dry ice (40 mL). The product was then centrifuged at 4k rpm for 10 mins to pellet the peptide. The supernatant was decanted, and another 40 mL of dry ice cooled diethyl ether was added, which was followed by another round of centrifugation. The diethyl ether was then decanted and the crude thioamide containing peptide was dried over a stream of N_2.

Peptide Purification

Peptides were dissolved in the minimal quantity of DMSO (1-4mL) and purified via reverse-phase high performance liquid chromatography (RP-HPLC; CSH C18 stationary phase, 5 µm, 19 mm x 250 mm, the solvent gradient was 10-40 % acetonitrile over 20 minutes). The purity of each peptide was assessed via ultra-high performance liquid chromatography (UPLC; CSH C18 stationary phase, 1.7 µm, 2.1 mm x 100 mm, the solvent gradient was 10-60% acetonitrile over 10 minutes). All peptides reported here were >95% pure, based on UPLC analysis monitoring at 220 nm. Peptide concentration was determined by absorbance at 205 nm (molar extinction coefficient = 28460 $M^{-1}cm^{-1}$). This value was determined using previously reported literature values of proteogenic amino acids at 205 nm[66]. The peptide solution was then aliquoted into Eppendorf tubes and lyophilized to provide a white powder. The material was stored at -80°C. When needed, a peptide sample was dissolved in DMSO at concentrations required for a given assay.

3.7.3 – Proteolysis of Thio-SIINFEKL Antigens

Stock solutions of Normal Mouse Serum (NMS) were prepared by resuspending lyophilized serum in 1 mL of sterile Nanopure water and allowing the mixture to incubate at 4°C for 2 hours with occasional stirring. NMS was then diluted to 25% serum and stored at -20°C until use. Immediately before use, NMS stock solution was allowed to thaw at room temperature for 30 minutes, and was then warmed at 37°C for 30 minutes before beginning the assay. For each proteolysis reaction, a 5.0 mM stock solution of each peptide in DMSO was diluted to 1.0 mM with DPBS. 20 µL of peptide solution was added to 180 µL of the pre-warmed NMS, and the solution was incubated at 37°C. This solution contained 100 µM peptide, 2% v/v DMSO, and 22.5% v/v NMS. At each timepoint, a 20 µL aliquot of the reaction solution was removed and mixed with 20 µL of the quenching solution: 42.5% H_2O/42.5% acetonitrile/15% TCA with 40 µg/mL of glibenclamide included as an internal standard. For each reaction, a "0 min" control was prepared by removing 20 µL from the proteolysis reaction mixture after vigorously mixing for 10 seconds. After quenching, the aliquots were allowed to incubate at 4°C for 40 minutes. The aliquots were then centrifuged to pellet precipitated serum proteins. 20 µL of the supernatant was mixed with 13.5 µL of 1 M NaOH to adjust the solution pH to 7, followed by a second round of centrifugation to separate any precipitate that formed. A 10 µL aliquot of each quenched and neutralized reaction solution was injected onto a Waters Acquity H-Class UPLC equipped with an Acquity BEH C18 column. 5-90% B (A: Nanopure water+0.1% TFA; B: HPLC-grade acetonitrile+0.1% TFA) gradient over 10 minutes with a 0.3 mL/minute flow rate. The relative amount of peptide was quantified by integration of the SII/analogue peak at 220 nm and normalization to the

glibenclamide internal standard. For each sample, the % peptide remaining was calculated by dividing the normalized peak area by the normalized peak area from the “0 min” control. Each reaction was run in triplicate. Half-life values were determined by fitting normalized data to a one-phase decay model in Origin 2020.

3.7.4 – Antigen MHC Binding

Assay was performed as previously described[98].

3.7.5 – *In Vitro* Stimulation of T Cells

Spleens collected from OT-1 mice were prepared in single cell suspension ($4x10^6$ cells per mL. 100 µL (400,000 cells) of cells were incubated per well in a 96 well plate. Biological triplicates for each treatment. 100 µL of plain media was added to negative control group and 100µl of media containing 2.5 µL of CD3-CD28 beads to the positive control groups. Splenocytes were incubated with peptide overnight at 37°C in incubator. (approx. 18 hrs). Next morning Golgi stop was added to all wells and incubated for six hours (diluted 1:100 in media and added 13µL per well). Cells were washed twice with FACS buffer and antibody mix was prepared for surface markers by diluting each antibody at 1:100 in BSB + FACS (1:1).

Protein monitored – surface marker used:

4-1BB – APC

CD8 – BV786

L/D – Ghost780

PD1 – BV650

CD4 – BUV395

50 µL of antibody mix was added to each well and incubated at 4°C for 30 minutes. The cells were washed once with PBS and resuspended in 200 µL of Fix/Perm Buffer [diluted 1:4; Concentrate: Diluent] (from FOXP3 staining kit). Celled were stored overnight in fixation buffer at 4°C in dark. The following day, the cells were spun down and washed once with 200 µL Perm-wash buffer (from FOXP3 kit) [diluted 1:10 in water] per well. Intra-cell antibody stain was then prepared as following: In 1:1 BSB and 1:10 Perm-wash add the following antibodies at 1:100 dilution, make up the volume with FACS wash.

Protein monitored – antibody used:

Perforin – FITC

Granzyme B – PE-CF594

TNFα – PE-Cy7

IFNγ – PE

50 µL of antibody mix was added per well and incubates for 30 minutes at 4°C. Cells were washed once with perm-wash buffer and then resuspended in 100 µL of FACS wash then Flow. Cells were analyzed using FlowJo version 10.10, for percentage of cells positive for expression of activation and checkpoint markers among all CD8 T cells.

3.7.6 – Excretion of Other Immune Response Associated Proteins

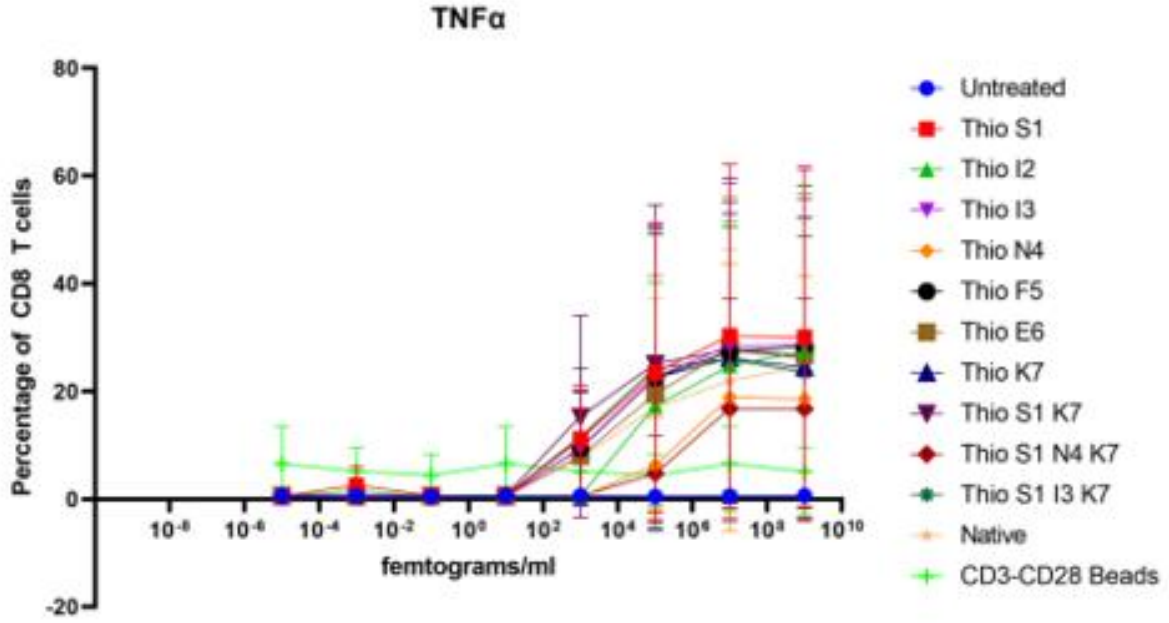


Figure 3.20: The percentage of CD8+ T Cells as function of peptide concentration and monitored via TNFα excretion. Position of the thioamide is indicated by the amino acid 1-letter code and position in the sequence.

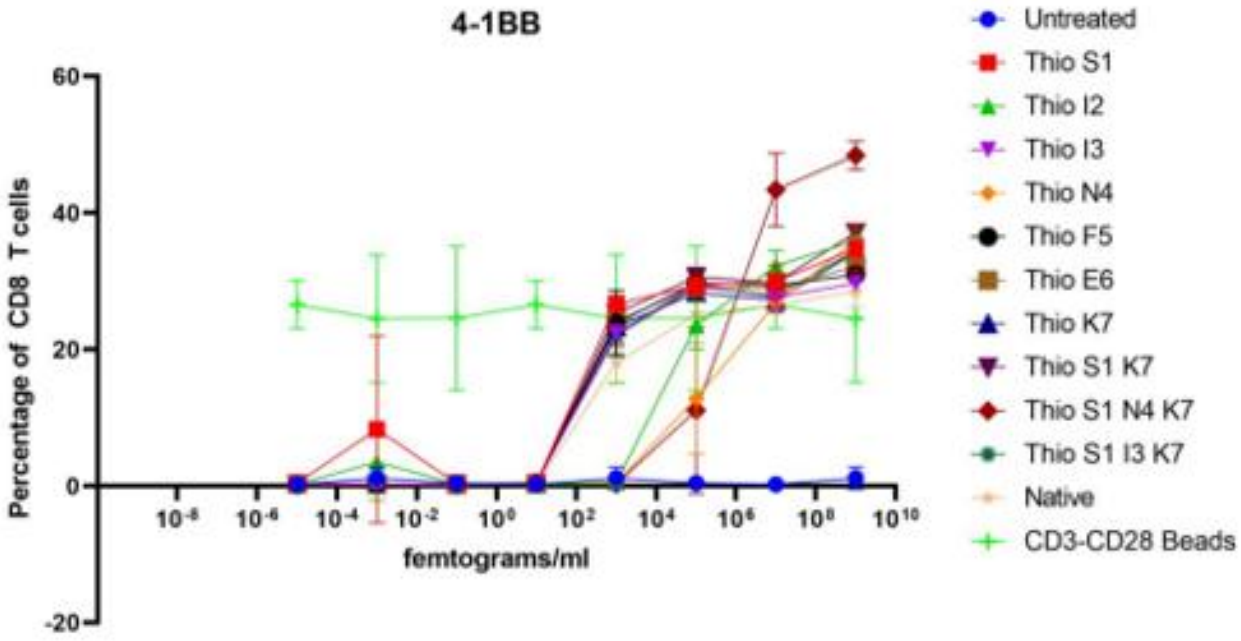


Figure 3.21: The percentage of CD8+ T Cells as function of peptide concentration and monitored via 4-1BB excretion. Position of the thioamide is indicated by the amino acid 1-letter code and position in the sequence.

