Doctoral Dissertation

博士論文

De novo discovery of bioactive disulfide-rich peptides

by means of *in vitro* selection

（多数のジスルフィド結合を持つペプチド薬剤探索）

A Dissertation Submitted for Degree of Doctor of Philosophy

December 2021

令和 3 年 12 月 博士（理学）申請

Department of Chemistry, Graduate School of Science,

The University of Tokyo

東京大学大学院理学系研究科化学専攻

LIU　WENYU

刘　文钰

Abstract

Disulfide-rich peptides (DRPs) are an expanding class of natural peptides containing more than 2 pairs of disulfide bonds. The knotted disulfide core is crucial for locking the peptide backbone into a constrained conformation and therefore it provides DRPs with remarkable proteolytic stabilities, target potency, selectivity, and/or cell permeability, making DRPs an attractive template for modern drug discovery. Current engineering methods for DRP molecules are mainly based on solid phase peptide synthesis (SPPS), which is capable of controlling the disulfide topology by taking advantage of protecting groups and auxiliaries. However, due to the structural complexity, chemical synthesis of DRPs requires massive laboratory work and leads to the difficulty of creating DRP library with large molecular diversity. On the contrary, display-based peptide screening (e.g. cell surface or phage display) is an alternative approach to devise ligands for pharmaceutical targets from a library of sequence-diverse DRPs. Nonetheless, the complexity of disulfide connectivity in DRPs leads to the difficulty of controlling the peptide architecture during cell-based display methods. The cell-free display method, Random non-standard Peptides Integrated Discovery (RaPID) system, is known to afford highly diverse libraries ($>10^{12}$ members) for affinity-based selection and may offer an improved route to identify DRPs with high potency and selectivity against proteins of interest. In addition, RaPID allows for incorporation of non-canonical amino acids (ncAAs) into peptides taking advantage of the Flexible *in vitro* translation (FIT) system, therefore making it possible to control disulfide conformation by introducing chemically modified cysteines during selection. The goal of this research is to design and construct diverse DRP-based peptide libraries with designed disulfide connectivity and develop bioactive DRP-based drug candidates against pharmaceutical targets using RaPID selection system.

In Chapter 2, I first assessed the feasibility of discovering bioactive DRPs using RaPID selection. I designed a peptide library based on the scaffold of a plant-derived cyclotide *Momordica cochinchinensis* trypsin inhibitor-II (MCoTI-II), which contains a cyclic backbone and 3 disulfide bonds. The library was linearized in loop 6 and contained randomized natural amino acids at all positions in loop 1 and 5, and 1 extra position in loop 6. RaPID selection was performed for affinity to human coagulation factor XIIa (FXIIa) immobilized on magnetic beads and following four rounds of selection, next-generation sequencing was performed to identify high frequency clonal sequences in the final cDNA library. To determine the significance of cyclic backbone on target binding and inhibition, both backbone acyclic (as panned during the selection) and cyclic (as existing in nature) forms of five most abundant selected peptides were synthesized by SPPS.

Surface plasmon resonance (SPR) and *in vitro* inhibitory assay revealed the most potent inhibitor, cyclic peptide MCoFx1 (cMCoFx1), with the dissociation constant (K_D) of 0.9 nM and the inhibition constant (K_i) of 0.37 nM toward FXIIa, which was 60-fold and 350-fold increase in FXIIa potency compared with native MCoTI-II. In general, backbone cyclic peptides were better inhibitors than their acyclic analogues (with less than 7-fold increase of potency upon cyclization), indicating that the knotted disulfide center had the most crucial influence on inhibitory potency. Further inhibitory specificity profiling of cMCoFx1 towards a panel of structurally and functionally related serine proteases, including human trypsin, FXa, FXIa, thrombin and plasmin, revealed cMCoFx1's high specificity for FXIIa. *In vitro* coagulation assay proved that cMCoFx1 selectively inhibited the intrinsic coagulation pathway in plasma. The superior stability of cMCoFx1 was then confirmed in human serum, with the degradation half-life of over 24 h, compared to a linear control peptide which was fully degraded within 1 h. Finally, the X-ray structure of co-crystalized cMCoFx1 and FXIIa was determined, revealing a tight binding of the peptide to the target in a substrate-like manner.

Although I successfully discovered a potent and selective FXIIa inhibitor without controlling the disulfide connectivity, the success was presumably due to the relatively easy folding nature of the template MCoTI-II. To utilize RaPID selection on other DRP scaffolds, the disulfide connectivity needs to be controlled during *in vitro* translation. Therefore, in Chapter 3, I decided to use stepwise deprotection method of protected cysteines and incorporated two acetamidomethyl-protected cysteines (Cys(Acm)) into a template peptide using FIT translation. A mild water-compatible Pd(2+) deprotection method was utilized to remove the Acm group and expose 2 free thiols to allow formation of the disulfide bond. Under the optimal deprotection conditions (20 mM PdCl$_2$ and 50 mM sodium diethyldithiocarbamate (DTC)), clean deprotection of both Acm groups and concomitant disulfide bond formation was observed. After confirming the utility of Cys(Acm) in FIT translation, bicyclic peptide libraries with globular and ribbon conformations were constructed. The library design called for spontaneous formation of the first macrocycle via a cyclization between N-terminal chloroacetyl (ClAc) group and a downstream Cys residue. Two remaining Cys, translated as Acm-protected derivatives, were then liberated upon the treatment with PdCl$_2$, which led to the formation of the second macrocycle. The mixed library which included 3, 4 or 5 random amino acids between the N-terminal ClAc-^{L}Tyr and each Cys was screened using RaPID selection against mouse/human asialoglycoprotein receptor (m/hASGR), a protein which promotes endocytosis and degradation of its ligand. As a result, the most potent identified peptide, masG1, bound to mASGR with the K_D of 1.37 nM. To further assess the significance of the bicyclic

constrains on target binding, a lariat form, masG1-lariat, was synthesized by substituting 2 Cys(Acm) with serine residues. Interestingly, masG1-lariat had over 100-fold weaker binding affinity compared with its parental molecule. Further evaluation of masG1 peptides in cell-internalization experiments is undergoing.

Next, in Chapter 4, I moved to designing a constrained tricyclic library containing 1 thioether and 2 disulfide bonds. In addition to Acm group, an orthogonal cysteine protecting group, phenylacetamidomethyl (Phacm), was introduced to facilitate stepwise deprotection and formation of 2 disulfide bonds. Phacm is labile under Acm deprotection conditions. Nevertheless, it can be selectively removed by penicillin G acylase (PGA) enzyme. I first assessed the compatibility of Cys(Phacm) in FIT translation by incorporating it into a model peptide and optimizing the deprotecting conditions. Addition of PGA enzyme resulted in full deprotection of both Phacm and the formation of a disulfide bond. Next, consecutive deprotection of Phacm and Acm was tested on another model peptide by incorporating 2 Cys(Phacm) and 2 Cys(Acm) during FIT translation. Mass shifts were observed after PGA and $PdCl_2$ treatment respectively, indicating a stepwise removal of 2 Phacm and 2 Acm groups and formation of the concomitant disulfide bonds accordingly. To construct the tricyclic libraries, similar to the bicyclic library design, a ClAc-^{D}Trp was incorporated at the N-terminus to facilitate the spontaneous ring closure with a downstream free cysteine. 2 Cys(Phacm) and 2 Cys(Acm) were incorporated to form a cystine knot conformation. In total, 3 tricyclic libraries were designed with diverse numbers of random amino acids between the ClAc-^{D}Tyr and each Cys (referred to loop in later context). In detail, Tricyclic_lib1 contained a mixture of 3, 4, and 5 random amino acids in each loop. Tricyclic_lib2 and 3 mimicked the loop length of natural occurred cyclotides, kalata B1 and MCoTI-II, respectively, with random amino acids substituting all residues in each loop. RaPID selection against SARS-CoV-2 main protease (M^{Pro}) was performed using 3 tricyclic libraries. Several target-binding tricyclic peptides were selected, and the binding affinity of chemically synthesized selected sequences was assessed by SPR measurement, with the most potent binder, MP327-peak *a*, showing a 38.6 nM K_D value. After inhibitory assay evaluation, I identified the most potent peptide, MP103-peak *a*, with IC50 value of 711 nM. Although all the selected peptides showed moderate binding and inhibitory ability, this work was a first trial in controlling the disulfide connectivity during high-throughput selection process.

In this research, I constructed 3 types of DRP-based peptide libraries with chemically protected cysteines and partially achieved disulfide connectivity control during FIT translation. The screening of cyclotide-based library proved the availability of RaPID selection on DRP

scaffolds and the following construction and screening of bicyclic and tricyclic peptide libraries with designed conformations suggested a further application of RaPID on *de novo* DRP discovery. In conclusion, this work is the first trial of structurally controlling peptide conformation via disulfide connectivity manipulation during *in vitro* translation and further application of this method would encourage discovery of more DRP based bioactive molecules.

Table of Contents

Chapter 3. Discovery of mASGR-binding bicyclic peptide with controllable globular conformation   52

Chapter 4. Construction of tricyclic peptide libraries using stepwise cysteine deprotection and RaPID selection against SARS-CoV-2 main protease   75

Chapter 1.

General introduction

1.1. Disulfide-rich peptides

Disulfide bonds play a significant role in defining both structure and function of peptides, stabilizing the peptide folding by decreasing entropy of the system.[1] Disulfide-rich peptides (DRPs), which are miniproteins containing more than two pairs of disulfide bonds, are particularly known for their constrained and well-defined three-dimensional structures. The compact structural traits endow DRPs with many beneficial properties, including exceptional thermal, chemical and proteolytic stability compared with linear peptides.[2] In addition, natural DRPs are reported to possess extraordinary target potency, selectivity, and/or cell permeability, making DRPs attractive scaffolds for designing peptide drug candidates.[2]

Natural DRPs are widely produced in bacteria, fungi, animals and plants. They normally contain up to 50 amino acids and incorporate 2–4 disulfide bonds. Depending on their folding patterns, DRPs can be roughly classified into several families, including knottins, hairpins, anti-microbial defensins, small conotoxins, and insulin.[3] Figure 1.1 shows some examples of natural DRPs. The knottin motif consists of more than six Cys residues to form disulfides, which adopt an interlocking arrangement often accompanied with backbone cyclization. The cyclic knottin form that has been extensively developed during the past two decades, cyclotide, has become a particular interest of researchers due to its distinct cyclic cystine knot (CCK) topology and its extensive bioactivities. Knottins have diverse functions ranging from plant defense[4] to predatory toxicity when expressed as toxins in venomous animals,[5] and have been reported to show low-immunogenicity.[6] The other widely studied DRP family, conotoxins, are a group of neurotoxic peptides containing multiple disulfide bonds, isolated from the venom of genus Conus.[7] As predatory venom peptides, conotoxins possess more diverse topologies and natural conotoxins normally inhibit voltage-dependent Na/K/Ca channels in the prey. This protein-interacting nature allows engineering conotoxins to be an efficient approach for therapeutic peptide discovery.

Although the majority of pharmacologically active DRPs is still in clinical stage, there are already several DRP drugs that are currently approved by the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMA).[8,9] In 2004, the FDA approved an atypical analgesic agent for the treatment of severe and chronic pain, Prialt, which is a synthetic form of the ω-conotoxin peptide.[10] In 2012, a three-disulfide containing DRP, linaclotide, was approved by the

FDA for the treatment of constipation-predominant irritable bowel syndrome and chronic idiopathic constipation.[11] In 2017, plecanatide containing two disulfide bonds was approved by the FDA for treatment of chronic idiopathic constipation.[12]

Since disulfide bonds play a crucial role in altering the conformations and functions of DRPs, the correct disulfide connectivity is essential for expressing the correct bioactivities. Folding of natural DRPs via disulfide formation can be affected by both intramolecular interactions between residues within the peptides, and intermolecular interactions with protein factors, such as the protein disulfide isomerases (PDIs). The inherent amino acid sequences induce noncovalent intramolecular interactions, such as hydrogen-bonding, hydrophobic or electrostatic interactions within the peptide, facilitating the folding and bringing close Cys residues to form disulfide bonds. On the other hand, in most organisms that natural DRPs are discovered, i.e. plants and toxin-producing predators, the intermolecular interactions between the peptide and PDI plays a significant role in folding and disulfide formation. PDI is a family of oxidoreductase enzyme that has a major role in oxidative folding of polypeptides in the endoplasmic reticulum (ER) of eukaryotic cells and functions as an ER chaperone.[13] Although the exact mechanism of PDI is still unclear, the enzyme is believed to interact with the substrate peptide at all stages along its folding pathway through hydrophobic interactions and forms (oxidizes), breaks (reduces), and/or shuffles (isomerizes) disulfide bonds in substrate via a thiol-disulfide exchange between its active-site and the substrate.[14,15]

In contrast, different from the enzyme-driven *in vivo* disulfide formation in eukaryotic ER, chemical synthesis of DRPs in laboratories requires efficient and robust synthetic methods for the formation of correct disulfide connectivity. In practice, laboratory chemical synthesis of peptides utilizes the solid phase peptide synthesis (SPPS). However, without catalytical reactions utilizing PDI or other oxidoreductases, the spontaneous disulfide formation takes place in several hours to a few days. Modern laboratory DRP synthesis takes advantage of SPPS incorporating various approaches to facilitate the disulfide formation. Most commonly, the peptide is subjected to oxidative folding conditions in a redox buffer, e.g. cysteine/cystine or reduced/oxidized glutathione. This oxidative folding procedure can produce the single thermodynamic peptide isomer and is especially common in the chemical synthesis of natural cyclotides. However, for the engineered sequences, the obtained products are often formed as a complex mixture of different topological isomers and need to be further purified, with extra effort to determine the correct folding fraction.

Another approach to chemically synthesize DRPs with the correct disulfide connectivity applies pairwise protection/deprotection of cysteines using orthogonal protective groups, which

ensures product formation as a single topological isomer. This strategy has been widely used in large-scale industrial synthesis of DRPs. Further discussion of commonly used Cys protecting groups can be found in Chapter 3.1 and 4.1. It can also be challenging to establish the disulfide pairing of the synthetic products. X-ray crystallography, NMR, and mass spectroscopy are the dominant experimental tools in elucidating such structures.
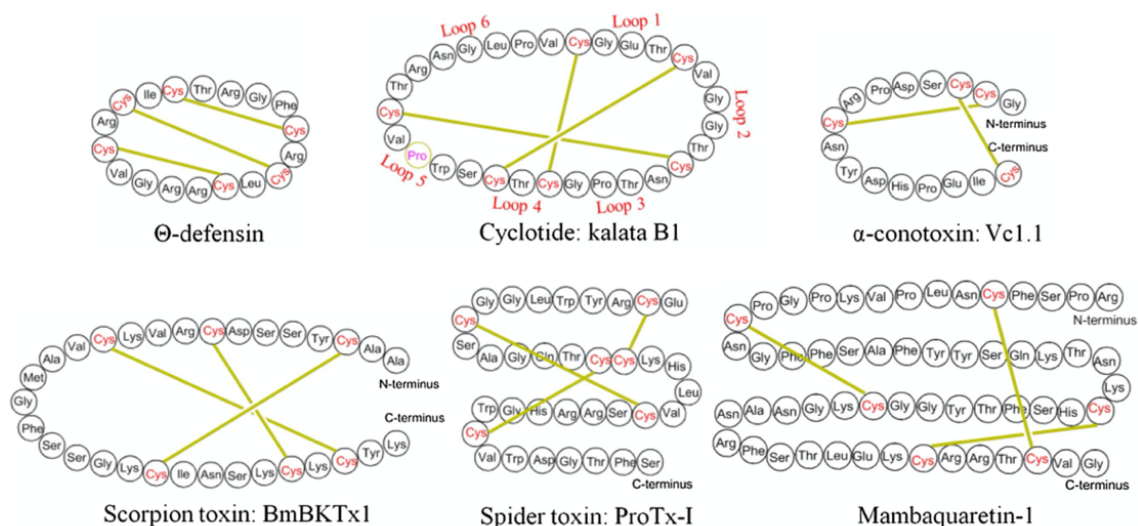
Figure 1.1: Examples of peptide sequence and disulfide connectivities of naturally occurring DRPs. Reproduced with permission from ref. 3 Copyright Elsevier.[3]

1.2. Current discovery methods of DRPs

DRPs are of particular interest in pharmaceutical compound discovery due to their rigid structures` and the resulting biological properties. The existing DRP scaffolds that have been studied and engineered can be divided into two categories: backbone acyclic and cyclic, with the former lacking the head-to-tail cyclization with the amide bond. Although the disulfide bonds in both the acyclic and cyclic DRPs play a key role in locking the peptide into unique folded conformations, the cyclic scaffolds are more commonly identified in nature. Natural cyclic DRPs include θ-defensins of 18 amino acids with three disulfide bonds in a laddered arrangement[16] and cyclotides of around 30 amino acids with three disulfide bonds in a cyclic cystine knot (CCK) configuration.[17,18] The natural cyclic DRPs are ribosomally synthesized as linear precursors and then post-translationally modified into their mature cyclic forms in the host organisms (Figure 1.2).[19] However, so far, we have little knowledge of the enzymes or other processes that cleave the

mature peptide from its precursor and facilitate formation of a cyclic backbone.[19] The ultrastable cyclic structures might have evolved because of the advantages they provide for stability, potency and adaptability to changing pest or pathogen pressures. However, some acyclic DRPs, especially members of the inhibitor cystine knot (ICK) family, possess similar level of bioactivity as their cyclic counterparts.[20,21]

To take advantage of this disulfide facilitated folded structure (acyclic or cyclic), many studies have been reported on engineering DRP scaffolds with extrinsic bioactivities. However, despite extensive efforts, the generation of designer DRPs with novel activity and specificity against pharmaceutical targets has proven to be challenging. One of the key obstacles is the complex nature of the DRP scaffold, with difficulties in high-throughput production of DRP libraries placing a limitation on their accessible sequence or chemical diversity. In the past two decades, the advanced synthetic methods, development of computational chemistry, and screening techniques have resulted in the discovery of many DRP-based bioactive molecules. In this chapter, I will briefly introduce the engineering and screening methods of the DRP scaffolds.
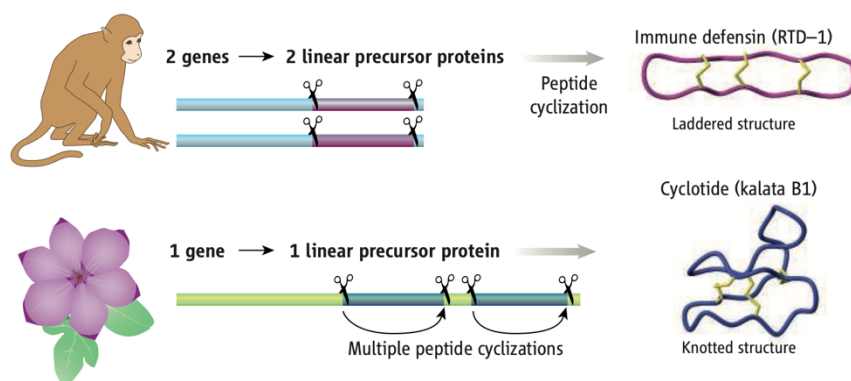
Figure 1.2: Gene-encoded cyclic DRPs are produced as fragments of linear precursor proteins that are excised and spliced head-to-tail. In the case of rhesus-defensin-1 (RTD-1) (top), two genes each code for half of the 18–amino acid mature peptide and a double head-to-tail ligation produces the circular peptide. In the case of cyclotides (bottom), a cystine knot embedded in the cyclic backbone provides extra stabilization. Reproduced with permission from ref. 19 Copyright Nature Publishing Group.[19]

1.2.1. Chemical synthesis based DRP discovery

In the past few decades, chemical synthesis-based approaches for DRP discovery have yielded many bioactive DRPs. These approaches include rational design, loop grafting or scaffold manipulation of existing DRP molecules and have been extensively explored in the past few

decades. For example, several cyclotide-based protease inhibitors have been generated through rational approaches designed to bias the specificity of potent natural trypsin inhibitors (the *Momordica cochinchinensis* derived cyclotides, MCoTI-I and MCoTI-II) towards other serine protease targets, for example the blood coagulation factors.[22-25] One particularly appealing target in this respect is coagulation factor XIIa (FXIIa), a trypsin-like serine protease that initiates the intrinsic pathway of the coagulation system. Previously, FXIIa inhibitors based on MCoTI-II were generated by screening a limited number of synthetic cyclotides bearing mutations in loop 1 (the primary binding loop of trypsin inhibitor cyclotides).[25] However, the potency of these molecules against FXIIa is ~1000-fold lower than that shown by MCoTI-II against trypsin (K_i = 30 pM) [23] and they have poor selectivity. Other cyclotide-based inhibitors of disease-related serine proteases, including β-tryptase, neutrophil elastase and matriptase, have also been reported[22–24] and, as with the FXIIa inhibitors, none have been sufficiently potent and selective to warrant further development.

Another chemical synthesis-based DRP engineering method is the 'grafting' method, which involves introducing peptide epitopes with known target-binding functionality into the loops of DRP scaffolds to generate variants with novel target tropism. This method has been broadly used in designing novel cyclotide variants, due to their relatively large size allowing more flexibility for loop grafting.[26–30] Several bioactive epitopes relevant to human diseases, including HIV,[26] cancer[27,30] and pain,[31] as well as some performance-boosting sequences, such as cell-penetrating peptides (CPPs),[32] have been engineering into various cyclotide scaffolds including the mobius and trypsin inhibitor cyclotides. However, such engineering has largely relied on pre-existing peptide epitopes that are not necessarily optimal for cyclotides, and the resulting molecules have exhibited relatively modest levels of activity, *e.g.* dissociation or inhibitory constants that generally fall in the micromolar or sub-micromolar range.

Some engineered DRPs were developed by altering their natural scaffolds. Since many of the discovered bioactive natural DRPs exhibit head-to-tail cyclic structure, which is accessed post-translationally from their linear analogues, it can be assumed that the cyclic structures have evolved because of the advantages they provide for stability, potency and adaptability to the natural targets. Some of these advantages can be transferred to natural backbone acyclic DRPs to create artificially engineered cyclic scaffolds by adding amino acid linkers to bridge their termini. A few DRP scaffolds that were synthetically cyclized include conotoxins MII and Vc1.1 with two disulfide bonds and the cyclic chlorotoxin with four disulfide bonds.[18,33–35] The cyclization was

5

proved to not interfere with the original target-binding properties, whereas improve the peptides' stability in serum.

1.2.2. Enzymatic cyclization and biosynthesis assisted DRP discovery

Due to the complex structure including disulfide guided folding and/or cyclic backbone, chemical synthesis of DRPs always require additional modifications. On the other hand, enzyme-assisted cyclization allows for fast and accurate ring closure with amide bond, providing more approaches for DRP discovery. Discovery of ligase type of asparaginyl endopeptidases (AEPs) in various plant species revealed one possibility of post-translational cyclization mechanism in plant-derived cyclic DRPs.[36] For example, the cyclization step during the biosynthesis of cyclotide kalata B1 was shown to be catalyzed by an AEP (OaAEP1b) in *Oldenlandia affinis*, a native African plant species.[37] Another example of a cystine knotted cyclic peptide is Cter-M, which is cyclized by an AEP in butterfly pea plants (*Clitoria ternatea*), dubbed butelase 1.[38] Other AEPs such as HaAEP1 from common sunflower (*Helianthus annuus*) and MCoAEP2 from gac fruit plants (*Momordica cochinchinensis*) have been reported to facilitate backbone cyclization during the biosynthesis of cyclic trypsin inhibitor peptides SFTI-1 and MCoTI-II, respectively.[39,40] Another type of post-translational transpeptidase sortase A, which is expressed in many Gram-positive bacteria, has been investigated for backbone cyclization in various protein engineering applications.[41] In bacteria, sortase A recognizes and cleaves a five-residue sequence, LP(X)T↓G, at the C-terminus of a target protein, generating an enzyme–substrate complex, which is subsequently replaced by a second peptide segment containing multiple N-terminal Gly residues via nucleophilic attack, resulting in ligation of the two fragments.[41,42] This mechanism can be adapted to cyclize peptides and proteins by engineering the target peptide sequence to contain the required N-terminal poly-Gly and C-terminal LP(X)TG motifs.[43] Using this strategy, linear kalata B1 and sunflower trypsin inhibitor-1 (STFI-1) were cyclized *in vitro* using recombinant *Staphylococcus aureus* sortase A,[44] and a cyclic conotoxin, PVIIA, was recently produced using an engineered sortase A variant, SrtA5o.[45] Since backbone cyclizing enzymes have only been discovered within the past decade, few novel engineered DRPs have been reported taking advantage of the chemoenzymatic synthesis. However, enzyme reaction provides possibility of creating efficient water-compatible backbone cyclization system for a large number of substrates, and therefore could encourage future cyclic DRP discovery.

Another biosynthetic cyclization method that has been extensively explored is the intein involved reaction. Intein was first identified in yeast where it catalyzes post-translational protein splicing reaction, during which an internal segment (intein) is excised from a protein and the two

flanking segments (exteins) are subsequently ligated.[46,47] To utilize intein as a DRP cyclizing enzyme, the cyclic peptide is expressed starting at a cysteine residue, with an upstream N-terminal methionine residue (required for translation), and a C-terminal intein segment (Figure 1.3a).[48] The N-terminal methionine is removed by Met aminopeptidase to expose the cysteine residue, which subsequently undergoes a thiol exchange with a thioester on the engineered intein, followed by S-N acyl transfer to yield backbone cyclization. Taking advantage of this cyclizing method, several DRPs, such as kalata B1, MCoTI-II, SFTI-I and RTD-1 have been successfully expressed in native or engineered *E. coli*.[48–51] Similar to the intein ligation method, another iteration of the intein-based method is protein trans-splicing. This strategy derives from the unusual biosynthesis of the DnaE protein in *Synechocystis* sp. PCC6803, which involves ligation of two separate proteins via a split intein.[52,53] In this strategy, two proteins are fused with the N or C-terminal segment of the DnaE intein, followed by self-association of the two intein fragments and formation of a functional intein unit, which undergoes self-processing and leads to excision of the intein domain followed by ligation of the two DnaE extein segments.[52] Using the same strategy, backbone cyclization can be achieved by expressing the split inteins on a single protein construct, where the target DRP peptide sequence is inserted between the N-terminal (I_N) and C-terminal (I_C) intein subunits (Figure 1.3b).[54] Cyclic DRPs, STFI-1 and MCoTI-II have been efficiently expressed in engineered *E. coli* with the correct backbone cyclization, and the *in vivo* production of a variant of MCoTI-I retrieved its native bioactivity of inhibiting cytotoxicity induced by cooverexpression of α-synuclein.[55] This development highlights the great potential of intein-based strategies, particularly protein trans-splicing, for generating genetically encoded libraries of cyclic peptides that can be screened for novel biological activities.[41]
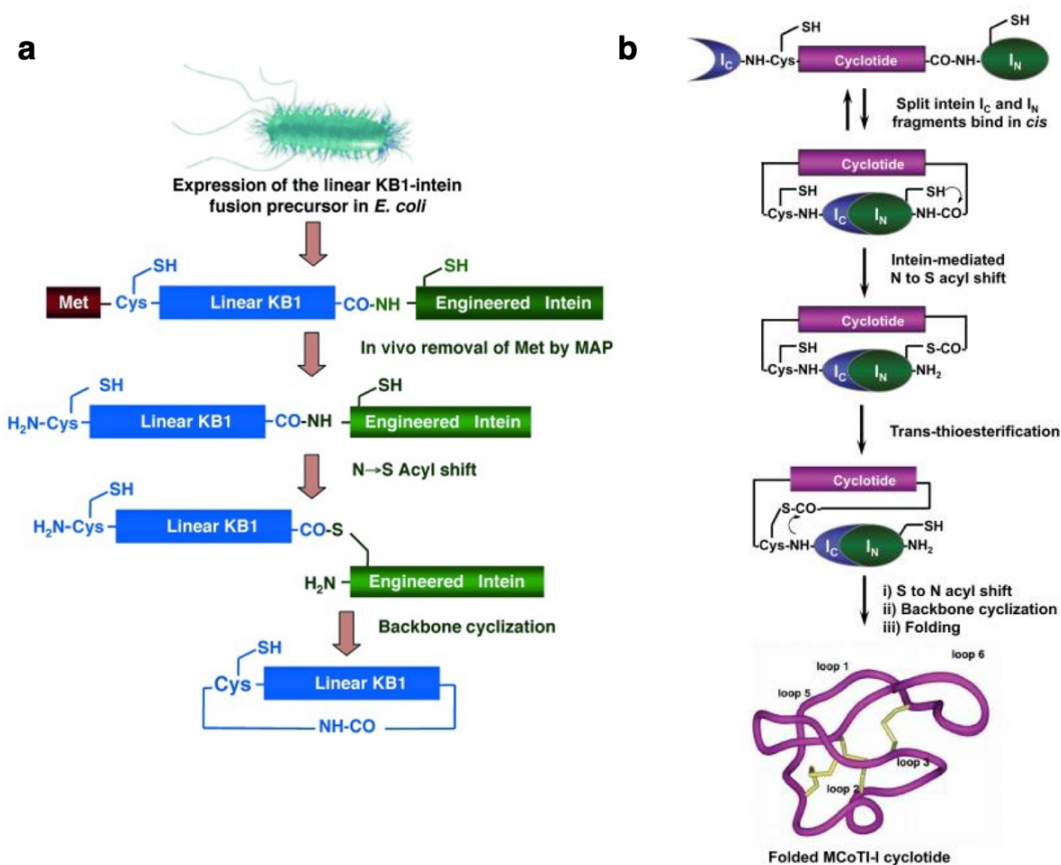
Figure 1.3: Intein-assisted biosynthesis of DRPs. (a) Recombinant expression of kalata B1 using intein ligation method. The engineered intein was expressed at the C-terminus; (b) Biosynthesis of MCoTI-I following the protein trans-splicing method. The MCoTI-I sequence was expressed in between C-terminal (I_C) and N-terminal (I_N) intein subunits. After self-association of two intein segments, the core peptide went through the intein cyclization mechanism. Figure (a) is reproduced with permission from ref. 48 Copyright 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim;[48] Figure (b) is reproduced with permission from ref. 54 Copyright 2017, Springer Science Business Media New York.[54]

1.2.3. Computational chemistry assisted rational design

In another aspect, in recent years, exploiting computational modeling to predict and solve problems in protein/peptide folding, protein-protein interaction, protein engineering and drug discovery has gained significant attention. With the development of efficient data mining algorithms and advanced dynamical force fields, the computational methods can model even the most complex systems, and evolve structure and dynamics with unabated accuracy.[56] Development

of computational technologies benefits the discovery of DRP-based pharmaceutical molecules in not only determining and analyzing structure-activity-relationship of the molecules, but also understanding and predicting theoretical folding pathways, isomer formation or interaction mechanism between natural/engineered DRPs and their pharmaceutical targets. Taking advantage of the publicly accessible or open-source DRP databases, *e.g.* ClanTox,[57] ConoServer,[58] and Cybase,[59] it has become more accessible and efficient to obtain the sequence resources, paving the way for studying the structure-function relationships of biological systems over wide temporal and spatial scales.[8]

One application of computational tools for DRP discovery is by combination with genome mining to predict functions of novel conotoxin sequences through machine learning. The details of predicting functions of conotoxins have been extensively reviewed recently.[60] In brief, among the 6254 identified conotoxin sequences in the ConoServer, only 243 peptides show identified pharmacological targets (the data was collected up to 2019).[61,62] Therefore, predicting bioactivities of the undefined conotoxins has become of great interest. Plenty of peptide representations and classification algorithms have been created and tested to optimize the performance in machine learning. For example, representing conotoxins by their amino acid composition and dipeptide composition, which is the composition based on neighboring amino acid pairs, incorporates more information about the correlation between peptide function and the amino acid composition.[63,64] Recent feature selecting approaches including the binomial distribution,[65] the relief algorithm[66] and the f-score algorithm[67] have improved the performance of the machine learning processes.[68] However, due to the limited number of conotoxins with defined functions, the current machine learning studies of conotoxins often utilize about 100–150 sequences as the training data, which is not sufficient due to the high variability of conotoxin sequences.[68]

Other than machine learning, docking studies and molecular dynamics (MD) simulations are computational simulation techniques used to study DRPs at the molecular level. Docking studies provide an in silico representation of the binding between a protein and its ligand (DRP peptide in this case), which can be used to predict binding energies and affinities.[69] On the other hand, MD simulations can complement docking approaches for the understanding and predicting the protein DRP interaction mechanism, entropy and dynamical fluctuations. A prevailing application of the two simulating methods on DRP discovery is the structure refinement of experimentally obtained X-ray crystal structure or NMR structure of DRP molecules or the complex with target proteins. For example, MD simulations were used to refine the three-dimensional structures predicted from NMR data of conotoxin BtIIIA[70] and a conotoxin EIVA[71] in complex with calcium atoms.[72] In

addition, docking and MD simulations have been employed to explain and predict the binding affinity and selectivity of proteins and DRP molecules, which provide hints for developing more potent DRP molecules by simulating the entropy and enthalpy around specific residues in the substrate. For example, docking studies were used to elucidate the preference of positively charged residues on conotoxin κM-RIIIK by a potassium ion channel, *TSha1*, due to the favored electrostatic interactions between the positive residues and potassium channel.[73]

1.2.4. Display-based bioactive DRP discovery

Peptide display-based selection is an alternative approach to discover ligands for pharmaceutical targets from a library of sequence-diverse DRPs. For example, the cell surface display technologies, *E. coli* and yeast surface display systems, have been used to obtain DRP mutants with binding affinity to integrin receptors and CTLA-4, and inhibitory activity to α-synuclein aggregation, matriptase-1 and neuropilin-1.[55,74–81] These approaches provide larger sequence diversity than chemical synthesis-based methods by introducing randomised codons at designated positions at the nucleotide level. However, for display purposes, these scaffolds are typically linearized no matter naturally acyclic or cyclic. For example, to construct cyclotide-based libraries, loops 2 or 6 of cyclotide scaffolds were disconnected to allow attachment to the display substrate.[78–81] Figure 1.4 indicates an example of a kalata B1-based peptide library displayed on *E. coli* surface, with loop 2 linearization.[79] Once an acyclic hit is obtained, it later can be cyclized to improve stability. These cell surface display approaches have been especially applied on cyclotide scaffolds, generating cyclotide variants against matriptase-1 and neuropilin-1 where the most active mutants have shown impressive activities (K_i = 0.83 nM to matriptase-1 using the MCoTI-II scaffold and K_D = 34 nM to neuropilin-1 using the kalata B1 scaffold).[79–81] Despite the remarkable inhibitory activity observed in the former case, the selectivity over trypsin (K_i = 36 nM) was limited to 40-fold. Yeast surface display has also been combined with rational design and loop grafting. For instance, several potent knottin-based integrin binders (with single digit nanomolar K_D) have been developed using display libraries designed by grafting an integrin-binding motif to a DPR scaffold, which was partially randomized and displayed on yeast surface.[74–76] These combinatorial applications take advantage of molecular grafting method, meanwhile allow sequence optimization endowed by cell-surface display, leading to discovery of potent integrin binders with the rigid DRP scaffold. Although the theoretical diversities of the *E. coli* or yeast display systems could be up to 10^9 variants, it is difficult to accurately estimate the expression of correctly folded cyclotide variants in such cellular systems. In addition, the periplasmic disulfide

formation or peptide folding may interfere with the cellular export efficiency.[79] This fact might have given a lower diversity than the theoretical value, yet resulting in non-optimal cyclotide variants.

Another display technology used for DRP discovery is phage display, which has been applied on several DRP scaffolds.[82-86] For example, a miniprotein Min-23, which was rational generated from *Ecballium elaterium* trypsin inhibitor II (EETI-II), was used as the phage display scaffold, yielding discovery of angiogenesis marker delta-like ligand 4 (Dll4)-binding sequence.[83] However, similar to the cell display approaches, phage display also takes advantage of cell expression systems. In order to display a properly folded and functional DRP, post-translational modifications by the expression host should be kept to a minimum.[84] On the other hand, the random sequences displayed in the DRP scaffold may not be optimal for accurate folding and conformational stability during cell expression. For DRP scaffolds with more than two disulfides, the expression host and selection environment need to be properly designed to permit correct disulfide formation.[85] It has been proved that the filamentous M13-fusion phage can be a reliable host system for displaying disulfide-constrained with complex secondary structures.[86]
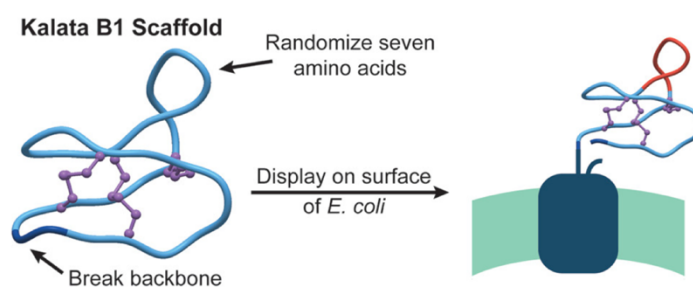
Figure 1.4: Display of kalata B1 based peptide library on *E. coli* surface. The kalata B1 scaffold was linearized in loop 2 and fused to an engineered bacterial display protein (eCPX). Reproduced with permission from ref. 79 Copyright 2011, American Chemical Society.[79]

1.3. mRNA display and RaPID selection
1.3.1. mRNA display

Other than phage display and the cell-based display approaches, I envisioned that messenger RNA (mRNA) *in vitro* display, capable of affording substantially higher library diversities than other display techniques, may offer an improved route to identify DRPs with high potency and selectivity against proteins of interest. mRNA-display is a peptide display technology taking

advantage of the covalent linkage of a peptide to its cognate mRNA molecule that can be retrieved and deconvoluted by nucleotide sequencing.[87] To efficiently link the phenotype (peptide) and the genotype (mRNA), mRNA display utilizes the antimicrobial natural product puromycin, which inhibits translation by mimicking the 3′ end of an aminoacyl-tRNA that can be recognized by ribosome during translation. To allow substantial molecular diversity, mRNA library with random nucleotide sequences is constructed and covalently modified with puromycin on its 3′ end, and then translated in an *in vitro* translation reaction. When approaching the end of translation reaction, the ribosome encounters the 3′ puromycin on each individual mRNA molecule and recognizes its aminoacyl-tRNA mimicking segments. The puromycin is transferred to the A-site (for aminoacyl) of the ribosome and is covalently linked to the nascent peptide on the P-site (for peptidyl) via an amide bond. In this manner, each individual mRNA is attached to the corresponding translated peptide via the puromycin. Usually the mRNA-peptide fusion molecule is reverse transcribed to give a library of cDNA-mRNA duplexes attached to their peptides. Affinity-based selection and pulldown of binding peptides is then processed against an immobilized target protein. cDNA sequences are then recovered from the binders and amplified by PCR, and transcribed to mRNA for the next round of mRNA display selection until an enrichment of sequences is observed.[88,89]

Compared with other screening technologies, the peptide library in mRNA display can reach a higher diversity because the *in vitro* systems do not require a transformation step, thus they are not limited by the library transformation efficiency. Theoretically, the peptide library complexity is nearly identical to that of the initial mRNA library complexity, which is always $>10^{12}$ in the reported practices. In contrast, a phage library may contain more than 10^{12} individual phage particles, but since a typical *E. coli* transformation efficiency is only 10^9 per microgram of DNA, the typical diversity that can be achieved using a phage display library is only around 10^9.[90]

This display system was previously applied on a backbone-acyclic knottin scaffold, *Ecballium elaterium* trypsin inhibitor-II (EETI-II), with *in vitro* translation achieved using the rabbit reticulocyte lysate system.[81] The library was screened against its natural target bovine trypsin, and the most potent binding sequence identified ($K_D = 16$ nM) was identical to wild type EETI-II, thus failing to produce improved *de novo* species. No other reports have emerged on this topic, and the use of mRNA display to evolve cystine knot peptides for new targets has not been demonstrated. One key question relates to the degree of sequence variation that is compatible with formation of the cystine knot, as randomizing multiple residues in the scaffold might prevent correct folding for certain non-native sequences. I hypothesized that an optimal *in vitro* translation system is needed to maintain high library diversity (and potentially compensate for folding losses) to enable selection

of potent and selective variants.

1.3.2. FIT translation and RaPID selection

Although mRNA display itself is an efficient and powerful platform for functional peptide discovery, the original translation conditions using cell lysate limits its application when it comes to the designing of non-canonical amino acid (ncAA) containing peptides. Several attempts have been reported to incorporate ncAA in translated peptide, including preloading the ncAA on tRNAs,[91–93] assigning ncAAs to nonsense codons,[94,95] or expanding genetic codons by using four- or five-nucleotide codons or unnatural base pairs.[96–101] The genetic code reprogramming was further benefited after 2001, by the development of fully reconstituted translation systems using highly purified *E. coli* ribosomes and recombinantly expressed translation factors, which is referred to as the Protein synthesis Using Recombinant Elements (PURE) system. The translation cocktail of this system contains 32 components, including initiation factors (IF1, IF2, IF3), elongation factors (EF-G, EF-Tu, EF-Ts), release factors (RF1, RF3), ribosome recycling factor (RRF), all 20 aaRSs, methionyl-tRNA transformylase (MTF), T7 RNA polymerase, and purified *E. coli* ribosomes. Furthermore, the system contains 46 tRNAs, NTPs, creatine phosphate, 10-formyl-5,6,7,8-tetrahydrofolic acid, all canonical 20 amino acids, creatine kinase, myokinase, nucleoside-diphosphate kinase, and pyrophosphate.[102] In this system, the translation machinery can be fully altered to include or exclude any component, such as tRNAs, amino acids, aaRSs, or release factors, allows the simultaneous reprogramming of several ncAAs by adding chemically charged tRNAs and omitting the corresponding aaRSs and canonical amino acids (cAAs). Szostak and co-workers showed in 2005 that this method of genetic code reprogramming and *in vitro* translation is compatible with mRNA display.[103] By combining the PURE system, mRNA display, and genetic reprogramming,[104–106] dozens of ncAA analogues have been incorporated for peptide binder selections.

Although the genetic reprogramming methods described above have significantly extended the versatility of peptide libraries used in mRNA display, the efficiency and fidelity of peptide synthesis in this system is highly dependent on the substrate promiscuity of recombinant aaRS toward ncAAs.[103] Insufficient aminoacylation may limit the diversity of the unnatural residues that can be used for library construction.[107] To further promote the scope of ncAAs that can be incorporated, Suga et al. established an efficient ribozyme-assisted translation system, called the Flexible *in vitro* Translation (FIT) approach.[108–110] In such approach, the ncAAs are precharged on tRNAs using the robust aminoacylating ribozymes, flexizymes. Taking advantage of the three

reported flexizymes, enhanced flexizyme (eFx), dinitro-flexizyme (dFx) and amino-flexizyme (aFx), all tRNAs with a 3′-terminal CCA motif are able to undergo aminoacylation if provided with a suitable amino acid substrate. By incorporating with PURE system, the flexizyme-assisted tRNA aminoacylation extensively expands the repertoire of ncAAs that can be incorporated in peptides. The mRNA display process combining with FIT translation, which is named random nonstandard peptide integrated discovery (RaPID) selection (Figure 1.5), allows the generation of peptide libraries with distinctive structural features. These libraries can contain D-amino acids, N-methylated amino acids, and also warhead-embedded linear or cyclic peptide libraries may be constructed. In this context, the RaPID system is an appealing platform, which allows the diversity-oriented selection of bioactive natural-product-like peptides with high target affinities.[89]
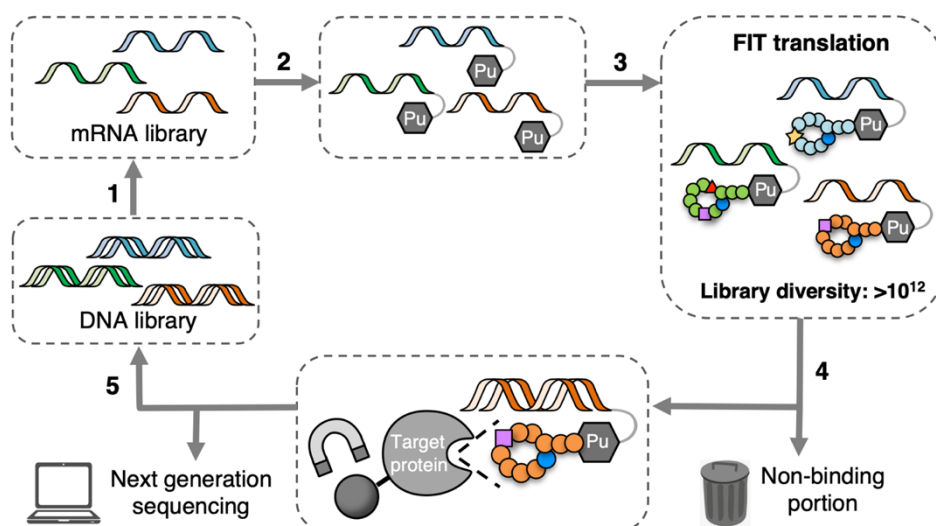
Figure 1.5: Standard RaPID selection scheme using macrocyclic peptide library. On round of RaPID selection includes (**1**) transcription of the DNA library to mRNA; (**2**) conjugation to puromycin (represented by **Pu**) leads to the formation of an mRNA library with linked 3′-puromycin; (**3**) translation of the library in FIT reaction leads to a library of macrocyclic peptides (green, blue and orange, with the ncAA indicated as blue circle, yellow star, red triangle or purple square) conjugated with the cognate mRNAs; (**4**) reverse transcription generates the cDNA of each mRNA, and the pooled mixture can then be panned against an immobilised protein target of interest to identify peptides with high affinity; and (**5**) recovery of the cognate DNA by PCR to generate an enriched library and the process can be repeated iteratively until only peptides with high affinity to the target are represented in the library.

