論文 / 著書情報
Article / Book Information

題目(和文)	再構成型無細胞タンパク質合成系を利用した機能性ポリペプチドの創出
Title(English)	Exploring sequence space of functional de novo polypeptides using a reconstituted cell-free system
著者(和文)	西川将太
Author(English)	Shota Nishikawa
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12749号, 授与年月日:2024年3月26日, 学位の種別:課程博士, 審査員:藤島 皓介,松浦 友亮,田口 英樹,丸山 厚,相澤 康則
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12749号, Conferred date:2024/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Exploring sequence space of functional *de novo* polypeptides using a reconstituted cell-free system

再構成型無細胞タンパク質合成系を利用した機能性ポリペプチドの創出

Shota Nishikawa

Doctor of Philosophy

School of Life Science and Technology

Tokyo Institute of Technology

2024

Abstract

In biology, polypeptides exhibit various physiochemical properties as they interact with other biomolecules and catalyze numerous biochemical reactions. Recent advances in synthetic biology have enabled the exploration of sequence space of functional *de novo* polypeptides. Especially, a reconstituted cell-free protein synthesis system (*i.e.*, PURE system) allows protein synthesis in a high-throughput way; thus, along with directed evolution techniques, the PURE system has enabled facile and efficient characterization of *de novo* functional polypeptides. The major goal of this thesis is to ***(i)*** develop a renovated mRNA display method for high-throughput identification of *de novo* peptides targeting single-stranded RNA with low complexity and ***(ii)*** open a new field in the cell-free system to handle O_2-labile metalloenzymes, particularly those employing iron-sulfur (Fe-S) cluster for their catalysis enabling post-translational modification of peptides. The tools developed in this thesis would enable the design and development of polypeptides that remain unexplored in nature and may even facilitate the future generation of a post-translational modified peptide library.

RNA-binding proteins participate in diverse cellular processes, including DNA repair, post-transcriptional modification, and cancer progression through their interactions with RNAs, making them attractive for biotechnological applications. While nature offers a variety of naturally occurring RNA-binding proteins, creating *de novo* RNA-binding peptides remains challenging due to the limited availability of structural information on RNA molecules and their modes of interaction with peptides. In particular, tailoring peptides to target single-stranded RNA with low complexity is challenging due to the inherent structural flexibility of RNA molecules. In this thesis, I have developed a codon-restricted mRNA display and identified multiple *de novo* peptides from a peptide library that bind to poly(C) and poly(A) RNA with

K_Ds ranging from micromolar to submicromolar concentrations. One of the newly identified peptides is capable of binding to the cytosine-rich sequences of the oncogenic Cdk6 3′UTR RNA and *MYU* lncRNA with an affinity comparable to that of the endogenous binding protein. Hence, I developed a novel platform for discovering *de novo* single-stranded RNA-binding peptides that offer promising avenues for regulating RNA functions.

Furthermore, to broaden the chemical properties of the peptide library generated *via* mRNA display, I have further explored the cell-free synthesis of Fe-S cluster proteins. The goal is to utilize the radical S-adenosyl-l-methionine (rSAM) enzymes for post-translational modification. However, the O_2-lability of the Fe-S cluster proteins has been a long-standing issue for applications in cell-free systems. In this thesis, I have developed a facile and efficient protocol to express mature Fe-S cluster proteins in the PURE system under aerobic conditions. To achieve this, one of the widely conserved Fe-S cluster assembly systems, known as the SUF system, has been reconstituted *in vitro* using a total of six recombinant SUF subunits (SufABCDSE) individually purified from *E. coli*. The incorporation of recombinant SUF proteins into the PURE system enabled mRNA translation-coupled Fe-S cluster assembly under the O_2-depleted conditions. To achieve a benchtop synthesis of mature Fe-S cluster proteins, an O_2-scavenging enzyme cascade was incorporated into the PURE system. The cascade begins with formate oxidation by formate dehydrogenase for NADH regeneration. The regenerated NADH is then consumed by flavin reductase for $FADH_2$ regeneration. Finally, bifunctional flavin reductase, along with catalase, removes O_2 from the solution while supplying $FADH_2$ to the SUF system, which is required for *de novo* Fe-S cluster assembly. These amendments

enabled a one-pot, two-step synthesis of mature Fe-S cluster proteins under aerobic conditions, achieving >90% of maturation efficiency.

In conclusion, this thesis reports a novel platform to discover *de novo* RNA-binding peptides as well as the renovated PURE system for the synthesis of O_2-labile Fe-S cluster proteins under aerobic conditions. Combining the two techniques developed in this thesis greatly expands the toolbox of biotechnology as well as the potential of the PURE system, paving the way for the future generation of post-translational modified peptide libraries *in vitro*.

Table of contents

Acknowledgements

Firstly, I would like to express my deepest gratitude to all the people who have helped me in my life and in the completion of this doctoral thesis, especially my family. With their support, I could pursue my interests and enjoy my life. Secondly, I am immensely thankful to my advisor, Prof. Kosuke Fujishima, for his unwavering guidance, invaluable insights, and constant encouragement throughout the entire research journey. His mentorship has been instrumental in shaping the direction of this work. Thirdly, I would like to acknowledge Prof. Po-Hsiang Wang at National Central University, Taiwan, for teaching me the essence of scientific thinking and experimental skills. I also would like to acknowledge my doctoral committee members, Prof. Hideki Taguchi, Prof. Atsushi Maruyama, Prof. Tomoaki Matsuura, and Prof. Yasunori Aizawa, for their valuable feedback and constructive criticism, which have greatly enriched the quality of this thesis. Special thanks go to my colleagues who have shared their expertise and engaged in insightful discussions, contributing to the intellectual stimulation of this research, especially, Mikhail Makarov and Prof. Wang's lab members; Jhen-Sheng Yan, Yi-Hsuan Lin, and Fang-I Yeh. Finally, I want to express my gratitude to my friends for offering encouragement and understanding during the challenging phases of this journey.

List of tables

List of figures

List of abbreviations

EDTA – ethylenediaminetetraacetic acid

FAD – flavin adenine dinucleotide

IPTG – Isopropyl β-D-1-thiogalactopyranoside

ESI-LC-MS – electrospray ionization-liquid chromatography-mass spectrometry

NAD – nicotinamide adenine dinucleotide

PCR – polymerase chain reaction

SDS-PAGE – sodium dodecyl sulfate-polyacrylamide gel electrophoresis

UV-vis – ultraviolet-visible

Chapter 1 – General introduction

This chapter includes a general introduction that covers previous studies regarding RNA-binding proteins, cell-free system, directed evolution, ribosomally synthesized and post-translationally modified peptides, and biosynthesis of the iron-sulfur cluster (**Section 1.1**). Based on the general introduction, two major research objectives, **Objective 1:** Develop methods to identify single-stranded RNA-binding peptides from random peptide sequence space using a renovated method combining codon-restricted DNA library and mRNA display and **Objective 2:** Develop methods to produce O2-labile [4Fe-4S] proteins under aerobic conditions using a reconstituted cell-free protein synthesis system, were proposed (**Section 1.2**). Finally, this chapter outlines the overall thesis structure (**Section 1.3**), along with the statement of authorship and publication status of the results reported in this thesis (**Section 1.4**).

1.1 Literature review

1.1.1 RNA-binding proteins.

A conventional RNA-binding protein (RBP) actively contributes to the assembly of ribonucleoprotein (RNP) complexes primarily participating in gene expression. This involvement occurs through the binding of the RBP to specific sequence and/or structural motifs within RNA, facilitated by modular combinations of well-defined RNA-binding domains (RBDs) (*e.g.*, RNA recognition motif (RRM) and hnRNP K homology domain (KH)) (**Figure 1-1**) (Beuth et al. 2005; Pérez-Cañadillas 2006). These insights, rooted in decades of cumulative knowledge encompassing cellular, biochemical, and structural data, have shaped

our understanding of gene expression processes. Nevertheless, recent studies have revealed intricate protein-RNA interactions that defy the conventional interactions observed in canonical RBDs (Albihlal and Gerber 2018). These results indicates that unconventional RNA binding constitutes a more widespread phenomenon than previously envisioned. The emerging paradigm challenges traditional perspectives and underscores the complexity inherent in RNA-protein interactions within cellular processes.

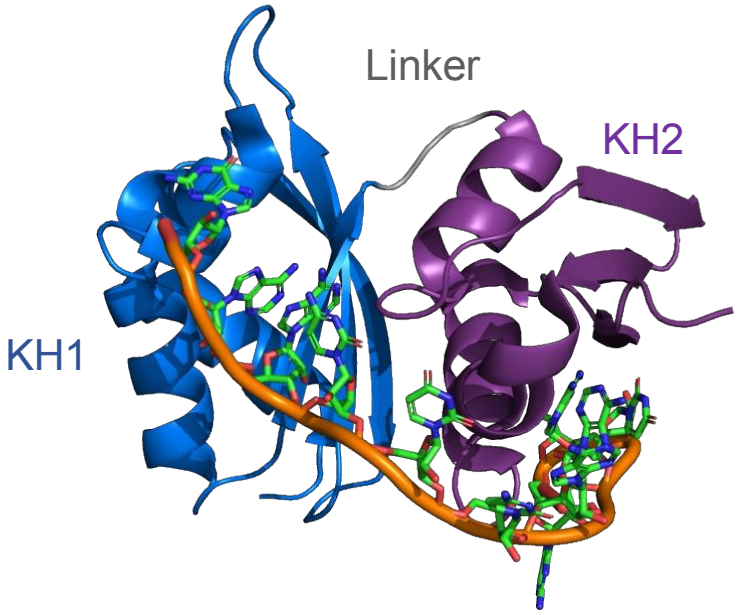


Figure 1-1. Structure of a *Mycobacterium tuberculosis* NusA–RNA complex. The crystal structure was obtained from RCSB protein data bank (PDB) (PDB ID: 2ASB). KH1 domain colored in blue, KH2 domain colored in purple, linker region colored in gray, RNA molecule in sticks.

So far, over 2000 RBPs have been reported (Corley, Burns, and Yeo 2020). The interaction modes of RBPs between RNA molecules are highly diverse owing to their ability to achieve nucleotide sequence-specific interactions through the combinations of multiple RNA-binding domains (RBDs) along with intrinsically disordered regions. Based on the literature, up until now, only about 17 RNA-binding domains with known interaction patterns have been characterized in detail (**Table 1-1**) (Corley, Burns, and Yeo 2020). RBPs typically harbor multiple RBDs, utilizing specific amino acid residues for direct interaction with RNA molecules. The interactions between these amino acid residues and RNA molecules involve multiple interactions such as hydrogen bonding, electrostatic interactions, stacking interactions, and hydrophobic interactions, enabling RNA sequence-specific recognition. Typically, the coexistence of multiple binding domains contributes to an increase in RNA-binding affinity (Lunde, Moore, and Varani 2007). This multi-step approach allows for nearly limitless combinations (although the variety is limited by the number of RBDs and the protein size), resulting in hundreds of different RNA-binding modes with diverse functions. Among them, the KH domain (K homology domain) has recently gained attention due to the highly conserved RNA-binding motif with primordial RNA-protein interactions (Pereira and Lupas 2018). The KH domain consisting of approximately 70 residues binds single-stranded nucleic acids and are built around a 45 residues carrying a conserved sequence composed of hydrophobic amino acids ((I/L/V)IGxxGxx(I/L/V)) which is highly conserved from prokaryotes to eukaryotes and responsible for RNA binding (Hollingworth et al. 2012).

The completion of the human genome project marked a significant turning point, revealing that while approximately three-quarters of the genome sequence is transcribed into

RNA, only a 1% encodes proteins (Hopkins and Groom 2002). This discovery highlighted the functional importance of long non-coding RNA (lncRNA) transcripts and their integral role in the extensive RNA-based regulatory networks influencing cellular processes (Wahlestedt 2013).

Humans are estimated to produce over 15,000 lncRNAs, some of which may emerge as viable drug targets (Harrow et al. 2012). This suggests that a significant number of mRNAs and non-coding RNAs could potentially expand the range of the human genome amenable to drug therapy. Considering RNA's pivotal role in most cellular functions, it presents a plethora of therapeutic opportunities. Experiments and clinical trials utilizing antisense oligonucleotides have established a robust proof of concept for RNA-targeting drugs (Liang et al. 2016, 2017). Furthermore, the potent use of synthetic RNAs to regulate the cellular RNA interference (RNAi) pathway (Fellmann and Lowe 2014) or CRISPR-based gene-editing systems (Fellmann et al. 2017), further validating the therapeutic promise of RNA-targeting therapy. These approaches can either suppress or enhance the expression of specific genes and can also modulate non-coding RNA functions. Currently, 11 RNA-based therapeutics have obtained clinical approval (Winkle et al. 2021). Despite their promise, these oligonucleotide-based approaches, which use large and often highly charged molecules, face significant delivery challenges due to the difficulty of passing through the plasma membrane and avoiding degradation (T. C. Roberts, Langer, and Wood 2020). Therefore, a small-molecule or peptide therapy that targets RNA would be preferable if it can be applied.

A number of small molecules capable of modulating RNA function have been identified, suggesting promising avenues for the development of RNA-targeting small molecule therapeutics. For example, small-molecule antibiotics (*e.g.*, streptomycin) produced by various

bacterial species have been found to bind to rRNA and disrupt the process of protein translation (Wilson 2014). Furthermore, there has been an increase in the identification of synthetic molecules that are capable of binding to RNA secondary structures. For example, the expanded repeat of *C9ORF72* RNA is the most common genetic cause of frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS). The expanded repeat adopts a hairpin structure in equilibrium with a quadruplex structure, and thus, bioactive small molecules targeting the secondary structures were designed and found to significantly ameliorate c9FTD/ALS-associated defects in a cellular model (Su et al. 2014). These molecules have demonstrated the ability to alter biological functions, indicating potential for therapeutic applications. However, for successful interaction, it is crucial that the RNA adopts a stable tertiary structure, creating pockets akin to those found in proteins. Nonetheless, targeting RNAs lacking these stable structures with small molecules remains a challenging task.

This strategy indicates that peptides, with their broader interaction surface, could be particularly effective in targeting RNA secondary structures. The broader interaction surface of peptides enables them to bind flexibly to RNA structures, resulting in high potency and specificity. This has been supported by various reports of RNA-binding peptides identified over the past two decades, particularly through phage display methods—a type of directed evolution (explained in detail in **section 1.1.3**) (Gc et al. 2019; Dremann and Chow 2016; Kaur et al. 2013; Lamichhane et al. 2011). For example, a 7-mer peptide discovered *via* phage display attaches to the 16S helix 34 of ribosomal RNA (rRNA), demonstrating a low micromolar dissociation constant (Gc et al. 2019). This peptide alters the structure of the rRNA upon binding, inhibiting the RsmC enzyme, which is responsible for the methylation of G1207's

exocyclic amine in rRNA. The introduction of this peptide was also observed to inhibit bacterial growth. This example suggests that peptides can be engineered to bind RNA effectively. With the appropriate approach and target, these molecules could achieve the necessary potency and selectivity for therapeutic use. By exploring the vast peptide sequence space with a combinatorial peptide library, we can investigate a wide range of RNA-binding peptides, potentially leading to the identification of therapeutic candidates.

Table 1-1. The list of RNA-binding domains (modified from Corley, Burns, and Yeo 2020).

Domain Name	PDB Structures	PDB Sructures with RNA	Size (a.a.)	Protein Families Containing Domain	Target RNA
Cold shock domain (CSD)	46	13	70	Cold shock proteins, Y-box proteins	ssRNA, ssDNA
Double-stranded RNA Binding Domain (dsRBD)	59	27	65	RNases, ADARs, Dicer	dsRNA (A-form RNA helix)
Helicase	701	114	350-400	DExH/D-box, Ski2-like, RIGI-like, NS3, UPF1-like RNA binding helicases	dsRNA, dsDNA
Intrinsically disordered region (IDR)	NA	NA	varies	Most RBPs	ssRNA, dsRNA, ssDNA, dsDNA
K homology (KH)	117	20	70	hnRNPs, translation regulation proteins, very common	ssRNA, ssDNA
La motif (LAM)	54	5	90	La proteins, La-related proteins (LARPs)	ssRNA
Piwi-Argonaute-Zwille (PAZ)	71	44	170	Argonaute proteins, Dicer	ssRNA (siRNA, gRNA)
P-element Induced Wimpy Testis (PIWI)	74	33	290	Argonaute proteins	ssRNA (siRNA, gRNA)
Pentatricopeptide repeat (PPR)	32	9	35 *n, 1 > n > 30	RNA editing proteins	ssRNA
Pseudouridine synthase and archaeosine transglycoslyase (PUA)	119	28	66-98	RNA modifying enzymes, metabolic enzymes	dsRNA
Pumillo-like repeat (PUM)	60	49	334	PUF proteins	ssRNA
Ribosomal S1-like (S1)	25	2	70	Ribosomal proteins, Translation initiation factors, RNase II, PNPase	ssRNA (mRNA), dsRNA (rRNA)
RNA Recognition Motif (RRM)	554	119	90	hnRNPs, splicing factors, very common	ssRNA
Sm and Like-Sm (Sm / Lsm)	31	23	80	U1 spliceosomal proteins, Hfq	ssRNA
thiouridine synthases, RNA methylases and pseudouridine synthases (THUMP)	12	3	100-110	tRNA modifying enzymes	dsRNA (tRNA)
YT521-B homology (YTH)	28	14	100-150	YTH family m^6A readers	N6-methyladenosine (m6A)-containing RNA
Zinc Finger (ZnF)	2677	63	30	Transcription factors, METTL enzymes, Very common	ssRNA, dsRNA, ssDNA, dsDNA

1.1.2 Cell-free system.

Over the past two decades, cell-free protein synthesis (CFPS) systems have emerged as a foundational platform in the field of synthetic biology. Traditional cell-based recombinant protein expression systems have been employed for the synthesis and characterization of a variety of proteins, as well as to produce valuable metabolites through the modification of cellular metabolic pathways (Nielsen and Keasling 2016). However, *in vivo* protein expression systems are constrained by the inherent limitations of living systems. Living systems operate effectively only within narrow ranges of conditions, including factors like pH, temperature, salt concentration, and solvent properties. Product toxicity further hinders the efficient production of recombinant products. In addition, the crosstalk with cellular metabolic pathways often reduces the substrate flux into synthetic pathways, resulting in low product yields and inefficient conversion rates. These challenges primarily arise from the requirement for living systems to maintain balanced homeostasis.