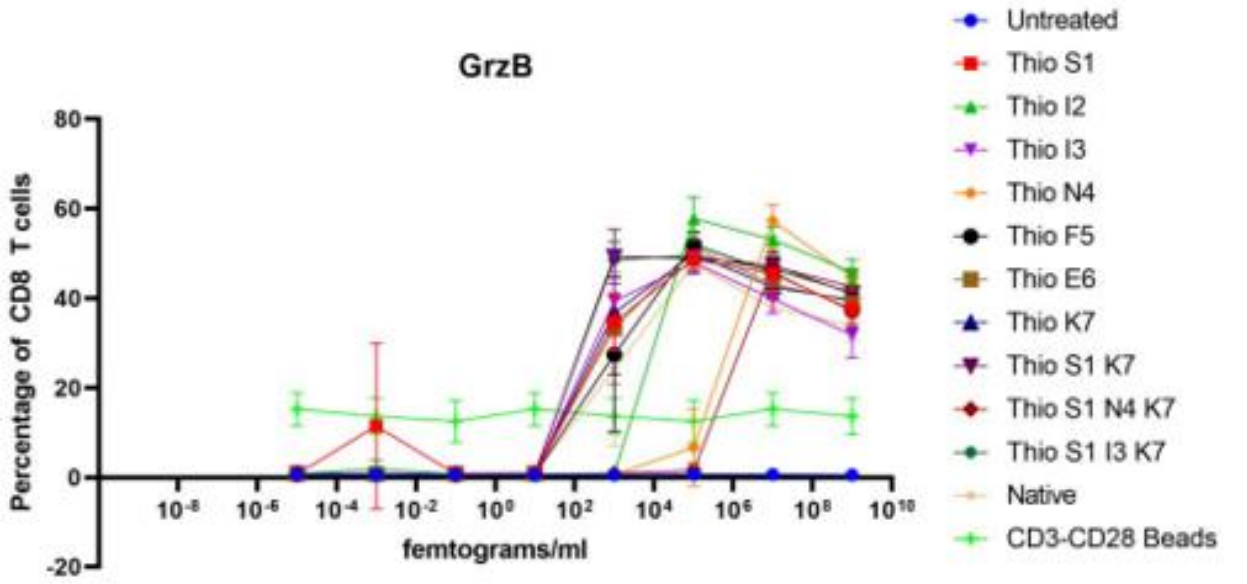


Figure 3.22: The percentage of CD8+ T Cells as function of peptide concentration and monitored via GrzB excretion. Position of the thioamide is indicated by the amino acid 1-letter code and position in the sequence.

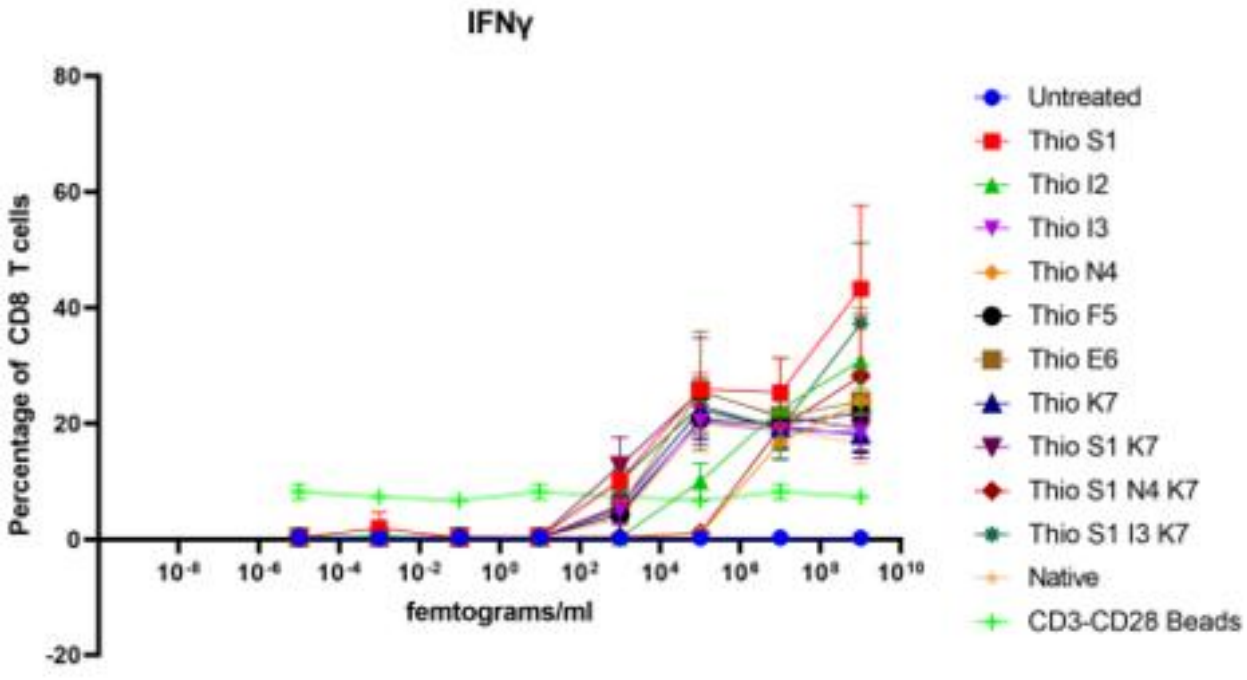


Figure 3.23: The percentage of CD8+ T Cells as function of peptide concentration and monitored via IFNγ excretion. Position of the thioamide is indicated by the amino acid 1-letter code and position in the sequence.

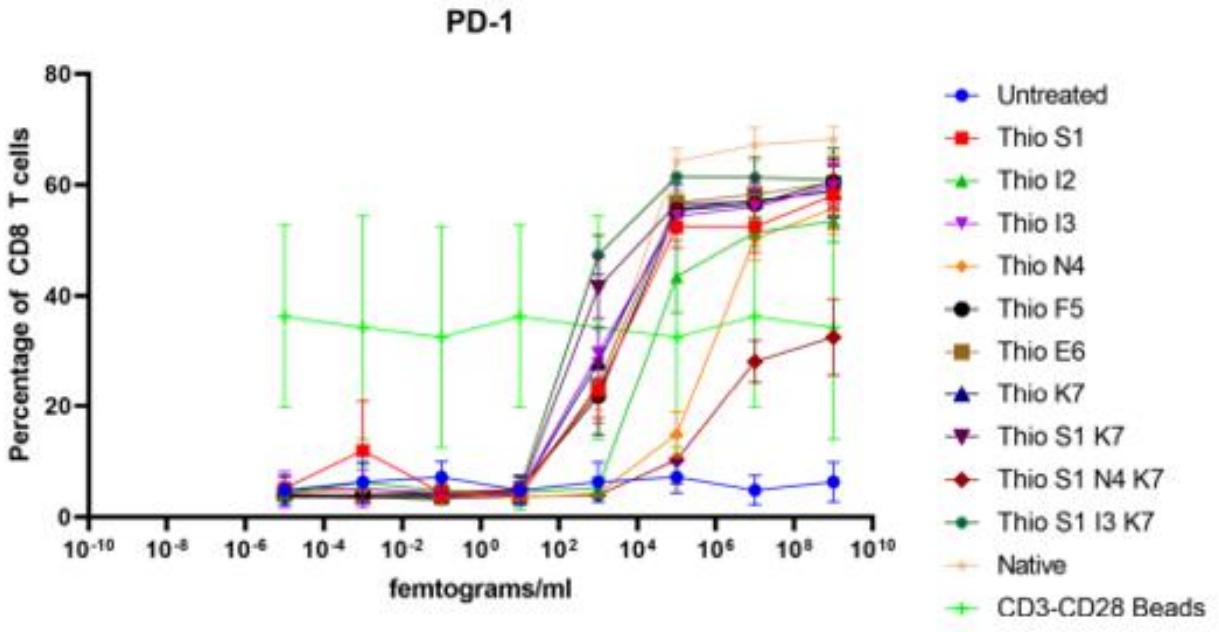


Figure 3.24: The percentage of CD8+ T Cells as function of peptide concentration and monitored via PD-1 excretion. Position of the thioamide is indicated by the amino acid 1-letter code and position in the sequence.

3.7.7 – *In Vivo* Experiments

Immunization Study

A total of 7 groups were formed as follows:

a. PBS + TLR3 + TLR9 agonists [200 µl SC] – twice one week apart; 4 mice

b. Thio(Ser1) + TLR3 + TLR9 agonists [200 µl SC] – twice one week apart; 5 mice

c. Thio(Ile2)+ TLR3 + TLR9 agonists [200 µl SC] – twice one week apart; 5 mice

d. Thio(Ile3)+ TLR3 + TLR9 agonists [200 µl SC] – twice one week apart; 5 mice

e. Thio(Ser1Lys7) + TLR3 + TLR9 agonists [200 µl SC] – twice one week apart; 5 mice

f. Thio(Ser1Ile3Lys7) + TLR3 + TLR9 agonists [200 µl SC] – twice one week apart; 5 mice

g. Native + TLR3 + TLR9 agonists [200 µl SC] – twice one week apart; 5 mice

After one week of second immunization, all mice were euthanized, and their spleens were collected. Splenocytes were prepared into single cell suspension and for each mouse fresh splenocytes were stained for flow cytometry the same day for analyzing SIINFEKL tetramer positive cells. Cells were analyzed using FlowJo version 10.10, for percentage of cells positive for SIINFEKL tetramer, global populations and expression of activation/exhaustion markers on T cells.

ELISPOT plate was prepared with 200,000 cells per well per animal (X4). Below were groups tested:

a. Untreated

b. Native pep

c. Specific Thio-antigen

d. CD3-CD28 beads

ELISPOT kit protocol was followed and the plate was coated with IFNγ antibody (one night before a mouse was sacrificed day). The cells were plated and treated as described above. The cells were undisturbed while incubating for 48 hours, then tagged with biotinylated-IFNγ, and finally with Streptavidin-AP. The plates were developed with BCIP and NBT, and allowed to dry overnight/over the weekend then read using the ELISPOT plate reader. Cells were plotted for number of spots counted.

Note: For immunizations: Peptide @ 100µg/mouse; TLR3 @ 100µg/mouse, TLR9 @ 50 µg/mouse.

For ELISPOT: coating antibody and biotin antibody used were the same as generally used for ELISA.

Tumor Growth Study

E.G7-OVA tumor cell line was cultured and expanded *in vitro*. A total of 35 C57Bl/6 mice were injected with $1x10^6$ EG7-OVA tumor cells (Day 0).

Total 7 groups [5 mice per group]:

a. PBS + TLR3 + TLR9

b. Thio(Ser1) + TLR3 + TLR9 agonists

c. Thio(Ile2 + TLR3 + TLR9 agonists

d. Thio(Ile3) + TLR3 + TLR9 agonists

e. Thio(Ser1Lys7) + TLR3 + TLR9 agonists

f. Thio(Ser1Ile3Lys7) + TLR3 + TLR9 agonists

g. SIINFEKL + TLR3 + TLR9 agonists

On the 9^{th} day of tumor inoculation, one OT-1 mouse was euthanized and its spleen was processed into a single cell suspension. $2x10^6$ OT-I spleen cells were injected per mouse via intraperitoneal injection. The following day (Day 10), the animals were euthanized with 100 µg of above listed peptide diluted in PBS, mixed with TLR3 and TLR9 agonists subcutaneously. Tumors were measured every other day and the animals were sacrificed

when the tumor volume reached or exceeded 2 cm^3. The survival percentage and tumor volume graphs were plotted for each animal.

3.7.8 – MALDI-TOF-MS and UPLC Data for Peptides Containing a Thioamide Substitution

For all mass spectrometry analysis, a 1 µL solution of 20 mg/mL 2,5-dihydroxybenzoic acid (DHB) matrix in 70:30 acetonitrile:water was allowed to evaporate on a Bruker Ultraflex MALDI plate. Once all solvent had evaporated, 1 µL of 50:50 acetonitrile:water solution containing a dissolved and purified thio-peptide was placed atop where the DHB matrix had dried. Once the peptide/matrix sample was fully dried, the peptides were ionized on a Bruker Ultraflex using 10% laser power. DHB, being a milder ionizing matrix, allows for detection of multiple +1 adducts of the thio- peptides; in this case [M+H]+, [M+Na]+, [M+K]+, and [M+2Na-H]+. Spectra are shown below.