1.4. Research objectives

In the above sections, I introduced the advantages, current engineering methods and perspectives of DRP-based drug discovery. In brief, due to the intramolecular disulfide bonds, DRPs commonly carry distinctive constrained folding structures, which provide them with remarkable proteolytic stabilities, target potency, selectivity, and/or cell permeability, making DRPs an attractive template for modern drug discovery. Current engineering method for DRP molecules are mainly based on chemical synthesis (SPPS), assisted by computational or structural analysis. Although chemical synthesis-derived methods have led to development of various bioactive DRPs, the relatively comprehensive synthesis route requires extensive efforts to create DRP-based peptide libraries with large diversity. On the other hand, other synthesis methods including enzymatic cyclization or recombinant expression, combined with several biological display approaches, such as cell or phage display, have been utilized in DRP discovery to create more bioactive miniproteins bearing the disulfide-rich scaffold. However, due to the complexity of disulfide connectivity in DRPs, it's hard to control the peptide architecture during a cell-based display method. In a different context, the cell-free display method, RaPID system, is known to afford highly diverse libraries ($>10^{12}$ members) for affinity-based selection and may offer an improved route to identify DRPs with high potency and selectivity against proteins of interest. In addition, RaPID allows for incorporation of non-natural amino acids into peptides taking advantage of the Flexible in vitro translation (FIT) system, therefore making it possible to control disulfide conformation by introducing chemically modified cysteines during selection.

In these contexts, the goal of my PhD research is to design and construct diverse DRP-based peptide libraries and develop bioactive DRP-based drug candidates against pharmaceutical targets using RaPID selection system. In brief, my first research objective is to examine whether the DRP scaffold can be applied to RaPID selection (or mRNA display) to establish peptides with potent target affinity. Next, after proving the feasibility of DRP scaffold in RaPID, my following objective is to control the disulfide connectivity during selection process to obtain peptides with desired conformation and bioactivities. To achieve these objectives, in Chapter 2, I first designed a library based on the scaffold of a plant-derived cyclotide MCoTI-II, containing 3 disulfide bonds, and screened it against human coagulation factor XIIa. This work has been published in ref. 111.[111] After identifying several potent target-binding cyclotide variants from the non-modified library, in Chapter 3 and 4, I further designed and screened several DRP libraries with bicyclic and tricyclic topologies, by controlling the disulfide connectivity. These libraries were processed by FIT translation to incorporate thiol-protected cysteine residues, which can be stepwise deprotected to

form desired disulfide bonds. The bicyclic and tricyclic libraries were screened against therapeutic asialoglycoprotein receptors and SARS-CoV-2 main protease, respectively, providing various DRP-based peptides with potent target affinity.

Chapter 2.

Discovery of potent and selective cyclotide-based FXIIa inhibitors using RaPID selection

2.1. Introduction

2.1.1. Cyclotides

Cyclotides are plant-derived head-to-tail cyclic peptides with a cystine knot topology (Figure 2.1a).[112] They contain three pairs of disulfide bonds which are formed between six Cys residues with a threaded conformation. The backbone regions between the Cys residues are referred to as loops. The constrained cyclic and knotted structures provide cyclotides with exceptional thermal, chemical and proteolytic stability compared with other linear peptides.[113] In addition, some cyclotides have been shown to be orally bioactive[31,114] and/or cell permeable.[115,116] Natural cyclotides are reported to display a wide range of host defense-related activities,[117] demonstrating their capability to exert potent biological effects. This combination of stability and potency, with the potential for oral activity, makes cyclotides appealing scaffolds for the development of next generation biologic therapies, agrichemical agents or chemical probes.[18] In nature, the sequences of cyclotides have been progressively optimized for specific functions over millions of years of evolution. By contrast, the production of engineered cyclotides with new functions or altered target specificity requires similar levels of optimization but over a dramatically shorter timeframe.

Based on the three-dimensional folding patterns, cyclotides are divided into three sub-families, Mobius, Bracelet and Trypsin inhibitor.[118] Among cyclotides, plant-derived *Momordica cochinchinensis* trypsin inhibitor-II, MCoTI-II, is known for its easy folding into the native conformation and is used as the engineering scaffold for epitope grafting or rational design.[118] As a potent natural trypsin inhibitor, MCoTI-II interacts with trypsin's S1 enzyme pocket in a substrate-like way. In general, MCoTI-II is docked into trypsin through non-covalent interaction between the cleavage sites P4-P1, P1'-P4' residues in loop 6 and loop 1 and the trypsin active site (Figure 2.1b).[119] The Lys residue on the P1 site is inserted into the S1 pocket and makes hydrogen bond with the catalytic Ser on the bottom of the S1 pocket.
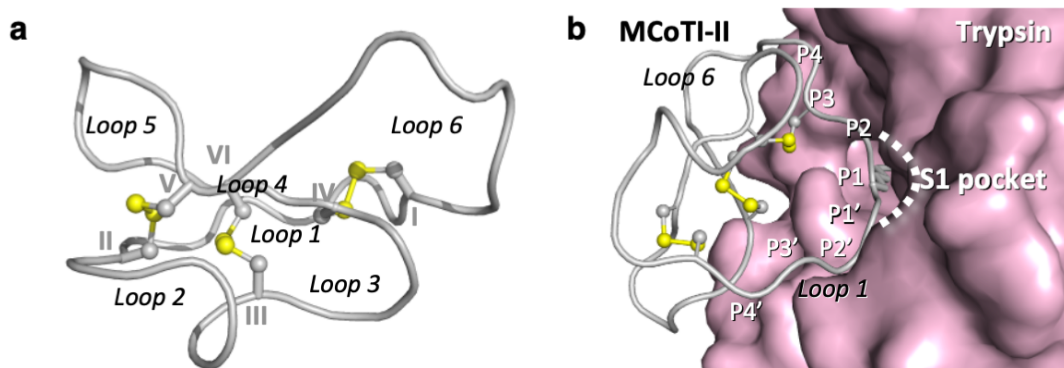
Figure 2.1: (a) Structure of the prototypic cyclotide, showing the head-to-tail cyclic backbone and knotted arrangement of three disulfide bonds deriving from six conserved Cys residues (labelled I–VI). Backbone regions between the Cys residues are referred to as loops 1–6; (b) Structure of the prototypic trypsin inhibitor cyclotide *Momordica cochinchinensis* trypsin inhibitor-II, MCoTI-II, interacting with the S1 enzyme pocket of trypsin (PDB 4GUX). The cleavage sites P4–P1, P1'–P4' (Schechter–Berger nomenclature) on MCoTI-II are indicated. Figure (a) is reproduced with permission from ref. 112 Copyright 1999 Academic Press.[112] Figure (b) is reproduced with permission from ref. 119 Copyright 2013, The American Society for Biochemistry and Molecular Biology, Inc.[119]

2.1.2. FXIIa and its inhibitor

Coagulation factor XIIa (FXIIa) is an activated multidomain serine protease of the trypsin-like family, which triggers the intrinsic pathway resulting in blood clotting through generation of thrombin and fibrin. In contrast to other coagulation factors, FXIIa deficiency in humans and animals is not associated with excessive bleeding,[120–122] indicating that the functionality of FXIIa is not essential for hemostasis. As such, Inhibition of FXIIa has been proposed as a strategy for developing safer therapies for thromboembolic and inflammatory diseases.[123,124] Other than thrombosis, FXIIa-directed therapies could also be useful for treatment of several diseases such as hereditary angioedema (HAE), sepsis, Alzheimer's disease or multiple sclerosis.[125]

Due to FXIIa involvement in several pathologies, over the past three decades, diverse FXIIa inhibitors such as small molecules, peptides, proteins, antibodies, antisense oligonucleotides, and siRNAs have been developed.[126] In particular, numerous natural peptides and mini-proteins derived from insects, bacteria, marine animals and plants have been found with mild to potent FXIIa inhibitory ability.[126] Most of the FXIIa inhibiting natural peptides possess more than one pair of disulfide bonds, which assist the peptides folding into a constrained three-dimensional structure.

Previously, several cyclotide-based protease inhibitors have been generated through rational approaches designed to bias the specificity of potent natural trypsin inhibitors (the *Momordica*

cochinchinensis derived cyclotides, MCoTI-I and MCoTI-II) towards other serine protease targets, including FXIIa.[22–25] FXIIa inhibitors based on MCoTI-II were generated by screening a limited number of synthetic cyclotides bearing mutations in loop 1 (the primary binding loop of trypsin inhibitor cyclotides).[25] However, the potency of these molecules against FXIIa is ~1000-fold lower than that shown by MCoTI-II against trypsin (K_i = 30 pM)[23] and they have poor selectivity. Other cyclotide-based inhibitors of disease-related serine proteases, including β-tryptase, neutrophil elastase and matriptase, have also been reported[22–24] and, as for the FXIIa inhibitors, none have been sufficiently potent and selective to warrant further development.

2.1.3. mRNA display of cyclotide-based library against FXIIa

As introduced before, mRNA display can afford substantially high library diversities ($>10^{12}$ variants) and may offer an improved route to identify cyclotides with high potency and selectivity against proteins of interest. This display system was previously applied on a backbone-acyclic knottin scaffold, *Ecballium elaterium* trypsin inhibitor II (EETI-II), with 3 disulfide bonds arranged similarly to MCoTI-II.[127] The library was screened against bovine trypsin. Interestingly, sequence of the most potent binding peptide (K_D = 16 nM) was identical to the wild type EETI-II, failing to obtain improved *de novo* species.

Therefore, starting from my master study, I applied the mRNA display for the affinity-based selection of an MCoTI-II derived cyclotide library against FXIIa and yielded in a few FXIIa binding peptides. I constructed a semi-randomized peptide library based on native MCoTI-II sequence which is linearized in loop 6, such that the residues predicted to interact with trypsin-like proteases (all of loops 1 and 5, and a V residue in loop 6) were randomized to allow the occurrence of any of the 20 canonical amino acids at these 12 positions (Figure 2.2a). The DNA encoding this library was assembled from degenerate oligonucleotides, transcribed into mRNA, ligated to a puromycin-linked oligonucleotide and translated *in vitro* to produce a library of mRNA-peptide fusion molecules. The library was screened for affinity to biotinylated FXIIa immobilized on streptavidin-coated magnetic beads. Following four rounds of selection (Figure 2.2b), next-generation sequencing was performed to identify high frequency clonal sequences in the final cDNA library.

After sequencing, a few peptides were chosen for solid phase peptide synthesis (SPPS). I conducted most of the synthesis work during a three-months internship in the University of Queensland in the first year of my PhD course and the following analysis were conducted during my PhD course.
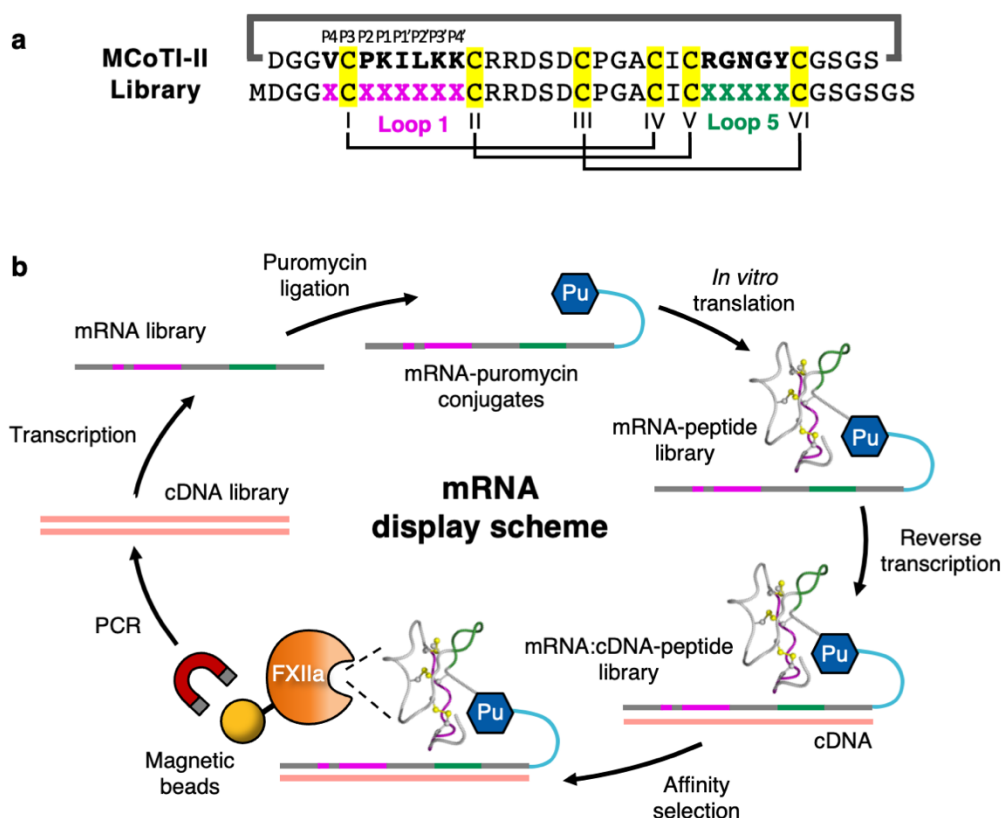
Figure 2.2: My master course study on this topic. (a) Library design based on native MCoTI-II scaffold.; (b) Experimental scheme of mRNA display using MCoTI-II-based library against FXIIa. Reproduced with permission from *J. Am. Chem. Soc.* **2021**, *143*, 18481–18489, Copyright 2021, American Chemical Society.

2.2. Results and discussion

2.2.1. Five peptides from selection were synthesized and possess native-like cyclotide conformation

Five peptides (designated MCoFx1–5, Figure 2.3a) from the top 19 selected sequences were chosen for solid-phase peptide synthesis (SPPS)[128] and characterization. Both backbone-acyclic (as selected during display screening) and cyclic forms of each sequence were synthesized to determine the effect of backbone cyclization on FXIIa binding and inhibitory activity. The synthetic route is shown in Figure 2.3b. The peptides were purified by RP-HPLC and characterized by analytical HPLC to give over 95% purity (Figure 2.3c).

Conformations of each peptide were assessed by ^{1}H-NMR spectroscopy, which showed comparable peak patterns to those of MCoTI-II (Figure 2.4a). Although the 1D NMR spectra

couldn't prove the exact peptide structure, the chemical shifts are valuable parameters for rapidly assessing cyclotide structural integrity. For example, for all acyclic and cyclic forms of MCoFx1–5, wide dispersion of amide NH chemical shifts at 6–10 ppm was observed and it could be used as a preliminary marker of correct folding in synthetic cyclotides.[129] Additionally, changes in α-proton NMR secondary chemical shift are routinely used to compare localised changes between different cyclotides.[129] The secondary shifts of cMCoFx1 (Figure 2.4b) indicated that cMCoFx1 adopted the native-like conformation of MCoTI-II. Some differences in chemical shifts were observed in loops 1 and 5, corresponding to sequence divergence from MCoTI-II.

a

Peptide	Loop 1 Sequence Loop 5
MCoTI-II	DGGVCPKILKKCRRDSDCPGACICRGNGYCGSGS
MCoFx1	DGGICPRIGRLCRRDSDCPGACICRATRFCGSGY
MCoFx2	DGGICPRILVYCRRDSDCPGACICIRRTYCGSGS
MCoFx3	DGGRCPRLLRWCRRDSDCPGACICARGGLCGSGS
MCoFx4	DGGVCPRVGWRCRRDSDCPGACICYPTKWCGSGS
MCoFx5	DGGRCCGGYLVCRRDSDCPGACICVFKKHCGSGS

5   10   15   20   25   30

b
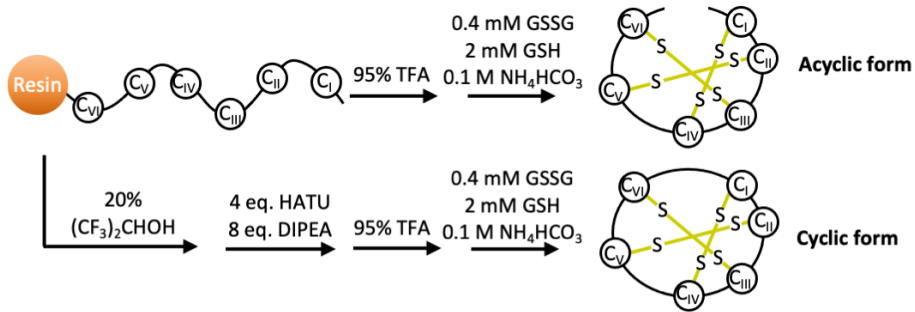
c
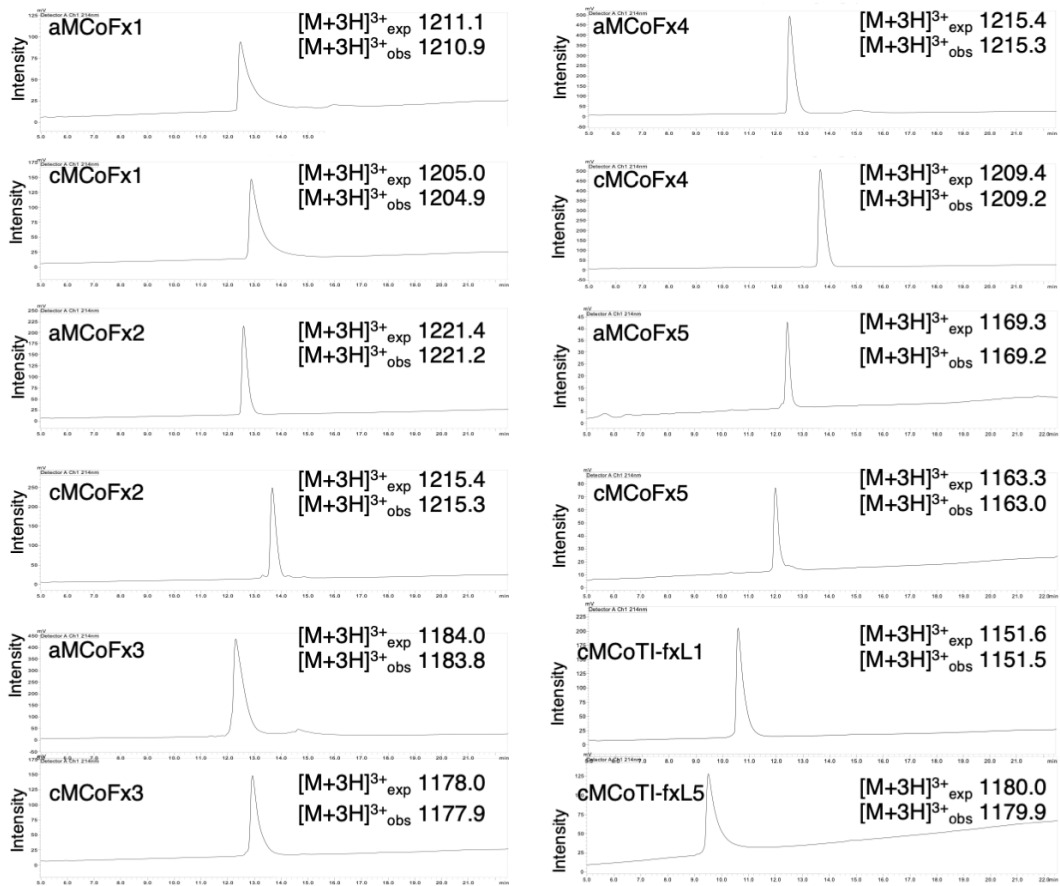
Figure 2.3: (a) Five peptides selected from the mRNA display against FXIIa. The residues randomized in the library are colored in RasMol coloring scheme; (b) Synthetic scheme of SPPS of the acyclic/acyclic forms of the selected peptides; (c) LC-MS spectra of all synthetic peptides. The expected and observed multi-charged ion mass of each peptide are listed on each spectrum. All the peptides have more than 95% purity. Reproduced with permission from *J. Am. Chem. Soc.* **2021**, *143*, 18481–18489, Copyright 2021, American Chemical Society.

a

MCoTI-II
aMCoFx1
aMCoFx2
aMCoFx3
aMCoFx4
aMCoFx5

MCoTI-II
cMCoFx1
cMCoFx2
cMCoFx3
cMCoFx4
cMCoFx5
cMCoTI-fxL1
cMCoTI-fxL5

10   8   6   4   2   0 ppm

24
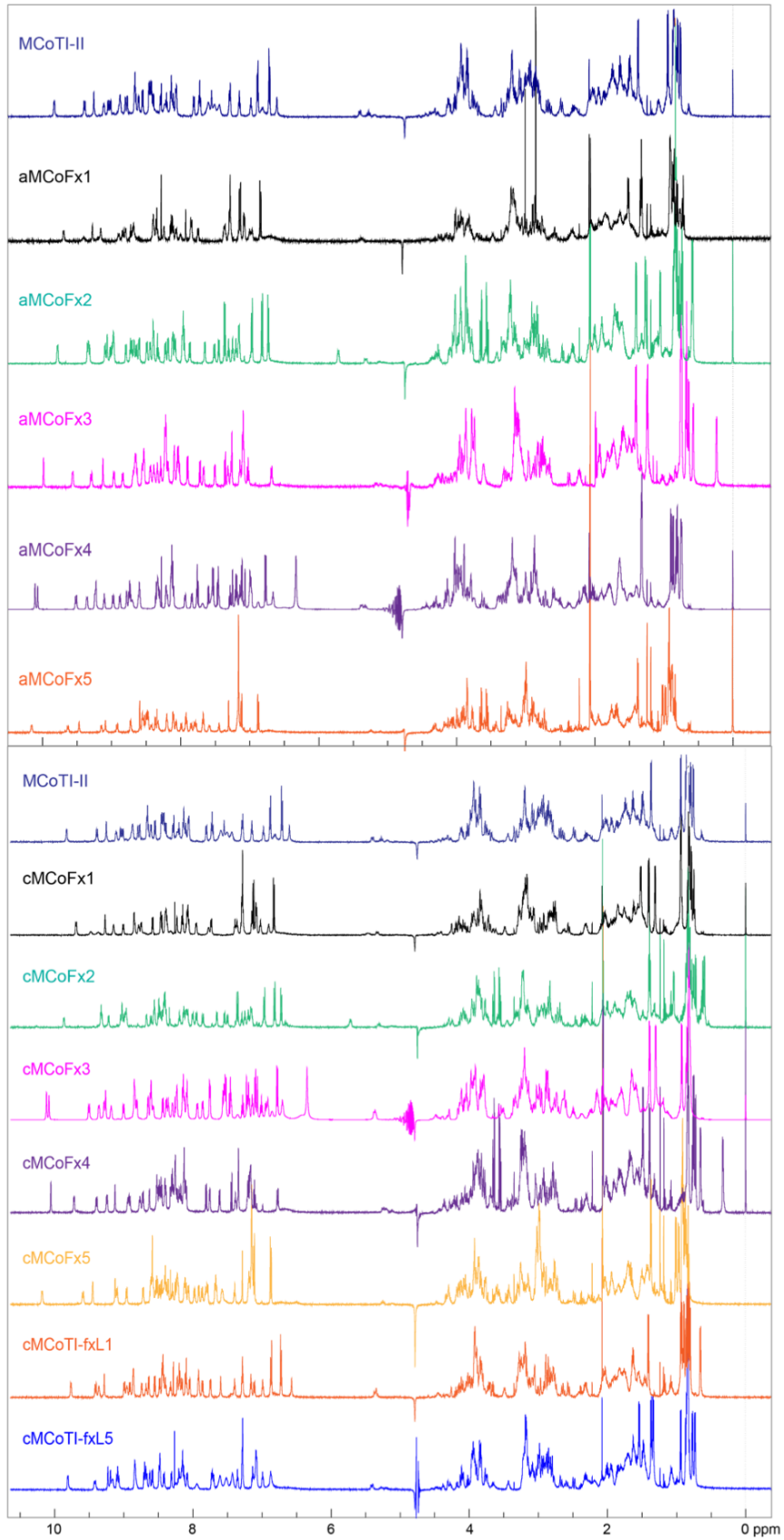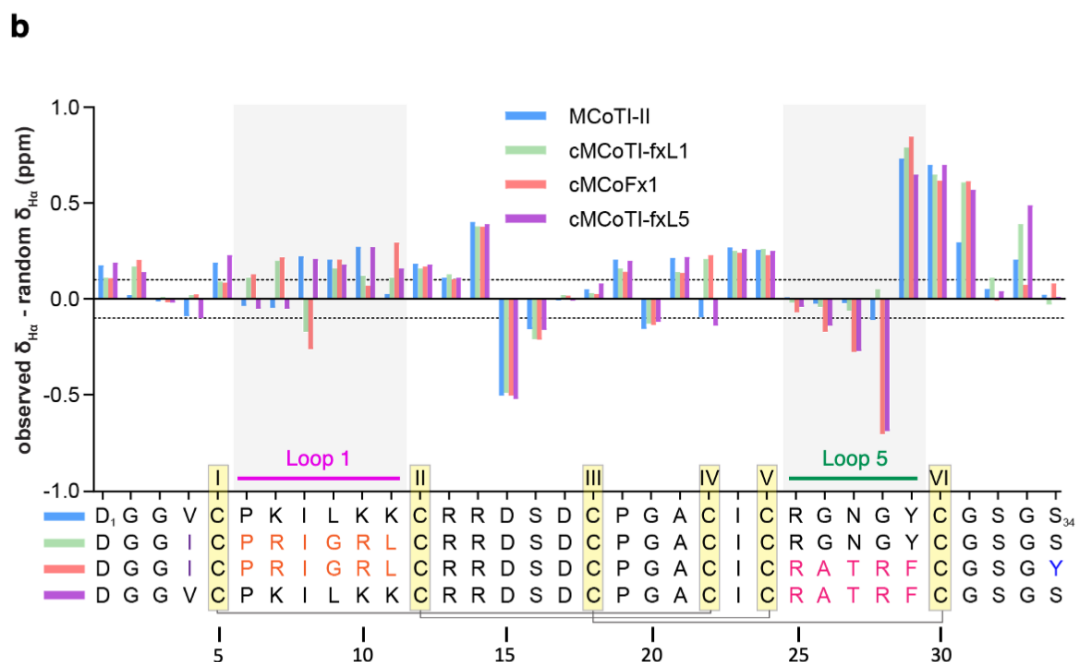
Figure 2.4: (a) 1D ^{1}H-NMR spectra of chemically synthesized cyclic MCoTI-II and acyclic and cyclic MCoFx1–5; (b) α-Proton secondary chemical shifts analysis of MCoTI-II (blue), cMCoFx1 (red) and loop-replacing variants cMCoTI-fxL1 (green) and cMCoTI-fxL5 (purple). The dotted lines represent secondary chemical shift values of − 0.1 and 0.1 ppm. Sequences of the four peptides are shown below the chart. Regions identical to cMCoFx1 in cMCoTI-fxL1 and cMCoTI-fxL5 are highlighted in orange and red, respectively. Six cysteines are highlighted in yellow, indicating the arrangement of three disulfide bonds. Reproduced with permission from *J. Am. Chem. Soc.* **2021**, *143*, 18481–18489, Copyright 2021, American Chemical Society.

2.2.2. Binding affinity and selectivity of selected peptides

Surface plasmon resonance (SPR) experiments were conducted, and the results demonstrated that all synthetic peptides exhibited FXIIa binding affinities in the nanomolar to picomolar range, regardless of backbone cyclization. The measurements were performed using a single-cycle kinetics, which measures association and dissociation of a series of concentrations in one cycle, without degenerating between each injection. The kinetic constants were assessed with a 1:1 binding model using the Biacore T200 (Cytiva) fitting software after measurement (Figure 2.5a).

Among all synthetic backbone acyclic and cyclic peptides, cMCoFx1 exhibited the highest binding affinity, with a dissociation constant (K_D) value of 900 pM, which is 60-fold higher affinity than native MCoTI-II. In general, the backbone cyclic analogues of each molecule (except

MCoFx5) displayed 2–7-fold higher binding affinity and inhibitory activity (*i.e.* lower K_D and K_i values) than their acyclic forms, indicating that although our screening was conducted using a backbone acyclic library, this strategy was not biased towards backbone acyclic analogues. The only exception was cMCoFx5, which displayed a slight loss of activity compared with aMCoFx5. However, since MCoFx5 contained an additional C6 and a G7 at P2 and P1 sites, replacing the conserved P6 and R7 residues, respectively, the sequence differed substantially from MCoFx1–4. In addition, both MCoFx5 displayed a subtle loss of activity compared with MCoTI-II (*vide infra*), these mutations did not appear to be beneficial for FXIIa binding.

MCoTI-II was originally found as a potent trypsin inhibitor (K_i = 30 pM). Although the sequence of FXIIa has 36% and 37% identity to human and bovine trypsin, respectively, the residues in the S1 enzymatic pocket have far higher homology (~90%). To assess the selectivity of the selected peptides, binding affinity against bovine trypsin, as an example of a related serine protease, were determined using SPR measurement. Consistent with literature, native MCoTI-II synthesized together with the selected peptides showed a K_D of less than 100 pM towards trypsin (Figure 2.5b). The most potent FXIIa binder, cMCoFx1, exhibited 60-fold greater affinity for FXIIa and 300-fold lower affinity for trypsin compared with those of native MCoTI-II. Interestingly, peptide cMCoFx2 showed better binding to FXIIa (K_D = 10 nM) compared with MCoTI-II (58 nM), however its binding affinity also remained high to trypsin (K_D less than 100 pM). Because cMCoFx1 and cMCoFx2 possessed identical residues at the P4, P2, P1 and P1′ sites, the significant selectivity of cMCoFx1 for FXIIa over trypsin must lie in the differences at the P2′–P4′ sites and/or in loop 5 between these two molecules. A similar trend was observed for cMCoFx3 and cMCoFx4, which exhibited 2–3-fold improved affinity for FXIIa compared to MCoTI-II, however possessed 2–8-fold higher affinity for trypsin.

Judging from the K_D values, neither aMCoFx5 nor cMCoFx5 showed detectable binding for trypsin (Figure 2.5b). As discussed earlier, the loop 1 sequence of these peptides diverged substantially from the other selected peptides. To look for the origin of this high selectivity, I speculate that the unique C6 mutation at P2 site as well as the consecutive G7 and G8 mutations at P1 and P1′ sites may alter the entire loop, resulting in a different loop structure or possibly altered the folding compared with native MCoTI-II. The original interaction with trypsin occurring via loop 1 might be eliminated by such an unexpected structural change, and instead create a possible interface between MCoFx5 and FXIIa, leading to the observed high selectivity even though the binding affinity is relatively modest (K_i = 23 nM).

a

Binding affinity towards FXIIa   **Binding affinity towards trypsin**

MCoTI-II $K_D = 58$ nM

MCoTI-II $K_D < 0.1$ nM

aMCoFx1 $K_D = 1.89$ nM
cMCoFx1 $K_D = 0.90$ nM
aMCoFx1 $K_D = 58$ nM
cMCoFx1 $K_D = 58 \pm 7.5$ nM

aMCoFx2
cMCoFx2
aMCoFx2
cMCoFx2

aMCoFx3
cMCoFx3
aMCoFx3
cMCoFx3

aMCoFx4
cMCoFx4
aMCoFx4
cMCoFx4

aMCoFx5
cMCoFx5
aMCoFx5
cMCoFx5

27

b
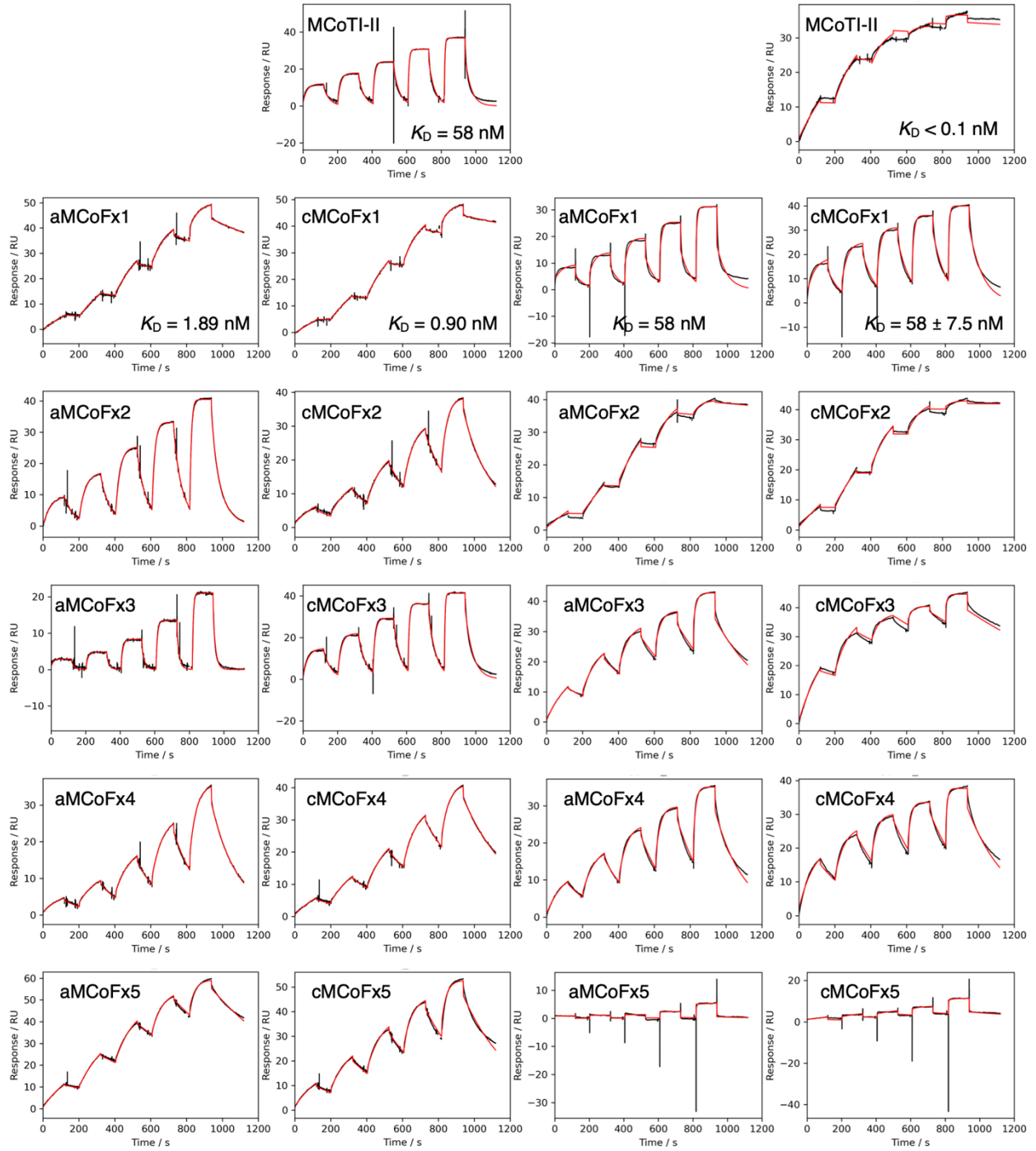
Peptide	Sequence	K_D (nM)			
		FXIIa		**Trypsin**	
	Loop 1 Sequence Loop 5	**Acyclic**	**Cyclic**	**Acyclic**	**Cyclic**
MCoTI-II	DGGVCPKILKKCRRDSDCPGACICRGNGYCGSGS		58 ± 7.5		< 0.1
MCoFx1	DGGICPRIGRLCRRDSDCPGACICRATRFCGSGY	1.89 ± 0.04	0.90 ± 0.04	59.6	31.5
MCoFx2	DGGICPRILVYCRRDSDCPGACICIRRTYCGSGS	45.4 ± 2.9	10 ± 1.3	< 0.1	< 0.1
MCoFx3	DGGRCPRLLRWCRRDSDCPGACICARGGLCGSGS	167 ± 42	22.8 ± 1.6	2.39	2.72
MCoFx4	DGGVCPRVGWRCRRDSDCPGACICYPTKWCGSGS	73.8 ± 6.3	14.4 ± 1.1	15.1	4.09
MCoFx5	DGGRCCGGYLVCRRDSDCPGACICVFKKHCGSGS	23.0 ± 0.4	81.8 ± 3.2	No binding	No binding

       5   10   15   20   25   30

Figure 2.5: (a) SPR sensorgrams of chemically synthesized cyclic MCoTI-II and acyclic and cyclic MCoFx1–5 binding to FXIIa or trypsin; (b) Dissociation constants (K_D) derived from the SPR measurement. Reproduced with permission from *J. Am. Chem. Soc.* **2021**, *143*, 18481–18489, Copyright 2021, American Chemical Society.

2.2.3. Inhibitory ability of selected peptides

Inhibitory ability of the selected peptides against FXIIa was determined using *in vitro* inhibition assay. FXIIa was pre-incubated with a fluorophore-labeled substrate (a tetramer peptide, Ac-QRFR-pNA), followed by the addition of selected peptides at different concentration. The inhibition constants (K_i) were fitted to the Michaelis-Menten kinetics. The K_i values of all backbone acyclic and cyclic peptides are shown in Figure 2.6.

The trend of K_i of the selected peptides was similar to that was seen for the K_D value. cMCoFx1 was the most active FXIIa inhibitor with a K_i of 370 pM, which is 350-fold more potent than native MCoTI-II. To the best of our knowledge, this molecule is amongst the most potent FXIIa peptide inhibitors reported to date. Interestingly, all the selected FXIIa binders possessed corresponding inhibitory ability, suggesting that all peptides interacted with the target protein at the enzymatic pocket.

Peptide	Loop 1 Sequence Loop 5	K_i (nM) Acyclic	Cyclic
MCoTI-II	DGGVCPKILKKCRRDSDCPGACICRGNGYCGSGS		129 ± 9
MCoFx1	DGGICPRIGRLCRRDSDCPGACICRATRFCGSGY	0.73 ± 0.2	0.37 ± 0.04
MCoFx2	DGGICPRILVYCRRDSDCPGACICIRRTYCGSGS	73 ± 3	12 ± 0.7
MCoFx3	DGGRCPRLLRWCRRDSDCPGACICARGGLCGSGS	525 ± 30	75 ± 3
MCoFx4	DGGVCPRVGWRCRRDSDCPGACICYPTKWCGSGS	17 ± 0.9	5.5 ± 0.5
MCoFx5	DGGRCCGGYLVCRRDSDCPGACICVFKKHCGSGS	106 ± 7	207 ± 9
	5 10 15 20 25 30		

Figure 2.6: Inhibitory constants (K_i) of chemically synthesized cyclic MCoTI-II and acyclic and cyclic MCoFx1–5 against FXIIa. Reproduced with permission from *J. Am. Chem. Soc.* **2021**, *143*, 18481–18489, Copyright 2021, American Chemical Society.

2.2.4. cMCoFx1 is a potent and selective inhibitor of FXIIa

To further expand the activity screen for the best selected inhibitor cMCoFx1 from FXIIa to other clinically relevant proteins, I performed the inhibitory assay using nine off-target serine proteases, including coagulation factors FXa, FXIa and thrombin. Inhibition was first tested in a broad range of cMCoFx1 concentrations, for proteases with IC_{50} < 5 μM, K_i was determined and shown in Figure 2.7b. The percentage value in brackets indicates percent inhibition at 5 μM for the off-target proteases.

These assays revealed that the difference in K_i for cMCoFx1 against FXIIa (0.37 nM) and trypsin (1,110 nM) exceeds three orders of magnitude. Remarkably, cMCoFx1 shows almost no inhibitory activity against other serine proteases, confirming that cMCoFx1 is highly specific for FXIIa. Since other off-target proteases including FXa and FXIa play important roles in biology, the FXIIa specificity of cMCoFx1 suggested its future application as a drug candidate.

Figure 2.7: (a) Sequences of MCoTI-II, cMCoFx1 and two loop-grafted variants, cMCoTI-fxL1 and cMCoTI-fxL5, where loop 1 or loop 5 in MCoTI-II was replaced by the optimized sequence in cMCoFx1; (b) Inhibitory activity of cMCoFx1, cMCoTI-fxL1 and cMCoTI-fxL5 against a panel of serine proteases. K_i was determined for inhibitors with $IC_{50} < 5$ µM against a given protease. The percentage value in brackets indicates percent inhibition at 5 µM for the off-target proteases. Reproduced with permission from *J. Am. Chem. Soc.* **2021**, *143*, 18481–18489, Copyright 2021, American Chemical Society.

2.2.5. Loop-grafted variants of cMCoFx1 proves that loop 1 is the key loop for inhibition and selectivity

To further analyze the inhibitory activity of cMCoFx1, I synthesized two loop-grafted peptides, referred to as cMCoTI-fxL1 and cMCoTI-fxL5, where the peptide sequence from either loop 1 or loop 5 of cMCoFx1 was grafted into the respective loop of MCoTI-II (Figure 2.7a). A comparison of the α-proton NMR chemical shifts of cMCoTI-fxL1 and cMCoTI-fxL5 (Figure 2.4b) showed their similarity to the parent peptides in respective regions. For instance, the secondary Hα shifts for loop 1 of cMCoTI-fxL1 were similar to those of cMCoFx1, whereas loop 1 of cMCoTI-fxL5 was comparable to MCoTI-II.

Similar to the previous section, inhibitory ability of the single loop-grafted variants against other relevant serine proteases was measured (Figure 2.7b). Strikingly, cMCoTI-fxL1 exhibited a comparable profile to cMCoFx1, although its potency against FXIIa (K_i = 0.69 nM) and selectivity with respect to trypsin and matriptase (K_i values of 996 and 2,210 nM, respectively) are slightly lower. By contrast, cMCoTI-fxL5 showed nearly the same potency and selectivity as native MCoTI-II, *i.e.* K_i values for FXIIa, trypsin and matriptase are 66, 0.08, and 17 nM, respectively. These results clearly demonstrated that the selected sequence of loop 1 played a major role in determining the potency and selectivity of MCoTI-II analogues, although the sequence of loop 5 did provide a subtle enhancement to the overall activity of cMCoFx1.

2.2.6. cMCoFx1 selectively inhibits the intrinsic pathway *in vitro*

Having identified two potent and selective FXIIa inhibitors (cMCoFx1 and cMCoTI-fxL1), *in vitro* coagulation assays were next performed to assess their biological activity in human plasma. Inhibition of FXIIa was examined in activated partial thromboplastin time (aPTT) assays, where addition of kaolin initiates the intrinsic pathway via activation of FXII. The clotting time in the aPTT assays was elongated by addition of both cMCoFx1 and cMCoTI-fxL1 in a dose-dependent manner and both inhibitors showed substantial activity in the nanomolar range (Figure 2.8a,b). The concentration of inhibitor required to double the clotting time observed in control assays ($EC_{2\times}$) was extracted from the fitting assay curves. By this measure, cMCoTI-fxL1 was slightly more potent than cMCoFx1 ($EC_{2\times}$ = 500 nM vs 702 nM), although this subtle difference between the two inhibitors could be due to variation in the experimental conditions between *in vitro* coagulation assays and previous enzyme inhibitory assays.

Besides the aPTT assays, prothrombin time (PT) assays were performed to confirm that both inhibitors are selective for the intrinsic pathway (Figure 2.8c). Since the coagulation cascade is classically divided into the intrinsic and extrinsic pathways, specific FXIIa inhibitors are expected to have no influence on the extrinsic pathway. The PT assays showed no significant differences in clotting time compared to control assays (without addition of the inhibitors) for cMCoFx1 and cMCoTI-fxL1 at 10 µM, demonstrating that the potent and selective activity of these inhibitors identified in biochemical assays extends to biological assays in human plasma.
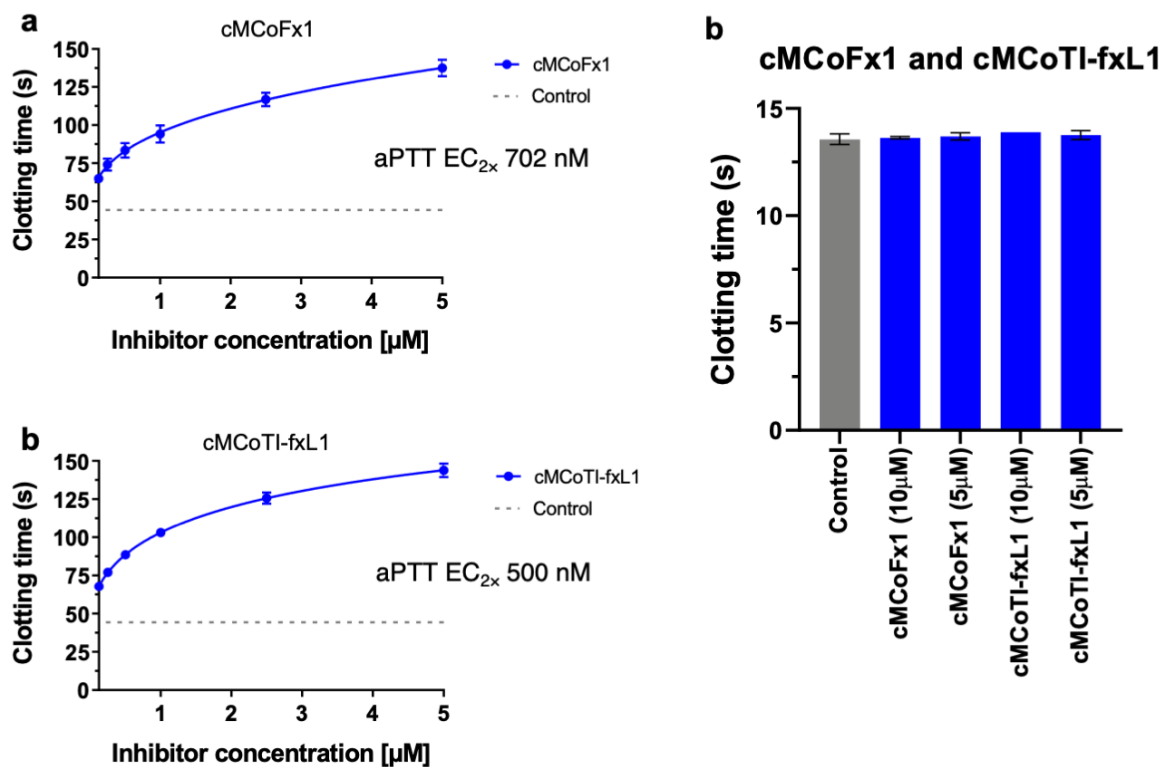
Figure 2.8: Inhibitory activity of (a) cMCoFx1 and (b) cMCoTI-fxL1 in activated partial thromboplastin time (aPTT) assays that measure clotting via the intrinsic pathway. The concentration of inhibitor required to double the clotting time observed in control assays (44.3 s, grey dashed line, buffer replaces addition of inhibitor) is shown as $EC_{2\times}$; (c) Inhibitory activity measurement of cMCoFx1 and cMCoTI-fxL1 show no activity at concentrations of 5 µM and 10 µM in prothrombin time (PT) assays, which measure clotting via the extrinsic pathway. The control bar indicates the clotting time where buffer replaces addition of inhibitors. Reproduced with permission from *J. Am. Chem. Soc.* **2021**, *143*, 18481–18489, Copyright 2021, American Chemical Society.

