Doctoral Dissertation (Censored)

博士論文　（要約）

Development of artificial MET activators based on engineered self-assembling protein capsules

(自己集合タンパク質カプセルを元にした人工 MET 活性化分子の開発)

A Dissertation Submitted for the Degree of Doctor of Philosophy December 2020

令和２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

Yamato Komatsu
小松　大和

Abstract

Mesenchymal-epithelial transition factor (MET), also called as cMET or Hepatocyte growth factor receptor (HGFR), is a transmembrane protein that regulates numerous downstream signaling proteins related to embryonic development, tissue regeneration and cancer metastasis. MET proteins bind to dimers of natural ligands, Hepatocyte growth factors (HGFs), and are activated by forming HGF-MET dimeric complexes. Activated MET promotes the downstream signaling cascade by recruiting downstream proteins and HGF-MET dimeric complexes are actively internalized into the endosomes by endocytosis pathways.

本部分については、特許申請に使用する予定のため非公開。

Chapter 4 describes *in vitro* selection toward discovery of an alcohol dehydrogenase peptidic enzyme. I constructed an *in vitro* selection system based on enzymatic activity and performed *in vitro* selection, which resulted isolation of several candidate sequences for peptidic enzymes. I analyzed obtained candidates and discussed the possibility of enzymatic activity of them.

Chapter 5 describes rationalization of the difference of the membrane permeability between two peptides, named as CP2 and CP2.3, by means of molecular dynamics (MD) calculation. The MD simulation of the CP2 and CP2.3 was performed and the results were analyzed. I have compared the dynamic of two peptides and discussed the reason and cause of the difference of the membrane permeability.

Chapter 6 is the general conclusion of this thesis. It summarizes achievements in this thesis and discusses perspectives.

Contents

Chapter1

General introduction

1. General introduction

1.1. Introduction

In this chapter, I would like to introduce (1) biology of the HGF-MET signaling pathway, (2) protein nanocapsules and (3) technologies utilized in this study. In the chapter 2 and 3, I focused on the HGF-MET signaling pathway as a target signaling pathway. In nature, this signaling pathway is regulated by sophisticated systems and molecules. I would like to introduce how this signaling pathway is significant for biology and medicine, especially for embryology, oncology and regenerative medicine, and how this research target has sophisticated regulation system in nature. Next, I would like to introduce the molecules which I focused on in this study and their benefits for artificial regulation of signaling pathways containing the HGF-MET pathway. As the third topic in this chapter, I would introduce technologies utilized in this study. This topic especially related to the chapter 4 and 5. Finally, I would clarify the novelties of this study and describe the aim of study.

1.2. HGF-MET pathway and its agonists

1.2.1. Hepatocyte growth factor (HGF)

HGF is a multi-functional cytokine secreted by mesenchymal cells and a starting molecule to drive the HGF-MET signaling pathway. Originally, HGF was detected in the serum of hepatectomized rats as a mitogen which helps the liver regeneration in 1984[1,2] and molecularly cloned as a growth factor for hepatocyte in 1989[3,4]. Currently, it is known that HGF is an identical molecule to the scatter factor (SF)[5], which was originally identified as a fibroblast-derived cell motility factor for epithelial cells, and plays an important role in various cellular activities such as embryonic development, wound healing and tissue regeneration.

Focusing on the regulation of HGF in nature, this bioactive protein is controlled by not only production and secretion but also extracellular processing and activation[6,7]. HGF is produced and secreted by mesenchymal cells as an inactive precursor protein (proHGF). The proHGF is a single chain protein comprised of 728 amino acids, having ~84 kDa of molecular weight. This inactive protein contains three regions, a signal sequence with 31 amino acids, a pre-α-subunit with 463 amino acids and a pre-β-subunit with 234 amino acids (Figure 1.1). After the secretion, proHGFs are recognized and processed by extracellular proteases, such as HGF activator (HGFA), urokinase-type plasminogen activator (uPA), plasma kallikrein, coagulation factors XI and XII, transmembrane protease serine S1 member 13 (TMPRSS13), matriptase and hepsin. These proteases detach the N-terminal signal sequence and cleave the peptide bond between Arg494 and Val495 of a single chain proHGF. The resulted molecule, activated HGF, has about 70-80 kDa of molecular weight and consists of two subunits, α-subunit and β-subunit linked via a disulfide bond. The α-subunit and the β-subunit consists of 463 and 234 amino acids and have ~54 kDa and ~26 kDa of molecular weight, respectively. They are glycosylated proteins and have many intramolecular disulfide bonds to form their active conformation.

An activated HGF consists of six domains, an amino-terminal hairpin loop (HL)domain, kringle 1-4 (K1-4) domains and a serine proteinase homology (SPH) domain. A HL domain and kringle domains are contained in an α-subunit and a SPH domain is contained in a β-subunit. Kringle domains are autonomous protein domains which fold into large loops stabilized by three disulfide bonds. The SPH domain is a domain which has homology with serine proteases but lacks enzymatic activity by mutations in essential residues[8]. Although the whole structure of HGF has not been defined, crystal structures of NK1 domain (the N-terminal HL and the K1 domain) and the β-subunit are shown in the Figure 1.2. Crystal structure analyses showed independently isolated NK1 forms head-to-tail dimers, which indicates two NK1 strongly interact each other[9]. This interaction can assist dimerization of HGFs necessary for MET activation (mentioned below).

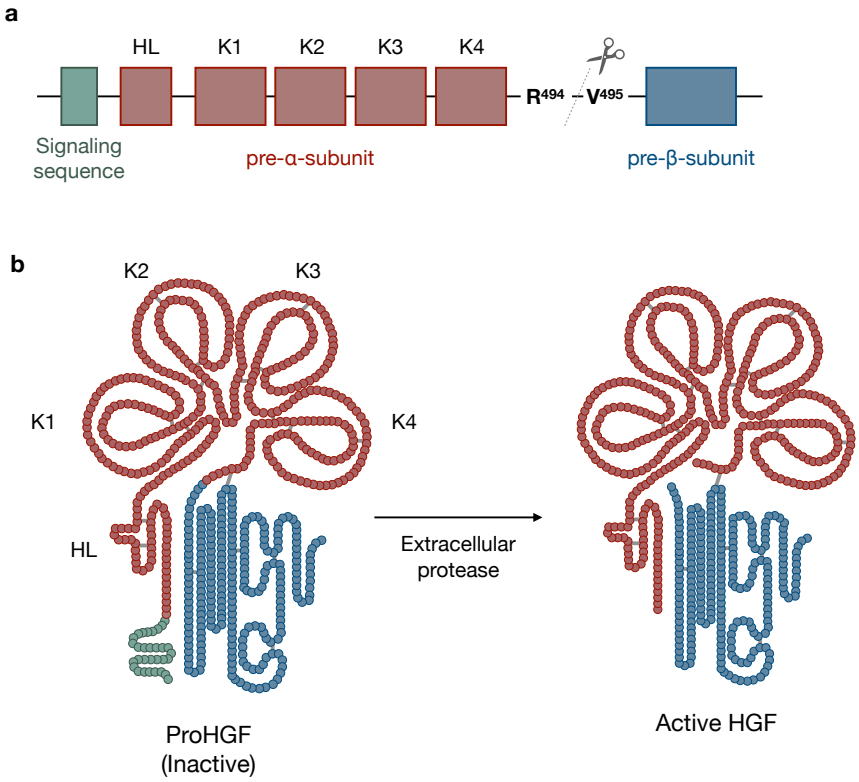


Figure 1.1 | a) Domain map of proHGF. The cleavage site is located between Arg494 and V495 residues. b) Schematic diagrams of the structure of proHGF and active HGF. HL: hairpin loop domain, K1-4: Kringle 1-4 domains.

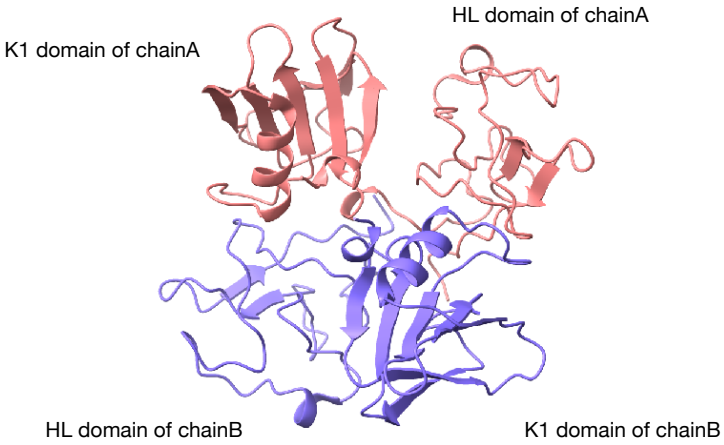


Figure 1.2 | Crystal structure of NK1 domains (the N-terminal HL and the K1 domain) of HGF. The crystal structure analysis shows NK1 domains of HGF form homodimer by themselves.

1.2.2. Mesenchymal-epithelial transition factor (MET)

MET (cMET), also known as hepatocyte growth factor receptor (HGFR), is a single-pass transmembrane receptor protein encoded by the *MET* gene and the specific receptor of HGF[6,10,11]. Although expression of HGF is restricted to cells of mesenchymal origin, MET is normally expressed by epithelial cells. The known ligands of MET are only HGF and its splicing isoforms, NK1 and NK2. MET belongs to the family of receptor tyrosine kinases (RTKs), which are generally activated by growth factors or hormones and regulate upstream signaling pathways related to cellular activities important for developmental biology, such as cellular migration, cell proliferation and morphogenesis[12–14].

The domain structure of MET is shown in the Figure 1.3. MET consists two subunits; α-subunit and β-subunit linked via a disulfide bond just like HGF. It is activated from a precursor by cleaving at a furin site located between residues 307 and 308 in the post-Golgi compartment[11]. The α-subunit is completely extracellular and the β-subunit is transmembrane subunits comprised of a Sema domain, a PSI (Plexins, Semaphorins and Integrins; other representative proteins that have the same domain) domain, IPT (Immunoglobulin-like fold shared by Plexins and Transcriptional factors) 1-4 domains and an intracellular domain. The Sema domain is a protein domain of ~450 amino acids identified in the semaphorin axon guidance proteins and the Sema domain of the MET adopts a β-propeller fold. The PSI domain is a cysteine-rich region, which consists of ~50 residues and contains four conserved disulfide bonds. IPT 1-4 domains form a stalk region of the MET protein. The intracellular domain of MET consists of a juxtamembrane domain, a kinase domain and a multifunctional docking site. The kinase domain has tyrosine kinase catalytic activity and its substrate tyrosine residues as with other RTK family proteins. The multifunctional docking site also has tyrosine residues to be phosphorylated in the step of MET activation and is related to the interactions with downstream signaling proteins.

MET protein has two forms; an active form where intracellular tyrosine residues were phosphorylated and an inactive form where the tyrosine residues were not phosphorylated. The activation of MET proteins was induced by the dimerization of MET proteins. Activated HGFs form a homodimeric complex and bind to two MET proteins. This binding property results the formation of HGF-MET dimers, comprised of two HGF proteins and two MET proteins, on the surface of hepatocytes and makes the intracellular domains of MET close to each other. Approached intracellular domains of MET automatically phosphorylates the other's tyrosine residues located in the kinase domain, *i.e.,* Tyr1234 and Tyr1235, and subsequent phosphorylation of tyrosine residues in the multifunctional domain, *i.e.,* Tyr1349 and Tyr1356, results an active form of MET, where downstream signaling proteins bind to the phosphorylated multifunctional docking domain.

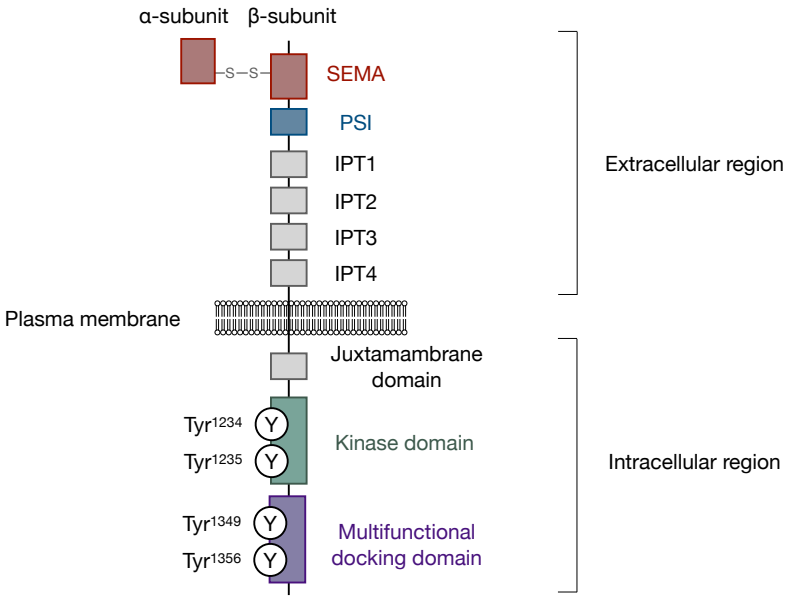


Figure 1.3 | Domain structure of a MET protein.

1.2.3. HGF-MET signaling pathway

The auto-phosphorylation of MET proteins activates the downstream MET signaling cascade and regulates a wide variety of cellular responses, such as cell migration, cell proliferation, morphogenesis, cell transformation and protection from apoptosis[6,11,15–19]. The phosphorylation of Tyr1234 and Tyr1235 residues induces subsequent phosphorylation of two additional tyrosine residues located in the multifunctional docking domain: Tyr1349 and Tyr1356 residues, which are involved in the recruitment of several transducers and adaptors[7,20,21] (Figure 1.4).

One of the most important cell responses induced by the HGF-MET signaling is cell proliferation, cell survival and cell migration. It is known that the mutation of loss of function of HGF (HGF knock-out mutation) is a lethal mutation, where organ development and embryogenesis are not completed and embryos with the HGF knock-out mutation die before the birth[22]. The phosphorylated/activated multifunctional docking domain of MET binds to many proteins related to cell signaling containing the Src-homology-2 domain (SH2 domain). The representative example of proteins with SH2 domain are PI3K, the non-receptor tyrosine kinase Src, the growth factor receptor-bound protein 2 (Grb2) and SH2 domain-containing transforming protein (SHC) adaptors, SHP2 (also known as PTPN11; an upstream activator of Src and Ras), phospholipase Cγ1 (PlCγ1)37 and the transcription factor STAT3[6,10,11]. These effector proteins are triggers of cellular responses related to organogenesis and regenerative pathways. Hence, the activation of the MET pathway play an important role in regenerative medicine.

Another important downstream phenomenon of HGF-MET signaling pathway is internalization of HGF-MET complexes into the endosomes[23–25]. The interaction of HGF and MET forms HGF-MET complexes, which are comprised of two activated HGF and two MET proteins, mentioned the above. This complex is actively internalized into the endosome by clathrin-mediated endocytosis pathways. After activation of MET, the HGF-MET complex is packed in a clathrin-coated vesicle and transported into peripheral early

endosomes. This internalization pathway is regulated by protein kinase Cε (PKCε).

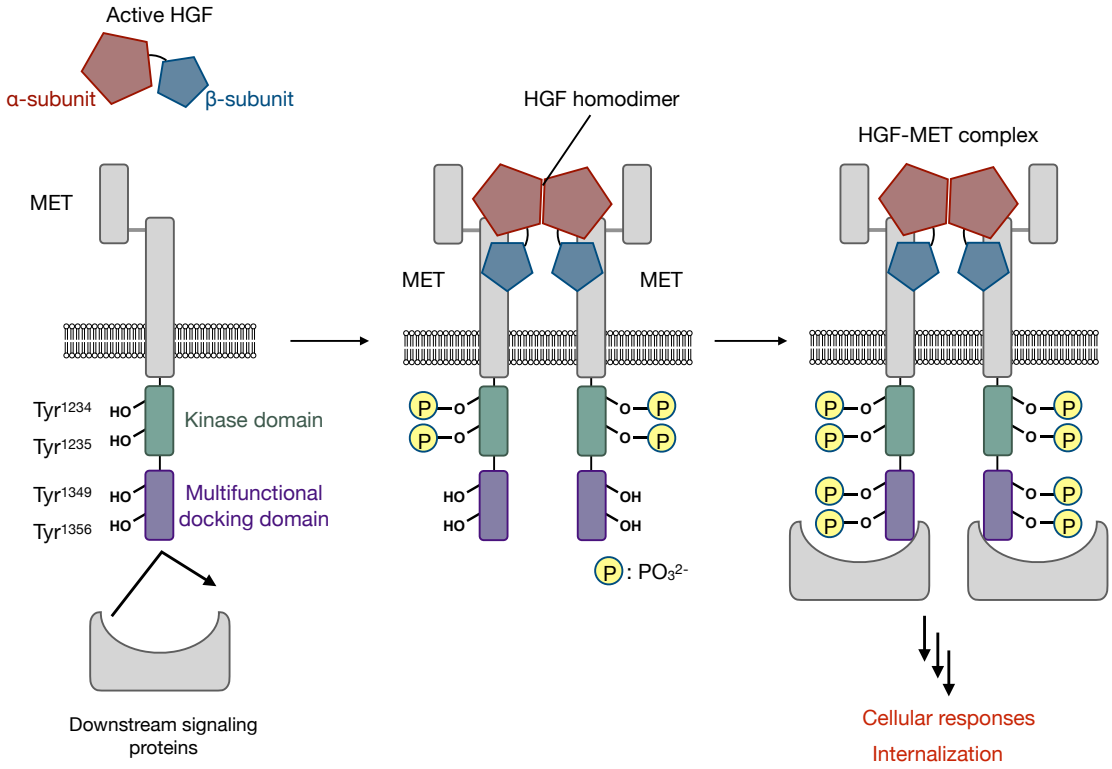


Figure 1.4 | The activation steps of MET induced by HGF homodimer. The dimerization of MET proteins induces auto-phosphorylation of Tyr1234 and Tyr1235 residues of MET and formation of HGF-MET complexes. Subsequent phosphorylation of Tyr1349 and Tyr1356 residues induces cellular responses related to regeneration and internalization of HGF-MET complexes.

1.2.4. Previous MET agonistic molecules

To date, as a promising option for regenerative medicine, several artificial MET agonistic activators have been developed. A representative example is recombinant hepatocyte growth factors (rHGF)[26,27]. A rHGF is produced by cultured mammalian cells that artificially express and secret HGFs by transfection. Although rHGF has been developed as a promising molecule for regenerative medicine, it has some problems in its production and stability. As the first problem, HGFs are glycosylated proteins and both of α-chains and β-chain are highly modified by sugar chains[7]. This feature makes it difficult to reproduce the "same" molecules; HGF proteins with the same sugar modification pattern. As the second problem, secreted HGF requires to be activated by extracellular protease to exhibit its MET dimerization activity as mentioned in the section 1.2. This step requires both of purified secreted HGF proteins (inactive) and purified proteases for activation of HGF, so the activation of HGF highly costs and is still obstacles for utilization as regenerative reagents. As an additional disadvantage, the production of proteins by using cultured mammalian cells requires high cost and technique in general. Both the production and the quality control of rHGF are difficult. Moreover, the stability of (recombinant) HGF is low. HGF itself is a huge protein; an α-chain and β-chain of HGF have ~54 kDa and ~26 kDa of molecular weight, respectively, and whole protein has about 80 kDa of molecular weight. As an experimental data in previous study, incubation of HGF in 37˚C for 3 days results the aggregation of up to 50% of HGF[28]. Low thermal stability of HGF is also a disadvantage as a potential molecule for regenerative medicine.

To overcome these problems, several synthetic MET agonists have been developed. They are designed as homodimers of MET binding motifs to dimerize MET proteins. They are comprised of dimers of chemically synthesized macrocyclic peptides[29] or a dimer of DNA aptamers[30]. Although these synthetic MET activators successfully induced MET dimerization and auto-phosphorylation, no artificial protein with MET agonistic

activity has been developed to the best of our knowledge. Thus, development of artificial MET activators comprised of *de novo* engineered proteins has been still challenging.

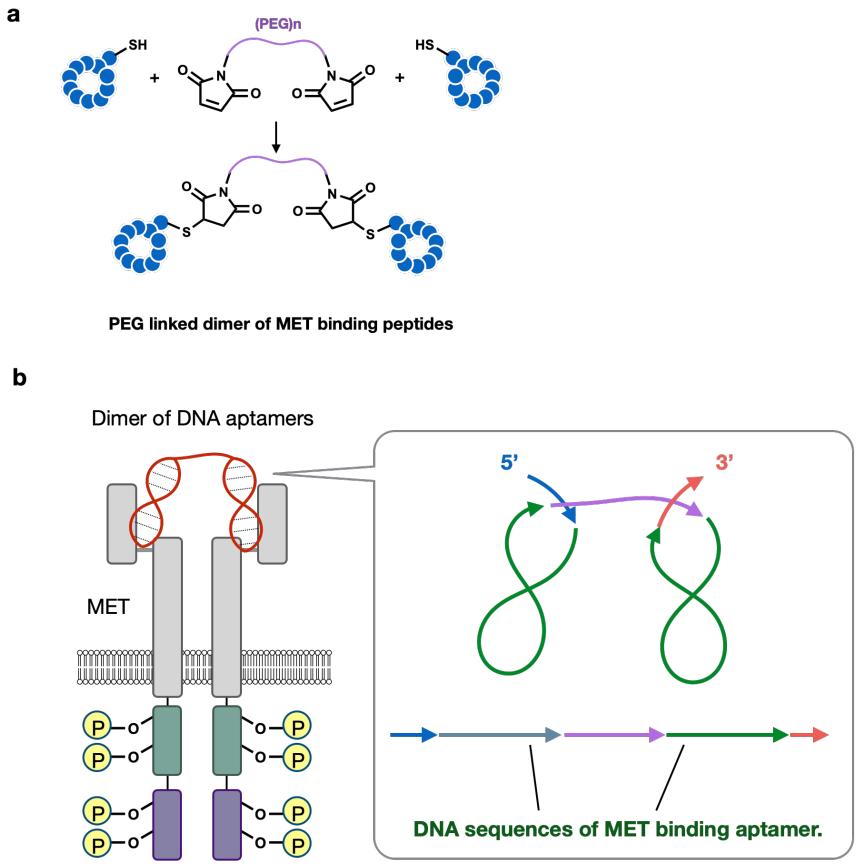


Figure 1.5 | Artificial MET agonists in previous studies. a) A MET agonist based on DNA aptamers. Dimeric DNA aptamer was designed as repeated sequences of MET binding DNA motifs. b) A MET agonist based on macrocyclic peptides. Two macrocyclic peptides were linked via PEG polymers by using the cysteine-maleimide conjugation.

1.3. Self-assembling protein capsules

1.3.1. Self-assembling proteins and bacterial capsids

Capsids are self-assembling proteins that automatically assemble into capsule-like form protein shells. The monomeric protein units, called as capsomeres, interact with each other and assembled into capsids by non-covalent bonds. Originally, though capsids had been discovered from virus and the term capsid had indicated “protein shells of virus”, after discoveries of non-viral capsids, it has indicated the capsule form molecules composed of protein units.