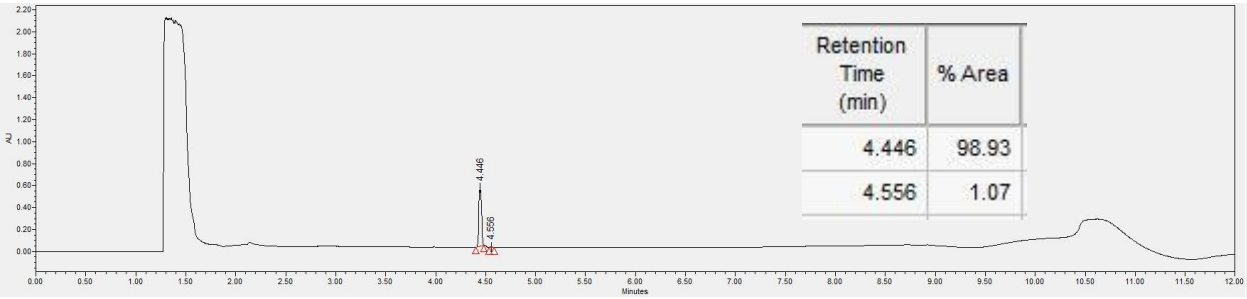


Figure 3.25: UPLC chromatogram of SIINFEKL at 220nm. Purity >98%.

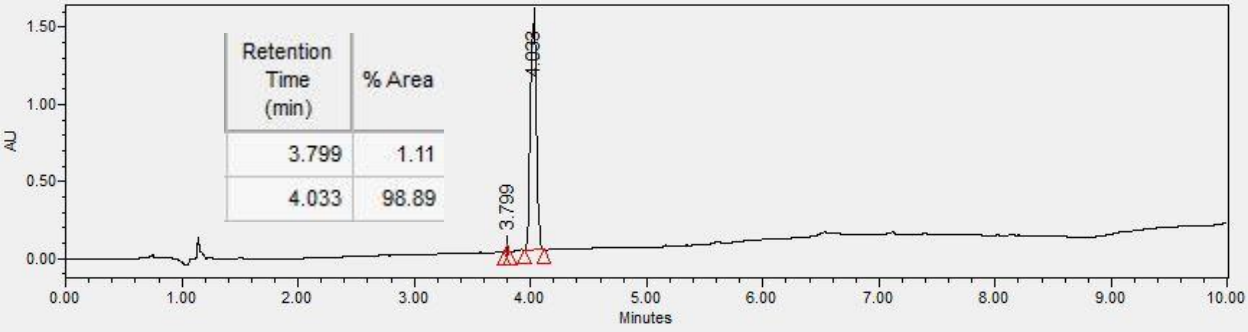


Figure 3.26: UPLC chromatogram of Thio(Ser1) SIINFEKL at 220nm. Purity >98%.

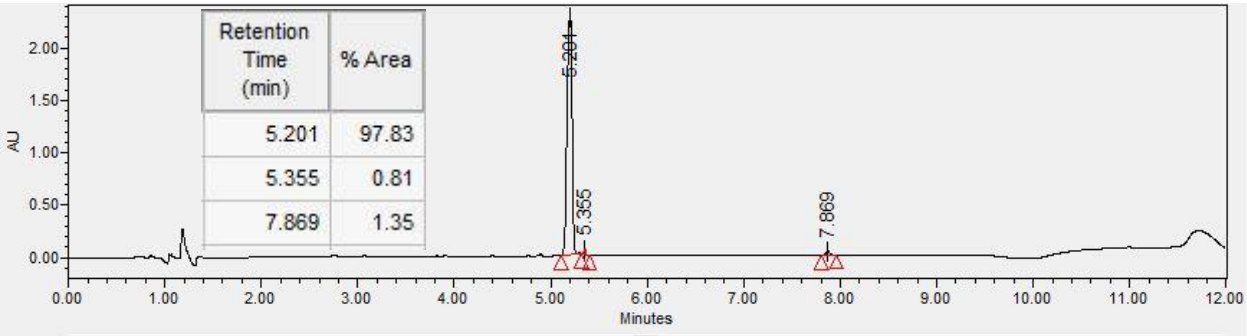


Figure 3.27: UPLC chromatogram of Thio(Ile2) SIINFEKL at 220nm. Purity >97%.

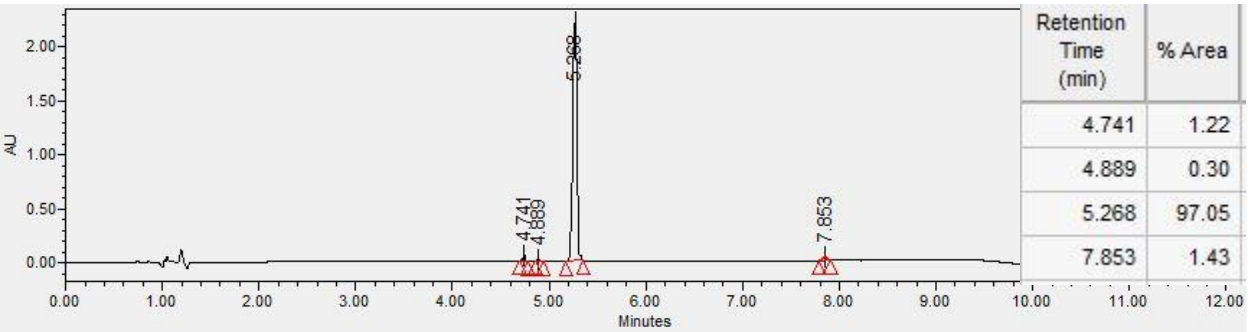


Figure 3.28: UPLC chromatogram of Thio(Ile3) SIINFEKL at 220nm. Purity >97%.

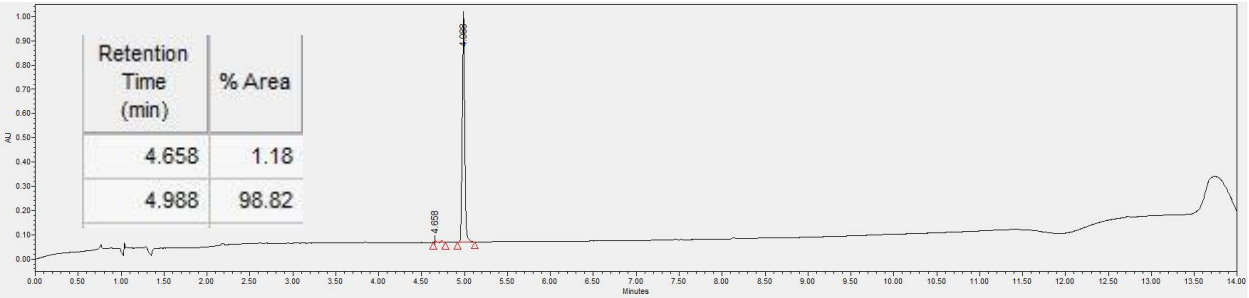


Figure 3.29: UPLC chromatogram of Thio(Asn4) SIINFEKL at 220nm. Purity >98%.

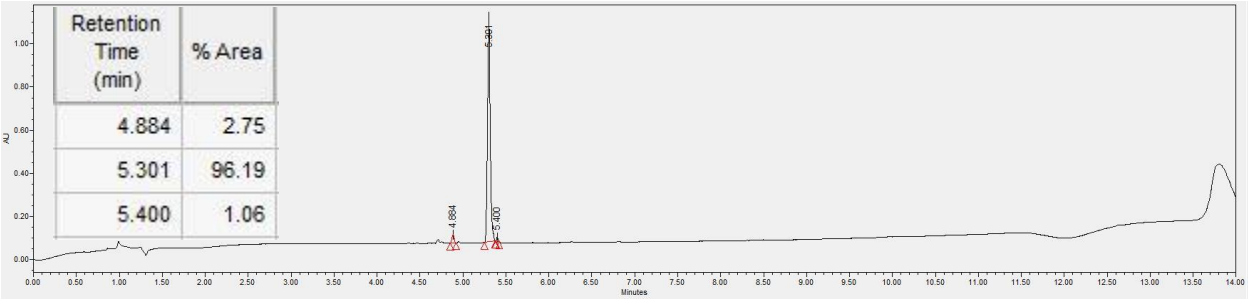


Figure 3.30: UPLC chromatogram of Thio(Phe5) SIINFEKL at 220nm. Purity >96%.

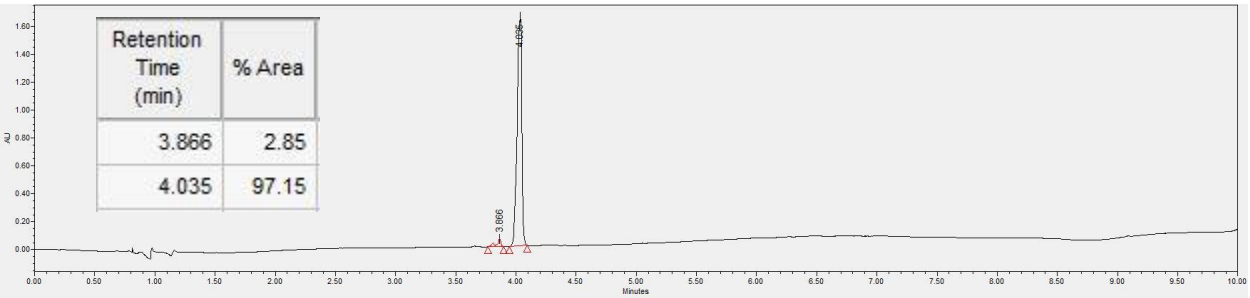


Figure 3.31: UPLC chromatogram of Thio(Glu6) SIINFEKL at 220nm. Purity >97%.

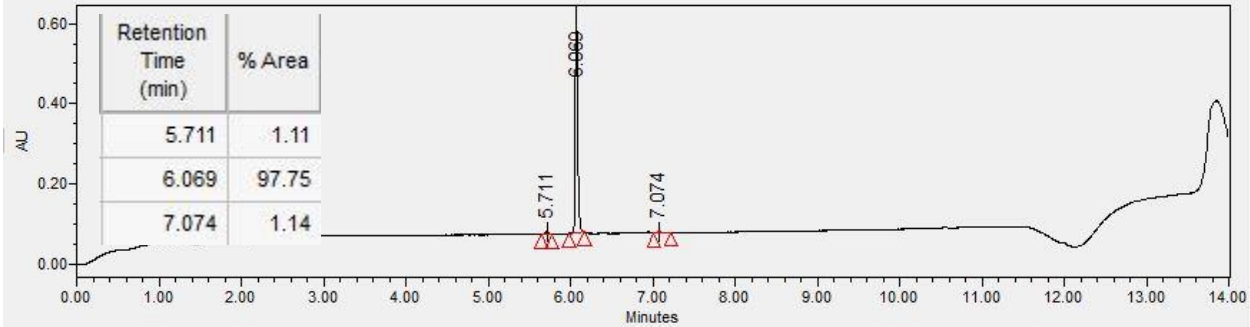
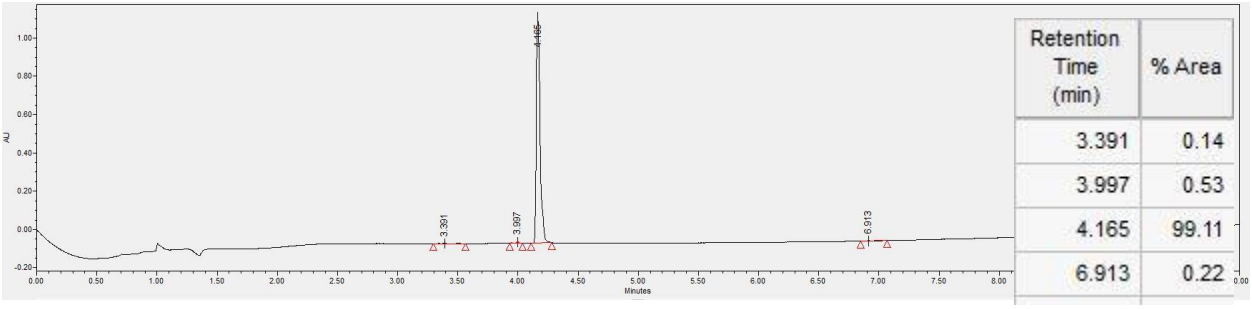
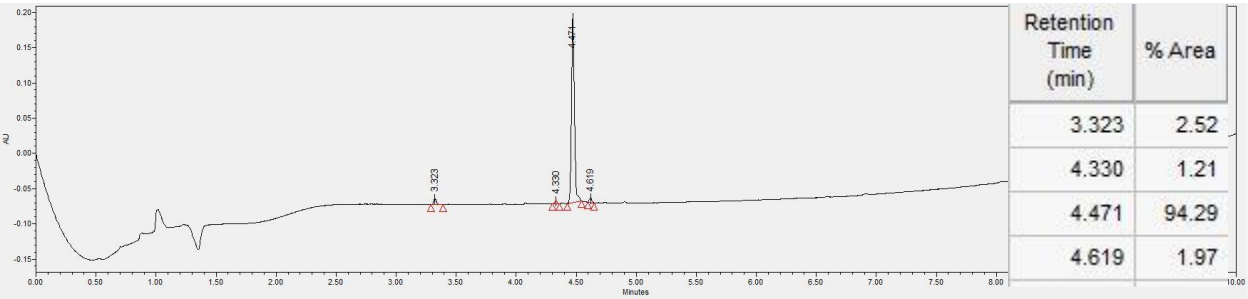