2.2.7. Co-crystal structure of cMCoFx1 and FXIIa

To understand the structural basis of the specific interaction between cMCoFx1 and FXIIa, I collaborated with Prof. Toru Sengoku in the Yokohama City University to determine the co-crystal structure of the cMCoFx1-FXIIa complex (Figure 2.9a). As expected, cMCoFx1 adopted a highly similar overall conformation to MCoTI-II[119] with a cyclic cystine knot topology (Figure 2.9b, PDB 4GUX, root-mean-square deviation of 0.47 Å for the 34 Cα atoms). The interaction mode of cMCoFx1-FXIIa complex was also comparable to the MCoTI-II-trypsin complex. R7 at the P1 site

of cMCoFx1 (corresponding to K7 of MCoTI-II) inserted into the S1 pocket of FXIIa, making an ionic interaction and two hydrogen bonds with Asp557 of FXIIa (Figure 2.9a,c).

cMCoFx1 formed intermolecular β-sheet-like hydrogen bonds with FXIIa at two regions (residues C5–P6 and G9–L11 in Figure 2.9c,d). The R7 carbonyl oxygen at the P1 site forms two hydrogen bonds with the mainchain nitrogen atoms of Gly561 and Ser563 of FXIIa. In addition, R7 carbonyl carbon is located 2.8 Å away from the catalytic Ser563 sidechain, suitable for nucleophilic attack (Figure 2.9c). The P6 carbonyl oxygen forms a hydrogen bond with the sidechain of FXIIa Gln560 (Figure 6c). These interactions are similar to those seen in the MCoTI-II-trypsin complex.

As mentioned above, the functional analysis of the loop-grafted variants cMCoTI-fxL1/L5 suggested that the high potency and selectivity of cMCoFx1 is primarily caused by residues in loop 1. On the other hand, the structural analysis proved that G9 and R10 of cMCoFx1 (corresponding to L9 and K10 of MCoTI-II, respectively) are the key determinants. G9 was located close to the sidechain of FXIIa Tyr515, diminishing the steric hinderance with its vacancy of bulky side chain. Substituting G9 with a larger residue could lead to a steric clash with Tyr515 (Figure 2.9d). On the contrary, in the MCoTI-II-trypsin complex, the corresponding trypsin Tyr151 made a hydrophobic interaction with L9 of MCoTI-II and adopted a different conformation (Figure 2.9e). R10 of cMCoFx1 forms hydrogen bonds with FXIIa Ser395 and Asp416, but no such interactions are formed by K10 in the MCoTI-II-trypsin complex (Figure 2.9d,e).

Co-crystal structure of FXIIa and another FXIIa-inhibiting mono-macrocyclic peptide, F3 (K_i = 1.5 nM), containing cyclic β-amino acids (cbAAs) was reported recently (Figure 2.9f,g).[130] A γ-turn was induced by incorporating two cbAAs, which stabilized the compact antiparallel β-sheet structure in F3. In the case of cMCoFx1, its rigid conformation is stabilized by the cyclic cystine knot scaffold. Structural comparison revealed that both F3 and cMCoFx1 bind to the active site of FXIIa in a similar manner, where the R sidechain (R7 in cMCoFx1 or R6 in F3) was deeply inserted into FXIIa's S1 pocket (Figure 2.9a,f). cMCoFx1 shows a slightly larger interaction area than F3 (962 Å^2 vs. 703 Å^2 in F3), which is consistent with its higher potency. The separate discoveries of F3 and cMCoFx1 demonstrate that mRNA display is broadly applicable to *de novo* discovery of not only macrocyclic peptides containing non-canonical amino acids but also nature-derived cystine-knot peptides.

a cMCoFx1-FXIIa

L11
G9
R10
R7
90°

G9 (P2') R7 (P1)
R10 (P3')
L11 (P4') I4 (P4)

I4
S1 pocket

Interface area: 962 Å²

b MCoTI-II-Trypsin
PDB 4GUX

L9 (P2') K7 (P1)
K10 (P3')
K11 (P4') V4 (P4)

c
Asp557
Ser563 Ser585 Gly587 Ser588
Gly561 R7 P6 G3
Gln560 C5

d
Tyr515 G9 I8
Ser395 Asp416
L11 R10

e
Tyr151 L9 I8
Phe44
K11 K10
Tyr42

f F3-FXIIa
PDB 6L63

ACHC8 R6
D-C17
ACHC13
Ac D-Y1

90°

ACHC8
R6 S1 pocket
ACHC13

Interface area: 703 Å²

g
F3 Ac y FAYDRR **1** LSNN **1** RNY c G

F4 Ac y RYT **1** NRLF **1** NA c G

1 (1*S*,2*S*)-2-ACHC

Figure 2.9: Structural analysis of the cMCoFx1-FXIIa complex. (a) Left; The cartoon model of the overall structure of cMCoFx1 (gray, with residues I4, P6-L11 highlighted in magenta and R25-F29 highlighted in green) bound with FXIIa (cyan). The loop 1 residues on cMCoFx1 that differ from MCoTI-II (I4, R7, G9, R10, and L11) and six C residues are shown as a stick model. Right; interaction surface. cMCoFx1 and FXIIa are shown as cartoon and surface models, respectively. The FXIIa atoms within 3.9 Å from any cMCoFx1 atoms are colored red; (b) Overall structure of MCoTI-II (slate) bound with trypsin (yellow) (PDB 4GUX). The corresponding loop 1 residues on MCoTI-II that differ from cMCoFx1 (V4, K7, L9, K10 and K11) and six C residues are shown as a stick model; (c) Interactions between cMCoFx1 and FXIIa near cMCoFx1 R7; (d) Interactions between cMCoFx1 and FXIIa near cMCoFx1 G9 and R10; (e) Interactions between MCoTI-II and trypsin near MCoTI-II L9 and K10; (f) Structural of the F3-FXIIa complex (PDB 6L63). Left; The cartoon model of the overall structure of F3 (orange) bound with FXIIa (cyan). R6, two ACHC residues, and the residues mediating macrocyclization (^{Ac}_D-Y1 and _D-C17) of F3 are shown as a stick model. Right; interaction surface. F3 and FXIIa are shown as cartoon and surface models, respectively. The FXIIa atoms within 3.9 Å from any F3 atoms are colored red; (g) Peptide sequences of cbAA-containing F3 and F4. Red sequence number: 1, (1*S*,2*S*)-2-ACHC; blue sequence: _D-amino acids. The peptides are macrocyclized with a thioether bond which was formed between ^{Ac}_D-Y1 and _D-C17 or _D-C13, respectively. Reproduced with permission from *J. Am. Chem. Soc.* **2021**, *143*, 18481–18489, Copyright 2021, American Chemical Society.
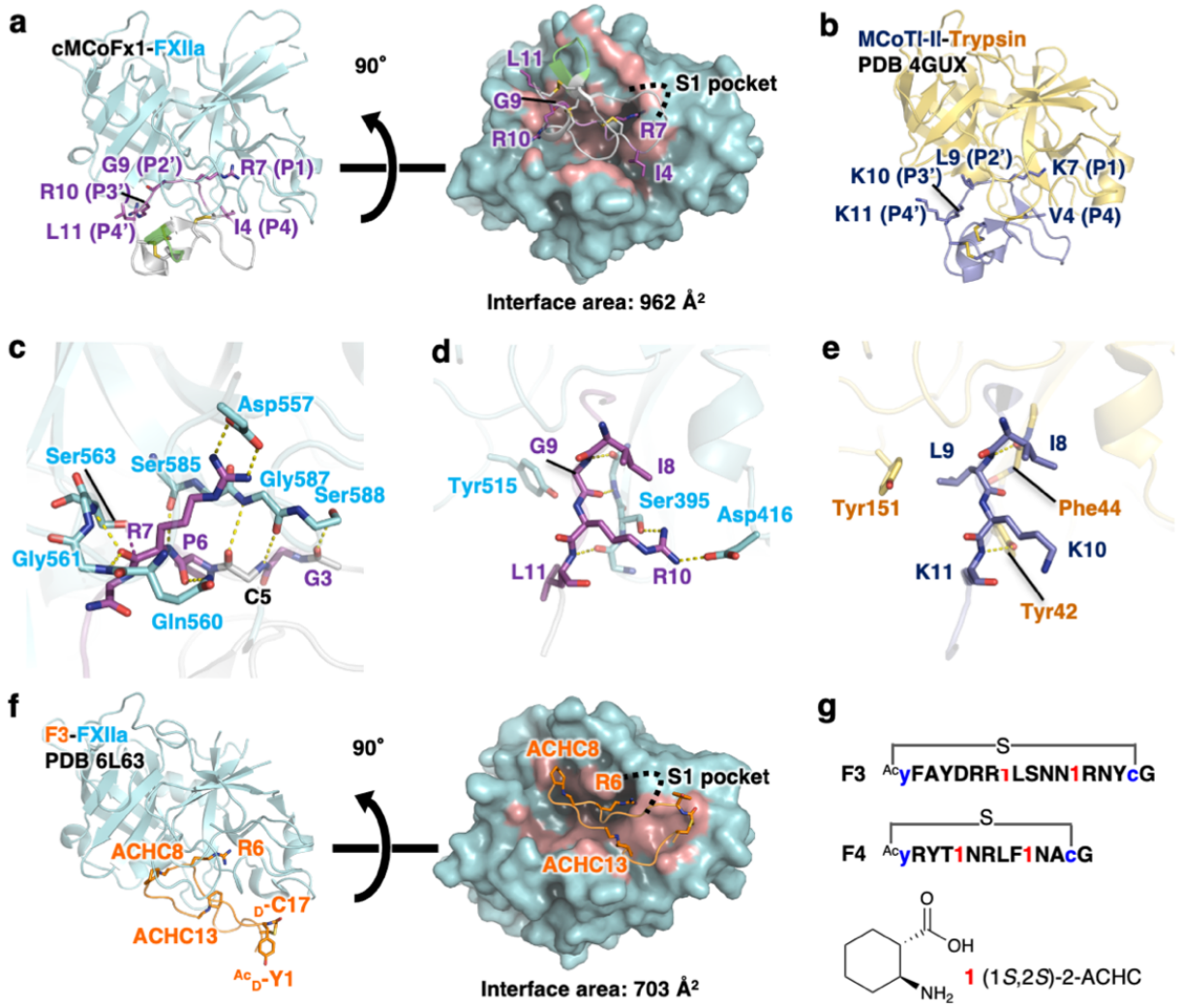
2.2.8. Serum stability of cMCoFx1 and its variants

To determine the proteolytic stability of selected peptides, stability of aMCoFx1, cMCoFx1, cMCoTI-fxL1 and native MCoTI-II was measured in human serum for up to 24 h (Figure 2.10). All tested peptides showed high serum stability with half-lives of more than 24 hours. On the contrary, a linear control peptide (EAIYAAPFAKKK) was tested in parallel and rapidly degraded in serum within 1 h. The results clearly showed that all three cystine-knotted variants have similar stability to navie MCoTI-II, irrespective of backbone cyclization.

As mentioned above, our laboratory has recently developed two nonstandard cbAA-containing mono-macrocyclic FXIIa inhibitors, F3 and F4 (Figure 2.9g), by means of the RaPID selection.[130] Strikingly, cMCoFx1 exhibits approximately 5-fold and 26-fold higher inhibitory activity than F3 and F4, respectively. The half-life ($t_{1/2}$) of cMCoFx1 was about 24 h, while $t_{1/2}$ of the nonstandard macrocycle F3 was about 60 h. The more compact macrocycle F4 exhibited even higher serum stability with $t_{1/2}$ over 285 h (>12 days). The greater proteolytic stability observed for F3 and F4 could be attributed to the existence of the unique β- and _D-amino acids as well as the structurally

compact monocycle scaffold. It should be emphasized even though cMCoFx1 consisted of only canonical amino acids, it had a remarkable serum stability. Clearly, the compact cystine knot structure of cMCoFx1 significantly contributes to the proteolytic resistance, yet its solvent exposed loops could leave the unwanted protease accessibility. I envision that the proteolytic stability of cMCoFx1 can be improved by either semi-rational mutations or a saturation mutagenesis display,[131] somewhat similar to recent elegant reports of bicyclic peptides conducted by the Heinis group.[132]
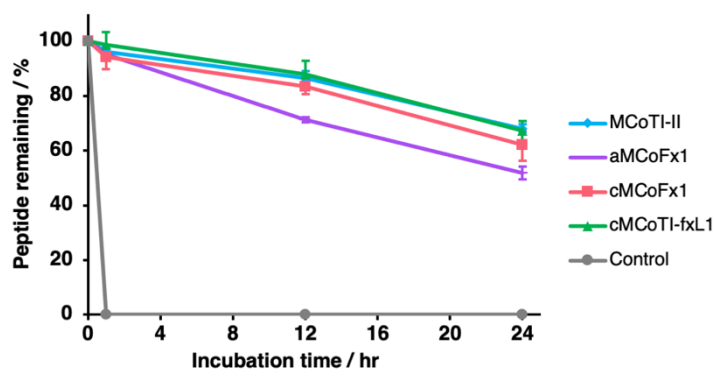
Figure 2.10: Stability of MCoTI-II (blue), aMCoFx1 (purple), cMCoFx1 (red), and loop-grafted variants cMCoTI-fxL1 (green) in human serum. Control (gray) indicates a linear peptide with sequence of EAIYAAPFAKKK which was fully degraded within 1 h. Time course of the percentage of remaining peptide were recorded when incubated in 100% human serum at 37 °C for up to 24 h. Results are the mean ± SEM from three replicates. Reproduced with permission from *J. Am. Chem. Soc.* **2021**, *143*, 18481–18489, Copyright 2021, American Chemical Society.

2.2.9. Saturation mutagenesis scanning of MCoFx1

Inspired by the discovery of potent FXIIa-inhibiting peptide cMCoFx1 derived from the previous mRNA display, I was interested in whether the inhibitory potency can be increased by further modifying residues except from loop 1 and 5. Therefore, I performed a saturation mutagenesis scanning of the selected MCoFx1 sequence. The mutant library was designed based on the mRNA display selected sequence, MCoFx1, with a random residue (containing any of 20 natural amino acids) replacing the parental residue in the cyclotide core positions in each mutant sequence. The parent and mutants were panned against FXIIa following the selection protocol, after which the library was separated into "binding" and "non-binding" fractions which was sequenced to reveal the bias of residues favored by FXIIa. As previously described,[133] I utilized

data from both the "binding" and "non-binding" fractions to increase the signal response and overall accuracy of the method. For every library mutant, Y-score was defined as the ratio of peptide's frequencies in "binding" and "non-binding" populations, and W-score was defined by subtracting the $\log_2 Y$ value of parent MCoFx1 from that of every mutant. W-scores of the single-position mutants were referred as Figure 2.10. Reporting of the W-scores makes for a uniform perception of the results, with higher scores corresponding to binding-beneficial mutations.

As indicated by the W-scores, the mutations in loop1 mainly decreased the binding affinity to FXIIa, whereas mutations at A26 to T/S/P/K/R and R28 to H/G in loop 5 slightly improved the mutant's target-binding affinity. The relatively low mutating tolerance of loop 1 and 5 indicated that our previous mRNA display process was efficient for generating the optimized FXIIa binders. However, interestingly, I observed several beneficial mutations in loop 2, 3, and 6, including R13, R14 and P19 to hydrophobic residues (I/L) or aromatic residues (F/Y), D1 to almost all other residues, and G2, G3, G33 and S34 to K/R. Combining with the co-crystal structure analysis of cMCoFx1-FXIIa, the saturation mutagenesis results suggested more possible binding interactions between MCoFx1 variants and FXIIa, therefore providing hints for our further inhibitor design.
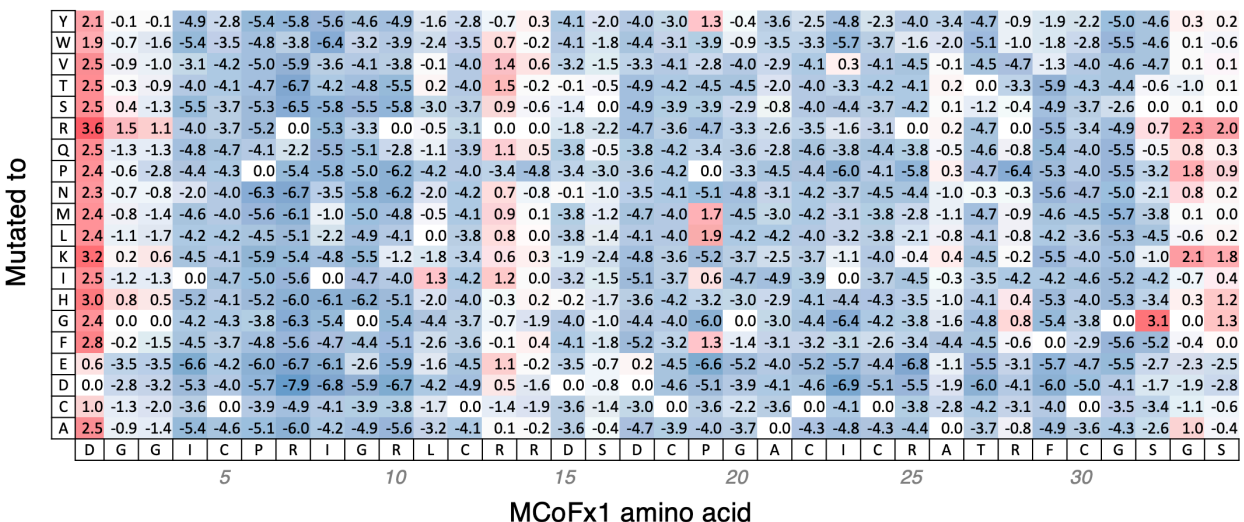
Mutated to	D	G	G	I	C	P	R	I	G	R	L	C	R	R	D	S	D	C	P	G	A	C	I	C	R	A	T	R	F	C	G	S	G	S
Y	2.1	-0.1	-0.1	-4.9	-2.8	-5.4	-5.8	-5.6	-4.6	-4.9	-1.6	-2.8	-0.7	0.3	-4.1	-2.0	-4.0	-3.0	1.3	-0.4	-3.6	-2.5	-4.8	-2.3	-4.0	-3.4	-4.7	-0.9	-1.9	-2.2	-5.0	-4.6	0.3	0.2
W	1.9	-0.7	-1.6	-5.4	-3.5	-4.8	-3.8	-6.4	-3.2	-3.9	-2.4	-3.5	0.7	-0.2	-4.1	-1.8	-4.4	-3.1	-3.9	-0.9	-3.5	-3.3	-5.7	-3.7	-1.6	-2.0	-5.1	-1.0	-1.8	-2.8	-5.5	-4.6	0.1	-0.6
V	2.5	-0.9	-1.0	-3.1	-4.2	-5.0	-5.9	-3.6	-4.1	-3.8	-0.1	-4.0	1.4	0.6	-3.2	-1.5	-3.3	-4.1	-2.8	-4.0	-2.9	-4.1	0.3	-4.1	-4.5	-0.1	-4.5	-4.7	-1.3	-4.0	-4.6	-4.7	0.1	0.1
T	2.5	-0.3	-0.9	-4.0	-4.1	-4.7	-6.7	-4.2	-4.8	-5.5	0.2	-4.0	1.5	-0.2	-0.1	-0.5	-4.9	-4.2	-4.5	-4.5	-2.0	-4.0	-3.3	-4.2	-4.1	0.2	0.0	-3.3	-5.9	-4.3	-4.4	-0.6	-1.0	0.1
S	2.5	0.4	-1.3	-5.5	-3.7	-5.3	-6.5	-5.8	-5.5	-5.8	-3.0	-3.7	0.9	-0.6	-1.4	0.0	-4.9	-3.9	-3.9	-2.9	-0.8	-4.0	-4.4	-3.7	-4.2	0.1	-1.2	-0.4	-4.9	-3.7	-2.6	0.0	0.1	0.0
R	3.6	1.5	1.1	-4.0	-3.7	-5.2	0.0	-5.3	-3.3	0.0	-0.5	-3.1	0.0	0.0	-1.8	-2.2	-4.7	-3.6	-4.7	-3.3	-2.6	-3.5	-1.6	-3.1	0.0	0.2	-4.7	0.0	-5.5	-3.4	-4.9	0.7	2.3	2.0
Q	2.5	-1.3	-1.3	-4.8	-4.7	-4.1	-2.2	-5.5	-5.1	-2.8	-1.1	-3.9	1.1	0.5	-3.8	-0.5	-3.8	-4.2	-3.4	-3.6	-2.8	-4.6	-3.8	-4.4	-3.8	-0.5	-4.6	-0.8	-5.4	-4.0	-5.5	-0.5	0.8	0.3
P	2.4	-0.6	-2.8	-4.4	-4.3	0.0	-5.4	-5.8	-5.0	-6.2	-4.2	-4.0	-3.4	-4.8	-3.4	-3.0	-3.6	-4.2	0.0	-3.3	-4.5	-4.4	-6.0	-4.1	-5.8	0.3	-4.7	-6.4	-5.3	-4.0	-5.5	-3.2	1.8	0.9
N	2.3	-0.7	-0.8	-2.0	-4.0	-6.3	-6.7	-3.5	-5.8	-6.2	-2.0	-4.2	0.7	-0.8	-0.1	-1.0	-3.5	-4.1	-5.1	-4.8	-3.1	-4.2	-3.7	-4.5	-4.4	-1.0	-0.3	-0.3	-5.6	-4.7	-5.0	-2.1	0.8	0.2
M	2.4	-0.8	-1.4	-4.6	-4.0	-5.6	-6.1	-1.0	-5.0	-4.8	-0.5	-4.1	0.9	0.1	-3.8	-1.2	-4.7	-4.0	1.7	-4.5	-3.0	-4.2	-3.1	-3.8	-2.8	-1.1	-4.7	-0.9	-4.6	-4.5	-5.7	-3.8	0.1	0.0
L	2.4	-1.1	-1.7	-4.2	-4.2	-4.5	-5.1	-2.2	-4.9	-4.1	0.0	-3.8	0.8	0.0	-3.8	-1.4	-4.1	-4.0	1.9	-4.2	-4.2	-4.0	-3.2	-3.8	-2.1	-0.8	-4.1	-0.8	-4.2	-3.6	-5.3	-4.5	-0.6	0.2
K	3.2	0.2	0.6	-4.5	-4.1	-5.9	-5.4	-4.8	-5.5	-1.2	-1.8	-3.4	0.6	0.3	-1.9	-2.4	-4.8	-3.6	-5.2	-3.7	-2.5	-3.7	-1.1	-4.0	-0.4	0.4	-4.5	-0.2	-5.5	-4.0	-5.0	-1.0	2.1	1.8
I	2.5	-1.2	-1.3	0.0	-4.7	-5.0	-5.6	0.0	-4.7	-4.0	1.3	-4.2	1.2	0.0	-3.2	-1.5	-5.1	-3.7	0.6	-4.7	-4.9	-3.9	0.0	-3.7	-4.5	-0.3	-3.5	-4.2	-4.2	-4.6	-5.2	-4.2	-0.7	0.4
H	3.0	0.8	0.5	-5.2	-4.1	-5.2	-6.0	-6.1	-6.2	-5.1	-2.0	-4.0	-0.3	0.2	-0.2	-1.7	-3.6	-4.2	-3.2	-3.0	-2.9	-4.1	-4.4	-4.3	-3.5	-1.0	-4.1	0.4	-5.3	-4.0	-5.3	-3.4	0.3	1.2
G	2.4	0.0	0.0	-4.2	-4.3	-3.8	-6.3	-5.4	0.0	-5.4	-4.4	-3.7	-0.7	-1.9	-4.0	-1.0	-4.4	-4.0	-6.0	0.0	-3.0	-4.4	-6.4	-4.2	-3.8	-1.6	-4.8	0.8	-5.4	-3.8	0.0	3.1	0.0	1.3
F	2.8	-0.2	-1.5	-4.5	-3.7	-4.8	-5.6	-4.7	-4.4	-5.1	-2.6	-3.6	-0.1	0.4	-4.1	-1.8	-5.2	-3.2	1.3	-1.4	-3.1	-3.2	-3.1	-2.6	-3.4	-4.4	-4.5	-0.6	0.0	-2.9	-5.6	-5.2	-0.4	0.0
E	0.6	-3.5	-3.5	-6.6	-4.2	-6.0	-6.7	-6.1	-2.6	-5.9	-1.6	-4.5	1.1	-0.2	-3.5	-0.7	0.2	-4.5	-6.6	-5.2	-4.0	-5.2	-5.7	-4.4	-6.8	-1.1	-5.5	-3.1	-5.7	-4.7	-5.5	-2.7	-2.3	-2.5
D	0.0	-2.8	-3.2	-5.3	-4.0	-5.7	-7.9	-6.8	-5.9	-6.7	-4.2	-4.9	0.5	-1.6	0.0	-0.8	0.0	-4.6	-5.1	-3.9	-4.1	-4.6	-6.9	-5.1	-5.5	-1.9	-6.0	-4.1	-6.0	-5.0	-4.1	-1.7	-1.9	-2.8
C	1.0	-1.3	-2.0	-3.6	0.0	-3.9	-4.9	-4.1	-3.9	-3.8	-1.7	0.0	-1.4	-1.9	-3.6	-1.4	-3.0	0.0	-3.6	-2.2	-3.6	0.0	-4.1	0.0	-3.8	-2.8	-4.2	-3.1	-4.0	0.0	-3.5	-3.4	-1.1	-0.6
A	2.5	-0.9	-1.4	-5.4	-4.6	-5.1	-6.0	-4.2	-4.9	-5.6	-3.2	-4.1	0.1	-0.2	-3.6	-0.4	-4.7	-3.9	-4.0	-3.7	0.0	-4.3	-4.8	-4.3	-4.4	0.0	-3.7	-0.8	-4.9	-3.6	-4.3	-2.6	1.0	-0.4

MCoFx1 amino acid

Figure 2.11: W-scores of the single-position mutants of cMCoFx1 towards FXIIa.

2.3. Conclusions

In this study, I used *in vitro* mRNA display for the discovery of potent FXIIa inhibitors from a cyclotide-based library containing two randomized loops. The most potent cyclotide analogue, cMCoFx1, inhibits FXIIa with a K_i of 370 pM, shows selectivity of more than three orders of magnitude over most other serine proteases, and potently inhibits coagulation via the intrinsic

pathway in human plasma. Analysis of a co-crystal structure of cMCoFx1-FXIIa revealed substrate-like tight binding, with interactions between loop 1 of cMCoFx1 and FXIIa providing the largest contribution to the inhibitor's high potency and exquisite specificity. The structural stability preinstalled in the cyclotide scaffold facilitated the generation of not only potent but also proteolytically stable FXIIa inhibitors. Most interestingly, cMCoFx1 exhibits an inhibitory potency that approaches the level of MCoTI-II for trypsin, a product of evolution over millions of years, highlighting the immense potential of mRNA display as a method for rapidly evolving potent cyclotide variants with potential pharmaceutical value.

2.4. Materials and methods

Chemicals and reagents

All chemical reagents were purchased from Watanabe Chemical Industry, Nacalai Tasque, Tokyo Chemical Industry, Sigma-Aldrich Japan or Wako. Unless otherwise noted, all the chemical reagents obtained from commercial sources were used without any purification. H_2O used for buffer preparations was from a Sartorius filtration system (18.2Ω).

MCoTI-II-based library generation

Oligonuclotides corresponding to the designed library were ordered from Eurofins Genomics, with the 12 random residues in loop 1, 5 and 6 encoded with NNK codon (N = A, T, G, C and K = G, T). The DNA library was constructed in a two-step PCR reaction using Q5 high-fidelity DNA polymerase (New England Biolabs), with the first extension step extending two pieces of oligos, MCoTI-II-NNK7.F80 and MCoTI-II-NNK5.R82, containing the whole peptide-coding region, while the second amplification step added upstream T7 promoter, GGG triplet, epsilon sequence and ribosome binding (Shine-Dalgarno) sequence and downstream puromycin linker binding sequence. The PCR reaction was conducted in $1\times$ Q5 reaction buffer (New England Biolabs), 200 µM each dNTPs, 0.5 µM forward and reverse primers, 1% (v/v) 1x Q5 High-Fidelity DNA Polymerase (New England Biolabs). To reach the library diversity of 10^{14} molecules, the first-step PCR was conducted in 650 µL scale and added directly to the second-step PCR (6500 µL scale), together with the other PCR recipes. The PCR products were extracted by phenol/chloroform, precipitated by ethanol, dissolved in 650 µL water, and used for *in vitro* transcription at 37 °C for 16 h in a 6500 µL reaction scale. The *in vitro* transcription reaction mixture contained 40 mM Tris-HCl, 1 mM spermidine, 0.01% (v/v) Triton X-100, 10 mM DTT, 30 mM $MgCl_2$, 5 mM NTPs, 30 mM KOH, 650 µL template DNA solution, 0.12 µM home-made T7 RNA polymerase at pH 8.0.

The resulting mRNA transcripts were precipitated by adding 10% (v/v) 3 M NaCl and 80% (v/v) of isopropanol followed by centrifuge. After wash with 70% (v/v) EtOH, the pellet was dissolved in water with 10% (v/v) volume of transcription reaction and equivalent volume of 2x RNA loading buffer (8 M urea, 2 mM Na$_2$EDTA.2H$_2$O, 2 mM Tris-HCl, pH 7.5) was added. After heating at 95 °C for 2 min, the mRNA was purified by 8% (v/v) polyacrylamide gel containing 6 M urea. The correct band was visualized by UV light, cut out, and extracted in 0.3 M NaCl solution for at least 5 hours. After removal of gel by centrifuge and filtration, twice volume of EtOH was added and the mRNA was precipitated by centrifuge at 13000 rpm for 15 min. The pellet was wash with 70% (v/v) EtOH, dried at room temperature, and dissolved in H$_2$O to 10 µM.

Oligonucleotides for library assembly and mRNA display	
MCoTI-II-NNK7.F80	CTTTAAGAAGGAGATATACATATGGATGGTGGCNNKTGCNNKNNKNNKNNKNNKTGCCGTCGTGATT CTGACTGCCCC
MCoTI-II-NNK5.R82	GCCCCCCGTCCTAGCTGCCCGAACCTGACCCGCAMNNMNNMNNMNNMNNGCAAATGCACGCACCGGGGCA GTCAGAATCACG
T7g10M.F46	TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATA
CGS3an13.R22	TTTCCGCCCCCCGTCCTAGCTG
Puromycin linker	CTCCCGCCCCCCGTCC-(PEG18)5-CC-Pu

mRNA display against FXIIa

mRNA template of MCoTI-II-based library was covalently linked to a puromycin linker using home-made T4 RNA ligase, before *in vitro* translated using a translation cocktail as previously described.[134] In brief, the translation mixture consisted of 50 mM HEPES–KOH (pH 7.6), 100 mM potassium acetate, 12.3 mM magnesium acetate, 2 mM ATP, 2 mM GTP, 1 mM CTP, 1 mM UTP, 20 mM creatine phosphate, 2 mM spermidine, 1 mM dithiothreitol, 100 µM 10-formyl-5,6,7,8-tetrahydrofolic acid, 1.5 mg ml^{-1} *E. coli* total tRNA, 1.2 µM *E. coli* ribosome, 0.6 µM methionyl-tRNA formyltransferase, 2.7 µM IF1, 0.4 µM IF2, 1.5 µM IF3, 0.26 µM EF-G, 10 µM EF-Tu/EF-Ts complex, 0.25 µM RF2, 0.17 µM RF3, 0.5 µM RRF, 4 µg ml^{-1} creatine kinase, 3 µg ml^{-1} myokinase, 0.1 µM inorganic pyrophosphatase, 0.1 µM nucleotide diphosphate kinase, 0.1 µM T7 RNA polymerase, 0.73 µM AlaRS, 0.03 µM ArgRS, 0.38 µM AsnRS, 0.13 µM AspRS, 0.02 µM CysRS, 0.06 µM GlnRS, 0.23 µM GluRS, 0.09 µM GlyRS, 0.02 µM HisRS, 0.4 µM IleRS, 0.04 µM LeuRS, 0.11 µM LysRS, 0.03 µM MetRS, 0.68 µM PheRS, 0.16 µM ProRS, 0.04 µM SerRS, 0.09 µM ThrRS, 0.03 µM TrpRS, 0.02 µM TyrRS and 0.02 µM ValRS, 0.5 mM each of Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr Trp, Tyr and Val, and 1.2 µM mRNA library conjugated to a puromycin linker. The *in vitro* translation was

performed at 37 °C for 45 min in 150 µl (for the first round of selection) or 10 µl (from the second to fourth rounds) scale. The reaction mixture was incubated at room temperature for 12 min, and a 0.2× volume of 100 mM EDTA (pH 8.0) was added and incubated at 37 °C for 30 min to induce the dissociation of ribosomes from the mRNA-peptide conjugates. Reverse transcription was carried out at 42 °C for 15 min using the CGS3an13.R22 primer and M-MLV reverse transcriptase lacking RNase H activity (Promega). Following reverse transcription, the mRNA:cDNA-peptide library was panned through a 2× translation volume of Dynabeads M-280 Streptavidin (Thermo Fisher, half-saturated with biotin) and incubated at 4 °C for 15 min, three times as a negative selection. Note that the negative selection was not performed at the first round. The supernatant retrieved from negative selection was added to 1× translation volume of Dynabeads immobilized with biotinylated human β-factor XIIa (Molecular Innovations, beads loading: 2 pmol protein/ul beads) and the mixture was incubated at 4°C for 30 min (positive selection). The beads were washed with 100 µL of cold PBST (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM of KH2PO4, 0.05% (v/v) Tween-20) for three times and the cDNA was eluted from the beads by heating to 95°C for 5 min in 100 µL of 1× PCR buffer (10 mM Tris–HCl (pH 9.0), 50 mM KCl, 0.1% (v/v) Triton X-100, 0.25 mM dNTP, 2.5 mM $MgCl_2$, 0.25 µM T7g10M.F46 and CGS3an13.R22 primers, and amplified by PCR. The elute (1 µl) was mixed with 19 µl of 1× PCR buffer that contained SYBR Green I and *Taq* DNA polymerase and the amount of cDNAs was quantified by real-time PCR. The rest elute was extracted by phenol/chloroform, precipitated by ethanol, dissolved in 10 µL water, and used for *in vitro* transcription of the subsequent round with the same recipe as preparation of the library. Scheme of an integrated round of selection is illustrated in Figure S2.

Solid phase peptide synthesis

Acyclic and cyclic forms of peptides:

    MCoFx1 DGGICPRIGRLCRRDSDCPGACICRATRFCGSGY,

    MCoFx2 DGGICPRILVYCRRDSDCPGACICIRRTYCGSGS,

    MCoFx3 DGGRCPRLLRWCRRDSDCPGACICARGGLCGSGS,

    MCoFx4 DGGVCPRVGWRCRRDSDCPGACICYPTKWCGSGS,

    MCoFx5 DGGRCCGGYLVCRRDSDCPGACICVFKKHCGSGS;

and cyclic form of peptides:

    cMCoTI-fxL1 DGGICPRIGRLCRRDSDCPGACICRGNGYCGSGS,

    cMCoTI-fxL5 DGGVCPKILKKCRRDSDCPGACICRATRFCGSGS

were synthesized using standard 9-fluorenylmethyl carbamate (Fmoc) solid-phase peptide synthesis chemistry as previously described.[128] In brief, each sequence was assembled on 2-chlorotrityl chloride resin (Chem-Impex, 0.45 mmol eq/g) using an automated peptide synthesizer (Symphony, Protein Technologies Inc.). Couplings were performed twice with 4 equiv of Fmoc-protected amino acids, 4 equiv of O-(6-chlorobenzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU), and 8 equiv of N,N-diisopropylethylamine (DIPEA) in DMF for 10 min. Removal of the Fmoc group was achieved using 30% piperidine in DMF (1 min). For cyclic form of each sequence, side-chain protected peptides were cleaved from the solid support using several resin bed volume washes with 1% trifluoracetic acid (TFA) in dichloromethane followed by lyophilization. Backbone cyclization was performed in DMF (50 mL per 0.1 mmol peptide) using 4 equiv of 1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) and 8 equiv of DIPEA for 6 h, followed by extraction and lyophilization. Cyclic precursor peptides were deprotected in a cocktail containing TFA/triisopropylsilane/water (95:2.5:2.5, v/v) and purified using RP-HPLC. Intramolecular disulfide bonds were formed in 0.1 M ammonium bicarbonate buffer (pH 8.5) by vigorous stirring at room temperature overnight. For acyclic form of each sequence, peptides were cleaved from the solid support and the side chains were deprotected using a cleavage cocktail containing TFA/triisopropylsilane/water (95:2.5:2.5, v/v) for 2 h, followed by precipitation in diethyl ether and purification using RP-HPLC. Formation of intramolecular disulfide bonds was performed as described above. Purification of all peptides was performed on a Prominence HPLC system (Shimadzu) using a water:acetonitrile gradient (with 0.05% TFA). The purity and integrity of the synthetic peptides were verified using LC-MS (LCMS-2020, Shimadzu) and the results are shown in Figure 2.3c.

¹H NMR spectroscopic characterization

Lyophilized peptides (purity > 95%) were dissolved in 90% H_2O/10% D_2O (v/v) to approximately 1 mM. ¹H one- and two-dimensional TOCSY (total correlation spectroscopy) and NOESY (nuclear Overhauser effect spectroscopy) spectra of MCoTI-II, MCoFx1-5 in both acyclic and cyclic forms, cMCoTI-fxL1, and cMCoTI-fxL5 were acquired using an Avance-600 MHz spectrometer (Bruker) at 25°C. The mixing time was 80 ms and 200 ms for TOCSY and NOESY, respectively. Spectra were internally referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) at 0.00 ppm and analyzed using CcpNMR Analysis V2. The α-proton secondary chemical

shifts of peptides were calculated by subtracting random coil chemical shifts as reported previously[135] from the experimentally observed shifts.

Surface plasmon resonance for binding affinity determination

Binding kinetics of each peptide towards human β-FXIIa were determined using a Biacore T200 machine (Cytiva). Running buffer was HBS-EP+ (10 mM HEPES, 150 mM NaCl, 3 mM EDTA and 0.05% (v/v) surfactant P20, pH 7.4). 1 μM biotinylated human β-FXIIa was immobilized on a Sensor Chip CAP (Cytiva) using Biotin CAPture Reagent (Cytiva) following the manufacturer's instructions. A series of concentrations of each peptide were injected as analyte and the binding kinetics were modeled using a 1:1 binding model.

Protease inhibition assays

The activity of acyclic and cyclic MCoFx1–5 was assessed in competitive inhibition assays, as previously described.[25] Human β-FXIIa was obtained from Molecular Innovations, together with human FXa, human FXIa, human α-thrombin, and human plasma kallikrein. Human cationic trypsin and human plasmin were obtained from Sigma-Aldrich, and recombinant human matriptase was sourced from R&D Systems. Recombinant human uPA and recombinant human tPA were expressed in Expi293 cells, as previously described.[136] FXIIa inhibition assays were performed in clear, low-binding 96-well plates using 250 μL (final volume) of assay buffer (0.1 M Tris-HCl pH 8.0, 0.1 M NaCl, 10 mM $CaCl_2$ and 0.005% Triton X-100). Inhibitors were serially diluted and incubated with 10 nM FXIIa at room temperature to reach equilibrium. After adding 100 μM Ac-QRFR-pNA substrate (K_M = 162 μM), enzymatic activity was measured by monitoring release of the pNA moiety using a TECAN infinite M1000 Pro plate reader (λ = 405 nm, reading interval, 10 s; assay time course, 7 min). Assays were conducted in three times in duplicate. Substrate kinetic constants (Michaelis-Menten) and inhibition constants (Morrison K_i) were determined by non-linear regression using Prism 7 (GraphPad). Selectivity assays with off-target proteases were performed in a similar way, except that different concentrations of enzyme and substrate were used, as indicated in Table 2.1. For assays using a fluorescent substrate (peptide-4-methylcoumaryl-7-amide, MCA), black low-binding 96-well plates were used and activity was measured using a TECAN infinite M1000 pro plate reader (λ_{ex} 360 nm, λ_{em} 460 nm, reading interval, 30 s; assay time course, 5–30 min).

Table 2.1: Protease-specific conditions and substrates for inhibition activity experiment.

Enzyme[a]	Conc.	Substrate	Conc.	K_M	Buffer variation[b]
Human β-FXIIa	10 nM[c]	Ac-QRFR-pNA	100 μM	162 μM	10 mM $CaCl_2$
Human trypsin	0.1 nM	Boc-VPR-MCA	5 μM	3.8 μM	10 mM $CaCl_2$
Human FXa	0.2 nM	Z-Pyr-GR-MCA	50 μM	-	10 mM $CaCl_2$
Human FXIa	1 nM	Boc-Glu(OBzl)-AR-MCA	100 μM	-	10 mM $CaCl_2$
Human thrombin	0.1 nM	Boc-VPR-MCA	40 μM	-	10 mM $CaCl_2$
Human plasma kallikrein	0.5 nM	Z-FR-MCA	50 μM	-	-
Human plasmin	1 nM	Ac-RM(O_2)YR-pNA	75 μM	24 μM	-
Rec. human matriptase	0.15 nM	Boc-QAR-MCA	10 μM	42 μM	-
Rec. human uPA	10 nM	Z-Pyr-GR-MCA	100 μM	-	10 mM $CaCl_2$
Rec. human tPA	10 nM	Z-Pyr-GR-MCA	100 μM	-	10 mM $CaCl_2$

[a]Enzymes purified from human plasma, human tissue (trypsin) or recombinantly expressed (indicated by rec.)

[b]Assay buffer was 0.1 M Tris-HCl pH 8.0, 0.1 M NaCl, 0.005% Triton X100 with 10 mM $CaCl_2$ included for the indicated enzymes

[c]Assays with high-affinity inhibitors (MCoFx1 and MCoFx1-L1) were performed with 4.2 nM enzyme

Coagulation assays

Citrated, platelet-poor plasma samples from healthy adults were collected, prepared, and stored using previously established collection protocols[137] under Research Agreement #18-03QLD-09 and University of Queensland Human Ethics Committee Approval #2016000256.

Standard coagulation assays were used to determine the effect of the inhibitors on the clotting time of human plasma. Activated partial thromboplastin time (aPTT) uses a Kaolin reagent (clay) to activate the intrinsic/contact pathway of the coagulation cascade, whereas prothrombin time (PT) uses Neoplastine (lyophilized thromboplastin prepared from rabbit cerebral tissue) to activate the extrinsic/tissue-factor pathway of the coagulation cascade. A minor adjustment was made to both tests (addition of 25 µL inhibitor or buffer; Table 2.2) to accommodate the addition of an inhibitor without changing the total volume (and therefore relative ratios of additives) of the assay. Concentrations of inhibitor are reported as the concentration at the point of incubation in the assay, i.e. not taking into account the addition of the start reagent (calcium for aPTT, or Neoplastine for PT).

Clotting time (seconds) of plasma was automatically measured using a STA-R Max® analyzer (Stago, Asnières sur Seine, France). Measurements were conducted using a viscosity-based (mechanical) detection system, whereby opposing magnets oscillate a small metal spherical pellet inside the test cuvette (250 µL total volume) until a clot is formed. Dilution of the inhibitor in Owren–Koller (OK) buffer for dose-response curves was performed automatically by the machine. Reagents were kept at 15–19 °C in the machine during experimentation and otherwise stored at 4 °C. All tests were performed in triplicate.

Table 2.2: Experimental procedures for aPTT and PT assays.

Experiment	Methodology
aPTT*	Step 1: 25 μL inhibitor (solubilized in Owren–Koller (OK) Buffer (isotonic saline, Stago # 00360)) + 50 μL kaolin/phospholipid (Stago # 00597) + 50 μL human plasma
	Step 2: 240 s incubation at 37°C
	Step 3: Addition of 50 μL 0.025 M calcium (Stago # 00367)
PT*	Step 1: 25 μL inhibitor (diluted with OK Buffer) + 50 μL human plasma
	Step 2: 240 s incubation at 37°C
	Step 3: Addition of 100 μL Neoplastine (Stago #00606)

*Control assays replaced 25 μL inhibitor with 25 μL OK buffer.
aPTT = Activated Partial Thromboplastin Time; PT = Prothrombin Time.

Crystallography

Untagged FXIIa (purchased from Molecular Innovations, catalog number HFXIIAB) was concentrated to 10 mg/ml in 10 mM HEPES-Na pH 7.0 and 150 mM NaCl, and then mixed with cMCoFx1 at a molar ratio of 1:3. Crystals of the cMCoFx1-FXIIa complex were grown at 10˚C from drops containing 100 nL of protein sample and 100 nL of reservoir solution (200 mM sodium acetate trihydrate pH 7.0, 20% PEG3350). Crystals were harvested with 200 mM sodium acetate trihydrate pH 7.0, 20% PEG3350, and 20% PEG400, and flash-frozen in liquid nitrogen.

X-ray diffraction data were collected at SPring-8 BL32XU using helical data collection method with a beam size of 10 x 15 mm and oscillation range of 0.1˚. The diffraction data from a single crystal were divided into small wedges (30 ˚ each), and then automatically processed and merged with the programs XDS and KAMO.

The structure of the FXIIa-cMCoFx1 complex was solved by molecular replacement method using the program molrep with the coordinates of free FXIIa (PDB 6B77) as a search model. The density for the bound cMCoFx1 molecule was clearly visible, and the atomic model was manually built with the program Coot. Structure refinement was carried out with the program Phenix. Data

collection and refinement statistics were shown in Table 2.3. Structural figures were prepared with the program PyMOL (Schrödinger, LLC). The interaction interface areas were calculated with the program PISA.

Table 2.3: Data collection and refinement statistics of cMCoFx1-FXIIa co-crystal structure.

	cMCoFx1-FXIIa complex
Data collection	
Space group	$P2_1$
Cell dimensions	
a, b, c (Å)	40.19 76.83 41.02
a, b, g (°)	90, 91.392, 90
Resolution (Å)	41.01–1.99 (2.11–1.99)
Total reflections	65611 (10753)
Unique reflections	17116 (2745)
Multiplicity	3.83 (3.92)
Completeness (%)	99.6 (99.4)
Mean I / sI	8.89 (1.42)
Wilson B-factor	28.87
R_{meas} (%)	13.2 (126.1)
$CC_{1/2}$	0.966 (0.572)
Refinement	
Resolution (Å)	41.01–1.99 (2.061–1.99)
Reflections used in refinement	17104 (1693)
Reflections used for R-free	1699 (165)
R-work	0.1983 (0.2983)
R-free	0.2475 (0.3498)
No. atoms	
FXIIa	1766
cMCoFx1	248
Sugar	75
Water	148
Average B-factors	
FXIIa	33.47
cMCoFx1	38.34
Sugar	53.14
Water	35.51
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.74
Ramachandran favored (%)	97.71
Ramachandran allowed (%)	1.91
Ramachandran outliers (%)	0.38
Rotamer outliers (%)	0.95
Number of TLS groups	6

Values in parentheses are for highest-resolution shell.