Complementing the well-established *in vivo* protein expression systems, CFPS system circumvents many of these limitations and facilitates robust and efficient protein production. CFPS system can be defined as a platform where protein synthesis occurs *in vitro* biochemical reactions. Unlike *in vivo* systems, the CFPS system lacks physical barriers like cell walls or membranes, allowing protein synthesis from linear DNA in customizable environments through the direct addition of ligands, cofactors, or molecular chaperones. This approach is particularly advantageous for challenging proteins with stability, solubility, folding, or assembly issues (Zigmond et al. 1997; Busso, Kim, and Kim 2004; Henrich et al. 2015; Kuruma and Ueda 2015).

The CFPS system can be classified into two types based on its preparation method: ***(i)*** cell extracts-based systems and ***(ii)*** reconstituted systems. The concept of the CFPS system was first demonstrated in 1897 when Eduard Buchner made a ground-breaking discovery. Buchner found that cell extracts derived from yeast exhibited the remarkable ability to ferment sugar independently of living yeast cells (Buchner and Rapp 1897). Nirenberg and Matthei later utilized *Escherichia coli* cell extracts to perform mRNA-dependent polypeptide synthesis and deciphered the genetic code (Nirenberg and Matthaei 1961; **Figure 1-2**). During its early stages, the CFPS systems were primarily based on "S30 extracts," which consisted of supernatants derived from various cell types through centrifugation of their lysates (Zubay 1973). The term "S30" referred to supernatants obtained after centrifugation at 30,000 × g. In the following decades, many researchers harnessed S30 extracts as a foundational platform for refining the CFPS system while the limited reaction time and inefficient protein yield have remained constraining the applications. In 1988, Spirin and his colleagues solved these problems and reported a prolonged and efficient cell-free translation system (Spirin et al. 1988). The renovated system has been achieved by continuously supplying the substrates (amino acids, ATP, and GTP) to the cell-free reaction mixture while simultaneously removing the synthesized polypeptides, enabling sustained protein synthesis. This approach was demonstrated in cell extracts derived from both prokaryotic (*Escherichia coli*) and eukaryotic (wheat germ and *Triticum* sp.) organisms. Remarkably, in both cases, the protein synthesis reaction continued for over 20 hours, consistently generating polypeptides at a high rate (hundreds of micrograms of proteins per milliliter) (Baranov et al. 1989; Endo et al. 1992). Nevertheless, the cost per individual reaction was high, making it impractical for commercial applications.

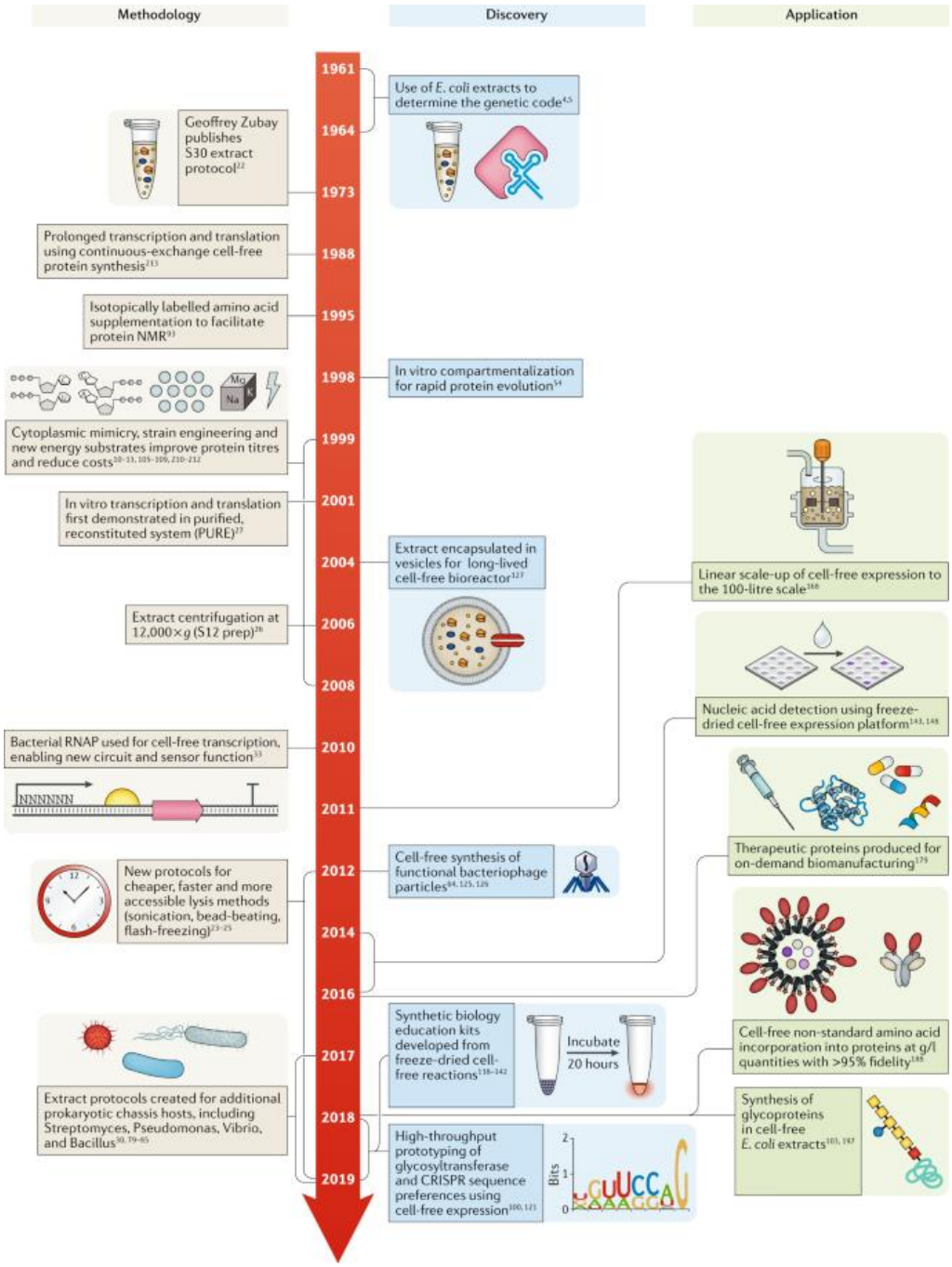


Figure 1-2. Timeline of cell-free gene expression systems. Adapted from Silverman, Karim, and Jewett 2020.

1.1.2.1 Cell extracts-based cell-free system.

Among the S30 extracts, the *Escherichia coli* cell extracts S30 has been extensively employed in CFPS system due to its quick and low-cost preparation along with efficient protein synthesis capabilities. Furthermore, the well-established genetic and metabolic engineering techniques available for *E. coli* offer opportunities to optimize the cell extracts (Jackson, Pratt, and Holland 1983; Seki, Matsuda, and Kigawa 2009).

To establish a low-cost and efficient cell-free system, Swartz and his colleagues have worked on improving the *E. coli* S30 CFPS system by modifying the endogenous metabolic pathway. They first reported that during cell-free protein synthesis reaction, phosphoenol pyruvate (PEP), the secondary energy source for ATP regeneration, and several amino acids are rapidly degraded, even in the absence of protein synthesis, which severely reduces protein yields. However, protein yields can be fully recovered by adding extra PEP and specific amino acids like arginine, cysteine, and tryptophan. Repeated additions of PEP, arginine, cysteine, and tryptophan can significantly extend the duration of protein synthesis. Subsequently, in 2001, they demonstrated that the yield of PEP-dependent protein synthesis could be more than doubled when coenzyme A and NAD^+ were added to activate endogenous ATP regeneration pathway (D.-M. Kim and Swartz 2001) (**Figure 1-3**). In this pathway, PEP is first converted to pyruvate while producing ATP via pyruvate kinase (PK). The addition of cofactors (CoA and NAD^+) further activates pyruvate dehydrogenase (PDH), which catalyzes the condensation of CoA and the produced pyruvate to synthesize acetyl-CoA and NADH; NAD^+ can be regenerated by lactate dehydrogenase (LDH). The synthesized acetyl-CoA is then converted to acetyl phosphate by phosphotransacetylase (PTA) while regenerating CoA, ultimately

generating ATP through acetate kinase (AK). The addition of these cofactors (CoA and NAD^+) activates the regeneration of ATP from pyruvate, supporting the cell-free protein synthesis reaction. This example illustrates how cell-free metabolic pathways can be adjusted to enhance protein yields, and the cofactor supplementation demonstrates the importance of cofactor recycling for prolonged cell-free protein synthesis reactions.

Until now, the *E. coli* S30 CFPS system has been utilized to produce and characterize proteins with unknown functions, including membrane proteins and glycosylated proteins, and prototype metabolic pathways for future *in vivo* metabolic engineering. However, there are limitations in the functional characterization of proteins due to background activities present in the cell extracts. Moreover, the exact composition of the S30 extracts remains unknown; yet, understanding the composition would be valuable for refining cell-free reactions, such as eliminating critical enzymes or modifying products directly in the reaction lysates.

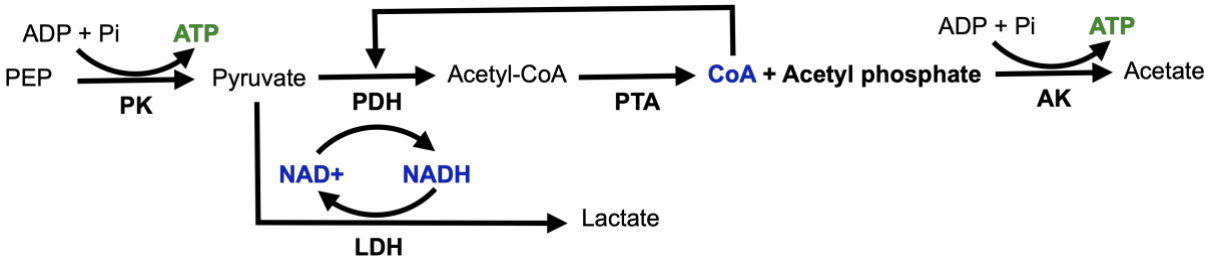


Figure 1-3. Phosphoenol pyruvate-dependent ATP regeneration pathway demonstrated in D.-M. Kim and Swartz 2001. AK, acetate kinase; LDH, lactate dehydrogenase; PDH, pyruvate dehydrogenase; PEP, phosphoenol pyruvate; Pi, inorganic phosphate; PK, pyruvate kinase; PTA, phosphotransacetylase.

1.1.2.2 Reconstituted cell-free system.

To address these challenges, the reconstituted cell-free system (PURE system) has been developed by reconstituting the *E. coli* translation machinery from individually purified components. Ueda's group at the University of Tokyo, established a fully reconstituted system with purified proteins necessary for transcription and translation. (Shimizu et al. 2001). In contrast to the cell-extracts based CFPS systems, the PURE system primarily comprises purified recombinant protein elements, providing a contamination-free alternative by minimizing unknown cellular components from the cell-free system. Recent research has focused on simplifying the system, enhancing its robustness, and lowering costs in the reconstruction process (Lavickova and Maerkl 2019) as well as modifying the energy regeneration pathway (P.-H. Wang et al. 2020). These advantages become particularly helpful when precise environmental control is needed for high-throughput functional characterization of proteins and the directed evolution of peptides/proteins (Contreras-Llano and Tan 2018; Tuckey et al. 2014).

The following three sub-sections report: ***(i)*** **directed evolution**, ***(ii)*** **peptide library with non-proteinogenic structures**, ***(iii)*** **ribosomally synthesized and post-translationally modified peptides, and** ***(iv)*** **biosynthesis of iron-sulfur cluster**.

1.1.3 Directed evolution.

Proteins follow the central dogma, where a protein's primary sequence of amino acids is determined by the genetic (DNA) and transcriptomic (RNA) information. This establishes a connection between the protein and the preceding nucleic acids. Consequently, the evolution of proteins necessitates corresponding changes in nucleic acid sequences. Advances in both *in vivo* and *in vitro* laboratory techniques, alongside next-generation sequencing (NGS), have led to the development of various selection methods for functional protein sequences, known as display methods. These techniques directly combine the genotype (nucleic acid) and phenotype (protein) (**Figure 1-4**).

For *in vivo* techniques, there are methods such as phage display (G. P. Smith 1985; Sidhu et al. 2000) and cell surface display (Francisco et al. 1993; Boder and Wittrup 1997; M. R. Smith, Khera, and Wen 2015), where protein molecules are displayed on phages or the surface of cells. These techniques have been widely utilized in creating various functional protein/peptide molecules but are often compromised by their toxicity to the host organism, significantly reducing the size of libraries, which can be vital, particularly when biological functions are evolved completely *de novo*.

In contrast, *in vitro* display technologies combined with CFPS system circumvent these constraints and offer access to vast sequence space without host interference. Generally, the primary format for *in vitro* selection is the crude cell extract-based CFPS system. For example, ribosome display using complexes mediated by ribosomes was developed by Mattheakis and colleagues using *E. coli* S30 (Mattheakis, Bhatt, and Dower 1994). In ribosome display,

complexes of mRNA, ribosomes, and proteins are formed *in vitro* by using mRNA lacking stop codons, linking the genotype with the phenotype. Typically, the dissociation of the translated peptide from the ribosome is catalyzed by a dissociation factor binding to a stop codon on the mRNA. However, in the absence of a stop codon on the mRNA, the dissociation factor cannot bind, resulting in the stalling of the ribosome on the mRNA. Consequently, a complex of mRNA-ribosome-peptide can be formed. In contrast, mRNA display links mRNA to the encoded peptide *via* a covalent bond mediated by an antibiotic, puromycin (R. W. Roberts and Szostak 1997; Nemoto et al. 1997). When polypeptide synthesis is performed using mRNA with puromycin bound to its 3' end as a template, puromycin forms a covalent bond with the C-terminal end of the nascent polypeptide, resulting in linking mRNA to the encoded protein (**Figure 1-5**). mRNA display is capable of screening the largest number of protein variants ($\sim 10^{13}$) and shows advantages towards molecular design and engineering given their *in vitro* natures (Cotten et al. 2012; Newton et al. 2020). Recently, a renovated mRNA display method called cDNA display has been developed by masking mRNA with cDNA through reverse transcription reactions (Yamaguchi et al. 2009; Mochizuki et al. 2011). The masking stabilizes the genotype into a less easily degraded double-stranded mRNA/cDNA hybrid. Meanwhile, liposome display has been developed, utilizing artificial lipid bilayers (liposomes) enabling the *in vitro* evolution of membrane proteins (Fujii et al. 2013). In this method, the gene (DNA) is encapsulated within liposomes at single-molecule level along with the PURE system. The single-gene encapsulation facilitates the genotype-phenotype linkage, allowing directed evolution.

The genotype-phenotype linkage allows peptide/protein library selection under various conditions. The selection process concludes by synthesizing a cDNA library through reverse transcription PCR from the obtained mRNA library, marking the end of one cycle of selection. By repeating the selection cycles several times, one can obtain proteins with the desired function from a vast peptide/protein sequence space. The *in vitro* display methods have been extensively used for the creation of novel antibodies (Kanamori, Fujino, and Ueda 2014; Nagumo et al. 2016), ligand proteins (Bashiruddin and Suga 2015), and especially for the discovery of *de novo* functional peptides/proteins (Keefe and Szostak 2001; Kang et al. 2015; Peacock and Suga 2021).

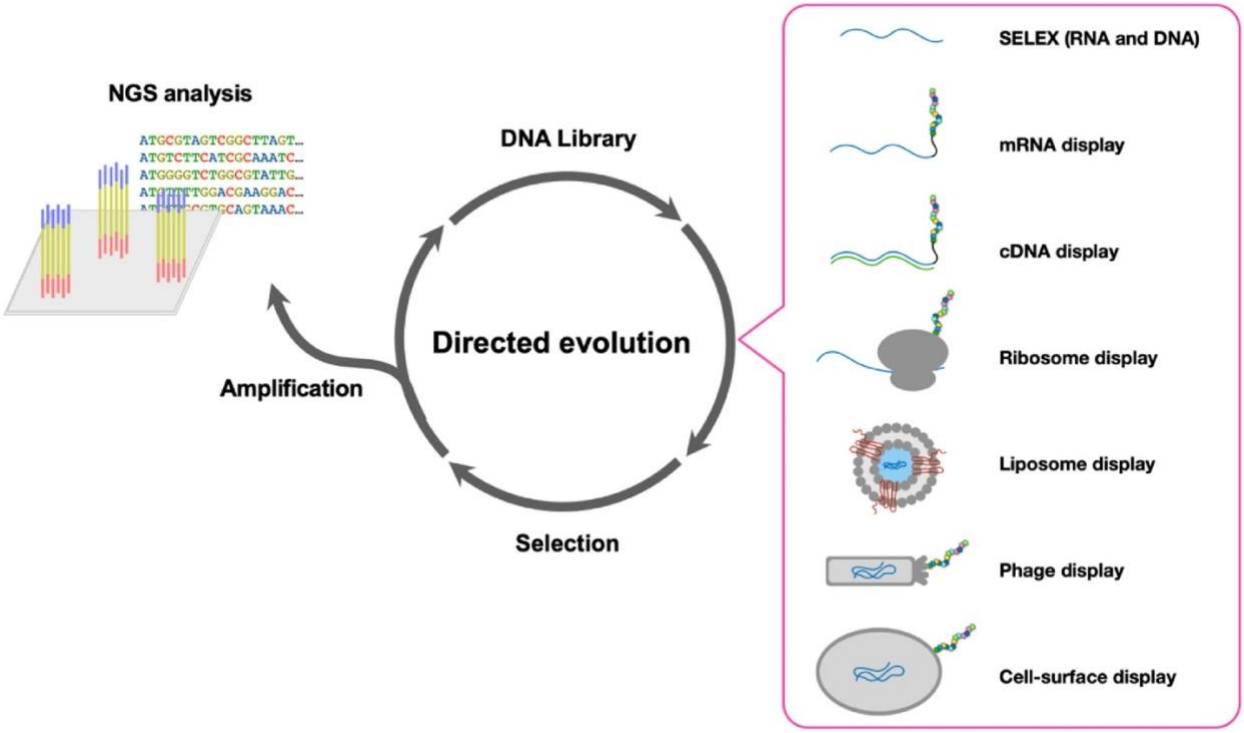


Figure 1-4. Various directed evolution strategies for peptides/proteins. The various strategies can be applied to peptides/proteins (display coupling nucleotide information to peptide information) and encompass both *in vitro* techniques (mRNA, cDNA, and liposome display and SELEX) as well as *in vivo* techniques (phage and cell-surface display). Adapted from Jia, Nishikawa, and Fujishima 2022

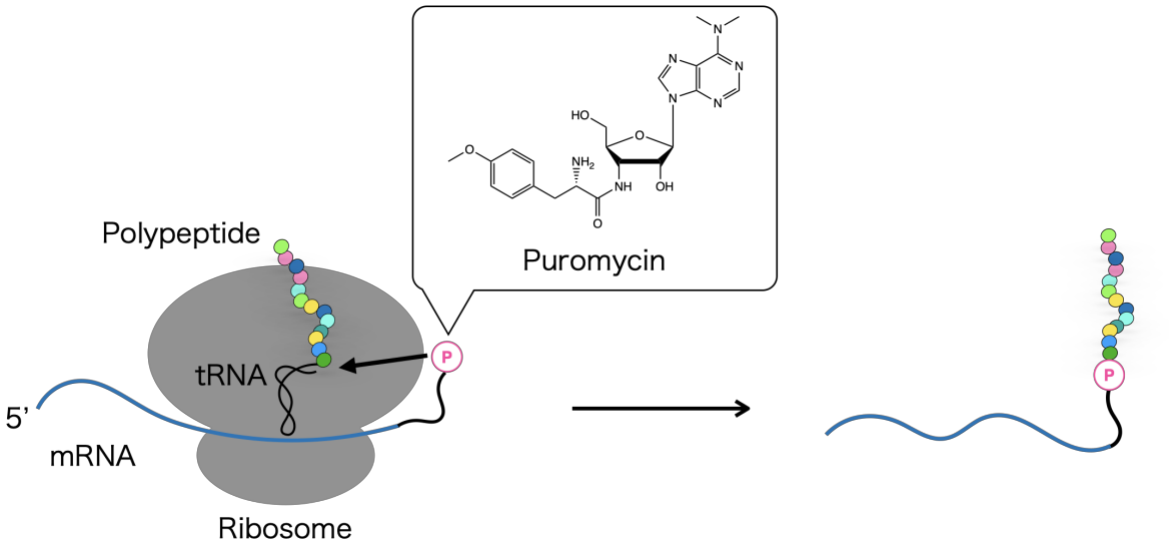


Figure 1-5. mRNA display method. Puromycin is utilized to incorporate itself into a nascent polypeptide during *in vitro* translation, leading to the formation of mRNA-peptide complexes.