Comparing to other self-assembling molecules, such as micelles[31–33], liposomes[34–36], lipid-nanoparticles[37] and inorganic nanoparticles[38–41], capsids have several advantages for an engineering target[42,43]. As the first advantage, capsids are easy to be edited by means of genetic engineering. They are exclusively comprised of proteins and amino acid sequences are coded on DNA and mRNA. We can customize their chemical properties by changing DNA sequences and try various capsids with various amin acid sequences by established cloning methodologies. As the second advantage, a capsid has the uniform size and is easy to control the size and purify particles with a uniformed size. The third advantage is its biocompatibility. Because capsids are exclusively comprised of proteogenic amino acids, the biodegradability is higher than other artificial, chemically synthesized molecules.

To date, various capsids including viral capsids and non-viral capsids have been discovered and engineered. The representative examples of engineering of viral capsids are based on the cowpea chlorotic mottle virus (CCMV)[42–45], the cowpea mosaic virus (CPMV)[46–49] and the Qβ bacterio phage[50–52] (Figure 1.6). Although the engineering of viral capsids has been well established and is promising to the development of molecules with novel functions, they have disadvantages to be used as a drug molecule. One is the limitation of encapsulation of molecules. Virus itself consists of not only viral capsid but also genome molecules, RNA or DNA inside a capsid. In principle, we cannot express

viral capsids without RNA/DNA. Moreover, these RNA/DNA molecules are often involved in assembling of capsids and it is difficult to isolate viral capsids themselves, *i.e.*, disassemble capsids that encapsulate RNA/DNA and then reassemble them without RNA/DNA. The second disadvantage is its potential viral toxicity. Although virus that does not infect to human or inactivated virus are used for the development of drug delivering molecules, in principle, they have possibilities to have unexpected mutation and/or hijack human translation systems. Comparing to viral capsids, non-viral capsids, such as ferritin[53–58] (iron stocking/delivering protein cage), sHSP[59,60] (small heat shock protein) (Figure 1.7) and bacterial capsids[61–70], have several advantages to be utilized for a drug delivering or drug-leading compartment. They have hollow body and no potential to have any viral toxicity.

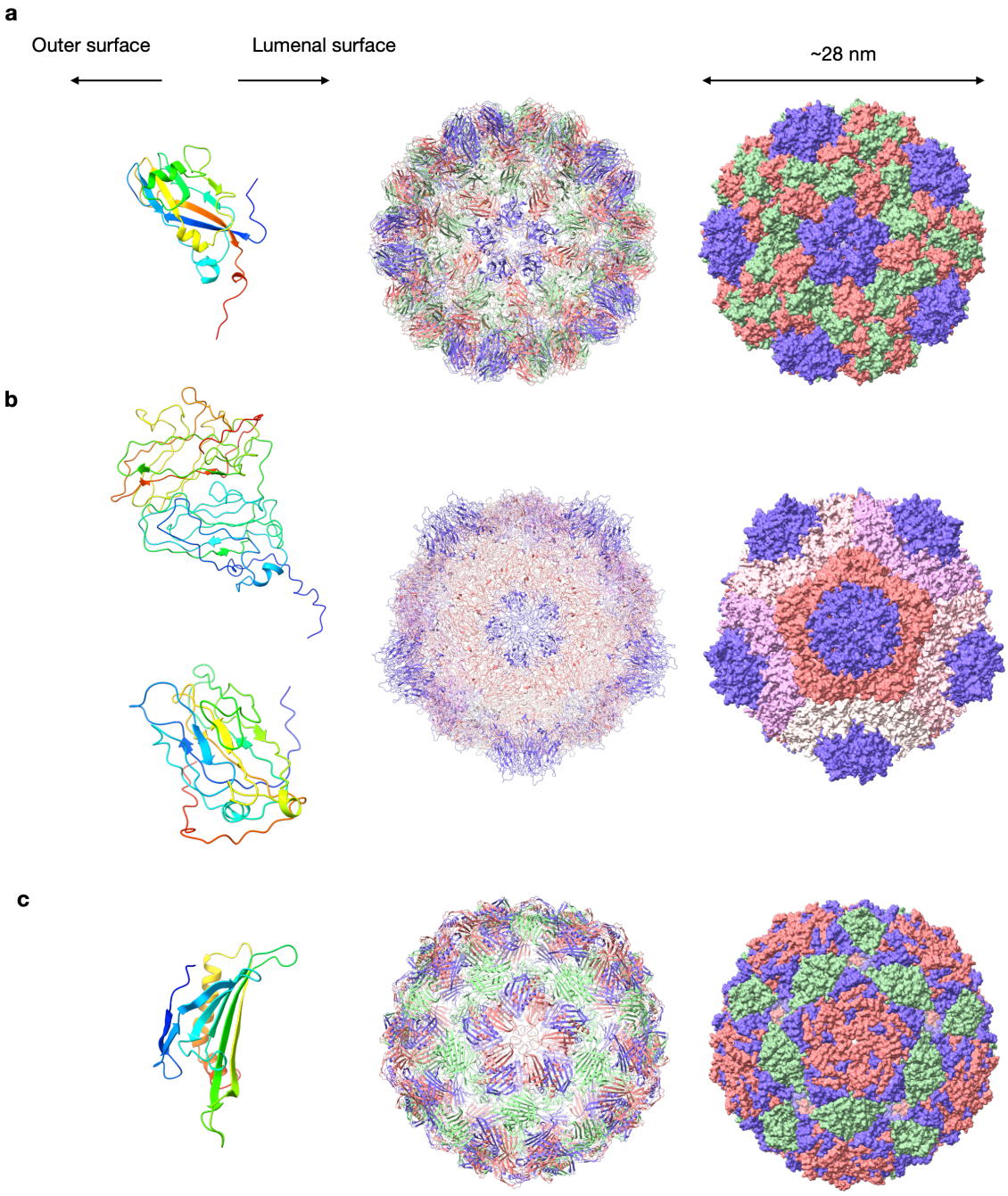


Figure 1.6 | Crystal structure of a) a cowpea chlorotic mottle virus (CCMV) capsid, b) a cowpea mosaic virus (CPMV) capsid and c) a Qβ bacterio phage capsid. Left: crystal structure of capsomeres; the outer surface is placed to the left side. The sequences are colored by a spectrum from blue (N-terminus) to red (C-terminus).

Middle: cartoon model of capsids. Right: Space filling model of capsids.

CCMV: 28 nm icosahedral capsid (T=3) comprised of 180 identical capsomeres (20 kDa). PDB ID: 1CWP

CPMV: 28 nm capsid comprised of 60 large capsomeres (upper, 41 kDa) and 60 small capsomeres (lower, 24 kDa). The large capsomeres are colored purple in the cartoon model and space filling model. PDB ID: 2BFU

Qβ bacterio phage: 28 nm icosahedral capsid (T=3) comprised of 180 identical capsomeres (14 kDa). PDB ID: 1QBE

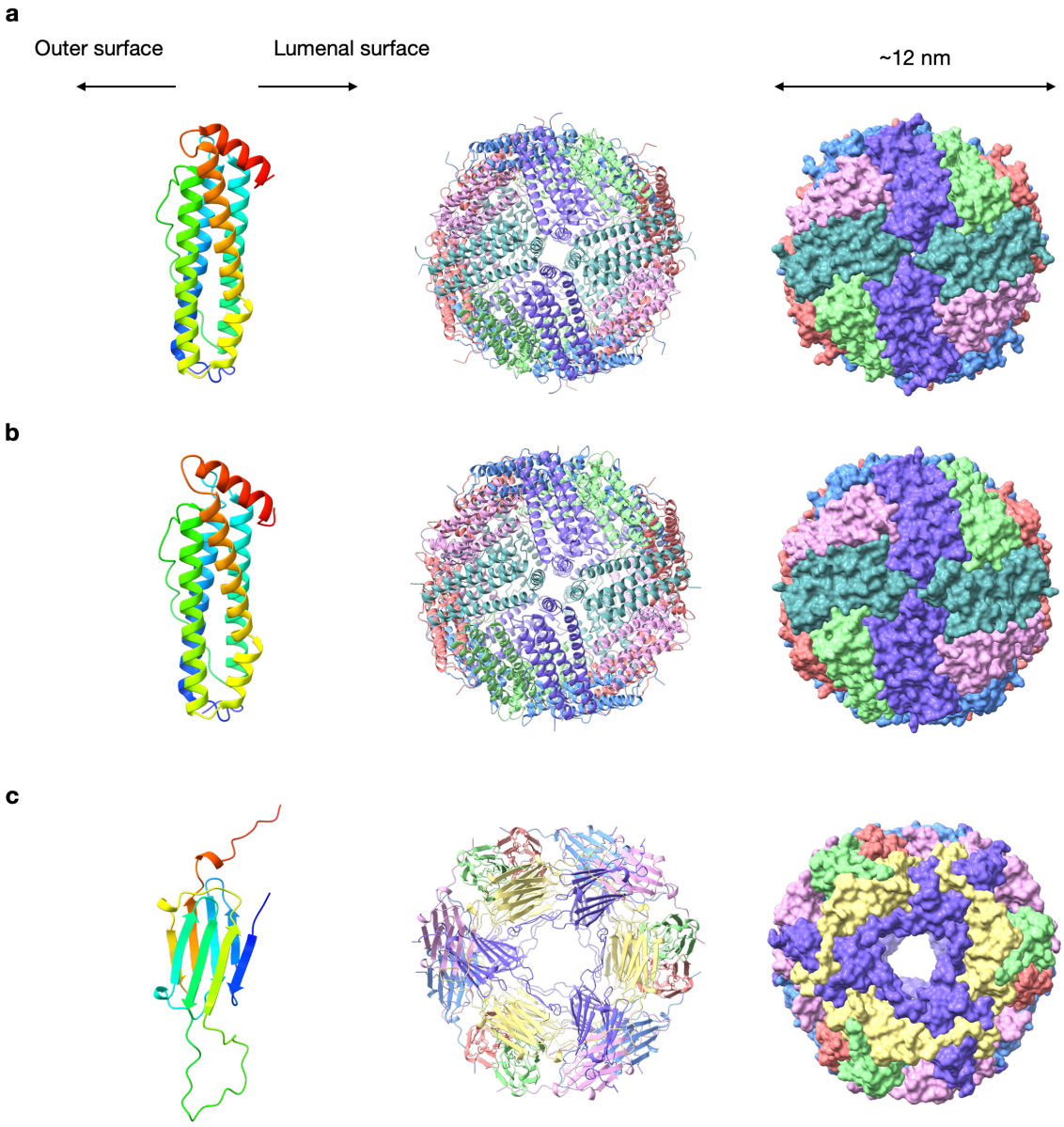


Figure 1.7 | Crystal structure of a human ferritin (a: heavy chain, b: light chain) and c) a small heat shock protein (sHSP) from *Methanococcus jannaschii*.

Left: crystal structure of capsomeres; the outer surface is placed to the left side. The sequences are colored by a spectrum from blue (N-terminus) to red (C-terminus).

Middle: cartoon model of capsids. Right: Space filling model of capsids.

Ferritin: 12 nm capsid comprised of 24 identical capsomeres (heavy chain: 21

kDa, light chain: 19 kDa). PDB ID: 5N27 (heavy chain), 2FFX (light chain)

sHSP: 12 nm capsid comprised of 24 identical capsomeres (16.5 kDa). PDB ID: 1SHS

1.3.2. *Aquifex aeolicus* lumazine synthase (AaLS) capsid

Aquifex aeolicus Lumazine Synthase (AaLS) is a promising bacterial capsid as a scaffold of artificial MET activators. It automatically self-assembles into capsids with a uniformed size composed of 60 identical subunits[71] (Figure 1.8). Because it is isolated from hyperthermophilic bacteria, it has thermal stability with a melting temperature of 119.9°C[71]. It also has a hollow body, which leads to further applications such as encapsulation of guest molecules and drug delivery[68,71]. In addition, previous studies succeeded customizing both of outside and inside of AaLS capsid to display targeting peptides or Fc-binging motifs on its surface[72,73] and to encapsulate various guest molecules[64,69,70,74,75] respectively (Figure 1.9). These studies indicate that the AaLS capsid is a robust scaffold for various genetic engineering and a promising candidate for displaying multiple MET binding motifs on its surface by means of genetic engineering. Interestingly, AaLS capsids engineered to encapsulate positive charged molecules, named as AaLS-Neg[64] and AaLS-Neg-13[76], have different morphology from AaLS wild type capsids[77]. In contrast that an AaLS wild type capsid is composed of 60 capsomeres, an AaLS-Neg capsid and AaLS-Neg-13 capsid are comprised of 180 and 360 capsomeres, respectively (Figure 1.10). These examples indicate that the size and morphology of AaLS capsids can be also customized.

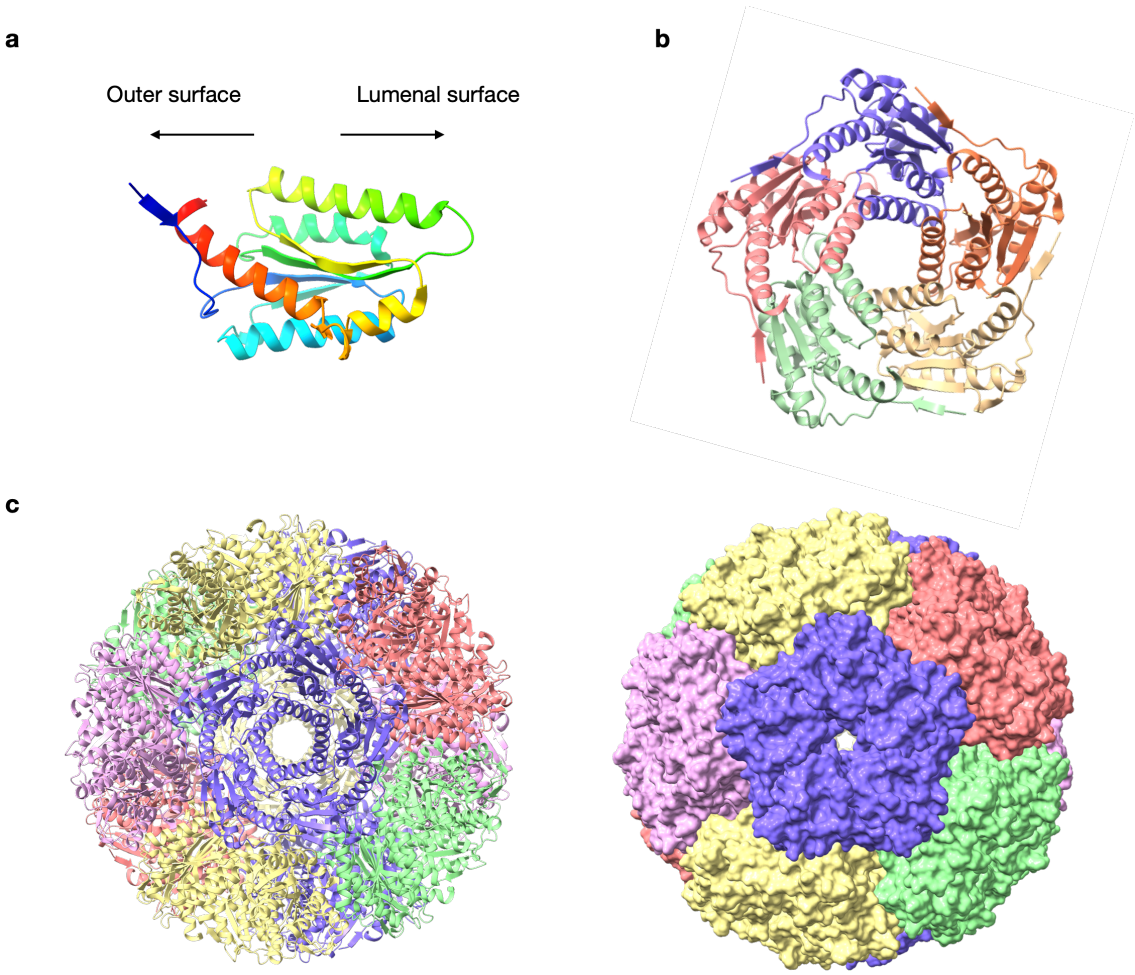


Figure 1.8 | Crystal structure of AaLS wild type. a) AaLS capsomere colored by a spectrum from blue (N-terminus) to red (C-terminus). b) Cartoon model of a pentamer of AaLS capsomeres. Capsomeres are separated by color. c) Cartoon model and space filling model of the AaLS capsid (60 mer). Pentamers are separated by color. PDB ID: 1HQK

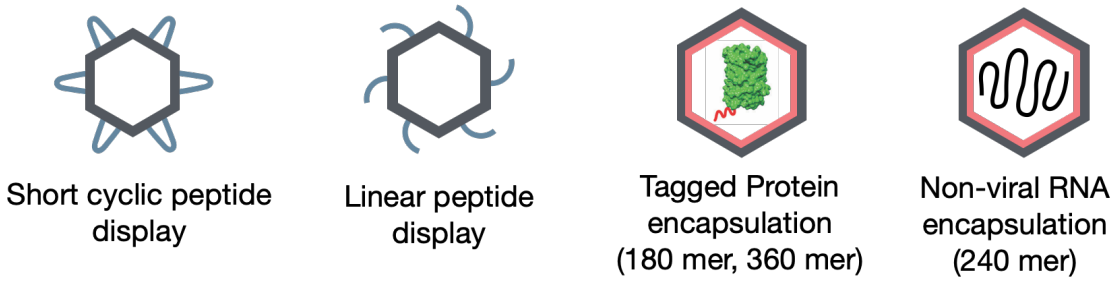


Figure 1.9 | Previous engineering of the AaLS capsid.

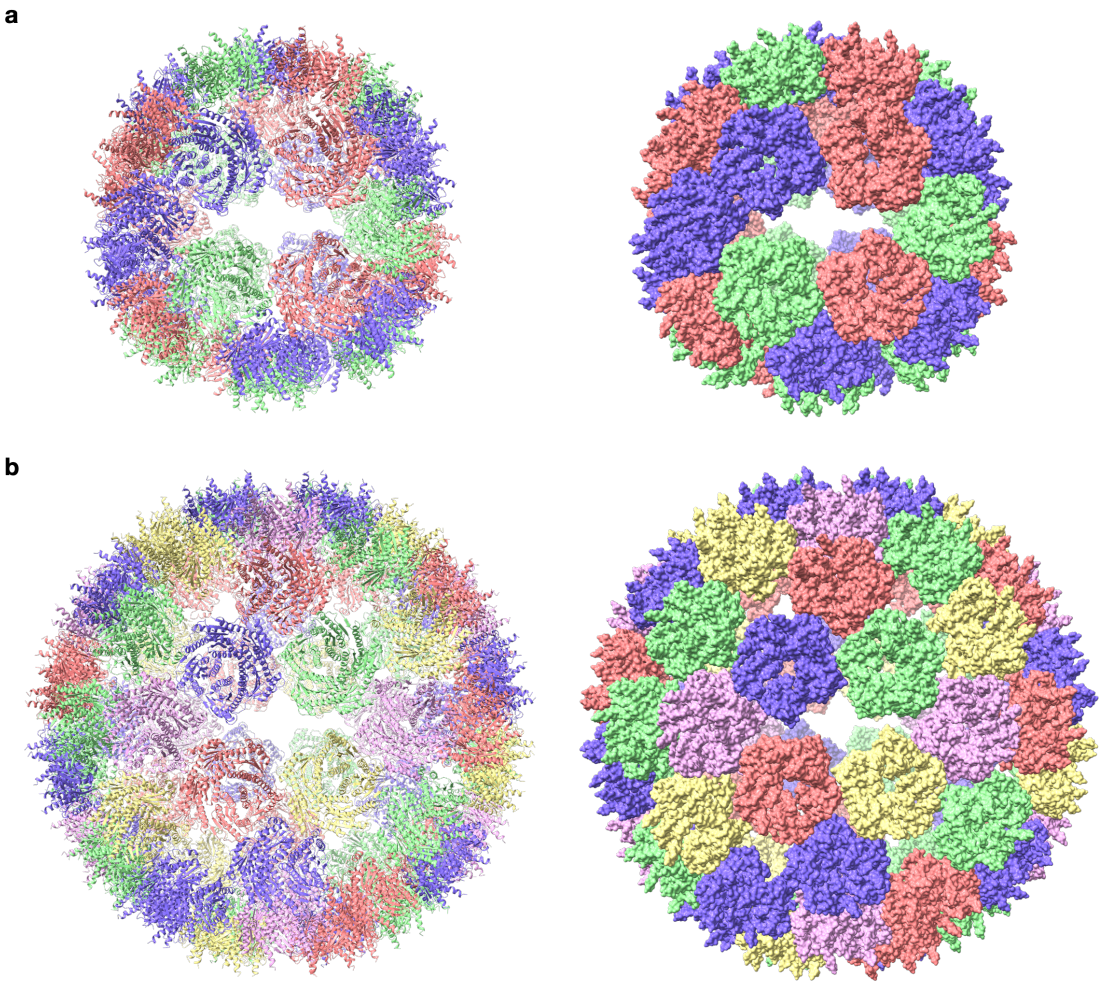


Figure 1.10 | Cartoon model and space filling model of engineered AaLS capsids that have different morphology. a) AaLS-Neg capsid (180 mer). PDB ID: 5MQ3 b) AaLS-Neg-13 capsid (360 mer). PDB ID: 5MQ7

1.4. Key technologies utilized in this study

1.4.1. RaPID (Random nonstandard peptides integrated discovery) system

The RaPID system is a methodology for the screening of peptides that bind to various proteins of interest. This system enables us to discover *de novo* peptides containing non-proteogenic amino acids that bind to any immobilized proteins of interest from libraries with more than 10^{12} of diversity. In the previous studies, many peptides with non-proteogenic amino acids that strongly bind to target proteins such as transporters[78,79], receptors[29,80], a histone deacetylase[81], a histone demethylase[82], a ubiquitin ligase[83], a kinase[84], a mutase[85] and a glycosidase[86] have been discovered. As a representative study, thioether cyclized peptide library was developed and utilized for the discovery of strong protein binders[87]. The peptides in these libraries were initiated by *N*-chloroacetyl-D/L-tyrosine and terminated by cysteine, which spontaneously cyclize via a thioether linkage (Figure 1.11).

The X-ray structural data of various RaPID macrocycles co-crystallized with the target proteins[78,79] have revealed that they can form a wide range of tertiary structures that interact with not only specific binding pockets but also shallow surface of proteins via the combination of hydrogen bonding and hydrophobic interactions (Figure 1.12). Importantly, the macrocycles could also make such interaction networks within the macrocyclic scaffolds and spontaneously fold into the active conformation by themselves. They form intricated hydrogen bonds within their backbone chain and some of them form even the structure where a water molecule seems to be chelated.