Serum stability

To provide a guide for their potential *in vivo* stability, the proteolytic stability of aMCoFx1, cMCoFx1, cMCoTI-fxL1 and the parent peptide MCoTI-II was evaluated in human serum as

described previously.[30] Briefly, peptides were incubated at a final concentration of 30 µM in human serum (human male AB serum, Sigma-Aldrich) at 37°C for 0, 1, 12 and 24 h. The reaction was stopped at designated times, with the serum proteins being denatured through the introduction of two volumes of acetonitrile (v/v). The samples were then spun at 17,000 g for 10 min and the amount of peptides in the supernatant was quantified using analytical HPLC. The percentage of peptide remaining at 1, 12 and 24 h was calculated using the area of the serum-treated peptide peak from 0 h as 100%. A linear peptide (EAIYAAPFAKKK) was included as a control which was faster degraded in serum within 1 h.

Library generation for saturation mutagenesis study

The saturation mutagenesis library was designed based on the mRNA display selected sequence, MCoFx1 (MDGGICPRIGRLCRRDSDCPGACICRATRFCGSGSGS), with a random residue replacing the parental residue at the core positions (DGGICPRIGRLCRRDSDCPGACICRATRFCGSGS) in each mutant sequence. Oligonuclotides corresponding to the parental MCoFx1 and single-position mutants were ordered from Eurofins Genomics, with the random residue encoding with NNK codon (N = A, T, G, C and K = G, T) (refer to Table xxx). The DNA library was constructed in a two-step PCR reaction using home-made KOD polymerase, with the first extension step extending two pieces of oligos, MCoFx1-reg1_original.F89 or MCoFx1-reg1_pn.F89 and MCoFx1.R70 for n = 1-16, or MCoFx1.F72 and MCoFx1-reg2_pn.R87 for n = 17-34. The second amplification PCR added T7g10M.F48 and Long_HA.R63 primers, with the product containing upstream T7 promoter, GGG triplet, epsilon sequence and ribosome binding (Shine-Dalgarno) sequence and downstream HA tag and puromycin linker binding sequence. For each parent or single-position mutated template, the PCR reaction was conducted in 100 mM Tris–HCl (pH 8.0), 6 mM $(NH_4)_2SO_4$, 10 mM KCl, 1% (v/v) Triton X-100, 0.25 mM dNTP, 2.5 mM $MgCl_2$, 0.5 µM forward and reverse primers, and 1% (v/v) 1× home-made KOD polymerase, and the PCR conditions were listed in Table xxx. The first-step extension PCR was conducted in 10 µL scale and added to 100 µL of the second-step amplification PCR reaction. The products were extracted by phenol/chloroform, precipitated by ethanol, dissolved in 20 µL water, and used for *in vitro* transcription at 37 °C for 16 h in a 200 µL reaction scale. The *in vitro* transcription reaction mixture contained 40 mM Tris-HCl, 1 mM spermidine, 0.01% (v/v) Triton X-100, 10 mM DTT, 30 mM $MgCl_2$, 5 mM NTPs, 30 mM KOH, 650 µL template DNA solution, 0.12 µM home-made T7 RNA polymerase at pH 8.0. The resulting mRNA transcripts were precipitated by adding 10% (v/v) 3M NaCl and 80% (v/v) of isopropanol followed

by centrifuge. After wash with 70% (v/v) EtOH, the pellet was dissolved in water and the concentration was adjusted to 10 µM. The parent and single-position mutated mRNA templates were equally mixed, yielding a 10 µM mRNA library for the following target scanning.

Oligonucleotides for saturation mutagenesis library generation	
Common primers	
T7g10M.F48	TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATG
Long_HA.R63	TTTCCGCCCCCCGTCCTAAGAACCAGAACCAGAACCTGCATAGTCGGGCACGTCGTATGGGTA
Short_HA.R36	TTTCCGCCCCCCGTCCTAAGAACCAGAACCAGAACC
MCo_HA_Rd1T7g10M.F55	CACTCTTTCCCTACACGACGCTCTTCCGATCTAACTTTAAGAAGGAGATATACAT
MCo_HA_Rd2R49c.R54	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGTCGGGCACGTCGTATGGGTA
For parental MCoFx1 and reg1_p1-16 templates	
MCoFx1-reg1_original.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGCGGATTGGGCGGCTTTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p1.F89	CTTTAAGAAGGAGATATACATATGNNKGGTGGCATTTGCCCGCGGATTGGGCGGCTTTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p2.F89	CTTTAAGAAGGAGATATACATATGGATNNKGGCATTTGCCCGCGGATTGGGCGGCTTTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p3.F89	CTTTAAGAAGGAGATATACATATGGATGGTNNKATTTGCCCGCGGATTGGGCGGCTTTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p4.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCNNKTGCCCGCGGATTGGGCGGCTTTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p5.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTNNKCCGCGGATTGGGCGGCTTTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p6.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCNNKCGGATTGGGCGGCTTTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p7.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGNNKATTGGGCGGCTTTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p8.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGCGGNNKGGGCGGCTTTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p9.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGCGGATTNNKCGGCTTTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p10.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGCGGATTGGGNNKCTTTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p11.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGCGGATTGGGCGGNNKTGCCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p12.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGCGGATTGGGCGGCTTNNKCGTCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p13.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGCGGATTGGGCGGCTTTGCNNKCGTGATTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p14.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGCGGATTGGGCGGCTTTGCCGTNNKGATTCTGACTGCCCCGGTGCGTG

MCoFx1-reg1_p15.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGCGGATTGGGCGGCTTTGCCGTC GTNNKTCTGACTGCCCCGGTGCGTG
MCoFx1-reg1_p16.F89	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGCGGATTGGGCGGCTTTGCCGTC GTGATNNKGACTGCCCCGGTGCGTG
MCoFx1.R70	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTAGTAGCCCTGCAAATGCACGCACCGGG GCAGTC

For reg2_p17-34 templates

MCoFx1.F72	CTTTAAGAAGGAGATATACATATGGATGGTGGCATTTGCCCGCGGATTGGGCGGCTTTGCCGTC GTGATTCT
MCoFx1-reg2_p17.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTAGTAGCCCTGCAAATGCACGCACCGGG GCAMNNAGAATCACGACGGCAAA
MCoFx1-reg2_p18.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTAGTAGCCCTGCAAATGCACGCACCGGG MNNGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p19.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTAGTAGCCCTGCAAATGCACGCACCMNN GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p20.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTAGTAGCCCTGCAAATGCACGCMNNGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p21.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTAGTAGCCCTGCAAATGCAMNNACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p22.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTAGTAGCCCTGCAAATMNNCGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p23.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTAGTAGCCCTGCAMNNGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p24.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTAGTAGCCCTMNNAATGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p25.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTAGTAGCMNNGCAAATGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p26.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTAGTMNNCCTGCAAATGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p27.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAACCTMNNAGCCCTGCAAATGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p28.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAAAAMNNAGTAGCCCTGCAAATGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p29.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCGCAMNNCCTAGTAGCCCTGCAAATGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p30.R87	GCACGTCGTATGGGTAGCTGCCGCTGCCMNNAAACCTAGTAGCCCTGCAAATGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p31.R87	GCACGTCGTATGGGTAGCTGCCGCTMNNGCAAAACCTAGTAGCCCTGCAAATGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p32.R87	GCACGTCGTATGGGTAGCTGCCMNNGCCGCAAAACCTAGTAGCCCTGCAAATGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p33.R87	GCACGTCGTATGGGTAGCTMNNGCTGCCGCAAAACCTAGTAGCCCTGCAAATGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA
MCoFx1-reg2_p34.R87	GCACGTCGTATGGGTAMNNGCCGCTGCCGCAAAACCTAGTAGCCCTGCAAATGCACGCACCGGG GCAGTCAGAATCACGACGGCAAA

Extension PCR		
94°C	1 min	1 cycle
52°C	1 min	5 cycles
68°C	1 min	
Amplification PCR		
94°C	40 s	12 cycles
50°C	40 s	
68°C	40 s	

HA affinity purification for saturation mutagenesis study

The mRNA library was covalently linked to a puromycin linker, *in vitro* translated, and reverse transcribed as previously described in the mRNA display protocol. To 10 μL of reverse transcription product, 1× volume of 2× blocking buffer was added (2× PBST supplemented with 2 mg/mL bovine serum albumin, where 2× PBST contained 274 mM NaCl, 5.4 mM KCl, 20 mM Na_2HPO_4, 3.6 mM of KH_2PO_4, and 0.1% (v/v) Tween-20). 20 μl of 10 mg/mL Anti-HA beads (Thermo Fisher) were washed twice with PBST before addition of the reverse transcription mixture. Incubation at 4 °C for 60 min ensued, after which the supernatant was discarded, and the beads were washed twice with PBST. Bound mRNA:cDNA-peptide conjugates were eluted from the beads with HA peptide (2 mg/mL in blocking buffer; sequence: NH_2-YPYDVPDYA-$CONH_2$) by incubating the suspension at 37 °C for 15 min, and collecting the supernatant. HA affinity purification was indispensable to remove peptide-unconjugated, puromycin-unligated mRNA/cDNA, and other translation side-products.

Screening of the saturation mutated library agasint FXIIa

To the HA-purified mRNA:cDNA-peptide library, 1× translation volume of Dynabeads immobilized with biotinylated human β-factor XIIa (Molecular Innovations, beads loading: 2 pmol protein/ul beads) was added and the mixture was incubated at 25°C for 3 h, allowing the bining reaction to reach equilibrium. The supernatant was retrieved and stored as "non-binding sample" at –20 °C. The beads were washed twice using 1× blocking buffer at 4°C for 16 h and 25°C for 3 h. Elution of the "binding" cDNA was carried out by heating the beads suspended in 0.1% (v/v) Triton X-100 at 95 °C for 5 min.

Next generation sequencing (NGS) and data analysis for saturation mutagenesis study

Concentrations of recovered "binding" cDNA, as well as the "non-binding" cDNA were determined from a qPCR assay. Recovered sample aliquots were analyzed by qPCR with Taq polymerase in 10 mM Tris–HCl (pH 9.0), 50 mM KCl, 0.1% (v/v) Triton X-100, 0.25 mM dNTP, 2.5 mM MgCl$_2$, 0.25 μM T7g10M.F48 and Short_HA.R36 primers, and home-made Taq polymerase.

Recovered "binding" cDNA, as well as the "non-binding" cDNA were PCR-amplified with Platinum SuperFi DNA Polymerase (Thermo Fisher) using manufacturer's protocol and 0.25 μM MCo_HA_Rd1T7g10M.F55 and MCo_HA_Rd2R49c.R54 primers (refer to Table xxx). Thermal cycling was performed based on the outcomes of qPCR so as to avoid cDNA overamplification. The product was carried forward to the second PCR step, using SuperFi Polymerase and Nextera XT v2 Set (sequences from Illumina) primers to install sequencing barcodes on each sample. The success of PCR was confirmed by 3% agarose gel electrophoresis. After, PCR products were combined and column-purified using a NucleoSpin kit (TaKaRa) adhering to manufacturer's protocol. Concentration of the combined cDNA sample was measured with Qubit (Thermo Fisher) using the dsDNA BR kit. The resulting DNA was appropriately diluted and analyzed by NGS.

Denatured cDNA library (10 pM containing 20% (mol/mol) PhiX Control v3 [Illumina] was sequenced on Illumina's MiSeq instrument in the single read 1×151 cycle mode using v3 chip, collecting data in the .fastq format. The details of the downstream data analysis can be found at https://github.com/avngrdv/FastqProcessor. Briefly, .fastq data files containing base calls were parsed to retrieve DNA sequences, which were in silico translated. Resulting peptide lists were filtered to discard sequences containing ambiguous symbols or not conforming to the library design criteria (overall peptide length). Constant region sequences were trimmed and the variable regions were further filtered to discard double or poly-mutants. The resulting peptides comprised the final data set for each sample. With these data sets, frequencies for each mutant were computed, and then "binding" and "non-binding" frequency matrices were compared to calculate W-scores of each mutant. The W-score is defined as below:

Define f_{mut}, *i.e.* frequency of an individual mutant mut in a sequencing sample as

$$f_{mut} = \frac{c_{mut}}{c_{total}}$$

where c_{mut} is the number of reads corresponding to mutant mut, and c_{total} is the total number of reads in the sample.

Then, Y-score for mutant mut is defined as

$$Y_{mut} = \frac{f_{mut}(\text{binding sample})}{f_{mut}(\text{non-binding sample})}$$

To compare the binding trend of mutants compared with the parental MCoFx1, W-score for mutant mut is defined as

$$W_{mut} = \log_2(Y_{mut}) - \log_2(Y_{MCoFx1})$$

where Y_{MCoFx1} is the Y-score corresponding to parental sequence MCoFx1.

Based on the definition, W-score of the parental MCoFx1 is 0; for mutant mut with $W_{mut} > 0$, the mutation is beneficial for target-binding compared with MCoFx1; for mutant mut with $W_{mut} < 0$, the mutation is disadvantageous for target-binding compared with MCoFx1.

Chapter 3.

# Discovery of mASGR-binding bicyclic peptide with controllable globular conformation

3.1. Introduction

3.1.1. Disulfide connectivity control in DRP discovery

Other than cyclotides, there are plenty of other types of disulfide-rich peptides (DRPs) in nature, derived from plants, insects, and marine organisms. As mentioned in Chapter 1, they are beneficial next-generation drug candidates due to their thermal stability and intrinsic bioactivities. The main discovery methods for DRP-based therapeutics could be divided into two categories, SPPS-based methods, and display technique-based methods. In the past few decades, bioactive DRPs have been discovered by SPPS-based rational design, structural engineering and point mutation scanning. However, due to DRPs' relatively small size (10–50 aa in length), the distance between each cysteine residue is short, leading to the possibility of forming non-native disulfide bonds. Therefore, the main difficulty in synthesizing these DRPs is the formation of the correct disulfide conformation. To solve this problem and yield correctly folded DRPs with higher purity, a few synthetic strategies including oxidative folding and application of cysteine protecting groups are broadly used.

Several DRP folding studies indicate that the correct (or native) formation of the disulfide bond is driven by protein folding.[9,138] In other words, intramolecular disulfide bond formation is considered as an additional element in stabilizing the three-dimensional protein architecture. Therefore, the peptide folding in DRPs mainly depends on their inherent amino acid sequences, meanwhile adjusted by the disulfide connectivity. To form correct disulfide connectivity during laboratory peptide synthesis, synthetic DRPs are subjected to oxidative folding in aqueous solutions that mimic physiological environments (Figure 3.1a).[139] The commonly used reaction buffer is slightly alkaline aqueous buffer (pH 7.5–8.5), with addition of 0.1–1.5 mol/L guanidine hydrochloride, in order to prevent peptide aggregation. The folding reaction is usually carried out in the presence of a redox buffer (GSH/GSSG), allowing intramolecular disulfide reshuffling to form the thermodynamically favored native conformation.[140] Although a large number of DRPs can fold correctly using the oxidative folding strategy, the folding yield could be relatively low. The folding of DRPs containing non-natural mutations could be more difficult than its native sequence.

The other common SPPS-based method of controlling the disulfide connectivity in DRPs is utilizing orthogonal cysteine protecting groups (PGs, Figure 3.1b).[139] In this method, cysteines protected with different thiol PGs are coupled into the peptide chain during SPPS, followed by stepwise deprotection treatments, releasing two free thiols at a time to form the desired disulfide bond. The commonly used cysteine PGs can be divided into several categories based on their deprotection conditions: acid-labile, reducing agent-labile, enzyme-labile, and photo-labile. For example, for synthesis of a cyclotide sequence, cysteines protected with trityl (Trt), acetamidomethyl (Acm), and p-methoxybenzyl (Mob) groups could be released in a stepwise manner using 95% TFA, I_2, and TFMSA/TFA, respectively (Figure 3.1b).[139]
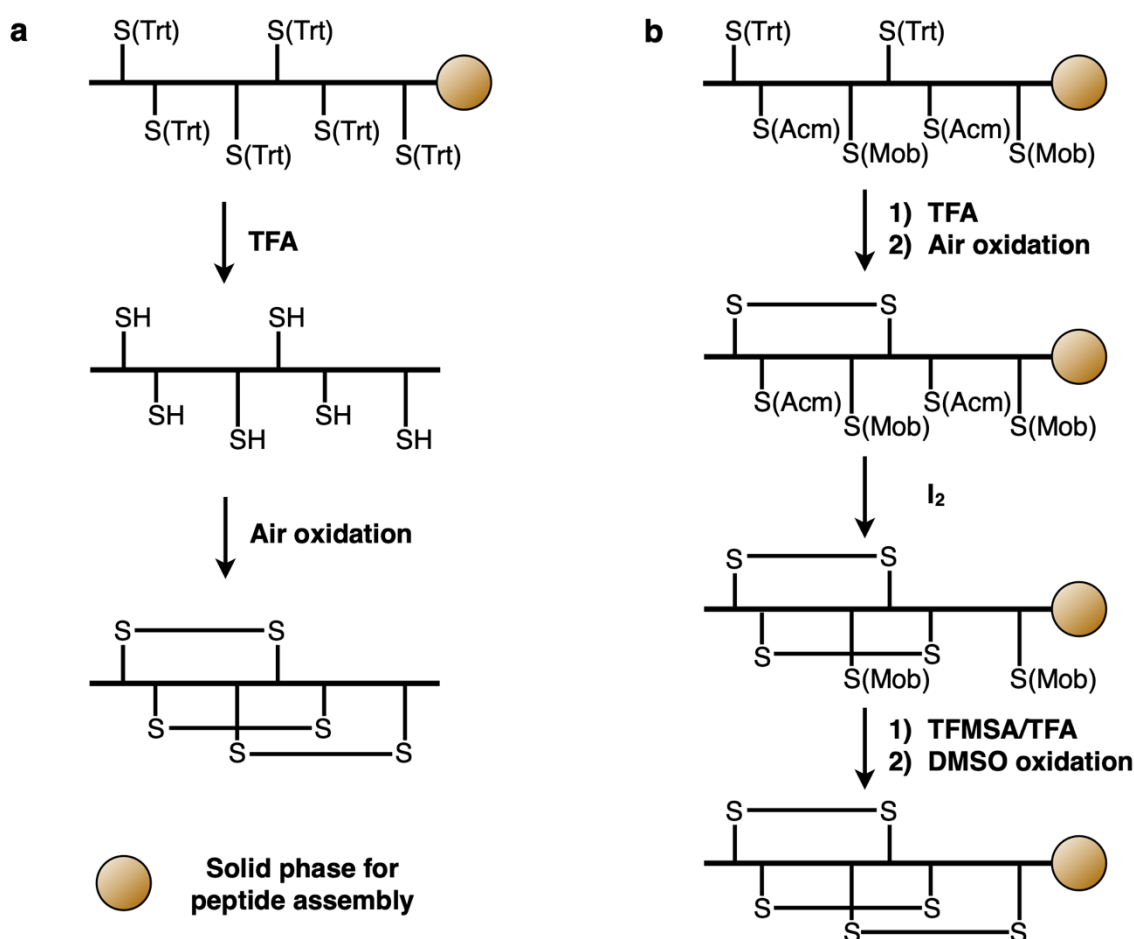
Figure 3.1: Laboratory methods for disulfide formation. (a) Oxidative folding in aqueous solutions mimicking physiological environments; (b) Stepwise formation of disulfide bonds utilizing Cys protecting groups. Reproduced with permission from ref. 139 Copyright 2014 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.[139]

### 3.1.2. Utilizing protected cysteines in FIT system

As introduced in Chapter 1, although SPPS-based rational design and peptide engineering have facilitated discovery of various bioactive DRPs, due to the labor associated with chemical synthesis, the synthetic libraries are always low in diversity (fewer than 10^6 compounds),[141] resulting in non-optimal DRP variants to warrant further therapeutic development. On the contrary, the random nonstandard peptides integrated discovery (RaPID) can afford peptide libraries with diversity of over 10^{12} molecules, meanwhile the genetic code reprogramming during flexible *in vitro* translation (FIT) allows the incorporation of various non-canonical amino acids into the library. I hypothesized that the RaPID selection system, capable of affording substantially higher library diversity, may offer an improved route to identify more bioactive DRPs. However, other than the work mentioned in Chapter 2, only very few studies have put an effort on discovering therapeutic DRPs using RaPID selection, and no DRP library with a certain disulfide connectivity has been reported up to date.

To facilitate the absolute control of disulfide connectivity in a DRP library, protected cysteines could be introduced by using FIT translation. As introduced above, there are plenty of cysteine protecting groups which are broadly used in SPPS, however, since the subsequent RaPID processes contain several enzymatic reactions (i.e. reverse transcription and PCR reactions), the deprotection conditions must be either mild or be isolated from the subsequent enzymatic reactions. The widely used acid-labile thiol PGs, such as Trt, Mmt or Mob, requires removal by strong acids, and therefore not applicable in this case. A novel water-compatible deprotection method of Cys(Acm) has been reported recently by Brik *et. al.*,[142,143] using transition metal palladium (Pd) as the deprotection reagent (Figure 3.2a,b). In order to efficiently recover the peptide, this method used excess amount of diethyldithiocarbamate (DTC) to scavenge the free Pd and Pd bound to the peptide side chains. However, the reaction mechanism is still not clear yet and the plausible mechanism is shown in Figure 3.2c.[143]

Another thiol PG, phenylacetamidomethyl (Phacm), is an analogue of the Acm group, and can be removed by all the Acm deprotection conditions as well. In addition, Phacm could be enzymatically removed by penicillin G acylase (PGA),[144] which catalyzes hydrolysis of the amide bond of the phenylacetamido group. This mild enzyme reaction is conducted at neutral pH in aqueous solution, making it suitable for RaPID system.

No research has been reported on incorporating Cys(Acm) and Cys(Phacm) into peptide using genetic code reprogramming. FIT translation of Cys(Acm) was demonstrated (data unpublished) but the deprotection conditions and the compatibility with *in vitro* selection have not been

illustrated at the outset of this work Based on these backgrounds, I designed this project on controlling the DRP disulfide connectivity using the Cys(Acm).
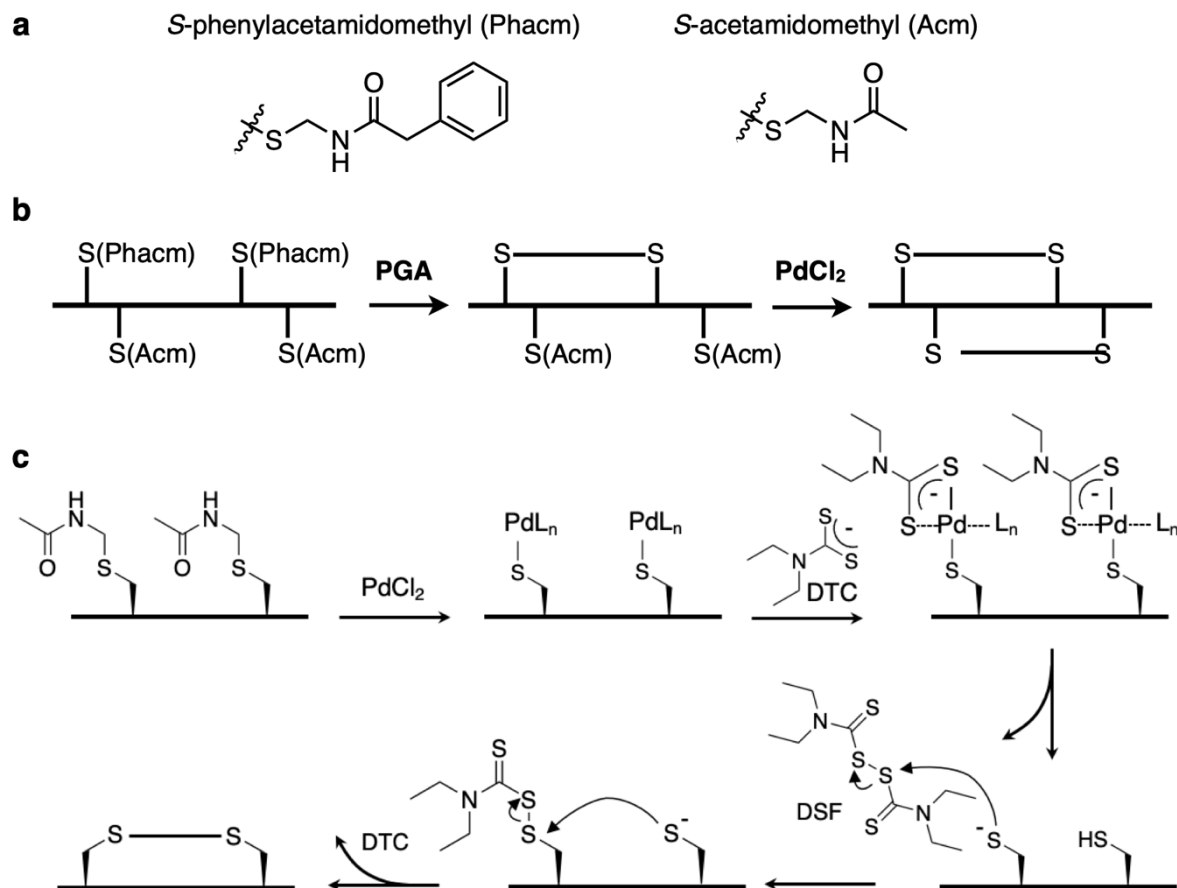
Figure 3.2: Stepwise deprotection strategies used in this work. (a) Structure of Cys-protecting *S*-phenylacetamidomethyl (Phacm) and *S*-acetamidomethyl (Acm) groups; (b) Deprotection methods for stepwise revealing free thiols to form designed disulfide bonds; (c) Plausible mechanism of Acm deprotection by PdCl$_2$. Figure (c) is reproduced with permission from ref. 143 Copyright 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.[143]

3.1.3. Asialoglycoprotein receptor (ASGR) as pharmaceutical target

To test whether *in vitro* translated peptide libraries utilizing Cys(Acm) could be used for discovery of therapeutic ligands, asialoglycoprotein receptor (ASGR) was chosen as the target protein for RaPID selection. ASGR is an integral membrane protein discovered in mammalian hepatocytes, recognizing glycoproteins bearing *N*-acetylgalactosamine (GalNAc) or galactose (Gal)

ligands and internalizes them via endocytosis.[145] Following internalization and endosomal acidification, ASGR releases GalNAc or Gal and recycles back to the plasma membrane, while glycoproteins proceed to the lysosome.[146] This internalization process requires the heterotrimerization of two ASGR subtypes, ASGR1 and 2, with the ratio of 2:1.[147] Though the genes encoding ASGR subtypes were identified in many mammalians, the protein sequences in mouse and human ASGR1/2 have been proved highly homologous: 39% of the amino acids are all identical in all four sequences.[145] The two subtypes from each species are more closely related, with 80% and 62% identical residues in m/hASGR1 and m/hASGR2, respectively.

ASGR has been exploited to achieve specific and efficient delivery of various cargo to the liver, due to its high expression in hepatocytes with 500,000 copies per cell.[145] Upon the high binding affinity of trivalent GalNAc towards ASGR, many tri-GalNAc-modified have been developed, most notably short interfering RNA (siRNA), anti-miRNAs, antisense oligonucleotide (ASO) therapeutics.[148,149] Recently, Bertozzi's group has developed an LYsosome TArgeting Chimeras (LYTACs) by conjugating tri-GalNAc on an antibody, which co-expresses a molecule that binds to the extracellular protein targets, enabling degradation of the undesired protein in liver cells.[150] Development of all these elegant devices has emphasized that ASGR is an impressive research target for generating ligands facilitating cargo internalization or degradation. However, other than the intrinsic ASGR ligands, GalNAc and Gal, there were only a few sugar-mimicking isomers that have been developed and assembled on therapeutic molecules. In this work, I aimed to develop a peptide-based ASGR ligand and create new therapeutics for target internalization or degradation.

3.2. Results and discussion

3.2.1. Optimization of Cys(Acm) deprotection conditions in translated peptide template

To test the translational fidelity of Cys(Acm) in FIT translation system, incorporation of Cys(Acm) into model peptide, TPL1 (Figure 3.3a), was assessed in a Met-deficient FIT reaction. Commercially purchased Boc-Cys(Acm)-OH was first activated with the C-terminal dinitrobenzyl (DNB) ester group, followed by removal of the Boc group using concentrated hydrochloric acid. The activated Cys(Acm)-DNB was charged on tRNA$^{EnAsn}_{CAU}$ via an ester bond using flexizyme (dFx) and incorporated into TPL1 during *in vitro* translation. An activated cyanomethyl biotinylated-Phe ester (^{Bio}Phe-CME) was charged on an initiator tRNA, tRNA^{fMet} and translated onto the N-terminal position of the peptide, to facilitate further pull-down test. The TPL1 peptide was expressed in the presence of 19 amino acids mixture, with exception of Met, and purified with

peptide cleanup SPE C-tip (Nikkyo Technos), to remove the large protein factors in the translation mixture. The translated TPL1 was tested with a MALDI-TOF mass spectrometer. Correct mass was observed with one [Bio]Phe and two Cys(Acm) residues incorporated into the peptide, however, a small peak with - 43 Da was shown next to the expected mass, indicating the loss of an acetyl group (Figure 3.3b). This might be caused either by the ionization effect of the mass spectroscopy, or due to the basic or acidic treatment during the activation reaction. Although amide bonds are extremely stable and their half-life in neutral pH under room temperature tends to be 350 years,[151] this loss of acetyl from the acetamido group might happen under strong base or acid condition.

According to literature, treatment with 10 eq. of $PdCl_2$ for 5 min, followed by 25 eq. of chelator sodium diethyldithiocarbamate (DTC) for 25 min can lead to the fully removal of Acm and cysteine oxidation to form the disulfide bond.[142,143] However, these reactions are performed on large scale (μmol scale) for synthetic peptides. To apply the deprotection reaction on the *in vitro* translated peptides, the reaction conditions needed to be optimized.

The peptide concentration in the FIT translation reaction is assumed to be 1.2 μM, however, at high concentration, Pd(2+) might coordinate with the oligonucleotides,[152] hindering the sequential PCR reactions. Therefore, low metal concentration is preferred. To FIT translated model peptide TPL1, $PdCl_2$ was added to the translation cocktail at a series of final concentrations from 2.5 to 20 mM and incubated for 2 h at 37 °C, followed by addition of 50 mM of DTC as the reaction quencher. After incubation at 37 °C for 30 min, yellow precipitate was formed and removed by centrifugation. The resulting solution was purified by C-tip and tested with mass spectroscopy (Figure 3.3b). Two intermediate peaks of - 43 Da and - 86 Da were observed, indicating the successive cleavage of two acetyl groups. The resulted mass spectra suggested that two Cys(Acm) residues were completely deprotected with treatment of 20 mM of $PdCl_2$ for 2h and a disulfide bond was formed under the deprotection conditions. Further optimization of the reaction time and chelator DTC concentration was not conducted in this Chapter.
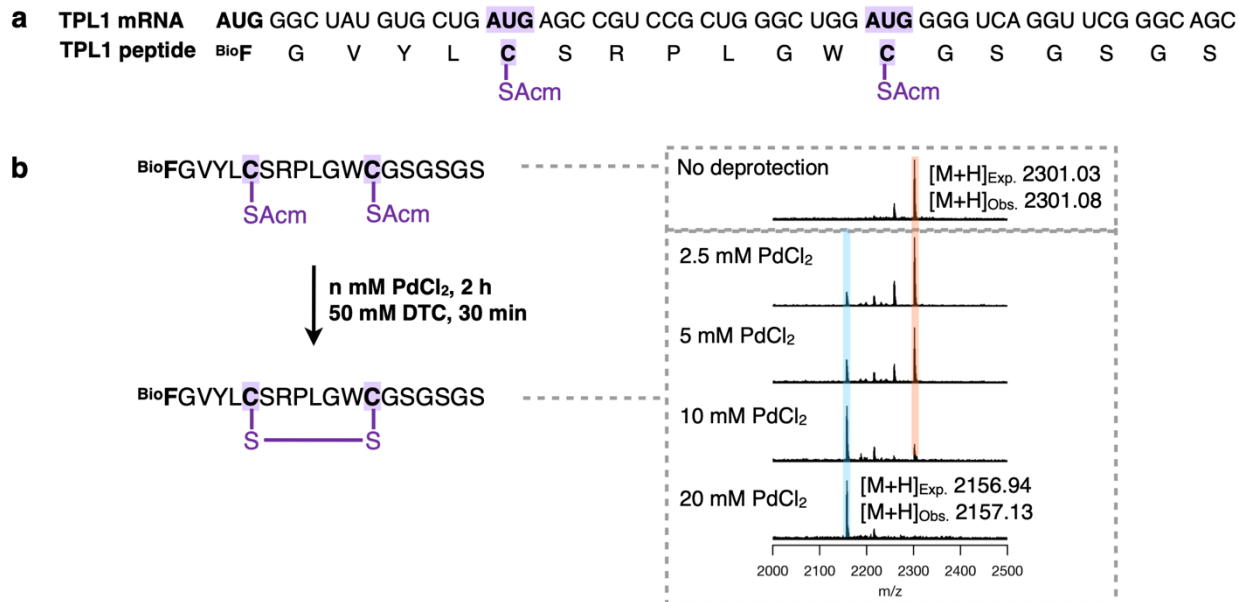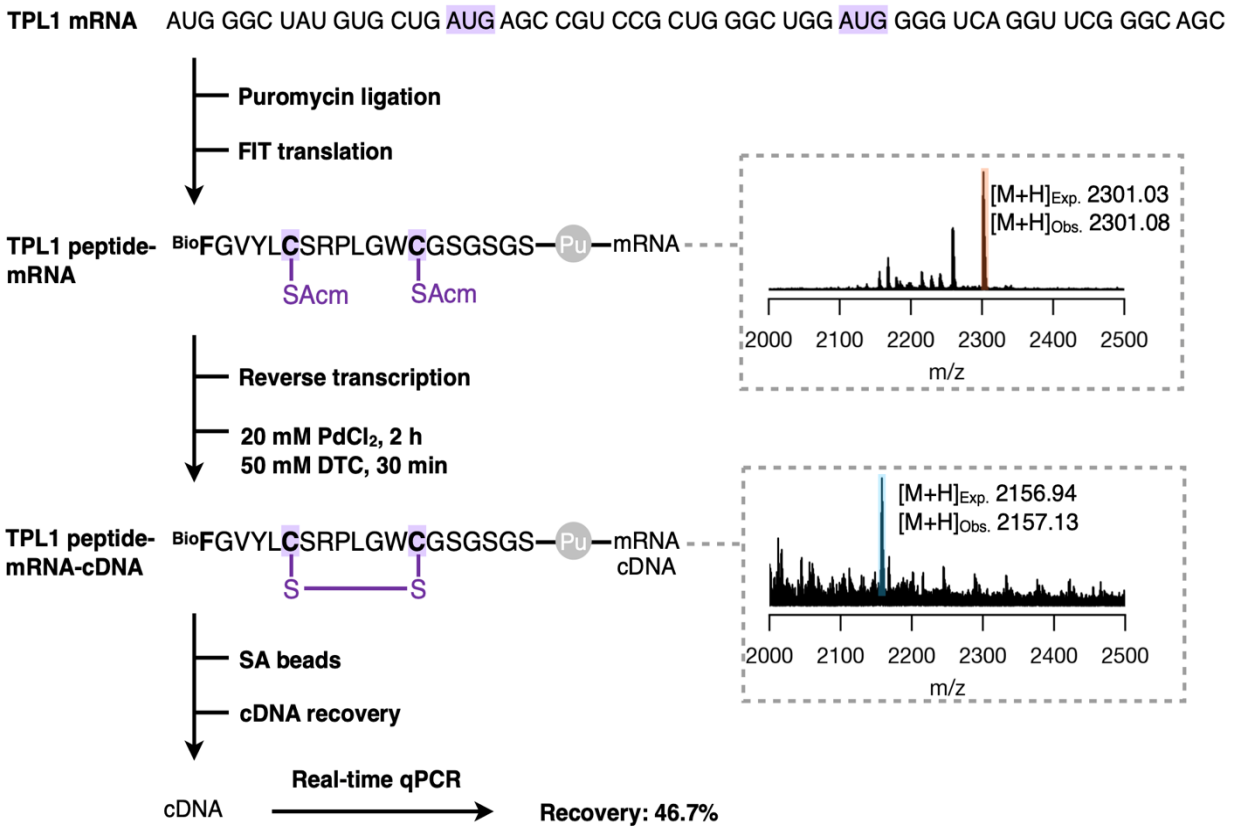
Figure 3.3: Optimization of Cys(Acm) deprotection conditions. (a) mRNA and peptide sequence of template peptide TPL1; (b) Deprotection of TPL1 under various $PdCl_2$ concentrations.

3.2.2. Compatibility of Cys(Acm) in RaPID selection system

After optimizing the deprotection conditions, the FIT translated TPL1 peptide was assessed in the RaPID selection procedure, to verify the feasibility of Cys(Acm) in RaPID selection system (Figure 3.4). Briefly, the mRNA of TPL1 was charged with 3' puromycin linker and *in vitro* translated as described in 3.2.1. Reverse transcription was performed after the translation reaction, followed by Acm deprotection with 20 mM of $PdCl_2$ for 2h and 50 mM of DTC for 30 min. The suspension was removed by centrifuging, followed by buffer exchange and desalting using size exclusion chromatography (Sephadex G-25 Fine resin, Cytiva, suspended in PBST buffer). The resulted solution was mixed with streptavidin embedded magnetic beads (Thermo Fisher) for 30 min at 4 °C, and the cDNAs of peptides that binds to the beads were retrieved by elution. Quantitative PCR (qPCR) reaction revealed the amount of peptides that recovered from the affinity-based beads pull-down reaction and the recovery rate was calculated based on the total amount of peptides before the pull down reaction. Because the TPL1 model peptide was translated with a biotinylated Phe at the N-terminus, theoretically, 100 % of the translated peptides should be recovered due to the strong biotin-streptavidin interaction. The actual recovery rate observed in the experiment was 46.7 %, which is consistent to the experimental value observed by other researchers in our laboratory.

Although the deprotection strategy makes no disturbance on the recovery rate, equally, the effect of RaPID procedures on the Acm deprotection needs to be assessed. Similar to the recovery rate determining experiment, mRNA of TPL1 was translated in a FIT translation cocktail, with addition of release factor 1 (RF1), dropping the translated peptide from mRNA, reverse transcribed and deprotected using the determined conditions. MALDI mass spectra were measured directly after the translation reaction and deprotection treatment, respectively (Figure 3.4), suggesting the correct mass with removal of two Acm groups and formation of the disulfide bond. However, the intensity of the mass spectra was relatively low compared with the translation check experiment in Section 3.2.1. In summary, these results suggested that Cys(Acm) can be introduced in FIT translated peptides and the deprotection conditions are compatible in RaPID selection system.


Figure 3.4: Verification of Cys(Acm) compatibility in RaPID selection process. (left) Experimental procedure of one round of RaPID selection; (right) Mass spectra of the translation product after reverse transcription and deprotection proved that the reverse transcription conditions had no deleterious effect on Acm deprotection.

3.2.3. Construction of bicyclic peptide libraries

After confirming the compatibility of Cys(Acm) in the RaPID system, two bicyclic peptide libraries containing one thioether and one disulfide were constructed and screened against mouse and human ASGR (m/hASGR) protein. Since bicyclic peptides containing two pairs of disulfide bonds are the simplest DRP models, which is found in many natural DRP families, including conotoxins, Kunitz-type inhibitors and scorpion toxin, I first designed the library based on this simplest architecture. As introduced previously, the actual three-dimensional structures of bicyclic DRPs are complicated due to the various loop lengths and amino acid sequences, I simplified the DRP structures into three main classes based on their disulfide connectivity (Figure 3.5a). Since the druggable protein targets commonly possess binding cavities or pockets, the globular and ribbon bicyclic peptides were chosen as scaffolds, for their more compact structures and easier interactions with the target protein.[153]

In detail, two bicyclic libraries with the globular and ribbon disulfide connectivity were designed (Figure 3.5b), such that a N-terminal chloroacetylated Tyr residue (ClAc-^{L}Tyr) was incorporated for spontaneous formation of a thioether bond with a downstream free Cys residue. Two Cys(Acm) residues were incorporated upstream and downstream (globular, G-library) or both upstream (ribbon, R-library) to the free Cys, using the elongator CAU codon as tested in the previous FIT translation test. A (Gly-Ser)$_3$ hexapeptide was embedded in the C-terminus as a spacer for linkage to the cognate mRNA template via puromycin. Either the globular or ribbon library contained three different lengths, which is a mixture of peptides with 3, 4 or 5 random amino acids between the ClAc-^{L}Tyr and each Cys. Although natural DRP utilized disulfides to bridge side-chain cyclization, the substitution with thioether bond improves the robustness of cyclic linkage and doesn't change the overall bond connectivity in the whole peptide.

To assess the feasibility of such a library design, I first translated two model peptides TPL2/3, corresponding to the G-library and R-library, respectively (Figure 3.5c). The MALDI mass spectroscopic analysis demonstrated that one ClAc-^{L}Tyr and two Cys(Acm) were correctly incorporated in the *in vitro* translated model peptides, and could be fully deprotected using the determined conditions, forming the thioether and disulfide bonds as designed.
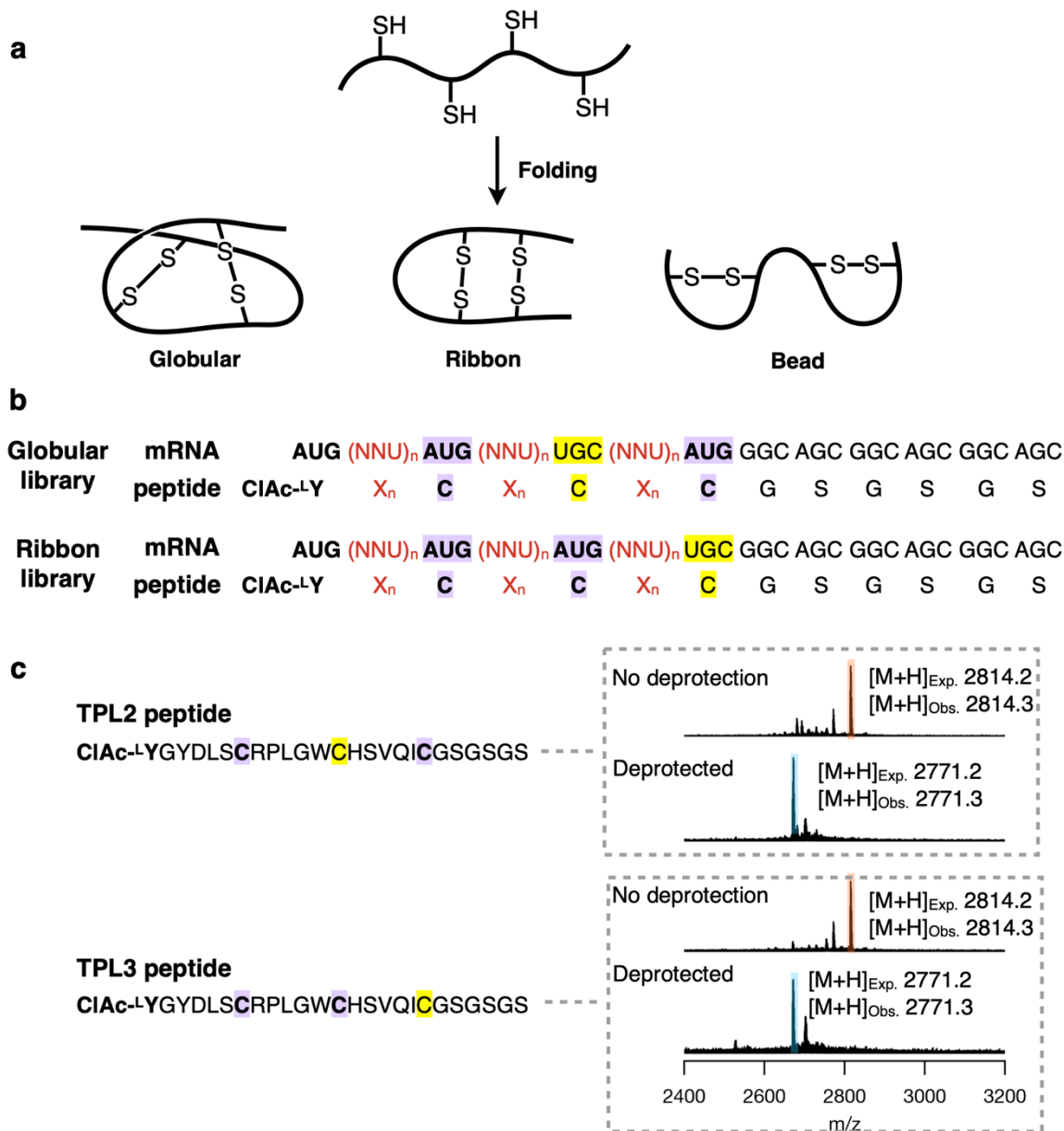
Figure 3.5: (a) Possible folding conformations of DRPs containing two disulfide bonds; (b) mRNA and peptide sequence of the globular and ribbon bicyclic library. The residues highlighted in purple indicates the AUG codon reprogrammed with Cys(Acm); (c) Verification of the library design.

3.2.4. RaPID selection of bicyclic libraries against m/hASGR

RaPID selection process was performed against protein targets m/hASGR using the G/R-libraries, making 4 sets of selections in total. The experimental scheme was similar to the standard RaPID protocol, adding the deprotection steps before the affinity-based beads recovery (Figure

61

3.6). In brief, the DNA encoding this library was assembled from degenerate oligonucleotides, transcribed into mRNA, ligated to a puromycin-linked oligonucleotide and FIT translated to produce a library of mRNA-peptide fusion molecules. Based on the maximum amount of ribosome in the *in vitro* translation cocktail, the theoretical diversity of the mRNA-peptide fusion library was more than 10^{14} variants. However, the diversity was likely decreased during mRNA display processes (*e.g.* mRNA library gel purification and puromycin ligation), therefore, it could be conservatively estimated that the mRNA-peptide fusion library comprised in excess of 10^{12} variants.[154] The libraries were reverse transcribed, followed by Acm deprotection using 20 mM of PdCl$_2$ for 2h and 50 mM of DTC for 30 min. After desalting, the library was screened sequentially for affinity to mock beads, a mock protein target adiponectin (APN, as negative control to the protein target) and biotinylated m/h-ASGR immobilized on streptavidin-coated magnetic beads. The cDNAs of the binding peptides were retrieved and the total recovery rate was assessed by qPCR and then amplified with PCR reaction. Following seven rounds of selection, next-generation sequencing was performed to identify high frequency clonal sequences in the final cDNA library.