1.1.3.1 Peptide library with non-proteinogenic structures.

Numerous peptides/proteins have been evolved *via in vitro* evolution (Ullman, Frigotto, and Cooley 2011). However, the limited chemical diversity inherent in canonical amino acids imposes constraints on both bioactivity and enzymatic functionality within the natural system. To push the boundaries of bioactivity and catalysis beyond what nature offers, researchers have introduced non-canonical amino acids, modified backbones, and macrocyclic frameworks into peptide and protein libraries, thereby greatly expanding the chemical diversity within these libraries (Goto and Suga 2021). Indeed, naturally occurring peptides produced by microorganisms often feature non-proteinogenic structures such as modified side chains, altered backbones, and macrocyclic frameworks (Driggers et al. 2008). These structural modifications play a pivotal role in enhancing their bioactivity by conferring resistance to proteolysis, increasing membrane permeability, and facilitating selective binding to specific target molecules. Consequently, these distinctive non-proteinogenic structures serve as the fundamental building blocks for the development of analogs of natural peptides, offering exciting prospects for the field of peptide and protein engineering.

mRNA display has been widely adopted to construct peptide libraries with non-proteinogenic structures, mainly due to the flexibility offered by CFPS system. To achieve the incorporation of non-proteinogenic structures, Flexizyme system using *in vitro* selected tRNA aminoacylating ribozyme has been developed (Goto, Katoh, and Suga 2011). These Flexizymes are capable of aminoacylating diverse non-canonical amino acids onto specific tRNA molecules, allowing the incorporation of non-canonical amino acids and macrocyclic peptide moieties. Flexizymes enable one of the most widely used cyclization methods in mRNA display

by facilitating the translation of an N-terminal alkyl chloride for subsequent reaction with downstream cysteine thiols (Goto and Suga 2021). However, the overall product yield is constrained by the fixed input concentration of aminoacyl tRNA charged with non-canonical amino acids during the translation process. In addition, non-canonical amino acids must undergo chemical modification with an activated leaving group for tRNA aminoacylation, posing an additional challenge.

As an alternative approach, the incorporation of non-canonical amino acids has been successfully achieved using orthogonal aminoacyl-tRNA synthetases (ORSs) (Chin 2017). These ORSs have been employed to introduce approximately 70 types of non-canonical amino acids *in vivo* (C. C. Liu and Schultz 2010), offering a promising alternative and complementary method for genetic code expansion. ORSs offer distinct advantages as they directly aminoacylate tRNA molecules, thus bypassing the need for chemical modification and preacylation steps required for Flexizymes. Moreover, the protein yield is not restricted by fixed concentrations of aminoacyl tRNA.

Traditionally, ORSs have required significant directed evolution efforts to enable the incorporation of a single non-canonical amino acid (L. Wang and Schultz 2001). However, some promiscuous ORSs have exhibited the remarkable ability to recognize multiple non-canonical amino acids (D. D. Young et al. 2011; Guo et al. 2014), thus making them potent versatile tools for *in vitro* display methods. For example, Bowers and his colleagues demonstrated that the promiscuous ORS p-CNF-RS can incorporate non-canonical amino acids including the novel substrate p-cyanopyridylalanine (p-CNpyrA) in CFPS systems (Iskandar et al. 2023). This enabled pyridine-thiazoline (pyr-thn) macrocyclization in mRNA display,

leading to the selection of a potent macrocyclic binder against the deubiquitinase USP15 (**Figure 1-6**). Their work illustrates how promiscuous ORSs can expand amino acid side-chain diversity and introduce non-proteinogenic structures into mRNA display libraries. However, both flexizyme and ORS technology still largely depend on the ribosome capacity to synthesize the polypeptide including non-canonical amino acids, leading to the limitation of the variety of chemical modification.

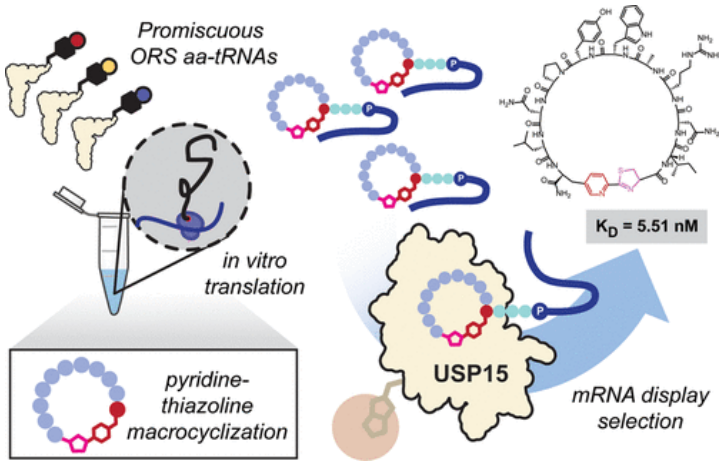


Figure 1-6. Macrocyclic peptide library generation and screening for USP15-binding peptides *via* mRNA display combined with promiscuous orthogonal aminoacyl-tRNA synthetases technique. Adapted from Iskandar et al. 2023. aa-tRNA, aminoacyl-tRNA; ORS, orthogonal aminoacyl-tRNA synthetases; USP15, ubiquitin-specific protease 15.

1.1.4 Ribosomally synthetized and post-translationally modified peptides.

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are peptide-derived natural products exhibiting antibiotic, antiviral, and anticancer properties. While the biosynthesis of RiPPs starts with the translation of a precursor peptide that contains a leader (or follower) sequence, the subsequent post-translational modifications and the cleavage of the leader peptide is characteristic features of RiPP biosynthesis (Arnison et al. 2013). This process leads to an exceptional structural diversity in RiPPs, primarily attributed to the multitude of post-translational modifications carried out by diverse enzymes. Previously, mRNA display has been demonstrated to be compatible with the lactazole and pantocin RiPP machineries. This method was utilized as a biochemical tool to study the substrate scope of the RiPP enzyme PaaA, which catalyzes the double dehydration/decarboxylation of two Glu residues in its peptide substrate PaaP, resulting in the formation of a fused-bicyclic core (Fleming et al. 2020). The modification of Glu allowed for the selection of modified versus unmodified substrate through a degradation step with the Glu-specific GluC protease, followed by enrichment using a streptavidin pulldown assay. The compatibility between RiPP enzymes and mRNA display opens the possibility of constructing artificial RiPP libraries *in vitro*, offering opportunities for the exploration of novel RiPPs and their functions.

1.1.4.1 Radical S-adenosyl-l-methionine enzymes.

Over the last decade, radical *S*-adenosyl-l-methionine (rSAM) enzymes have emerged as the most versatile biocatalysts responsible for catalyzing various types of post-translational modifications in RiPPs. Based on the UniProt database, the rSAM enzyme comprises one of

the largest superfamilies with 750,000 entries (Precord et al. 2023). Nevertheless, the majority of rSAM enzymes have not been characterized by their functions. rSAM enzymes all contain a [4Fe–4S] cluster ligated by three cysteine residues. Cysteines residing in a CX_3CX_2C motif are bound to three of the four iron ions of the cluster. The remaining, unligated iron ion of the cluster binds to the α-amino and α-carboxylate groups of SAM (Sofia et al. 2001). This binding mode facilitates inner-sphere electron transfer from the reduced form of the cluster into the sulfur atom of SAM, resulting in a reductive cleavage of SAM to methionine and a 5′-deoxyadenosyl radical (dAdo•). The generated dAdo• initiates radical-mediated transformations by abstracting a hydrogen atom from the substrate (Cosper et al. 2000; Henshaw, Cheek, and Broderick 2000). Protein structure analysis shows that all structurally characterized rSAM enzymes possess a triose-phosphate isomerase mutase (TIM) barrel fold, either in its full or partial form, as a binding site for the SAM/[4Fe–4S] cluster (Grell, Goldman, and Drennan 2015) (**Figure1-7**).

Despite sharing a common fold, rSAM enzymes display significant divergence in the post-translational modifications, including the macrocyclization *via* thioether bond (Flühe et al. 2012; Benjdia et al. 2016) and carbon-carbon bonds (Schramma, Bushin, and Seyedsayamdost 2015; Barr et al. 2016), methylation (Freeman et al. 2012; Pierre et al. 2012; Parent et al. 2016; Freeman et al. 2017), and epimerization for the regioselective introduction of D- and β-amino acids into RiPPs (Parent et al. 2018; Benjdia et al. 2017; Morinaka et al. 2018) **(Table 1-2)**. The remarkable diversity of modifications catalyzed by rSAM enzymes has expanded the potential of RiPPs in generating structurally diverse molecules with a wide range of biological activities. Moreover, the substrates of rSAM enzymes are diverse including small molecules (*e.g.*, amino

acids, NTP, and fatty acids) (Wecksler et al. 2009; Hover et al. 2015; McLaughlin et al. 2016), DNA (Chandor et al. 2006), RNA (Hernández et al. 2007; A. P. Young and Bandarian 2018), peptides (Hudson et al. 2019; Parent et al. 2018; Gagsteiger et al. 2022; Sugiyama et al. 2022), and proteins (Crain and Broderick 2014; Shisler and Broderick 2014). While the versatility of rSAM enzymes in terms of substrate specificity and catalyzed reactions holds great potential for the generation of artificial peptide libraries with non-proteinogenic structure, their O_2-lability of the [4Fe-4S] cluster has hindered their applications in CFPS systems. The [4Fe-4S] cluster is highly sensitive to O_2, requiring anaerobic conditions for their purification and handling. Even brief exposure to O_2 can lead to significant degradation of the [4Fe–4S] clusters, converting them into the less stable [3Fe–4S] cluster. Extended exposure to O_2 can result in complete cluster degradation. These challenges related to cluster stability posed difficulties for the application of rSAM in CFPS systems.

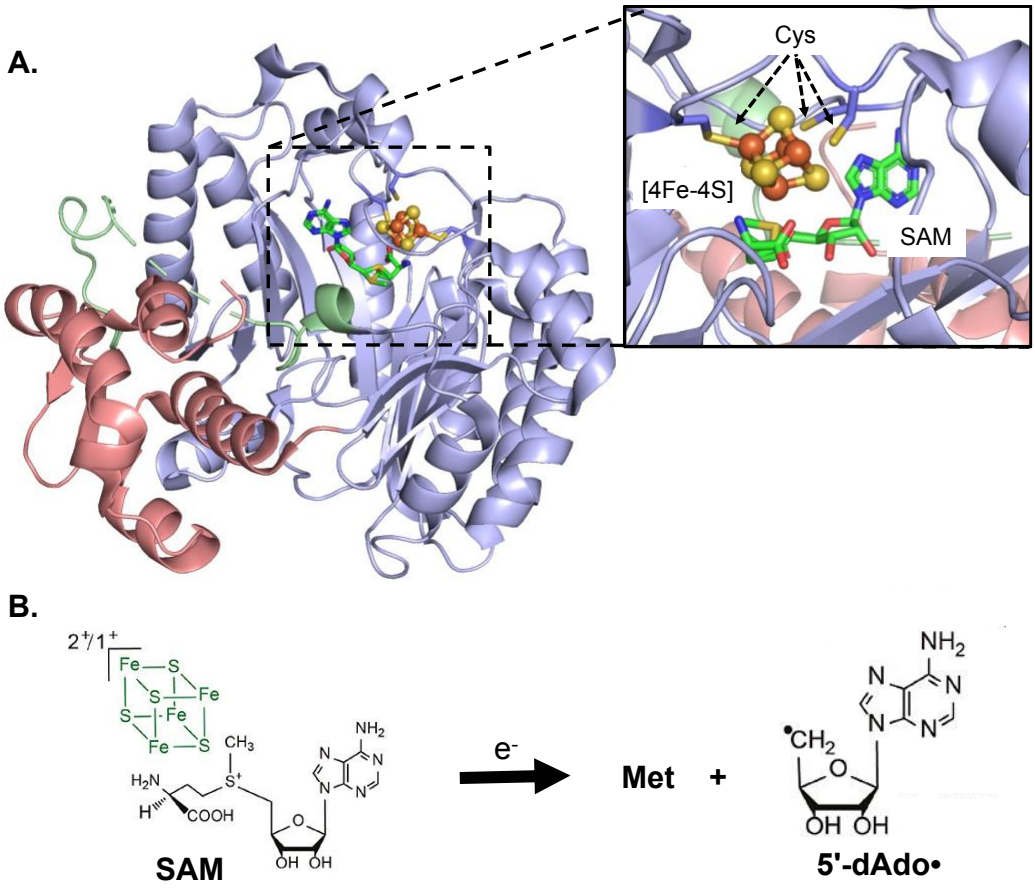


Figure 1-7. The structure and catalytic mechanism of radical S-adenosyl-l-methionine enzyme. (A) HemN crystal structure (PDB ID 1OLT). N-terminal domain colored in palegreen, radical S-adenosyl-lmethionine (rSAM) domain and triose-phosphate isomerase (TIM) barrel fold in light blue, C-terminal domain in salmon pink, [4Fe–4S] clusters in yellow and brown spheres, SAM in green sticks. (B) The reductive cleavage of SAM to generate methionine (Met) and the 5′-deoxyadenosyl radical (dAdo•).

Table 1-2. List of the representative radical S-adenosyl-lmethionine-catalyzed reactions.

Class	Reaction	Enzyme
C-C bond formation	Val-Tyr crosslink	MftC (Ayikpoe and Latham 2019)
	Glu-Tyr crosslink	PqqE (Barr et al. 2016)
	Arg-Tyr crosslink	RrrB (Caruso et al. 2019)
	Lys-Trp crosslink	StrB (Schramma, Bushin, and Seyedsayamdost 2015)
Thioether formation	Sulfur-C_α thioether crosslink	AlbA (Flühe et al. 2012)
	Sulfur-C_β thioether crosslink	PapB (Hudson et al. 2019)
C-O bond formation	O-C_α ether crosslink	DarE and TqqB (Clark, Bushin, and Seyedsayamdost 2019; Ma et al. 2023)
C-methylation	Methyltransfer	PoyBCE TsrM, and Mmp10 (Freeman et al. 2012, 2017; Pierre et al. 2012; Parent et al. 2016; Deobald et al. 2018)
D-amino acids introduction	Epimerization	PoyD, OspD, AvpD, PlpD, YydG (Morinaka et al. 2014; Freeman et al. 2012; Benjdia et al. 2017)
β-amino acid introduction	Peptide splicing	PlpXY (Morinaka et al. 2018)

1.1.5 Biosynthesis of iron-sulfur cluster.

Iron-sulfur (Fe–S) clusters are ancient prosthetic groups essential for fundamental life processes. For example, in small Fe-S cluster protein ferredoxin (Fd) is involved in various important biological processes such as photosynthesis, facilitates electron transfer during redox reactions (Mortenson, Valentine, and Carnahan 1962; Tagawa and Arnon 1962). On the other hand, aconitase, which is an enzyme participating in the citric acid cycle, employs Fe-S clusters as Lewis acids within their active sites (Dennis H. Flint and Allen 1996). Aconitase catalyzes the stereo-specific conversion of citrate to isocitrate through the intermediate cis-aconitate, showcasing the versatile functions of Fe-S clusters in various biological processes. The most common types of clusters found in nature are [2Fe−2S], [3Fe−4S], and [4Fe−4S] cluster (**Figure 1-8**). The three types of clusters exhibit different standard reduction potentials from −650 mV to −130 mV, thus catalyzing a wide range of redox reactions (Jafari et al. 2022). These clusters are typically coordinated by cysteine residues, with some exceptions employing aspartic acid, glutamic acid, and histidine residues (Ohta and Ohki 2017). Beyond the canonical Fe-S clusters, unique Fe-S cluster enzymes exist in nature that possess unconventional Fe-S clusters due to their distinctive cluster derivative chemistry. Notable examples include nitrogenases (J. Kim and Rees 1992), [FeFe]-hydrogenase (Berggren et al. 2013), and Ni-Fe carbon monoxide dehydrogenase (CODH) (Lindahl 2002).