Figure 3.32: UPLC chromatogram of Thio(Lys7) SIINFEKL at 220nm. Purity >97%.



Retention Time (min)	% Area
3.391	0.14
3.997	0.53
4.165	99.11
6.913	0.22

Figure 3.33: UPLC chromatogram of Thio(Ser1Lys7) SIINFEKL at 220nm. Purity >99%.



Retention Time (min)	% Area
3.323	2.52
4.330	1.21
4.471	94.29
4.619	1.97

Figure 3.34: UPLC chromatogram of Thio(Ser1Ile3Lys7) SIINFEKL at 220nm. Purity >94%.

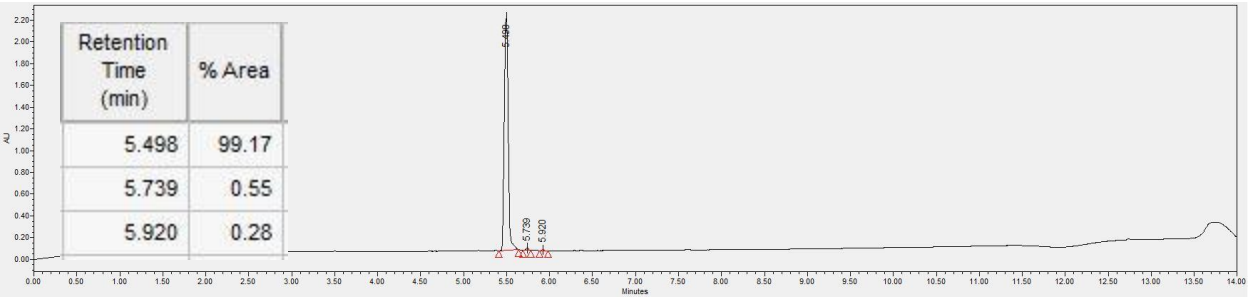


Figure 3.35: UPLC chromatogram of Thio(Ser1Asn4Lys7) SIINFEKL at 220nm. Purity >99%.

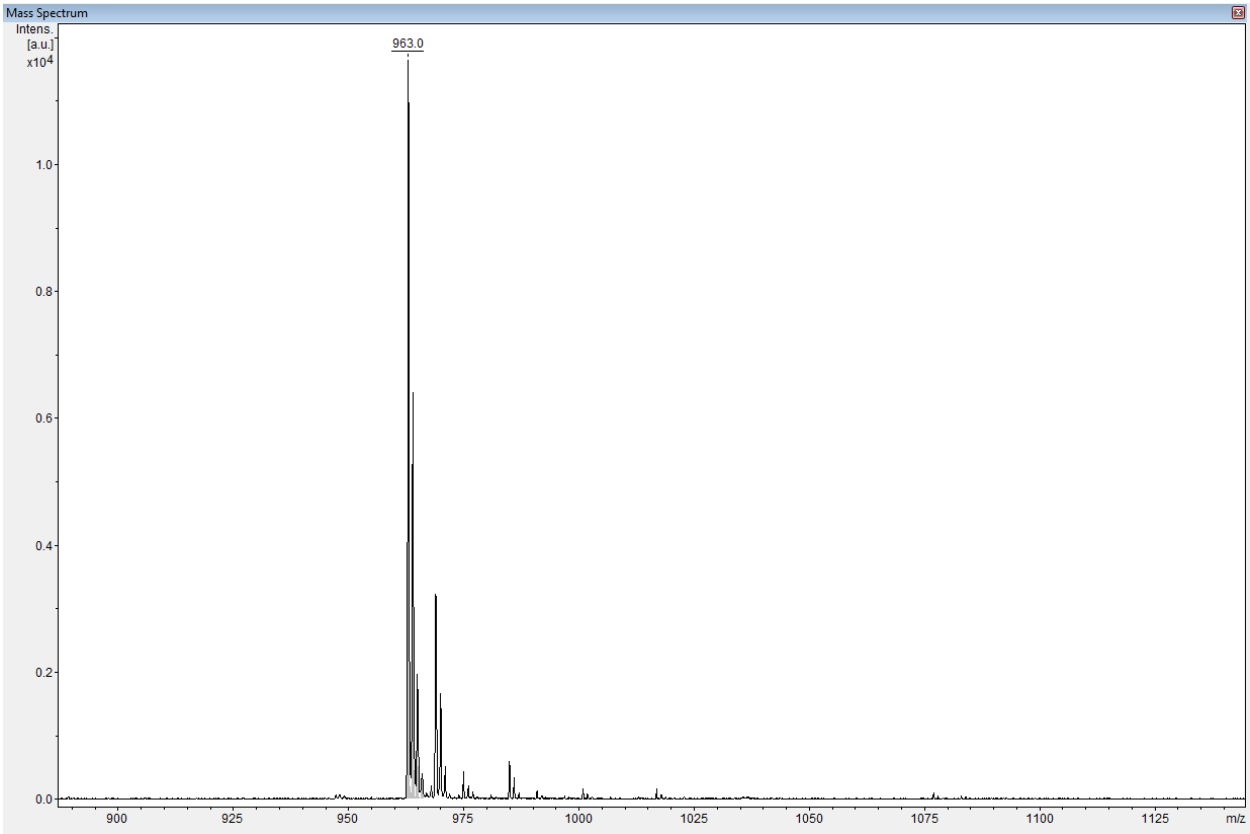


Figure 3.36: MALDI-TOF-MS of Native SIINFEKL. Observed Mass $[M+H]^+$ = 963.0. Calculated mass: $[M+H]^+$ = 964.1**.**

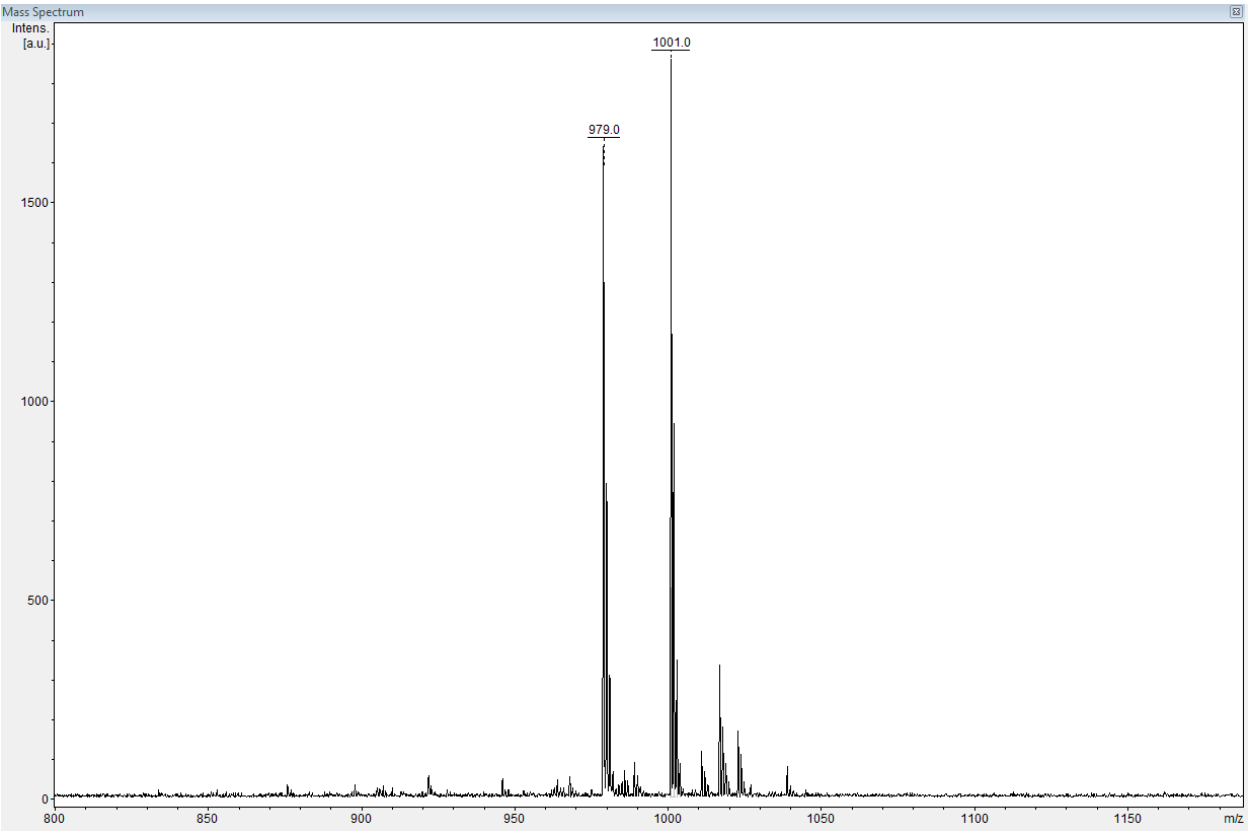


Figure 3.37: MALDI-TOF-MS of Thio(Ser1) SIINFEKL. Observed Mass $[M+H]^+$ = 978.9 and $[M+Na]^+$ = 1001.0. Calculated mass: $[M+H]^+$ = 980.1, $[M+Na]^+$ = 1001.5, and $[M+K]^+$= 1017.5.