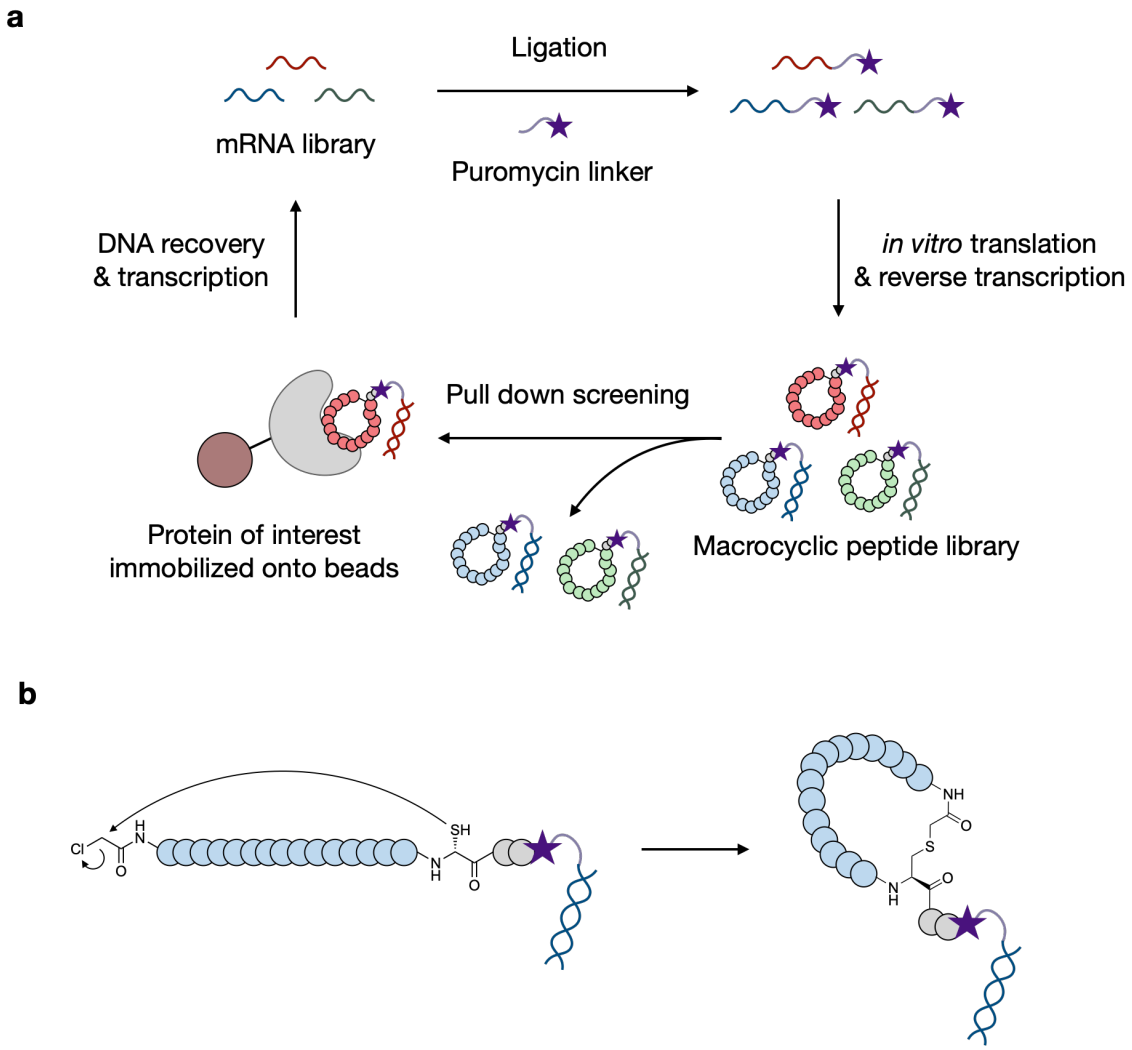


Figure 1.11 | a) Schematic illustration of the RaPID system. b) Spontaneous cyclization between *N*-chloroacetyl group and downstream cysteine thiol in macrocyclic peptide libraries.

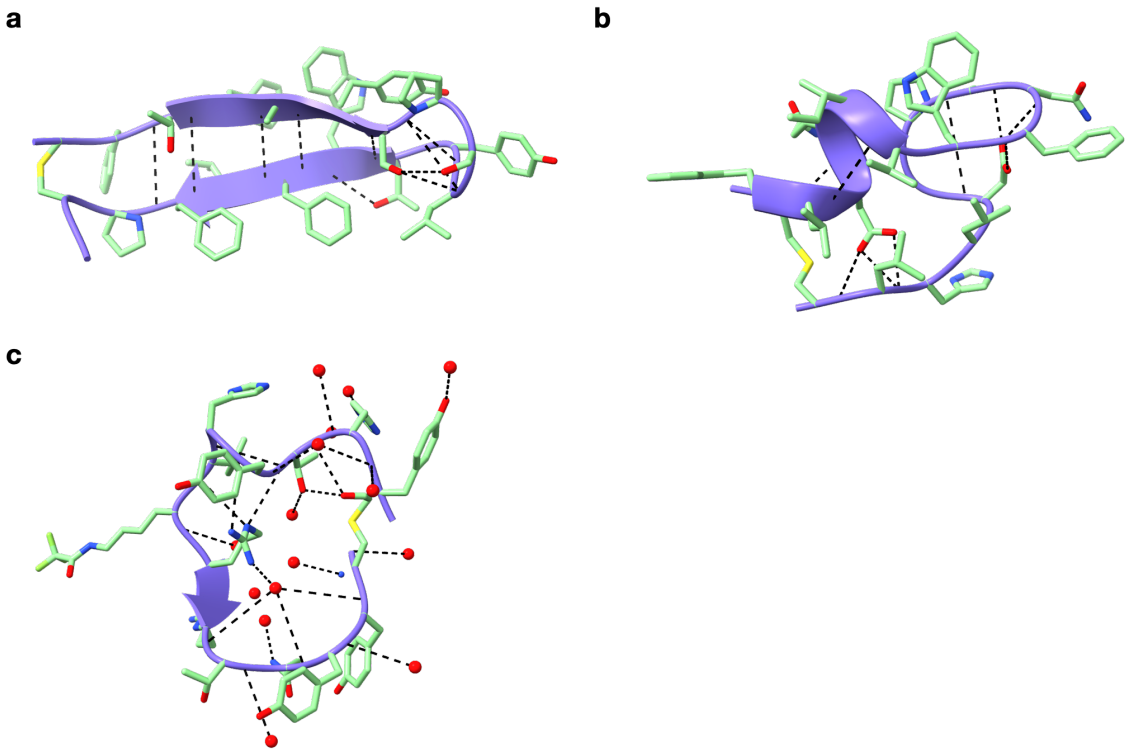


Figure 1.12 | Examples of peptides discovered by the RaPID system.

a) MATE binding macrocyclic peptide. It forms a β-hairpin like structure by intramolecular hydrogen bonds. PDB ID: 3WBN

b) CmABCB1 binding macrocyclic peptide. It forms an α-helix like structure and a intricated hydrogen bond network. PDB ID: 3WMG

c) SIRT2 binding macrocyclic peptide. It forms a structure where a water molecule seems to be chelated. PDB ID: 4L3O

Black dotted line: hydrogen bond, Red: Oxygen atom, Blue: Nitrogen atom, Yellow: Sulfur atom

1.4.2. Lasso-grafting technology

Macrocyclic peptides are an attractive modality as protein binders due to wider molecular surface area than that of small molecules[88]. In addition to the compatibility to the RaPID system, macrocyclic peptides also have the compatibility to the protein engineering as an application. As mentioned in the section 1.4.3, thioether-closing RaPID macrocycles seems to spontaneously form their certain conformation necessary for binding to the target proteins by themselves and the thioether linkages seem not to be involved in the formation. Based on this notion, a novel protein engineering concept, lasso-grafting technology, has been developed[89]. In this method, the thioether bond of a macrocycle of interest discovered by RaPID system was disconnect and removed and the resulted core motif of peptide, *i.e.*, pharmacophore, is reconnected to a part of protein, *i.e.*, scaffold (Figure 1.13). In the previous study[89], a Fc domain of antibody, an alkaline phosphatase, human serum albumin, an adeno-associated virus capsid and others have been used as a scaffold and a loop of scaffold proteins was replaced by the peptidic pharmacophores.

The lasso-grafting technology offers us to separate the discovery process of de novo pharmacophore from the protein engineering, avoiding the protein folding issue potentially occurring in the library construction with random sequence insertions. Consequently, it reduces the efforts and increases the success rate of devising a protein with a desired binding function to a target of interest.

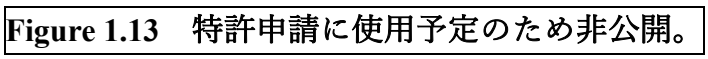


Figure 1.13 | Schematic illustration of the lasso-grafting technology.

1.5. Purpose of this study

The lasso-grafting technology can add new binding properties with maintaining the original function of a scaffold protein. In the first report[89], it is used for development of multi-specific binding proteins and targeting of AAV (adeno-associated viral) vector.

本部分については、特許申請に使用する予定のため非公開。

In addition to the work on the lasso-grafting against AaLS capsids, I also worked on the construction of *in vitro* selection system toward discovery of ADH (alcohol dehydrogenase)-like peptidic enzymes and rationalization of the membrane permeability of macrocyclic peptides. In chapter 4, I constructed an *in vitro* selection system that detects aldehyde residues contained in library peptides and the chemical properties of peptides obtained by the constructed *in vitro* selection system was assayed by various methods. In chapter 5, the molecular dynamics calculation of two peptides has been performed and the results were analyzed to reveal the difference of membrane permeability between two peptides.

Chapter2

Development of MET agonists based on AaLS capsids

第 2 章については、特許申請に使用する予定のため非公開。

Chapter3

Internalization of MET agonistic capsids into endosomes and further engineering toward drug delivering

第 3 章については、特許申請に使用する予定のため非公開。

Chapter4

In vitro selection of an alcohol dehydrogenase peptidic enzyme

2. *In vitro* selection of an alcohol dehydrogenase peptidic enzyme

2.1. Introduction

Alcohol dehydrogenation is one of the most important chemical reactions for living organisms. A protein enzyme, called alcohol dehydrogenase (ADH), catalyzes this reaction using zinc ion and oxidized form of nicotinamide adenine dinucleotide (NAD^+) as co-factors[95,96]. In order to achieve precise recognition of the substrates/co-factors and high catalytic activity, the ADH has to form the sophisticated active site, where the side-chain functional groups are displayed in appropriate orientations and intermolecular intricated hydrogen bonds are formed to capture and fix substrates/co-factors[97–100]. For these reasons, the size of ADH needs to be significantly large, generally composed of over 400 amino acids. However, such huge and sophisticated proteins could not appear straightaway in the process of evolution[101–103]. More likely, a smaller primitive ADH with a weak activity first emerged and have evolved to the modern ADHs. Therefore, I aimed to discover peptide sequences capable of catalyzing alcohol dehydrogenation.

In the active site of such small peptidic ADHs, necessary side-chain functional groups should be displayed in appropriate orientations, which is important for binding and capturing NAD^+ and efficiently using it for oxidation of the target alcohol. I conceived that macrocyclic peptides, which generally adopt rigid and constrained conformations, could be promising candidates for peptidic ADHs. To obtain macrocyclic peptides with the ADH activity, I aimed to construct an *in vitro* selection system based on the RaPID system in this chapter.

As mentioned in the section 1.4.1, the RaPID system is one of the most powerful method for the screening of macrocyclic peptides. However, the RaPID system can discover only protein binders and cannot screen peptide library based on enzymatic activity. To achieve I constructed a macrocyclic peptide library that contains a substrate alcohol residue and macrocyclic peptide that can be an ADH peptidic enzyme and an *in vitro* selection system that can pick up an aldehyde residue resulted from alcohol

dehydrogenation catalyzed by active species.

2.2. Results and discussion

2.2.1. Design of macrocyclic peptide library containing a substrate alcohol residue

An mRNA-encoded peptide library, in which a macrocyclic peptide region with randomized sequences (enzyme candidates) is connected to benzyl alcohol moiety (substrate) via a tetra-glycine linker was designed (Figure 4.1). To obtain this library displayed on mRNA templates, two artificial residues, an Nγ-(chloroacetyl)-α,γ-diaminobutyric acid (Cab) and a phenylalanine derivative bearing benzyl alcohol moiety (BzAlc-Phe), were utilized and designated at the upstream position of the randomized region and the N-terminal position, respectively, by means of genetic code reprogramming technology. The activated derivatives of artificial amino acids (Fig 4.2) have been utilized for the aminoacylation onto the corresponding tRNA, which is required the genetic code reprogramming method. The incorporated Cab residue can undergo posttranslational reaction with a downstream cysteine (Cys) to form a macrocyclic scaffold.

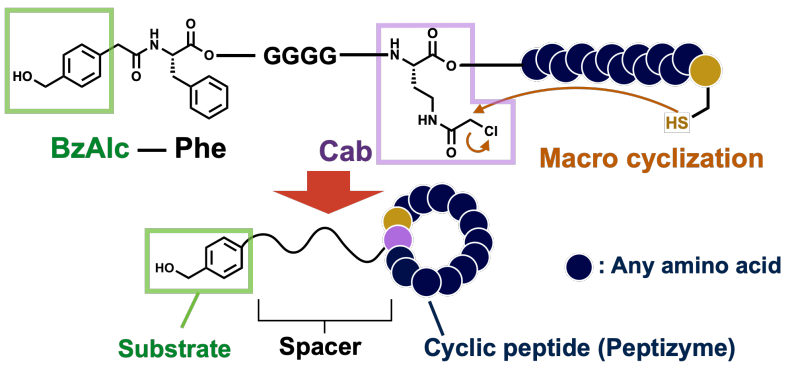


Figure 4.1 | The design of macrocyclic peptide library containing a substrate alcohol residue. The spontaneous cyclization produces macrocyclic peptides with random amino acid sequences.

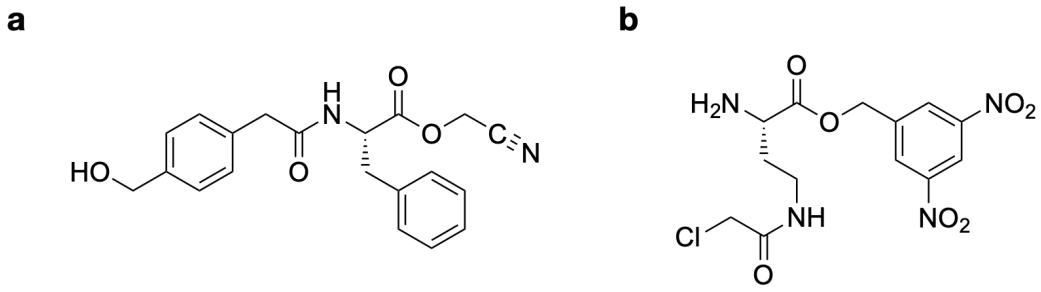


Figure 4.2 | The structures of the activated derivatives of artificial amino acids utilized for the library construction. a) BzAlc-Phe-CME b) Cab-DBE. CME: Cyanomethyl ester, DBE: 2,4-dinitrobenzyle ester. The CME activated amino acids were used with eFx flexizyme and the DBE activated amino acids were used with dFx flexizyme in the step of aminoacylation onto the corresponding tRNA for genetic code reprogramming.

To generate the above designed peptide library, two mRNA libraries (Fig. 4.3) were designed and utilized for further experiments. The short library and long library have random amino acid sequence regions comprised of 8-16 amino acids and 18-26 amino acids, respectively. In general, the translation efficiency of longer peptides generated from longer mRNA is lower than shorter peptides. In this study, in order to normalize the difference of translation efficiency derived from the difference of library lengths, I used two libraries with different lengths. Two libraries also have different linker sequence between the downstream cysteine and the cognate mRNA to prevent cross contamination.

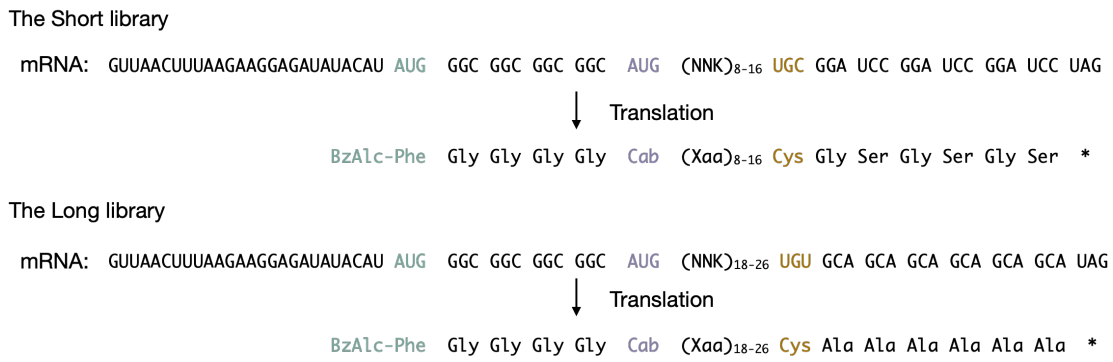


Figure 4.3 | The sequences of mRNA libraries. The initiating AUG codon is translated into BzAlc-Phe (colored green) and the elongating AUG codon is translated into Cab (colored purple) by using genetic reprogramming technology. The codon of the consensus downstream cysteine is colored yellow.

Xaa: Randomized amino acid, * : stop codon.

2.2.2. Construction of *in vitro* selection system based on ADH enzymatic activity

Next, an *in vitro* selection system based on ADH enzymatic activity was designed (Figure 4.4). To select peptides with ADH activity, I aimed to construct an *in vitro* selection system that detects a product of chemical reaction catalyzed by ADH, *i.e.,* an aldehyde moiety. To collect macrocyclic peptides with aldehyde moieties, I conjugated a biotin and aldehyde moiety of a peptide by using biotin-hydrazide and pulled down by streptavidin beads.

The optimized procedures of *in vitro* selection are following; first, a peptide-mRNA-cDNA conjugate library where macrocyclic peptides with the above design and their cognate mRNA-cDNA was linked by puromycin linker based on mRNA display technology was generated by *in vitro* transcription, puromycin linker ligation, translation and reverse transcription reactions. The resulted library was incubated in the presence of NAD^+, zinc and magnesium ions to execute alcohol dehydrogenation reaction. In case that the macrocyclic peptide region exhibits ADH-like activity, BzAlc-Phe could be oxidized to the corresponding aldehyde; BzAld-Phe. To remove peptide sequences that bind to streptavidin or the surface of magnetic beads, the resulted library molecules were incubated with streptavidin-immobilized magnetic beads. After the removal of bead binders, the aldehyde group of active species was specifically biotinylated by biotin-hydrazide. After the removal of un-reacted residual biotin-hydrazide molecules by twice acetone precipitations, the resulted molecules were incubated with streptavidin-immobilized magnetic beads so that the active fraction of the library members could be pulled down. The beads were washed with the EK-T buffer containing 6 M Urea in order to prevent non-specific binding except for avidin-biotin complex formation and the resulted beads were incubated at 95°C to elute the cDNA of pulled down molecules. After PCR amplification, the next peptide-mRNA-cDNA conjugate library was re-constructed by repeating the same *in vitro* transcription, puromycin linker ligation, translation and reverse transcription reactions. I defined this series of manipulations as a round and firstly

aimed to validate how one around can give a selection pressure against the libraries and integrate the macrocyclic peptides with aldehyde groups in the libraries.

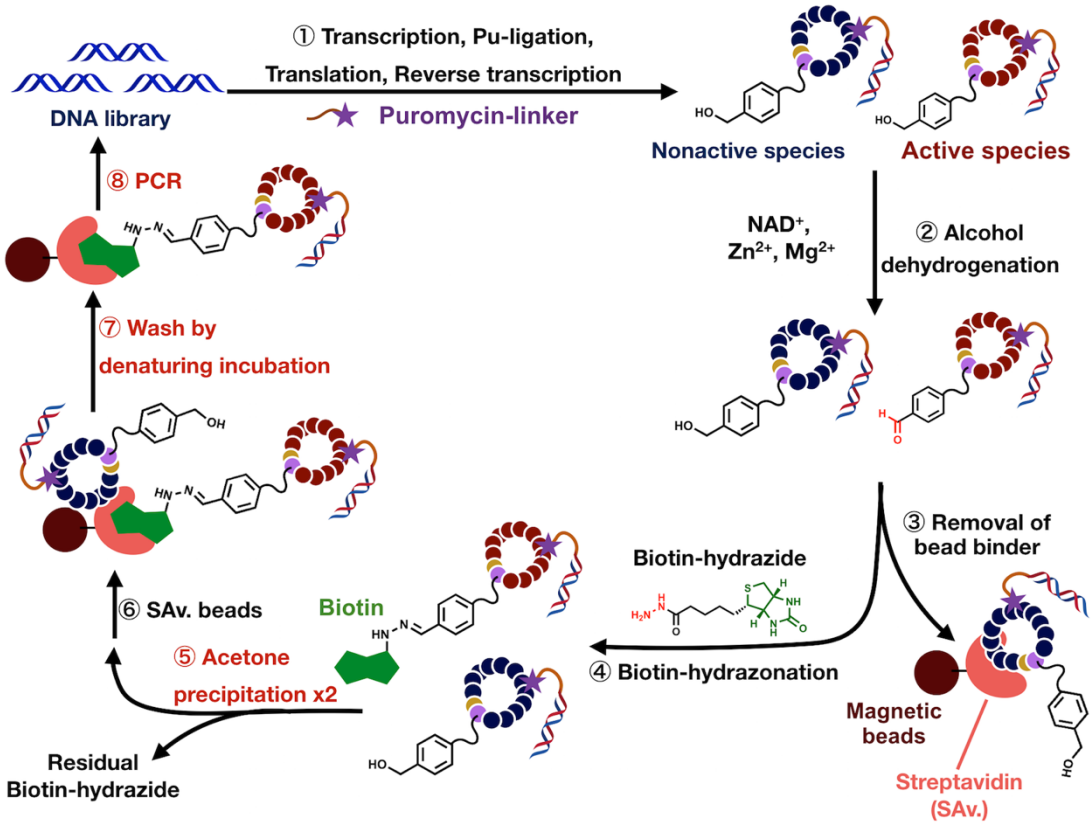


Figure 4.4 | The design of an *in vitro* selection system based on ADH enzymatic activity.

To validate the constructed *in vitro* selection system collects aldehyde-containing macrocyclic peptides selectively, I measured and compared the recovery rates of macrocyclic peptides containing benzyl aldehyde and macrocyclic peptides containing benzyl alcohol. The macrocyclic peptides containing benzyl aldehyde was prepared by using the BzAld-Phe-CME instead of the BzAlc-Phe-CME (Fig4.5). The recovery rate is calculated by comparing the number of cDNA contained in the library solution just after the reverse transcription and the number of cDNA eluted from the pulled down fraction. The number of cDNA molecules were measured by quantitative PCR with the same primer sets used for the reconstruction of the DNA library, *i.e.,* the set of T7g10M.F36 and CGS3ggatcc.Sgca.R28 for the short library and the set of T7g10M.F42 and CA6tgc6.Saca.R25 for the long library. Because it is expected very rare event that an amino acid sequence from random sequences has ADH enzymatic activity, the recovery rate of macrocyclic peptides initiated by benzyl alcohol expected to be quite low. If the constructed selection system has enough selectivity to be utilized for the discovery of peptidic enzymes, only the recovery rates of macrocyclic peptides initiated by benzyl aldehyde expect to be high. The results are shown in the Table 4.1. As the result, the 13.9% of benzyl aldehyde containing peptides were collected by the constructed selection system, whereas only 0.02% of peptides initiated by benzyl alcohol were recovered.