The recovery rate was calculated as the ratio of peptides binding to the target protein divided by the initial peptides in the library. Raising of the recovery rate over selection rounds suggested the enrichment of the target-binding peptides (Figure 3.7). Both G/R-libraries showed considerable recovery rate ascent against mASGR, with a positive/negative ratio of ~400 fold in the seventh round, which indicated that the selected peptides exclusively bind to the target protein. However, for the selection using G/R-libraries against hASGR, the total recovery rate didn't show significant rise, whereas the positive/negative ratio was relatively low compared to selection against its mouse counterpart.

Unfortunately, sequencing results of the last round revealed a severe contamination of the libraries during selection (Figure 3.8). Since the selections using G/R-libraries were performed in parallel with libraries with macrocyclic structure design, the sequences recovered from the G/R-libraries contained a large component of the macrocyclic peptide sequences. This might be caused by the same linker sequence (corresponding to the T7g10M.F46 and CGS3an13.R22 primers) designed in the G/R-libraries and the macrocyclic peptide library. Since the contaminated macrocyclic sequences were relatively shorter than the designed libraries, the PCR amplification yield might be higher, therefore resulting in better amplification of the macrocyclic sequences. On the other hand, the selected macrocyclic sequences might possess stronger binding affinity than the G/R-peptides, therefore overwhelming the designed bicyclic libraries in the selection.

Another factor that might lead to the failure of the selection is the massive loss of input library after $PdCl_2$ treatment. The standard RaPID selection protocol calculates the recovery rate of each selection round using the ratio of the amount of cDNA retrieved from the positive beads (Output) and cDNA after reverse transcription (Input). However, in this selection, the Input sample was taken after the $PdCl_2$ deprotection, therefore, the input loss was underestimated. Table 3.1 showed the difference between calculated and experimental Input cDNA amount. The low input recovery (less than 10%) in the second, third, fourth and seventh rounds indicated that $PdCl_2$ deprotection step was deleterious for the PCR reaction of cDNA, possibly due to Pd(2+) ion coordinating to the nucleotide and impeding the polymerization reaction. Therefore, the recovered cDNA might not be sufficiently amplified during the later recovery PCR, thus leading to sharp decrease of the library diversity and failure of selecting potent binding sequences. In addition, since $PdCl_2$ was dissolved in DMSO and directly added to the reaction mixture after reverse transcription (2% v/v), the pH changed caused by addition of $PdCl_2$ was neglected. However, addition of $PdCl_2$ DMSO solution in water led to severe pH decrease and therefore might led to hinderance of later PCR reactions. After observing input loss during the selection process, sample input during the method verification experiment in Section 3.2.2 was reevaluated and it turned out that the input recovery was only 2.73%. Therefore, in order to eliminate the input loss, the deprotection conditions need to be further optimized.

Table 3.1: calculated and experimental input cDNA amount in four selections

Globular-mASGR	rnd1	rnd2	rnd3	rnd4	rnd5	rnd6	rnd7
Input$_{Cal.}$	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13
Input$_{Exp.}$	8.E+12	3.E+10	1.E+12	6.E+11	6.E+12	6.E+12	1.E+11
Input recovery (%)	68.05	0.26	9.04	5.02	53.37	51.20	0.97

Globular-hASGR	rnd1	rnd2	rnd3	rnd4	rnd5	rnd6	rnd7
Input$_{Cal.}$	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13
Input$_{Exp.}$	1.E+13	5.E+10	5.E+11	7.E+11	3.E+12	7.E+12	5.E+11
Input recovery (%)	89.33	0.38	4.55	5.63	28.82	55.18	3.81

Ribbon-mASGR	rnd1	rnd2	rnd3	rnd4	rnd5	rnd6	rnd7
Input$_{Cal.}$	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13
Input$_{Exp.}$	8.E+12	5.E+09	1.E+12	5.E+11	7.E+12	5.E+12	7.E+11
Input recovery (%)	64.20	0.04	9.42	3.95	58.63	40.50	5.99

Ribbon-hASGR	rnd1	rnd2	rnd3	rnd4	rnd5	rnd6	rnd7
Input$_{Cal.}$	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13	1.2E+13
Input$_{Exp.}$	7.E+12	1.E+10	1.E+12	1.E+12	6.E+12	3.E+12	7.E+11
Input recovery (%)	59.67	0.11	8.59	9.35	46.52	25.81	5.64

Only the selection of G-library against mASGR displayed one dominating sequence (48 % over the total sequenced library) with the expected globular bicyclic design. The dominating sequences in all the other three selections were the macrocyclic peptides. This severe contamination might be caused by spills of the macrocyclic library solutions on the working bench or pipettes, or the contamination of the common buffer. Several attempts were performed to avoid the contamination, including changing all the reaction buffers, redesign the library with different spacer sequence, etc. However, none of the approaches resulted in successful selection or removal of the contamination. Therefore, I finally gave up on selecting for more bicyclic ASGR-binding peptides, but focused on evaluating the binding affinity of the single globular candidate binding to mASGR.
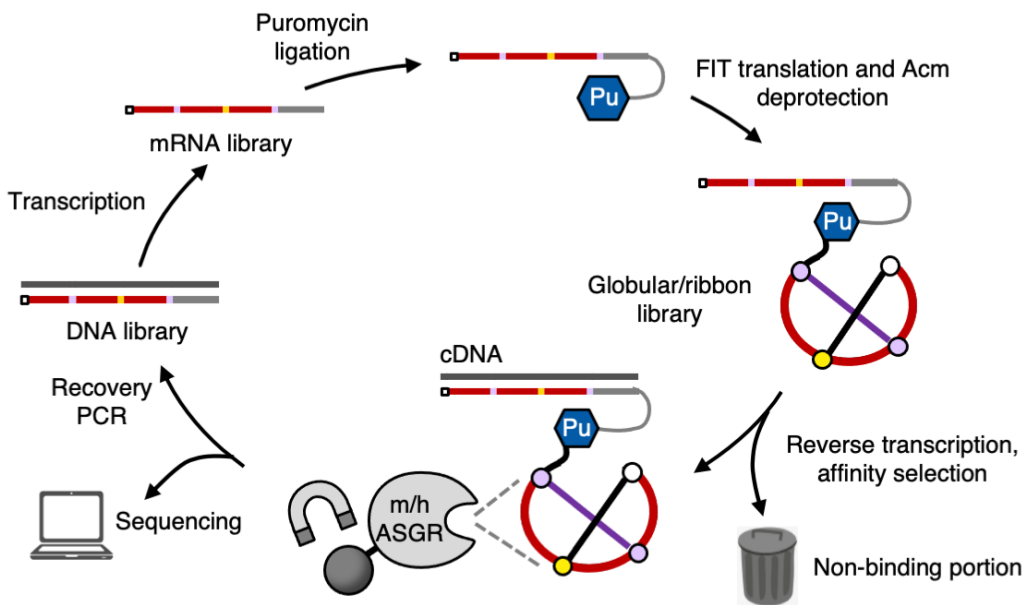
Figure 3.6: Selection scheme of tricyclic libraries against SARS-CoV-2 MPro.
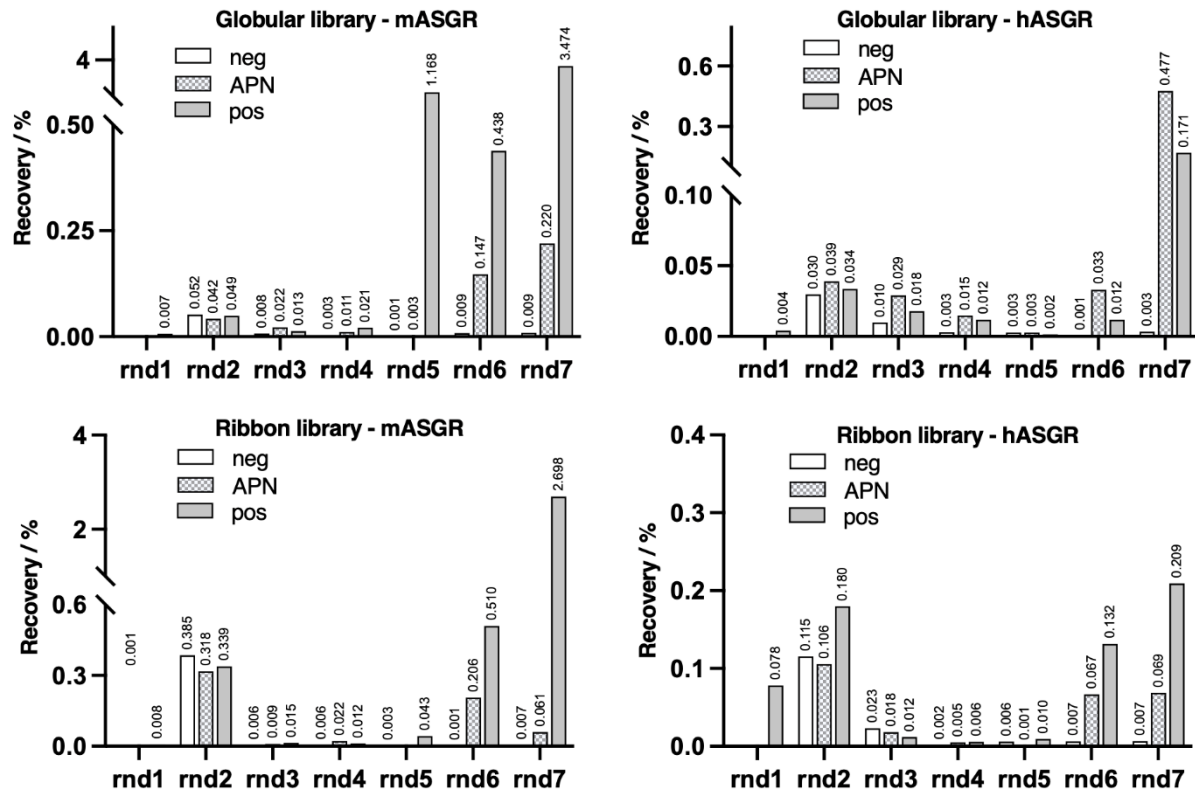
Figure 3.7: Percentage recovery of cDNAs in each library after every round of selection.

Figure 3.8: Sequences of the top 30 most abundant peptides recovered from affinity selection against ASGR from the globular and ribbon libraries.
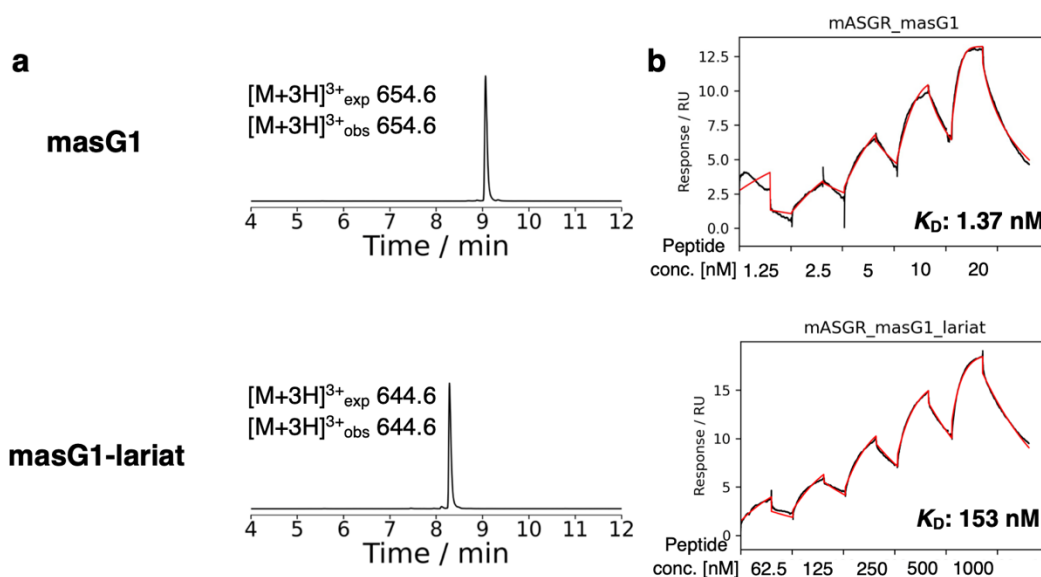
3.2.5. Binding affinity of masG1 and its lariat variant towards mASGR

The selected mASGR-binding peptide with designed globular conformation, masG1, was synthesized using SPPS, with the thioether and disulfide bonds generated similar to the approaches conducted during the RaPID selection. In brief, residues of masG1 were assembled on NovaPEG Rink Amide resin (Novabiochem), with two Cys protected with Acm and one Cys protected with Trt (translated as free Cys during FIT translation). The ClAc group was added to the N-terminal free Tyr using N-(chloroacetoxy)succinimide (ClAc-NHS), followed by deprotection of Cys(Trt) using 95 % TFA and spontaneous ring closure via formation of the thioether bond. The peptide was lyophilized and dissolved in 6M guanidine hydrochloride solution, with addition of 10 eq. of PdCl$_2$ to remove the Acm group (37 °C, 2 h). Extra Pd(2+) ion was removed using 100 eq. of DTC precipitation, and the resulting peptide was purified with HPLC spectroscopy. Analytical HPLC of the synthetic masG1 showed the pure compound with correct mass (Figure 3.9a), indicating successful synthesis of the bicyclic globular masG1. To assess the effect of the bicyclic structure on the target-binding property, a lariat variant of masG1 was synthesized as well, by replacing two Cys(Acm) with Ser residues. I hypothesized that the removal of the disulfide bond could lead to the complete change of the overall structure from a compact bicyclic conformation to a macrocyclic peptide with a long tail (lariat). Target binding affinity of both the parent masG1 and the lariat variant masG1-lariat were determined using SPR and similar to that was discussed in Chapter 2, peptides in a series of concentration were flowed through biotinylated target protein and the resonance response was fitted using a single-cycle kinetics (Figure 3.9b).

The SPR measurement suggested that masG1 was a single-digit nanomolar binder of mASGR (K_D = 1.37 nM), whereas it solely bound to mASGR, with no binding affinity against its human counterpart (no response detected for hASGR). The comparison of masG1 (K_D = 1.37 nM) and masG1-lariat (K_D = 153 nM) suggested that the bicyclic structure could be critical for the potent binding affinity, by forming a more compact structure to interact with the binding convex on ASGR. Further experiments are underway to determine the cell-internalization ability of masG1. However, since our final goal of this research is to create a peptide ligand that can be internalized into the human liver cell for the elimination of malignant proteins, the exclusive reactivity of masG1 against mASGR is not enough for future therapeutical perspectives. Therefore, to raise the cross-reactivity for the future study, point mutation engineering or deep mutagenesis scanning[131] could be performed on the masG1 molecule.

Figure 3.9: (a) LC-MS trace of synthetic bicyclic peptide masG1 and its lariat variant masG1-lariat; (b) Biacore sensorgram of masG1 and masG1-lariat binding with mASGR. The K_D values are indicated on the right bottom.

3.3. Conclusions

In this work, I successfully incorporated protected cysteines into peptide libraries using FIT translation and constructed bicyclic peptide libraries with globular and ribbon conformations. RaPID selection using the bicyclic libraries was conducted against m/hASGR, revealing one potent mASGR binding peptide, masG1, with the expected globular conformation. Comparison of masG1 ($K_D = 1.37$ nM) and its lariat variant masG1-lariat ($K_D = 153$ nM) revealed that the bicyclic structure is critical for the potent binding affinity. This study was the first trial on applying protected cysteines in RaPID selection, proving the feasibility for controlling peptide conformation *in situ* during RaPID process, highlighting the immense potential of the approach as a method for rapidly evolving potent DRP-based pharmaceuticals.

3.4. Materials and methods
Chemicals and reagents

All chemical reagents were purchased from Watanabe Chemical Industry, Nacalai Tasque, Tokyo Chemical Industry, Sigma-Aldrich Japan or Wako. Unless otherwise noted, all the chemical reagents obtained from commercial sources were used without any purification. H_2O used for buffer preparations was from a Sartorius filtration system (18.2Ω).

List of DNA oligomers

All PCR reactions were conducted using Taq DNA polymerase unless otherwise noted. All oligonucleotides used for PCR reactions in this Chapter are listed in this table.

For flexizyme preparation	
Fx5'.F36	GTAATACGACTCACTATAGGATCGAAAGATTTCCGC
T7ex5.F22	GGCGTAATACGACTCACTATAG
eFx.R45	ACCTAACGCTAATCCCCTTTCGGGGCCGCGGAAATCTTTCGATCC
eFx.R18	ACCTAACGCTAATCCCCT
dFx.R46	ACCTAACGCCATGTACCCTTTCGGGGATGCGGAAATCTTTCGATCC
dFx.R19	ACCTAACGCCATGTACCCT
For tRNA preparation	
Ini-3'.R20-Me	TGmGTTGCGGGGGCCGGATTT
Ini-3'.R38	TGGTTGCGGGGGCCGGATTTGAACCGACGATCTTCGGG
Ini1-1G-5'.F49	GTAATACGACTCACTATAGGCGGGGTGGAGCAGCCTGGTAGCTCGTCGG
Ini cat.R44	GAACCGACGATCTTCGGGTTATGAGCCCGACGAGCTACCAGGCT
EnAsn-3'.R20-Me	TGmGCGGCTCTGACTGGACTC
EnAsn-3'.R38	TGGCGGCTCTGACTGGACTCGAACCAGTGACATACGGA
EnAsn-5'.F49	GTAATACGACTCACTATAGGCTCTGTAGTTCAGTCGGTAGAACGGCGGA
EnAsn CAT.R43	GAACCAGTGACATACGGATTNNNAGTCCGCCGTTCTACCGACT
For peptide template and library synthesis	
T7g10M.F46	TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATA
CGS3an13.R22	TTTCCGCCCCCCGTCCTAGCTG
TLP1.F91	CTTTAAGAAGGAGATATACATATGGGCTATGTGCTGATGAGCCGTCCGCTGGGCTGGATGGGG TCAGGTTCGGGCAGCTAGGACGGGGGGGC
TLP2.R99	CTAGCTGCCGCTGCCGCTGCCCATGATCTGCACGCTATGGCACCAGCCCAGCGGGCGCATGCT CAGATCATAGCCCATATGTATATCTCCTTCTTAAAG
TLP3.R99	CTAGCTGCCGCTGCCGCTGCCGCAGATCTGCACGCTATGCATCCAGCCCAGCGGGCGCATGCT CAGATCATAGCCCATATGTATATCTCCTTCTTAAAG
Bicyc-G-NNK9.R81	CTAGCTGCCGCTGCCGCTGCCCATMNNMNNMNNGCAMNNMNNMNNCATMNNMNNMNNCATATG TATATCTCCTTCTTAAAG

Bicyc-G-NNK12.R90	CTAGCTGCCGCTGCCGCTGCCCATMNNMNNMNNMNNGCAMNNMNNMNNMNNCATMNNMNNMNN MNNCATATGTATATCTCCTTCTTAAAG
Bicyc-G-NNK15.R99	CTAGCTGCCGCTGCCGCTGCCCATMNNMNNMNNMNNMNNGCAMNNMNNMNNMNNMNNCATMNN MNNMNNMNNMNNCATATGTATATCTCCTTCTTAAAG
Bicyc-R-NNK9.R81	CTAGCTGCCGCTGCCGCTGCCGCAMNNMNNMNNCATMNNMNNMNNCATMNNMNNMNNCATATG TATATCTCCTTCTTAAAG
Bicyc-R-NNK12.R90	CTAGCTGCCGCTGCCGCTGCCGCAMNNMNNMNNMNNCATMNNMNNMNNMNNCATMNNMNNMNN MNNCATATGTATATCTCCTTCTTAAAG
Bicyc-R-NNK15.R99	CTAGCTGCCGCTGCCGCTGCCGCAMNNMNNMNNMNNMNNCATMNNMNNMNNMNNMNNCATMNN MNNMNNMNNMNNCATATGTATATCTCCTTCTTAAAG

For peptide template and library synthesis

Puromycin linker	CTCCCGCCCCCCGTCC-(PEG18)5-CC-Puromycin

Preparation of flexizymes (eFx/dFx)

Template DNAs that have a T7 promoter (TAATACGACTCACTATAG) prior to transcripts were prepared by extension of forward and reverse primer pairs (eFx: Fx5'.F36 and eFx.R45; dFx: Fx5'.F36 and dFx.R46), followed by addition of forward and reverse primers (eFx: T7ex5.F22 and eFx.R18; dFx: T7ex5.F22 and dFx.R19) with PCR reactions. PCR reactions were conducted in 1x PCR reaction mixture containing 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% (v/v) Triton X-100, 2.5 mM $MgCl_2$, 250 μM each dNTPs, 0.5 μM forward and reverse primers, and 1.5% (v/v) 1× Taq DNA polymerase at pH 9.0. Sequential PCRs were performed with several cycles (depending on the reaction yield checked by agarose gel analysis) of PCR reactions with 95 °C for 40s, 50 °C for 40s, 72 °C for 40s. The PCR products were extracted by phenol/chloroform, precipitated by ethanol, and used for *in vitro* transcription at 37 °C for 16 h in a 100-2000 uL reaction scale. The *in vitro* transcription reaction mixture contained 40 mM Tris-HCl, 1 mM spermidine, 0.01% (v/v) Triton X-100, 10 mM DTT, 30 mM $MgCl_2$, 5 mM NTPs, 30 mM KOH, 10% (v/v) template DNA solution, 0.12 μM T7 RNA polymerase at pH 8.0. The resulting RNA transcripts were treated with RQ1 DNase (Promega) for 1 hour at 37∘C. The RNA transcripts were precipitated by adding 10% (v/v) 3M NaCl and 80% v/v of isopropanol followed by centrifuge. After wash with 70% (v/v) EtOH, the RNA pellet was dissolved in ultrapure H_2O with 10% (v/v) volume of transcription reaction and equivalent volume of 2x RNA loading buffer (8M urea, 2 mM $Na_2EDTA.2H_2O$, 2 mM Tris-HCl, pH 7.5) was added. After heating at 95 °C for 2 min, RNA was purified by 8% (v/v) polyacrylamide gel containing 6 M urea. The RNA band was visualized by UV light, cut out, and extracted in the 0.3 M NaCl solution for at least 5 hours. After removal of gel by centrifuge and filtration, twice volume of EtOH was added and RNA was precipitated by centrifuge at 13000 rpm

for 15 min. After wash with 70% (v/v) EtOH and drying at room temperature, the product was dissolved in H$_2$O and the concentration was assessed and diluted to 250 μM.

Preparation of tRNAs (tRNA^{fMet} and tRNA$^{EnAsn}_{CAU}$)

The preparation of tRNAs was similar to flexizymes, which was described above. For tRNAs, the DNA templates were prepared by extension of forward and reverse primer pairs (tRNA^{fMet}: Ini1-1G-5'.F49 and Ini cat.R44; tRNA$^{EnAsn}_{CAU}$: EnAsn-5'.F49 and EnAsn CAU.R43), followed by the first PCR reaction adding the T7 promoter (tRNA^{fMet}: T7ex5.F22 and Ini-3'.R38; tRNA$^{EnAsn}_{CAU}$: T7ex5.F22 and EnAsn-3'.R38), and the second PCR reaction adding forward and reverse primers (tRNA^{fMet}: T7ex5.F22 and Ini-3'.R20-Me; tRNA$^{EnAsn}_{CAU}$: T7ex5.F22 and EnAsn-3'.R20-Me). The transcription reaction for tRNAs contained 40 mM Tris-HCl, 1 mM spermidine, 0.01% (v/v) Triton X-100, 10 mM DTT, 22.5 mM MgCl$_2$, 3.75 mM NTPs, 5 mM GMP 22.5 mM KOH, 10% (v/v) template DNA solution, 0.12 μM T7 RNA polymerase with pH 8.0. All the other reaction conditions including PCR reaction cycles, extraction, precipitation, and polyacrylamide gel purifications were identical to the preparation of flexizymes. The product was dissolved in H$_2$O and the concentration was assessed and diluted to 250 μM.

Synthesis of peptide template and libraries

Synthesis of mRNA templates for TPL1–3 and the bicyclic libraries was similar to preparation of flexizymes, as described above. Briefly, the DNA templates were prepared by a two-step PCR reaction including the extension of forward primer T7g10M.F46 and the corresponding reverse primers, followed by a PCR amplification adding the forward and reverse primers T7g10M.F46 and CGS3an13.R22. Sequential PCRs were performed using Taq polymerase (95 °C for 40s, 58 °C for 40s, 72 °C for 40s) for minimal number of cycles (4 cycles for extension and 5 cycles for PCR amplification), and the resulting DNA was used as a template for T7 RNA polymerase mediated transcription at 37°C for 16 h prior to polyacrylamide gel purification. *In vitro* transcription was performed in 200 μL and 6500 μL scale for the model templates, TPL1–3, and the libraries, respectively. All the other reaction conditions including extraction, precipitation, and polyacrylamide gel purifications were identical to the preparation of flexizymes. The mRNA template was dissolved in H$_2$O and the concentration was assessed and diluted to 10 μM.

Synthesis of aminoacylated tRNAs

Aminoacylation of tRNAs using activated amino acids was performed as follows: 25 μM each flexizyme (eFx for ^{Bio}Phe-CME and ClAc-^{L}Tyr-CME; dFx for Cys(Acm)-DBE) and the according tRNA in 50 mM HEPES-KOH pH7.5 were heated to 95 °C for 2 min and cooled slowly to 25 °C. For reactions using eFx or dFx, 600 mM or 100 mM of MgCl$_2$ (final concentration) was added and the reaction was incubated at 25 °C for 5 min. After cooled on ice, the activated amino acid was added to a final concentration of 5 mM, and the reaction was allowed to proceed on ice for the optimal time determined for each amino acid (2 h for ^{Bio}Phe-tRNA^{fMet}, ClAc-^{L}Tyr-tRNA^{fMet} and Cys(Acm)-tRNA$^{EnAsn}_{CAU}$). Aminoacylated tRNA was recovered by twice volume of ethanol precipitation and washed with 70% EtOH and stored dry at -80 °C until use.

In vitro translation (FIT translation)

In vitro translation was performed using a purified recombinant expression system as described in Chapter 2. For testing the translation and deprotection of TLP1–3, individual mRNA template was added at a concentration of 2 μM, all three aminoacylated tRNAs (^{Bio}Phe-tRNA^{fMet}, ClAc-^{L}Tyr-tRNA^{fMet} and Cys(Acm)-tRNA$^{EnAsn}_{CAU}$) was included at a concentration of 25 μM, respectively. Methionine and 10-formyl-5,6,7,8-tetrahydrofolic acid were excluded from the translation. Translation products were purified by SPE C-tip (Nikkyo Technos) and eluted with 1.2 μL of half-saturated α-cyano-4-hydroxycinnamic acid matrix substance dissolved in 0.5% (v/v) acetic acid, 80 % (v/v) acetonitrile, followed by analyzation using MALDI-TOF mass spectroscopy (Bruker UltraFlex mass spectrometer).

RaPID selection of bicyclic libraries agasint m/hASGR

mRNA templates of the globular/ribbon bicyclic libraries were covalently linked to a puromycin linker (as shown in the oligonucleotides list) using home-made T4 RNA ligase, before translated in FIT translation reactions containing without Met and its 10-formyl-5,6,7,8-tetrahydrofolic acid. Each reaction contained 1.2 μM mRNA-puro, 25 μM ClAc-^{L}Tyr-tRNA^{fMet} and 25 μM Cys(Acm)-tRNA$^{EnAsn}_{CAU}$. For the first round of selection, translation was performed at 150 uL scale, to give an initial theoretical library diversity of approximately 10^{14} molecules at the mRNA level. The mRNA-peptide conjugate libraries were isolated from the ribosomes by EDTA and reverse transcribed using the CGS3an13.R22 primer and RNase H- reverse transcriptase (Promega). The reaction mixtures were then desalted using Sephadex G-25 Fine resin (Cytiva) suspended in PBST buffer, before deprotection with 0.4× translation volume of 100 mM PdCl$_2$ and incubated at 37°C for 2 h. After quenching the reaction with 0.4× translation volume 500 mM DTC

and incubating at 37 °C for 30 min, the mixture was centrifuged to remove the precipitate. The supernatant mixture was desalted and panned against 200 nM m/hASGR immobilized on Dynabeads M280 Streptavidin (Invitrogen) at 4 °C for 30 min. The beads were washed three times with cold PBST (10 mM phosphate, 130 mM NaCl, 0.05% (v/v) Tween-20, pH 7.4) and cDNA recovered by PCR using the primers T7g10m.F46 and CGS3an13.R22, followed by transcription using home-made T7 RNA polymerase to generate the mRNA for the subsequent round. For the second and later rounds of selection, FIT translation was performed at 10 uL scale, and libraries were subjected to three rounds of counter selection against mock beads, and one round of negative selection against trimerized APN protein, before being panned against m/hASGR. Assessment of recovery following each round was performed by quantitative real time PCR (qPCR) using Sybr Green I dye (Lonza Japan Ltd) on a light cycler nano thermal cycler (Roche). cDNA recovered from each round was sequenced using a MiSeq high-throughput sequencer (Illumina) and analyzed using CLC workbench software (Qiagen).

Solid phase peptide synthesis

Both the masG1 and masG1-lariat peptides were synthesized at 50 μmol scale using standard Fmoc solid phase peptide synthesis methodology. In brief, each sequence was assembled on NovaPEG Rink Amide Resin (Novabiochem, 0.44 mmol/g) using an automated peptide synthesizer (Syro I Parallel Peptide Synthesizer, Biotage). Fmoc-protected amino acids, Fmoc-Ala-OH, Fmoc-Asp(tBu)-OH, Fmoc-Glu(tBu)-OH, Fmoc-Phe-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Asn(Trt)-OH, Fmoc-Pro-OH, Fmoc-Gln(Trt)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Val-OH, Fmoc-Trp(Boc)-OH, Fmoc-Tyr(Boc)-OH, Fmoc-Cys(tBu)-OH, and Fmoc-Cys(Acm)-OH were purchased from Watanabe Chemical Ltd. Couplings were performed twice with 6 equivalent of Fmoc-protected amino acids (in DMF), 6 equivalent of *O*-(Benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HBTU, in DMF), 6 equivalent of 1-hydroxybenzytriazole (HOBt, in DMF), and 12 equivalent of *N,N*-diisopropylethylamine (DIPEA, in NMP) for 40 min. Removal of the Fmoc group was achieved using 40% (v/v) piperidine in DMF at 30 °C for 12 min. To each side-chain protected peptide, 2 ml of 0.2 M chloroacetyl *N*-hydroxysuccinimide ester (ClAc-NHS) in DMF was added to the resin and incubated for 1 h at room temperature. After washing the resin with DMF for five times, the peptides were deprotected and cleaved from the resin using 92.5% (v/v) TFA, 2.5% (v/v) water, 2.5% (v/v) triisopropylsilane (TIS) and 2.5% (v/v) 3,6-dioxa-1,8-octanedithiol (DODT) for 3 h, followed by precipitation using diethyl ether and dissolved in 5 ml

of 50% (v/v) DMSO/0.1% TFA (v/v) in water. The pH of the solution was adjusted by triethylamine (TEA) to pH 8 to induce spontaneous macrocycle enclosure between the N-terminal chloroacetamide group and the sulfhydryl group of downstream free Cys. The cyclized peptides were purified on RP-HPLC (Shimadzu) using a water:acetonitrile gradient (with 0.1% TFA). After lyophilization, masG1 was dissolved 6 M guanidine hydrochloride (guanidine-HCl) to a final concentration of 0.5 mM, and treated with 10 equivalent of 100 mM $PdCl_2$ in 6M guanidine-HCl solution at 37 °C for 15 min, before quenched with 100 equivalent of 400 mM sodium diethyldithiocarbamate (DTC, in aqueous solution) at 37 °C for 30 min. The deprotected peptides were purified on RP-HPLC and lyophilized, followed by dissolving in PBS buffer with 10% (v/v) isopropanol. Oxidized and reduced glutathione were added at a final concentration of 2 mM and 0.8 mM, respectively, and the mixture was incubated at room temperature for 2 h, before purified on RP-HPLC. The purity and integrity of the synthetic peptides were verified using LC-MS (Xevo TQ-S, Waters).

Surface plasmon resonance for binding affinity determination

Binding kinetics of each peptide towards m/hASGR were determined using a Biacore T200 machine (Cytiva). Running buffer was HBS-EP+ (10 mM HEPES, 150 mM NaCl, 3 mM EDTA and 0.05% (v/v) surfactant P20, pH 7.4). 500 nM biotinylated m/hASGR was immobilized on a Sensor Chip CAP (Cytiva) using Biotin CAPture Reagent (Cytiva) following the manufacturer's instructions. A series of concentrations of each peptide were injected as analyte and the binding kinetics were modeled using a 1:1 binding model.

Chapter 4.

Construction of tricyclic peptide libraries using stepwise cysteine deprotection and RaPID selection against SARS-CoV-2 main protease

4.1. Introduction

4.1.1. Orthogonal Cys protecting groups

As mentioned in Chapter 3, there are two main approaches for chemical synthesis of multiple disulfide bonds during SPPS: folding during air oxidation, and stepwise removal of protected Cys residues. In the latter method, cysteines protected with different thiol PGs are coupled into the peptide chain during SPPS, followed by stepwise deprotection treatments, releasing two free thiols at one time to form the desired disulfide bond. Due to the requirement of stepwise releasing a certain PG after each deprotection treatment, the PG-deprotection pair needs to be orthogonal. The standard Fmoc-based SPPS protocols generally utilize acid-labile PGs, such as trityl (Trt), *tert*-butyl (tBu), or *tert*-butoxycarbonyl (Boc), for protection of amino acids with reactive side chains. Therefore, Cys thiol groups protected with acid-labile PGs can be removed together with other side-chain protection and cleavage from the resin.[155] Photo-labile protecting groups (PPGs) such as *o*-Nitrobenzyloxycarbonyl (*o*NZ), 4,5-Dimethoxy-2-nitrobenzyl (DMNB), or 2-nitrobenzyl (NBzl) are another class of chemoselective PGs, which are recently applied in SPPS to provide more alternatives for orthogonal disulfide formation. Compared with chemical and light-dependent deprotection conditions, enzymatic deprotection is relatively used less frequently in large-scale SPPS, due to the comparatively high cost of the enzyme production. As mentioned previously in Section 3.1.2, thiol PG phenylacetamidomethyl (Phacm) could be enzymatically removed by penicillin G acylase (PGA), which catalyzes hydrolysis of the amide bond of the phenylacetamido group.[144] Thanks to the industrial usage and production of PGA, it is commercially available, making it accessible to be used in SPPS and reconstituted *in vitro* translation reaction.

For the synthesis of DRPs containing more than two pairs of disulfide bonds, conventional SPPS often utilizes Cys PGs that can be chemically deprotected, i.e. acid-labile or metal-labile PGs. For example, cysteines protected with trityl (Trt), acetamidomethyl (Acm), and p-methoxybenzyl (Mob) groups could be released in a stepwise manner using 95% TFA, I_2, and TFMSA/TFA, respectively.

As mentioned in Chapter 3, Brik et al. has reported several novel Cys PGs suitable for DRP synthesis.[142,143,156] One of their recent reports introduced an ultrafast and regioselective one-pot synthesis of DRPs containing three disulfides using the combination of an acid-labile and a photo-labile PGs (Figure 4.1).[156] This work utilized a three-step deprotection and disulfide formation, consisting of step 1, oxidation of two free thiols (less than 10 s); step 2, UV deprotection of NBzl in an oxidative condition; and step 3, Pd(2+) deprotection of Acm and oxidation with disulfiram (DSF). This ultrafast and water compatible synthetic methodology provides hints for the future DRP synthetic study, and other than the Acm group, which was proved compatible with the RaPID selection, photo-labile NBzl could also be a potential PG for *in vitro* selection.
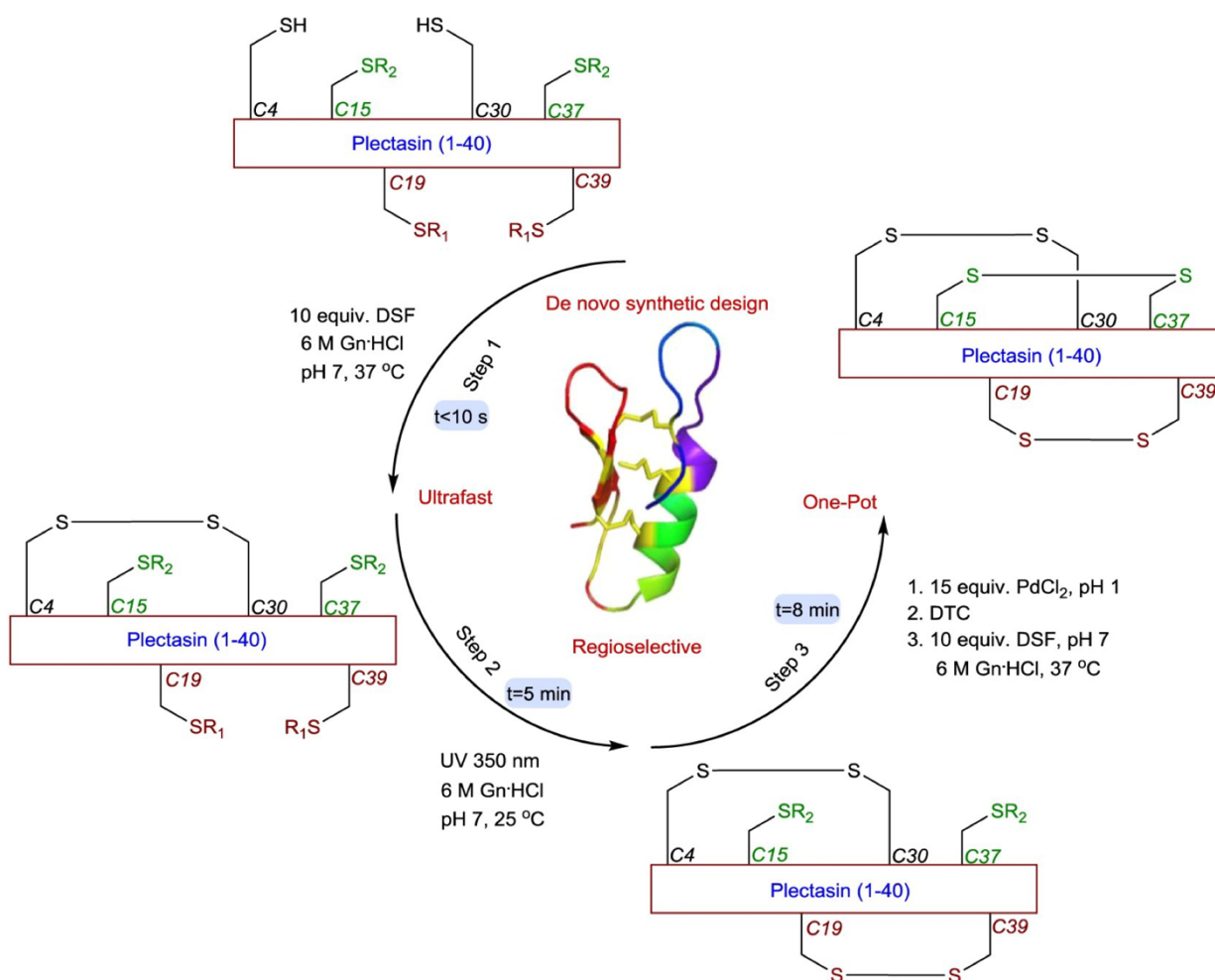
Figure 4.1: An example of one-pot synthesis of spectasin containing three pairs of disulfide bonds. R_1 and R_2 indicate Cys protecting groups NBzl and Acm, respectively. Reproduced with permission from ref. 156 Copyright 2021, Shay Laps et al.[156]

4.1.2. Utilization of Cys(Phacm) in peptide synthesis

Another thiol PG, phenylacetamidomethyl (Phacm), is an analogue of the Acm group, and can be removed by all the Acm deprotection conditions as well. In addition, Phacm could be enzymatically removed by penicillin G acylase (PGA), which catalyzes hydrolysis of the amide bond of the phenylacetamido group (Figure 3.2a,b).[144] The PGA enzyme was originally discovered in the biosynthesis of β-lactam antibiotics and have gained a unique position among enzymes used by pharmaceutical industry, since it's environmentally friendly and economic alternative to chemical processes. Indeed, the annual consumption of PGA has recently been estimated to range from 10 to 30 million tons.[157] PGAs belong to the structural superfamily of N-terminal nucleophile hydrolases and, so far, 21 PGA producers were described among bacteria, fungi, and yeasts.[158,159] The enzyme acts as a transferase or a hydrolase in aqueous reactions and has as its catalytic residue, the N-terminal serine on β-subunit (Serβ1), is activated by a bridging water molecule. It recognizes the side chains of substrates composed of hydrophobic moiety (phenyl-, pyridyl-, thienyl-, tetrazolyl-, CN-, etc.), connected with a short spacer (–CH2–, –OCH2–, –SCH2–, –CHOH–, –CHNH2–) to carboxyl group or its derivative (–COOH, –COOR, –CONHR). The acyl donor on the substrate first associates with the enzyme active site that is highly specific for hydrophobic groups[160] and acylates the Serβ1 to give the acyl-enzyme intermediate, which is further attacked by a water molecule (hydrolysis) or β-lactam nucleus (synthesis). The proposed mechanism of PGA catalyzing hydrolysis of penicillin G is demonstrated in Figure 4.2.[161] Although the reaction mechanisms for hydrolysis and synthesis catalyzed by PGAs are identical, the reaction conditions are different.[161]

Since the phenylacetamido group in Phacm can be recognized by PGA and this mild enzyme reaction is conducted at neutral pH in aqueous solution, it could be suitable for RaPID system. No research has been reported on incorporating Cys(Phacm) into peptide using genetic reprogramming. As mentioned above, a former postdoctoral researcher in our laboratory, Dr. Yichao Huang, synthesized Cys(Acm) and Cys(Phacm) activated with dinitrobenzyl ester and successfully charged them on tRNA using flexizyme, followed by incorporation into peptides taking advantage of the FIT translation. However, he left without optimizing the deprotection conditions. Based on his discovery, I designed this project on controlling the DRP disulfide connectivity using the Cys(Acm) and Cys(Phacm).
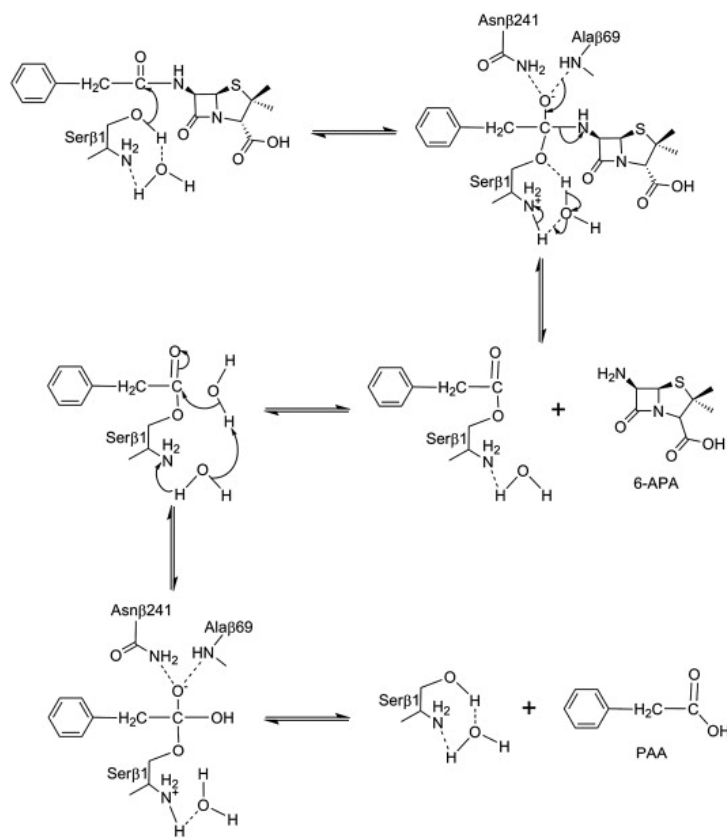
Figure 4.2: Proposed mechanism of hydrolysis of penicillin G mediated by PGA to yield a β-lactam antibiotic 6-APA and phenylacetic acid (PAA). Reproduced with permission from ref. 161 Copyright 2013 Elsevier Inc.[161]

4.1.3. SARS-CoV-2 main protease (M[Pro]) as pharmaceutical target

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the pathogen that causes the disease COVID-19, produces replicase polyproteins pp1a and pp1ab that contain, respectively, 11 or 16 nonstructural proteins (nsp).[162] Nsp5, the main protease (M[pro]), is one of two cysteine proteases necessary for viral replication and assembly, with analogous functional counterparts in earlier SARS-CoV-1 and Middle Eastern Respiratory Syndrome (MERS) coronavirus variants.[163] The M[pro] is a three-domain cysteine protease, which presents a Cys145-His41 catalytic dyad located in the cleft between domains I and II. Several common features are shared among M[pro] substrates, including the presence of a Gln residue at P1 (Schechter-Berger nomenclature).[164] No known human cysteine protease cleaves after Gln, thus offering potential selectivity for this viral target over the human proteome.[165]

Although several COVID-19 vaccines have been released in the market and more than 40% of the worldwide population have been fully vaccinated (the data was collected from Wikipedia on Nov.24, 2021), there is still no FDA proved specific medications for COVID-19 treatment. Several antiviral or autoimmune disease medications, such as tocilizumab, remdesivir, or dexamethasone, have been used in the treatment to reduce the risk or of hospitalization or death. However, most of the current medicines applied on hospitalized patients are immune immunosuppressant that can release the severe immune responses caused by the virus. Compared with the surface spike protein, which is easily mutated via viral spreading, the viral proteases are more reliable and therefore tractable targets for small molecule oral therapies. Moreover, a recent report has demonstrated oral activity of an M^{pro} inhibitor in a transgenic mouse model of SARS-CoV-2 infection.[166] Given the pivotal role of SARS-CoV-2 M^{pro} in viral replication and its potential for mechanistic safety, SARS-CoV-2 M^{pro} inhibition represents an attractive antiviral therapy to treat COVID-19.[165]

4.2. Results and discussion

4.2.1. Reoptimization of Cys(Acm) translation and deprotection conditions

As mentioned in Chapter 3, one possible reason of the failure of the selection was due to the library loss caused by $PdCl_2$ treatment. However, although the divalent Cu(2+) ion was deleterious for nucleotides,[167] there was no report on harmful effect of Pd(2+). Therefore, I decided to further optimize the Acm deprotection conditions using Pd(2+), aiming at reducing the library loss.