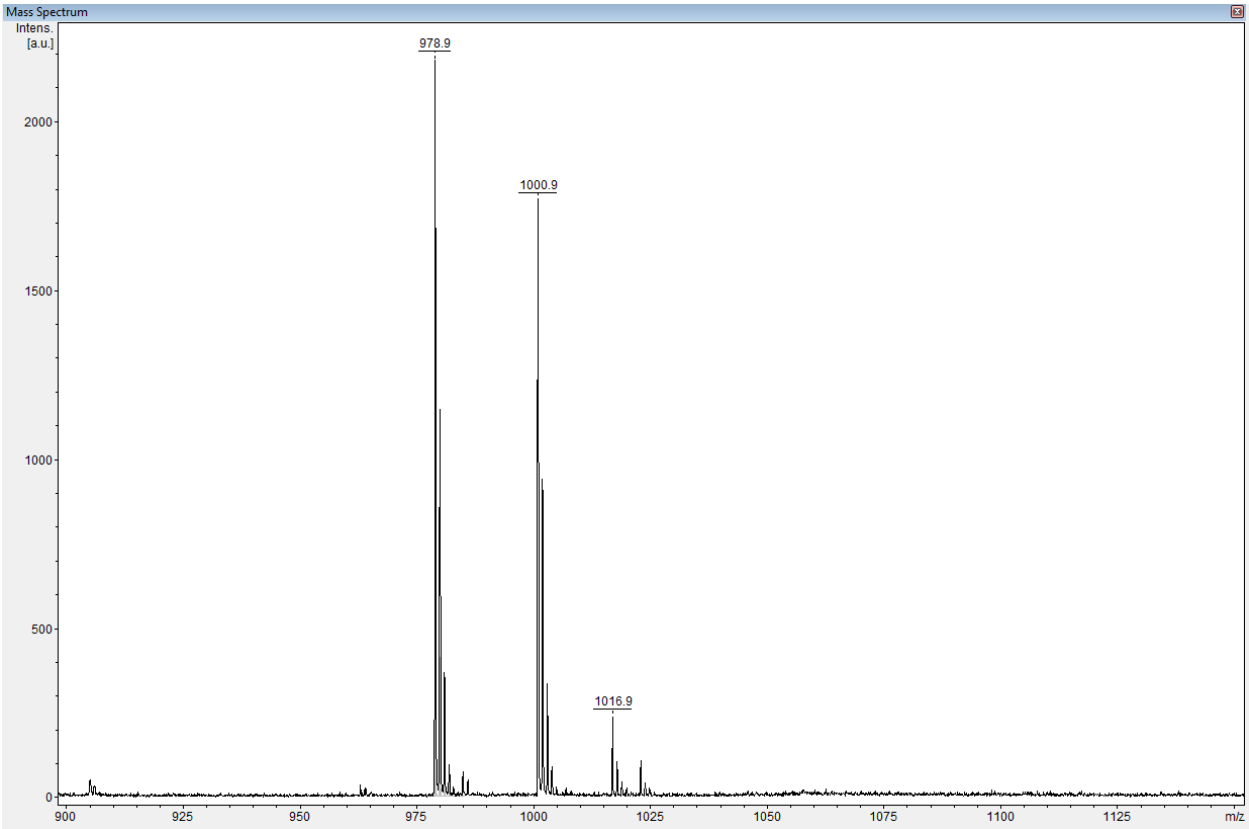


Figure 3.38: MALDI-TOF-MS of Thio(Ile2) SIINFEKL. Observed Mass $[M+H]^+$ = 978.9, $[M+Na]^+$ = 1000.9, and $[M+K]^+$= 1016.9. Calculated mass: $[M+H]^+$ = 980.1, $[M+Na]^+$ = 1001.5, and $[M+K]^+$= 1017.5.

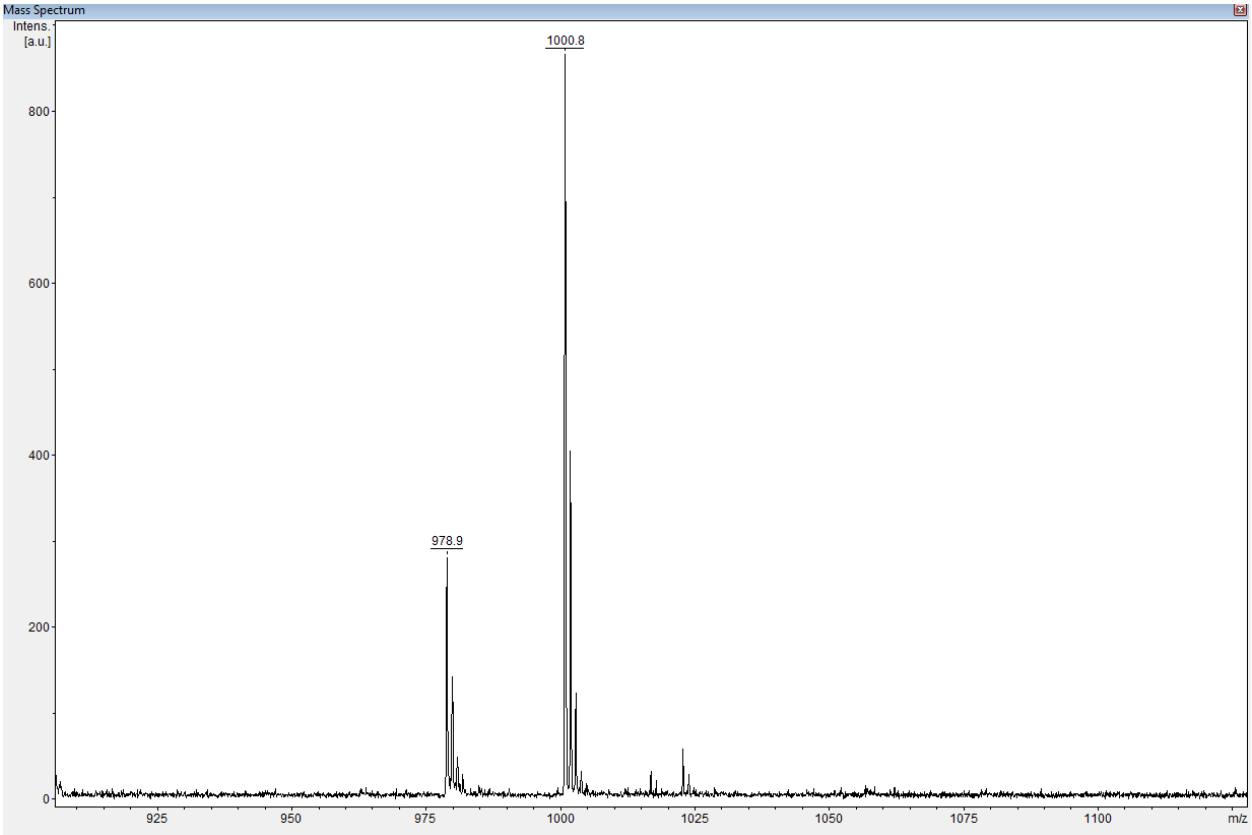


Figure 3.39: MALDI-TOF-MS of Thio(Ile3) SIINFEKL. Observed Mass $[M+H]^+$ = 978.9, and $[M+Na]^+$ = 1000.8. Calculated mass: $[M+H]^+$ = 980.1, $[M+Na]^+$ = 1001.5, and $[M+K]^+$= 1017.5.

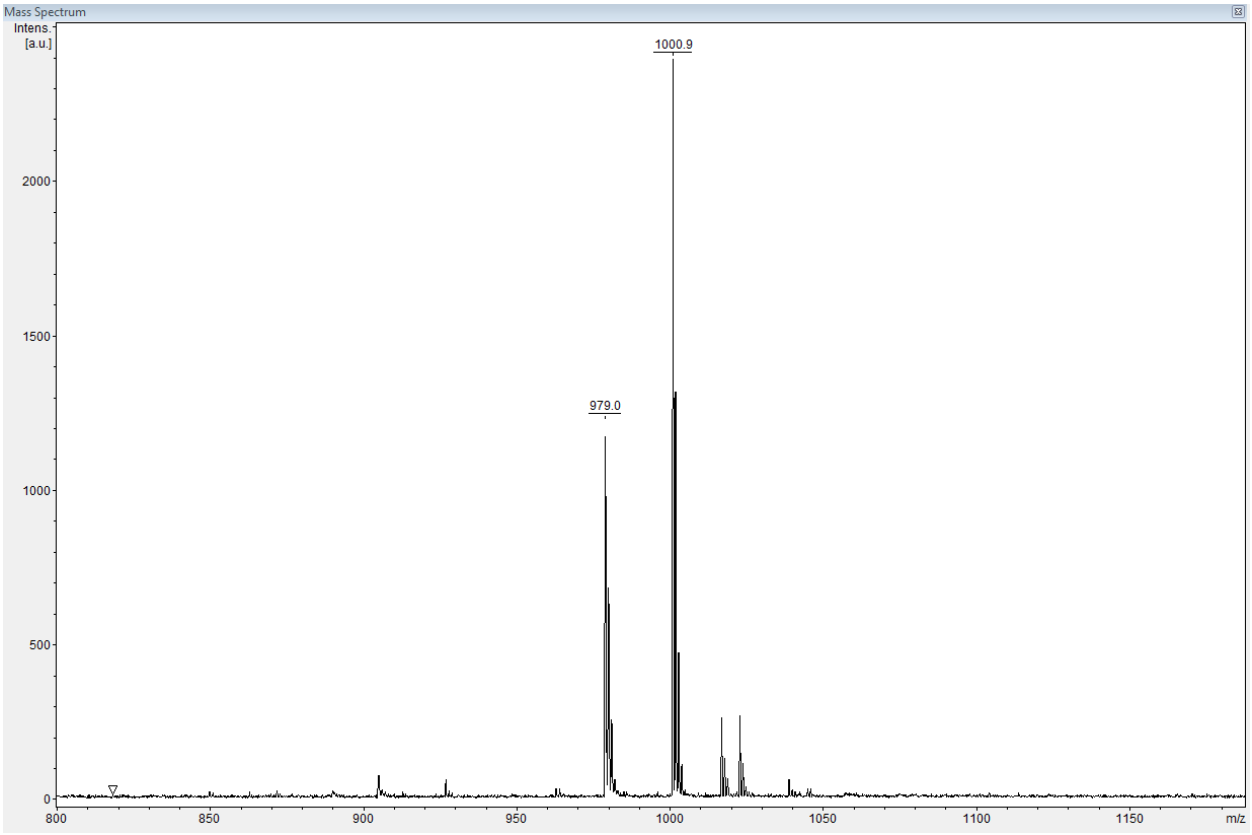


Figure 3.40: MALDI-TOF-MS of Thio(Asn4) SIINFEKL. Observed Mass $[M+H]^+$ = 979.0, and $[M+Na]^+$ = 1000.9. Calculated mass: $[M+H]^+$ = 980.1, $[M+Na]^+$ = 1001.5, and $[M+K]^+$= 1017.5.