In a cell, the Fe-S clusters are assembled *de novo* by a helper protein system. In bacteria, the cluster assembly is facilitated by three distinct systems known as NIF (**NI**trogen **F**ixation), ISC (**I**ron-**S**ulfur **C**luster), and SUF (mobilization of **SU**l**F**ur). The NIF system is responsible for cubane-type Fe-S cluster assembly, specifically in nitrogenase for nitrogen fixation

(Jacobson et al. 1989) while the ISC system serves as the primary Fe-S cluster assembly system in bacteria (Zheng et al. 1998). The third Fe-S cluster assembly system, SUF system functions as a stress-response cluster assembly system that becomes active under conditions of oxidative stress and iron starvation (Takahashi and Tokumoto 2002). The three assembly systems are widely present in microbial biota, with many organisms possessing more than one system (Tokumoto et al. 2004). The assembly of Fe-S clusters follows a stepwise process: ***(i)*** sulfide group extraction from precursor cysteine, ***(ii)*** Fe-S cluster assembly on a scaffold protein, and ***(iii)*** transfer of the assembled cluster to apo-form of Fe-S cluster proteins. Among the three steps, the systems responsible for Fe-S cluster assembly differ between each machinery while all assembly systems commonly employ an enzyme called cysteine desulfurase for the sulfide extraction.

The ISC system consists of seven components encoded in *isc* operon (*iscSUA-hscBA-fdx-iscX*) (Zheng et al. 1998; Takahashi and Nakamura 1999; Tokumoto and Takahashi 2001). Among the encoded proteins, IscS functions as a pyridoxal 5'-phosphate (PLP)-dependent cysteine desulfurase, which extracts sulfide group from cysteine (D. H. Flint 1996), while IscU is responsible for the cluster assembly (Agar et al. 2000). Meanwhile, IscA serves as a carrier protein, facilitating the transfer of the cluster to apo-proteins (Ollagnier-de-Choudens et al. 2001). On the other hand, the *E. coli* SUF system comprises six components encoded in the *suf* operon (*sufABCDSE*) (Takahashi and Tokumoto 2002). Although the SUF system also employs PLP-dependent cysteine desulfurase SufS (Ollagnier-de-Choudens et al. 2003) and a carrier protein SufA (Gupta et al. 2009), the other components are significantly distinct from those in the ISC system. Especially, SufB, SufC, and SufD form a protein complex known as $SufBC_2D$,

serving as the scaffold for Fe-S cluster assembly (Wollers et al. 2010a). Furthermore, SufE functions as a sulfur-transfer protein, delivering inorganic sulfide extracted by SufS to the $SufBC_2D$ complex (Layer et al. 2007). These distinctive features set the SUF system apart from the ISC system, highlighting their differences (**Table 1-3**). The NIF system also diverges from both the ISC and SUF systems. In contrast to the complexity of the ISC and SUF systems, the NIF system consists of only two components: PLP-dependent cysteine desulfurase NifS (Zheng et al. 1993) and the Fe-S cluster-assembling scaffold NifU (Fu et al. 1994; Dos Santos et al. 2004), which are encoded by *nifSU* (Jacobson et al. 1989). Unlike the ISC and SUF system, NifU also serves as a carrier protein to transfer the Fe-S cluster to apo-proteins.

Traditionally, Fe-S cluster proteins have been overexpressed and purified *via* heterologous expression in *E. coli*. Generally, one of the Fe-S cluster assembly systems is co-expressed for the Fe-S cluster assembly and maturation of apo-Fe-S cluster proteins of interest. For example, Fd from different organisms, including [2Fe-2S] Fd (PetF and FdxH) from *P. boryanum*, 2[4Fe-4S] Fd (FdxN) from *E. coli*, [4Fe-4S] Fd from *B. thermoproteolyticus*, and [3Fe-4S][4Fe-4S] Fd (FdII) from *R. capsulatus*, were co-expressed with all ISC subunits in *E. coli* (Nakamura, Saeki, and Takahashi 1999). The co-expression of ISC system resulted in a significant increase in the yields of all five holo-Fds, reaching up to 7.6-fold of the maximum yields. This result highlights the ISC system's remarkable versatility for the assembly of Fe-S clusters in a wide range of Fe-S cluster proteins since it demonstrated no apparent specificity towards particular Fds or Fe-S cluster types. Another previous study also confirmed that increased expression of SUF subunits led to a substantial increase (approximately 3-fold) in the production of the recombinant [4Fe-4S] cluster protein (Corless et al. 2020). The broad

substrate specificity of these assembly systems has supported the overexpression of versatile Fe-S cluster proteins and thus facilitated their structural and functional characterization.

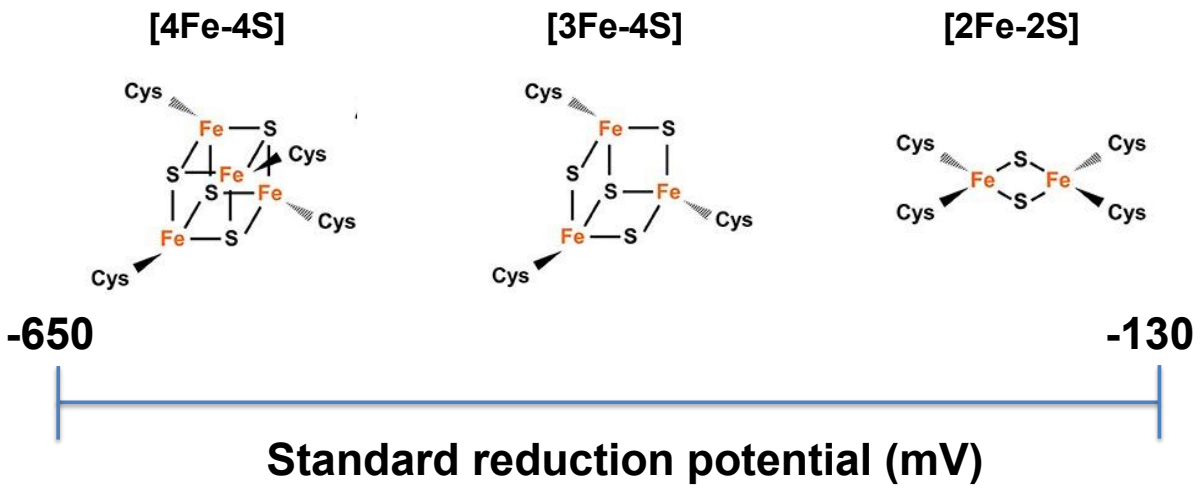


Figure 1-8. The most common types of iron-sulfur clusters observed in nature (Jafari et al. 2022). Standard reduction potentials are with respect to the standard hydrogen electrode.

Table 1-3. Bacterial iron-sulfur (Fe-S) cluster assembly systems.

Fe-S cluster assembly system	Fe-S cluster assembly process		
	(i) Sulfide extraction	*(ii)* Fe-S cluster assembly	*(iii)* Cluster transfer to apo-proteins
ISC	IscS	IscU	IscA
NIF	NifS	NifU	
SUF	SufES	SufBCD	SufA

1.2 Rationale and research objectives

In this thesis, I employed mRNA display method along with PURE system to explore the sequence space of *de novo* functional peptides. Especially, I tried to identify *de novo* RNA-binding peptide in a high-throughput way since RBPs play crucial roles in a range of cellular processes such as DNA repair, post-transcriptional modification, and cancer progression by interacting with RNAs. This multifunctionality makes them appealing for various biotechnological applications. Despite the abundance of naturally occurring RBPs in nature, the creation of *de novo* RNA-binding peptides poses challenges due to the limited structural information available on RNA molecules and their interaction modes with peptides. Reports of RBPs discovered using display methods have mainly utilized phage display, with limited targets such as U1 nuclear low molecular weight RNA and GU base-paired mini-helices RNA (Danner and Belasco 2001; Frugier and Schimmel 1997). The primary challenge in these cases has been the influence of unfavorable base-pairing formation between the target RNA chain and mRNA. Both ribosome display and mRNA display methods involve mRNA in the complex formed during selection.

In **Chapter 2**, I developed a codon-restricted mRNA display method to screen for *de novo* RNA-binding peptides, demonstrating that peptides with limited amino acid diversity can still represent RNA-binding capability. While the limited amino acid diversity can be considered as a downside, we can diversify the chemical property of peptides *via* RiPP enzymes. However, as described in **General introduction** (**Section 1.1**), the most versatile RiPP enzyme, rSAM enzyme, requires the O_2-labile metallocofactor, Fe-S cluster, which hinders the application in CFPS system. Therefore, in **Chapter 3**, I worked on developing a

method to produce Fe-S cluster proteins using a cell-free system under aerobic conditions, which in the future can be applied to produce various rSAM enzymes catalyzing epimerization, methylation, and peptide cyclization reactions in CFPS systems. The work was conducted in Earth-Life Science Institute, where I was based in Prof. Kosuke Fujishima's research group, while cooperating with Prof. Po-Hsiang Wang's environmental engineering and enzymology research group at National Central University, Taiwan. The major objectives of this thesis were as follows:

Objective 1. Develop methods to identify single-stranded RNA-binding peptides from random peptide sequence space using a renovated method combining codon-restricted DNA library and mRNA display.

Objective 2. Develop methods to produce O_2-labile [4Fe-4S] proteins under aerobic conditions using a reconstituted cell-free protein synthesis system.

1.3 Thesis outline and structure

Chapter 1: General introduction (this section also includes research objectives and overall thesis structure.)

Chapter 2: *De novo* single-stranded RNA-binding peptides discovered by codon-restricted mRNA display

This chapter addresses **Objective 1** and reports the development of a renovated mRNA display method that allows high-throughput screening of *de novo* single-stranded RNA-binding peptides, and the potent application of the renovated method for the identification of therapeutic peptides. The data present in this chapter has been published in **Biomacromolecules** (**See Section 1.4**) where I am the first author.

Chapter 3: One-pot *de novo* synthesis of [4Fe-4S] proteins using recombinant SUF system under aerobic conditions

This chapter addresses **Objective 2** and reports the development of a facile and efficient protocol for expressing mature [4Fe-4S] proteins using a reconstituted cell-free translation system under aerobic conditions. The incorporation of recombinant SUF helper proteins into the PURE system facilitated mRNA translation and [4Fe-4S] cluster assembly under anaerobic conditions. To address the vulnerability of [4Fe-4S] clusters to O_2, O_2-scavenging cascade was introduced into the PURE system, enabling the one-pot synthesis of mature [4Fe-4S] proteins under aerobic conditions with a high maturation rate (>90%). The content in this chapter has been published in **ACS synthetic biology** (**See Section 1.4**) where I am a co-first author.

Chapter 4: Summary, significance, and future perspectives

This chapter integrates and summarizes the work of this thesis and describes future research directions.

1.4 Statement of authorship and publication status

In this thesis, **Chapter 2** is mainly based on the results from a published paper in **Biomacromolecules** where I am a first author. **Chapters 3** is mainly based on the results from the published papers in **ACS synthetic biology** where I am a co-first author. The detailed citations and statement of authorship are provided below:

Chapter 2. *De novo* single-stranded RNA-binding peptides discovered by codon-restricted mRNA display

Shota Nishikawa, Hidenori Watanabe, Takayuki Katoh, and Kosuke Fujishima. (2023) *De novo* single-stranded RNA-binding peptides discovered by codon-restricted mRNA display. *Biomacromolecules*, 25, 1, 355–365. doi: 10.1021/acs.biomac.3c01024.

Contributions: S.N., N.T., T.K., and K.F. designed research; S.N. and H.W. performed research; S.N. and K.F. analyzed data; and all authors wrote the paper.

Chapter 3: One-pot *de novo* synthesis of [4Fe-4S] proteins using recombinant SUF system under aerobic conditions

Po-Hsiang Wang, **Shota Nishikawa**, Shawn Erin McGlynn, and Kosuke Fujishima. (2023) One-pot *de novo* synthesis of [4Fe-4S] proteins using recombinant SUF system under aerobic conditions. *ACS synthetic biology*, 20;12(10):2887-2896. doi: 10.1021/acssynbio.3c00155.

Contributions: P.-H.W. and K.F. conceptualized this research; P.-H.W. and S.N performed the experiments; all authors participated in experimental design and data analysis; P.-H.W. and S.N. wrote this paper with assistance from all authors.

Reference

1. Agar, J. N., C. Krebs, J. Frazzon, B. H. Huynh, D. R. Dean, and M. K. Johnson. 2000. "IscU as a scaffold for iron-sulfur cluster biosynthesis: Sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU." *Biochemistry* 39 (27): 7856–62.

2. Albihlal, Waleed S., and André P. Gerber. 2018. "Unconventional RNA-binding proteins: An uncharted zone in RNA biology." *FEBS Letters* 592 (17): 2917–31.

3. Arnison, Paul G., Mervyn J. Bibb, Gabriele Bierbaum, Albert A. Bowers, Tim S. Bugni, Grzegorz Bulaj, Julio A. Camarero, et al. 2013. "Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature." *Natural Product Reports* 30 (1): 108–60.

4. Ayikpoe, Richard S., and John A. Latham. 2019. "MftD catalyzes the formation of a biologically active redox center in the biosynthesis of the ribosomally synthesized and post-translationally modified redox cofactor mycofactocin." *Journal of the American Chemical Society* 141 (34): 13582–91.

5. Baranov, V. I., Morozov IYu, S. A. Ortlepp, and A. S. Spirin. 1989. "Gene expression in a cell-free system on the preparative scale." *Gene* 84 (2): 463–66.

6. Barr, Ian, John A. Latham, Anthony T. Iavarone, Teera Chantarojsiri, Jennifer D. Hwang, and Judith P. Klinman. 2016. "Demonstration that the radical S-adenosylmethionine (SAM) enzyme PqqE catalyzes *de novo* carbon-carbon cross-linking within a peptide substrate PqqA in the presence of the peptide chaperone PqqD." *Journal of Biological Chemistry* 291 (17): 8877–84.

7. Bashiruddin, Nasir K., and Hiroaki Suga. 2015. "Construction and screening of vast libraries of natural product-like macrocyclic peptides using *in vitro* display technologies." *Current Opinion in Chemical Biology* 24 (February): 131–38.

8. Benjdia, Alhosna, Alain Guillot, Benjamin Lefranc, Hubert Vaudry, Jérôme Leprince, and Olivier Berteau. 2016. "Thioether bond formation by SPASM domain radical SAM enzymes: Cα H-atom abstraction in subtilosin a biosynthesis." *Chemical Communications (Cambridge, England)* 52 (37): 6249–52.

9. Benjdia, Alhosna, Alain Guillot, Pauline Ruffié, Jérôme Leprince, and Olivier Berteau. 2017. "Post-translational modification of ribosomally synthesized peptides by a radical SAM epimerase in *Bacillus subtilis*." *Nature Chemistry* 9 (7): 698–707.

10. Berggren, G., A. Adamska, C. Lambertz, T. R. Simmons, J. Esselborn, M. Atta, S. Gambarelli, et al. 2013. "Biomimetic assembly and activation of [FeFe]-hydrogenases." *Nature* 499 (7456): 66–69.

11. Beuth, Barbara, Simon Pennell, Kristine B. Arnvig, Stephen R. Martin, and Ian A. Taylor. 2005. "Structure of a *Mycobacterium tuberculosis* NusA-RNA complex." *The EMBO Journal* 24 (20): 3576–87.

12. Boder, E. T., and K. D. Wittrup. 1997. “Yeast surface display for screening combinatorial polypeptide libraries.” *Nature Biotechnology* 15 (6): 553–57.

13. Buchner, Eduard, and Rudolf Rapp. 1897. “Alkoholische gährung ohne hefezellen.” *Berichte Der Deutschen Chemischen Gesellschaft* 30 (3): 2668–78.

14. Busso, Didier, Rosalind Kim, and Sung-Hou Kim. 2004. “Using an *Escherichia coli* cell-free extract to screen for soluble expression of recombinant proteins.” *Journal of Structural and Functional Genomics* 5 (1–2): 69–74.

15. Caruso, Alessio, Ryan J. Martinie, Leah B. Bushin, and Mohammad R. Seyedsayamdost. 2019. “Macrocyclization *via* an arginine-tyrosine crosslink broadens the reaction scope of radical S-adenosylmethionine enzymes.” *Journal of the American Chemical Society* 141 (42): 16610–14.

16. Chandor, Alexia, Olivier Berteau, Thierry Douki, Didier Gasparutto, Yannis Sanakis, Sandrine Ollagnier-de-Choudens, Mohamed Atta, and Marc Fontecave. 2006. “Dinucleotide spore photoproduct, a minimal substrate of the DNA repair spore photoproduct lyase enzyme from *Bacillus subtilis*.” *Journal of Biological Chemistry* 281 (37): 26922–31.

17. Childs-Disney, Jessica L., and Matthew D. Disney. 2016. “Approaches to validate and manipulate RNA targets with small molecules in cells.” *Annual Review of Pharmacology and Toxicology* 56 (1): 123–40.

18. Chin, Jason W. 2017. “Expanding and reprogramming the genetic code.” *Nature* 550 (7674): 53–60.

19. Clark, Kenzie A., Leah B. Bushin, and Mohammad R. Seyedsayamdost. 2019. “Aliphatic ether bond formation expands the scope of radical SAM enzymes in natural product biosynthesis.” *Journal of the American Chemical Society* 141 (27): 10610–15.

20. Contreras-Llano, Luis E., and Cheemeng Tan. 2018. “High-throughput screening of biomolecules using cell-free gene expression systems.” *Synthetic Biology* 3 (1): ysy012.

21. Corless, Elliot I., Erin L. Mettert, Patricia J. Kiley, and Edwin Antony. 2020. “Elevated expression of a functional Suf pathway in *Escherichia coli* BL21(DE3) enhances recombinant production of an iron-sulfur cluster-containing protein.” *Journal of Bacteriology* 202 (3): e00496-19.

22. Corley, Meredith, Margaret C. Burns, and Gene W. Yeo. 2020. “How RNA-binding proteins interact with RNA: Molecules and mechanisms.” *Molecular Cell* 78 (1): 9–29.

23. Cosper, N. J., S. J. Booker, F. Ruzicka, P. A. Frey, and R. A. Scott. 2000. “Direct FeS cluster involvement in generation of a radical in lysine 2,3-aminomutase.” *Biochemistry* 39 (51): 15668–73.

24. Cotten, Steve W., Jianwei Zou, Rong Wang, Bao-Cheng Huang, and Rihe Liu. 2012. “mRNA display-based selections using synthetic peptide and natural protein libraries.” *Methods in Molecular Biology* 805: 287–97.