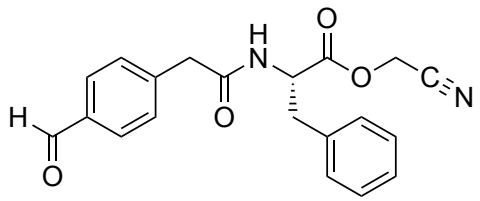


Figure 4.5 | The structures of the BzAld-Phe-CME utilized for the validation of constructed *in vitro* selection system.

Initaiator	The number of total cDNA in the library	The number of total cDNA in the pulled down fraction	Recovery rate [%]
BzAlc-Phe	2.89×10^{12}	6.61×10^{8}	2.28×10^{-2}
BzAld-Phe	1.83×10^{12}	2.55×10^{11}	1.39×10^{1}

Table 4.1 | The number of cDNA molecules and recovery rate in the validation. The numbers of cDNA were calculated from the results of quantitative PCR and the recovery rate was calculated by dividing the number of pulled down cDNA by the number of synthesized cDNA.

2.2.3. *In vitro* selection system based on ADH enzymatic activity

After repeating the *in vitro* selection round for six times, the cDNA of the obtained library were sequenced by a next generation sequencer to identify the sequences of macrocyclic peptides that might have ADH-like activity. Among the identified macrocyclic peptides, three sequences abundant in the selected library were used for further validation. All of the three peptides underwent efficient pulled down by streptavidin only after incubation in the presence of co-factors followed by biotin-hydrazide. This suggested that the benzyl alcohol moiety on the peptides might be converted to the corresponding aldehyde during the incubation with co-factors (Fig. 4.6).

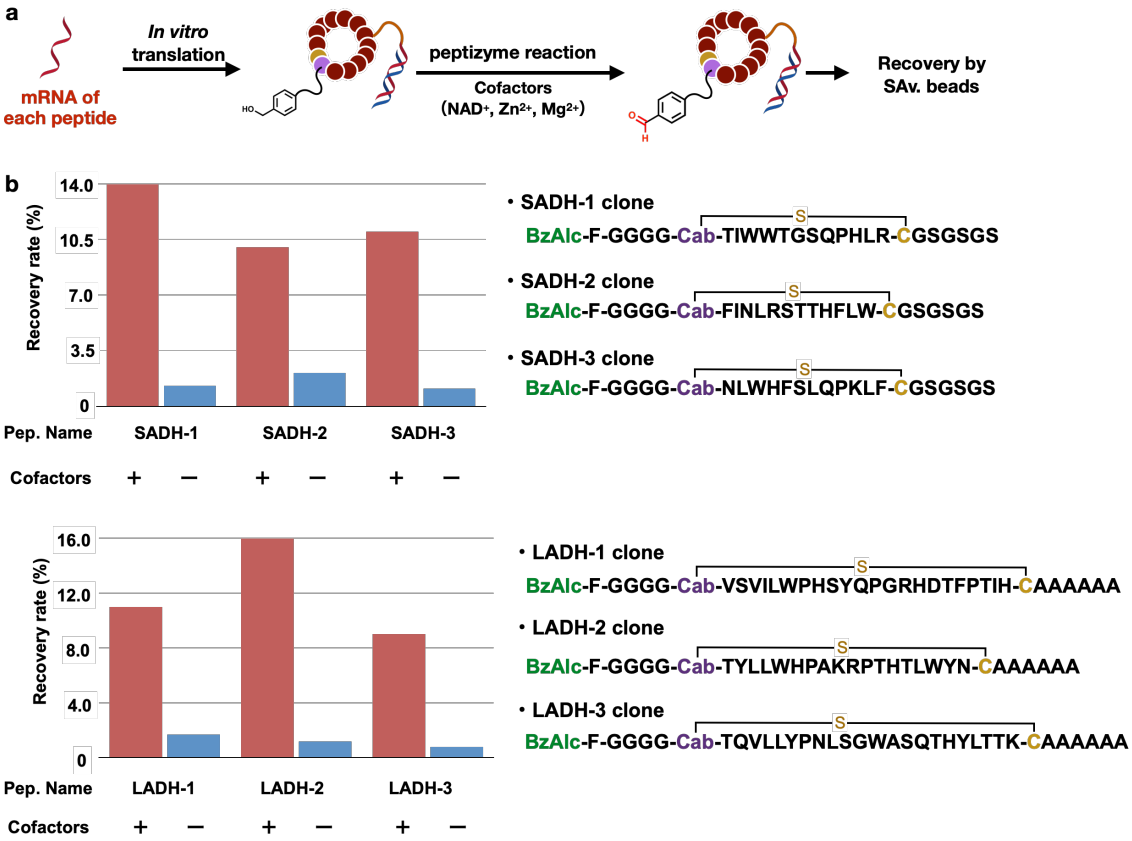


Figure 4.6 | a) Schematic procedure of pull down assay. b) Comparison of the pulling down efficiency of macrocyclic peptides with small (the upper column) and large (the lower column) ring sizes. The sequences of six peptides are also described.

As a further validation of the ADH-like enzymatic activity of obtained peptides, I aimed to trace the mass change originated from ADH reaction and biotin-hydrazonation (Figure 4.7). Each peptide was independently translated by *in vitro* translation system and incubated in the presence of NAD^+, zinc and magnesium ions to execute alcohol dehydrogenation reaction. The mass of resulted molecules was measured by MALDI-TOF-MS. As the results, no macrocyclic peptides gave a mass peak correspond to the dehydrogenated and biotinylated product.

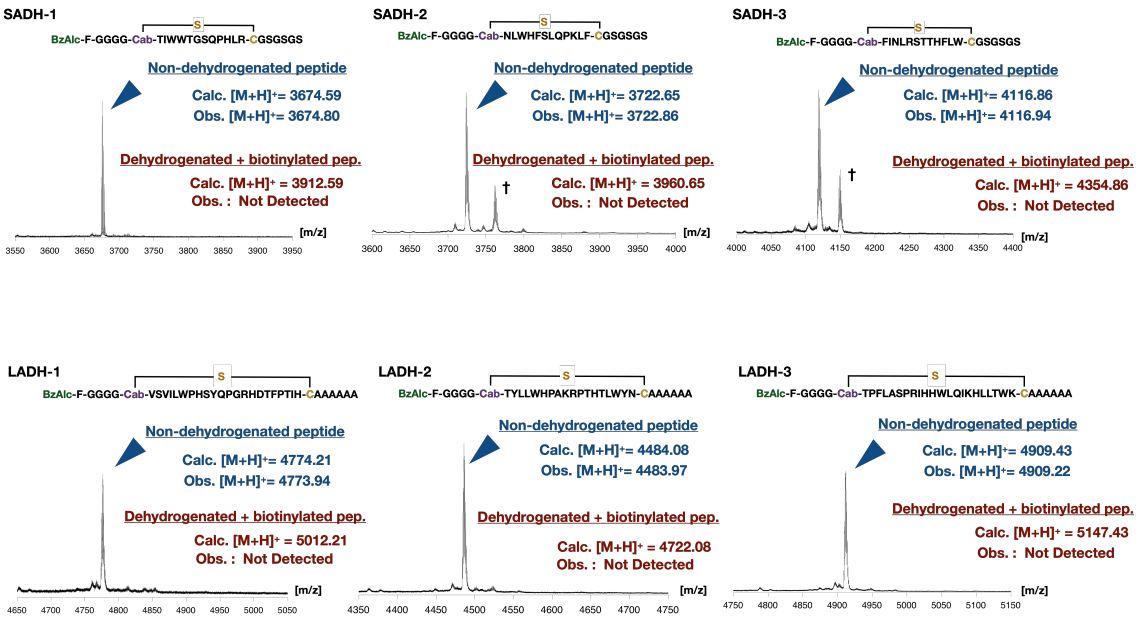


Figure 4.7 | The MALDI-TOF-MS spectra of macrocyclic peptides after the incubation with co-factors and biotin-hydrazonation. †: potassium adducts.

2.3. Conclusion

In this chapter, I aimed the construction of an *in vitro* selection system based on ADH-like enzymatic activity and discovery of macrocyclic peptides with ADH-like activity. However, the ADH-like activity of obtained peptides was not observed by means of mass spectrometry.

Although the aimed active species has not discovered, the constructed selection system successfully collected macrocyclic peptides with benzyl aldehyde moieties, whereas peptides without an aldehyde residue was not recovered. The constructed system and strategy used in this study can potentially be applicable for the development of an *in vitro* selection system based on another enzymatic activity.

2.4. Material and method

Equipments and materials

MALDI-TOF-MS ultrafleXtreme-SG2 was purchased from Bruker Daltonics and utilized for mass analysis. Thermalcycler for PCR was TC-312 or TC-3000 purchased from TECHNE, PC815 or PC818 purchased from ASTEC, or iCycler purchased from BIO-RAD. Light cycler and Light cycler 96 were purchased from Roche. All oligonucleotides were purchased from Eurofins Genomics Inc. and GeneDesign, Inc. ces was analyzed by Fasmac Inc. 1H NMR spectra were measured by Bruker AV300.. All chemical reagents were purchased from Watanabe Chemical Industry, Nacalai Tasque, Tokyo Chemical Industry, Wako Fuji-film, Sigama-Aldrich Japan, and Roche and Thermo Scientific. They were used without further purification. MiSeq for next generation sequencing was purchased from illumina.

Preparation of flexizyme (eFx and dFx)

1. Extension reaction

Forward primer and reverse primer (eFx: Fx5'.F36 and eFx.R45, dFx: Fx5'.F36 and dFx.R46, 1 μM each) were mixed in the PCR miture (10 mM Tris·HCl (pH 9.0), 50 mM KCl, 2.5 mM $MgCl_2$, 0.25 mM each dNTPs, 0.1% (v/v) Triton X-100, and 45 nM Taq DNA polymerase). Primer extension was conducted in a 10-100 μL reaction mixture by denaturing (95°C for 1 min) followed by 5 cycles of annealing (50°C for 1 min) and extending (72°C for 1 min).

2. Amplification

The extension product was 10-fold diluted with the PCR mixture and primers for amplification (eFx: T7ex5.F22 and eFx.R18, dFx: T7ex.F22 and dFx.R19, 1 μM each) were added. Amplification by PCR was conducted in up to 200 μL reaction mix by 12-17

cycles of denaturing (95°C for 40 sec), annealing (50°C for 40 sec), and extending (72°C for 40 sec). Amplification PCR product was confirmed by 3% agarose gel electrophoresis and ethidium bromide staining.

3. Phenol/chloroform/isoamylalcohol extraction

The same volume of phenol/chloroform/isoamylalcohol (25:24:1) was added into the PCR product. The solution was mixed by voltex and centrifuged at 4°C at 13000 rpm for 5 min. The water layer was recovered and the same volume of chloroform/ isoamylalcohol (24:1) was added to the water layer. The solution was mixed by voltex and centrifuged at 4°C at 13000 rpm for 2 min. The water layer was recovered as the PCR product.

4. Ethanol precipitation

1/10 volume of 3 M NaCl and the 2.2 times volume of ethanol were added to the PCR product. The solution was centrifuged at 25°C at 13000 rpm for 15 min. The supernatant was discarded and 70% ethanol was added. After being centrifuged at 25°C at 13000 rpm for 3 min, the supernatant was discarded. The pellet was and dried and dissolved in 1/10 volume of PCR solution of water.

5. In vitro transcription and gel purification

Transcription reaction was conducted by incubating the transcription mixture (40 mM Tris·HCl (pH 8.0), 1 mM spermidine, 0.01% (v/v) Triton X-100, 10 mM DTT, 22.5 mM $MgCl_2$, 3.75 mM each NTPs, 5 mM GMP, 22.5 mM KOH, 10% (v/v) DNA template prepared above, and 120 nM T7 RNA polymerase) at 37°C overnight. The transcription mixture was mixed with RQ1 RNase-Free DNase (1 U/μL, 1μL, Promega), and incubated at 37°C for 60 min. The resultant tRNA transcript was precipitated by isopropanol precipitation and dissolved in water. The flexizyme transcript was purified by 8% denaturing PAGE and ethanol precipitation, and then dissolved in water.

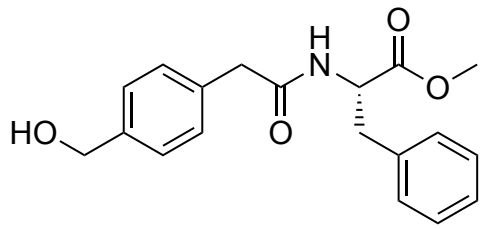
Preparation of tRNA

$tRNA^{Ini}_{cau}$ and $tRNA^{EnAsn}_{cau}$ were prepared by using the same protocol of preparation of flexizymes except for primers transcription mixture. Appropriate forward and reverse primers ($tRNA^{Ini}_{cau}$: Ini1-1G-5'.F49, Ini cat.R44, T7ex5.F22, Ini-3'.R38, and Ini-3'.R20, $tRNA^{EnAsn}_{cau}$: EnAsn-5'.F49, EnAsn AUG.R43, T7ex5.F22, EnAsn-3'.R38 and EnAsn-3'.R20) were used for the preparation of each DNA template of tRNAs. The transcription reaction of tRNAs was conducted by incubating the transcription mixture (40 mM Tris·HCl (pH 8.0), 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) DNA template, and 120 nM T7 RNA polymerase) at 37°C overnight.

Synthesis of non-natural amino acids

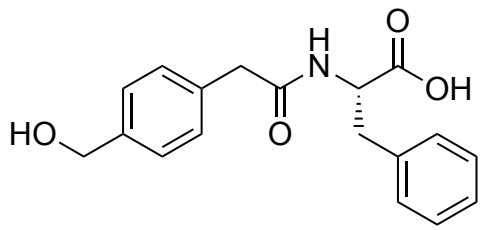
Non-natural amino acids activated with an appropriate ester groups (BzAlc-Phe-CME, BzAld-Phe-CME, Ac-Phe-CME and Cab-DBE) were synthesized by the procedure reported previously[104,105].

Synthesis of (S)-methyl 2-(2-(4-(hydroxymethyl)phenyl)acetamido)-3-phenylpropanoate (1)



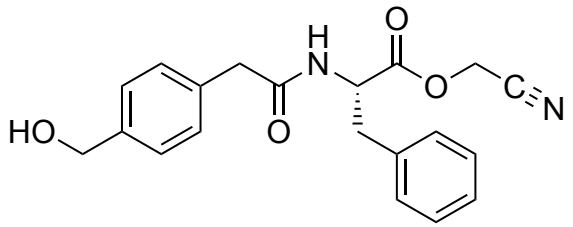
p-(hydroxymethyl)-phenylacetic acid (302.2 mg, 1.82 mmol) was dissolved in 10mL THF. Et_3N (202.4 mg, 2.73 mmol), L-phenylalanine methyl ester hydrochloride (432.2 mg, 2.00 mmol) and DMT-MM (755.9 mg, 2.73 mmol) were added to the solution and stirred for 22 hours. The solution was concentrated and diluted with AcOEt. The organic layer was washed with sat. $NaHCO_3$ aq., 1M HCl, water and brine, then dried over Na_2SO_4. The solvent was removed and flash column chromatography ($CHCl_3$ / AcOEt = 2 / 1) afforded 2 (617mg, 93%). 1H NMR (300 MHz, CDCl3) δ 1.65 (1H, t, J = 6.0 Hz), 3.05 (2H, d, J = 6.0 Hz), 3.55 (2H, s), 3.70 (3H, s), 4.70 (2H, dd, J = 5.4 Hz), 4.86 (1H, td, J = 5.7 Hz, J = 8.1 Hz), 5.80 (1H, br d, 1H), 6.89-6.92 (2H, m), 7.17-7.22 (5H, m), 7.33 (2H, d, J = 8.1 Hz)

Synthesis of (S)-2-(2-(4-(hydroxymethyl)phenyl)acetamido)-3-phenylpropanoic acid (2)



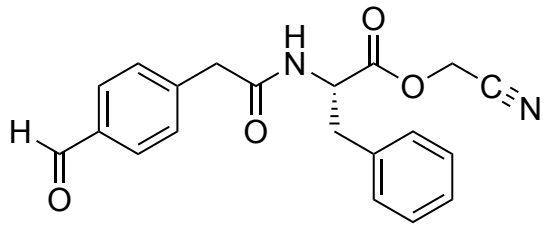
Compound 1 (261.7 mg, 0.80 mmol) was dissolved in 2 mL KOH 1M in MeOH. This solution was stirred for 5 hours, concentrated, dissolved in water, then washed with Et_2O. The aqueous phase was acidified with 1M HCl, and extracted with Et_2O. The organic layer was dried Na_2SO_4, concentrated and crude product 2 (257.6 mg) was obtained. This compound was used in the next step without purification. 1H NMR (300 MHz, MeOD) δ 2.95 (1H, dd, J = 9.0 Hz, J = 14.1 Hz), 3.21 (1H, dd, J = 4.8 Hz, J = 13.8 Hz), 3.42 (2H, s), 4.57 (2H, s), 4.64 (1H, td, J = 4.8 Hz, J = 8.7 Hz), 7.10-7.14 (4H, m), 7.18-7.25 (5H, m).

Synthesis of (S)-cyanomethyl 2-(2-(4-(hydroxymethyl)phenyl)acetamido)-3 -phenylpropanoate (BzAlc-Phe-CME)



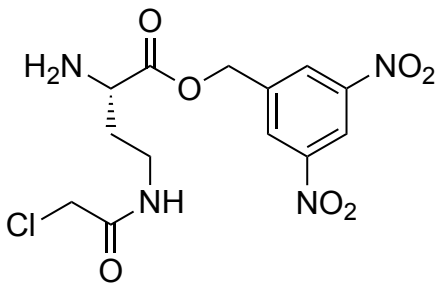
Crude product 2 (257.6 mg, 0.82 mmol) was dissolved in 3 mL DMF. Chloroacetonitrile (310 μL, 4.29 mmol) and ethyldiisopropylamine (156 μL, 0.90 mmol) were added to the mixture and stirred for 17 hours. Additional chloroacetonitrile (52μL, 0.81mmol) and ethyldiisopropylamine (71 μL,0.41mmol) were added and this solution was stirred for more 2 hours. The reaction was then diluted with AcOEt and the organic layer was was washed with 1M HCl, sat. $NaHCO_3$ aq. and brine, then dried over Na_2SO_4. Solvent was concentrated and purified by flash column chromatography (CHCl3 / AcOEt = 1 / 2). The desired product 1 ((189.6mg, 67% yield of 2 steps) as a white solid. 1H NMR (300 MHz, CDCl3) δ 1.68 (1H, t, J = 6.0 Hz), 2.99-3.12 (2H, m), 3.56 (2H, s), 4.71 (2H, d, J = 5.7 Hz), 4.76 (2H, q, J = 15.6 Hz), 4.87 (1H, q, J = 6.3Hz), 5.70 (1H, br d), 6.92-6.95 (2H, m), 7.17 (2H, d, J = 8.1 Hz), 7.35 (2H, d, J = 8.1Hz). ESIMS: m/z 353.1470 (M+H). calc: 353.1496.

Synthesis of (S)-cyanomethyl 2-(2-(4-formylphenyl)acetamido)-3-phenylpropanoate (BzAld-Phe-CME)



BzAlc-Phe-CME (18.6 mg, 53 μmol) was dissolved in 1.8 mL CH_2Cl_2 / THF (5 / 4) and to this suspension $NaHCO_3$ (44.6 mg, 530 μmol) and Dess-Martin periodinane (34.0 mg, 80 μmol) were added, then this solution was stirred for 1 hour under argon. The suspension was treated with $Na_2S_2O_3$ aq., diluted with CH_2Cl_2 and the layers were separated. The aqueous phase was extracted with CH_2Cl_2. The combined organic layers were washed with brine and dried over $MgSO_4$. Solvent was concentrated and purification by PTLC (hexane / AcOEt = 2 / 3) provided BzAld-Phe-CME (13.5 mg, 73%) as a white solid. 1H NMR (300 MHz, $CDCl_3$) δ 3.01-3.16 (2H, m), 3.63 (2H, s), 4.75 (2H, q, J = 15.6 Hz), 4.90 (1H, q, J = 6.3 Hz), 5.70 (1H, br d), 6.91-6.96 (2H, m), 7.20-7.7.25 (3H, m), 7.35 (2H, d, J = 8.1 Hz), 7.85 (2H, d, J = 8.1 Hz). ESIMS: m/z 351.1306 (M+H). calc: 351.1339.

Synthesis of (S)-3,5-dinitrophenyl 2-amino-4-(2-chloroacetamido)butanoate (Cab-DBE)



A mixture of (S)-2-((tert-butoxycarbonyl)amino)-4-(2-chloroacetamido) butanoic acid (472.5 mg, 1.5 mmol), 3,5-dinitrobenzyl chloride(422.3 mg, 1.95 mmol), and DIPEA (337 μL, 1.95 mmol) in DMF (5.0 mL) was stirred at room temperature for 12 hours. After the reaction, AcOEt (15 mL) was added to the solution and the organic layer was washed with 1M aqueous HCl (15 mL, twice). This layer was washed with saturated $NaHCO_3$ (15 mL, twice), then washed with brine and dried over $MgSO_4$. After evaporation and silica gel column chromatography (hexane : AcOEt = 1 : 1), (S)-3,5-dinitrophenyl 2-((tert-butoxycarbonyl)amino)-4-(2-chloroacetamido)butanoate (327.3 mg, 0.70 mmol, 46%, 2 steps) was obtained. This compound was deprotected using HCl/AcOEt, (S)-3,5-dinitrophenyl 2-amino-4-(2-chloroacetamido)butanoate (Cab-DBE, 119 mg, 45%)wasfinallyobtained. 1HNMR (DMSO-d6, 300MHz) δ8.83(s, 1H), 8.75(s, 2H), 8.71 (br, 2H), 8.53 (br, 1H), 5.55 (m, 2H), 4.20 (t, 1H), 4.08 (s, 2H), 2.05 (m, 2H). ESIMS: m/z 374.0606 (M+H). calc: 374.0629.

Synthesis of non-natural aminoacyl-tRNAs

Aminoacyl-tRNAs for 5 μL translation reactions were prepared as follows: 12 μL HEPES-KOH buffer (pH 7.5, 83 mM) that contained 42 μM tRNA and 42 μM flexizyme (eFx for CME-activated amino acids and dFx for DBE-activated ones) was heated at 95°C for 2 min and cooled to 25°C over 5 min. $MgCl_2$ (3 M, 4 μL) was added and the mixture was incubated at 25 °C for 5 min. The acylation reaction was initiated by the addition of activated amino acid substrate in DMSO (25 mM, 4 μL) and the mixture was incubated on ice for 2 h. After acylation, the reaction was stopped by the addition of 80 μL of 0.3 M sodium acetate (pH 5.2), and the RNA was precipitated by ethanol precipitation. The pellet was rinsed twice with 70% ethanol containing 0.1 M sodium acetate (pH 5.2), and once with 70% ethanol. The resulting amonoacyl-tRNA was dissolved in 1 mM sodium acetate (pH 5.2) just before adding it to the translation solution.