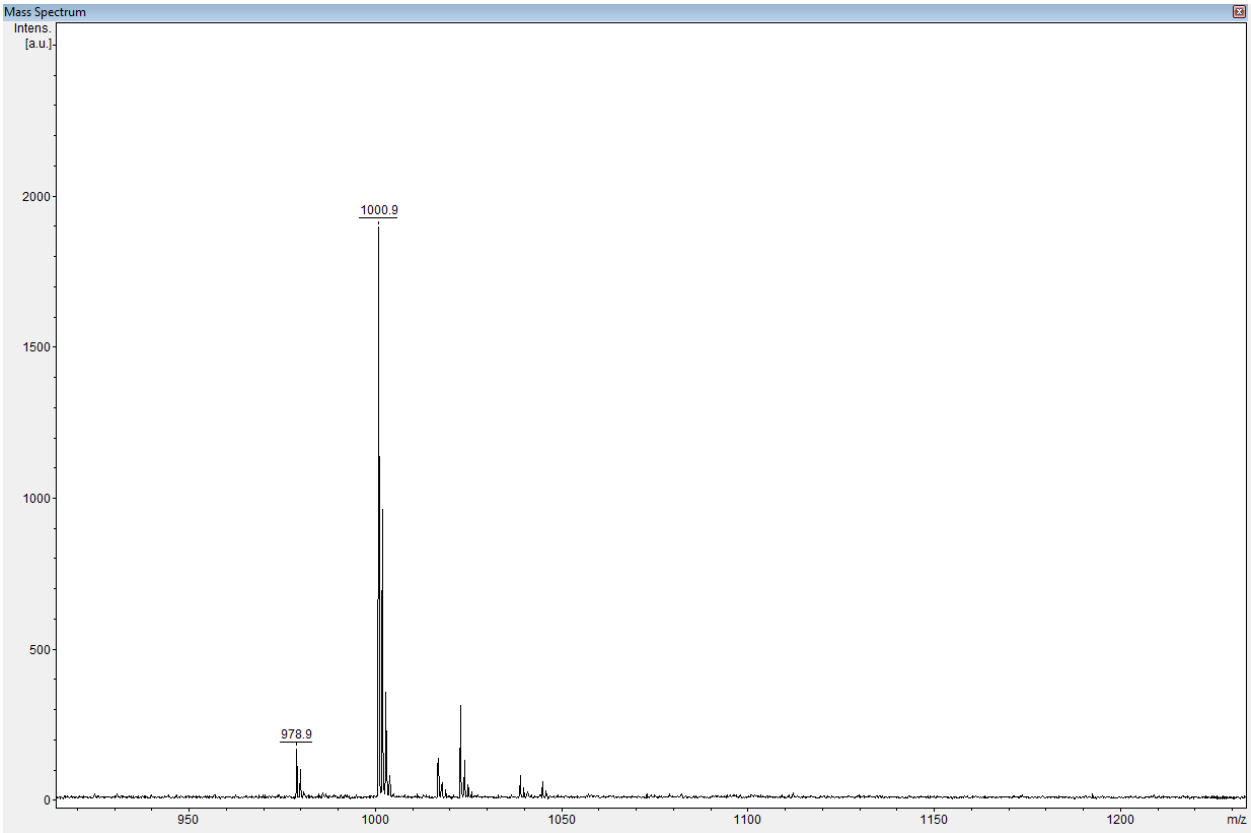


Figure 3.41: MALDI-TOF-MS of Thio(Phe5) SIINFEKL. Observed Mass $[M+H]^+$ = 978.9 and $[M+Na]^+$ = 1000.9. Calculated mass: $[M+H]^+$ = 980.1, $[M+Na]^+$ = 1001.5, and $[M+K]^+$= 1017.5.

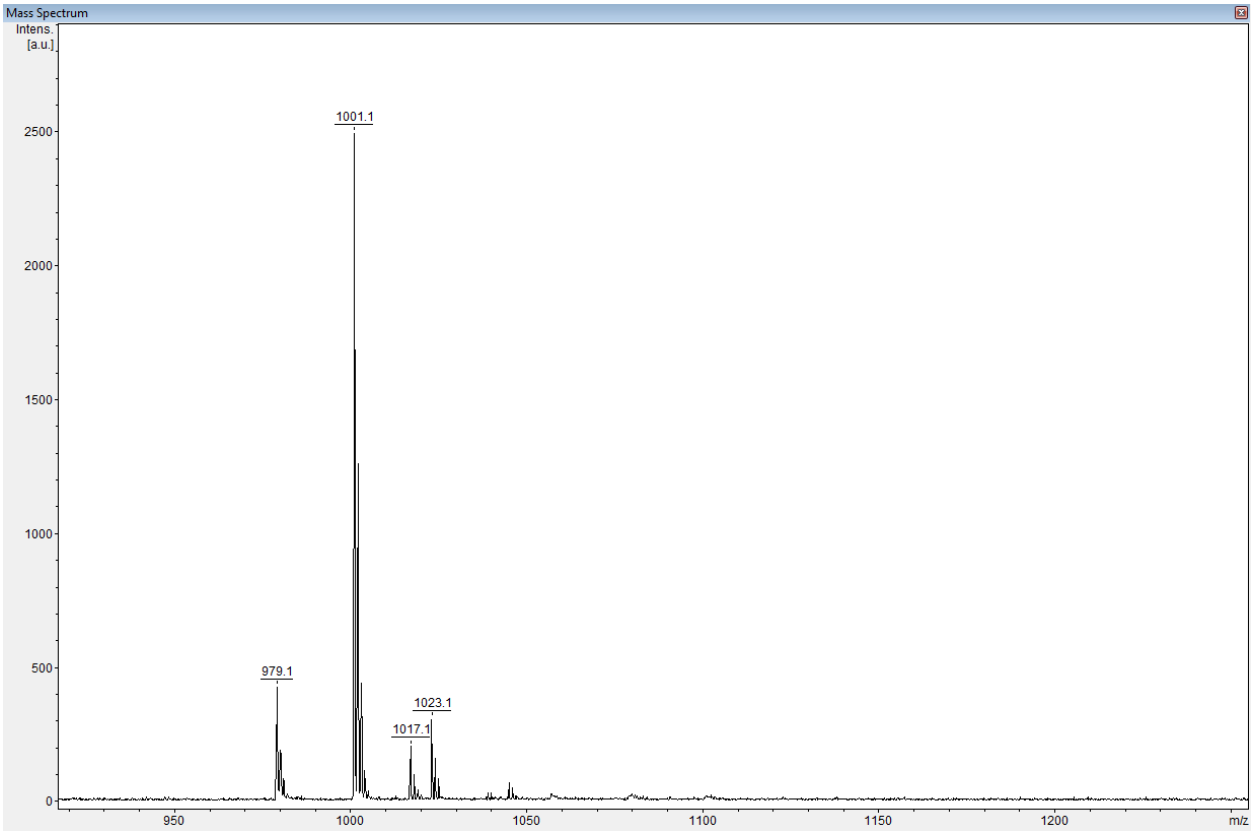


Figure 3.42: MALDI-TOF-MS of Thio(Glu6) SIINFEKL. Observed Mass $[M+H]^+$ = 979.1, $[M+Na]^+$ = 1001.1, and $[M+K]^+$= 1017.1. Calculated mass: $[M+H]^+$ = 980.1, $[M+Na]^+$ = 1001.5, and $[M+K]^+$= 1017.5.

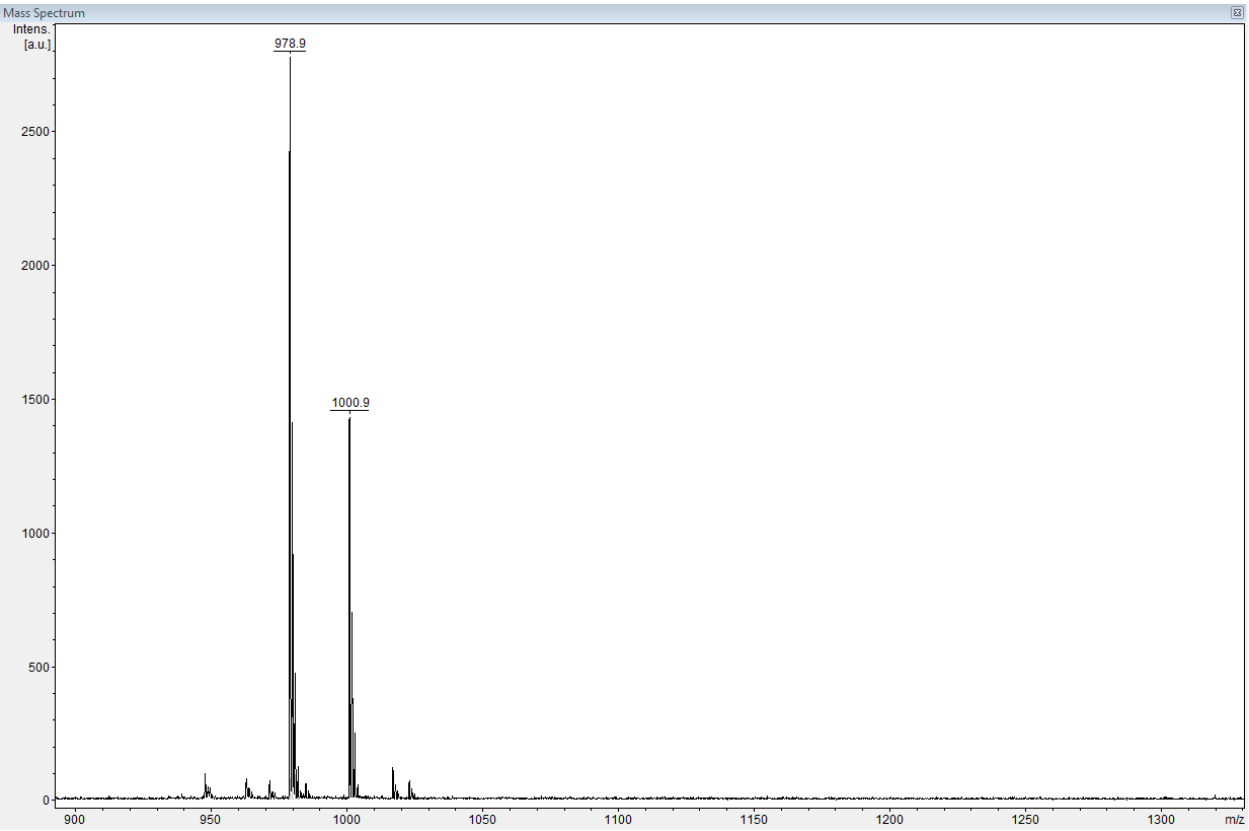


Figure 3.43: MALDI-TOF-MS of Thio(Lys7) SIINFEKL. Observed Mass $[M+H]^+$ = 978.9, and $[M+Na]^+$ = 1000.9. Calculated mass: $[M+H]^+$ = 980.1, $[M+Na]^+$ = 1001.5, and $[M+K]^+$= 1017.5.

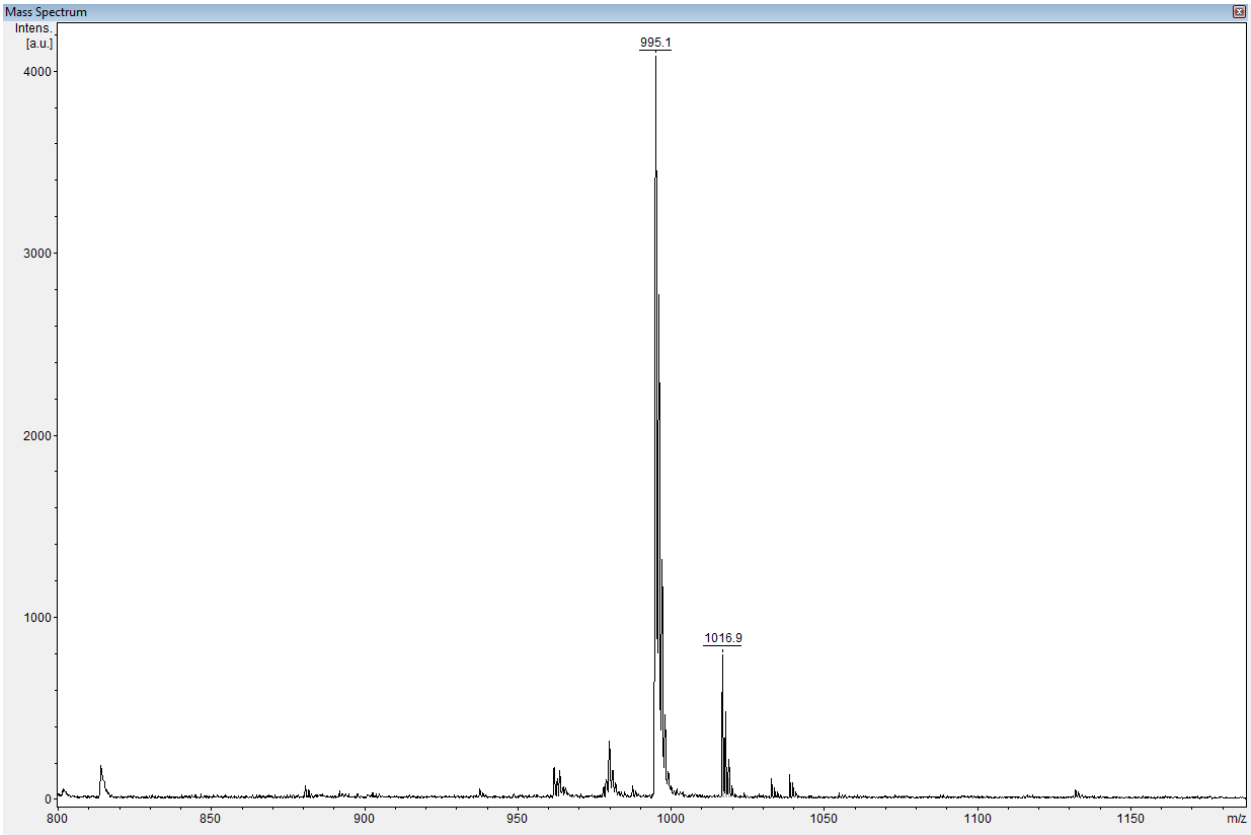


Figure 3.44: MALDI-TOF-MS of Thio(Ser1Lys7) SIINFEKL. Observed Mass $[M+H]^+$ = 995.1 and $[M+Na]^+$ = 1016.9. Calculated mass: $[M+H]^+$ = 995.5 and $[M+Na]^+$ = 1017.9