25. Crain, Adam V., and Joan B. Broderick. 2014. “Pyruvate formate-lyase and its activation by pyruvate formate-lyase activating enzyme.” *Journal of Biological Chemistry* 289 (9): 5723–29.

26. Danner, Stefan, and Joel G. Belasco. 2001. “T7 phage display: A novel genetic selection system for cloning RNA-binding proteins from cDNA libraries.” *Proceedings of the National Academy of Sciences* 98 (23): 12954–59.

27. Deobald, Darja, Lorenz Adrian, Christian Schöne, Michael Rother, and Gunhild Layer. 2018. “Identification of a unique radical SAM methyltransferase required for the Sp^3-C-methylation of an arginine residue of methyl-coenzyme M reductase.” *Scientific Reports* 8 (1): 7404.

28. Dos Santos, Patricia C., Archer D. Smith, Jeverson Frazzon, Valerie L. Cash, Michael K. Johnson, and Dennis R. Dean. 2004. “Iron-sulfur cluster assembly.” *Journal of Biological Chemistry* 279 (19): 19705–11.

29. Dremann, Danielle N., and Christine S. Chow. 2016. “The development of peptide ligands that target helix 69 rRNA of bacterial ribosomes.” *Bioorganic & Medicinal Chemistry* 24 (18): 4486–91.

30. Driggers, Edward M., Stephen P. Hale, Jinbo Lee, and Nicholas K. Terrett. 2008. “The exploration of macrocycles for drug discovery — an underexploited structural class.” *Nature Reviews. Drug Discovery* 7 (7): 608–24.

31. Endo, Yaeta, Shiro Otsuzuki, Kimio Ito, and Kin-Ichiro Miura. 1992. “Production of an enzymatic active protein using a continuous flow cell-free translation system.” *Journal of Biotechnology* 25 (3): 221–30.

32. Fellmann, Christof, Benjamin G. Gowen, Pei-Chun Lin, Jennifer A. Doudna, and Jacob E. Corn. 2017. “Cornerstones of CRISPR–Cas in drug discovery and therapy.” *Nature Reviews. Drug Discovery* 16 (2): 89–100.

33. Fellmann, Christof, and Scott W. Lowe. 2014. “Stable RNA interference rules for silencing.” *Nature Cell Biology* 16 (1): 10–18.

34. Fleming, Steven R., Paul M. Himes, Swapnil V. Ghodge, Yuki Goto, Hiroaki Suga, and Albert A. Bowers. 2020. “Exploring the post-translational enzymology of PaaA by mRNA display.” *Journal of the American Chemical Society* 142 (11): 5024–28.

35. Flint, D. H. 1996. “*Escherichia coli* contains a protein that is homologous in function and N-terminal sequence to the protein encoded by the NifS gene of *Azotobacterv vinelandii* and that can participate in the synthesis of the Fe-S cluster of dihydroxy-acid dehydratase.” *Journal of Biological Chemistry* 271 (27): 16068–74.

36. Flint, Dennis H., and Ronda M. Allen. 1996. “Iron−sulfur proteins with nonredox functions.” *Chemical Reviews* 96 (7): 2315–34.

37. Flühe, Leif, Thomas A. Knappe, Michael J. Gattner, Antje Schäfer, Olaf Burghaus, Uwe Linne, and Mohamed A. Marahiel. 2012. “The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilosin A.” *Nature Chemical Biology* 8 (4): 350–57.

38. Francisco, J. A., R. Campbell, B. L. Iverson, and G. Georgiou. 1993. “Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface.” *Proceedings of the National Academy of Sciences of the United States of America* 90 (22): 10444–48.

39. Freeman, Michael F., Cristian Gurgui, Maximilian J. Helf, Brandon I. Morinaka, Agustinus R. Uria, Neil J. Oldham, Hans-Georg Sahl, Shigeki Matsunaga, and Jörn Piel. 2012. “Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides.” *Science (New York, N.Y.)* 338 (6105): 387–90.

40. Freeman, Michael F., Maximilian J. Helf, Agneya Bhushan, Brandon I. Morinaka, and Jörn Piel. 2017. “Seven enzymes create extraordinary molecular complexity in an uncultivated bacterium.” *Nature Chemistry* 9 (4): 387–95.

41. Frugier, M., and P. Schimmel. 1997. “Subtle atomic group discrimination in the RNA minor groove.” *Proceedings of the National Academy of Sciences of the United States of America* 94 (21): 11291–94.

42. Fu, W., R. F. Jack, T. V. Morgan, D. R. Dean, and M. K. Johnson. 1994. “NifU gene product from *Azotobacter vinelandii* is a homodimer that contains two identical [2Fe-2S] clusters.” *Biochemistry* 33 (45): 13455–63.

43. Fujii, Satoshi, Tomoaki Matsuura, Takeshi Sunami, Yasuaki Kazuta, and Tetsuya Yomo. 2013. “*In vitro* evolution of α-hemolysin using a liposome display.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (42): 16796–801.

44. Gagsteiger, Jana, Sören Jahn, Lorenz Heidinger, Lukas Gericke, Jennifer N. Andexer, Thorsten Friedrich, Christoph Loenarz, and Gunhild Layer. 2022. “A cobalamin-dependent radical SAM enzyme catalyzes the unique Cα-methylation of glutamine in methyl-coenzyme M reductase.” *Angewandte Chemie (International Ed. in English)* 61 (32): e202204198.

45. Gc, Keshav, Davidnhan To, Kumudie Jayalath, and Sanjaya Abeysirigunawardena. 2019. “Discovery of a novel small molecular peptide that disrupts helix 34 of bacterial ribosomal RNA.” *RSC Advances* 9 (69): 40268–76.

46. Goto, Yuki, Takayuki Katoh, and Hiroaki Suga. 2011. “Flexizymes for genetic code reprogramming.” *Nature Protocols* 6 (6): 779–90.

47. Goto, Yuki, and Hiroaki Suga. 2021. “The RaPID platform for the discovery of pseudo-natural macrocyclic peptides.” *Accounts of Chemical Research* 54 (18): 3604–17.

48. Grell, Tsehai A. J., Peter J. Goldman, and Catherine L. Drennan. 2015. “SPASM and twitch domains in S-adenosylmethionine (SAM) radical enzymes.” *Journal of Biological Chemistry* 290 (7): 3964–71.

49. Guo, Li-Tao, Yane-Shih Wang, Akiyoshi Nakamura, Daniel Eiler, Jennifer M. Kavran, Margaret Wong, Laura L. Kiessling, Thomas A. Steitz, Patrick O'Donoghue, and Dieter Söll. 2014. "Polyspecific pyrrolysyl-tRNA synthetases from directed evolution." *Proceedings of the National Academy of Sciences of the United States of America* 111 (47): 16724–29.

50. Gupta, Vibha, Maïté Sendra, Sunil G. Naik, Harsimranjit K. Chahal, Boi Hanh Huynh, F. Wayne Outten, Marc Fontecave, and Sandrine Ollagnier de Choudens. 2009. "Native *Escherichia coli* SufA, coexpressed with SufBCDSE, purifies as a [2Fe−2S] protein and acts as an Fe−S transporter to Fe−S target enzymes." *Journal of the American Chemical Society* 131 (17): 6149–53.

51. Harrow, Jennifer, Adam Frankish, Jose M. Gonzalez, Electra Tapanari, Mark Diekhans, Felix Kokocinski, Bronwen L. Aken, et al. 2012. "GENCODE: The reference human genome annotation for the ENCODE project." *Genome Research* 22 (9): 1760–74.

52. Henrich, Erik, Christopher Hein, Volker Dötsch, and Frank Bernhard. 2015. "Membrane protein production in *Escherichiac coli* cell-free lysates." *FEBS Letters* 589 (15): 1713–22.

53. Henshaw, Timothy F., Jennifer Cheek, and Joan B. Broderick. 2000. "The [4Fe-4S]$^{1+}$ cluster of pyruvate formate-lyase activating enzyme generates the glycyl radical on pyruvate formate-lyase: EPR-detected single turnover." *Journal of the American Chemical Society* 122 (34): 8331–32.

54. Hernández, Heather L., Fabien Pierrel, Eric Elleingand, Ricardo García-Serres, Boi Hanh Huynh, Michael K. Johnson, Marc Fontecave, and Mohamed Atta. 2007. "MiaB, a bifunctional radical-S-adenosylmethionine enzyme involved in the thiolation and methylation of tRNA, contains two essential [4Fe-4S] clusters." *Biochemistry* 46 (17): 5140–47.

55. Hollingworth, David, Adela M. Candel, Giuseppe Nicastro, Stephen R. Martin, Paola Briata, Roberto Gherzi, and Andres Ramos. 2012. "KH domains with impaired nucleic acid binding as a tool for functional analysis." *Nucleic Acids Research* 40 (14): 6873–86.

56. Hopkins, Andrew L., and Colin R. Groom. 2002. "The druggable genome." *Nature Reviews. Drug Discovery* 1 (9): 727–30.

57. Hover, Bradley M., Nam K. Tonthat, Maria A. Schumacher, and Kenichi Yokoyama. 2015. "Mechanism of pyranopterin ring formation in molybdenum cofactor biosynthesis." *Proceedings of the National Academy of Sciences of the United States of America* 112 (20): 6347–52.

58. Hudson, Graham A., Brandon J. Burkhart, Adam J. DiCaprio, Christopher J. Schwalen, Bryce Kille, Taras V. Pogorelov, and Douglas A. Mitchell. 2019. "Bioinformatic mapping of radical S-adenosylmethionine-dependent ribosomally synthesized and post-translationally modified peptides identifies new Cα, Cβ, and Cγ-linked thioether-containing peptides." *Journal of the American Chemical Society* 141 (20): 8228–38.

59. Iskandar, Sabrina E., Jarrett M. Pelton, Elizaveta T. Wick, Derek L. Bolhuis, Albert S. Baldwin, Michael J. Emanuele, Nicholas G. Brown, and Albert A. Bowers. 2023. "Enabling

genetic code expansion and peptide macrocyclization in mRNA display *via* a promiscuous orthogonal aminoacyl-tRNA synthetase." *Journal of the American Chemical Society* 145 (3): 1512–17.

60. Jackson, M., J. M. Pratt, and I. B. Holland. 1983. "Enhanced polypeptide synthesis programmed by linear DNA fragments in cell-free extracts lacking exonuclease V." *FEBS Letters* 163 (2): 221–24.

61. Jacobson, M. R., V. L. Cash, M. C. Weiss, N. F. Laird, W. E. Newton, and D. R. Dean. 1989. "Biochemical and genetic analysis of the NifUSVWZM cluster from *Azotobacterv vinelandii*." *Molecular & General Genetics: MGG* 219 (1–2): 49–57.

62. Jafari, Sonia, Yakini A. Tavares Santos, Justin Bergmann, Mehdi Irani, and Ulf Ryde. 2022. "Benchmark study of redox potential calculations for iron-sulfur clusters in proteins." *Inorganic Chemistry* 61 (16): 5991–6007.

63. Jia, Tony Z., Shota Nishikawa, and Kosuke Fujishima. 2022. "Sequencing the origins of life." *BBA Advances* 2 (100049): 100049.

64. Kanamori, Takashi, Yasuhiro Fujino, and Takuya Ueda. 2014. "PURE ribosome display and its application in antibody technology." *Biochimica et Biophysica Acta* 1844 (11): 1925–32.

65. Kang, Shou-Kai, Bai-Xue Chen, Tian Tian, Xi-Shuai Jia, Xin-Yi Chu, Rong Liu, Peng-Fei Dong, Qing-Yong Yang, and Hong-Yu Zhang. 2015. "ATP selection in a random peptide library consisting of prebiotic amino acids." *Biochemical and Biophysical Research Communications* 466 (3): 400–405.

66. Kaur, Moninderpal, Chamila N. Rupasinghe, Edvin Klosi, Mark R. Spaller, and Christine S. Chow. 2013. "Selection of heptapeptides that bind helix 69 of bacterial 23S ribosomal RNA." *Bioorganic & Medicinal Chemistry* 21 (5): 1240–47.

67. Keefe, A. D., and J. W. Szostak. 2001. "Functional proteins from a random-sequence library." *Nature* 410 (6829): 715–18.

68. Kim, Dong-Myung, and James R. Swartz. 2001. "Regeneration of adenosine triphosphate from glycolytic intermediates for cell-Free protein synthesis." *Biotechnology and Bioengineering*. 74 (4): 309-16.

69. Kim, J., and D. C. Rees. 1992. "Structural models for the metal centers in the nitrogenase molybdenum-iron protein." *Science (New York, N.Y.)* 257 (5077): 1677–82.

70. Kuruma, Yutetsu, and Takuya Ueda. 2015. "The PURE system for the cell-free synthesis of membrane proteins." *Nature Protocols* 10 (9): 1328–44.

71. Lamichhane, Tek N., N. Dinuka Abeydeera, Anne-Cécile E. Duc, Philip R. Cunningham, and Christine S. Chow. 2011. "Selection of peptides targeting helix 31 of bacterial 16S ribosomal RNA by screening M13 phage-display libraries." *Molecules (Basel, Switzerland)* 16 (2): 1211–39.

72. Lavickova, Barbora, and Sebastian J. Maerkl. 2019. “A simple, robust, and low-cost method to produce the PURE cell-free system.” *ACS Synthetic Biology* 8 (2): 455–62.

73. Layer, Gunhild, S. Aparna Gaddam, Carla N. Ayala-Castro, Sandrine Ollagnier-de Choudens, David Lascoux, Marc Fontecave, and F. Wayne Outten. 2007. “SufE transfers sulfur from SufS to SufB for iron-sulfur cluster assembly.” *Journal of Biological Chemistry* 282 (18): 13342–50.

74. Liang, Xue-Hai, Wen Shen, Hong Sun, Michael T. Migawa, Timothy A. Vickers, and Stanley T. Crooke. 2016. “Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames.” *Nature Biotechnology* 34 (8): 875–80.

75. Liang, Xue-Hai, Hong Sun, Wen Shen, Shiyu Wang, Joyee Yao, Michael T. Migawa, Huynh-Hoa Bui, et al. 2017. “Antisense oligonucleotides targeting translation inhibitory elements in 5′ UTRs can selectively increase protein levels.” *Nucleic Acids Research* 45 (16): 9528–46.

76. Lindahl, Paul A. 2002. “The Ni-containing carbon monoxide dehydrogenase family: Light at the end of the tunnel?” *Biochemistry* 41 (7): 2097–2105.

77. Liu, Chang C., and Peter G. Schultz. 2010. “Adding new chemistries to the genetic code.” *Annual Review of Biochemistry* 79 (1): 413–44.

78. Lunde, Bradley M., Claire Moore, and Gabriele Varani. 2007. “RNA-binding proteins: modular design for efficient function.” *Nature Reviews. Molecular Cell Biology* 8 (6): 479–90.

79. Ma, Suze, Wenhui Xi, Shu Wang, Heng Chen, Sijia Guo, Tianlu Mo, Wenxue Chen, et al. 2023. “Substrate-controlled catalysis in the ether cross-link-forming radical SAM enzymes.” *Journal of the American Chemical Society*, 145 (42): 22945-22953.

80. Mattheakis, L. C., R. R. Bhatt, and W. J. Dower. 1994. “An *in vitro* polysome display system for identifying ligands from very large peptide libraries.” *Proceedings of the National Academy of Sciences of the United States of America* 91 (19): 9022–26.

81. McLaughlin, Martin I., Nicholas D. Lanz, Peter J. Goldman, Kyung-Hoon Lee, Squire J. Booker, and Catherine L. Drennan. 2016. “Crystallographic snapshots of sulfur insertion by lipoyl synthase.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (34): 9446–50.

82. Mochizuki, Yuki, Manish Biyani, Sachika Tsuji-Ueno, Miho Suzuki, Koichi Nishigaki, Yuzuru Husimi, and Naoto Nemoto. 2011. “One-pot preparation of mRNA/cDNA display by a novel and versatile puromycin-linker DNA.” *ACS Combinatorial Science* 13 (5): 478–85.

83. Morinaka, Brandon I., Edgars Lakis, Marjan Verest, Maximilian J. Helf, Thibault Scalvenzi, Anna L. Vagstad, James Sims, Shinichi Sunagawa, Muriel Gugger, and Jörn Piel. 2018.

"Natural noncanonical protein splicing yields products with diverse β-amino acid residues." *Science (New York, N.Y.)* 359 (6377): 779–82.

84. Morinaka, Brandon I., Anna L. Vagstad, Maximilian J. Helf, Muriel Gugger, Carsten Kegler, Michael F. Freeman, Helge B. Bode, and Jörn Piel. 2014. "Radical S-adenosyl methionine epimerases: Regioselective introduction of diverse D-amino acid patterns into peptide natural products." *Angewandte Chemie (International Ed. in English)* 53 (32): 8503–7.

85. Mortenson, L. E., R. C. Valentine, and J. E. Carnahan. 1962. "An electron transport factor from *Clostridium pasteurianum*." *Biochemical and Biophysical Research Communications* 7 (6): 448–52.

86. Nagumo, Yu, Kei Fujiwara, Kenichi Horisawa, Hiroshi Yanagawa, and Nobuhide Doi. 2016. "PURE mRNA display for *in vitro* selection of single-chain antibodies." *Journal of Biochemistry* 159 (5): 519–26.

87. Nakamura, M., K. Saeki, and Y. Takahashi. 1999. "Hyperproduction of recombinant ferredoxins in *Escherichia coli* by coexpression of the ORF1-ORF2-IscS-IscU-IscA-HscB-Hs cA-Fdx-ORF3 gene cluster." *The Journal of Biochemistry* 126 (1): 10–18.

88. Nemoto, N., E. Miyamoto-Sato, Y. Husimi, and H. Yanagawa. 1997. "*In vitro* virus: Bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome *in vitro*." *FEBS Letters* 414 (2): 405–8.

89. Newton, Matilda S., Yari Cabezas-Perusse, Cher Ling Tong, and Burckhard Seelig. 2020. "*In vitro* selection of peptides and proteins-advantages of mRNA display." *ACS Synthetic Biology* 9 (2): 181–90.

90. Nielsen, Jens, and Jay D. Keasling. 2016. "Engineering cellular metabolism." *Cell* 164 (6): 1185–97.

91. Nirenberg, M. W., and J. H. Matthaei. 1961. "The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides." *Proceedings of the National Academy of Sciences of the United States of America* 47 (10): 1588–1602.