***In vitro* Translation and cyclization**

In vitro translation reaction was conducted by incubating the translation mixture (2 mM ATP, 2 mM GTP, 1 mM CTP, 1 mM UTP, 20 mM creatine phosphate, 50 mM HEPES-KOH pH 7.6, 100 mM potassium acetate, 12 mM Mg acetate, 2 mM Spermidine, 1 mM DTT, 15 mg/mL tRNA mix, 0.3 mM magnesium acetate, 1.2 μM E.coli ribosome, 0.6 μM MTF, 2.7 μM IF1, 0.4 μM IF2, 1.5 μM IF3, 0.26 μM EF-G, 10 μM EF-Tu, 0.66 μM EF-Ts, 0.25 μM RF2, 0.17 μM RF3, 0.5 μM RRF, 4 μg/mL crearine kinase, 3 μg/mL myokinase, 0.1 μM inorganic pyrophosphatase, 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 His RS, 0.4 μM IleRS, 0.04 μM LeuRS, 0.11 μM LysRS, 0.03 μ 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, 0.1 μM DNA template, 0.5 mM 19 amino acid mix except for Met, 50 μM acyl tRNA, 200 μM AcONa for formyl donor(-)) at 37°C for 30 min. Then, the translation product was incubated at

25°C for 12 min and 1/5 volume of 100 mM ETDA (pH 8.0) was added. The spontaneous cyclization was conducted by incubating the above solution at 37°C for 30 min.

Acetone precipitation

Three times volume of -20°C acetone was added into the peptide solution. Then, the solution was mixed by voltex and incubated at -20°C for over 20 min. After incubation, the solution was centrifuged at 4°C, 13000 rpm for 10 min. The supernatant was removed and 70% ethanol was added. After being centrifuged at 4°C, 13000 rpm for 3 min, the supernatant was discarded and the pellet was dissolved into water.

Isopropanol precipitation

1/10 volume of 3M NaCl was added into the peptide solution. Then, 1.3 times volume of isopropanol was added. After mixing by voltex, the solution was centrifuged at 25°C, 13000 rpm for 10 min. The supernatant was removed and 70% ethanol was added. After being centrifuged at 25 °C, 13000 rpm for three min, the supernatant was discarded and the pellet was dissolved into water.

Primer sequence lists

Templates of flexizyme and tRNA

Fx5'.F36

5′- GTAATACGACTCACTATAGGATCGAAAGATTTCCGC-3′

T7ex5.F22

5′-GGCGTAATACGACTCACTATAG-3′

eFx.R45

5′-ACCTAACGCTAATCCCCTTTCGGGGCCGCGGAAATCTTTCGATCC-3′

eFx.R18

5′-ACCTAACGCTAATCCCCT-3′

dFx.R46

5′-ACCTAACGCCATGTACCCTTTCGGGGATGCGGAAATCTTTCGATCC-3′

dFx.R19

5′-ACCTAACGCCATGTACCCT-3′

Ini1-1G-5'.F49

5′-GTAATACGACTCACTATAGGCGGGGTGGAGCAGCCTGGTAGCTCGTCGG- 3′

Ini cat.R44

5′- GAACCGACGATCTTCGGGTTATGAGCCCGACGAGCTACCAGGCT-3′

Ini-3'.R38

5′- TGGTTGCGGGGGCCGGATTTGAACCGACGATCTTCGGG-3′

Ini-3'.R20

5′-TGGTTGCGGGGGCCGGA TTT-3′

EnAsn-5'.F49

5′-GTAATACGACTCACTATAGGCTCTGTAGTTCAGTCGGTAGAACGGCGGA- 3′

EnAsn AUG.R43

5′- GAACCAGTGACATACGGATTATGAGTCCGCCGTTCTACCGACT-3′

EnAsn-3'.R38

5′- TGGCGGCTCTGACTGGACTCGAACCAGTGACATACGGA -3′

EnAsn-3'.R20

5′- TGGCGGCTCTGACTGGACTC -3′

Primers for the quantitative PCR and reconstruction of DNA library of the selection

T7g10M.F46

5′-TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATA-3′

T7g10M.F42

5′-TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATA-3′

T7g10M.F36

5′-TAATACGACTCACTATAGGGTTAACTTTAAGAAGGA-3′

CGS3ggatcc.Sgca.R39

5′-TTTCCGCCCCCCGTCCTAGGATCCGGATCCGGATCCGCA-3′

CGS3ggatcc.Sgca.R28

5′-TTTCCGCCCCCCGTCCTAGGATCCGGAT-3′

CGS3ggatcc.Sgca.R22

5′- TTTCCGCCCCCCGTCCTAGGAT -3′

CA6tgc6.Saca.R39

5′-TTTCCGCCCCCCGTCCTATGCTGCTGCTGCTGCTGCACA-3′

CA6tgc6.Saca.R25

5′- TTTCCGCCCCCCGTCCTATGCTGCT -3′

CA6tgc6.Saca.R22

5′- TTTCCGCCCCCCGTCCTATGCT -3′

Primers for DNA templates of selected peptides for validation by MALDI-TOF-MS

Short families (SADH-1, SADH-2 and SADH-3)

T7g10M.F60

5′-TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATGGG

TGGTGGTGGC-3′

S-ADH-1.Rv1.R48

5′-AGGCTGAGAACCAGTCCACCAAATAGTCATGCCACCACCACCCATATG-3′

S-ADH-1.Rv2.R48

5′-GGATCCGGATCCGGATCCGCACCTAAGATGAGGCTGAGAACCAGTCCA-3′

S-ADH-2.Rv1.R48

5′-AGGCTGCAACGAAAAATGCCAAAGATTCATGCCACCACCACCCATATG-3′

S-ADH-2.Rv2.R48

5′-GGATCCGGATCCGGATCCGCAAAAAAGCTTAGGCTGCAACGAAAAATG-3′

S-ADH-3.Rv1.R48

5′-ATGCGTAGTAGAACGCAGATTAATAAACATGCCACCACCACCCATATG-3′

S-ADH-3.Rv2.R48

5′-GGATCCGGATCCGGATCCGCACCAAAGAAAATGCGTAGTAGAACGCAG-3′

Long families (LADH-1, LADH-2 and LADH-3)

T7g10M.F60

5′-TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATGGG
TGGTGGTGGC-3′

L-ADH-1.Rv1.R63

5′-ACGCCCCGGCTGATAAGAATGAGGCCACAAAATCACAGACACCATCCCA
CCACCACCCATATG-3′

L-ADH-1.Rv2.R63

5′-TGCTGCTGCTGCTGCTGCACAATGAATAGTCGGAAAAGTATCATGACGCC
CCGGCTGATAAGA-3′

L-ADH-2.Rv1.R57

5′-CTCCCAAGGCCACACAAACCGAGCA TGA TGAA TCAACA TCCCACCACCA
CCCATATG-3′

L-ADH-2.Rv2.R57

5′-TGCTGCTGCTGCTGCTGCACACAACTGCATATTATCAGTCTCCCAAGGCC
ACACAAA-3′

L-ADH-3.Rv1.R63

5′-CTGAAGCCAATGATGAATCCGCGGAGAAGCCAGAAAAGGCGTCA TCCCA
CCACCACCCATATG-3′

L-ADH-3.Rv2.R63

5′-TGCTGCTGCTGCTGCTGCACACTTCCAAGTCAGAAGATGCTTAATCTGAA
GCCAA TGA TGAA T-3′

Forward primers for next-generation sequencing

Rd1T7g10M.F70

5′-CACTCTTTCCCTACACGACGCTCTTCCGATCTTAATACGACTCACTATAGG

GTTAACTTTAAGAAGGAGA-3′

an13Rd2.R49

5′-GACTGGAGTTCAGACGTGTGCTCTTCCGA TCTTTTCCGCCCCCCGTCCT-3′

P5S502Rd1.F57

5′-AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCT

ACACGAC-3′

P5S503Rd1.F57

5′-AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCT

ACACGAC-3′

P5S505Rd1.F57

5′-AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCT

ACACGAC-3′

P5S508Rd1.F57

5′-AATGATACGGCGACCACCGAGATCTACACCTAAGCCTACACTCTTTCCCT

ACACGAC-3′

P5S510Rd1.F57

5′-AATGATACGGCGACCACCGAGATCTACACCGTCTAATACACTCTTTCCCT

ACACGAC-3′

P5S511Rd1.F57

5′-AATGATACGGCGACCACCGAGATCTACACTCTCTCCGACACTCTTTCCCT

ACACGAC-3′

Reverse primers for next-generation sequencing

Rd2N701P7.R52

5′-CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGAC

GTG-3′

Rd2N702P7.R52

5′-CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGAC

GTG-3′

Rd2N703P7.R52

5′-CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGAC

GTG-3′

Rd2N704P7.R52

5′-CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGAC

GTG-3′

Rd2N705P7.R52

5′-CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGAC

GTG-3′

Rd2N706P7.R52

5′-CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGAC

GTG-3′

Rd2N707P7.R52

5′-CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGAC

GTG-3′

Rd2N710P7.R52

5′-CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGAC

GTG-3′

Rd2N711P7.R52

5′-CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTTCAGAC

GTG-3′

Rd2N712P7.R52

5′-CAAGCAGAAGACGGCATACGAGATTCCTCTACGTGACTGGAGTTCAGAC

GTG-3′

Rd2N714P7.R52

5′-CAAGCAGAAGACGGCATACGAGATTCATGAGCGTGACTGGAGTTCAGAC

GTG-3′

Rd2N715P7.R52

5′-CAAGCAGAAGACGGCATACGAGATCCTGAGATGTGACTGGAGTTCAGAC

GTG-3′

***In vitro* selection scheme by means of the RaPID system**

Translation scale at each round

In each selection, the selection was conducted as 150 μL translation scale for 1st round, as 5.0 μL translation scale for 2nd round and 2.5 μL scale for 3rd-7th round. In this experimental section, only the scheme of 2.5 μL scale was shown. 1st round was conducted according to this section as same way.

1. Ligation of the puromycin linker to the mRNA

Puromycin-ligation was conducted by incubating the ligation mixture (50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM DTT, 10 mM ATP, 20% DMSO, 1.5 μM CC-puromycin linker, 1 μM T4 RNA ligase and 1 μM mRNA library for 40 μL) at 25°C for 30 min. After ligation, the same volume of the solution containing 0.6 M NaCl and 10 mM EDTA (pH 7.5) was added. This mRNA-CC-puromycin-linker was purified by phenol/chloroform/isoamylalcohol extraction and ethanol precipitation according to above. The pellet was dissolved with H_2O and diluted to 6 μM.

2. Translation

Translation mix was prepared to contain 1 μM mRNA linker, 0.5 mM 19 amino acid mix except for Met, 50 μM acyl tRNA, 200 μM AcONa in addition to the same components written above. The solution was incubated at 37℃ for 30 min for translation, and then at 25℃ for 12 min. After adding 1/5 volume of 100 mM ETDA (pH 8.0) solution, the spontaneous cyclization was conducted by incubating at 37 ˚C for 30 min.

3. Reverse transcription

The resulting peptide-mRNA conjugates solution was added into reverse transcription mixture containing 0.26 mM dNTP, 26.7 mM EPPS (pH 8.3), 1.79 μM reverse primer (CGS3ggatcc.Sgca.R39 for the Short library, CA6tgc6.Saca.R39 for the

Long library), 10.6 mM KOH, 15.9 mM Mg(OAc)2 and M-MLV H(-) Reverse Transcriptase (Promega Corp.) in final concentration. The reverse transcription was conducted by incubating this solution at 42°C for 1 h. 1 μL of the product was diluted with 399 μL of H_2O to make a sample for the calculation of the whole amount of synthesized peptide-mRNA-cDNA conjugates.

4. Alcohol dehydrogenase (ADH) reaction

The ADH reaction was conducted by incubating the ADH solution (50 mM EPPS (pH 7.5), 500 mM KCl, 100 mM $MgCl_2$, 0.5 mM $ZnCl_2$ and 100 mM NAD^+) at 25°C for 18 h. Before adding NAD^+, peptides were incubated with $MgCl_2$ and $ZnCl_2$ at room temperature for 5 min.

5. Removal of bead binder

Preparation of Negative beads

The same volume of slurry of Dynabeads® M-280 was used. The three times volume of EK-T buffer (50 mM EPPS (pH 7.5), 500 mM KCl and 0.1% Tween (v/v)) was added and the solution was suspended. After being centrifuged, supernatant was discarded. Then, beads were rinsed out by repeating this manipulation for three times. After washing, the half amount of beads was incubated in masking buffer (50 μM biotin in EK-T buffer) at 4°C for 30 min. After washing masked beads, unmasked beads were added and the solution was suspended. Three sets of this beads solution were prepared for one round.

Capturing bead binding peptides

The same volume of EK-T buffer and negative beads was added to the solution after ADH reaction. The suspension was rotated at 4°C for 30 min. In order to remove bead binder, this manipulation was repeated for three times. After rotation, supernatant was recovered. For the conservation of the diversity of the initial peptide libraries, this step was skipped in the 1st round.

6.Biotin-hydrazonation

After translation reaction at a 2.5 μL scale, 2.5 μL of AEB buffer (5 mM biotin-hydrazide, 100 mM AcONa, 2 mM EDTA, pH 5.2) was added to this solution and incubated at 25°C for 1 hour.

7. Acetone precipitation

Three times volume of -20°C acetone was added into the resulting solution. Then, the solution was mixed by voltex and incubated at -20°C for over 20 min. After incubation, the solution was centrifuged at 4°C, 13000 rpm for 10 min. The supernatant was removed and the same volume of the negative selection solution of water was added. And then, 3 times volume of -20°C acetone was added into the solution. The solution was mixed by voltex and incubated at -20°C for over 20 min. After incubation, the solution was centrifuged at 4 °C, 13000 rpm for 10 min. The supernatant was removed and 70% ethanol was added. After being centrifuged at 4°C, 13000 rpm for 3 min, the supernatant was discarded and the pellet was dissolved into appropriate volume of EK-T buffer.

8. Positive selection

Preparation of Positive beads

100 μL slurry of Dynabeads® M-280 per 2.5 μL translation were used for the selection 1st-3rd round. 50 μL slurry of beads per 2.5 μL translation were used for the selection 4th-7th round. three times volume of EK-T buffer was added and the solution was suspended. After being centrifuged, supernatant was discarded. Then, beads were rinsed out by repeating this manipulation for three times.

Capturing biotinylated peptides

The pellet after acetone precipitation was dissolved into the eight times volume of bead-slurry of EK-T buffer. After suspending the solution with beads, the suspension was rotated at 4°C for 30 min. After rotation, supernatant was discarded.

9. Washing non-specifically binding peptides

Resulting beads were suspended in 3 times volume of bead-slurry of EK-T buffer containing 6 M Urea as the final conentration. Then, the suspension was rotated at 4°C for 30 min. After being centrifuged, supernatant was discarded. So as to wash away non-specifically binding peptides, this manipulation was repeated for three times.

10. Recovery and quantification of cDNA

1x PCR master mix (-Taq) containing 250 nM each forward or reverse primer (T7g10M.F36 and CGS3ggatcc.Sgca.R28 were used for the short library, T7g10M.F42 and CA6tgc6.Saca.R25 were used for the long library). 0.75% (v/v) of 1x Taq DNA polymerase and 0.5% (v/v) SYBR Green I was added to 1x PCR master mix (-Taq) to make 1x qPCR master mix. 1x PCR master mix (-Taq) 100 μL (400 μL for the selection 1st round) was added to the negative and positive beads, and then after incubating the beads at 95℃ for 5 min, the supernatant was recovered and called “Negative” or “Positive” samples. 1 μL of samples called “Input”, “Negative” and “Positive” was added to 19 μL of 1x qPCR master mix. Using these samples, real time qPCR was conducted under the condition as below. Short library: (94°C 10 sec → 61.0°C 10 sec → 72 ℃ 30 sec, 0.5 ℃/sec) x45 cycles to determine the binding fraction. Long library: (94 ℃ 10 sec → 62.8 ℃ 10 sec → 72 ℃ 40 sec, 0.5℃/sec) x45 cycles to determine the binding fraction.

11. Amplification of cDNA by PCR (reconstruction of DNA library)

0.75% (v/v) of 1x Taq DNA polymerase was added to 1x PCR elution mix (-Taq) 100 μL called “Negative” or “Positive” samples. Then, PCR primers were added (short library: T7g10M.F36 and CGS3ggatcc.Sgca.R28 250 μM in final concentration, long library: T7g10M.F42 and CA6tgc6.Saca.R25 250 μM in final concentration). PCR amplification of cDNA from short library was conducted by appropriate cycles of denaturing (94 °C for 40 sec), annealing (61 °C for 40 sec), and extending (72 °C for 40

sec). PCR amplification of cDNA from long library was conducted by appropriate cycles of denaturing (94 °C for 40 sec), annealing (62.8 °C for 40 sec), and extending (72 °C for 40 sec). Resulting cDNA was purified by phenol/chloroform/isoamylalcohol extraction and ethanol precipitation. The pellet was dissolved with 50 mM KCl 10 μL.

12. Transcription

The reagents were mixed to make 40 mM Tris, 1 mM spermidine, 0.01% (v/v) Triton X-100, 10 mM DTT, 25 mM $MgCl_2$, 2.3 mM NTPs, 1% (v/v) RNase inhibitor, 0.24 μM T7 RNA polymerase, 20% (v/v) cDNA solution (pH 8.0) for 10 μL scale. The solution was incubated at 37℃ overnight. After adding H_2O 10 μL to the soution, 0.6 M NaCl, 50 mM EDTA 20 μL was added. The product was purified by phenol/chloroform/isoamylalcohol extraction once and isopropanol precipitation twice. The pellet was dissolved with H_2O as 10 μM. The mRNA was used for ligation of the puromycin linker in next round of the selection.

Preparation of DNA samples and nextgeneration sequcing using MiSeq

PCR master mix 50 μL, 100 μM Rd1T7g10M.F70 0.25 μL, 100 μM an13Rd2.R49 0.25 μL and each 100 μM the cDNA library for “Input”, “Negative” and “Positive” of each round were mixed. Then, extension reaction (95℃ 1 min → (95℃ 1 min → 63℃ 1 min → 72℃ 1 min) x4 cycles → 72℃ 2 min) was carried out. PCR master mix 100 μL, the extension product 1.0 μL, 100 μM P5S5xxRd1.F57 0.5 μL and 100 μM Rd2N7xxP7.F52 0.5 μL were mixed. Then, PCR (95℃ 1 min → (95℃ 1 min → 52℃ 1 min → 72℃ 1 min) x8 cycles → 72℃ 2 min) was carried out. The amplification of DNA was analyzed using 3% agarose gel electrophoresis. The product was purified by phenol/chloroform extraction and ethanol precipitation. These procedures were following to the protocol written in Section 2.8. The same amount of each samples were mixed and purified using NucleoSpin Gel and PCR Clean-up® (Takara Bio, Inc.). The concentration of the samples was measured using Qubit dsDNA BR Assay kit (Thermo Fisher Scientific K.K.). The

samples were diluted with 10 mM Tris-HCl (pH 8.5) / 0.1% Tween20 as 10 nM. The concentration of the samples was measured using Qubit dsDNA HS Assay kit (Thermo Fisher Scientific K.K.) again. The samples were diluted with 10 mM Tris-HCl (pH 8.5) / 0.1% Tween20 as 4 nM. The samples were analyzed using MiSeq.

Preparation of DNA templates of SADH1-3 and LADH1-3.

1. Extension reaction

Forward primer and reverse primer (T7g10M.F60 and S/L-ADH-n.Rv1, n: the number corresponding to each peptide, 1 μM each) were mixed in the PCR miture (10 mM Tris·HCl (pH 9.0), 50 mM KCl, 2.5 mM $MgCl_2$, 0.25 mM each dNTPs, 0.1% (v/v) Triton X-100, and 45 nM Taq DNA polymerase). Primer extension was conducted in a 10-100 μL reaction mixture by denaturing (95°C for 1 min) followed by 5 cycles of annealing (50°C for 1 min) and extending (72°C for 1 min).

2. Amplification

The extension product was 10-fold diluted with the PCR mixture and primers for amplification (T7g10M.F42 and S/L-ADH-n.Rv2, n: the number corresponding to each peptide, 1 μM each) were added. Amplification by PCR was conducted in up to 200 μL reaction mix by 12-17 cycles of denaturing (95°C for 40 sec), annealing (50°C for 40 sec), and extending (72°C for 40 sec). Amplification PCR product was confirmed by 3% agarose gel electrophoresis and ethidium bromide staining.

3. Phenol/chloroform/isoamylalcohol extraction

The same volume of phenol/chloroform/isoamylalcohol (25:24:1) was added into the PCR product. The solution was mixed by voltex and centrifuged at 4°C at 13000 rpm for 5 min. The water layer was recovered and the same volume of chloroform/ isoamylalcohol (24:1) was added to the water layer. The solution was mixed by voltex and centrifuged at 4°C at 13000 rpm for 2 min. The water layer was recovered as the PCR

product.

4. Ethanol precipitation

1/10 volume of 3 M NaCl and the 2.2 times volume of ethanol were added to the PCR product. The solution was centrifuged at 25°C at 13000 rpm for 15 min. The supernatant was discarded and 70% ethanol was added. After being centrifuged at 25°C at 13000 rpm for 3 min, the supernatant was discarded. The pellet was and dried and dissolved in 1/10 volume of PCR solution of water.

***In vitro* translation, cyclization, ADH reaction and bitoin-hydrazonation of the selected peptides and MALDI-TOF-MS analysis**

In vitro translation reaction was conducted by incubating the translation mixture (2 mM ATP, 2 mM GTP, 1 mM CTP, 1 mM UTP, 20 mM creatine phosphate, 50 mM HEPES-KOH pH 7.6, 100 mM potassium acetate, 12 mM Mg acetate, 2 mM Spermidine, 1 mM DTT, 15 mg/mL tRNA mix, 0.3 mM magnesium acetate, 1.2 μM E.coli ribosome, 0.6 μM MTF, 2.7 μM IF1, 0.4 μM IF2, 1.5 μM IF3, 0.26 μM EF-G, 10 μM EF-Tu, 0.66 μM EF-Ts, 1μM RF1, 0.25 μM RF2, 0.17 μM RF3, 0.5 μM RRF, 4 μg/mL crearine kinase, 3 μg/mL myokinase, 0.1 μM inorganic pyrophosphatase, 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 His RS, 0.4 μM IleRS, 0.04 μM LeuRS, 0.11 μM LysRS, 0.03 μ 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, 0.1 μM DNA template, 0.5 mM 19 amino acid mix except for Met, 50 μM acyl tRNA, 200 μM AcONa for formyl donor(-)) at 37°C for 30 min. Then, the translation product was incubated at 25°C for 12 min and 1/5 volume of 100 mM ETDA (pH 8.0) was added. The spontaneous cyclization was conducted by incubating the above solution at 37°C for 30 min.

The resulted peptides were incubated in the ADH solution (50 mM EPPS (pH 7.5), 500 mM KCl, 100 mM $MgCl_2$, 0.5 mM $ZnCl_2$ and 100 mM NAD^+) at 25°C for 18 h. Before adding NAD^+, peptides were incubated with $MgCl_2$ and $ZnCl_2$ at room temperature for 5 min. After the ADH reaction, the same volume of AEB buffer (5 mM biotin-hydrazide, 100 mM AcONa, 2 mM EDTA, pH 5.2) was added to this solution and incubated at 25°C for 1 hour. The resulted products were purified by SPE C-TIP and analyzed by MALDI-TOF-MS.