The first improvement to reduce the effect of Pd(2+) ion on nucleotides was reducing the Pd concentration in the deprotection reaction. In addition, due to the low solubility of $PdCl_2$ in water, DMSO solution was used to dissolve $PdCl_2$ in Chapter 3. Therefore, deprotection conditions with less DMSO portion and lower Pd(2+) concentration were to our interest.

To increase the solubility of Pd(2+) ion, the tetrachloride coordination complex of Pd(2+), Na_2PdCl_4 was tested in this trial. Although the mechanism of Pd(2+) deprotection of Acm group is yet unclear, the plausible reaction mechanism shown in Figure 3.2c implies that the Pd(0/2+) core is the most essential element for deprotection. In addition, using tetrachloride coordinated Pd(2+) transferred the soluble Pd form from $Pd(H_2O)_n^{2+}$ to $PdCl_4^{2-}$, and therefore increased its solubility in water. The Na_2PdCl_4 was prepared by dissolving $PdCl_2$ in saturated NaCl solution to a final concentration of 20 mM. The pH of the Na_2PdCl_4 solution was adjusted to pH7 using 200 mM phosphate buffer (NaH_2PO_4 and Na_2HPO_4).

Various deprotection conditions were assessed to determine the deprotection conditions of Acm using Na_2PdCl_4. A template peptide bearing two elongator AUG codon, TPL4, was designed

and Cys(Acm) was incorporated using FIT translation. During the actual RaPID selection, volume of the reaction mixture is altered after translation due to the addition of reverse transcription reagent and selection buffer. To resemble this procedure, the dilution ratio was optimized by diluting the translation product for 2.5×, 5× or 10× with selection buffer (PBST) (Figure 4.3b). The diluted reaction mixture was incubated with 1 mM Na_2PdCl_4 at 37°C for 15 min, before quenched by 10 equivalent of DTC (prepared as 200 mM solution in 200 mM phosphate buffer) and incubated at 37°C for 30 min. The resulting deprotection products were desalted and measured by MALDI-TOF and the mass spectra are shown in Figure 4.3a. Next, concentration of Na_2PdCl_4 was assessed in 10× diluted translation product (Figure 4.3b). The optimized deprotection conditions were determined to be: 10× dilution of translation reaction using PBST, followed by incubation with 1 mM Na_2PdCl_4 (final concentration, abbreviated as fc.) at 37°C for 15 min, before incubated with 10 mM of DTC (fc.) at 37°C for 30 min.

After determining the optimized deprotection conditions using Na_2PdCl_4, its effect on the Input library loss was assessed. Samples from the sixth round of the selection using bicyclic library against the mASGR were taken as example for the test. The mRNA from the sixth round was puromycin ligated, translated with Cys(Acm), reverse transcribed, before setting "Input 1" sample by diluting 0.5 µL of the reaction in 250 µL of water. The reverse transcription mixture was desalted, diluted to 100 µL using PBST containing 1 mg/mL BSA, deprotected with the above determined deprotection conditions, and desalted, followed by setting "Input 2" sample by diluting 0.5 µL of the reaction in 250 µL of water. Concentrations of the cDNA in both Input samples were quantified using qPCR and the percentage input recovery (ratio of Input 2/Input 1) was calculated to be 40.6%. The results indicated that the new deprotection conditions using 1 mM Na_2PdCl_4 could retain ~40% of the input library, which is sufficient for practical selection. A negative control using the old deprotection conditions (20 mM $PdCl_2$) was conducted in parallel and the percentage input recovery was only 2%. The lower input loss achieved by the new deprotection conditions using Na_2PdCl_4 proved that the input loss was at least partially caused by $PdCl_2$ treatment.
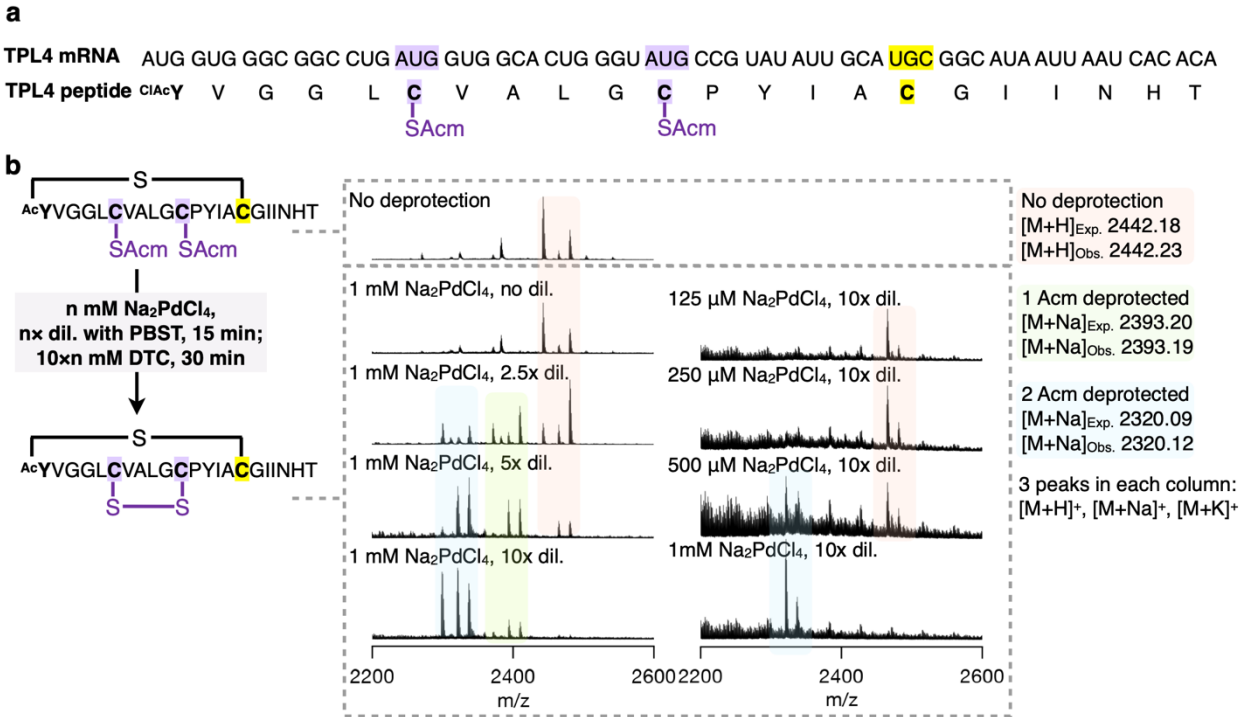
Figure 4.3: Optimization of Cys(Acm) deprotection using Na$_2$PdCl$_4$. (a) mRNA and peptide sequence of the template peptide TPL4; (b) Deprotection of Cys(Acm) incorporated TPL4 under various dilution factors and Na$_2$PdCl$_4$ concentrations.

4.2.2. Optimization of Cys(Phacm) translation and deprotection conditions

After determining the new deprotection conditions of Cys(Acm), the incorporation and deprotection conditions of Cys(Phacm) were evaluated.

To test the translational fidelity of Cys(Phacm) in FIT translation system, incorporation of Cys(Phacm) into model peptide, TPL5 (Figure 4.4a), was assessed in a Met deficient FIT reaction. Since the final goal of this work is to control disulfide formation order of DRP containing 2 pairs of disulfide bonds, both Cys(Phacm) and Cys(Acm) need to be incorporated into the same peptide. Therefore, in this section, elongator AUG codon and tryptophan TGG codon were assigned to Cys(Phacm) and Cys(Phacm), respectively. Since Cys(Phacm) has been successfully applied in the FIT translation system in our laboratory previously (unpublished work), the aminoacylation and translational incorporation conditions were not further optimized. Cys(Phacm)-DBE was charged on tRNA$^{\mathrm{EnAsn}}_{\mathrm{CAU}}$ using flexizyme (dFx) and incorporated into TPL5 during *in vitro* translation. An activated cyanomethyl biotinylated-Phe ester ($^{\mathrm{Bio}}$Phe-CME) was charged on an initiator tRNA, tRNA$^{\mathrm{fMet}}$ and translated onto the N-terminal position of the peptide, to facilitate further pull-down

test. The TPL5 peptide was *in vitro* translated in the presence of 18 amino acids mixture, with exception of Met and Trp, and purified with peptide cleanup SPE C-tip to remove the large protein factors in the translation mixture. The translated TPL5 was tested with MALDI-TOF mass spectrometer and the correct mass was observed with one [Bio]Phe and two Cys(Phacm) residues incorporated into the peptide (Figure 4.4b).

Based on the previous study, deprotection of Cys(Phacm) was conducted by adding 1 units/ml penicillin G acylase (PGA, Iris Biotech) to the translation reaction. However, since this product was no more available, a substitute, penicillin amidase, was purchased from Sigma Aldrich. Although with different name, penicillin amidase possesses similar functionalities in selectively hydrolyzing benzylpenicillin, therefore could be utilized for deprotection of Cys(Phacm). Deprotection conditions were optimized for the new enzyme using *in vitro* translated template peptide TPL5. First, the reaction time and enzyme concentration were investigated first. Deprotection was performed on translated TPL5 by addition of 1 units/ml new PGA and incubation at 37°C for 0.5, 1, 2 and 5 h; or by addition of 1, 10, or 100 units/ml diluted new PGA and incubation at 37°C for 1 h (Figure 4.4c). The deprotected product was tested with MALDI-TOF mass spectroscopy. However, an amidomethyl adduct peak was observed as the main product under all the tested conditions. According to the Phacm deprotection mechanism (Figure 4.2), after hydrolysis of phenylacetyl by PGA, the amidomethyl group is supposed to spontaneously hydrolyze due toits unstable nature. MS-MS spectrum was measured to confirm the partially deprotected product, and the result shown in Figure 4.5 indicated that the amidomethyl product was truly obtained. Considering that a water molecule reacts as the nucleophilic attacking group in the hydrolysis mechanism, replacing it with a stronger nucleophilic molecule might improve the reaction yield. Therefore, the deprotection reaction was performed with various pHs, hydroxylamine concentrations and reaction times. Deprotection was conducted on translated TPL5 by addition of 1 units/ml new PGA and 10 mM hydroxylamine under pH of 5.5, 7.3, 8.3 (100 mM of MES, phosphate, and tris buffer, respectively) and incubation at 37°C for 1 h; or by addition of 1 units/ml new PGA under pH 7.3 with 1, 2.5, 5, or 10 mM hydroxylamine and incubation at 37°C for 0.5 or 1 h (Figure 4.4d). Although deprotection with 1 mM hydroxylamine at pH7 for 60 min can totally remove Phacm, however, a negative control without addition of hydroxylamine also showed clean fully deprotected peak. This might be due to the pH adjusting step by additionally adding 100 mM phosphate buffer, thus, further deprotection optimization with various phosphate concentrations was conducted. Deprotection was conducted on translated TPL5 by addition of 1 units/ml new PGA and 10, 20, 50 or 100 mM pH7.3 phosphate buffer and incubation at 37°C for

2 h (Figure 4.4e). After optimizing the rection pH, time, and adducts, the deprotection conditions were eventually determined to be addition of 1 units/ml new PGA and 50 mM pH7.3 phosphate buffer and incubation at 37°C for 2 h.
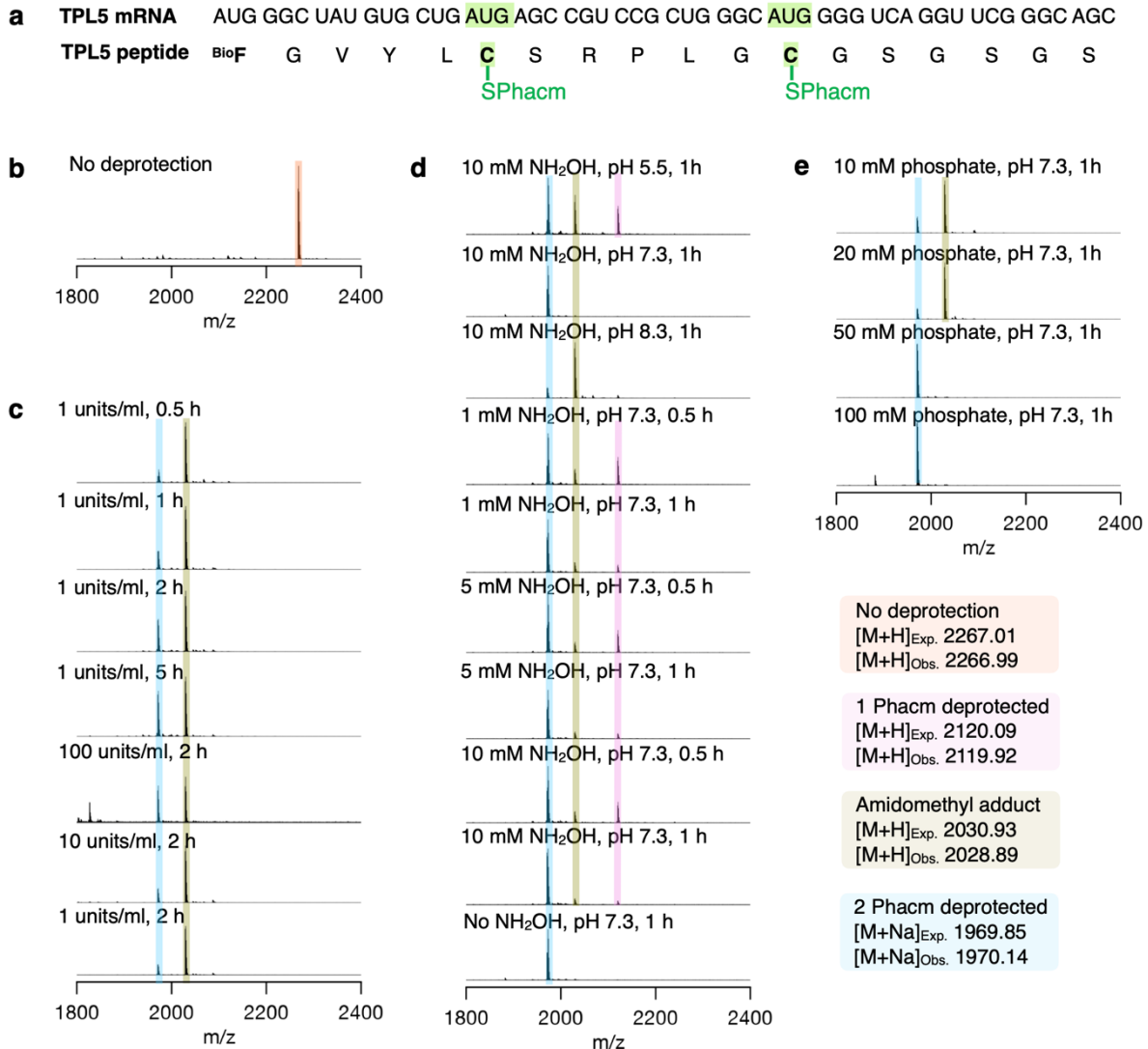
Figure 4.4: Optimization of Cys(Phacm) deprotection. (a) mRNA and peptide sequence of the template peptide TPL5; (b) *In vitro* translated TPL5 incorporates two Cys(Phacm) residues; (c) Deprotection of TPL4 under various PGA dilution factors and reaction times; (d) Deprotection of TPL4 by addition of various concentrations of hydroxylamine under pH 5.5, 7.3 and 8.3; (e) Deprotection of TPL4 by diluting with various concentrations of phosphate buffer.
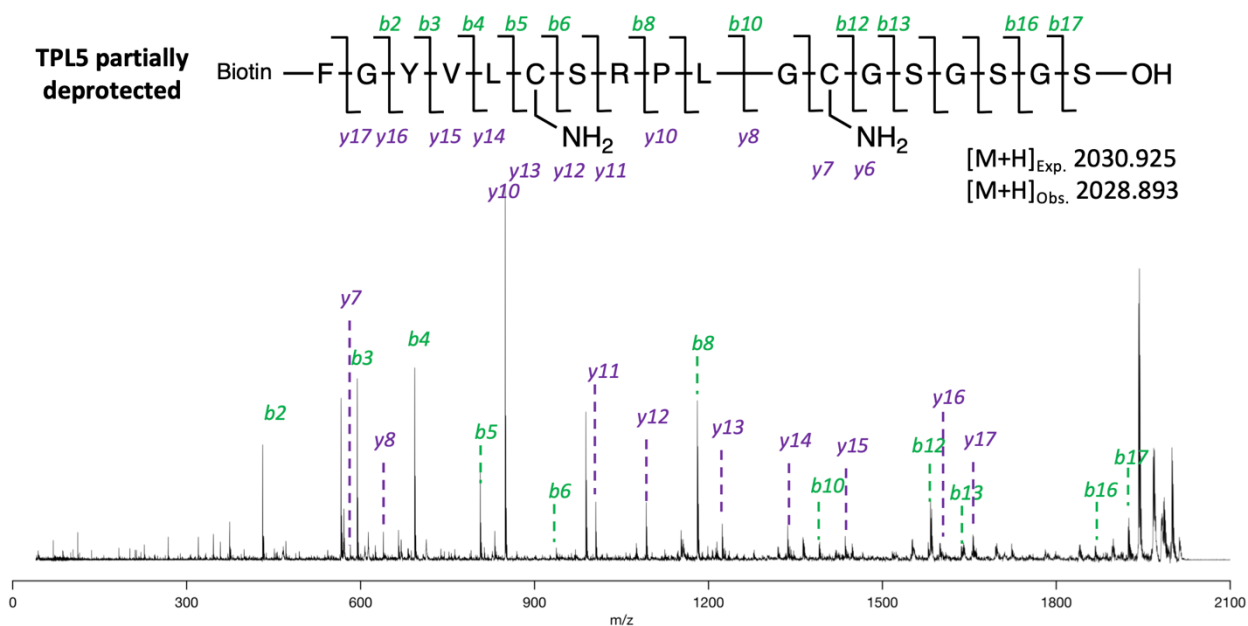
Figure 4.5: MS/MS analysis of the partially deprotected TPL5 peptide indicating the formation of the amidomethyl product.

4.2.3. Tricyclic library construction

After optimizing the deprotection conditions for Cys(Acm) and Cys(Phacm), three tricyclic peptide libraries containing one thioether and two disulfides were constructed and screened against m/hASGR protein. Since two types of protected Cys were simultaneously incorporated in the library, Cys(Phacm) and Cys(Acm) were encoded using the elongator CAU codon and tryptophan UGG codons, respectively. To verify the incorporation efficiency using the UGG codon, activated Cys(Acm)-DBE was acylated on tRNA$^{Pro1E2}_{CCA}$ and translated into template peptide TPL6 (Figure 4.6a), followed by Acm deprotection using the optimized deprotection conditions determined above. Mass spectra indicated that both the translation and Acm deprotection proceeded as expected (Figure 4.6b), therefore, this design was applicable for the library construction.

Considering that the most common disulfide connectivity of natural DRPs containing 3 disulfides is the threaded conformation, bridging between Cys I-IV, II-V and III-VI, I decided to construct the tricyclic libraries only with the knotted connectivity. The Tricyclic_lib1 was designed in a similar way as the bicyclic library in Chapter 3. In details, an N-terminal chloroacetylated Tyr residue (ClAc-^{D}Trp) was introduced for spontaneous formation of a thioether bond with a downstream free Cys residue. One Cys(Phacm) and one Cys(Acm) residues were incorporated upstream to the free Cys, as well as downstream, as indicated in Figure 4.7. The Cys(Phacm) and Cys(Acm) were incorporated using the elongator CAU and tryptophan UGG codon respectively.

A (Gly-Ser)₃ hexapeptide was embedded in the C-terminus as a spacer for linkage to the cognate mRNA template via puromycin. Similar to the bicyclic libraries introduced in Chapter 3, Tricyclic_lib1 contained three different lengths, which is a mixture of peptides with 3, 4 or 5 random amino acids between the ClAc-^{D}Trp and each protected/free Cys.

Other than Tricyclic_lib1, Tricyclic_lib2 and Tricyclic_lib3 were designed based on natural cyclotide sequences, kalata B1 and MCoTI-II. Considering that natural cyclotide structure significantly depends on the length of cyclotide loops, designing libraries with natural loop lengths could lead to development of variants possessing the natural bioactivities. The library design for Tricyclic_lib2 and Tricyclic_lib3 are similar to Tricyclic_lib1, except that all the natural loop residues were replaced by random amino acids encoded with NNK codon. In all the tricyclic libraries, one disulfide bond was substituted with a thioether bond to improves the robustness of cyclic linkage and this change was expected to not change the overall bond connectivity in the whole peptide.

To assess the feasibility of such a library design, model peptide TPL7, corresponding to the Tricyclic_lib1, was designed and translated by incorporating Cys(Phacm) and Cys(Acm) (Figure 4.8). The MALDI mass spectroscopy analysis demonstrated that one ClAc-^{D}Trp, two Cys(Phacm) and two Cys(Acm) were correctly incorporated in the *in vitro* translated model peptides. Stepwise deprotections using the optimized conditions were conducted and the mass spectra suggested the possibility of full deprotection of Phacm and Acm and formation of the disulfide bonds.

Figure 4.6: Evaluation of Cys(Acm) incorporation reprogrammed with UGG codon. (a) mRNA and peptide sequence of template peptide TPL6; (b) Cys(Acm) reprogrammed with UGG codon and tRNA$^{Pro1E2}_{CCA}$ led to clean translation product.

Figure 4.7: Tricyclic library design. All three libraries were initiated with a ClAc-DTrp to form a thioether bond with a downstream free Cys residue. One Cys(Phacm) and one Cys(Acm) residues were incorporated upstream to the free Cys, as well as downstream. Tricyclic_lib1 was constructed with a mixture of 3, 4, or 5 random amino acids in each loop. Tricyclic_lib2 and 3 mimicked kalata B1 and MCoTI-II by substituting each loop with random residues.
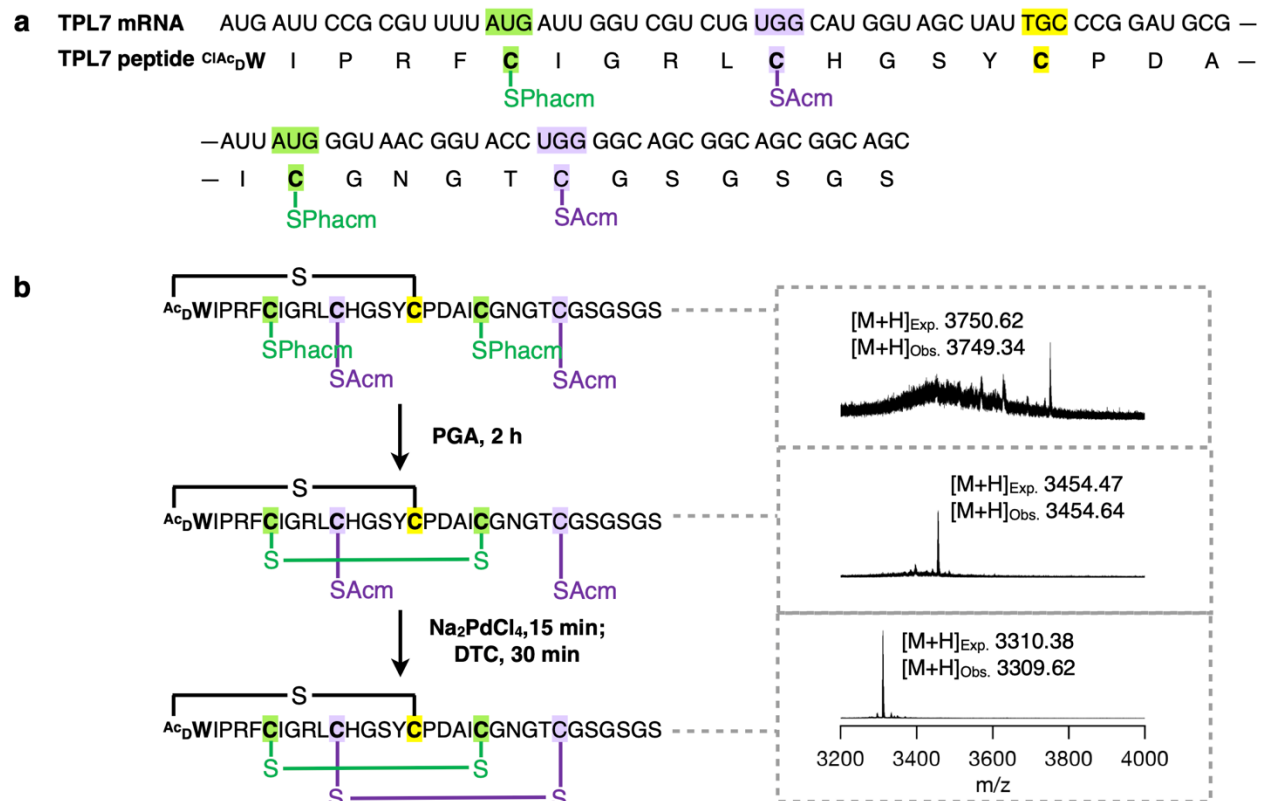
Figure 4.8: Examination of stepwise deprotection of Cys(Phacm) and Cys(Acm). (a) mRNA and peptide sequence of template peptide TPL7; (b) Stepwise deprotection using the optimized deprotection conditions.

4.2.4. RaPID selection against SARS-CoV-2 M^{Pro}

RaPID selection was performed against protein targets M^{Pro} using three tricyclic libraries, with experimental scheme shown in Figure 4.9. In brief, the DNA encoding this library was assembled from degenerate oligonucleotides, transcribed into mRNA, ligated to a puromycin-linked oligonucleotide and FIT translated to produce a library of mRNA-peptide fusion molecules. The libraries were reverse transcribed, followed by pH adjusting by addition of 1/40× (v/v) of 400 mM H$_3$PO$_4$ solution and 1/40× (v/v) of 100 mM phosphate buffer (pH7.3). The H$_3$PO$_4$ solution was additionally added to neutralize the basic condition used in reverse transcription reaction. Phacm deprotection was conducted by adding 1 units/ml PGA enzyme and incubating at 37°C for 2 h. The reaction mixture was desalted and diluted with PBST, before Acm deprotection by incubating with 1 mM Na$_2$PdCl$_4$ (final concentration) at 37°C for 15 min, followed by incubating with 10 mM of DTC at 37°C for 30 min.

 After desalting, the library was screened sequentially for affinity to mock M280 Streptavidin Dynabeads (Invitrogen, half-saturated with biotin) and biotinylated M^{Pro} immobilized on the same type of beads. The cDNAs of the binding peptides were retrieved and the total recovery rate was assessed by qPCR and then amplified with PCR reaction. Notably, to monitor the possible input library loss during the deprotection processes, Input samples were taken after reverse transcription, after Phacm deprotection and after Acm deprotection as Input 1, 2, and 3, respectively. Input recovery, calculated by ratio of Input 3/Input 1, in each selection was shown in Table 4.1. For all rounds in the three selections, the percentage input recovery values were above 20%, indicating that Phacm and Acm deprotection processes were not impeding the library recovery or qPCR reactions. Following seven rounds of selection for Tricyclic_lib1/3 and nine rounds of selection for Tricyclic_lib2, next-generation sequencing was performed to identify high frequency clonal sequences in the recovered final cDNA library.

The recovery rate was calculated as the ratio of peptides binding to the target protein divided by the initial peptides in the library, in this case Input 1. Raising of the recovery rate over selection rounds suggested the enrichment of the target-binding peptides (Figure 4.10). All the three tricyclic libraries showed considerable recovery rate ascent against M^{Pro}, with a positive/negative ratio of over 100-fold in the last round, indicating that the selected peptides exclusively bind to the target protein.

Table 4.1: Input loss in each round caused by the deprotection treatments

Input recovery (%)	rnd 1	rnd 2	rnd 3	rnd 4	rnd 5	rnd 6	rnd 7	rnd 8	rnd 9
Tricyclic_lib1	228.6	26.9	21.9	22.9	27.3	45.0	21.9		
Tricyclic_lib2	148.8	24.1	20.0	43.1	32.9	29.5	28.3	42.0	33.2
Tricyclic_lib3	163.1	22.1	14.7	36.0	30.9	36.2	43.0		

cDNA recovered after all rounds from the three selections were sequenced using next generation sequencing (NGS) and alignment of the top 30 most abundant sequences from the last round of each library was listed in Figure 4.11. In general, the selected sequences were highly converged in all three libraries, especially in Tricyclic_lib3, nearly all the selected peptides shared the identical sequence of MP301. Although both Cys(Phacm) and Cys(Acm) were detected in nearly all the selected sequences, the deletion or addition of a deprotected/free Cys residue was observed in all the libraries, suggesting that the target protein prefers a non-paired free Cys. The consistency of an odd number of Cys residues in the selected peptides might be due to the nature of the target protein, which has a Cys on the proteolytic active center. The extra free Cys in the selected peptides might form a covalent linkage with the active center and therefore inhibit the activity of M^{Pro}.

Sequences with the shortest library design containing 3 random amino acids between each protected/free Cys was preferred in Tricyclic_lib1, whereas the longest design with 5 random residues was almost filtered out. Six sequences (MP101, 102, 103, 104, 106 and 108) were picked up from a few sorted families of the aligned NGS result, with the number indicating the ranking of population in the total NGS dataset (i.e. MP101 and 102 were the most and second abundant sequences in Tricyclic_lib1). The selected peptides preferred negatively charged Lys and Arg throughout the random loops, suggesting some potential electrostatic interactions between the selected peptides and target protein. An additional CAU or UGG codon was detected in the random region near the C-terminus in MP101 and MP104 families, respectively, indicating the presence of an additional free thiol that might assist in target binding by forming a covalent bond with the catalytic Cys. However, since M^{Pro} is a cysteine protease whose substrate specificity was fairly broad and included enzyme-substrate interaction at several enzymatic pockets, the prediction of inhibitory substrate loop would still remain ambiguous.

Similar to Tricyclic_lib1, sequences yielded from Tricyclic_lib2 also preferred an odd number of Cys, providing a free thiol for potential covalent bonding with the target protein. From the several converged families, 5 presentative sequences with the highest abundancy were assigned to

MP201, 202, 203, 204 and 205. Except for an additional Cys(Phacm) found in Loop 3 and 5 in MP201 and 203,respectively, deletion of the Cys(Acm) or Cys(Phacm) was observed in MP202, 204 and 205, resulting in formation of a bicyclic peptide structure after all the deprotection steps.

Compared to Tricyclic_lib1 and lib2, the sequences selected from lib3 were highly converged.Therefore, only 3 peptides, named MP301, 306, 327, were picked up from this selection. The most abundant MP301 contained 1 free Cys in Loop 1 and possibly forming a small 4-residue ring closed with a thioether bond. Both MP306 and 327 possessed the constrained cystine knot connectivity as the library design.
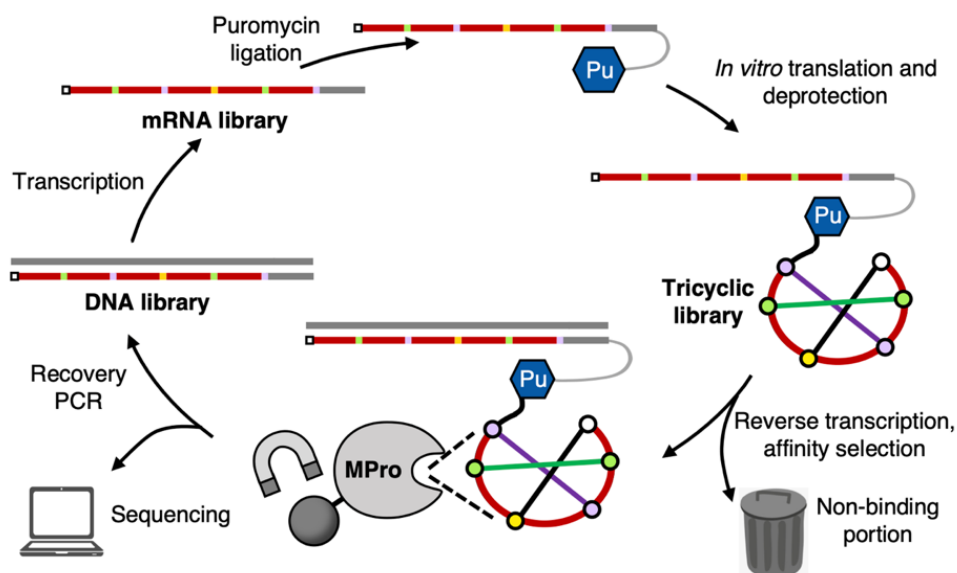
Figure 4.9: Selection scheme of tricyclic libraries against SARS-CoV-2 M^{Pro}.
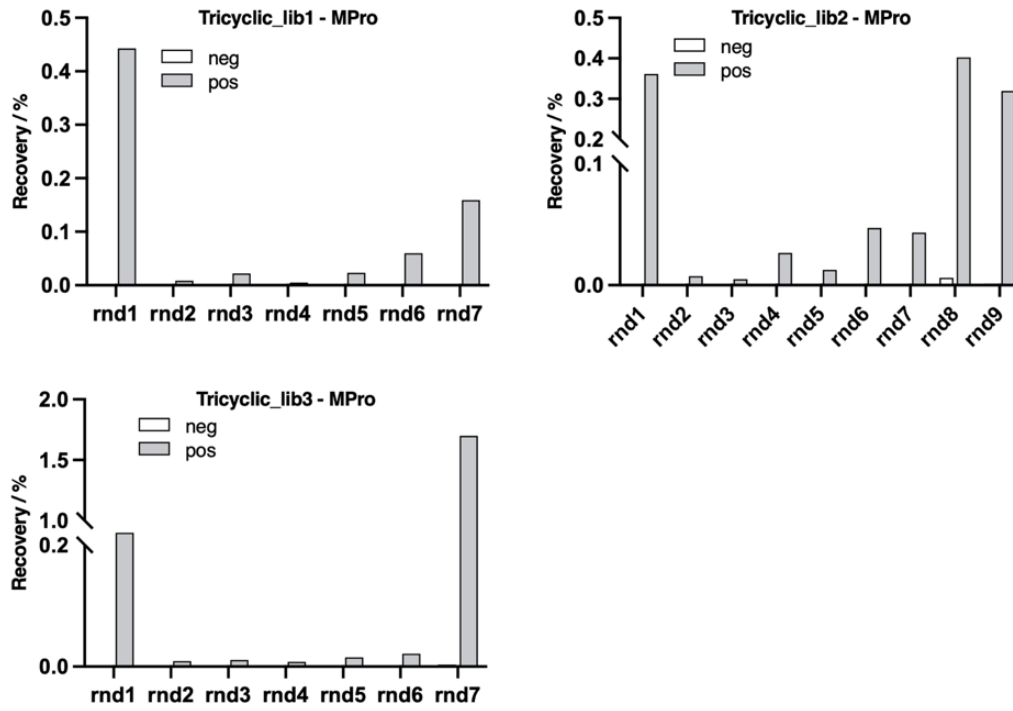
Figure 4.10: Percentage recovery of cDNAs in each library after every round of selection.
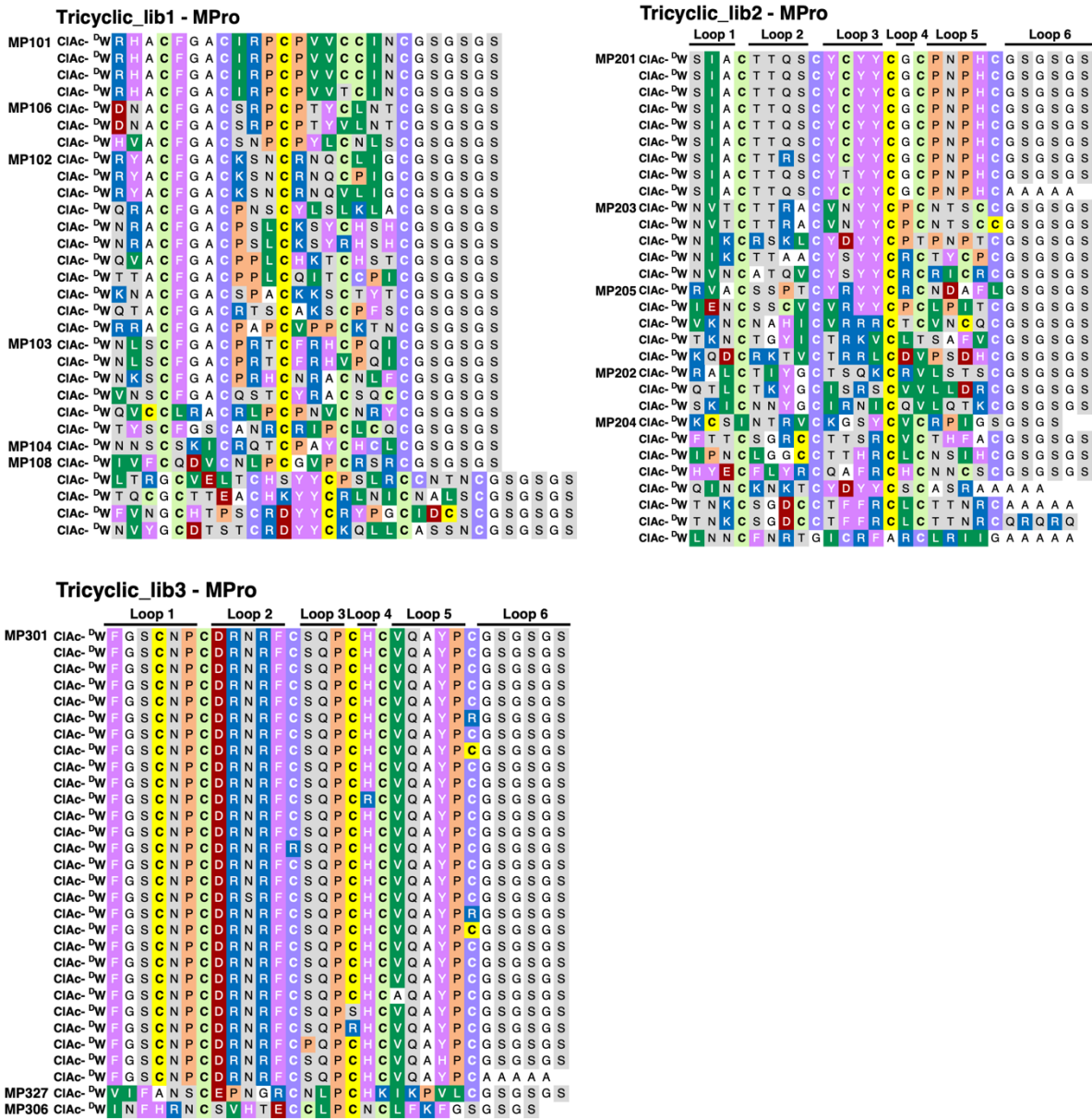
Figure 4.11: Sequences of the top 30 most abundant peptides recovered from affinity selection against M^{Pro} from Tricyclic_lib1, 2, and 3.

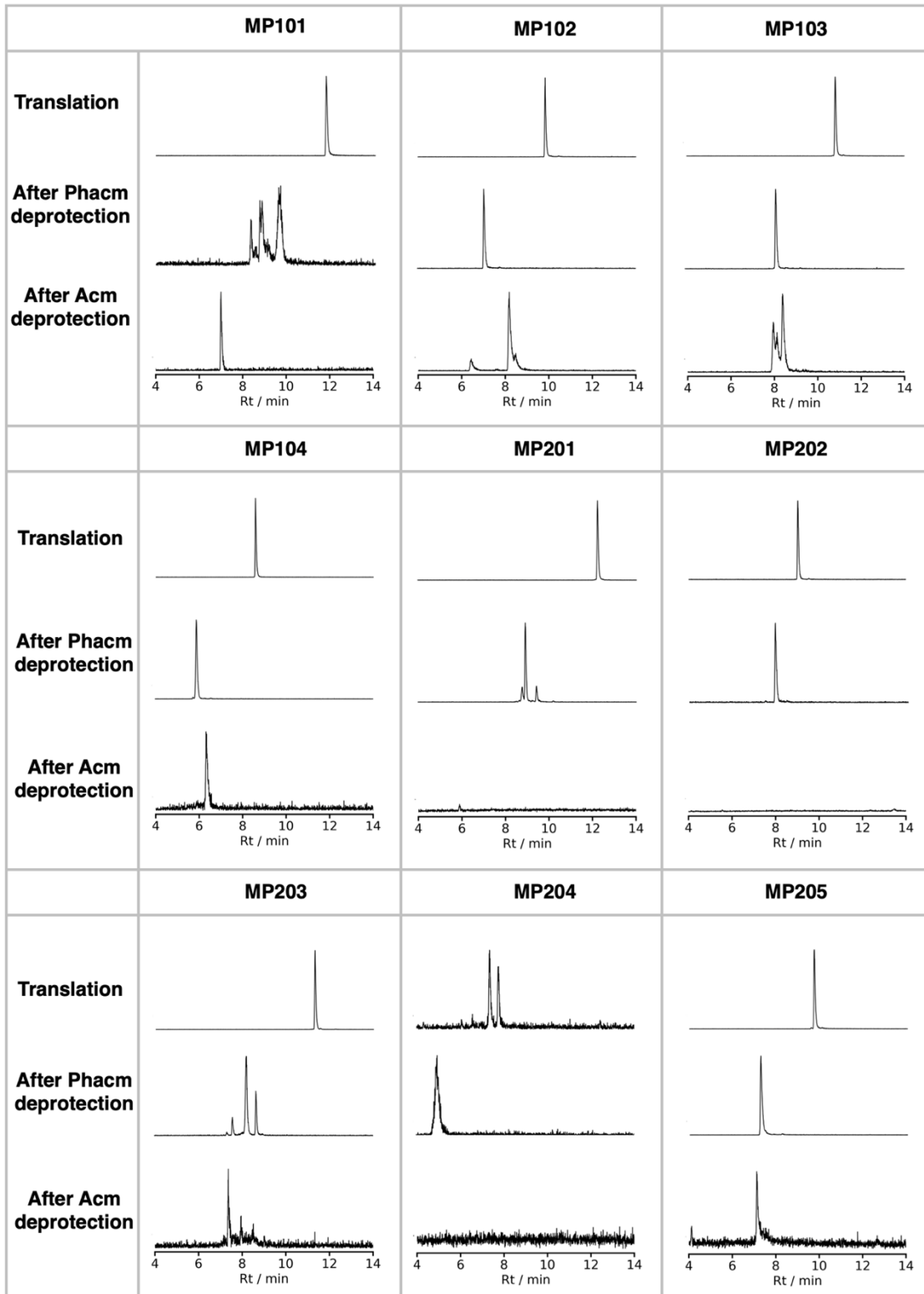## 4.2.5 Clone assay of selected peptides and LC-MS of FIT translated clones

DNA oligomers encoding the selected sequences were ordered from Eurofins and amplified with PCR in the same way as the libraries. The amplified DNAs were individually transcribed, puromycin ligated, *in vitro* translated and deprotected following the selection protocol, followed by flowing against the immobilized target protein to determine the rough binding affinity of each selected peptide. This process mimicking one round of RaPID selection using nucleotide of a single
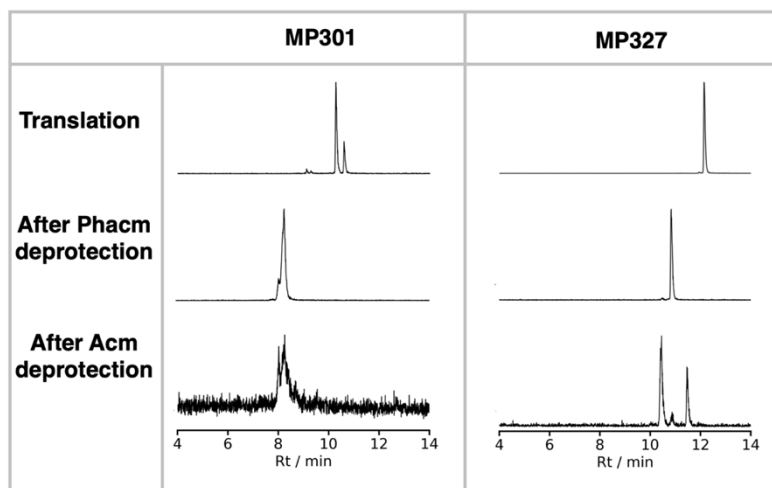
selected sequence was defined as "clone assay". Similar to the real selection procedure, the recovery rate was calculated as the ratio of the amount of the target-binding portion divided by the initial amount of cDNA that have been reverse transcribed. Recovery rate of each selected sequence (referred to as clone) was shown in Table 4.2. Although the clone assay measurement was not strictly quantitative (since the translation efficiency could vary for different sequences), the trend of the recovery rate might to some extend explain for the binding affinity of the selected peptides. All of the selected sequences showed sufficient recovery rate after panning through M^{Pro} immobilized beads, indicating that the selection approach was able to identify M^{Pro} binding peptides. For sequences lacking Cys(Phacm) or Cys(Acm) on the designed position, i.e. MP202 and 205, a control clone named MP2022 or 2052 was constructed by mutating the wrong residue back into Cys(Phacm) or Cys(Acm). Similarly, for selected sequences with additional Cys(Phacm), i.e. MP203 and 306, the Cys(Phacm) was mutated to Ile by changing the encoding oligonucleotides from AUG to AUC. The mutated control clones were named as MP2032 and 3062, respectively. Except 3062, all the other mutated controls showed lower recovery than the selected parent clones, indicating that the non-paired free thiol could benefit the target binding.

In addition to the target-affinity scanning, the clones with relatively high recovery (MP101−104, 201−205, 301, 327) were individually translated and deprotected using the selection conditions. The translation product, Phacm deprotection product, and Acm deprotection product of each clone were tested using LC-MS, to determine the deprotection yield and peptide conformation in the selection procedure (Figure 4.12). The extracted ion chromatograph (EIC) of the expected mass for each peptide was shown in Figure 4.11, with the extraction window of ± 0.1 Da. For MP102, 103 and 327 without additional Cys/Cys(Phacm)/Cys(Acm), a single peak is expected after each deprotection treatment, however, 3 split peaks were observed after the Acm deprotection, suggesting the possible disulfide shuffling. For the other sequences, due to the existence of an additional Cys/Cys(Phacm)/Cys(Acm), the possible disulfide formation at different positions might lead to various peptide conformation, therefore resulting in the splitting of retention times. Although LC-MS analysis of the translated peptides suggested the possibility of disulfide reshuffling caused by Acm deprotection, due to the low translation efficiency, the total ion intensity of all peptides was fairly low (not shown). Therefore, further chemical synthesis was conducted to determine the binding affinity of each selected peptide.