92. Ohta, Shun, and Yasuhiro Ohki. 2017. "Impact of ligands and media on the structure and properties of biological and biomimetic iron-sulfur clusters." *Coordination Chemistry Reviews* 338 (May): 207–25.

93. Ollagnier-de-Choudens, Sandrine, David Lascoux, Laurent Loiseau, Frédéric Barras, Eric Forest, and Marc Fontecave. 2003. "Mechanistic studies of the SufS-SufE cysteine desulfurase: Evidence for sulfur transfer from SufS to SufE." *FEBS Letters* 555 (2): 263–67.

94. Ollagnier-de-Choudens, Sandrine, Tony Mattioli, Yasuhiro Takahashi, and Marc Fontecave. 2001. "Iron-sulfur cluster assembly." *Journal of Biological Chemistry* 276 (25): 22604–7.

95. Parent, Aubérie, Alhosna Benjdia, Alain Guillot, Xavier Kubiak, Clémence Balty, Benjamin Lefranc, Jérôme Leprince, and Olivier Berteau. 2018. "Mechanistic investigations of PoyD, a radical S-adenosyl-l-methionine enzyme catalyzing iterative and directional epimerizations

in polytheonamide A biosynthesis." *Journal of the American Chemical Society* 140 (7): 2469–77.

96. Parent, Aubérie, Alain Guillot, Alhosna Benjdia, Gwladys Chartier, Jérôme Leprince, and Olivier Berteau. 2016. "The B12-radical SAM enzyme PoyC catalyzes valine Cβ-methylation during polytheonamide biosynthesis." *Journal of the American Chemical Society* 138 (48): 15515–18.

97. Peacock, Hayden, and Hiroaki Suga. 2021. "Discovery of *de novo* macrocyclic peptides by messenger RNA display." *Trends in Pharmacological Sciences* 42 (5): 385–97.

98. Pereira, Joana, and Andrei N. Lupas. 2018. "The ancestral KH peptide at the root of a domain family with three different folds." *Bioinformatics (Oxford, England)* 34 (23): 3961–65.

99. Pérez-Cañadillas, José Manuel. 2006. "Grabbing the message: Structural basis of mRNA 3′UTR recognition by Hrp1." *The EMBO Journal* 25 (13): 3167-3178–3178.

100. Pierre, Stéphane, Alain Guillot, Alhosna Benjdia, Corine Sandström, Philippe Langella, and Olivier Berteau. 2012. "Thiostrepton tryptophan methyltransferase expands the chemistry of radical SAM enzymes." *Nature Chemical Biology* 8 (12): 957–59.

101. Precord, Timothy W., Sangeetha Ramesh, Shravan R. Dommaraju, Lonnie A. Harris, Bryce L. Kille, and Douglas A. Mitchell. 2023. "Catalytic site proximity profiling for functional unification of sequence-diverse radical S-adenosylmethionine enzymes." *ACS Bio & Med Chem Au* 3 (3): 240–51.

102. Roberts, R. W., and J. W. Szostak. 1997. "RNA-peptide fusions for the *in vitro* selection of peptides and proteins." *Proceedings of the National Academy of Sciences of the United States of America* 94 (23): 12297–302.

103. Roberts, Thomas C., Robert Langer, and Matthew J. A. Wood. 2020. "Advances in oligonucleotide drug delivery." *Nature Reviews. Drug Discovery* 19 (10): 673–94.

104. Schramma, Kelsey R., Leah B. Bushin, and Mohammad R. Seyedsayamdost. 2015. "Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink." *Nature Chemistry* 7 (5): 431–37.

105. Seki, Eiko, Natsuko Matsuda, and Takanori Kigawa. 2009. "Multiple inhibitory factor removal from an *Escherichia coli* cell extract improves cell-free protein synthesis." *Journal of Bioscience and Bioengineering* 108 (1): 30–35.

106. Shimizu, Y., A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda. 2001. "Cell-free translation reconstituted with purified components." *Nature Biotechnology* 19 (8): 751–55.

107. Shisler, Krista A., and Joan B. Broderick. 2014. "Glycyl radical activating enzymes: structure, mechanism, and substrate interactions." *Archives of Biochemistry and Biophysics* 546 (March): 64–71.

108. Sidhu, S. S., H. B. Lowman, B. C. Cunningham, and J. A. Wells. 2000. “Phage display for selection of novel binding peptides.” *Methods in Enzymology* 328: 333–63.

109. Silverman, Adam D., Ashty S. Karim, and Michael C. Jewett. 2020. “Cell-free gene expression: An expanded repertoire of applications.” *Nature Reviews. Genetics* 21 (3): 151–70.

110. Smith, G. P. 1985. “Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface.” *Science (New York, N.Y.)* 228 (4705): 1315–17.

111. Smith, Mason R., Eshita Khera, and Fei Wen. 2015. “Engineering novel and improved biocatalysts by cell surface display.” *Industrial & Engineering Chemistry Research* 54 (16): 4021–32.

112. Sofia, H. J., G. Chen, B. G. Hetzler, J. F. Reyes-Spindola, and N. E. Miller. 2001. “Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: Functional characterization using new analysis and information visualization methods.” *Nucleic Acids Research* 29 (5): 1097–1106.

113. Spirin, A. S., V. I. Baranov, L. A. Ryabova, S. Y. Ovodov, and Y. B. Alakhov. 1988. “A continuous cell-free translation system capable of producing polypeptides in high yield.” *Science (New York, N.Y.)* 242 (4882): 1162–64.

114. Su, Zhaoming, Yongjie Zhang, Tania F. Gendron, Peter O. Bauer, Jeannie Chew, Wang-Yong Yang, Erik Fostvedt, et al. 2014. “Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in C9FTD/ALS.” *Neuron* 83 (5): 1043–50.

115. Sugiyama, Ryosuke, Angelica Faith L. Suarez, Yohei Morishita, Thi Quynh Ngoc Nguyen, Yi Wei Tooh, Muhammad Nur Hadi Bin Roslan, Justin Lo Choy, et al. 2022. “The biosynthetic landscape of triceptides reveals radical SAM enzymes that catalyze cyclophane formation on Tyr- and His-containing motifs.” *Journal of the American Chemical Society* 144 (26): 11580–93.

116. Tagawa, K., and D. I. Arnon. 1962. “Ferredoxins as electron carriers in photosynthesis and in the biological production and consumption of hydrogen gas.” *Nature* 195 (4841): 537–43.

117. Takahashi, Y., and M. Nakamura. 1999. “Functional assignment of the ORF2-IscS-IscU-IscA-HscB-HscA-Fdx-ORF3 gene cluster involved in the assembly of Fe-S clusters in *Escherichia coli*.” *The Journal of Biochemistry* 126 (5): 917–26.

118. Takahashi, Yasuhiro, and Umechiyo Tokumoto. 2002. “A third bacterial system for the assembly of iron-sulfur clusters with homologs in archaea and plastids.” *Journal of Biological Chemistry* 277 (32): 28380–83.

119. Tokumoto, U., and Y. Takahashi. 2001. “Genetic analysis of the Isc operon in *Escherichia coli* involved in the biogenesis of cellular iron-sulfur proteins.” *The Journal of Biochemistry* 130 (1): 63–71.

120. Tokumoto, Umechiyo, Seiichi Kitamura, Keiichi Fukuyama, and Yasuhiro Takahashi. 2004. "Interchangeability and distinct properties of bacterial Fe-S cluster assembly systems: functional replacement of the Isc and Suf operons in *Escherichia coli* with the NifSU-like operon from *Helicobacter pylori*." *Journal of Biochemistry* 136 (2): 199–209.

121. Tuckey, Corinna, Haruichi Asahara, Ying Zhou, and Shaorong Chong. 2014. "Protein synthesis using a reconstituted cell-free system." *Current Protocols in Molecular Biology* 108 (1): 16.31.1-16.31.22.

122. Ullman, Christopher G., Laura Frigotto, and R. Neil Cooley. 2011. "*In vitro* methods for peptide display and their applications." *Briefings in Functional Genomics* 10 (3): 125–34.

123. Wahlestedt, Claes. 2013. "Targeting long non-coding RNA to therapeutically upregulate gene expression." *Nature Reviews. Drug Discovery* 12 (6): 433–46.

124. Wang, L., and P. G. Schultz. 2001. "A general approach for the generation of orthogonal tRNAs." *Chemistry & Biology* 8 (9): 883–90.

125. Wang, Po-Hsiang, Kosuke Fujishima, Samuel Berhanu, Yutetsu Kuruma, Tony Z. Jia, Anna N. Khusnutdinova, Alexander F. Yakunin, and Shawn E. McGlynn. 2020. "A bifunctional polyphosphate kinase driving the regeneration of nucleoside triphosphate and reconstituted cell-free protein synthesis." *ACS Synthetic Biology* 9 (1): 36–42.

126. Wecksler, Stephen R., Stefan Stoll, Ha Tran, Olafur T. Magnusson, Shu-Pao Wu, David King, R. David Britt, and Judith P. Klinman. 2009. "Pyrroloquinoline quinone biogenesis: demonstration that PqqE from *Klebsiella pneumoniae* is a radical S-adenosyl-L-methionine enzyme." *Biochemistry* 48 (42): 10151–61.

127. Wilson, Daniel N. 2014. "Ribosome-targeting antibiotics and mechanisms of bacterial resistance." *Nature Reviews. Microbiology* 12 (1): 35–48.

128. Winkle, Melanie, Sherien M. El-Daly, Muller Fabbri, and George A. Calin. 2021. "Noncoding RNA therapeutics - challenges and potential solutions." *Nature Reviews. Drug Discovery* 20 (8): 629–51.

129. Wollers, Silke, Gunhild Layer, Ricardo Garcia-Serres, Luca Signor, Martin Clemancey, Jean-Marc Latour, Marc Fontecave, and Sandrine Ollagnier de Choudens. 2010. "Iron-sulfur (Fe-S) cluster assembly." *Journal of Biological Chemistry* 285 (30): 23331–41.

130. Yamaguchi, Junichi, Mohammed Naimuddin, Manish Biyani, Toru Sasaki, Masayuki Machida, Tai Kubo, Takashi Funatsu, Yuzuru Husimi, and Naoto Nemoto. 2009. "cDNA display: A novel screening method for functional disulfide-rich peptides by solid-phase synthesis and stabilization of mRNA-protein fusions." *Nucleic Acids Research* 37 (16): e108.

131. Young, Anthony P., and Vahe Bandarian. 2018. "TYW1: A radical SAM enzyme involved in the biosynthesis of wybutosine bases." *Methods in Enzymology* 606 (June): 119–53.

132. Young, Douglas D., Travis S. Young, Michael Jahnz, Insha Ahmad, Glen Spraggon, and Peter G. Schultz. 2011. “An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity.” *Biochemistry* 50 (11): 1894–1900.

133. Zheng, L., V. L. Cash, D. H. Flint, and D. R. Dean. 1998. “Assembly of iron-sulfur clusters. Identification of an IscSUA-HscBA-Fdx gene cluster from *Azotobacter vinelandii*.” *Journal of Biological Chemistry* 273 (21): 13264–72.

134. Zheng, L., R. H. White, V. L. Cash, R. F. Jack, and D. R. Dean. 1993. “Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis.” *Proceedings of the National Academy of Sciences of the United States of America* 90 (7): 2754–58.

135. Zigmond, Sally H., Michael Joyce, Jane Borleis, Gary M. Bokoch, and Peter N. Devreotes. 1997. “Regulation of actin polymerization in cell-free systems by GTPγS and Cdc42.” *The Journal of Cell Biology* 138 (2): 363–74.

136. Zubay, G. 1973. “*In vitro* synthesis of protein in microbial systems.” *Annual Review of Genetics* 7 (1): 267–87.

Chapter 2 – *De novo* single-stranded RNA discovered by codon-restricted mRNA display

2.1 Introduction

RBPs are essential in core biological processes such as transcription, translation, gene expression, *etc*. More than 2,000 RBPs have been reported and found to play critical roles in diverse cellular processes including DNA repair, post-transcriptional modification, and cancer progression through their interactions with RNAs (Hentze et al. 2018). Moreover, RNA-protein interactions regulate viral replication and are often misregulated in neurological disorders and cancers (Anczuków et al. 2012; Linya Wang, Jeng, and Lai 2011; Kechavarzi and Janga 2014; H. J. Kim et al. 2013). Therefore, RBPs have garnered considerable attention in both biological research as well as biotechnological applications.

RBPs commonly exhibit modular structures characterized by the presence of multiple repeats of a limited set of domains (Lunde, Moore, and Varani 2007). The RNA-binding domains can assemble in diverse configurations, thereby creating adaptable RNA-binding surfaces. The modular architecture of the RBPs imparts the capacity to engage RNA with enhanced specificity and affinity when compared to individual domains. Previously, significant advancements have been achieved in tailoring the RNA sequence specificity of Pumilio/fem-3-binding factor (PUF) repeat domains (Cheong and Hall 2006). A comprehensive recognition amino acid code for each of the four RNA bases has been determined, granting substantial flexibility in the design of protein specificity (Filipovska et al. 2011). Engineered PUF domains have been effectively integrated with diverse effector domains to serve various purposes, such as tracking RNA localization within cells (Ozawa et al. 2007; Tilsner et al. 2009), creating site-specific RNA endonucleases (Choudhury et al. 2012), and regulating the translation and stability of specific mRNA molecules *in vivo* (Cooke et al. 2011). These developments suggest

a wide range of potential applications for RBPs as versatile tools for biotechnological applications. For example, peptides originating from RNA-binding domains such as RBM39, SAFA, or hnRNPK, have been conjugated with the Penetratin cell-penetrating peptide (CPP) to effectively inhibit cancer cell proliferation by disrupting RNA-protein interactions (Puvvula et al. 2021; Puvvula, Buczkowski, and Moon 2021).

While nature has provided a repertoire of naturally occurring RBPs, these existing RBPs represent only a fraction of the potential repertoire of RBPs given a large fraction of the possible folds remains unexplored (Minami et al. 2023). RBDs that remain unexplored in nature not only hold promise for the expansion of our biotechnological toolkit but also harbor the potential to deepen our understanding of the fundamental mechanisms governing these molecular interactions. These profound insights, in turn, serve to facilitate the rational design of *de novo* RBPs. Therefore, high-throughput screening for *de novo* RNA-binding peptides from a random peptide library offers a novel repertoire of RNA-binding peptides that have not existed in nature. Especially, targeting single-stranded RNA (ssRNA) regions providing ground insights regarding nucleotide-specific recognition, which are highly beneficial for precise RNA probing and mRNA regulation. It is worth noting that ssRNA regions play a crucial role in regulating virus infection and oncogene expression, emphasizing the importance of the development of *de novo* RNA-binding peptides (Z. Luo et al. 2014; Kawasaki et al. 2016). For example, cytosine-rich regions are present in the 5' untranslated region of the RNA of hepatitis-C virus and poliovirus (Linya Wang, Jeng, and Lai 2011; Blyn et al. 1996). The cytosine-rich regions regulate transcription levels *via* the interaction with the host poly(C)-binding proteins (PCBP) (Linya Wang, Jeng, and Lai 2011; Ghanem et al. 2016; Z. Li, Wang, and Jia 2022). Poly(A)-

binding proteins (PABP) are also known as essential eukaryotic mRNA regulators promoting the circularization of mRNA to enhance translation efficiency (Blobel 1973; Gallie 1991; Wells et al. 1998).

In vitro display methods such as phage display, ribosome display, and mRNA display have been used as a powerful tool to screen novel peptide aptamers against versatile targets (Mattheakis, Bhatt, and Dower 1994; Hanes and Plückthun 1997; Nemoto et al. 1997; R. W. Roberts and Szostak 1997). Indeed, Li *et al.* recently reported a *de novo* hydrophobic-cationic RNA-binding peptide using phage display (P. Li, Holliger, and Tagami 2022). mRNA display is a widely used *in vitro* selection technique to rapidly isolate functional peptides from libraries of more than 10^{13} molecules. It has been applied to directed evolution studies, drug development, and cellular proteomics analysis (Nagumo et al. 2016; Bashiruddin and Suga 2015; Oikonomou, Salatino, and Tavazoie 2020). mRNA display uses a stable covalent link between mRNA and its cognate polypeptide, thus enabling a selection for versatile protein functions such as protein binding, ATP-binding, and catalytic function including RNA ligation (Fukuda et al. 2006; Keefe and Szostak 2001; Seelig and Szostak 2007). mRNA display platform has also been diversified, for example using a cell-free system to analyze protein-protein interaction and screening enzyme inhibitors (Yamagishi et al. 2011; Miyamoto-Sato et al. 2005). More recently, mRNA display combined with PURE systems (PURE mRNA display) has achieved high efficiency in *in vitro* selection by minimizing the carry-over of intrinsic cellular components (Shimizu et al. 2001; Reyes et al. 2021). However, there is an apparent downside for the mRNA display when it comes to RNA-binding. Especially if the target RNA contains a stretch of single strand region (>5 nt), it can form a stable base-pairing between the

mRNA, and therefore, mRNA harboring complementary sequences of the targeted RNA will be selected and hinder the selection process. To overcome this issue, Dr. Nemoto's group invented the cDNA display method to mask mRNA with reverse-transcribed cDNA (Yamaguchi et al. 2009; Mochizuki et al. 2011). They also applied this method with a codon-restricted library to screen for a *de novo* transfer RNA (tRNA) binding peptides consisting of four prebiotically relevant amino acids (Kumachi, Husimi, and Nemoto 2016). Similarly, Giacobelli *et al.* explored sequence variants of the C-terminal domain of ribosomal protein uL11 using PURE mRNA display and selected a primitive uL11 protein analog only composed of ten prebiotically plausible amino acids (Giacobelli et al. 2022). These previous studies demonstrate that peptides with a limited amino acid sequence can still interact with structurally conserved tRNA and ribosomal RNA (rRNA). In this study, I mitigate the base-pairing problem by restricting the codon usage of the DNA library and demonstrate the efficiency of a codon-restricted mRNA display technique in targeting low-complexity ssRNA, which can be further implemented to target biological RNA sequences.