The Python script for analyzing the results of high-throuput sequencing

Mr. Naoya Ozawa wrote this python script.

For analyzing new designed library, the primers were replaced as each proper one.

```
# Copyright (c) 2015 Naoya Ozawa

import sys
import re

def main(args):
    # The names of input files are provided
    # as arguments to the script
    input_names = args
    # Count the number of input files
    input_num = len(input_names)
    # Prepare a table for counting
    # {
    #     translated region: [
    #         count in the 1st file,
    #         count in the 2nd file,...],
    #     ...
    # }
    dna_counts = {}
    # Do the following for each file
    for i, input_name in enumerate(input_names):
        sys.stderr.write("Reading ¥"{0}¥"...".format(input_name))
        sys.stderr.flush()
        # Get list of DNAs of translated regions
```

```
        dnas = read_file(input_name)
        # Do the following for each DNA
        for dna in dnas:
            # Find the DNA in the table
            counts = dna_counts.get(dna, None)
            # If not found make a new entry
            if counts is None:
                counts = [0] * input_num
                dna_counts[dna] = counts
            # Increment the counter for the current file
            counts[i] += 1
        sys.stderr.write(" done.¥n")

# Write out results to stdout
# Write header
sys.stdout.write("DNA,length,fshift,peptide,{0}¥n".format(
        ",".join(map(
            lambda name: ".".join(name.split("/")[-1].split(".")[:-1]),
            input_names))))
# Write each entry
for dna, counts in dna_counts.iteritems():
        # Write following entries separated by a comma to stdout:
        #       - DNA of translated region,
        #       - length of DNA of translated region,
        #       - frame shift (number of extra bp),
        #       - peptide,
        #       - count in the 1st file,
        #       - count in the 2nd file,
```

```
            #         - ...
            sys.stdout.write("{0},{1},{2},{3},{4}¥n".format(
                dna,
                len(dna),
                len(dna)%3,
                translate(dna),
                ",".join(map(str, counts))))

        sys.stderr.write("Finished successfully.    Enjoy your sequences!¥n")

# Forward primer used in selection (T7g10M.F48)
forward_primer                                                              =
"TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATG"
# Reverse complement of reverse primer used in selection (CGS3an13.R39)
c_reverse_primer                                                            =
"TGCGGCAGCGGCAGCGGCAGCTAGGACGGGGGGCGGAAA"

# Returns a list of DNAs for cyclic regions which has passed a filter.
# Cyclic region is AUG-random-TGC
def read_file(input_name):
        input_reader = FASTQReader(open(input_name, "rU"))
        # Prepare arrays to save DNAs
        dnas = []
        # Do the following for each record in the file
        for name, full_dna, qualities in input_reader:
            # Discard sequences without complete forward primer
            if not full_dna.startswith(forward_primer):
                continue
```

```
            # Find reverse primer
            reverse_primer_pos = full_dna.find(c_reverse_primer)
            # Discard sequences without complete reverse primer
            if reverse_primer_pos == -1:
                continue
            # Caclulate the start of start codon
            start_pos = len(forward_primer) - 3
            # Calculate the end of cycteine codon
            end_pos = reverse_primer_pos + 3
            # Get the cyclic region
            dna = full_dna[start_pos:end_pos]
            # If any nucleotide has lower quality than 30,
            # discard this sequence
            if min(qualities[start_pos:end_pos]) < 30:
                continue
            # Save the DNA for cyclic region
            dnas.append(dna)
        return dnas

# Codon table
codon_table = {
        "TTT": "F", "TCT": "S", "TAT": "Y", "TGT": "C",
        "TTC": "F", "TCC": "S", "TAC": "Y", "TGC": "C",
        "TTA": "L", "TCA": "S", "TAA": "*", "TGA": "*",
        "TTG": "L", "TCG": "S", "TAG": "+", "TGG": "W",

        "CTT": "L", "CCT": "P", "CAT": "H", "CGT": "R",
        "CTC": "L", "CCC": "P", "CAC": "H", "CGC": "R",
```

```
        "CTA": "L", "CCA": "P", "CAA": "Q", "CGA": "R",
        "CTG": "L", "CCG": "P", "CAG": "Q", "CGG": "R",

        "ATT": "I", "ACT": "T", "AAT": "N", "AGT": "S",
        "ATC": "I", "ACC": "T", "AAC": "N", "AGC": "S",
        "ATA": "I", "ACA": "T", "AAA": "K", "AGA": "R",
        "ATG": "M", "ACG": "T", "AAG": "K", "AGG": "R",

        "GTT": "V", "GCT": "A", "GAT": "D", "GGT": "G",
        "GTC": "V", "GCC": "A", "GAC": "D", "GGC": "G",
        "GTA": "V", "GCA": "A", "GAA": "E", "GGA": "G",
        "GTG": "V", "GCG": "A", "GAG": "E", "GGG": "G"}

# Initiator amino acid
initiator_aa = "w"

def translate(dna):
        aa_list = []
        dna_cursor = 0
        dna_length = len(dna)
        if dna[0:3] == "ATG":
                aa_list.append(initiator_aa)
                dna_cursor = 3
        while dna_cursor + 3 <= dna_length:
                aa = codon_table[dna[dna_cursor : dna_cursor + 3]]
                aa_list.append(aa)
                dna_cursor += 3
        return "".join(aa_list)
```

```
class FASTQReader:
    def __init__(self, input_file):
        self.input_file = input_file

    def __iter__(self):
        return self

    # Returns (name, sequence, quality_list)
    def next(self):
        name_line = self.input_file.readline()
        seq_line = self.input_file.readline()
        plus_line = self.input_file.readline()
        quality_line = self.input_file.readline()

        if quality_line == "":
            raise StopIteration()

        return (
            name_line.rstrip("¥n")[1:],
            seq_line.rstrip("¥n"),
            map(lambda letter: ord(letter) - 33, quality_line.rstrip("¥n")))

    def close(self):
        self.input_file.close()

if __name__ == "__main__":
    main(sys.argv[1:])
```

Chapter5

Rationalization of the difference between the membrane permeability of CP2 and CP2.3

3. Rationalization of the difference of the membrane permeability of CP2 and CP2.3 peptides

3.1. Introduction

The membrane permeability is one of the most important chemical property for the development of drug molecules. The molecules that have membrane permeability can reach to the intracellular target molecules by passive diffusion. Especially in the development of peptidic drugs, the membrane permeability is one of the biggest obstacles in the application steps. In general, the discovery of peptidic drug candidates with the membrane permeability is still very difficult[88,106,107].

Cyclosporin A (CsA) is known as a peptide with the membrane permeability. It is used as a potent immunosuppressive drug[108,109] and it can reach and bind to the target molecule, cyclophilin, and this event leads to suppression of the response from T-cells and mast cells[110,111]. Interestingly, CsA has relatively high molecular weight and lipophilicity, which does not match to the conventional physicochemical guidelines for membrane permeability and bioavailability. Although many previous studies have explored correlations between various properties of small molecules and their membrane permeability[112–119], it is difficult to rationalize and explain the high membrane permeability and bioavailability of CsA.

Recently, methodologies of rationalization by means of molecular dynamics (MD) simulation have been developed and utilized for the rationalization and modeling of membrane permeation[120]. In particular, a study[121] has suggested kinetic models of CsA to rationalize its high membrane permeability by simulating multiple conformational states using MD simulation software, GROMOS (GROningen MOlecular Simulation); whose version is GROMOS11/GROMOS++, which is written in C++ and one of the fastest software for the molecular dynamics calculation of proteins, lipids and nucleic acids. In this report, the CsA is hypothesized to have conformational change from a conformation in water to a conformation in the membrane interior[121,122]. The MD simulation of the CsA

in water and chloroform as representative conditions of hydrophilic and hydrophobic environment, respectively, performed and the resulted dynamic structures and trajectories were compared with experimental data of NMR measurements. As the results, it suggests a generalized hypothesis of membrane permeation of molecules, where the conformations common in hydrophilic conditions, such as the extracellular and intracellular solution, and in hydrophobic environment, such as the inside of cell membranes, named as "congruent" conformation, is important for the membrane permeability and the higher frequency and longer lifetime of the congruent conformations lead to the membrane permeability (Figure 5.1).

In addition to the kinetic models of the CsA, the dynamics of several backbone-cyclized decapeptides in water and chloroform have been simulated by GROMOS and formation of intramolecular hydrogen bonds has been analyzed[123]. In this report, the formation of the transannular hydrogen bonds that promote the β-turn structures in water leads to the membrane permeability and the population of conformations that form β-turn structures in water showed positive correlation with PAMPA log P_e, comparing to the experimental results. Because the β-turn forming conformations were major in the hydrophobic conditions, this result supports the kinetics model of membrane permeation suggested by the MD simulation results of the CsA in the previous study, where congruent conformation is important for the membrane permeability.

In this study, I hypothesized that the above kinetic models can be applicable for not only backbone-cyclized peptides but also thioether-cyclized peptides. If it is simulated that thioether-cyclized peptides with different bioactivities expected to be resulted from the difference of the membrane permeability have conformational properties related to the congruent structures, it might provide us a point of view for the rationalization of the membrane permeability difference of macrocyclic peptides.

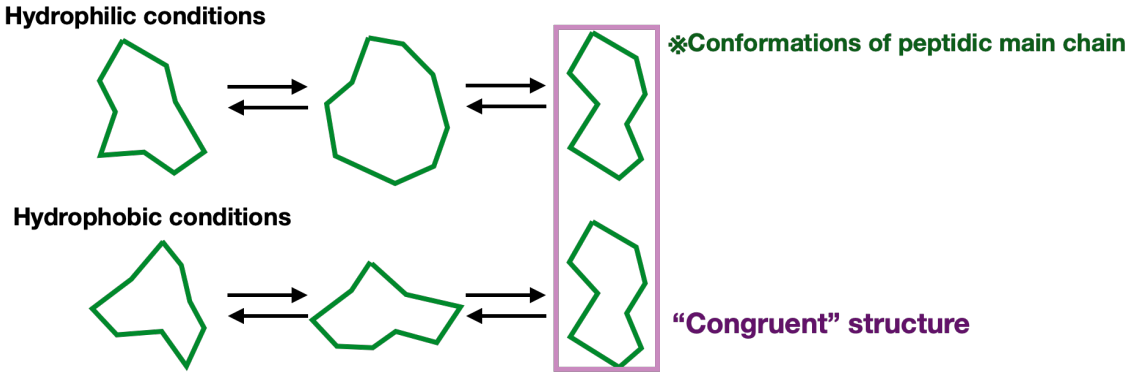


Figure 5.1 | The generalized hypothesis of membrane permeation of macrocyclic peptides. The “congruent” structure means common structures both in hydrophilic conditions and hydrophobic conditions.

In this study, I focused on the two macrocyclic peptides discovered as a strong binder against a histone demethylase: KMD4A, named as CP2 and CP2.3[82]. They have been discovered by the RaPID system and cyclized between a N-chloroacetyl residue of the N-terminal tyrosine and a thiol residue of C-terminus cysteine via a thioether linkage. The previous study revealed that the CP2.3 exhibits its inhibitory activity against the target protein, KMD4A, both *in vitro* and *in vivo* conditions, whereas the CP2 exhibits its activity only *in vitro* conditions[82]. Moreover, it is revealed that CP2 forms the β-turn like structure in the co-crystal with the KMD4A. I hypothesized the difference of the *in vivo* inhibitory activity of CP2 and CP2.3 is caused by the difference of the membrane permeability of them and aimed to rationalize the difference of the membrane permeability of CP2 and CP2.3 by using GROMOS.

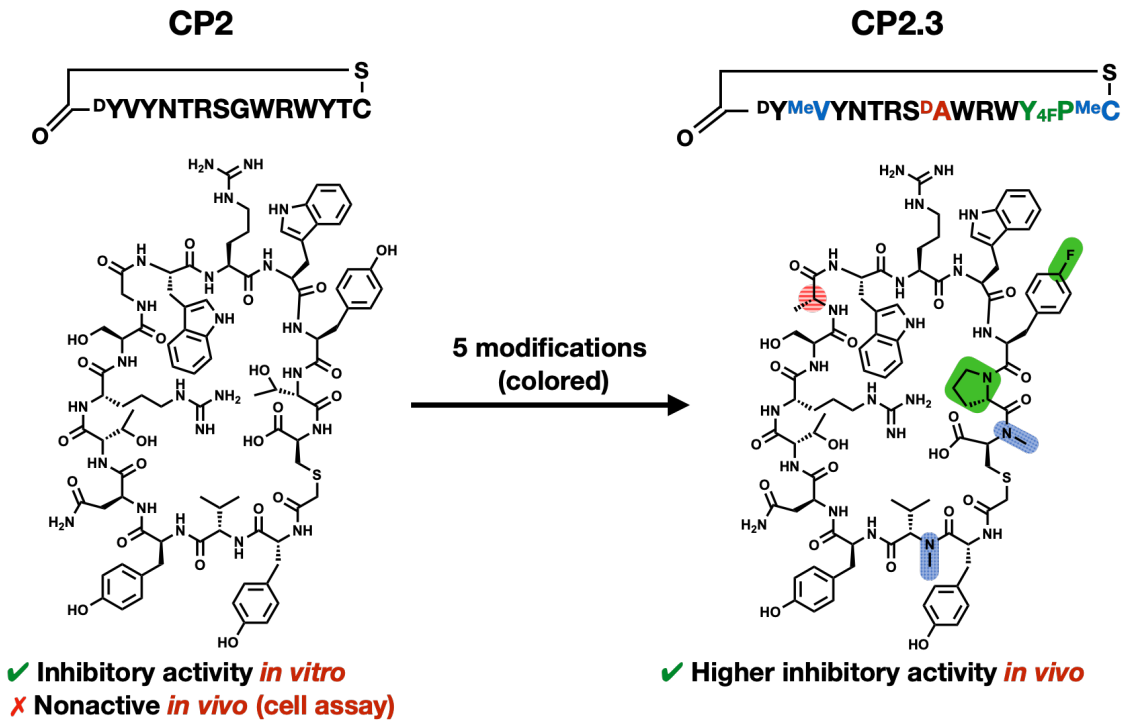


Figure 5.2 | The structures of CP2 and CP2.3 peptide. The CP2.3 has been generated by modifying the CP2. The five additional chemical modifications against CP2 were colored as red, blue and green.

3.2. Results and discussion

3.2.1. The results of the MD simulation of CP2

First, I simulated the conformations of CP2 and their transitions in water and chloroform as hydrophilic conditions and hydrophobic conditions, respectively. The detail conditions of simulation are described in the section of the material and method. Summarizing the simulation conditions, the simulation temperature was 300K and the total simulation time was 100 ns. I used a Hamiltonian exchange replicas method with 11 of the number of replica and 2 ps of the frequency of Hamiltonian replica exchanging.

The resulted trajectorieswere clustered and classified by its RMSD (root mean square deviation) with 0.06 nm of cutoff. The result is shown in the Figure 5.3. We can see the trajectories of CP2 in hydrophilic conditions have more variety than that of CP2 in hydrophobic conditions.

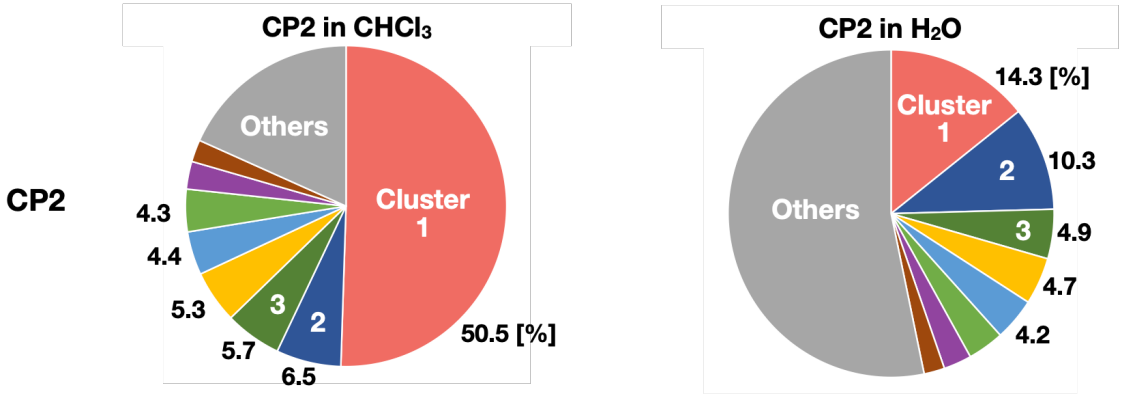


Figure 5.3 | The clusters of trajectories of CP2.

Next, as the more detail analysis, I analyzed the intramolecular backbone hydrogen bonds of eight representative clusters of CP2 in hydrophilic conditions (Figure 5.4). Because the congruent conformation is common in hydrophilic and hydrophobic conditions, I hypothesized that the common hydrogen bonds both in hydrophilic and hydrophobic conditions are involved in the congruent structure and uncommon hydrogen bonds prevent the peptide to form the congruent conformation. The results are shown in Figure 5.4. The hydrogen bonds are described as blue arrow from the donner to the acceptor. The center structure of each cluster calculated by RMSD are also shown under the diagram of each cluster. The percentages of clusters in all trajectories are written in parentheses.

As the result, the center structure of clusters 3 and 6 (occupied only 8.5% of all trajectories) had the G^9 to T^6 hydrogen bond, which is common in hydrophobic conditions. Moreover, the center of cluster 7 and 8 formed hydrogen bond that induce the “twisted” conformations and the percentage of trajectories belonging to the cluster 7 and 8 was 5%. These results indicates that the low frequency of the “congruent” hydrogen bond and formation of twisted structures are related to the membrane permeability of the CP2.

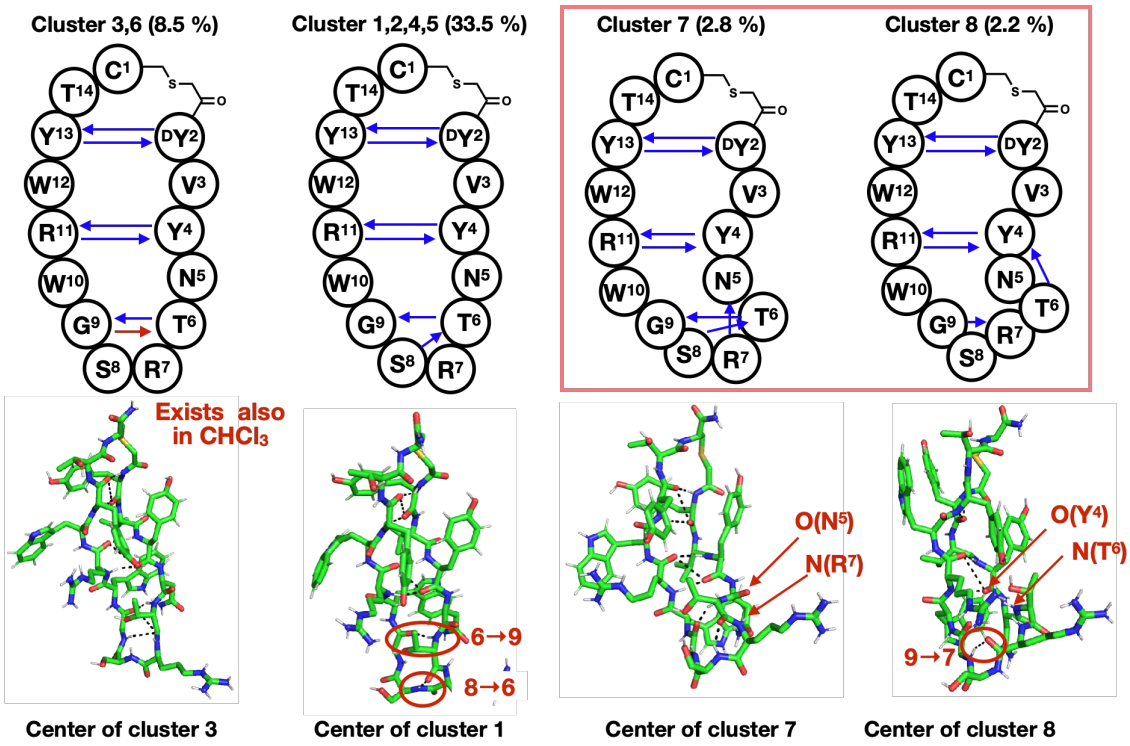


Figure 5.4 | The diagram of hydrogen bond formations of the clusters of CP2 in hydrophilic conditions. Mainchain-mainchain hydrogen bonds were analyzed and the formation is described as arrow from the donner to the acceptor. The clusters in red box are assumed as twisted structures based on the three-dimensional trajectory of each center of cluster.

5.2.1. The results of the MD simulation of CP2.3

Next, I simulated the conformations and conformational transitions of the CP2.3 with the same simulation conditions of the CP2. The result of clustering of trajectories is shown Figure 5.5. Comparing to the clustering result of the CP2, the variation of the clusters of CP2.3 is less than that of the CP2. The cluster 1, which has the most members of trajectories, of the CP2 in hydrophilic conditions occupied only 14% among all trajectories, whereas the percentage of cluster 1 of the CP2.3 in hydrophilic conditions was more than 60%. This result indicates that the conformation of the CP2.3 at a certain time has less than 0.06 nm of RMSD from the center of cluster 1 with more than 60% of probability. This probabilistic limitation leads to the low cost of entropy in conformation transitions. (The details are mentioned in the next section.)

Moreover, not only the variety of conformations of the CP2.3 in hydrophilic conditions but also the variety in hydrophobic conditions is lower than that of the CP2. The cluster 1 of the CP2 in hydrophobic conditions occupied 50.5% among all trajectories, however, the cluster of the CP2.3 occupied 53.7%. The percentages of the cluster 2 and 3 of the CP2.3 in hydrophobic conditions is also higher than that of the CP2. This result indicates that the CP2 can form diverse structures probabilistically even in hydrophobic conditions.

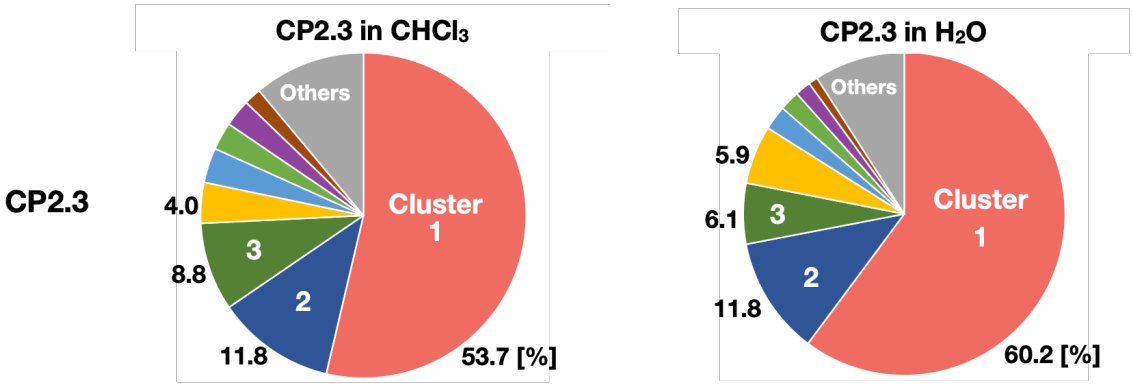


Figure 5.5 | The clusters of trajectories of CP2.3.