Table 4.2: Clone assay results of the selected peptides

		neg [%]	pos [%]	pos/neg	
MP101	ClAc-DWRHACFGACIRPCPVVCCINCGSGSGS	0.002	0.836	430	
MP102	ClAc-DWRYACFGACKSNCRNQCLIGCGSGSGS	0.002	0.709	422	
MP103	ClAc-DWNLSCFGACPRTCFRHCPQICGSGSGS	0.003	0.767	223	
MP104	ClAc-DWNNSCSKICRQTCPAYCHCLCGSGSGS	0.002	0.959	602	
MP106	ClAc-DWDNACFGACSRPCPTYCLNTCGSGSGS	0.001	0.232	165	
MP108	ClAc-DWIVFCQDVCNLPCGVPCRSRCGSGSGS	0.003	0.350	134	
MP201	ClAc-DWSIACTTQSCYCYYCGCPNPHCGSGSGS	0.003	0.961	283	
MP202	ClAc-DWRALCTIYGCTSQKCRVLSTSCGSGSGS	0.001	1.267	875	
MP2022	ClAc-DWRALCTIYGCTSQKCRCLSTSCGSGSGS	0.002	0.368	198	
MP203	ClAc-DWNVTCTTRACVNYYCPCNTSCCGSGSGS	0.003	1.013	332	
MP2032	ClAc-DWNVTCTTRACVNYYCPCNTSICGSGSGS	0.001	0.374	284	
MP204	ClAc-DWKCSINTRVCKGSYCVCRPIGSGSGS	0.000	0.596	1301	
MP205	ClAc-DWRVACSSPTCYRYYCRCNDAFLGSGSGS	0.001	1.064	772	
MP2052	ClAc-DWRVACSSPTCYRYYCRCNDAFCGSGSGS	0.002	0.472	201	
MP301	ClAc-DWFGSCNPCDRNRFCSQPCHCVQAYPCGSGSGS	0.012	8.927	770	
MP306	ClAc-DWINFHRNCSVHTECCLPCNCLFKFGCGSGSGS	0.004	0.465	114	
MP3062	ClAc-DWINFHRNCSVHTECILPCNCLFKFGCGSGSGS	0.007	1.249	174	
MP327	ClAc-DWVIFANSCEPNGRCNLPCHCIKPVLCGSGSGS	0.004	1.447	402	

	MP101	MP102	MP103
Translation			
After Phacm deprotection			
After Acm deprotection			

	MP104	MP201	MP202
Translation			
After Phacm deprotection			
After Acm deprotection			

	MP203	MP204	MP205
Translation			
After Phacm deprotection			
After Acm deprotection			

Figure 4.12: LC-MS analysis of each clone peptide after translation, Phacm deprotection and Acm deprotection.

4.2.6. Reshuffling observed during solid phase peptide synthesis

Peptides MP101−104, 201−205, 301, 3062 and 327 were divided into 3 groups (Table 4.3) and synthesized with solid phase peptide synthesis using 3 different strategies. Sequences without unpaired Cys, including MP102, 103, 204, 205, 3062 and 327, were defined as Group 1. In detail, Group 1 peptides were either as the library designed (MP102, 103, 3062 and 327) or had deletion of 1 or 2 Cys(Phacm)/Cys(Acm) (MP204 and 205). Sequences with unpaired Cys, including MP101, 201, 203, 301, were defined as Group 2. In details, Group 2 peptides contained either an additional Cys(Phacm) (MP101, 201 and 203) or a free Cys (MP301). In addition, 4 variants of MP104 were defined as Group 3. Since the observed MP104 sequence was ClAc-^{D}WNNSC(Phacm)SKIC(Acm)RQTCPAYC(Phacm)HC(Acm)LC(Acm)GSGSGS, there were only 2 possible disulfide connectivity patterns after Acm deprotection. The C9-C19 and C9-C21 variants were named as MP1041 and MP1042, respectively. To test the significance of the free Cys in MP1041 and 1042, the free Cys was mutated to Ser and the new variants were named MP1043 and MP1044. Since PGA enzyme was a relatively high-cost enzyme, for the large scale SPPS synthesis, Cys(Dpm) was used instead of Cys(Phacm) for the Group 1 peptides, which were expected to possess expected a certain conformation. For Group 2 peptides, which were expected to have uncertain conformation, due to the complexity of their possible disulfide connectivity, Cys(Phacm) was used in SPPS to mimic the deprotection conditions during the selection. For the MP104 variants in Group 3, both Cys(Dpm) and Cys(Phacm) were used to achieve the designed conformation. In all peptide, a Cys(Mmt) was assembled on the free Cys position in the selected

sequences, with the selective removal of Mmt allowing formation of the thioether bond between the N-terminal ClAc group and the deprotected Cys. All the three group peptides were synthesized on NovaPEG Rink Amide resin (Novabiochem) by automated peptide synthesizer and a ClAc group was attached to the N-terminal ^{D}Trp using N-(chloroacetoxy)succinimide (ClAc-NHS), followed by deprotection of Cys(Mmt) using 1% TFA and spontaneous ring closure via formation of the thioether bond. 95% TFA was used to deprotect Cys(Dpm) and other acid-labile protecting groups, as well as release the peptide from resin. For Group 1 and 3 peptides, air oxidation was performed in 0.1 M NH$_4$HCO$_3$ solution, followed by lyophilization to remove the salt. For Group 2 and 3 peptides, the Phacm group was removed by PGA enzyme in 50 mM phosphate, 75% (v/v) DMSO solution. The Acm group was removed by 20 mM Na$_2$PdCl$_4$, followed by quenching with 200 mM DTC. The intermediate products after every cleavage/deprotection step were tested with MALDI-TOF and LC-MS spectroscopy.

Due to the complexity of the multistep SPPS, the desired final products with correct mass were only observed for peptides MP102, 103, 3062, 327, 204, 205, 203, 1044 after the final Acm deprotection. However, LC-MS analysis of the Group 1 peptides before and after Acm deprotection revealed a peak splitting effect caused by the Na$_2$PdCl$_4$ treatment (Figure 4.13), which was consistent to the observation of the translated clone peptides. The splitting effect might be due to the disulfide reshuffling, however, since mechanism of the reaction is not yet clear, I could only speculate that it was caused by the Na$_2$PdCl$_4$ treatment. A recent report by Brik *et al*. claimed about the same disulfide reshuffling effect using PdCl$_2$ as the Acm deprotection reagent.[156] The deprotection reaction was conducted with PdCl$_2$ in pH 1 buffer in their SPPS method, and they claimed that carrying out this step at higher pHs led to significant disulfide reshuffling. Although the Pd-assisted Acm deprotection method has been frequently reported by Brik *et al*. since 2016, the DRPs that were synthesized using this method were mainly natural products with optimized disulfide connectivity. Therefore, the reshuffling issue was omitted in their early studies. Judged by LC-MS spectra, all peaks of each peptide were separated and collected using HPLC, lyophilized and dissolved in DMSO for further binding affinity test. The separated HPLC fragments were named peak *a*, *b*, or *c* based on their retention time, with peak *a* the earliest and peak *c* the latest retention product.

**Table 4.3:** Selected peptide sequences for SPPS synthesis

Group 1

```
MP102  ClAc-DWRYACFGACKSNCRNQCLIGCGSGSGS-NH2
MP103  ClAc-DWNLSCFGACPRTCFRHCPQICGSGSGS-NH2
MP3062 ClAc-DWINFHRNCSVHTECILPCNCLFKFGCGSGSGS-NH2
MP327  ClAc-DWVIFANSCEPNGRCNLPCHCIKPVLCGSGSGS-NH2
MP204  ClAc-DWKCSINTRVCKGSYCVCRPIGSGSGS-NH2
MP205  ClAc-DWRVACSSPTCYRYYCRCNDAFLGSGSGS-NH2
```

Group 2

```
MP101  ClAc-DWRHACFGACIRPCPVVCCINCGSGSGS-NH2
MP201  ClAc-DWSIACTTQSCYCYYCGCPNPHCGSGSGS-NH2
MP203  ClAc-DWNVTCTTRACVNYYCPCNTSCCGSGSGS-NH2
MP301  ClAc-DWFGSCNPCDRNRFCSQPCHCVQAYPCGSGSGS-NH2
```

Group 3

```
MP1041 ClAc-DWNNSCSKICRQTCPAYCHCLCGSGSGS-NH2
MP1042 ClAc-DWNNSCSKICRQTCPAYCHCLCGSGSGS-NH2
MP1043 ClAc-DWNNSCSKICRQTCPAYCHCLSGSGSGS-NH2
MP1044 ClAc-DWNNSCSKICRQTCPAYCHSLCGSGSGS-NH2
```
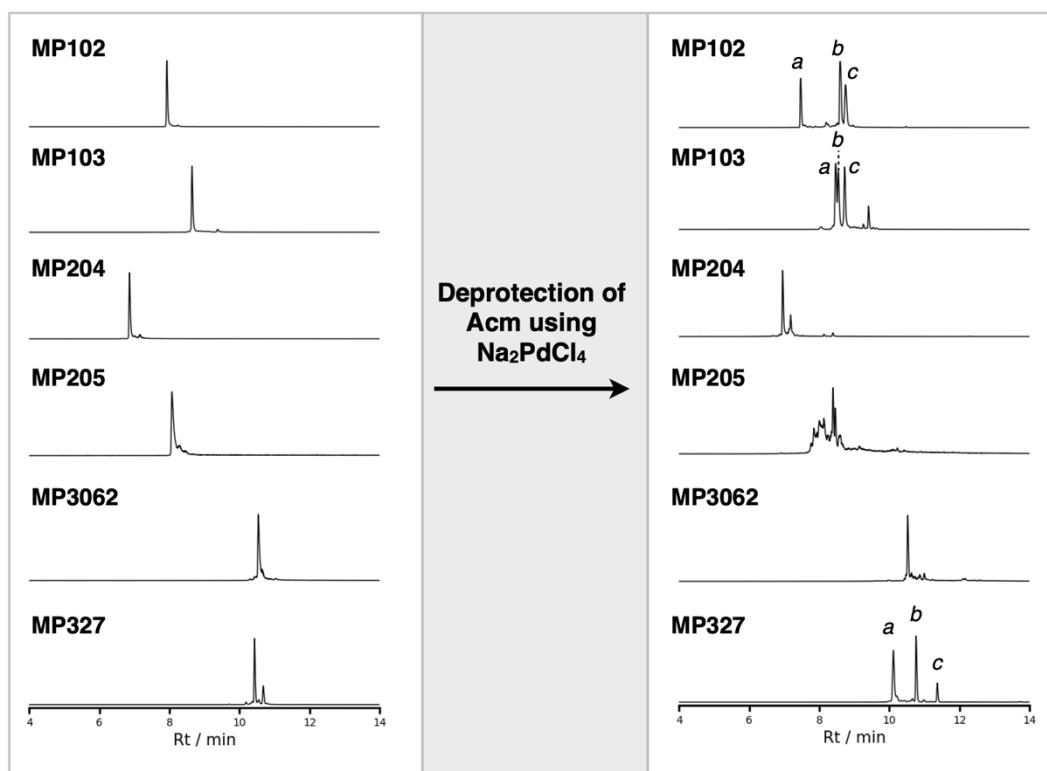

Figure 4.13: LC-MS spectra of Group 1 peptides before and after Na2PdCl4 deprotection. The shown spectra were EICs extracted with the expected product mass.

4.2.7. Binding affinity and inhibitory ability of selected peptides

Target binding affinity of all fragments for each selected sequences were determined using SPR. Similar to the SPR measurement in previous chapters, each peptide at a series of concentrations were flowed through the target protein immobilized on sensorchip and the resonance response was fitted using a single-cycle kinetics. The SPR fitting curves were shown in Figure 4.14, and summary of the estimated K_D values of all synthetic fractions are shown in Table 4.4. K_D values of most of the synthetic peptide fractions were within the high-nanomolar range, indicating their moderate binding affinity to M^{Pro}. MP102-peak *c* and MP327-peak *a* showed two-digit nanomolar binding affinity, with K_D of 84.9 and 38.6 nM, respectively. All fractions of MP102, 103, 3062, 327 and 203 were sent to the collaborators' lab in the University of Oxford for the inhibitory ability test and the IC$_{50}$ values are shown in Table 4.4. Except from MP102-peak *b* and MP103-peak *c*, most of the binding fractions were also moderate low-micromolar range inhibitors. To our surprise, the only Group 2 sequence that was successfully synthesized, MP203, has no inhibitory activity against M^{Pro}. Due to the non-paired Cys in MP203, it was expected to inhibit M^{Pro} by forming covalent disulfide bond with the Cys residue at the catalytic center. The MALDI-TOF measurement during inhibition assay suggested the formation of a covalent bond between M^{Pro} and MP203, with a mass shift of 2972.15 Da based on the original M^{Pro} mass (calculated mass of MP203 is 2971.06 Da, mass spectra not shown). It can be concluded that the non-paired Cys in MP203 must have formed the covalent bond with an M^{Pro} Cys residue other than the catalytic Cys.
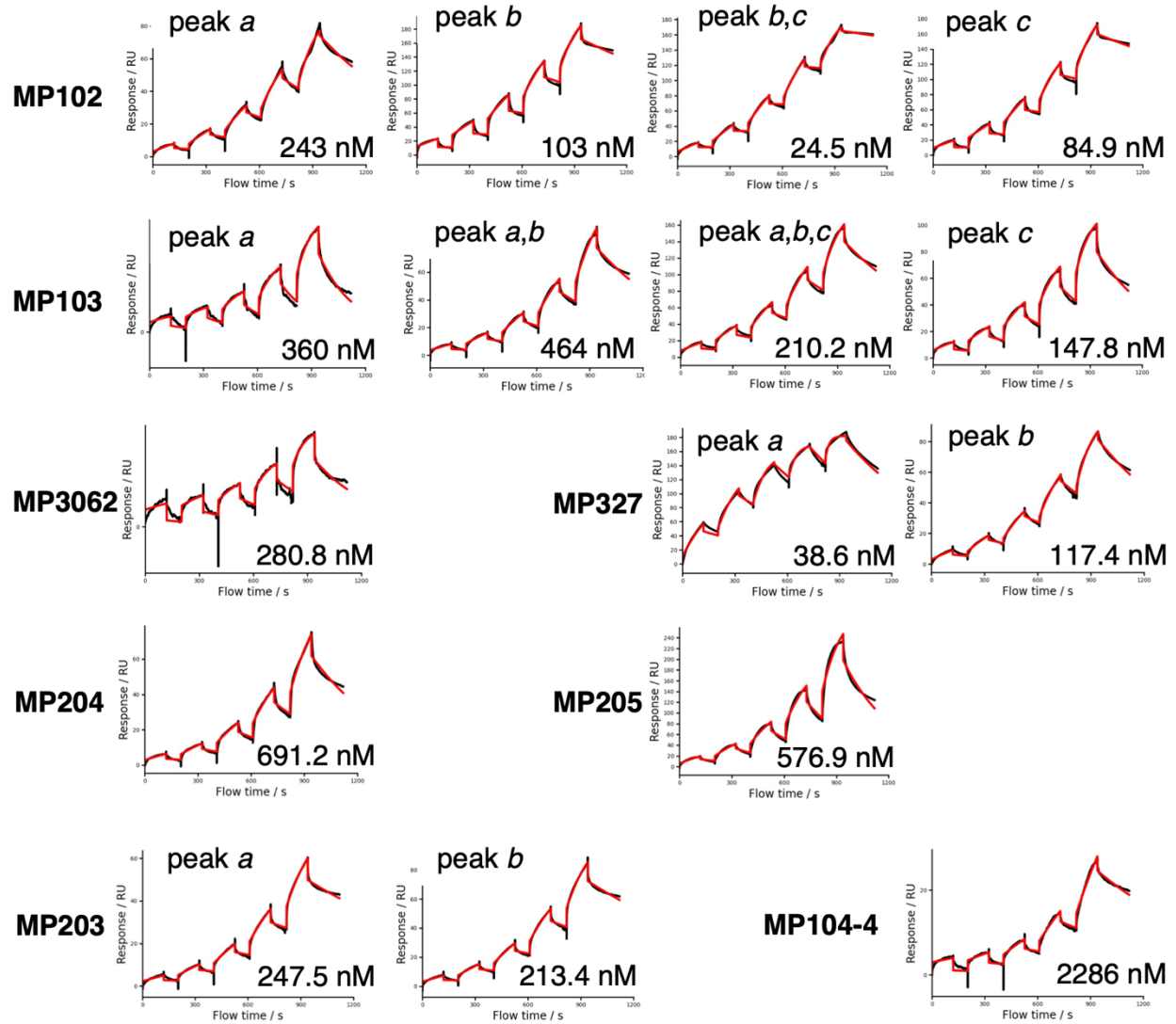
Figure 4.14: Biacore sensorgram of selected MPro binding tricyclic peptides synthesized by SPPS.

Table 4.4: Dissociation constant (K_D) and inhibitory IC$_{50}$ of each HPLC fraction of synthetic peptides

Peptide in DMSO solution			Purity [%]	K_D [nM]	IC$_{50}$ [µM]
MP102	peak a	Ac-^{D}WRYACFGACKSNCRNQCLIGCGSGSGS-NH$_2$	1	243	4.666
	peak b		90	103	0.918
	peak b,c		N.A.	24.5	1.330
	peak c		50	84.9	2.013
MP103	peak a	Ac-^{D}WNLSCFGACPRTCFRHCPQICGSGSGS-NH$_2$	36	360	2.447
	peak a,b		N.A.	464	3.998
	peak a,b,c		N.A.	147.8	3.078
	peak c		20	210.2	0.711
MP3062	single peak	Ac-^{D}WINFHRNCSVHTECILPCNCLFKFGCGSGSGS-NH$_2$	28	280.8	1.631
MP327	peak a	Ac-^{D}WVIFANSCEPNGRCNLPCHCIKPVLCGSGSGS-NH$_2$	90	38.6	1.432
	peak b		90	117.4	2.574
MP204	single peak	Ac-^{D}WKCSINTRVCKGSYCVCRPIGSGSGS-NH$_2$	54	691.2	N.A.
MP205	single peak	Ac-^{D}WRVACSSPTCYRYYCRCNDAFLGSGSGS-NH$_2$	N.A.	576.9	N.A.
MP203	peak a	Ac-^{D}WNVTCTTRACVNYYCPCNTSCCGSGSGS-NH$_2$	80	247.5	No inhibition
	peak b		N.A.	213.4	No inhibition
MP1044	single peak	Ac-^{D}WNNSCSKICRQTCPAYCHSLCGSGSGS-NH$_2$	40	2286	N.A.

4.3. Conclusions and perspectives

In this chapter, protected Cys residues, Cys(Phacm) and Cys(Acm) were successfully incorporated peptide libraries using FIT translation. RaPID selection using the constructed tricyclic libraries was conducted against SARS-CoV-2 M^{Pro}, revealing several target binding peptides with nanomolar range binding affinity. However, LC-MS analysis of the *in vitro* translated and SPPS synthesized selected peptides indicates that the Acm deprotection procedure using Pd(2+) reagents caused severe disulfide reshuffling, which is against the aim of controlling disulfide connectivity during selection procedure. Although 2 HPLC purified peptide fractions of selected MP102 and 103 showed submicromolar IC$_{50}$ values in the inhibitory assays, the original novelty of this work was proved to be inappropriate. Therefore, further investigation of the selected peptides will focus on determining their disulfide connectivity and improving the binding and inhibitory activity, rather than improving the disulfide controlling methodology.

4.4. Materials and methods
List of DNA oligomers

All PCR reactions were conducted using Taq DNA polymerase unless otherwise noted. All oligonucleotides used for PCR reactions in this Chapter are listed in this table.

For tRNA^Pro1E2^_CCA_ preparation	
T7ex5.F22	GGCGTAATACGACTCACTATAG
Pro1E2-3'.R20-Me	TGmGCGGGTGATAGGGGATTC
Pro1E2-5'.F50	GTAATACGACTCACTATAGGGTGATTGGCGCAGCCTGGTAGCGCACTTCG
Pro1E2 CCA.R62	TGGCGGGTGATAGGGGATTCGAACCCCTGACCCCTTCGTCTGGAACGAAGTGCGCTACCAGG
For peptide template synthesis	
T7g10M.F46	TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATA
CGS3an13.R36	TTTCCGCCCCCGTCCTAGCTGCCGCTGCCGCTGCC
TPL4.R90	CTACCCGGTGTTATTGCTTATGCATGCAATATACGGCATACCCAGTGCCACCATCAGGCCGCCCACCATATGTATATCTCCTTCTTAAAG
TPL4-3'.R36	TTTCCGCCCCCGTCCTACCCGGTGTTATTGCTTAT
TPL5.R75	GCTGCCGCTGCCGCTGCCCATGCCCAGCGGACGGCTCATCAGCACATAGCCCATATGTATATCTCCTTCTTAAAG
TPL6.R75	GCTGCCGCTGCCGCTGCCCCAGCCCAGCGGACGGCTCCACAGCACATAGCCCATATGTATATCTCCTTCTTAAAG
TPL7.R120	CTAGCTGCCGCTGCCGCTGCCCCAGGTACCGTTACCCATAATCGCATCCGGGCAATAGCTACCATGCCACAGACGACCAATCATAAAACGCGGAATCATATGTATATCTCCTTCTTAAAG
For tricyclic library synthesis	
tricyclic_lib1-NNK3. R105	CTAGCTGCCGCTGCCGCTGCCCCAMNNMNNMNNCATMNNMNNMNNGCAMNNMNNMNNCCAMNNMNNMNNCATMNNMNNMNNCATATGTATATCTCCTTCTTAAAG
tricyclic_lib1-NNK4. R120	CTAGCTGCCGCTGCCGCTGCCCCAMNNMNNMNNMNNCATMNNMNNMNNMNNGCAMNNMNNMNNMNNCCAMNNMNNMNNMNNCATMNNMNNMNNMNNCATATGTATATCTCCTTCTTAAAG
tricyclic_lib1-NNK5. R135	CTAGCTGCCGCTGCCGCTGCCCCAMNNMNNMNNMNNMNNCATMNNMNNMNNMNNMNNGCAMNNMNNMNNMNNMNNCCAMNNMNNMNNMNNMNNCATMNNMNNMNNMNNMNNCATATGTATATCTCCTTCTTAAAG
tricyclic_lib2. R108	CTAGCTGCCGCTGCCGCTGCCCCAMNNMNNMNNMNNCATMNNGCAMNNMNNMNNMNNCCAMNNMNNMNNMNNCATMNNMNNMNNCATATGTATATCTCCTTCTTAAAG
tricyclic_lib3. R120	CTAGCTGCCGCTGCCGCTGCCCCAMNNMNNMNNMNNMNNCATMNNGCAMNNMNNMNNCCAMNNMNNMNNMNNMNNCATMNNMNNMNNMNNMNNMNNCATATGTATATCTCCTTCTTAAAG
For selected clone synthesis	
MP101.R105	CTAGCTGCCGCTGCCGCTGCCCCAATTAATCATCATAACAACAGGGCACGGACGAATCCAAGCACCAAACATAGCATGACGCATATGTATATCTCCTTCTTAAAG
MP102.R105	CTAGCTGCCGCTGCCGCTGCCCCAACCAATAAGCATCTGATTACGGCAATTACTCTTCCAAGCACCAAACATAGCATACCGCATATGTATATCTCCTTCTTAAAG
MP103.R105	CTAGCTGCCGCTGCCGCTGCCCCAAATCTGAGGCATATGACGAAAGCACGTACGCGGCCACGCACCAAACATACTCAGATTCATATGTATATCTCCTTCTTAAAG
MP104.R105	CTAGCTGCCGCTGCCGCTGCCCCACAGCCAATGCATATAAGCAGGGCACGTCTGCCTCCAAATCTTAGACATACTATTATTCATATGTATATCTCCTTCTTAAAG
MP106.R105	CTAGCTGCCGCTGCCGCTGCCCCAAGTATTCAGCATATACGTCGGGCAAGGACGACTCCAAGCACCAAACATAGCATTATCCATATGTATATCTCCTTCTTAAAG
MP108.R105	CTAGCTGCCGCTGCCGCTGCCCCAACGAGAACGCATCGGAACACCGCAAGGCAGATTCCACACATCCTGCATAAAAACAATCATATGTATATCTCCTTCTTAAAG

MP201.R108	CTAGCTGCCGCTGCCGCTGCCCCAATGAGGATTCGGCATACCGCAATAATACATATACCAAGACT GAGTCGTCATCGCAATACTCATATGTATATCTCCTTCTTAAAG
MP202.R108	CTAGCTGCCGCTGCCGCTGCCCCAACTAGTACTGAGCACACGGCACTTCTGAGAAGTCCAACCAT AAATAGTCATCAAAGCACGCATATGTATATCTCCTTCTTAAAG
MP2022.R108	CTAGCTGCCGCTGCCGCTGCCCCAACTAGTACTGAGCATACGGCACTTCTGAGAAGTCCAACCAT AAATAGTCATCAAAGCACGCATATGTATATCTCCTTCTTAAAG
MP203.R108	CTAGCTGCCGCTGCCGCTGCCCCACATACTAGTATTCATAGGGCAATAATAATTAACCCACGCAC GAGTCGTCATAGTAACATTCATATGTATATCTCCTTCTTAAAG
MP2032.R108	CTAGCTGCCGCTGCCGCTGCCCCAGATACTAGTATTCATAGGGCAATAATAATTAACCCACGCAC GAGTCGTCATAGTAACATTCATATGTATATCTCCTTCTTAAAG
MP204.R102	CTAGCTGCCGCTGCCGCTGCCAATAGGACGCATCACGCAATAAGAACCCTTCCAAACACGAGTAT TTATCGAACACTTCATATGTATATCTCCTTCTTAAAG
MP205.R108	CTAGCTGCCGCTGCCGCTGCCCAAAAAAGCATCATTCATACGGCAATAATACCGATACCACGTAG GACTCGACATCGCAACCCTCATATGTATATCTCCTTCTTAAAG
MP2052.R108	CTAGCTGCCGCTGCCGCTGCCCCAAAAAGCATCATTCATACGGCAATAATACCGATACCACGTAG GACTCGACATCGCAACCCTCATATGTATATCTCCTTCTTAAAG
MP301.R120	CTAGCTGCCGCTGCCGCTGCCCCAAGGATAAGCCTGAACCATATGGCAAGGCTGAGACCAAAACC TATTCCTATCCATAGGATTACAACTACCAAACATATGTATATCTCCTTCTTAAAG
MP306.R120	CTAGCTGCCGCTGCCGCTGCCCCAACCAAACTTAAACAGCATATTGCACGGAAGCATCCACTCCG TATGAACAGACATATTACGATGAAAATTAATCATATGTATATCTCCTTCTTAAAG
MP3062.R120	CTAGCTGCCGCTGCCGCTGCCCCAACCAAACTTAAACAGCATATTGCACGGAAGGATCCACTCCG TATGAACAGACATATTACGATGAAAATTAATCATATGTATATCTCCTTCTTAAAG
MP327.R120	CTAGCTGCCGCTGCCGCTGCCCCAAAGAACAGGCTTAATCATATGGCACGGAAGATTCCACCTAC CATTAGGCTCCATACTATTCGCAAAAATAACCATATGTATATCTCCTTCTTAAAG

Preparation of tRNA$^{Pro1E2}_{CCA}$

The preparation of tRNA^{fMet} and tRNA$^{EnAsn}_{CAU}$ was described in Chapter 3. For tRNA$^{Pro1E2}_{CCA}$, the DNA templates were prepared by extension of forward and reverse primer pairs Pro1E2-5'.F50 and Pro1E2 CCA.R62, followed by PCR reaction adding the T7 promoter T7ex5.F22 and Pro1E2-3'.R20-Me. The transcription reaction for tRNAs contained 40 mM Tris-HCl, 1 mM spermidine, 0.01% (v/v) Triton X-100, 10 mM DTT, 22.5 mM MgCl$_2$, 3.75 mM NTPs, 5 mM GMP 22.5 mM KOH, 10% (v/v) template DNA solution, 0.12 μM T7 RNA polymerase with pH 8.0. All the other reaction conditions including PCR reaction cycles, extraction, precipitation, and polyacrylamide gel purifications were identical to the preparation of tRNAs in Chapter 3. The product was dissolved in H$_2$O and the concentration was assessed and diluted to 250 μM.

Synthesis of peptide template and libraries

Synthesis of mRNA templates for TPL4–7 and the tricyclic libraries was similar to preparation of flexizymes and tRNAs, as described in Chapter 3. Briefly, the DNA templates were prepared by a two-step PCR reaction including the extension of forward primer T7g10M.F46 and the

corresponding reverse primers, followed by a PCR amplification adding the forward and reverse primers T7g10M.F46 and CGS3an13.R36 or TPL4-3'.R36 (for preparation of TPL4). Sequential PCRs were performed using Taq polymerase (95 °C for 40s, 58 °C for 40s, 72 °C for 40s) for minimal number of cycles (4 cycles for extension and 5 cycles for PCR amplification), and the resulting DNA was used as a template for T7 RNA polymerase mediated transcription at 37°C for 16 h prior to polyacrylamide gel purification. *In vitro* transcription was performed in 200 μL and 6500 μL scale for TPL4–7 and the tricyclic libraries, respectively. All the other reaction conditions including extraction, precipitation, and polyacrylamide gel purifications were identical to the preparation of flexizymes. The mRNA template was dissolved in H_2O and the concentration was assessed and diluted to 10 μM.

Synthesis of aminoacylated tRNAs

Aminoacylation of tRNAs using activated amino acids was performed as follows: 25 μM each flexizyme (eFx for ^{Bio}Phe-CME, ClAc-^{L}Tyr-CME and ClAc-^{D}Trp-CME; dFx for Cys(Acm)-DBE and Cys(Phacm)-DBE) and the according tRNA in 50 mM HEPES-KOH pH7.5 were heated to 95 °C for 2 min and cooled slowly to 25 °C. For reactions using eFx or dFx, 600 mM or 100 mM of $MgCl_2$ (final concentration) was added and the reaction was incubated at 25 °C for 5 min. After cooled on ice, the activated amino acid was added to a final concentration of 5 mM, and the reaction was allowed to proceed on ice for the optimal time determined for each amino acid (2 h for ^{Bio}Phe-tRNA^{fMet}, ClAc-^{L}Tyr-tRNA^{fMet}, ClAc-^{D}Trp-tRNA^{fMet} and Cys(Acm)-tRNA$^{Pro1E2}_{CCA}$; overnight for Cys(Phacm)-tRNA$^{EnAsn}_{CAU}$). Aminoacylated tRNA was recovered by twice volume of ethanol precipitation and washed with 70% EtOH and stored dry at -80 °C until use.

***In vitro* translation (FIT translation)**

In vitro translation was performed using a purified recombinant expression system as described in Chapter 2. For testing the translation and deprotection of TLP4–7, individual mRNA template was added at a concentration of 2 μM, aminoacylated tRNAs ^{Bio}Phe-tRNA^{fMet}, ClAc-^{L}Tyr-tRNA^{fMet}, ClAc-^{D}Trp-tRNA^{fMet} and Cys(Acm)-tRNA$^{Pro1E2}_{CCA}$ were included at a concentration of 25 μM, Cys(Phacm)-tRNA$^{EnAsn}_{CAU}$ was included at a concentration of 50 μM. Translation products were purified by SPE C-tip (Nikkyo Technos) and eluted with 1.2 μL of half-saturated α-cyano-4-hydroxycinnamic acid matrix substance dissolved in 0.5% (v/v) acetic acid, 80v/v% acetonitrile, followed by analyzation using MALDI-TOF mass spectroscopy (Bruker UltraFlex mass spectrometer).

RaPID selection of tricyclic libraries agasint M^{Pro}

mRNA templates of the tricyclic libraries were covalently linked to a puromycin linker using home-made T4 RNA ligase, before translated in FIT translation reactions as previously explained in Chapter 2. In brief, each reaction contained 1.2 μM mRNA-puro, 25 μM ClAc-^{D}Trp-tRNA^{fMet} and 25 μM Cys(Acm)-tRNA$^{Pro1E2}_{CCA}$ and 50 μM Cys(Phacm)-tRNA$^{EnAsn}_{CAU}$. The *in vitro* translation was performed at 37 °C for 30 min in 150 μl (for the first round of selection) or 10 μl (from the second to fourth rounds) scale. The mRNA-peptide conjugate libraries were incubated at room temperature for 12 min, and a 0.2× volume of 100 mM EDTA (pH 8.0) was added and incubated at 37 °C for 30 min to induce the dissociation of ribosomes from the mRNA-peptide conjugates. Reverse transcription was carried out at 42 °C for 15 min using the CGS3an13.R36 primer and M-MLV reverse transcriptase lacking RNase H activity (Promega). The pH was adjusted by adding 1/20× translation volume of 400 mM H$_3$PO$_4$, followed by dilution with × translation volume of 100 mM phosphate buffer (pH 7.3). 0.4× translation volume of 1 units/ml PGA enzyme (Sigma-Aldrich, 8.2 unit / mg protein) was added and incubated at 37 °C for 2 h for the deprotection of Phacm, followed by desalting using Sephadex G-25 Fine resin (Cytiva) suspended in PBST buffer. The resulting mixture was diluted with 5.6× translation volume of PBST, before adding 0.1× translation volume of 100 mM Na$_2$PdCl$_4$ and incubating at 37 °C for 15 min for the deprotection of Acm. 0.2× translation volume of 500 mM DTC was added and incubated at 37 °C for 30 min to quench the reaction, followed by centrifuge to remove the precipitate. The supernatant mixture was desalted and panned through a 2× translation volume of Dynabeads M-280 Streptavidin (Thermo Fisher, half-saturated with biotin) and incubated at 4 °C for 15 min, three times as a negative selection. Note that the negative selection was not performed at the first round. The supernatant retrieved from negative selection was added to 1× translation volume of Dynabeads immobilized with biotinylated SARS-CoV-2 M^{Pro} (recombinantly expressed in the collaborator's lab) and the mixture was incubated at 4°C for 30 min (positive selection). The beads were washed with 100 μL of cold PBST for three times and the cDNA was eluted from the beads by heating to 95°C for 5 min in 100 μL of 1× PCR buffer (10 mM Tris–HCl (pH 9.0), 50 mM KCl, 0.1% (v/v) Triton X-100, 0.25 mM dNTP, 2.5 mM MgCl$_2$, 0.25 μM T7g10M.F46 and CGS3an13.R36 primers, and amplified by PCR. The elute (1 μl) was mixed with 19 μl of 1× PCR buffer that contained SYBR Green I (Lonza Japan Ltd) and *Taq* DNA polymerase and the amount of cDNAs was quantified by real-time PCR (Roche). The rest elute was extracted by phenol/chloroform, precipitated by ethanol, dissolved in 10 μL water, and used for *in vitro*

transcription of the subsequent round with the same recipe as preparation of the library. cDNA recovered from each round was sequenced using a MiSeq high-throughput sequencer (Illumina) and analyzed using CLC workbench software (Qiagen).

Solid phase peptide synthesis

All the peptides selected from tricyclic libraries were synthesized at 25 μmol scale using standard Fmoc solid phase peptide synthesis methodology. In brief, each sequence was assembled on NovaPEG Rink Amide Resin (Novabiochem, 0.44 mmol/g) using an automated peptide synthesizer (Syro I Parallel Peptide Synthesizer, Biotage). Fmoc-protected amino acids, Fmoc-Ala-OH, Fmoc-Asp(tBu)-OH, Fmoc-Glu(tBu)-OH,` Fmoc-Phe-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Asn(Trt)-OH, Fmoc-Pro-OH, Fmoc-Gln(Trt)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Val-OH, Fmoc-Trp(Boc)-OH, Fmoc-Tyr(Boc)-OH, Fmoc-Cys(Mmt)-OH, Fmoc-Cys(Dpm)-OH, Fmoc-Cys(Phacm)-OH and Fmoc-Cys(Acm)-OH were purchased from Watanabe Chemical Ltd. Couplings were performed twice with 6 equivalent of Fmoc-protected amino acids (in DMF), 6 equivalent of *O*-(Benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate (HBTU, in DMF), 6 equivalent of 1-hydroxybenzytriazole (HOBt, in DMF), and 12 equivalent of *N,N*-diisopropylethylamine (DIPEA, in NMP) for 40 min. Removal of the Fmoc group was achieved using 40% (v/v) piperidine in DMF at 30 °C for 12 min. To each side-chain protected peptide, 2 ml of 0.2 M chloroacetyl *N*-hydroxysuccinimide ester (ClAc-NHS) in DMF was added to the resin and incubated for 1 h at room temperature. After washing the resin with DMF and CH_2Cl_2 for five times, respectively, the resin was dried in vacuum. Cys(Mmt) was deprotected using 1% (v/v) TFA and 2.5% (v/v) TIS in CH_2Cl_2 for 30 min and the resin was washed with CH_2Cl_2 and DMF for five times, respectively, followed by on-resin thioether cyclization between the N-terminal chloroacetamide group and the sulfhydryl group of the deprotected Cys with 5% DIPEA in DMF for overnight. The peptide was then cleaved from the resin and side-chain deprotected using 92.5% (v/v) TFA, 2.5% (v/v) water, 2.5% (v/v) triisopropylsilane (TIS) and 2.5% (v/v) 3,6-dioxa-1,8-octanedithiol (DODT) for 3 h, followed by precipitation using diethyl ether and lyophilized. For the sequences containing Cys(Dpm) (Group 1 and 3 peptides), air-oxidation was conducted in 0.1 M ammonium bicarbonate solution. The reaction was monitored using Ellman test, MALDI-TOF and LC-MS spectroscopy, until the disulfide bond was completely formed. The reaction mixture was lyophilized, and for sequences containing Cys(Phacm) (Group 2 and 3 peptides), the peptide was dissolved in 50 mM phosphate (pH 7.3), 70% (v/v) DMSO and deprotected by adding

approximately 25 unit of PGA enzyme (Sigma-Aldrich, 8.2 unit / mg protein). The reaction was monitored using MALDI-TOF and LC-MS spectroscopy, until complete deprotection and formation of the disulfide bond. The deprotected peptide was lyophilized and for all sequences containing Cys(Acm) (Group 1, 2 and 3 peptides), the dry peptide was dissolved in 1xPBS in 75% (v/v) DMSO and reacted by addition of final concentration of 20 mM Na_2PdCl_4 solution at 37 °C for 3 h. The reaction was monitored using MALDI-TOF and LC-MS spectroscopy, and after fully deprotection and formation of disulfide bonds, the reaction was quenched by addition of final concentration of 200 mM DTC and incubated at 37 °C for 30 min. The resulting product was centrifuged to remove precipitate and purified by HPLC, followed by lyophilization to obtain each peptide as powder. The purity and integrity of the synthetic peptides were verified using LC-MS (Xevo TQ-S, Waters).

Surface plasmon resonance for binding affinity determination

Binding kinetics of each peptide towards M^{Pro} was determined using a Biacore T200 machine (Cytiva). Running buffer was PBST (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM of KH_2PO_4, 0.05% (v/v) Tween-20). Biotinylated M^{Pro} was immobilized on a Sensor Chip CAP (Cytiva) using Biotin CAPture Reagent (Cytiva) following the manufacturer's instructions. A series of concentrations of each peptide were injected as analyte and the binding kinetics were modeled using a 1:1 binding model.

Chapter 5.

General conclusions

In my PhD study, I aimed at designing and constructing diverse disulfide-rich peptide (DRP) based libraries and developing bioactive DRP-based drug candidates against pharmaceutical protein targets. As a background, naturally occurring DRPs have distinctlyconstrained structures due to the intramolecular disulfide bonds, which provide them with remarkable proteolytic stabilities, target potency, selectivity, and/or cell permeability. However, current engineering methods for DRP discovery, including synthetic or cell-based display methods, either require extensive efforts to create DRP-based libraries with large diversity, or confront difficulties in controlling the peptide architecture during a cell-based expression. Along such background, my PhD study focused on developing bioactive DRP-based peptides with controllable disulfide conformation using an mRNA display derived *in vitro* peptide screening method, random nonstandard peptide integrated discovery (RaPID) selection. Taking advantage of RaPID selection, I explored 3 DRP-based library designs, 1) native cyclotide scaffold, 2) bicyclic scaffold with controlled disulfide connectivity, and 3) tricyclic scaffold with knotted disulfide connectivity, and selected several *de novo* DRP molecules binding to various pharmaceutical protein targets.

In Chapter 2, I first examined whether the DRP scaffold can be applied in RaPID selection (or mRNA display) to establish peptides with potent target affinity. I first designed a library based on the scaffold of a plant-derived cyclotide MCoTI-II, containing a cystine knot core threaded by 3 disulfide bonds. Diversity of the library was derived from 2 randomized loops and the library was screened against coagulation factor XIIa (FXIIa) using *in vitro* mRNA display. The most potent cyclotide analogue, cMCoFx1, inhibits FXIIa with a K_i of 370 pM, shows selectivity of more than three orders of magnitude over most other serine proteases, and potently inhibits coagulation via the intrinsic pathway in human plasma. Analysis of a co-crystal structure of cMCoFx1-FXIIa revealed substrate-like tight binding, with interactions between loop 1 of cMCoFx1 and FXIIa providing the largest contribution to the inhibitor's high potency and exquisite specificity. The experiments in this Chapter has proved the compatibility of DRP scaffold in *in vitro* selection system, meanwhile demonstrated that the structural stability preinstalled in the cyclotide scaffold has benefited the generation of not only potent but also proteolytically stable FXIIa inhibitors.

In Chapter 3, I successfully incorporated protected cysteine, Cys(Acm), into peptide libraries using FIT translation and controlled the peptide folding by adjusting the disulfide topology. By stepwise forming a thioether bond between an N-terminal chloroacetyl group and a non-protected

downstream Cys, followed by a disulfide bond between 2 deprotected Cys residues, two bicyclic peptide libraries were constructed with globular and ribbon conformations. RaPID selection using the bicyclic libraries was conducted against mouse/human asialoglycoprotein receptor (m/hASGR), revealing one potent mASGR binding peptide, masG1, with the expected globular conformation. Comparison of masG1 (K_D = 1.37 nM) and its lariat variant masG1-lariat (K_D = 153 nM) revealed that the bicyclic structure is critical for the potent binding affinity. This study was the first trial on applying protected cysteines in RaPID selection, proving the feasibility for controlling peptide conformation *in situ* during RaPID process, highlighting the immense potential of the approach as a method for rapidly evolving potent DRP-based pharmaceuticals.

Along with the usage of Cys(Acm) in Chapter 3, in Chapter 4, I incorporated another protected Cys species, Cys(Phacm), into peptide libraries using FIT translation and constructed 3 tricyclic peptide libraries using stepwise deprotection method. RaPID selection was conducted against SARS-CoV-2 main protease (M^{Pro}), revealing several target binding peptides with nanomolar range binding affinity. However, LC-MS analysis of the *in vitro* translated and SPPS synthesized peptides indicates that the Acm deprotection procedure using Pd(2+) reagents caused severe disulfide reshuffling, which is against the aim of controlling disulfide connectivity during selection procedure. Although 2 HPLC purified peptide fractions of selected MP102 and 103 showed submicromolar IC_{50} values in the inhibitory assays, the original novelty of this work was proved to be inappropriate. Therefore, further investigation of the selected peptides will focus on determining their disulfide connectivity and improving the binding and inhibitory activity, rather than improving the disulfide controlling methodology.

This work has utilized mRNA display and RaPID selection to discover DRP-based bioactive molecules. It is the first utilization of protected Cys in *in vitro* selection and provides an example of developing DRP molecules with controlled disulfide connectivity during display methods. Although the application of Cys(Acm) in the tricyclic library might have caused disulfide shuffling during deprotection, it could still be used in libraries with one disulfide or in combination with other robust chemical linkages, such as triazole bridge achieved by the click reaction or other cross-coupling reactions.

To summarize, in my PhD work, by utilizing *in vitro* selection technologies, including mRNA display and RaPID selection, I discovered several bioactive DRP molecules with affinities towards various pharmaceutical protein targets. In addition, I controlled the disulfide connectivity by incorporating protected Cys residues in the DRP libraries and stepwise deprotection. This work

provide an efficient and controllable way of DRP discovery, which might lead to development of more DRP scaffold based drug candidates.

Reference list

1.   Hartig, G. R.; Tran, T. T.; Smythe, M. L. Intramolecular disulphide bond arrangements in nonhomologous proteins. *Protein Sci.* **2005**, *14* (2), 474–482.

2.   Northfield, S. E.; Wang, C. K.; Schroeder, C. I.; Durek, T.; Kan, M. W.; Swedberg, J. E.; Craik, D. J. Disulfide-rich macrocyclic peptides as templates in drug design. *Eur. J. Med. Chem.* **2014**, *77*, 248–257.

3.   Barkan, D. T.; Cheng, X.-L.; Celino, H.; Tran, T. T.; Bhandari, A.; Craik, C. S.; Sali, A.; Smythe, M. L. Clustering of disulfide-rich peptides provides scaffolds for hit discovery by phage display: application to interleukin-23. *BMC Bioinform.* **2016**, *17* (1), 1–16.

4.   Poth, A. G.; Mylne, J. S.; Grassl, J.; Lyons, R. E.; Millar, A. H.; Colgrave, M. L.; Craik, D. J. Cyclotides associate with leaf vasculature and are the products of a novel precursor in petunia (Solanaceae). *J. Biol. Chem.* **2012**, *287* (32), 27033–27046.

5.   Smith, J. J.; Cummins, T. R.; Alphy, S.; Blumenthal, K. M. Molecular interactions of the gating modifier toxin ProTx-II with Nav1. 5: implied existence of a novel toxin binding site coupled to activation. *J. Biol. Chem.* **2007**, *282* (17), 12687–12697.

6.   Craik, D. J.; Clark, R. J.; Daly, N. L. Potential therapeutic applications of the cyclotides and related cystine knot mini-proteins. *Expert Opini. Investig. Drugs* **2007**, *16* (5), 595–604.

7.   Terlau, H.; Olivera, B. M. Conus venoms: a rich source of novel ion channel-targeted peptides. *Physiol. Rev.* **2004,** *84*, 41–68.