2.2 Materials and Methods

All oligos were ordered through FASMAC Inc. (Atsugi-shi, Kanagawa, Japan). All peptides were chemically synthesized and purified with high purity (>90%) by GenScript (Piscataway, NJ, USA) unless specified. (Methods for *in vitro* transcription.; *in vitro* RNA transcription; puromycin-DNA-tag ligation; synthesis of mRNA-peptide conjugate by *in vitro* translation; gel-purification; negative selection of mRNA-tag by the affinity with polyRNA; peptide selection based on the affinity with polyRNA; reverse transcription PCR (RT-PCR); sequence analysis; principal component analysis; calculation of hydrophobicity and net charge

of peptide; enrichment analysis of combinatorial amino acid pairs; microscale thermophoresis are shown in **Supporting Information**. The complete list of DNA and RNA oligos used in this work is available in **Table 2-1**).

2.2.1 DNA library preparation.

DNA libraries were prepared by annealing two single-stranded DNA (ssDNA) oligos and then extension *via* Klenow fragment. Codon-restricted DNA libraries were produced from forward DNA oligo (PolyAGCU-binding-F oligo) (2 μM) and the corresponding reverse DNA oligo (VMM/VRR, SBB, RDK, or HHY codon-library-R oligo) (2 μM). The oligo DNA pairs were mixed with dNTP (200 μM) in Klenow buffer (Takara Bio, Kusatsu, Shiga, Japan) and annealed by heating at 90℃ for 1 min and then cooling slowly to room temperature. Subsequently, 4 U of Klenow fragment (Takara Bio, Kusatsu, Shiga, Japan) was added to the reaction mixture. Primer was extended at 25℃ for 5 min and then 37℃ for 1 h followed by inactivation of the enzyme at 50℃ for 15 min. The linearized DNA libraries, VMM/VRR codon library, and HHY codon library were then purified with NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, Germany). The DNA library was quantified by agarose gel electrophoresis (agarose gel (2.5%, w/v) in 1x tris-acetate-EDTA (TAE)). The electrophoresis was run at 100 V for 40 min. The gel was then stained with SYBR™ Gold Nucleic Acid Gel Stain (Invitrogen, Waltham, MA USA). Gel images were captured and analyzed using Amersham Imager 680 (GE Healthcare, Chicago, IL, USA). Nucleic acid concentrations were quantified with NanoDrop™2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

2.2.2 *In vitro* transcription.

RNA libraries were produced by *in vitro* transcription from a DNA library. ScriptMax® Thermo T7 Transcription Kit (TOYOBO, Osaka, Osaka, Japan) was used for *in vitro* transcription. The transcription was carried out with 200 ng (4 pmol, 2.4×10^{12} molecules) of the linear DNA molecules according to the manufacturer's protocol. The reaction mixture was incubated at 40℃ for 2 h. The RNA transcripts were then purified using NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, Germany)

2.2.3 Puromycin-DNA-tag ligation.

The mRNA libraries attached with a puromycin-DNA-tag were generated by the Y-ligation method (Nishigaki et al. 1998). Puromycin-DNA-tag labeled with fluorescein isothiocyanate (FITC) was synthesized by Japan Bio Services Co., LTD (Asaka, Saitama, Japan). Initially, the mRNA library and the puromycin-DNA-tag were partially annealed. A reaction mixture containing the transcribed mRNA (20 μg), puromycin-DNA tag (8 μM), ATP (1 mM), and T4 ligation buffer (New England Biolabs, Inc., Ipswich, MA, USA) was then incubated at 90℃ for 30 s and cooled to room temperature. Subsequently, 3 U of T4 polynucleotide kinase (PNK; New England Biolabs, Inc., Ipswich, MA, USA) and 20 U of T4 RNA ligase (New England Biolabs, Inc., Ipswich, MA, USA) were added to the mixture and incubated at 25℃ for 30 min to ligate the 5′ end of puromycin-DNA-tag to the 3′ end of the mRNA library. Next, mRNA-tag ligation products were isolated by urea polyacrylamide gel electrophoresis. For the electrophoresis, 40 mL of a separating gel consisting of urea (8 M), polyacrylamide (6%, w/v), and 1x tris-borate-EDTA (TBE) was run at 50 mA for 1 h. The

ligation products in the gel were detected by the fluorescence specific for FITC with a blue LED transilluminator (FUJIFILM Wako Pure Chemicals, Osaka, Osaka, Japan) and excised with a clean scalpel. Finally, the mRNA tag was recovered by gel-purification.

2.2.4 Synthesis of mRNA-peptide conjugate by *in vitro* translation.

mRNA-peptide conjugates were synthesized from the mRNA-tag *in vitro*. PURExfrex 1.0 (GeneFrontier Corp., Kashiwa, Chiba, Japan) was used for the *in vitro* translation. Initially, a reaction mixture (100 μL) containing 800 ng of the mRNA tag was prepared according to the manufacturer's protocols. The reaction mixture was incubated at 37°C for 30 min. Then, the salt mixture was added to the reaction to bring its final concentration to $MgCl_2$ (32.5 mM) and KCl (375 mM). Following that, the reaction mixture was incubated again at 37°C for 1 h. An equal amount of 2x laemmli sample buffer (Bio-Rad, Hercules, CA, USA) was then added to the reaction and centrifuged at 10,000 × g for 3 min to remove the precipitants. The supernatant was collected and loaded on a SDS-Urea-PAGE gel (SDS-PAGE stacking gel (3.5%, w/v), SDS-6 M Urea-PAGE resolving gel (10%, w/v)). The electrophoresis was conducted at 30 mA for the stacking gel and at 50 mA for the resolving gel. Note that a gradient SDS-PAGE gel (4–12%, w/v; TEFCO, Hachioji, Tokyo, Japan) was used from round 3 to round 6 of the VMM/VRR codon libraries instead of the SDS-Urea-PAGE gel, since we could separate the mRNA-tag and mRNA-peptide conjugates with higher resolution. For a gradient SDS-PAGE gel, electrophoresis was conducted at 60 mA for 90 min. Finally, the mRNA-peptide conjugates were isolated by excision as previously described (Reyes et al. 2021), and purified by gel-purification.

2.2.5 Gel-purification.

A gel slice containing the desired band was purified from a gel through electroelution followed by ethanol precipitation. For electroelution, Model 422 Electro-Eluter (Bio-Rad, Hercules, CA, USA) apparatus was used along with a glass tube, a frit, and a membrane cap (molecular weight cut-off: 12 kDa; Bio-Rad, Hercules, CA, USA). After this apparatus was properly assembled, the kneaded gel slice was loaded into the glass tube filled with 1x TBE buffer. The sample was eluted at 10 mA/glass tube for 45 min. At the end of elution, the polarity of the initial current was reversed for 30 s to dissociate the purified products from the dialysis membrane. Then, the eluate was carefully collected into the microcentrifuge tube. As for ethanol precipitation, sodium acetate (3 M) equivalent to 10% volume of the obtained volume was added to the eluate and mixed well. A volume of pre-chilled ethanol (99.5%, v/v) equal to three times the total sample was further added to the mixture. The sample mixture was then centrifuged at 4℃, 15,000 × g for 1 h. After removing the supernatant, 1 mL of the chilled ethanol (70%, v/v) was added to the microcentrifuge tube. The dislodged pellet in ethanol (70%, v/v) was centrifuged again at 4℃, 15, 000 × g for 15 min. Subsequently, the supernatant was removed, and the pellet was air-dried at room temperature to evaporate the residual ethanol. Finally, the pellet was resuspended in nuclease-free water and the mRNA concentration was measured with NanoDrop^{TM}2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

2.2.6 Negative selection of mRNA-tag by the affinity with polyRNA.

The mRNA-tag with affinity to biotinylated polyRNA was removed in this pretreatment. First, 20 μL of 25-μm strep-tactin-immobilized magnetic beads (MagStrep "type3" XT Beads; IBA Lifesciences, Göttingen, Germany) was captured by a magnetic stand (DynaMag-2 Magnet; Thermo Fisher Scientific, Waltham, MA, USA) and washed twice with RNA-binding buffer (Tris-HCl (10 mM; pH 7.4), EDTA (0.5 mM), NaCl (500 mM), Tween 20 (0.05%, v/v)). After that, the polyRNA was immobilized on the magnetic beads. Biotinylated polyRNA (2.5 μM) in 200 μL of the RNA-binding buffer was added to the magnetic beads and mixed at 4°C for 30 min with a rotator (80 rpm). After discarding the supernatant, the magnetic beads were washed three times with the RNA-binding buffer. Subsequently, 25 μL of the mRNA-tag (100 ng/μL) in the RNA-binding buffer (Tris-HCl (10 mM; pH 7.4), EDTA (0.5 mM), NaCl (500 mM), Tween 20 (0.05%, v/v)) was added to the magnetic beads and mixed for 30 min at room temperature with a rotator (80 rpm). Finally, the supernatant was collected for the following *in vitro* translation.

2.2.7 Peptide selection based on the affinity with polyRNA.

The mRNA-peptide conjugates with affinity to polyRNA were selected over the affinity of its mRNA-tag portion. First, two sets of strep-tactin-immobilized magnetic beads (20 μL) were washed twice with the RNA-binding buffer by magnetic stand. Next, the target polyRNA was immobilized to one of those magnetic beads as described in the previous section. Subsequently, 100 ng of mRNA-peptide conjugates (1.7 pmol, 1.0×10^{12} molecules) was prepared in the RNA-binding buffer (200 μL), and the solution was first added to the intact

magnetic beads and mixed for 10 min at room temperature with a rotator (80 rpm) to remove the conjugates with affinity to the surface of the beads. The supernatant, containing mostly the nonspecific binding-free mRNA-peptide conjugates, was then recovered and used for the affinity selection with the target RNA. The entire supernatant was added to the target RNA-immobilized magnetic beads and mixed for 30 min at room temperature with a rotator (80 rpm). After the removal of the supernatant, the magnetic beads were washed three times with 1x Buffer W (IBA Lifesciences, Göttingen, Germany). The elution buffer (25 μL) (Tris-HCl (100 mM; pH 8.0), NaCl (150 mM), EDTA (1 mM), biotin (50 mM)) was then added and vigorously mixed. Finally, the mixture was incubated for 10 min at room temperature. The eluate was collected and used for reverse transcription polymerase chain reaction (RT-PCR).

2.2.8 RT-PCR.

The eluate mRNA-peptide conjugates (10 μL) were used as a substrate for the RT reaction using ReverTra Ace-α- (TOYOBO, Osaka, Osaka, Japan) with RT-PCR-F and RT-PCR-R primers (each 0.2 μM) by following the manufacturer's protocol. RT reactions (20 μL) were started by adding the reverse transcriptase (reverse transcription reaction conditions: 50℃, 20 min, 99℃, 5 min, and 4℃, 5 min). The reaction was then mixed with 20 μL of Q5 High-Fidelity 2x Master Mix (New England BioLabs, Ipswich, MA, USA) and aliquoted into a total of six tubes to recover samples at different PCR cycles (0, 10, 15, 20, 25, and 30 cycles) to check the amplification efficiency. PCR condition was set as follows: initial denaturation at 98℃ for 30 s, followed by 30 cycles of 98℃ for 10 s, 65℃ for 10 s, and 72℃ for 10 s, then final extension at 72℃ for 30 s. To avoid over-amplification of DNA, these reaction samples were run on a TAE agarose gel (3%, w/v) stained by 1x SYBR Gold, and the optimal amplification cycle was

determined based on the band intensity and the absence of non-specific amplified products. Finally, RT-PCR reaction (40 μL) was once again conducted with optimal PCR cycles, and the product was purified using NucleoSpin™ Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, Germany) for the next round of selection and sequencing.

2.2.9 Next-generation sequencing.

Next-generation sequencing (NGS; Next-generation sequencing) was applied to the DNA library from each round (round 0 to round 7) using a MiSeq system (Illumina; San Diego, CA, USA). Qubit 2.0 fluorometer (Invitrogen, Waltham, MA, USA) with Qubit dsDNA HS assay kit (Thermo Fisher Scientific, Waltham, MA, USA) was used to quantify sequencing libraries precisely. DNA sequencing libraries were generated using NEBNext Ultra II DNA Library Prep kit (New England Biolabs, Inc., Ipswich, MA, USA). To enable multiplex analysis, unique index sequences were added to the DNA sequencing library by NEBNext Multiplex Oligos for Illumina (Index Primers Set 1, 2; New England Biolabs, Inc., Ipswich, MA, USA). Finally, the lengths of the constructed DNA libraries were confirmed by Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and paired-end sequencing was performed using Miseq Reagent Kit v3 (600 cycles).

2.2.10 Sequence analysis.

Sequence files were obtained from the MiSeq Illumina platform as FASTQ files. A self-made pattern search program written in Perl script was used to extract the coding region, which is the region between the upstream of the start codon and the fixed sequence (8 nucleotide sequence) of the random region. Then, by using the count option of the software called

FASTAptamer (Alam, Chang, and Burke 2015), the number of identical sequences was counted and ranked for each round (the populations of raw DNA sequences obtained *via* MiSeq Illumina platform, extracted peptide sequences, and identical peptide sequences were listed in **Table 2-2**). Since the number of reads acquired in each round was different, we followed the progression of enrichment based on the normalized reads per million (RPM) value. For round 7, we performed clustering analysis on peptide sequences with RPM values of 10 or higher to create clusters based on their sequence similarity. In this case, we used the cluster option of FASTAptamer to define a cluster as a group of peptide sequences with a distance of 7 or less between peptide sequences. The sequence logo of the selected RNA-binding peptides was created using WebLogo 3 server (Crooks et al. 2004), and log2-fold changes of amino acid frequencies after selection was calculated from the average amino acid frequencies in round 7 and 0.

2.2.11 Calculation of hydrophobicity and net charge of peptides.

The hydrophobicity and net charge index of peptides were calculated using "Peptides" ver 2.4.3 package 50 of R ver 4.0.2 (Osorio, Rondón-Villarreal, and Torres 2015). The GRAVY score (Grand average of hydropathy), which is a hydrophobicity index of peptides, was calculated based on Kyte and Doolittle's amino acid hydrophobicity index (Kyte and Doolittle 1982). The net charge index at pH 7.4 was calculated using the Henderson-Hasselbalch equation using Lehninger's pKa scale (Butterworth 2005).

2.2.12 Enrichment analysis of combinatorial amino acid pairs.

The frequency of combinatorial dipeptide motifs of peptide libraries from round 0 and round 7 was calculated, and the enrichment score was derived from the following equation. Hierarchical clustering was then used to visualize the enrichment score of dipeptide motifs by heat mapping.

$$Enrichment\ score\ (\%) = \frac{Average\ frequency\ of\ dipeptide\ motif\ in\ \ round\ 7\ -\ Average\ frequency\ of\ \ dipeptide\ motif\ in\ round\ 0}{Average\ frequency\ of\ dipeptide\ motif\ in\ round\ 0} \times 100$$

2.2.13 Spectral shift experiment.

Chemically synthesized cy5-labeled oligonucleotides were obtained from FASMAC. While AP1 and AP2 were completely dissolved in the Binding buffer (Tris-HCl (10 mM; pH 8.0) and Tween 20 (0.05%, v/v)), both CP1 and CP2 formed aggregates when the concentration reached above 10 μM. Therefore, aggregation-free spectral shift (SpS) measurement was applied to determine the dissociation constant (K_D). SpS experiment was conducted with the identified RNA-binding peptides in three independent experimental trials on a Monolith X (NanoTemper Technologies, Munich, Bayern, Germany) (Langer et al. 2022). The oligonucleotide solutions were prepared in Tris-HCl (20 mM; pH 8.0), Tween 20 (0.1%, v/v), and the peptide solutions were prepared in RNase-free water. A two-fold dilution series of the unlabeled peptides was prepared in 20 nM oligonucleotide, with the final concentrations of peptides ranging from 100 μM to 3 nM or 500 μM to 15 nM. Samples were incubated for 30 min at room temperature. Following incubation, the samples were filled into standard treated

capillaries (NanoTemper Technologies, Munich, Bayern, Germany) and subsequently subjected to SpS detection analysis. General settings were applied for all SpS experiments as follows: manual temperature control: 25℃, LED laser: red. LED power settings were chosen individually for each sample by adjusting the LED power to yield fluorescence signals of at least 200 units. For SpS detection, the fluorescence was recorded simultaneously at 650 and 670 nm for 5 s without applying a temperature change. The values obtained were normalized and plotted against the peptide concentration. The dissociation constant was then determined using a single-site model to fit the curve.

2.2.14 Microscale thermophoresis.

Chemically synthesized fluorescein (FAM)-labeled CP1 peptide was obtained from genscript (Piscataway, NJ, USA). The oncogenic RNA motifs were synthesized by *in vitro* transcription described in the previous section. The RNA transcripts were then purified using Monarch RNA Cleanup Kit (New England Biolabs, Inc., Ipswich, MA, USA). Microscale thermophoresis experiment was conducted with a set of biologically derived RNAs in three independent experimental trials on a Monolith NT.115 (NanoTemper Technologies, Munich, Bayern, Germany) unless specified. Both the FAM-labeled CP1 peptide and RNA were prepared in a human plasma-like medium (Gibco, Billings, MT, USA) supplemented with 0.1 mg/mL bovine serum albumin (Sigma, St. Louis, MO, USA). A two-fold dilution series of the unlabeled RNAs was prepared in 50 nM FAM-labeled CP1 peptide, with the final concentrations of RNAs ranging from 2 μM to 0.06 nM. Samples were incubated for 30 min at room temperature. Following incubation, the samples were filled into standard treated capillaries (NanoTemper Technologies, Munich, Bayern, Germany) and subsequently subjected

to microscale thermophoresis. General settings were applied for all microscale thermophoresis experiments as follows: manual temperature control: 25℃, LED laser: blue, fluorescence measurement before microscale thermophoresis: 5 s, microscale thermophoresis (IR laser) on: 45 s, fluorescence after microscale thermophoresis: 15 s, delay: 25 s. LED and microscale thermophoresis power settings were chosen individually for each sample by adjusting the LED power to yield fluorescence signals of at least 200 units and the microscale thermophoresis power to achieve an appropriate thermophoretic response (standard setting 20% microscale thermophoresis power, for a few experiments 40 to 60%). The values obtained were normalized and plotted against the peptide concentration. The dissociation constant was then determined using a single-site model to fit the curve.

2.2.15 Circular dichroism spectroscopy.

The circular dichroism (CD) spectra were recorded using a CD spectrophotometer J-1100 (JASCO, Hachioji, Tokyo, Japan) over the wavelength range 190–250 nm in steps of 1 nm with an averaging time of 1 s per step. Peptides were prepared in a buffer containing Tris-HCl (pH 8.0; 10 mM) and Tween 20 (0.05%, v/v) with the final concentration of 0.2 mg/mL and the CD spectra were measured in 1 mm path-length quartz cells at 25℃. The CD signal was obtained as ellipticity in units of millidegrees, and the resulting spectra were averaged from two scans and buffer-spectrum subtracted. All CD measurements were performed twice.