3.2.2. Comparison of the probability where the CP2/CP2.3 forms congruent conformations

In this section, I would compare the simulation results of the CP2 and CP2.3. To rationalize the difference of the membrane permeability between these two peptides, I would like to discuss the possibility where the CP2 or CP2.3 forms congruent conformations. If the CP2.3 forms the congruent conformations with a high probability, we can adopt the generalized hypothesis suggested in the previous study[121] for the explanation of the difference of the membrane permeability. First, I would compare the probability of hydrogen bond formation of each peptide in hydrophilic and hydrophobic conditions. Next, I would like to discuss why the difference of distribution of trajectories can leads to the difference of the membrane permeability, discussing the entropic cost of conformation transitions. Finally, I would discuss which modifications performed against CP2 are critical for the membrane permeability of the CP2.3.

As the first comparison between the CP2 and CP2.3, I would compare hydrogen bonds formed in hydrophilic conditions and hydrophobic conditions (Figure 5.6). The upper column of the Figure 5.6 summarizes the frequency of hydrogen bonds of CP2 in hydrophilic or hydrophobic environment. The frequency of each hydrogen bond was calculated and colored blue with the corresponding intensity. We can see some hydrogen bonds, Y^{13} to $^{D}Y^{2}$, R^{11} to Y^{4}, Y^{4} to R^{11} and T^{6} to G^{9}, are common in both hydrophilic and hydrophobic conditions. However, the region from T^{6} to G^{9} has different hydrogen patterns in different conditions. Although the major hydrogen bonds formed in hydrophobic conditions are G^{9} to T^{6}, W^{10} to S^{8} and G^{9} to R^{7} hydrogen bonds, in hydrophilic conditions, the G^{9} to T^{6} bond and S^{8} to T^{6} bond were frequently formed. The former is common in hydrophobic conditions, however, the latter is not observed in hydrophobic conditions. Because the formation of congruent, *i.e.,* common in both conditions, hydrogen bonds is necessary to form congruent conformation, at least, we can think that the formation of the S^{8} to T^{6} bond prevent the CP2 to form congruent

conformations. In fact, the standard absolute difference (SADF) of CP2 between the trajectories formed in hydrophilic and hydrophobic conditions was higher than that of CP2.3. The SADF was calculated by the below equation.

$$SDF(water, chloroform) = \frac{1}{N_{donor} \times N_{acceptor}} \sum_{j}^{N_{donor}} \sum_{i}^{N_{acceptor}} | n_{ij}^{water} - n_{ij}^{chloroform} |$$

n_{ij} = the total number of hydrogen bond from the j-th residue to the i-th residue

N_{donor}, $N_{acceptor}$ = the number of NH or CO that can be donor or acceptor of h-bond

This difference of SADFs supports the hypothesis that the formation probability of congruent conformation of the CP2 is lower than the CP2.3 and lead to the low membrane permeability and *in vivo* activity of the CP2.

Comparing to the CP2, the hydrogen bond pattern of CP2.3 in hydrophobic conditions is similar to that in hydrophilic conditions (Figure 5.6 b). Focusing on the region from T^6 to G^9 in hydrophilic conditions, the $^{D}A^9$ to T^6 and T^6 to $^{D}A^9$ bond are the major hydrogen bonds. These two hydrogen bonds are common in hydrophobic conditions and these congruent hydrogen bonds can be related to the higher membrane permeability and *in vivo* activity of the CP2.3. In addition to the region from T^6 to G^9, the $^{D}Y^2$ to $Y_{4F}{}^{13}$ hydrogen bond is dominantly formed, whereas the CP2 forms both of the C^1 to Y^{13} and $^{D}Y^2$ to Y^{13} hydrogen bonds. The formation of two hydrogen bonds toward Y^{13} indicates that this region of the CP2 have more flexibility than that of the CP2.3. As the modifications performed against the CP2 to generate the CP2.3, the C^1 residues of the CP2 had been N-methylated and resulted $^{Me}C^1$ cannot be a donner of a hydrogen bond in principle. We can conclude that the low flexibility of this region also save the entropic cost of the CP2.3 during hydrophilic-hydrophobic environment change of membrane permeation.

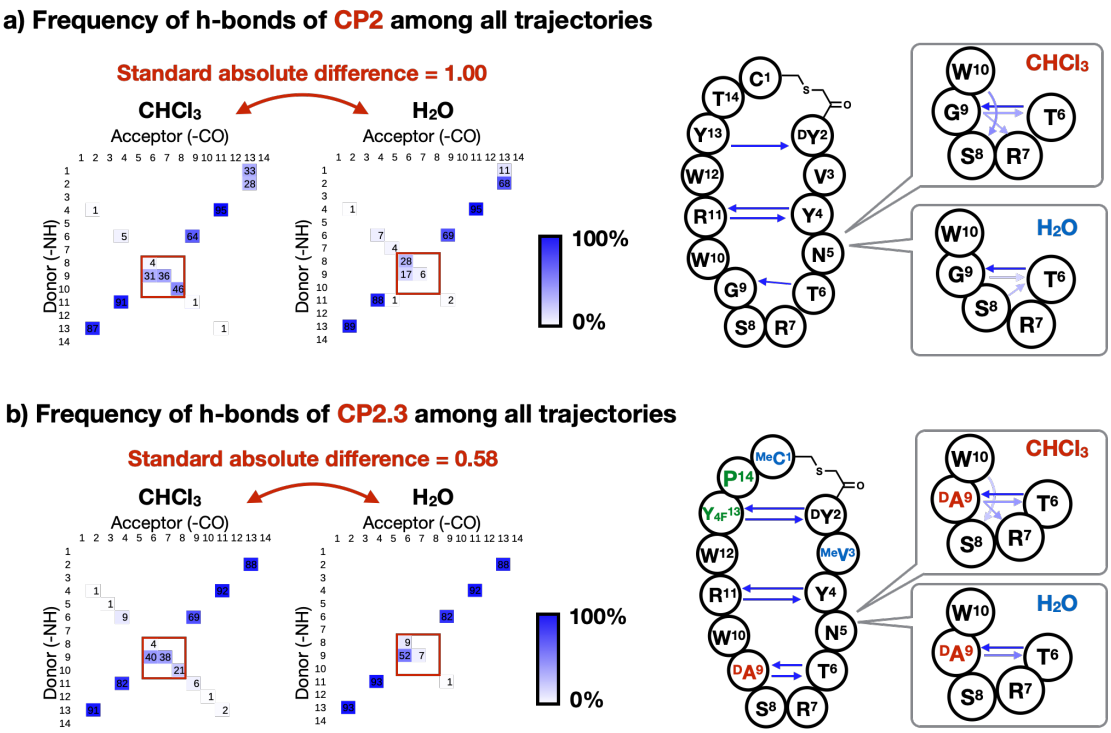


Figure 5.6 | Analyses of the hydrogen bond formation of the CP2/CP2.3 in hydrophobic/hydrophilic conditions.

Finally, to assume the difference of the entropic cost of the CP2 and CP2.3 during hydrophilic-hydrophobic environment change, I compared the RMSF (root mean square fluctuation) of backbones of the CP2 and CP2.3. RMSF of a certain atom is calculated as below.

$$RMSF(\mathbf{r}_i) = \left(\frac{1}{N_t} \sum_{t}^{N_t} (\mathbf{r}_i(t_n) - \langle \mathbf{r}_i \rangle)^2 \right)^{\frac{1}{2}}$$

Shown as the above equation, the RMSF indicates how an atom had violently moved during simulation. In this section, I calculated the RMSF of α-carbon atoms of amino acids over the whole simulation. Using this calculation, we can assume the flexibility of backbones of the macrocyclic peptides. The results are shown in the Figure 5.7. Comparing the RMSF in hydrophilic conditions and the RMSF in hydrophobic conditions, the 3rd-8th region of the CP2 has much higher RMSF in hydrophilic conditions than RMSF in hydrophobic conditions. This difference means that when the environment around the CP2 change from hydrophilic to hydrophobic, the 3rd-8th region of the CP2 can be unstabilized by the entropic cost. Considering the situation where the peptides permeate into the cell membrane, the environment surrounding the peptides changes from hydrophilic to hydrophobic. We can assume the high entropic cost of the CP2 during this permeation process leads to the low membrane permeability and low *in vivo* activity.

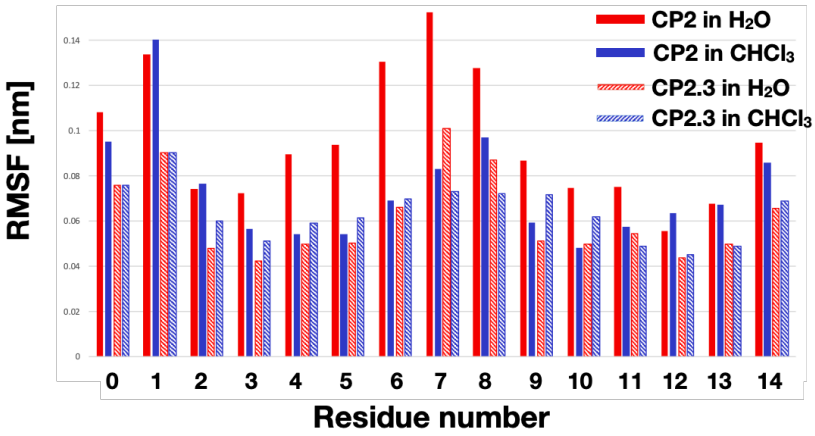


Figure 5.7 | RMSF of α-carbon atoms of backbone amino acids.

3.3. Conclusion

In this chapter, I aimed to rationalize the difference between the membrane permeability of the CP2 and CP2.3 macrocyclic peptides by means of MD simulation. In the section 5.2 and 5.3, I clustered the trajectories of the CP2 and CP2.3 in water and chloroform as hydrophilic environments and hydrophobic environments, respectively. As the results, it has been revealed that the CP2 can form more various conformations than the CP2.3. Moreover, the center structures of several clusters formed by CP2 in hydrophilic conditions had a twisted form and several unique hydrogen bonds accompanied by the twisted conformation appeared.

In the section 5.4, I compared the hydrogen bond pattern of the CP2 and CP2.3 in hydrophobic/hydrophilic conditions. The CP2.3 had more hydrogen bonds common in both hydrophilic and hydrophobic conditions than the CP2 had. This result may indicate that the CP2.3 tends to form the congruent structures than the CP2. Finally, I calculated and compared the RMSF of the CP2 and CP2.3 peptides. The difference between RMSF values in hydrophilic conditions and hydrophobic conditions can explain the difference of membrane permeability of two peptides.

3.4. Material and method

GROMOS version and force field parameter set

The GROMOS11 (the version released in May 2011, written in C++) was used for the simulations in this chapter. The 54A7 parameter set was utilized for the force field parameter sets, which is optimized from the 53A5 force field with adjusted torsional angle terms to better reproduce helical propensities, altered N–H, C=O repulsion, new CH_3 charge group, parameterisation of Na^+ and Cl^- to improve free energy of hydration and new improper dihedrals for the expression of aqueous or apolar solutions of proteins, nucleotides and sugars[124,125]. In particular, the epsRF (reaction field relative dielectric permittivity) of water and chloroform in this parameter sets are designated as 61 and 4.8, respectively.

The details of simulations by GROMOS11 are described the bellow; basic manipulations are in accordance with the GROMOS documentations listed on the http://www.gromos.net/page.pl?page=about page. The words highly related to the manipulations and/or mentions of GROMOS software, *e.g.*, program commands, variable specifications, arguments, important notes in the scripts and so on, are written in `Monaco font`.

1. Preparation of topology files

The topology files of CP2 and CP2.3 were generated by the packaged program in GROMOS11, `make_top`, with the amino acid unit sets and parameter sets of 54A7 by mentioning `@build` as 54a7.mtb and `@param` as 54a7.ifp parameter package files with the corresponding peptide sequences. For the simulations of the CP2 and CP2.3 in hydrophilic conditions, charged arginine residues (packaged as `ARG`) were used for the topology files. On the other hand, for the simulations of the peptides in hydrophobic conditions, deprotonated arginine residues (packaged as `ARGN`) were utilized.

The topology files of charged peptides were fused with a topology file of counter chloride ions, which are generated with the same parameter sets and `@seq CL- CL-` argument due to the compatibility with neutralized simulation boxes in further steps. The combination of topology files was conducted by the `com_top` program.

2. Energy minimization

As the co-crystal structure of the CP2 with KMD4A would be used for the initial state of MD (molecular dynamics) simulations, the energy minimization of the initial state was performed. The simulation temperature was set as 0 K and 2000 steps of MD simulation were performed with 2 fs of the integration time step. The parameters of the input file were set as `NSTLIM` (the maximum number of steps): `2000` and `DT`: `0.002`; the total simulation time was 4 ps. This energy minimization was conducted in gas phase (in vacuum).

3. Solvation of peptides into simulation solvent boxes

The peptides were put in the simulation boxes filled with the corresponding solvent, *i.e.*, water (SPC) or chloroform. The coordinates of atoms in the co-crystal structure of KMD4A and the CP2 peptides was used for initial state of two peptides. For the elimination of solvent molecules in the space occupied by inserted peptide molecules, the

minimal wall and threshold were argued as 0.8 nm and 0.23 nm, respectively, by the mention of `@minwall 0.8 @thresh 0.23` with `@rotate` argument.

The solvation of a charged peptide was conducted by two steps. First, a peptide without counter ion had been put in the simulation box. Second, the counter chloride ions were installed into a simulation box by using the `ion` program with `@negative 2 CL- @potential 0.8 @mindist 0.35` arguments.

4. Equilibration in the simulation solvent box

In order to relax the unfavorable atom-atom contacts between the solute(s) and the solvent, energy minimization of the solvent was conducted. The simulation temperature was set as 0 K and 10000 steps of MD simulation were performed with 2 fs of the integration time step, *i.e.*, 20 ps as the total time.. The parameters of the input file were set as `NSTLIM` (the maximum number of steps): `10000` and `DT`: `0.002`.

5. Thermalization and constant pressure equilibration

The temperature of simulation box was thermalized up to 300 K, *i.e.,* the initial velocity from a Maxwell-Boltzmann distribution was given to the constructed topology and initial coordinates. The thermalization steps were performed every 20 ps with the weak-coupling scheme algorithm, setting the parameter of `ALGORITHM` as 0. Every 20 ps, the system was thermalized 60.0 K and finally the system was thermalized up to 300 K with 5 steps. After each thermalization step, the MD simulation: the constant pressure equilibration was performed 20 ps with 2 fs of the integration time step, setting the parameter `NSTLIM` and `DT` as `10000` and `0.002`, respectively. After the system was thermalized to 300 K, the 100 ns of equilibration with 2 fs of the integration time step was performed.

6. Hamiltonian replica exchange MD simulation

To extend the sampling field, Hamiltonian replica exchange MD simulations[126,127] were conducted against four conditions; the CP2/CP2.3 in hydrophilic/hydrophobic environment. The imaginary Hamiltonian was calculated from the peptide with flexible dihedral angles. The number of Hamiltonian exchange replicas was set as 11 and the replica of Hamiltonian ($\mathscr{H}*$) was calculated as the below.

$$\mathscr{H}* = \lambda\, \mathscr{H}_{\text{constrained dihedral angle}} + (1-\lambda)\, \mathscr{H}_{\text{flexible dihedral angle}}$$
$$(\lambda = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)$$

The frequency of Hamiltonian replica exchanging was once every 2 ps and the total simulation time was 100 ns with 2 fs of the integration time step (the number of total step: 5×10^7). The trajectory with minimal Hamiltonian is adopted as the initial state of the next step, which leads to find a conformation with lower energy unexpected to be found in the sampling field without imaginary Hamiltonian replica generation.

7. Trajectory recording

The coordinate trajectory was exported every 100 steps during MD simulations, *i.e.*, the trajectories were recorded every 0.2 ps and 500001 trajectories have been obtained. The trajectories of Hamiltonian replica with λ = 1.0 were used for the further analyses.

8. Calculation of RMSD (Root-mean-square difference)

The RMSD is defined as the below;

$$RMSD(\mathbf{r}, \mathbf{r}^{\text{ref}}) = \left(\frac{1}{N_a} \sum_{\text{i}=1}^{N_a} (\mathbf{r}_i - \mathbf{r}_i^{ref})^2 \right)^{\frac{1}{2}}$$

N_a = the number of atoms

$\mathbf{r}_{\text{i}}$ = the coordinate of the i-th particle in the one configuration

$\mathbf{r}^{\text{ref}}$ = the reference position

In this study, **r** was substituted by the coordinate of a trajectory at the certain time point. $\mathbf{r}^{\text{ref}}$ was substituted by the coordinate of the trajectory of the initial state, the trajectory resulted from the previous equilibration.

For the calculation, `rmsd` and `rsmdmat` program was used with `@atomsrmsd 1:CA,N,C` argument, where the backbone atoms of peptide (α-carbon, nitrogen and carbon of the backbone amide moieties) was substituted into **r**, which generate the rmsd matorix of arbitrary two time points.

9. Clusterization of trajectories

The clusterization of trajectories was based on rmsd matrix generated by the rmsd calculation step. The cutoff was set as 0.06 nm with `@cutoff` argument, which results to specify the similarity criterion for two structures. The program `cluster` classifies the all trajectories and generates the clusters with the corresponding cutoff value.

10. The analyses of hydrogen bonds

The formations of intramolecular hydrogen bonds were analyzed by `hbond` program packed in the GROMOS simulation. Several previous studies[121,123] have succeeded to analyze the hydrogen bond formations in the same solvents, SPC water and chloroform. Based on these studies, the parameters of hydrogen bond analyses were set by `@Hbparas` as `0.25` and `135`, where a maximum distance between the hydrogen and the acceptor is 0.25 nm and a minimum angle between donor, hydrogen and acceptor atoms is 135°. The positional relations that fit the above conditions are counted as the formation of intramolecular hydrogen bonds. The frequency was calculated by the number of the above hydrogen formation during the entire 100 ns MD simulation.

11. Calculation of RMSF (Root-mean-square)

The definition of RMSF is given by the below equation;

$$RMSF(\mathbf{r}_i) = \left(\frac{1}{N_t} \sum_{t}^{N_t} (\mathbf{r}_i(t_n) - \langle \mathbf{r}_i \rangle)^2 \right)^{\frac{1}{2}}$$

N_t = the number of time points

$\mathbf{r}_i(t_n)$ = the position of atom *i* at time t_n

$\langle \mathbf{r}_i \rangle$ = mean position

The `rmsf` program was used with `@atomsrmsf 1:CA,N,C` argument, where the backbone atoms of peptide (α-carbon, nitrogen and carbon of the backbone amide moieties) was substituted into **r**.

Chapter6

General conclusion

第 6 章については、特許申請に使用する予定のため非公開。

References

1. Nakamura, T. & Ichihara, K. N. A. Partial purification and characterization of hepatocyte growth factor from serum of hepatectomized rats. **122**, 10 (1984).
2. Michalopoulos, G., Houck, K. A., Dolan, M. L. & Luetteke, N. C. Control of Hepatocyte Replication by Two Serum Factors. **44**, 7 (1984).
3. Nakamura, T. *et al.* Molecular cloning and expression of human hepatocyte growth factor. **342**, 4 (1989).
4. Keiji, M. *et al.* Molecular Cloning and Sequence Analysis of cDNA for Human Hepatocyte Growth Factor. *Biochem. Biophys. Res. Commun.* **163**, 967–973 (1989).
5. Weidner, K. M. *et al.* Evidence for the identity of human scatter factor and human hepatocyte growth factor. *Proc. Natl. Acad. Sci.* **88**, 7001–7005 (1991).
6. Trusolino, L., Bertotti, A. & Comoglio, P. M. MET signalling: principles and functions in development, organ regeneration and cancer. *Nat. Rev. Mol. Cell Biol.* **11**, 834–848 (2010).
7. Matsumoto, K., Funakoshi, H., Takahashi, H. & Sakai, K. HGF–Met Pathway in Regeneration and Drug Discovery. *Biomedicines* **2**, 275–300 (2014).
8. Lokker, N. A. *et al.* Structure-function analysis of hepatocyte growth factor: identification of variants that lack mitogenic activity yet retain high affinity receptor binding. *EMBO J.* **11**, 2503–2510 (1992).
9. Chirgadze, D. Y. *et al.* Crystal structure of the NK1 fragment of HGF/SF suggests a novel mode for growth factor dimerization and receptor binding. *Nat. Struct. Biol.* **6**, 8 (1999).
10. Birchmeier, C., Birchmeier, W., Gherardi, E. & Vande Woude, G. F. Met, metastasis, motility and more. *Nat. Rev. Mol. Cell Biol.* **4**, 915–925 (2003).
11. Organ, S. L. & Tsao, M.-S. An overview of the c-MET signaling pathway. *Ther. Adv. Med. Oncol.* **3**, S7–S19 (2011).

12. Lemmon, M. A. & Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **141**, 1117–1134 (2010).
13. Mele, S. & Johnson, T. K. Receptor Tyrosine Kinases in Development: Insights from Drosophila. *Int. J. Mol. Sci.* **21**, 188 (2019).
14. Szilveszter, K. P., Németh, T. & Mócsai, A. Tyrosine Kinases in Autoimmune and Inflammatory Skin Diseases. *Front. Immunol.* **10**, 1862 (2019).
15. Wang, H. *et al.* The Function of the HGF/c-Met Axis in Hepatocellular Carcinoma. *Front. Cell Dev. Biol.* **8**, 55 (2020).
16. Thayaparan, T., Spicer, J. F. & Maher, J. The role of the HGF/Met axis in mesothelioma. *Biochem. Soc. Trans.* **44**, 363–370 (2016).
17. Cheng, F. & Guo, D. MET in glioma: signaling pathways and targeted therapies. *J. Exp. Clin. Cancer Res.* **38**, 270 (2019).
18. Mulcahy, E. Q. X., Colón, R. R. & Abounader, R. HGF/MET Signaling in Malignant Brain Tumors. *Int. J. Mol. Sci.* **21**, 7546 (2020).
19. Miranda, O., Farooqui, M. & Siegfried, J. Status of Agents Targeting the HGF/c-Met Axis in Lung Cancer. *Cancers* **10**, 280 (2018).
20. Moosavi, F., Giovannetti, E., Saso, L. & Firuzi, O. HGF/MET pathway aberrations as diagnostic, prognostic, and predictive biomarkers in human cancers. *Crit. Rev. Clin. Lab. Sci.* **56**, 533–566 (2019).
21. Liang, H. & Wang, M. MET Oncogene in Non-Small Cell Lung Cancer: Mechanism of MET Dysregulation and Agents Targeting the HGF/c-Met Axis. *OncoTargets Ther.* **Volume 13**, 2491–2510 (2020).
22. Uehara, Y. *et al.* Placental defect and embryonic lethality in mice lacking hepatocyte growth factor/scatter factor. *Nature* **373**, 702–705 (1995).
23. Hammond, D. E. *et al.* Endosomal Dynamics of Met Determine Signaling Output. *Mol. Biol. Cell* **14**, 1346–1354 (2003).
24. Joffre, C. *et al.* A direct role for Met endocytosis in tumorigenesis. *Nat. Cell Biol.* **13**, 827–837 (2011).