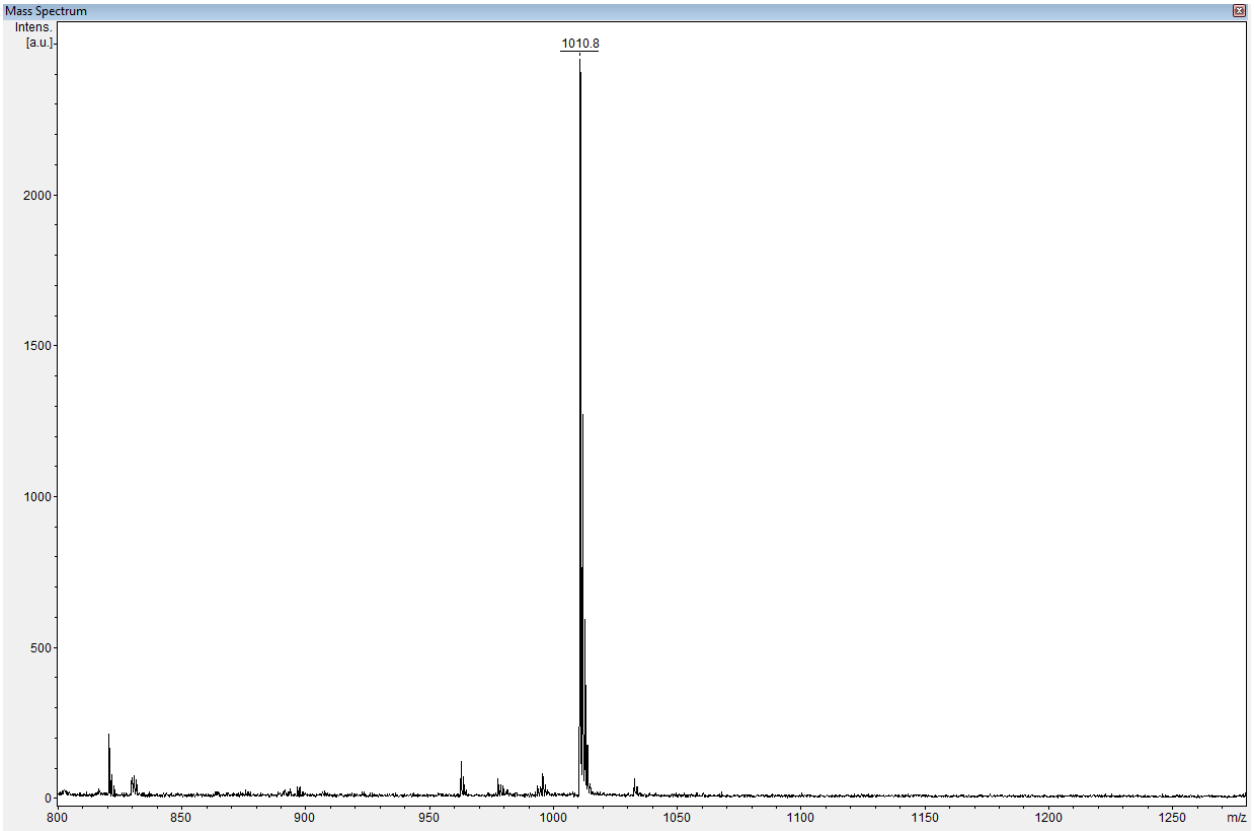


Figure 3.45: MALDI-TOF-MS of Thio(Ser1Ile3Lys7) SIINFEKL. Observed Mass $[M+H]^+$ = 1010.8. Calculated mass: $[M+H]^+$ = 1011.5.

3.8 – References

1. Schlake, T.; Thess, A.; Fotin-Mleczek, M. Kallen, K.J., Developing mRNA-vaccine technologies. *RNA biology*,**2012**, *9*(11), 319-1330.
2. Pardi, N.; Hogan, M.J.; Weissman, D.; Recent advances in mRNA vaccine technology. *Current opinion in immunology*,**2020**, *65*, 14-20.
3. Minor, P.D., Live attenuated vaccines: Historical successes and current challenges. *Virology*, **2015**b *479*, 379-392.
4. Torchilin, V.P.; Lukyanov, A.N., Peptide and protein drug delivery to and into tumors: challenges and solutions. *Drug discovery today*, **2003**, *8*(6), 259-266.
5. Henninot, A.; Collins, J.C.; Nuss, J.M., The current state of peptide drug discovery: back to the future?. *Journal of medicinal chemistry*,**2018**, *61*(4), 1382-1414.
6. Otvos Jr, L.; Wade, J.D., 2014. Current challenges in peptide-based drug discovery. *Frontiers in chemistry*, **2014**, *2*, 62.
7. Bergmann, N.C.; Davies, M.J.; Lingvay, I.; Knop, F.K., Semaglutide for the treatment of overweight and obesity: a review. *Diabetes, Obesity and Metabolism*, **2023**, *25*(1), 18-35.
8. Christou, G.A.; Katsiki, N.; Blundell, J.; Fruhbeck, G.; Kiortsis, D.N,. Semaglutide as a promising antiobesity drug. *Obesity Reviews*, **2019**, *20*(6), 805-815.
9. Frías, J.P.; Davies, M.J.; Rosenstock, J.; Pérez Manghi, F.C.; Fernández Landó, L.; Bergman, B.K.; Liu, B.; Cui, X.; Brown, K., Tirzepatide versus semaglutide once weekly in patients with type 2 diabetes. *New England Journal of Medicine*, **2021**, *385*(6), 503-515.

10. Jastreboff, A.M.; Aronne, L.J.; Ahmad, N.N.; Wharton, S.; Connery, L.; Alves, B.; Kiyosue, A., Zhang, S.; Liu, B.; Bunck, M.C.; Stefanski, A., Tirzepatide once weekly for the treatment of obesity. *New England Journal of Medicine*, **2022**, *387*(3), 205-216.
11. Karagiannis, T.; Avgerinos, I.; Liakos, A.; Del Prato, S.; Matthews, D.R.; Tsapas, A.; Bekiari, E., Management of type 2 diabetes with the dual GIP/GLP-1 receptor agonist tirzepatide: a systematic review and meta-analysis. *Diabetologia*, **2022**, *65*(8), 1251-1261.
12. Guichard, G.; Zerbib, A.; Le Gal, F.A.; Hoebeke, J.; Connan, F.; Choppin, J.; Briand. J.P.; Guillet, J.G., Melanoma peptide MART-1 (27−35) analogues with enhanced binding capacity to the human class I histocompatibility molecule HLA-A2 by introduction of a β-amino acid residue: implications for recognition by tumorinfiltrating lymphocytes. *J Med Chem.,* **2000**, *43(20)*, 3803–3808
13. Reinelt, S.; Marti, M.; Dédier, S.; Reitinger, T.; Folkers, G.; de Castro, J.; Rognan, D., β-amino acid Scan of a class I Major histocompatibility complex-restricted alloreactive T-cell epitope. *J Biol Chem.,* **2001**, *276(27)*, 24525–24530
14. Webb, AI.; Dunstone, MA.; Williamson, NA.; Price, JD.; de Kauwe, A.; Chen, W.; Oakley, A.; Perlmutter, P.; McCluskey, J.; Aguilar, M.I., T cell determinants incorporating β-amino acid residues are protease resistant and remain immunogenic in vivo, *The J Immunol*. **2005**, *175(6)*, 3810–3818.
15. Cheloha, RW.; Woodham, A.W.; Bousbaine, D.; Wang, T.; Liu, S.; Sidney, J.; Sette, A.; Gellman, S.H.; Ploegh, H.L., Recognition of class II MHC peptide ligands that contain beta-amino acids, *J Immunol.*, **2019**, *203(6)*, 1619–1628.

16. Gibadullin, R.; Randall, CJ.; Sidney, J.; Sette, A.; Gellman, S.H.; Backbone modifications of HLA-A2-Restricted antigens induce diverse binding and T cell activation outcomes. *J Am Chem Soc*, **2021**, *143 (17)*, 6470–6481.
17. Outlaw, V.K.; Cheloha, R.W.; Jurgens, E.M.; Bovier, F.T.; Zhu, Y.; Kreitler, D.F.; Harder, O.; Niewiesk, S.; Porotto, M.; Gellman, S.H.; Moscona, A., Engineering protease-resistant peptides to inhibit human parainfluenza viral respiratory infection. *Journal of the American Chemical Society*, **2021** *143*(15), 5958-5966.
18. Carmona, G.; Rodriguez, A.; Juarez, D.; Corzo, G.; Villegas, E., Improved protease stability of the antimicrobial peptide Pin2 substituted with D-amino acids. *The protein journal*, **2013**, *32*, 456-466.
19. Hamamoto, K.; Kida, Y.; Zhang, Y.; Shimizu, T.; Kuwano, K., Antimicrobial activity and stability to proteolysis of small linear cationic peptides with D-amino acid substitutions. *Microbiology and immunology*, **2002**, *46*(11), 741-749.
20. Molhoek, E.M.; Van Dijk, A.; Veldhuizen, E.J.; Haagsman, H.P.; Bikker, F.J., Improved proteolytic stability of chicken cathelicidin-2 derived peptides by D-amino acid substitutions and cyclization. *Peptides*, **2011**, *32*(5), 875-880.
21. Miller, S.M.; Simon, R.J.; Ng, S.; Zuckermann, R.N.; Kerr, J.M.; Moos, W.H., Comparison of the proteolytic susceptibilities of homologous L-amino acid, D-amino acid, and N-substituted glycine peptide and peptoid oligomers. *Drug Development Research*, **1995**, *35*(1), 20-32.
22. Mitchell, J.B.; Smith, J., D-amino acid residues in peptides and proteins. *Proteins: Structure, Function, and Bioinformatics*, **2003**, *50*(4), 563-571.