8.   Imhof, D.; Roy, D.; Albericio, F. Chemical Design and Biomedical Applications of Disulfide-rich Peptides: Challenges and Opportunities. *Front. Chem.* **2020**, *8*.

9.   Fang, G.-M.; Chen, X.-X.; Yang, Q.-Q.; Zhu, L.-J.; Li, N.-N.; Yu, H.-Z.; Meng, X.-M. Discovery, structure, and chemical synthesis of disulfide-rich peptide toxins and their analogs. *Chin. Chem. Lett.* **2018**, *29* (7), 1033–1042.

10.   McGivern, J. G. Ziconotide: a review of its pharmacology and use in the treatment of pain. *Neuropsychiatr. Dis. Treat.* **2007**, *3* (1), 69–85.

11.   Yu, S. W.; Rao, S. S. Advances in the management of constipation-predominant irritable bowel syndrome: the role of linaclotide. *Therap Adv. Gastroenterol.* **2014**, *7* (5), 193–205.

12.   Thomas, R. H.; Luthin, D. R. Current and emerging treatments for irritable bowel syndrome with constipation and chronic idiopathic constipation: focus on prosecretory agents. *Pharmacotherapy* **2015**, *35* (6), 613–630.

13.   Wilkinson, B.; Gilbert, H. F. Protein disulfide isomerase. *Biochim. Biophys. Acta, Proteins and Proteomics* **2004**, *1699* (1–2), 35–44.

14.   Irvine, A. G.; Wallis, A. K.; Sanghera, N.; Rowe, M. L.; Ruddock, L. W.; Howard, M. J.; Williamson, R. A.; Blindauer, C. A.; Freedman, R. B. Protein disulfide-isomerase interacts with a substrate protein at all stages along its folding pathway. *PloS One* **2014**, *9* (1), e82511.

15.   Hatahet, F.; Ruddock, L. W. Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. *Antioxid. Redox Signal.* **2009**, *11* (11), 2807–2850.

16.   Lehrer, R. I.; Cole, A. M.; Selsted, M. E. θ-Defensins: cyclic peptides with endless potential. *J. Biol. Chem.* **2012,** *287* (32), 27014–27019.

17.   Craik, D. Plant cyclotides, *Botanical Research*, Vol. 76; Academic Press: 2015.

18.   Wang, C. K.; Craik, D. J. Designing macrocyclic disulfide-rich peptides for biotechnological applications. *Nat. Chem. Biol.* **2018**, *14* (5), 417–427.

19.   Craik, D. J. Seamless proteins tie up their loose ends. *Science* **2006**, *311* (5767), 1563–1564.

20.   Hara, S.; Makino, J.; Ikenaka, T. Amino acid sequences and disulfide bridges of serine proteinase inhibitors from bitter gourd (Momordica charantia LINN.) seeds. *J. Biochem.* **1989**, *105* (1), 88–92.

21.   Le Nguyen, D.; Heitz, A.; Chiche, L.; Castro, B.; Boigegrain, R.; Favel, A.; Coletti-Previero, M. Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. *Biochimie* **1990**, *72* (6–7), 431–435.

22.   Thongyoo, P.; Bonomelli, C.; Leatherbarrow, R. J.; Tate, E. W. Potent inhibitors of β-tryptase and human leukocyte elastase based on the MCoTI-II scaffold. *J. Med. Chem.* **2009**, *52* (20), 6197–6200.

23.   Sommerhoff, C. P.; Avrutina, O.; Schmoldt, H.-U.; Gabrijelcic-Geiger, D.; Diederichsen, U.; Kolmar, H. Engineered cystine knot miniproteins as potent inhibitors of human mast cell tryptase β. *J. Mol. Biol.* **2010**, *395* (1), 167–175.

24.   Quimbar, P.; Malik, U.; Sommerhoff, C. P.; Kaas, Q.; Chan, L. Y.; Huang, Y.-H.; Grundhuber, M.; Dunse, K.; Craik, D. J.; Anderson, M. A. High-affinity cyclic peptide matriptase inhibitors. *J. Biol. Chem.* **2013**, *288* (19), 13885–13896.

25.   Swedberg, J. E.; Mahatmanto, T.; Abdul Ghani, H.; de Veer, S. J.; Schroeder, C. I.; Harris, J. M.; Craik, D. J. Substrate-guided design of selective FXIIa inhibitors based on the plant-derived Momordica cochinchinensis trypsin inhibitor-II (MCoTI-II) scaffold. *J. Med. Chem.* **2016**, *59* (15), 7287–7292.

26.   Aboye, T. L.; Ha, H.; Majumder, S.; Christ, F.; Debyser, Z.; Shekhtman, A.; Neamati, N.; Camarero, J. A. Design of a novel cyclotide-based CXCR4 antagonist with anti-human immunodeficiency virus (HIV)-1 activity. *J. Med. Chem.* **2012**, *55* (23), 10729–10734.

27.   Ji, Y.; Majumder, S.; Millard, M.; Borra, R.; Bi, T.; Elnagar, A. Y.; Neamati, N.; Shekhtman, A.; Camarero, J. A. In vivo activation of the p53 tumor suppressor pathway by an engineered cyclotide. *J. Am.Chem. Soc.* **2013**, *135* (31), 11623–11633.

28.   Wang, C. K.; Gruber, C. W.; Cemazar, M. a.; Siatskas, C.; Tagore, P.; Payne, N.; Sun, G.; Wang, S.; Bernard, C. C.; Craik, D. J. Molecular grafting onto a stable framework yields novel cyclic peptides for the treatment of multiple sclerosis. *ACS Chem. Biol.* **2014**, *9* (1), 156–163.

29.   Chan, L. Y.; Craik, D. J.; Daly, N. L. Cyclic thrombospondin-1 mimetics: grafting of a thrombospondin sequence into circular disulfide-rich frameworks to inhibit endothelial cell migration. *Biosci. Rep.* **2015**, *35* (6), e00270.

30.   Huang, Y.-H.; Henriques, S. T.; Wang, C. K.; Thorstholm, L.; Daly, N. L.; Kaas, Q.; Craik, D. J. Design of substrate-based BCR-ABL kinase inhibitors using the cyclotide scaffold. *Sci. Rep.* **2015**, *5* (1), 1–15.

31.   Wong, C. T.; Rowlands, D. K.; Wong, C. H.; Lo, T. W.; Nguyen, G. K.; Li, H. Y.; Tam, J. P. Orally active peptidic bradykinin B1 receptor antagonists engineered from a cyclotide scaffold for inflammatory pain treatment. *Angew. Chem. Int. Ed.* **2012**, *51* (23), 5620–5624.

32.   Huang, Y.-H.; Chaousis, S.; Cheneval, O.; Craik, D. J.; Henriques, S. T. Optimization of the cyclotide framework to improve cell penetration properties. *Front. Pharmacol.* **2015**, *6*, 17.

33.   Clark, R. J.; Jensen, J.; Nevin, S. T.; Callaghan, B. P.; Adams, D. J.; Craik, D. J. The engineering of an orally active conotoxin for the treatment of neuropathic pain. *Angew. Chem. Int. Ed.* **2010**, *49* (37), 6545–6548.

34.   Akcan, M.; Stroud, M. R.; Hansen, S. J.; Clark, R. J.; Daly, N. L.; Craik, D. J.; Olson, J. M. Chemical re-engineering of chlorotoxin improves bioconjugation properties for tumor imaging and targeted therapy. *J. Med. Chem.* **2011**, *54* (3), 782–787.

35.   Clark, R. J.; Fischer, H.; Dempster, L.; Daly, N. L.; Rosengren, K. J.; Nevin, S. T.; Meunier, F. A.; Adams, D. J.; Craik, D. J. Engineering stable peptide toxins by means of backbone cyclization: stabilization of the α-conotoxin MII. *Proc. Natl. Acad. Sci.* **2005**, *102* (39), 13767–13772.

36.   Tang, T. M. S. L.; Luk, L. Y. Asparaginyl endopeptidases: enzymology, applications and limitations. *Org. Biomol. Chem.* **2021**, *19*, 5048–5062.

37.   Harris, K. S.; Durek, T.; Kaas, Q.; Poth, A. G.; Gilding, E. K.; Conlan, B. F.; Saska, I.; Daly, N. L.; Van Der Weerden, N. L.; Craik, D. J. Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat. Commun.* **2015**, *6* (1), 1–10.

38.   Poth, A. G.; Colgrave, M. L.; Lyons, R. E.; Daly, N. L.; Craik, D. J. Discovery of an unusual biosynthetic origin for circular proteins in legumes. *Proc. Natl. Acad. Sci.* **2011**, *108* (25), 10127–10132.

39.   Du, J.; Yap, K.; Chan, L. Y.; Rehm, F. B.; Looi, F. Y.; Poth, A. G.; Gilding, E. K.; Kaas, Q.; Durek, T.; Craik, D. J. A bifunctional asparaginyl endopeptidase efficiently catalyzes both cleavage and cyclization of cyclic trypsin inhibitors. *Nat. Commun.* **2020**, *11* (1), 1–11.

40.   Mylne, J. S.; Colgrave, M. L.; Daly, N. L.; Chanson, A. H.; Elliott, A. G.; McCallum, E. J.; Jones, A.; Craik, D. J. Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nat. Chem.l Biol.* **2011**, *7* (5), 257–259.

41.   de Veer, S. J.; Craik, D. J. Naturally occurring disulfide-rich cyclic peptides from plants and animals: Synthesis and biosynthesis. *Chemical Biology of Natural Products*; CRC Press, 2017.

42.   Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *Proc. Natl. Acad. Sci.* **1999**, *96* (22), 12424–12429.

43.   Antos, J. M.; Popp, M. W.-L.; Ernst, R.; Chew, G.-L.; Spooner, E.; Ploegh, H. L. A straight path to circular proteins. *J. Biol. Chem.* **2009**, *284* (23), 16028–16036.

44.   Jia, X.; Kwon, S.; Wang, C.-I. A.; Huang, Y.-H.; Chan, L. Y.; Tan, C. C.; Rosengren, K. J.; Mulvenna, J. P.; Schroeder, C. I.; Craik, D. J. Semienzymatic cyclization of disulfide-rich peptides using Sortase A. *J. Biol. Chem.* **2014**, *289* (10), 6627–6638.

45.   Kwon, S.; Bosmans, F.; Kaas, Q.; Cheneval, O.; Conibear, A. C.; Rosengren, K. J.; Wang, C. K.; Schroeder, C. I.; Craik, D. J. Efficient enzymatic cyclization of an inhibitory cystine knot-containing peptide. *Biotechnol. Bioeng.* **2016**, *113* (10), 2202–2212.

46.   Kane, P. M.; Yamashiro, C. T.; Wolczyk, D. F.; Neff, N.; Goebl, M.; Stevens, T. H. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H (+)-adenosine triphosphatase. *Science* **1990**, *250* (4981), 651–657.

47.   Xu, M.-Q.; Perler, F. B. The mechanism of protein splicing and its modulation by mutation. *EMBO J.* **1996**, *15* (19), 5146–5153.

48.   Kimura, R. H.; Tran, A. T.; Camarero, J. A. Biosynthesis of the cyclotide kalata B1 by using protein splicing. *Angew. Chem. Int. Ed.* **2006**, *45* (6), 973–976.

49.   Camarero, J. A.; Kimura, R. H.; Woo, Y.; Cantor, J.; Shekhtman, A. Biosynthesis of a fully functional cyclotide inside living bacterial cells. *ChemBiochem* **2007**, *8* (12), 1363–1366.

50.   Austin, J.; Kimura, R. H.; Woo, Y.-H.; Camarero, J. A. In vivo biosynthesis of an Ala-scan library based on the cyclic peptide SFTI-1. *Amino Acids* **2010**, *38* (5), 1313–1322.

51.   Gould, A.; Li, Y.; Majumder, S.; Garcia, A. E.; Carlsson, P.; Shekhtman, A.; Camarero, J. A. Recombinant production of rhesus θ-defensin-1 (RTD-1) using a bacterial expression system. *Mol. BioSyst.* **2012**, *8* (4), 1359–1365.

52.   Wu, H.; Hu, Z.; Liu, X.-Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. *Proc. Natl. Acad. Sci.* **1998**, *95* (16), 9226–9231.

53.   Scott, C. P.; Abel-Santos, E.; Wall, M.; Wahnon, D. C.; Benkovic, S. J. Production of cyclic peptides and proteins in vivo. *Proc. Natl. Acad. Sci.* **1999**, *96* (24), 13638–13643.

54.   Jagadish, K.; Camarero, J. A. Recombinant expression of cyclotides using split inteins. *Split Inteins. Methods in Molecular Biology*, Vol. 1495. Humana Press, New York, NY, 2017.

55.   Jagadish, K.; Gould, A.; Borra, R.; Majumder, S.; Mushtaq, Z.; Shekhtman, A.; Camarero, J. A. Recombinant Expression and Phenotypic Screening of a Bioactive Cyclotide Against α-Synuclein-Induced Cytotoxicity in Baker′s Yeast. *Angew. Chem. Int. Ed.* **2015**, *54* (29), 8390–8394.

56.   Huang, P.-S.; Boyken, S. E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016**, *537* (7620), 320–327.

57.   Naamati, G.; Askenazi, M.; Linial, M. ClanTox: a classifier of short animal toxins. *Nucleic Acids Res.* **2009**, *37*, W363–W368.

58.   Kaas, Q.; Westermann, J.-C.; Halai, R.; Wang, C. K.; Craik, D. J. ConoServer, a database for conopeptide sequences and structures. *Bioinformatics* **2008**, *24* (3), 445–446.

59.   Wang, C. K.; Kaas, Q.; Chiche, L.; Craik, D. J. CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. *Nucleic Acids Res.* **2007**, *36*, D206–D210.

60.   Dao, F.-Y.; Yang, H.; Su, Z.-D.; Yang, W.; Wu, Y.; Hui, D.; Chen, W.; Tang, H.; Lin, H. Recent advances in conotoxin classification by using machine learning methods. *Molecules* **2017**, *22* (7), 1057.

61.   Mir, R.; Karim, S.; Amjad Kamal, M.; M Wilson, C.; Mirza, Z. Conotoxins: Structure, therapeutic potential and pharmacological applications. *Curr. Pharm. Des.* **2016**, *22* (5), 582–589.

62.   Kaas, Q.; Yu, R.; Jin, A.-H.; Dutertre, S.; Craik, D. J. ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res.* **2012**, *40* (D1), D325–D330.

63.   Wang, X.; Wang, J.; Wang, X.; Zhang, Y. Predicting the types of ion channel-targeted conotoxins based on avc-svm model. *Biomed Res. Int.* **2017**, *2017*, 2929807.

64.   Wu, Y.; Zheng, Y.; Tang, H. Identifying the types of ion channel-targeted conotoxins by incorporating new properties of residues into pseudo amino acid composition. *Biomed Res. Int.* **2016**, *2016*, 3981487.

65.   Yuan, L.-F.; Ding, C.; Guo, S.-H.; Ding, H.; Chen, W.; Lin, H. Prediction of the types of ion channel-targeted conotoxins based on radial basis function network. *Toxico. In Vitro* **2013**, *27* (2), 852–856.

66.   Zhang, L.; Zhang, C.; Gao, R.; Yang, R.; Song, Q. Using the SMOTE technique and hybrid features to predict the types of ion channel-targeted conotoxins. *J. Theor. Biol.* **2016**, *403*, 75–84.

67.   Ding, H.; Deng, E.-Z.; Yuan, L.-F.; Liu, L.; Lin, H.; Chen, W.; Chou, K.-C. iCTX-type: a sequence-based predictor for identifying the types of conotoxins in targeting ion channels. *Biomed Res. Int.* **2014**, *2014*, 286419.

68.   Mansbach, R. A.; Travers, T.; McMahon, B. H.; Fair, J. M.; Gnanakaran, S. Snails in silico: A review of computational studies on the conopeptides. *Mar. Drugs* **2019**, *17* (3), 145.

69.   Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Protein–protein docking dealing with the unknown. *J. Comput. Chem.* **2010**, *31* (2), 317–342.

70.   Akcan, M.; Cao, Y.; Chongxu, F.; Craik, D. J. The three-dimensional solution structure of mini-M conotoxin BtIIIA reveals a disconnection between disulfide connectivity and peptide fold. *Bioorg. Med. Chem.* **2013**, *21* (12), 3590–3596.

71.   Chi, S.-W.; Park, K.-H.; Suk, J.-E.; Olivera, B. M.; McIntosh, J. M.; Han, K.-H. Solution conformation of αA-conotoxin EIVA, a potent neuromuscular nicotinic acetylcholine receptor antagonist from Conus ermineus. *J. Biol. Chem.* **2003**, *278* (43), 42208–42213.

72.   Rigby, A. C.; Baleja, J. D.; Li, L.; Pedersen, L. G.; Furie, B. C.; Furie, B. Role of γ-carboxyglutamic acid in the calcium-induced structural transition of conantokin G, a conotoxin from the marine snail conus geographus. *Biochemistry* **1997**, *36* (50), 15677–15684.

73.   Verdier, L.; Al-Sabi, A.; Rivier, J. E.; Olivera, B. M.; Terlau, H.; Carlomagno, T. Identification of a novel pharmacophore for peptide toxins interacting with K+ channels. *J. Biol. Chem.* **2005**, *280* (22), 21246–21255.

74.   Kimura, R. H.; Levin, A. M.; Cochran, F. V.; Cochran, J. R. Engineered cystine knot peptides that bind αvβ3, αvβ5, and α5β1 integrins with low-nanomolar affinity. *Proteins* **2009**, *77* (2), 359–369.

75.   Silverman, A. P.; Levin, A. M.; Lahti, J. L.; Cochran, J. R. Engineered cystine-knot peptides that bind αvβ3 integrin with antibody-like affinities. *J. Mol. Biol.* **2009**, *385* (4), 1064–1075.

76.   Moore, S. J.; Leung, C. L.; Norton, H. K.; Cochran, J. R. Engineering agatoxin, a cystine-knot peptide from spider venom, as a molecular probe for in vivo tumor imaging. *PloS One* **2013**, *8* (4), e60498.

77.   Maaß, F.; Wüstehube-Lausch, J.; Dickgießer, S.; Valldorf, B.; Reinwarth, M.; Schmoldt, H. U.; Daneschdar, M.; Avrutina, O.; Sahin, U.; Kolmar, H. Cystine-knot peptides targeting cancer-relevant human cytotoxic T lymphocyte-associated antigen 4 (CTLA-4). *J. Pept. Sci.* **2015**, *21* (8), 651–660.

78.   Kimura, R. H.; Teed, R.; Hackel, B. J.; Pysz, M. A.; Chuang, C. Z.; Sathirachinda, A.; Willmann, J. K.; Gambhir, S. S. Pharmacokinetically stabilized cystine knot peptides that bind alpha-v-beta-6 integrin with single-digit nanomolar affinities for detection of pancreatic cancer. *Clin. Cancer Res.* **2012**, *18* (3), 839–849.

79.   Getz, J. A.; Rice, J. J.; Daugherty, P. S. Protease-resistant peptide ligands from a knottin scaffold library. *ACS Chem. Biol.* **2011**, *6* (8), 837–844.

80.   Getz, J. A.; Cheneval, O.; Craik, D. J.; Daugherty, P. S. Design of a cyclotide antagonist of neuropilin-1 and-2 that potently inhibits endothelial cell migration. *ACS Chem. Biol.* **2013**, *8* (6), 1147–1154.

81.   Glotzbach, B.; Reinwarth, M.; Weber, N.; Fabritz, S.; Tomaszowski, M.; Fittler, H.; Christmann, A.; Avrutina, O.; Kolmar, H. Combinatorial optimization of cystine-knot peptides towards high-affinity inhibitors of human matriptase-1. *PloS One* **2013**, *8* (10), e76956.

82.   Kocsis, A.; Kékesi, K. A.; Szász, R.; Végh, B. M.; Balczer, J.; Dobó, J.; Závodszky, P.; Gál, P.; Pál, G. Selective inhibition of the lectin pathway of complement with phage display selected peptides against mannose-binding lectin-associated serine protease (MASP)-1 and-2: significant contribution of MASP-1 to lectin pathway activation. *J. Immunol.* **2010**, *185* (7), 4169–4178.

83.   Zoller, F.; Markert, A.; Barthe, P.; Zhao, W.; Weichert, W.; Askoxylakis, V.; Altmann, A.; Mier, W.; Haberkorn, U. Combination of phage display and molecular grafting generates highly specific tumor-targeting miniproteins. *Angew. Chem. Int. Ed.* **2012**, *51* (52), 13136–13139.

84.	Zoller, F.; Haberkorn, U.; Mier, W. Miniproteins as phage display-scaffolds for clinical applications. *Molecules* **2011**, *16* (3), 2467–2485.

85.	Hanes, J.; Plückthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci.* **1997**, *94* (10), 4937–4942.

86.	McLafferty, M.; Kent, R.; Ladner, R.; Markland, W. M13 bacteriophage displaying disulfide-constrained microproteins. *Gene* **1993**, *128* (1), 29–36.

87.	Roberts, B. L.; Markland, W.; Ley, A. C.; Kent, R. B.; White, D. W.; Guterman, S. K.; Ladner, R. C. Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. *Proc. Natl. Acad. Sci.* **1992**, *89* (6), 2429–2433.

88.	Josephson, K.; Ricardo, A.; Szostak, J. W. mRNA display: from basic principles to macrocycle drug discovery. *Drug Discov. Today* **2014**, *19* (4), 388–399.

89.	Huang, Y.; Wiedmann, M. M.; Suga, H. RNA display methods for the discovery of bioactive macrocycles. *Chem. Rev.* **2018**, *119* (17), 10360–10391.

90.	Odegrip, R.; Coomber, D.; Eldridge, B.; Hederer, R.; Kuhlman, P. A.; Ullman, C.; FitzGerald, K.; McGregor, D. CIS display: in vitro selection of peptides from libraries of protein–DNA complexes. *Proc. Natl. Acad. Sci.* **2004**, *101* (9), 2806–2810.

91.	Heckler, T. G.; Chang, L. H.; Zama, Y.; Naka, T.; Chorghade, M. S.; Hecht, S. M. T4 RNA ligase mediated preparation of novel" chemically misacylated" tRNAPheS. *Biochemistry* **1984**, *23* (7), 1468–1473.

92.	Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **1989**, *244* (4901), 182–188.

93.	Bain, J.; Diala, E. S.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* **1989**, *111* (20), 8013–8014.

94.	Wang, L.; Brock A.; Herberich B.; Schultz P. G. Expanding the genetic code of Escherichia coli. *Science* **2001**, *292* (5516), 498–500.

95.	Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **2004**, *126* (44), 14306–14307.

96.	Anderson, D. G.; Peng, W.; Akinc, A.; Hossain, N.; Kohn, A.; Padera, R.; Langer, R.; Sawicki, J. A. A polymer library approach to suicide gene therapy for cancer. *Proc. Natl. Acad. Sci.* **2004**, *101* (45), 16028–16033.

97.   Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M. Incorporation of nonnatural amino acids into proteins by using various four-base codons in an Escherichia coli in vitro translation system. *Biochemistry* **2001**, *40* (37), 11060–11064.

98.   Magliery, T. J.; Anderson, J. C.; Schultz, P. G. Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in Escherichia coli. *J. Mol. Biol.* **2001**, *307* (3), 755–769.

99.   Hohsaka, T.; Ashizuka, Y.; Murakami, H.; Sisido, M. Five-base codons for incorporation of nonnatural amino acids into proteins. *Nucleic Acids Res.* **2001**, *29* (17), 3646–3651.

100.   Bain, J.; Switzer, C.; Chamberlin, R.; Benner, S. A. Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature* **1992**, *356* (6369), 537–539.

101.   Hirao, I.; Ohtsuki, T.; Fujiwara, T.; Mitsui, T.; Yokogawa, T.; Okuni, T.; Nakayama, H.; Takio, K.; Yabuki, T.; Kigawa, T. An unnatural base pair for incorporating amino acid analogs into proteins. *Nat. Biotechnol.* **2002**, *20* (2), 177–182.

102.   Forster, A. C.; Weissbach, H.; Blacklow, S. C. A simplified reconstitution of mRNA-directed peptide synthesis: activity of the epsilon enhancer and an unnatural amino acid. *Anal. Biochem.* **2001**, *297* (1), 60–70.

103.   Josephson, K.; Hartman, M. C.; Szostak, J. W. Ribosomal synthesis of unnatural peptides. *J. Am. Chem. Soc.* **2005**, *127* (33), 11727–11735.

104.   Hartman, M. C.; Josephson, K.; Lin, C.-W.; Szostak, J. W. An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLoS One* **2007**, *2* (10), e972.

105.   Matthew, C.; AshtonáCropp, T. Genetic incorporation of 4-fluorohistidine into peptides enables selective affinity purification. *Org. Biomol. Chem.* **2017**, *15* (21), 4536–4539.

106.   Iqbal, E. S.; Dods, K. K.; Hartman, M. C. Ribosomal incorporation of backbone modified amino acids via an editing-deficient aminoacyl-tRNA synthetase. *Org. Biomol. Chem.* **2018**, *16* (7), 1073–1078.

107.   Hartman, M. C.; Josephson, K.; Szostak, J. W. Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc. Natl. Acad. Sci.* **2006**, *103* (12), 4356–4361.

108.   Murakami, H.; Ohta, A.; Goto, Y.; Sako, Y.; Suga, H. Flexizyme as a versatile tRNA acylation catalyst and the application for translation, *Nucleic Acids Symp. Ser. (Oxf.)* **2006**, *50*, 35–36.

109.   Murakami, H.; Saito, H.; Suga, H. A versatile tRNA aminoacylation catalyst based on RNA. *Chem. Biol.* **2003**, *10* (7), 655–662.

110.   Niwa, N.; Yamagishi, Y.; Murakami, H.; Suga, H. A flexizyme that selectively charges amino acids activated by a water-friendly leaving group. *Bioorg. Med. Chem. Lett.* **2009**, *19* (14), 3892–3894.

111.   Liu, W.; de Veer, S. J.; Huang, Y.-H.; Sengoku, T.; Okada, C.; Ogata, K.; Zdenek, C. N.; Fry, B. G.; Swedberg, J. E.; Passioura, T. An Ultrapotent and Selective Cyclic Peptide Inhibitor of Human β-Factor XIIa in a Cyclotide Scaffold. *J. Am. Chem. Soc.* **2021**, *143* (44), 18481–18489.

112.   Craik, D. J.; Daly, N. L.; Bond, T.; Waine, C. Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J. Mol. Biol.* **1999**, *294* (5), 1327–1336.

113.   Colgrave, M. L.; Craik, D. J. Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. *Biochemistry* **2004**, *43* (20), 5965–5975.

114.   Thell, K.; Hellinger, R.; Sahin, E.; Michenthaler, P.; Gold-Binder, M.; Haider, T.; Kuttke, M.; Liutkevičiūtė, Z.; Göransson, U.; Gründemann, C. Oral activity of a nature-derived cyclic peptide for the treatment of multiple sclerosis. *Proc. Natl. Acad. Sci.* **2016**, *113* (15), 3960–3965.

115.   Cascales, L.; Henriques, S. T.; Kerr, M. C.; Huang, Y.-H.; Sweet, M. J.; Daly, N. L.; Craik, D. J. Identification and characterization of a new family of cell-penetrating peptides: cyclic cell-penetrating peptides. *J. Biol. Chem.* **2011**, *286* (42), 36932–36943.

116.   Henriques, S. T.; Huang, Y.-H.; Chaousis, S.; Sani, M.-A.; Poth, A. G.; Separovic, F.; Craik, D. J. The prototypic cyclotide kalata B1 has a unique mechanism of entering cells. *Chem. Biol.* **2015**, *22* (8), 1087–1097.

117.   Huang, Y.-H.; Du, Q.; Craik, D. J. Cyclotides: Disulfide-rich peptide toxins in plants. *Toxicon* **2019**, *172*, 33–44.

118.   de Veer, S. J.; Kan, M.-W.; Craik, D. J. Cyclotides: from structure to function. *Chem. Rev.* **2019**, *119* (24), 12375–12421.

119.   Daly, N. L.; Thorstholm, L.; Greenwood, K. P.; King, G. J.; Rosengren, K. J.; Heras, B.; Martin, J. L.; Craik, D. J. Structural insights into the role of the cyclic backbone in a squash trypsin inhibitor. *J. Biol. Chem.* **2013**, *288* (50), 36141–36148.

120.   Avrutina, O.; Schmoldt, H.-U.; Gabrijelcic-Geiger, D.; Le Nguyen, D.; Sommerhoff, C. P.; Diederichsen, U.; Kolmar, H. Trypsin inhibition by macrocyclic and open-chain variants of the squash inhibitor MCoTI-II. *Biol. Chem.* 2005, *386* (12), 1301–1306.

121.   Pauer, H.-U.; Renné, T.; Hemmerlein, B.; Legler, T.; Fritzlar, S.; Adham, I.; Müller-Esterl, W.; Emons, G.; Sancken, U.; Engel, W. Targeted deletion of murine coagulation factor XII gene-a model for contact phase activation in vivo. *J. Thromb. Haemost.* **2004**, *92* (3), 503–508.

122.   Maas, C.; Renné, T. Coagulation factor XII in thrombosis and inflammation. *Blood* **2018**, *131* (17), 1903–1909.

123.   Renné, T.; Pozgajová, M.; Grüner, S.; Schuh, K.; Pauer, H.-U.; Burfeind, P.; Gailani, D.; Nieswandt, B. Defective thrombus formation in mice lacking coagulation factor XII. *J. Exp. Med.* **2005**, *202* (2), 271–281.

124.   Nickel, K. F.; Long, A. T.; Fuchs, T. A.; Butler, L. M.; Renné, T. Factor XII as a therapeutic target in thromboembolic and inflammatory diseases. *Arterioscler. Thromb. Vasc. Biol.* **2017**, *37* (1), 13–20.

125.   Davoine, C.; Bouckaert, C.; Fillet, M.; Pochet, L. Factor XII/XIIa inhibitors: Their discovery, development, and potential indications. *Eur. J. Med. Chem.* **2020**, *208*, 112753.

126.   Kalinin, D. V. Factor XII(a) inhibitors: a review of the patent literature. *Expert Opin. Ther. Pat.* **2021**, *31* (12), 1155–1176.

127.   Baggio, R.; Burgstaller, P.; Hale, S. P.; Putney, A. R.; Lane, M.; Lipovsek, D.; Wright, M. C.; Roberts, R. W.; Liu, R.; Szostak, J. W. Identification of epitope-like consensus motifs using mRNA display. *J. Mol. Recognit.* **2002**, *15* (3), 126–134.

128.   Cheneval, O.; Schroeder, C. I.; Durek, T.; Walsh, P.; Huang, Y.-H.; Liras, S.; Price, D. A.; Craik, D. J. Fmoc-based synthesis of disulfide-rich cyclic peptides. *J. Org. Chem.* **2014**, *79* (12), 5538–5544.

129.   Handley, T. N. G.; Wang, C. K.; Harvey, P. J.; Lawrence, N.; Craik, D. J. Cyclotide Structures Revealed by NMR, with a Little Help from X-ray Crystallography. *Chembiochem* **2020**, *21* (24), 3463–3475.

130.   Katoh, T.; Sengoku, T.; Hirata, K.; Ogata, K.; Suga, H. Ribosomal synthesis and de novo discovery of bioactive foldamer peptides containing cyclic β-amino acids. *Nat. Chem.* **2020**, *12* (11), 1081–1088.

131.   Rogers, J. M.; Passioura, T.; Suga, H. Nonproteinogenic deep mutational scanning of linear and cyclic peptides. *Proc. Natl. Acad. Sci.* **2018**, *115* (43), 10959–10964.

132.   Wilbs, J.; Kong, X.-D.; Middendorp, S. J.; Prince, R.; Cooke, A.; Demarest, C. T.; Abdelhafez, M. M.; Roberts, K.; Umei, N.; Gonschorek, P. Cyclic peptide FXII inhibitor provides safe anticoagulation in a thrombosis model and in artificial lungs. *Nat. Commun.* **2020**, *11* (1), 1–13.

133.   Vinogradov, A. A.; Nagai, E.; Chang, J. S.; Narumi, K.; Onaka, H.; Goto, Y.; Suga, H. Accurate Broadcasting of Substrate Fitness for Lactazole Biosynthetic Pathway from Reactivity-Profiling mRNA Display. *J. Am. Chem. Soc.* **2020**, *142* (48), 20329–20334.

134.   Goto, Y.; Katoh, T.; Suga, H., Flexizymes for genetic code reprogramming. *Nat. Protoc.* **2011**, *6* (6), 779–790.

135.   Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. 1 H, 13 C and 15 N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **1995**, *6* (2), 135–140.

136.   Li, C. Y.; de Veer, S. J.; Law, R. H.; Whisstock, J. C.; Craik, D. J.; Swedberg, J. E. Characterising the Subsite Specificity of Urokinase‑Type Plasminogen Activator and Tissue‑Type Plasminogen Activator using a Sequence‑Defined Peptide Aldehyde Library. *ChemBioChem* **2019,** *20* (1), 46-50.

137.   Zdenek, C. N.; Hay, C.; Arbuckle, K.; Jackson, T. N.; Bos, M. H.; op den Brouw, B.; Debono, J.; Allen, L.; Dunstan, N.; Morley, T. Coagulotoxic effects by brown snake (Pseudonaja) and taipan (Oxyuranus) venoms, and the efficacy of a new antivenom. *Toxicol. In Vitro* **2019**, *58*, 97–109.

138.   Schrimpf, A.; Linne, U.; Geyer, A. Eight at one stroke - a synthetic tetra-disulfide peptide epitope. *Org. Biomol. Chem.* **2017**, *15* (12), 2512–2521.

139.   Postma, T. M.; Albericio, F. Disulfide formation strategies in peptide synthesis. *European J. Org. Chem.* **2014**, *2014* (17), 3519–3530.

140.   Akondi, K. B.; Muttenthaler, M.; Dutertre, S.; Kaas, Q.; Craik, D. J.; Lewis, R. J.; Alewood, P. F. Discovery, synthesis, and structure-activity relationships of conotoxins. *Chem. Rev.* **2014**, *114* (11), 5815–5847.

141.   Obexer, R.; Walport, L. J.; Suga, H. Exploring sequence space: harnessing chemical and biological diversity towards new peptide leads. *Curr. Opin. Chem. Biol.* **2017**, *38*, 52–61.

142.   Maity, S. K.; Jbara, M.; Laps, S.; Brik, A. Efficient Palladium-Assisted One-Pot Deprotection of (Acetamidomethyl)Cysteine Following Native Chemical Ligation and/or Desulfurization To Expedite Chemical Protein Synthesis. *Angew. Chem. Int. Ed.* **2016**, *55* (28), 8108–8112.

143.   Laps, S.; Sun, H.; Kamnesky, G.; Brik, A. Palladium-Mediated Direct Disulfide Bond Formation in Proteins Containing S-Acetamidomethyl-cysteine under Aqueous Conditions. *Angew. Chem. Int. Ed.* **2019**, *58* (17), 5729–5733.

144.   Gongora-Benitez, M.; Basso, A.; Bruckdorfer, T.; Royo, M.; Tulla-Puche, J.; Albericio, F. Eco-friendly combination of the immobilized PGA enzyme and the S-Phacm protecting group for the synthesis of Cys-containing peptides. *Chemistry* **2012**, *18* (50), 16166–16176.

145. Spiess, M. The asialoglycoprotein receptor: a model for endocytic transport receptors. *Biochemistry* **1990**, *29* (43), 10009–10018.

146. Stockert, R. J. The asialoglycoprotein receptor: relationships between structure, function, and expression. *Physiol. Rev.* **1995**, *75* (3), 591–609.

147. Spiess, M.; Lodish, H. F. Sequence of a second human asialoglycoprotein receptor: conservation of two receptor genes during evolution. *Proc. Natl. Acad. Sci.* **1985**, *82* (19), 6465–6469.

148. Glazier, D. A.; Liao, J.; Roberts, B. L.; Li, X.; Yang, K.; Stevens, C. M.; Tang, W. Chemical synthesis and biological application of modified oligonucleotides. *Bioconjug. Chem.* **2020**, *31* (5), 1213–1233.

149. Huang, Y. Preclinical and clinical advances of GalNAc-decorated nucleic acid therapeutics. *Mol. Ther. Nucleic Acids* **2017**, *6*, 116–132.

150. Ahn, G.; Banik, S. M.; Miller, C. L.; Riley, N. M.; Cochran, J. R.; Bertozzi, C. R. LYTACs that engage the asialoglycoprotein receptor for targeted protein degradation. *Nat. Chem. Biol.* **2021**, 1–10.

151. Sun, Y.; Frenkel-Pinter, M.; Liotta, C. L.; Grover, M. A. The pH dependent mechanisms of non-enzymatic peptide bond cleavage reactions. *Phys. Chem. Chem. Phys.* **2019**, *22* (1), 107–113.

152. Pillai, C.; Nandi, U. Interaction of palladium (II) with DNA. *Biochim. Biophys. Acta* **1977**, *474* (1), 11–16.

153. Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **2009**, *49* (2), 377–389.

154. Yamagishi, Y.; Shoji, I.; Miyagawa, S.; Kawakami, T.; Katoh, T.; Goto, Y.; Suga, H. Natural product-like macrocyclic N-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed de novo library. *Chem. Biol.* **2011**, *18* (12), 1562–1570.

155. Shinbara, K.; Liu, W.; van Neer, R. H. P.; Katoh, T.; Suga, H. Methodologies for Backbone Macrocyclic Peptide Synthesis Compatible With Screening Technologies. *Front. Chem.* **2020,** *8* (447).

156. Laps, S.; Atamleh, F.; Kamnesky, G.; Sun, H.; Brik, A. General synthetic strategy for regioselective ultrafast formation of disulfide bonds in peptides and proteins. *Nat. Commun.* **2021**, *12* (1), 1–9.

157. Marešová, H.; Plačková, M.; Grulich, M.; Kyslík, P. Current state and perspectives of penicillin G acylase-based biocatalyses. *Appl. Microbiol. Biotechnol.* **2014,** *98* (7), 2867–2879.

158.   Tishkov, V.; Savin, S.; Yasnaya, A. Protein engineering of penicillin acylase. *Acta Naturae* **2010**, *2* (3), 47–61.

159.   Grulich, M.; Štěpánek, V.; Kyslík, P. Perspectives and industrial potential of PGA selectivity and promiscuity. *Biotechnol. Adv.* **2013**, *31* (8), 1458–1472.

160.   Duggleby, H. J.; Tolley, S. P.; Hill, C. P.; Dodson, E. J.; Dodson, G.; Moody, P. C. Penicillin acylase has a single-amino-acid catalytic centre. *Nature* **1995**, *373* (6511), 264–268.

161.   Srirangan, K.; Orr, V.; Akawi, L.; Westbrook, A.; Moo-Young, M.; Chou, C. P. Biotechnological advances on penicillin G acylase: pharmaceutical implications, unique expression mechanism and production strategies. *Biotechnol. Adv.* **2013**, *31* (8), 1319–1332.

162.   Lee, J.; Worrall, L. J.; Vuckovic, M.; Rosell, F. I.; Gentile, F.; Ton, A.-T.; Caveney, N. A.; Ban, F.; Cherkasov, A.; Paetzel, M. Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nat. Commun.* **2020,** *11* (1), 1–9.

163.   Hilgenfeld, R. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J.* **2014**, *281* (18), 4085–4096.

164.   Schecter, I.; Berger, A. On the size of the active site in proteases. I. Papain. 1967. *Biochem. Biophys. Res. Commun.* **2012**, *425*, 497–502.

165.   Owen, D. R.; Allerton, C. M.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R. D.; Carlo, A.; Coffman, K. J. An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021**, eabl4784.

166.   Qiao, J.; Li, Y.-S.; Zeng, R.; Liu, F.-L.; Luo, R.-H.; Huang, C.; Wang, Y.-F.; Zhang, J.; Quan, B.; Shen, C. SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model. *Science* **2021**, *371* (6536), 1374–1378.

167.   Aruoma, O. I.; Halliwell, B.; Gajewski, E.; Dizdaroglu, M. Copper-ion-dependent damage to the bases in DNA in the presence of hydrogen peroxide. *Biochem. J.* **1991**, *273* (3), 601–604.

List of accomplishments

【Publications related to this thesis】

1. "An Ultrapotent and Selective Cyclic Peptide Inhibitor of Human β-Factor XIIa in a Cyclotide Scaffold", <u>Wenyu Liu</u>, Simon J. de Veer, Yen-Hua Huang, Toru Sengoku, Chikako Okada, Kazuhiro Ogata, Christina N. Zdenek, Bryan G. Fry, Joakim E. Swedberg, Toby Passioura, David J. Craik, and Hiroaki Suga, *J. Am. Chem. Soc.* **2021**, *143*, 18481–18489

【Publications not related to this thesis】

1. "Display Selection of Exotic Macrocyclic Peptides Expressed under a Radically Reprogrammed 23 Amino Acid Genetic Code", Toby Passioura, <u>Wenyu Liu</u>, Daniel Dunkelmann, Takashi Higuchi, and Hiroaki Suga, *J. Am. Chem. Soc.* **2018**, *140*, 11551–11555

2. "Potent Cyclic Peptide Inhibitors of FXIIa Discovered by mRNA Display with Genetic Code Reprogramming", Daniel J. Ford, Nisharnthi M. Duggan, Sarah E. Fry, Jorge Ripoll-Rozada, Stijn M. Agten, <u>Wenyu Liu</u>, Leo Corcilius, Tilman M. Hackeng, Rene van Oerle, Henri M. H. Spronk, Anneliese S. Ashhurst, Vishnu Mini Sasi, Joe A. Kaczmarski, Colin J. Jackson, Pedro José Barbosa Pereira, Toby Passioura, Hiroaki Suga, and Richard J. Payne, *J. Med. Chem.* **2021**, *64*, 7853–7876

3. "Methodologies for Backbone Macrocyclic Peptide Synthesis Compatible With Screening Technologies", Koki Shinbara, <u>Wenyu Liu,</u> Renier H. P. van Neer, Takayuki Katoh, and Hiroaki Suga, *Front. Chem.* **2020**, *8*, 447

【Oral presentations】

1. "Discovery of a cyclotide-based coagulation Factor XIIa inhibitor by mRNA display", <u>Wenyu Liu</u>, Simon J. de Veer, Yen-Hua Huang, Toru Sengoku, Toby Passioura, David Craik and Hiroaki Suga, The 101st CSJ Annual Meeting, Online Meeting, March 2021

2. "Development of an ultrapotent and selective human β-factor XIIa inhibitor in a cyclotide scaffold", <u>Wenyu Liu</u>, Simon J. de Veer, Yen-Hua Huang, Toru Sengoku, Toby Passioura, David Craik and Hiroaki Suga, The 58th Japanese Peptide Symposium, Online Meeting, October 2021

【Poster presentations】

1．"Ribosomal synthesis of high diversity macrocyclic peptide libraries for selection of protein-protein　interaction in drug-like chemical space", <u>Wenyu Liu</u>, Toby Passioura, Takashi Higuchi, Hiroaki Suga, 第 49 回 若手ペプチド夏の勉強会，長崎市，2017 年 8 月

2. "Development of cyclotide scaffold-based inhibitors of coagulation factor XIIa", <u>Wenyu Liu</u>, Joakim Swedberg, Toby Passioura, David Craik and Hiroaki Suga, 4th International Conference on Circular Proteins and Peptides, Kawasaki, November 2018

3. "Identification of Cyclotide-Like Human Factor XIIa Ligands Through mRNA-Display", <u>Wenyu Liu</u>, Joakim Swedberg, Toby Passioura, David Craik and Hiroaki Suga, 10th International Peptide Symposium, Kyoto, Japan, December 2018

4. "Discovery of Cyclotide-like Human Factor Xlla Ligands through mRNA-display", <u>Wenyu Liu</u>, Simon de Veer, Joakim Swedberg, Yen-Hua Huang, Toby Passioura, David Craik and Hiroaki Suga, The 27th ZESTY Network Seminar, Tokyo, June 2019

5. "Development of a cyclotide-based ultrapotent, selective coagulation Factor XIIa inhibitor using mRNA display", <u>Wenyu Liu</u>, Simon J. de Veer, Joakim E. Swedberg, Toby Passioura, Yen-Hua Huang, David Craik and Hiroaki Suga, Pacifichem 2021, Virtual Symposium, December 2021

【**Awards and fellowships**】

1. Global Science Graduate Course Scholarship (2016-2021)

2. Travel Grant for a 3-months internship by the Graduate Research Abroad in Science Program, School of Science, the University of Tokyo (January-March, 2019)

3. Research Fellowship for Young Scientists, the Japan Society of the Promotion of Science (2020-2021)

4. CSJ Student Presentation Award 2021 at the 101st Chemical Society of Japan Annual Meeting (March 2021)

5. Good Stone Award at the 58th Japanese Peptide Symposium (November 2021)

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Hiroaki Suga, who has the attitude and the substance of genius, for the invaluable guidance and advice throughout the past five years. Since I joined the bioorganic laboratory in 2016, Prof. Suga's has been offering tremendous support and generous help. Without his wonderful supervision, I would not be able to achieve my doctoral dissertation.

Secondly, I would like to thank Dr. Toby Passioura, Dr. Alexander Vinogradov and Dr. Yichao Huang for the essential experiment support and deep discussions in my doctoral works.

Thirdly, I would like to acknowledge Dr. Simon de Veer, Dr. Yen-Hua Huang and Prof. David Craik, from the Institute of Molecular Science, the University of Queensland, for their amazing collaborations. Thanks to their inspiration and technical support, my doctoral research went smoothly.

Next, I would like to thank Prof. Kazuhiro Ogata and Dr. Toru Sengoku for the immense contributions to co-crystal structure measurement part of my thesis.

Then, I would like to express my sincere gratefulness to all the other Suga laboratory members, including staff members, Prof. Yuki Goto, Prof. Takayuki Katoh, and Dr. Naohiro Terasaka, as well as all the post-doctoral fellows and students. They have always been very supportive and helpful, and they mean a lot to me.

Next, I would like to thank the Global Science Graduation Course (GSGC) program, by the School of Science, the University of Tokyo, for the generous and heart-warming support. Furthermore, I would like to acknowledge Japan Society for the Promotion of Science (JSPS) Research Fellowship for Young Scientists DC2 for the research grant, Grant-in-Aid for JSPS fellows No. 20J11284, and the allowance.

Last but not the least, I would like to thank my parents, my husband and other family members for their selfless giving, continuous support, trust and companion throughout my doctoral study.

December 2021 Wenyu Liu