25. Barrow-McGee, R. & Kermorgant, S. Met endosomal signalling: In the right place, at the right time. *Int. J. Biochem. Cell Biol.* **49**, 69–74 (2014).
26. Mitchell, A. C., Briquez, P. S., Hubbell, J. A. & Cochran, J. R. Engineering growth factors for regenerative medicine applications. *Acta Biomater.* **30**, 1–12 (2016).
27. Wang, Z. *et al.* Novel biomaterial strategies for controlled growth factor delivery for biomedical applications. *NPG Asia Mater.* **9**, e435–e435 (2017).
28. Ross, J. *et al.* Protein Engineered Variants of Hepatocyte Growth Factor/Scatter Factor Promote Proliferation of Primary Human Hepatocytes and in Rodent Liver. *Gastroenterology* **142**, 897–906 (2012).
29. Ito, K. *et al.* Artificial human Met agonists based on macrocycle scaffolds. *Nat. Commun.* **6**, 6373 (2015).
30. Ueki, R. *et al.* A chemically unmodified agonistic DNA with growth factor functionality for in vivo therapeutic application. *Sci. Adv.* **6**, eaay2801 (2020).
31. Brinkhuis, R. P., Rutjes, F. P. J. T. & van Hest, J. C. M. Polymeric vesicles in biomedical applications. *Polym. Chem.* **2**, 1449 (2011).
32. Elsabahy, M. & Wooley, K. L. Design of polymeric nanoparticles for biomedical delivery applications. *Chem. Soc. Rev.* **41**, 2545 (2012).
33. Kamaly, N., Xiao, Z., Valencia, P. M., Radovic-Moreno, A. F. & Farokhzad, O. C. Targeted polymeric therapeutic nanoparticles: design, development and clinical translation. *Chem. Soc. Rev.* **41**, 2971 (2012).
34. Zasadzinski, J. A., Wong, B., Forbes, N., Braun, G. & Wu, G. Novel methods of enhanced retention in and rapid, targeted release from liposomes. *Curr. Opin. Colloid Interface Sci.* **16**, 203–214 (2011).
35. Chandrawati, R. & Caruso, F. Biomimetic Liposome- and Polymersome-Based Multicompartmentalized Assemblies. *Langmuir* **28**, 13798–13807 (2012).
36. Akbarzadeh, A. *et al.* Liposome: classification, preparation, and applications. *Nanoscale Res. Lett.* **8**, 102 (2013).

37. Kumar, S. & Randhawa, J. K. High melting lipid based approach for drug delivery: Solid lipid nanoparticles. *Mater. Sci. Eng. C* **33**, 1842–1852 (2013).

38. Kumar, A., Zhang, X. & Liang, X.-J. Gold nanoparticles: Emerging paradigm for targeted drug delivery system. *Biotechnol. Adv.* **31**, 593–606 (2013).

39. Wang, S., Wen, Y., Cui, Z. & Xue, Y. Size dependence of adsorption kinetics of nano-MgO: a theoretical and experimental study. *J. Nanoparticle Res.* **18**, 15 (2016).

40. Valizadeh, A. *et al.* Quantum dots: synthesis, bioapplications, and toxicity. *Nanoscale Res. Lett.* **7**, 480 (2012).

41. Probst, C. E., Zrazhevskiy, P., Bagalkot, V. & Gao, X. Quantum dots as a platform for nanoparticle drug delivery vehicle design. *Adv. Drug Deliv. Rev.* **65**, 703–718 (2013).

42. Speir, J. A., Munshi, S., Wang, G., Baker, T. S. & Johnson, J. E. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo- electron microscopy. 33 (2014).

43. Gillitzer, E., Suci, P., Young, M. & Douglas, T. Controlled Ligand Display on a Symmetrical Protein-Cage Architecture Through Mixed Assembly. *Small* **2**, 962–966 (2006).

44. Suci, P. A., Varpness, Z., Gillitzer, E., Douglas, T. & Young, M. Targeting and Photodynamic Killing of a Microbial Pathogen Using Protein Cage Architectures Functionalized with a Photosensitizer. *Langmuir* **23**, 12280–12286 (2007).

45. Minten, I. J., Hendriks, L. J. A., Nolte, R. J. M. & Cornelissen, J. J. L. M. Controlled Encapsulation of Multiple Proteins in Virus Capsids. *J. Am. Chem. Soc.* **131**, 17771–17773 (2009).

46. Lin, T. *et al.* The Refined Crystal Structure of Cowpea Mosaic Virus at 2.8 Å Resolution. *Virology* **265**, 20–34 (1999).

47. Ochoa, W. F., Chatterji, A., Lin, T. & Johnson, J. E. Generation and Structural Analysis of Reactive Empty Particles Derived from an Icosahedral Virus. *Chem.*

Biol. **13**, 771–778 (2006).

48. Steinmetz, N. F., Lomonossoff, G. P. & Evans, D. J. Cowpea Mosaic Virus for Material Fabrication: Addressable Carboxylate Groups on a Programmable Nanoscaffold. *Langmuir* **22**, 3488–3490 (2006).
49. Strable, E. *et al.* Unnatural Amino Acid Incorporation into Virus-Like Particles. *Bioconjug. Chem.* **19**, 866–875 (2008).
50. Golmohammadi, R., Fridborg, K., Bundule, M., Valegård, K. & Liljas, L. The crystal structure of bacteriophage Qβ at 3.5 Å resolution. *Structure* **4**, 543–554 (1996).
51. Prasuhn, D. E. *et al.* Plasma Clearance of Bacteriophage Qβ Particles as a Function of Surface Charge. *J. Am. Chem. Soc.* **130**, 1328–1334 (2008).
52. Banerjee, D., Liu, A. P., Voss, N. R., Schmid, S. L. & Finn, M. G. Multivalent Display and Receptor-Mediated Endocytosis of Transferrin on Virus-Like Particles. *ChemBioChem* **11**, 1273–1279 (2010).
53. Uchida, M. *et al.* Targeting of Cancer Cells with Ferrimagnetic Ferritin Cage Nanoparticles. *J. Am. Chem. Soc.* **128**, 16626–16633 (2006).
54. Lin, X. *et al.* Chimeric Ferritin Nanocages for Multiple Function Loading and Multimodal Imaging. *Nano Lett.* **11**, 814–819 (2011).
55. Li, X. *et al.* Epidermal Growth Factor-Ferritin H-Chain Protein Nanoparticles for Tumor Active Targeting. *Small* **8**, 2505–2514 (2012).
56. Zhen, Z. *et al.* RGD-Modified Apoferritin Nanoparticles for Efficient Drug Delivery to Tumors. *ACS Nano* **7**, 4830–4837 (2013).
57. Khoshnejad, M., Parhiz, H., Shuvaev, V. V., Dmochowski, I. J. & Muzykantov, V. R. Ferritin-based drug delivery systems: Hybrid nanocarriers for vascular immunotargeting. *J. Controlled Release* **282**, 13–24 (2018).
58. He, J., Fan, K. & Yan, X. Ferritin drug carrier (FDC) for tumor targeting therapy. *J. Controlled Release* **311–312**, 288–300 (2019).
59. Flenniken, M. L. *et al.* Selective attachment and release of a chemotherapeutic

agent from the interior of a protein cage architecture. *Chem. Commun.* 447 (2005) doi:10.1039/b413435d.

60. Guan, X., Chang, Y., Sun, J., Song, J. & Xie, Y. Engineered Hsp Protein Nanocages for siRNA Delivery. *Macromol. Biosci.* **18**, 1800013 (2018).

61. Wagner, H. J., Capitain, C. C., Richter, K., Nessling, M. & Mampel, J. Engineering bacterial microcompartments with heterologous enzyme cargos. *Eng. Life Sci.* **17**, 36–46 (2017).

62. Rurup, W. F., Snijder, J., Koay, M. S. T., Heck, A. J. R. & Cornelissen, J. J. L. M. Self-Sorting of Foreign Proteins in a Bacterial Nanocompartment. *J. Am. Chem. Soc.* **136**, 3828–3832 (2014).

63. Moon, H., Lee, J., Min, J. & Kang, S. Developing Genetically Engineered Encapsulin Protein Cage Nanoparticles as a Targeted Delivery Nanoplatform. *Biomacromolecules* **15**, 3794–3801 (2014).

64. Seebeck, F. P., Woycechowsky, K. J., Zhuang, W., Rabe, J. P. & Hilvert, D. A Simple Tagging System for Protein Encapsulation. *J. Am. Chem. Soc.* **128**, 4516–4517 (2006).

65. Gabashvili, A. N. *et al.* Encapsulins—Bacterial Protein Nanocompartments: Structure, Properties, and Application. *Biomolecules* **10**, 966 (2020).

66. Erlendsson, S. *et al.* Structures of virus-like capsids formed by the Drosophila neuronal Arc proteins. *Nat. Neurosci.* **23**, 172–175 (2020).

67. Han, X. & Woycechowsky, K. J. Encapsulation and Controlled Release of Protein Guests by the *Bacillus subtilis* Lumazine Synthase Capsid. *Biochemistry* **56**, 6211–6220 (2017).

68. Azuma, Y., Zschoche, R. & Hilvert, D. The C-terminal peptide of *Aquifex aeolicus* riboflavin synthase directs encapsulation of native and foreign guests by a cage-forming lumazine synthase. *J. Biol. Chem.* **292**, 10321–10327 (2017).

69. Lilavivat, S., Sardar, D., Jana, S., Thomas, G. C. & Woycechowsky, K. J. *In Vivo* Encapsulation of Nucleic Acids Using an Engineered Nonviral Protein Capsid. *J.*

Am. Chem. Soc. **134**, 13152–13155 (2012).

70. Azuma, Y., Edwardson, T. G. W., Terasaka, N. & Hilvert, D. Modular Protein Cages for Size-Selective RNA Packaging in Vivo. *J. Am. Chem. Soc.* **140**, 566–569 (2018).

71. Zhang, X., Meining, W., Fischer, M., Bacher, A. & Ladenstein, R. X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 Å resolution: determinants of thermostability revealed from structural comparisons. *J. Mol. Biol.* **306**, 1099–1114 (2001).

72. Min, J., Kim, S., Lee, J. & Kang, S. Lumazine synthase protein cage nanoparticles as modular delivery platforms for targeted drug delivery. *RSC Adv* **4**, 48596–48600 (2014).

73. Kim, H., Kang, Y. J., Min, J., Choi, H. & Kang, S. Development of an antibody-binding modular nanoplatform for antibody-guided targeted cell imaging and delivery. *RSC Adv.* **6**, 19208–19213 (2016).

74. Wörsdörfer, B., Pianowski, Z. & Hilvert, D. Efficient in Vitro Encapsulation of Protein Cargo by an Engineered Protein Container. *J. Am. Chem. Soc.* **134**, 909–911 (2012).

75. Terasaka, N., Azuma, Y. & Hilvert, D. Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. *Proc. Natl. Acad. Sci.* **115**, 5432–5437 (2018).

76. Worsdorfer, B., Woycechowsky, K. J. & Hilvert, D. Directed Evolution of a Protein Container. *Science* **331**, 589–592 (2011).

77. Sasaki, E. *et al.* Structure and assembly of scalable porous protein cages. *Nat. Commun.* **8**, 14663 (2017).

78. Tanaka, Y. *et al.* Structural basis for the drug extrusion mechanism by a MATE multidrug transporter. *Nature* **496**, 247–251 (2013).

79. Kodan, A. *et al.* Structural basis for gating mechanisms of a eukaryotic P-

glycoprotein homolog. *Proc. Natl. Acad. Sci.* **111**, 4049–4054 (2014).

80. Matsunaga, Y., Bashiruddin, N. K., Kitago, Y., Takagi, J. & Suga, H. Allosteric Inhibition of a Semaphorin 4D Receptor Plexin B1 by a High-Affinity Macrocyclic Peptide. *Cell Chem. Biol.* **23**, 1341–1350 (2016).
81. Morimoto, J., Hayashi, Y. & Suga, H. Discovery of Macrocyclic Peptides Armed with a Mechanism-Based Warhead: Isoform-Selective Inhibition of Human Deacetylase SIRT2. *Angew. Chem. Int. Ed.* **51**, 3423–3427 (2012).
82. Kawamura, A. *et al.* Highly selective inhibition of histone demethylases by de novo macrocyclic peptides. *Nat. Commun.* **8**, 14773 (2017).
83. Yamagishi, Y. Natural Product-Like Macrocyclic N-Methyl-Peptide Inhibitors against a Ubiquitin Ligase Uncovered from a Ribosome-Expressed De Novo Library. **18**, 1562–1570 (2011).
84. Hayashi, Y., Morimoto, J. & Suga, H. *In Vitro* Selection of Anti-Akt2 Thioether-Macrocyclic Peptides Leading to Isoform-Selective Inhibitors. *ACS Chem. Biol.* **7**, 607–613 (2012).
85. Yu, H. *et al.* Macrocycle peptides delineate locked-open inhibition mechanism for microorganism phosphoglycerate mutases. *Nat. Commun.* **8**, 14932 (2017).
86. Jongkees, S. A. K. Rapid Discovery of Potent and Selective Glycosidase-Inhibiting De Novo Peptides. *Cell Chem. Biol.* **24**, 381–390 (2017).
87. Goto, Y. *et al.* Reprogramming the Translation Initiation for the Synthesis of Physiologically Stable Cyclic Peptides. **3**, 120–129 (2008).
88. Vinogradov, A. A., Yin, Y. & Suga, H. Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges. *J. Am. Chem. Soc.* **141**, 4167–4181 (2019).
89. Mihara, E. *et al. Lasso-grafting of macrocyclic peptide pharmacophores yields multi-functional proteins*. https://www.researchsquare.com/article/rs-85705/v1 (2020) doi:10.21203/rs.3.rs-85705/v1.
90. Azuma, Y., Herger, M. & Hilvert, D. Diversification of Protein Cage Structure

Using Circularly Permuted Subunits. *J. Am. Chem. Soc.* **140**, 558–561 (2018).

91. Fujii, Y. *et al.* PA tag: A versatile protein tagging system using a super high affinity antibody against a dodecapeptide derived from human podoplanin. *Protein Expr. Purif.* **95**, 240–247 (2014).
92. Mazzone, M. *et al.* An uncleavable form of pro–scatter factor suppresses tumor growth and dissemination in mice. *J. Clin. Invest.* **114**, 1418–1432 (2004).
93. Gherardi, E. *et al.* Structural basis of hepatocyte growth factor/scatter factor and MET signalling. *Proc. Natl. Acad. Sci.* **103**, 4046–4051 (2006).
94. Miao, W. *et al.* Impaired ligand-dependent MET activation caused by an extracellular SEMA domain missense mutation in lung cancer. *Cancer Sci.* **110**, 3340–3349 (2019).
95. Eklund, H. *et al.* Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *J. Mol. Biol.* **102**, 27–59 (1976).
96. Eklund, H. *et al.* Structure of a triclinic ternary complex of horse liver alcohol dehydrogenase at 2.9 Å resolution. *J. Mol. Biol.* **146**, 561–587 (1981).
97. Eklund, H. *et al.* Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *J. Mol. Biol.* **102**, 27–59 (1976).
98. Eklund, H. *et al.* Structure of a triclinic ternary complex of horse liver alcohol dehydrogenase at 2.9 Å resolution. *J. Mol. Biol.* **146**, 561–587 (1981).
99. Akagawa, K., Fujiwara, T., Sakamoto, S. & Kudo, K. One-pot sequential alcohol oxidation and asymmetric α-oxyamination in aqueous media using recyclable resin-supported peptide catalyst. *Chem. Commun.* **46**, 8040 (2010).
100. Raj, S. B., Ramaswamy, S. & Plapp, B. V. Yeast Alcohol Dehydrogenase Structure and Catalysis. *Biochemistry* **53**, 5791–5803 (2014).
101. James P., F., Aubrey R., H. J., Rihe, L. & Leslie E., O. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59–61 (1996).
102. Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci.* **86**, 7054–7058 (1989).

103. Bernhardt, H. S. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)a. *Biol. Direct* **7**, 23 (2012).
104. Takatsuji, R. *et al.* Ribosomal Synthesis of Backbone-Cyclic Peptides Compatible with In Vitro Display. *J. Am. Chem. Soc.* **141**, 2279–2287 (2019).
105. Ryota, Y. *In Vitro Selection* of Peptide Catalysts with Alcohol Dehydrogenase Activity. *Master Thesis* (2014).
106. Doak, B. C., Zheng, J., Dobritzsch, D. & Kihlberg, J. How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets. *J. Med. Chem.* **59**, 2312–2327 (2016).
107. Fosgerau, K. & Hoffmann, T. Peptide therapeutics: current status and future directions. *Drug Discov. Today* **20**, 122–128 (2015).
108. Rüegger, A. *et al.* Cyclosporin A, ein immunsuppressiv wirksamer Peptidmetabolit aus *Trichoderma polysporum* (L INK ex P ERS .) *Rifai*. *Helv. Chim. Acta* **59**, 1075–1092 (1976).
109. Wenger, R. M. Synthesis of Cyclosporine and Analogues: Structural Requirements for Immunosuppressive Activity. *Angew. Chem. Int. Ed. Engl.* **24**, 77–85 (1985).
110. Liu, J. *et al.* Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* **66**, 807–815 (1991).
111. Liu, J. *et al.* Inhibition of T cell signaling by immunophilin-ligand complexes correlates with loss of calcineurin phosphatase activity. *Biochemistry* **31**, 3896–3901 (1992).
112. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settingsq. *Adv. Drug Deliv. Rev.* 24 (2001).
113. Refsgaard, H. H. F. *et al.* In Silico Prediction of Membrane Permeability from Calculated Molecular Parameters. *J. Med. Chem.* **48**, 805–811 (2005).
114. Ekins, S. & Rose, J. In silico ADME/Tox: the state of the art. *J. Mol. Graph.*

Model. **20**, 305–309 (2002).

115. Ekins, S. *et al.* Progress in predicting human ADME parameters in silico. *J. Pharmacol. Toxicol. Methods* **44**, 251–272 (2000).
116. Ekins, S. *et al.* Three-Dimensional Quantitative Structure-Permeability Relationship Analysis for a Series of Inhibitors of Rhinovirus Replication. *J. Chem. Inf. Comput. Sci.* **41**, 1578–1586 (2001).
117. Mälkiä, A., Murtomäki, L., Urtti, A. & Kontturi, K. Drug permeation in biomembranes. *Eur. J. Pharm. Sci.* **23**, 13–47 (2004).
118. Fujikawa, M., Ano, R., Nakao, K., Shimizu, R. & Akamatsu, M. Relationships between structure and high-throughput screening permeability of diverse drugs with artificial membranes: Application to prediction of Caco-2 cell permeability. *Bioorg. Med. Chem.* **13**, 4721–4732 (2005).
119. Veber, D. F. *et al.* Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).
120. Leung, S. S. F., Mijalkovic, J., Borrelli, K. & Jacobson, M. P. Testing Physical Models of Passive Membrane Permeation. *J. Chem. Inf. Model.* **52**, 1621–1636 (2012).
121. Witek, J. *et al.* Kinetic Models of Cyclosporin A in Polar and Apolar Environments Reveal Multiple Congruent Conformational States. *J. Chem. Inf. Model.* **56**, 1547–1562 (2016).
122. Rezai, T. *et al.* Conformational Flexibility, Internal Hydrogen Bonding, and Passive Membrane Permeability: Successful in Silico Prediction of the Relative Permeabilities of Cyclic Peptides. *J. Am. Chem. Soc.* **128**, 14073–14080 (2006).
123. Witek, J. *et al.* Rationalization of the Membrane Permeability Differences in a Series of Analogue Cyclic Decapeptides. *J. Chem. Inf. Model.* **59**, 294–308 (2019).
124. Poger, D., Van Gunsteren, W. F. & Mark, A. E. A new force field for simulating phosphatidylcholine bilayers. *J. Comput. Chem.* **31**, 1117–1125 (2010).

125. Nathan, S. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J* **40**, 843–856 (2011).
126. Hritz, J. & Oostenbrink, C. Hamiltonian replica exchange molecular dynamics using soft-core interactions. *J. Chem. Phys.* **128**, 144121 (2008).
127. Meli, M. & Colombo, G. A Hamiltonian Replica Exchange Molecular Dynamics (MD) Method for the Study of Folding, Based on the Analysis of the Stabilization Determinants of Proteins. *Int J Mol Sci* 13 (2013).

List of accomplishments

【Publication related to the thesis】

1. “In vitro selection of an alcohol dehydrogenase peptidic enzyme”, Yamato Komatsu, Ryota Yokosuka, Yuki Goto, Hiroaki Suga, *Peptide Science*, **2020**, 99-100

【Publications not related to the thesis】

1. “Ribosomal Synthesis of Backbone-Cyclic Peptides Compatible with In Vitro Display”, Ryo Takatsuji, Koki Shinbara, Takayuki Katoh, Yuki Goto, Toby Passioura, Ryo Yajima, Yamato Komatsu, and Hiroaki Suga, *J. Am. Chem. Soc.,* **2019**, 141, 2279-2287

2. 小松 大和, 後藤 佑樹, 菅 裕明 「擬天然ペプチドの試験管内人工生合成系による合成技術及びその医薬品候補分子探索への応用」,『第 80 号酵素工学研究会トピックス』, 酵素工学研究会, p23-28 (2018)

【Poster presentations】

1. “タンパク質酵素の起源に迫る!?アルコール酸化活性を持つペプチド酵素のスクリーニング系構築”, 小松 大和, 後藤 祐樹, 菅 裕明, 若手ペプチド夏の勉強会, 長崎, 2017 年 8 月

2. “アルコール脱水素触媒活性を有する短鎖ペプチド酵素の試験管内セレクション”, 小松 大和, 横須賀 亮太, 後藤 祐樹, 菅 裕明, 日本ケミカルバイオロジー学会（JSCB）第 13 回年会, 東京, 2018 年 6 月

3. “In vitro selection of an alcohol dehydrogenase peptidic enzyme”, Yamato Komatsu, Ryota Yokosuka, Yuki Goto, Hiroaki Suga, 第 56 回ペプチド討論会, 東京, 2019 年, 10 月

Acknowledgements

This thesis was completed under the supervision of Prof. Hiroaki Suga. I would like to express my sincere gratitude to Prof. Suga for his kind guidance, valuable suggestions and encouragement throughout this work. I would like to thank Asst. Prof. Naohiro Terasaka and Assoc. Prof. Yuki Goto for valuable discussions and teaching me various concepts and techniques for organic chemistry and biochemistry. I also thank Assoc. Prof. Takayuki Kato, Assoc. Prof. Naokazu Kano and Asst. Prof. Toby Passioura for their helpful advice and discussions. I thank all members in Suga Lab.

I appreciate the work done by Prof. Kunio Matsumoto's group in Kanazawa University for biological assays in chapter 2 and 3. I also really thank Prof. Riniker Sereina's group, especially Prof. Riniker Sereina and Dr. König Gerhard, for the usage authorization and lecture of GROMOS software in chapter 5.

I thank Materials Education program for the future leaders in Research, Insudtry, and Technology (MERIT) for financial support. I also thank Japan Society for the Promotion of Science (JSPS) for a grant and financial support.

Finally, I would like to thank my family for invaluable assistance and cheering words.