23. Insaidoo, F.K.; Borbulevych, O.Y.; Hossain, M.; Santhanagopolan. S.M.; Baxter, T.K.; Baker, B.M., Loss of T cell antigen recognition arising from changes in peptide and major histocompatibility complex protein flexibility: implications for vaccine design, *J Biol Chem.*, **2011,** *286(46)*, 40163–40173.
24. Signorelli, K.L.; Watts, L.M.; Lambert, L.E., The importance of DR4Dw4beta chain residues 70, 71, and 86 in peptide binding and T cell recognition, *Cell Immunol.* **1995**, *162(2)*, 217–224
25. Nietfield, W.; Bauer, M.; Fevrier, M.; Maier, R.; Holzwarth, B.; Frank, R.; Maier, B.; Riviere, Y.; Meyerhans, A., Sequence constraints and recognition by CTL of an HLA-B27-restricted HIV-1 gag epitope. *The J. Immunol.* **1995**, *154(5)*, 2189–2197.
26. Ostrov, D.; Krieger, J.; Sidney, J.; Sette, A.; Concannon, P., T cell receptor antagonism mediated by interaction between T cell receptor junctional residues and peptide antigen analogues. *J Immunol.* **1993**, *150(10),* 4277–4283.
27. Rock, K.L.,.; Reits, E.; Neefjes, J., Present yourself! By MHC class I and MHC class II molecules. *Trends in immunology*, **2016**, 37(11), 724-737.
28. Neefjes, J.; Jongsma, M.L.; Paul, P.; Bakke, O., Towards a systems understanding of MHC class I and MHC class II antigen presentation, *Nature reviews immunology*, **2011**, *11*(12), 823-836.
29. Rastogi, I.; Mannone, J.A.; Gibadullin, R.; Moseman, J.E.; Sidney, J.; Sette, A.; McNeel, D.G.; Gellman, S.H., β-amino acid substitution in the SIINFEKL antigen alters immunological recognition. *Cancer Biology & Therapy*, **2025**, *26*(1), 2486141.

30. Liu, C.; Barrett, T. M.; Chen, X.; Ferrie, J. J.; Petersson, E. J. Fluorescent Probes for Studying Thioamide Positional Effects on Proteolysis Reveal Insight into Resistance to Cysteine Proteases. *ChemBioChem*, **2019**, *20 (16)*, 2059−2062.
31. Asboth, B.; Polgar, L., Transition-State Stabilization at the Oxyanion Binding Sites of Serine and Thiol Proteinases: Hydrolyses of Thiono and Oxygen Esters. *Biochemistry* B, *22 (1)*, 117−122.
32. Yao, S.; Zutshi, R.; Chmielewski, J. Endothiopeptide Inhibitors of HIV-1 Protease, *Bioorg. Med. Chem. Lett.,* **1998**, *8 (6)*, 699−704.
33. Bartlett, P. A.; Spear, K. L.; Jacobsen, N. E. A Thioamide Substrate of Carboxypeptidase A. *Biochemistry,* **1982**, *21 (7)*, 1608− 1611.
34. Foje, K. L.; Hanzlik, R. P. Peptidyl Thioamides as Substrates and Inhibitors of Papain, and as Probes of the Kinetic Significance of the Oxyanion Hole, *Biochim. Biophys. Acta,* **1994,** *1201 (3)*, 447−453.
35. Verma, H.; Khatri, B.; Chakraborti, S.; Chatterjee, J., Increasing the Bioactive Space of Peptide Macrocycles by Thioamide Substitution, *Chem. Sci.* **2018***, 9 (9)*, 2443−2451.
36. Barrett, T. M.; Chen, X. S.; Liu, C.; Giannakoulias, S.; Phan, H. A. T.; Wang, J.; Keenan, E. K.; Karpowicz, R. J.; Petersson, E. J. Studies of Thioamide Effects on Serine Protease Activity Enable TwoSite Stabilization of Cancer Imaging Peptides. *ACS Chem. Biol*. **2020**, *15 (3)*, 774−779.
37. Chen, X.; Mietlicki-Baase, E. G.; Barrett, T. M.; McGrath, L. E.; Koch-Laskowski, K.; Ferrie, J. J.; Hayes, M. R.; Petersson, E. J. Thioamide Substitution Selectively Modulates

Proteolysis and Receptor Activity of Therapeutic Peptide Hormones. *J. Am. Chem. Soc.* **2017**, *139 (46)*, 16688−16695.

38. Gibadullin, R.; Morris, R.K.; Niu, J.; Sidney, J.; Sette, A.; Gellman, S.H., Thioamide Analogues of MHC I Antigen Peptides. *Journal of the American Chemical Society*, **2023,** *145*(47), 25559-25569.
39. Rosendahl Huber, S. K.; Luimstra, J. J.; Van Beek, J.; Hoppes, R.; Jacobi, R. H. J.; Hendriks, M.; Kapteijn, K.; Ouwerkerk, C.; Rodenko, B.; Ovaa, H.; De Jonge, J. Chemical Modification of Influenza CD8+ T-Cell Epitopes Enhances Their Immunogenicity Regardless of Immunodominance, *PLoS One,* **2016**, *11 (6)*, No. e0156462.
40. Soema, P. C.; Rosendahl Huber, S. K.; Willems, G. J.; Jacobi, R.; Hendriks, M.; Soethout, E.; Jiskoot, W.; de Jonge, J.; van Beek, J.; Kersten, G. F. A.; Amorij, J. P. Whole-Inactivated Influenza Virus Is a Potent Adjuvant for Influenza Peptides Containing CD8+ T Cell Epitopes. *Front. Immunol.,* **2018**, *9 (MAR)*, 525.
41. Jäger, E.; Höhn, H.; Necker, A.; Förster, R.; Karbach, J.; Freitag, K.; Neukirch, C.; Castelli, C.; Salter, R. D.; Knuth, A.; Maeurer, M. J. Peptide-Specific CD8+ T-Cell Evolution in Vivo: Response to Peptide Vaccination with Melan-A/MART-1, *Int. J. Cancer,* **2002**, *98 (3)*, 376−388.
42. Bins, A.; Mallo, H.; Sein, J.; Van Den Bogaard, C.; Nooijen, W.; Vyth-Dreese, F.; Nuijen, B.; De Gast, G. C.; Haanen, J. B. A. G. Phase I Clinical Study with Multiple Peptide Vaccines in Combination with Tetanus Toxoid and GM-CSF in Advanced-Stage HLA-A*0201-Positive Melanoma Patients, *J. Immunother.*, **2007**, *30 (2)*, 234−239.

43. Valmori, D.; Fonteneau, J.-F.; Lizana, C. M.; Gervois, N.; Liénard, D.; Rimoldi, D.; Jongeneel, V.; Jotereau, F.; Cerottini, J.-C.; Romero, P. Enhanced Generation of Specific Tumor-Reactive CTL In Vitro by Selected Melan-A/MART-1 Immunodominant Peptide Analogues., *J. Immunol.*, **1998**, *160 (4)*, 1750−1758.
44. Ebeling, W.; Hennrich, N.; Klockow, M.; Metz, H.; Orth, H.D.; Lang, H. Proteinase K from Tritirachium Album Limber, *Eur. J.Biochem*., **1974**, *47 (1)*, 91−97.
45. Tsoras, A.N.; Wong, K.M.; Paravastu, A.K.; Champion, J.A., Rational design of antigen incorporation into subunit vaccine biomaterials can enhance antigen-specific immune responses. *Frontiers in Immunology*, **2020**, *11,* 1547.
46. Dersh, D.; Yewdell, J.W.; Wei, J.; A SIINFEKL-based system to measure MHC class I antigen presentation efficiency and kinetics. *Antigen Processing: Methods and Protocols*, **2019**, 109-122.
47. Hulseberg, P.D.; Zozulya, A.; Chu, H.H.; Triccas, J.A.; Fabry, Z.; Sandor, M.; The same well-characterized T cell epitope SIINFEKL expressed in the context of a cytoplasmic or secreted protein in BCG induces different CD8+ T cell responses. *Immunology letters*, **2010**, *130*(1-2), 36-42.
48. Kamalov, M.; Kählig, H.; Rentenberger, C.; Müllner, A.R.; Peterlik, H.; Becker, C.F., Ovalbumin epitope SIINFEKL self-assembles into a supramolecular hydrogel. *Scientific Reports*, **2019**, *9*(1), 2696.
49. Falk, K.; Rötzschke, O.; Faath, S.; Goth, S.; Graef, I.; Shastri, N.; Rammensee, H.G., Both human and mouse cells expressing H-2Kb and ovalbumin process the same peptide, SIINFEKL. *Cellular immunology*, **1993** *150*(2), 447-452.

50. Mukherjee, S.; Verma, H; Chatterjee, J., Efficient site-specific incorporation of thioamides into peptides on a solid support. *Organic letters*, **2015**, *17*(12), 3150-3153.
51. Khatri, B.; Bhat, P.; Chatterjee, J., Convenient synthesis of thioamidated peptides and proteins. *Journal of Peptide Science*, **2020**, *26*(4-5), e3248.
52. Goldberg, J.M.; Batjargal, S.; Chen, B.S.; Petersson, E.J.; Thioamide quenching of fluorescent probes through photoinduced electron transfer: mechanistic studies and applications. *Journal of the American Chemical Society*, 2013, *135*(49), 18651-18658.
53. Ozturk, T.; Ertas, E.; Mert, O.; Use of Lawesson's reagent in organic syntheses. *Chemical reviews*, **2007** *107*(11), 5210-5278.
54. Khatoon, H.; Abdulmalek, E., A focused review of synthetic applications of Lawesson's reagent in organic synthesis. *Molecules*, **2021**, *26*(22), 6937.
55. Gayen, K.S.; Chatterjee, N., Diversity of Lawesson's reagent: Advances and scope. *Asian Journal of Organic Chemistry*, **2020**, *9*(4), 508-528.
56. Thomsen, I.; Clausen, K.; Scheibye, S.; Lawesson S.O., *Organic Syntheses*, 1990, *7*, 372.
57. Byerly-Duke, J.; VanVeller, B., Thioimidate Solutions to Thioamide Problems During Thionopeptide Deprotection, *Organic Letters*, **2024**, *26*, 1452-1457
58. Camacho, L.A.; Lampkin. B.; VanVeller, B., A Bottom-Up Approach to Preserve Thioamide Residue Stereochemistry During Fmoc Solid-Phase Peptide Synthesis, *Organic Letters*, **2019**, *21*, 7015-7018
59. Gjonaj, L.; Roelfes, G., Selective chemical modification of DNA with alkoxy-and benzyloxyamines. *Organic & Biomolecular Chemistry*, **2015** *13*(21), 6059-6065.

60. Lampkin, B.J.; VanVeller, B., Hydrogen bond and geometry effects of thioamide backbone modifications. *The Journal of organic chemistry*, **2021**, *86*(24), 18287-18291.
61. Deres, K.; Beck, W.; Faath, S.; Jung, G.; Rammensee, H.G., MHC/peptide binding studies indicate hierarchy of anchor residues. *Cellular immunology*, **1993** *151*(1), 158-167.
62. Koch, C.P.; Perna, A.M.; Pillong, M.; Todoroff, N.K.; Wrede, P.; Folkers, G.; Hiss, J.A.; Schneider, G., Scrutinizing MHC-I binding peptides and their limits of variation. *PLoS computational biology*, **2013,** *9*(6), e1003088.
63. Patsoukis, N.; Wang, Q.; Strauss, L.; Boussiotis, V.A., Revisiting the PD-1 pathway. *Science advances*, **2020** *6*(38), eabd2712.
64. Riella, L.V.; Paterson, A.M.; Sharpe, A.H.; Chandraker, A., Role of the PD-1 pathway in the immune response. *American Journal of Transplantation*, **2012** *12*(10), 2575-2587.
65. Jeon, D.; Hill, E.; Moseman, J.E.; McNeel, D.G., Combining toll-like receptor agonists with immune checkpoint blockade affects antitumor vaccine efficacy. *Journal for Immunotherapy of Cancer*, **2024** *12*(5), e008799.
66. Anthis, N.J.; Clore, G. M., Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm, *Protein Science*, **2013**, *22*, 851-8.