Table 2-1. Complete list of DNA and RNA oligos used in this work.

Name	Sequence (5′ to 3′)
PolyAGCU-binding-F	GGCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATG
(VMM/VRR)n-library-R	TGGTTCACCTGATCCGCTGCTGGGTCC**KKBYYBKKBKKBYYBKKBYYBKKBYYBYYBKKB YYBKKBKKBYYBKKBYYBKKBYYBYYBKKBYYBKKBKKBYYBKKBYYBKKBYYBYYBKK BYYBKKB**CATGTTTTTCTCCTTGTTAAAG *B = C/G/T, K = G/T, Y = C/T
(SBB)n-library-R	TGGTTCACCTGATCCGCTGCTGGGTCC**VVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSV VSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVSVVS VVS**CATGTTTTTCTCCTTGTTAAAG *V = A/C/G, S = C/G
(RDK)n-library-R	TGGTTCACCTGATCCGCTGCTGGGTCC**MHYMHYMHYMHYMHYMHYMHYMHYMHYMHYMH YMHYMHYMHYMHYMHYMHYMHYMHYMHYMHYMHYMHYMHYMHYMHYMHYMHYMHYMH YMHYMHYMHY**CATGTTTTTCTCCTTGTTAAAG *M = A/C, H = A/C//T, Y = C/T
(HHY)n-library-R	TTTTTCACCTGATCCGCTGCTGGGTGA**RDDRDDRDDRDDRDDRDDRDDRDDRDDRDDRDD RDDRDDRDDRDDRDDRDDRDDRDDRDDRDDRDDRDDRDDRDDRDDRDDRDDRDDRDDR DDRDDRDD**CATGTTTTTCTCCTTGTTAAAG *D = A/G/T, R = A/G
RT-PCR-F (PolyAGCU-binding-F)	GGCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATG
RT-PCR-R	TTTTTCACCTGATCCGCTGC
Puromycin-DNA-tag	pCCCTTCACCTGATCCGCTGAAAAAAAAAAAAAAAAAA (Spacer 18) (Spacer 18) (F-dT) (Spacer 18) CC (Puro) (Spacer 18): Hexaethylene glycol, (F-dT) = Fluorescein-dT, (Puro) = Puromycin,
5′ biotin-Poly(C) RNA	(Biotin)-CCCCCCCCCCCC
5′ biotin-Poly(A) RNA	(Biotin)-AAAAAAAAAAAA
5′ Cy5-Poly(A) DNA/RNA	(Cy5)-AAAAAAAAAAAA
5′ Cy5-Poly(C) DNA/RNA	(Cy5)-CCCCCCCCCCCC

Table 2-2. The populations of raw DNA sequences, extracted peptide sequences, and identical peptide sequences obtained from individual round *via* next-generation sequencing.

VMM/VRR codon library

Round	Raw DNA sequences	Extracted peptide sequences	Identical peptide sequences
0	3,163,494	1,651,906	1,649,290
1	2,851,143	1,227,819	662,161
2	2,919,380	1,242,922	492,709
3	3,214,793	1,801,552	722,051
4	741,515	433,352	205,768
5	700,994	362,686	135,282
6	690,762	325,866	98,284
7	671,754	453,480	131,443

HHY codon library

Round	Raw DNA sequences	Extracted peptide sequences	Identical peptide sequences
0	1,078,498	523,504	522,199
1	1,664,773	1,333,261	350,788
2	3,110,416	874,731	329,568
3	4,110,723	1,206,002	362,360
4	4,365,343	1,112,944	271,727
5	4,525,495	1,210,037	247,523
6	3,777,180	1,197,653	200,787
7	6,247,854	572,922	86,826

2.3 Results

2.3.1 Peptide library design.

The scheme for codon-restricted mRNA display is outlined in (**Figure 2-1**). I designed two types of DNA libraries encoding 33-mer randomized polypeptide sequences, each targeting a 12-mer poly(A) and poly(C) RNA (**Figure 2-2A**). To avoid unwanted base-pairing between mRNA and the target RNA, codons including the complementary nucleotide were eliminated from the coding sequence**.** For example, the DNA library targeting poly(A) RNA comprises two types of degenerate "VMM" and "VRR" codons both lacking thymine nucleotide, which encodes a total of nine (Ala/Thr/Asn/Asp/Glu/Gln/Lys/His/Pro) and five (Gly/Gln/Glu/Arg/Lys) types of amino acids, respectively. Similarly, the DNA library targeting poly(U) RNA comprises a single degenerate "SBB" codons lacking adenine nucleotide, which encodes a total of six (Gly/Ala/Leu/Val/Arg/Pro) types of amino acids. The DNA library targeting poly(G) RNA comprises a single degenerate "RDK" codons lacking cytosine nucleotide, which encodes a total of ten (Gly/Ile/Met/Val/Ser/Asn/Asp/Glu/Arg/Lys) types of amino acids. Finally, the DNA library targeting poly(C) RNA comprises a single degenerate codon "HHY" lacking guanine nucleotide, which encodes a total of nine types of amino acids (Leu/Ile/Phe/Tyr/Ser/Thr/Asn/His/Pro).

2.3.2 *In vitro* selection of *de novo* RNA-binding peptides.

The four DNA libraries were generated (**Figure 2-2B**) and subjected to my renovated codon-restricted mRNA display method. A puromycin-DNA tag was attached to the 3' ends of the mRNA library created *in vitro* **(Figure 2-3)**. Next, through *in vitro* translation, we isolated

the mRNA-peptide conjugates **(Figure 2-4A)**. Due to insufficient efficiencies in forming RNA-peptide conjugates for the SBB and RDK codon library, the following selection procedures were only applied to the VMM/VRR and HHY codon library. The mRNA-peptide conjugates were subjected to affinity selection against poly(A) or poly(C) RNA. Finally, to generate a cDNA library for the next round of selection and deep sequencing, the eluted molecules were subjected to RT-PCR (**Figure 2-4B**). Seven rounds of selection were performed, with the cDNA sequences obtained from each round analyzed by NGS to monitor the enrichment of specific amino acids or peptide motifs expected to play a role in RNA binding.

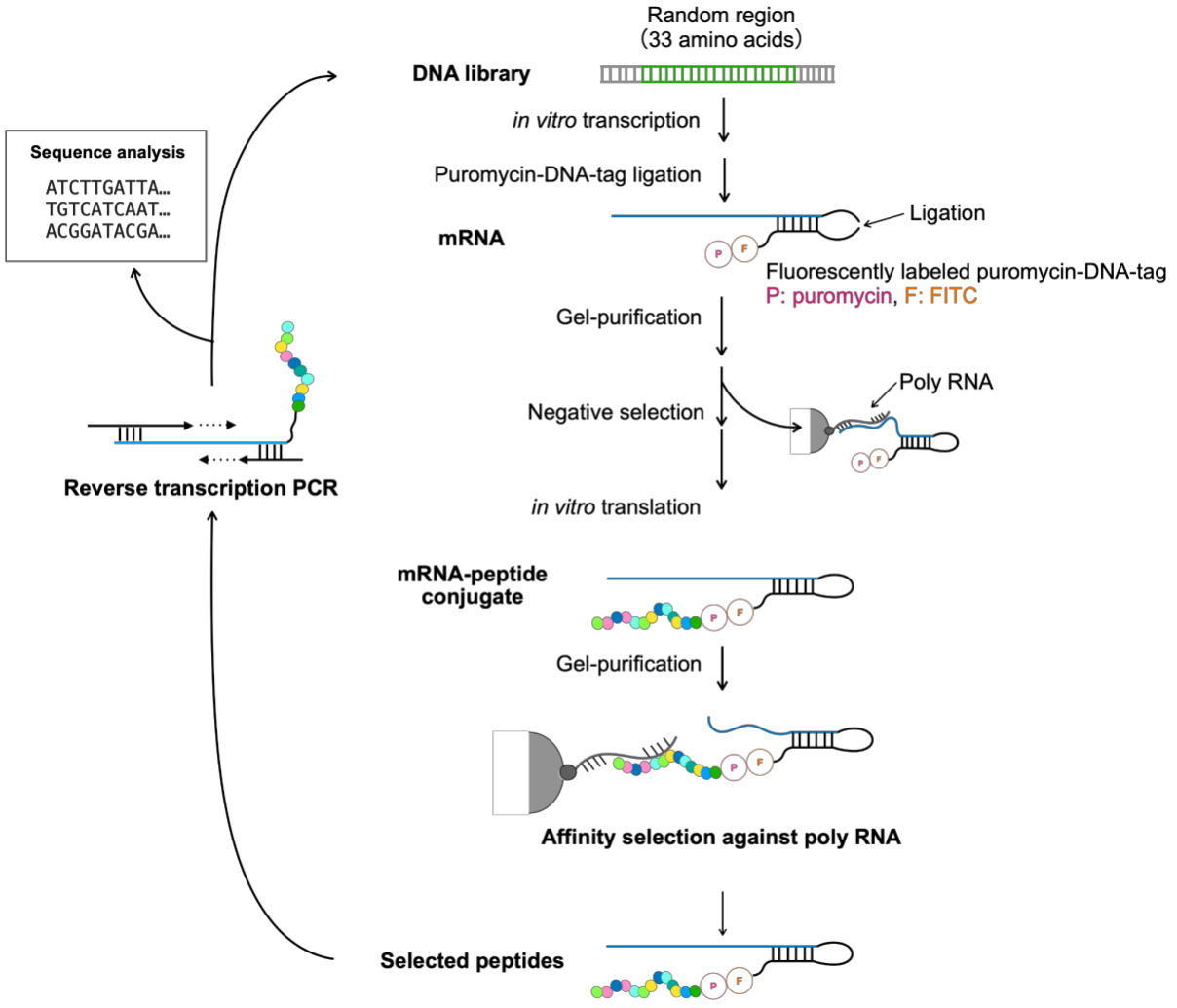


Figure 2-1. Schematic overview of codon-restricted mRNA display. This method starts with a designed codon-restricted DNA library (VMM/VRR or HHY codon library) encoding 33-amino acid random peptides (see Figure 2-2A). The DNA library is *in vitro* transcribed to mRNA library and undergoes ligation with the puromycin-DNA-tag. The tagged products are gel-purified and subjected to negative selection. The remaining tagged products are *in vitro* translated using a cell-free system and the resulting mRNA-peptide conjugates with a variety of 10^{12} sequences are gel-purified and subjected to affinity selection against polyRNA. The

selected peptide-mRNA conjugates are subjected to reverse transcription polymerase chain reaction (PCR) to the DNA library and used in the next round of selection. Also, DNA libraries from each round (round 0 to round 7) are subjected to Illumina MiSeq sequencing. FITC, fluorescein isothiocyanate.

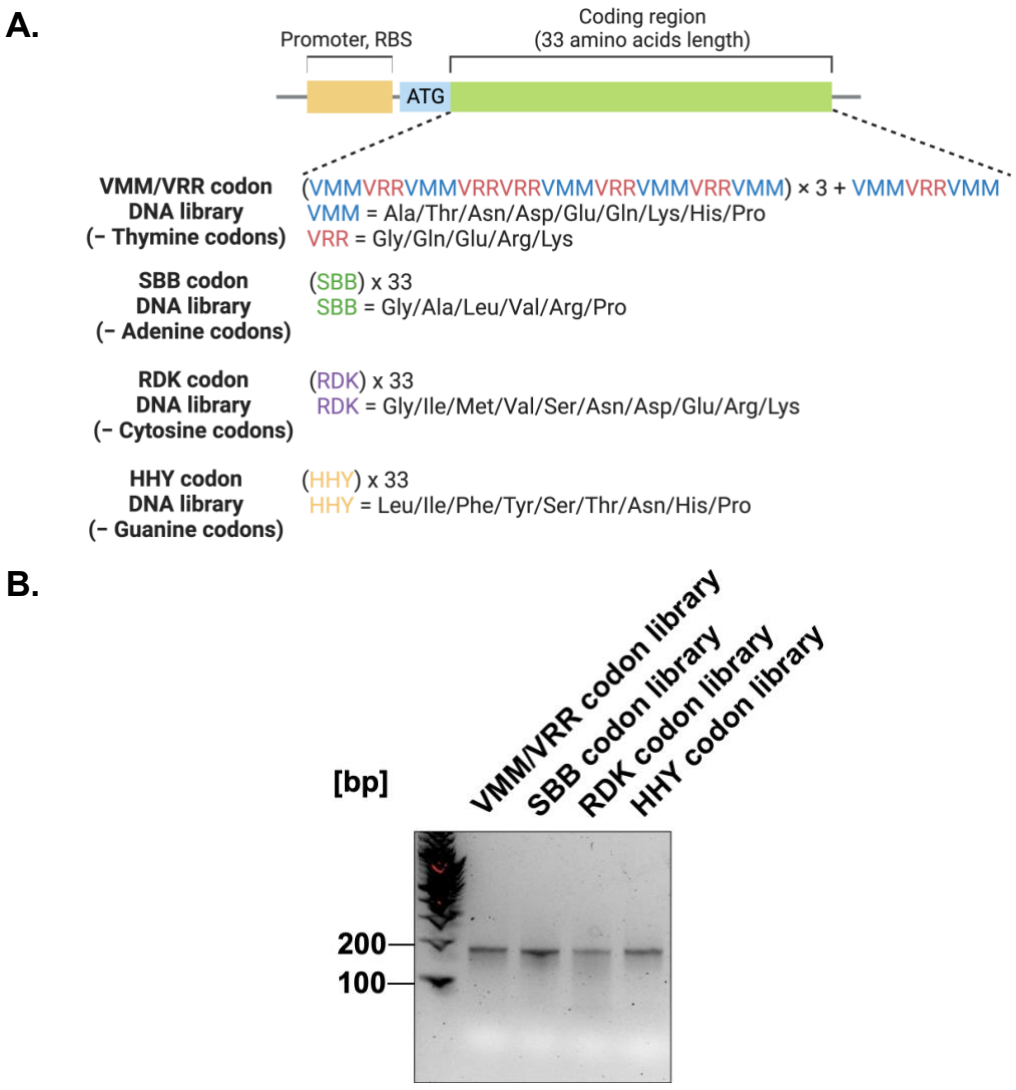


Figure 2-2. Selection of *de novo* peptide libraries against single-stranded RNA using codon-restricted mRNA display. (A) The construct of codon-restricted DNA libraries. RBS, ribosome binding site. (B) DNA libraries were synthesized *via* Klenow extension reaction. The gel images were taken after electrophoresis in a tris-acetate-EDTA (TAE) agarose gel (3%, w/v).

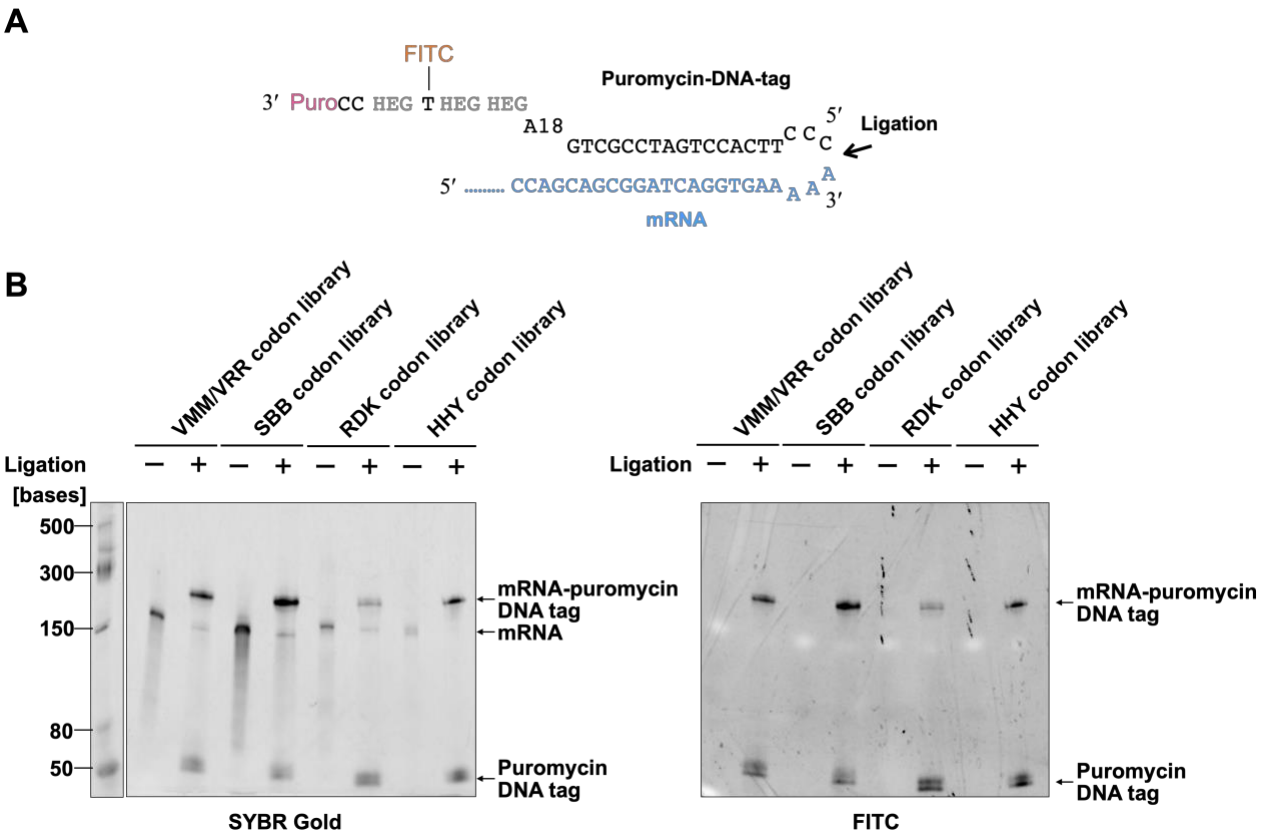


Figure 2-3. Generation of puromycin-DNA-tagged mRNA library *via* Y-ligation reaction. (A) Schematic overview of Y-ligation reaction. FITC, fluorescein isothiocyanate. (B) The gel images of mRNA and Y-ligation reaction mixture after electrophoresis on 8 M urea tris-borate-EDTA (TBE) gels (6%, w/v) (left: SYBR Gold staining, right: FITC fluorescence detection).

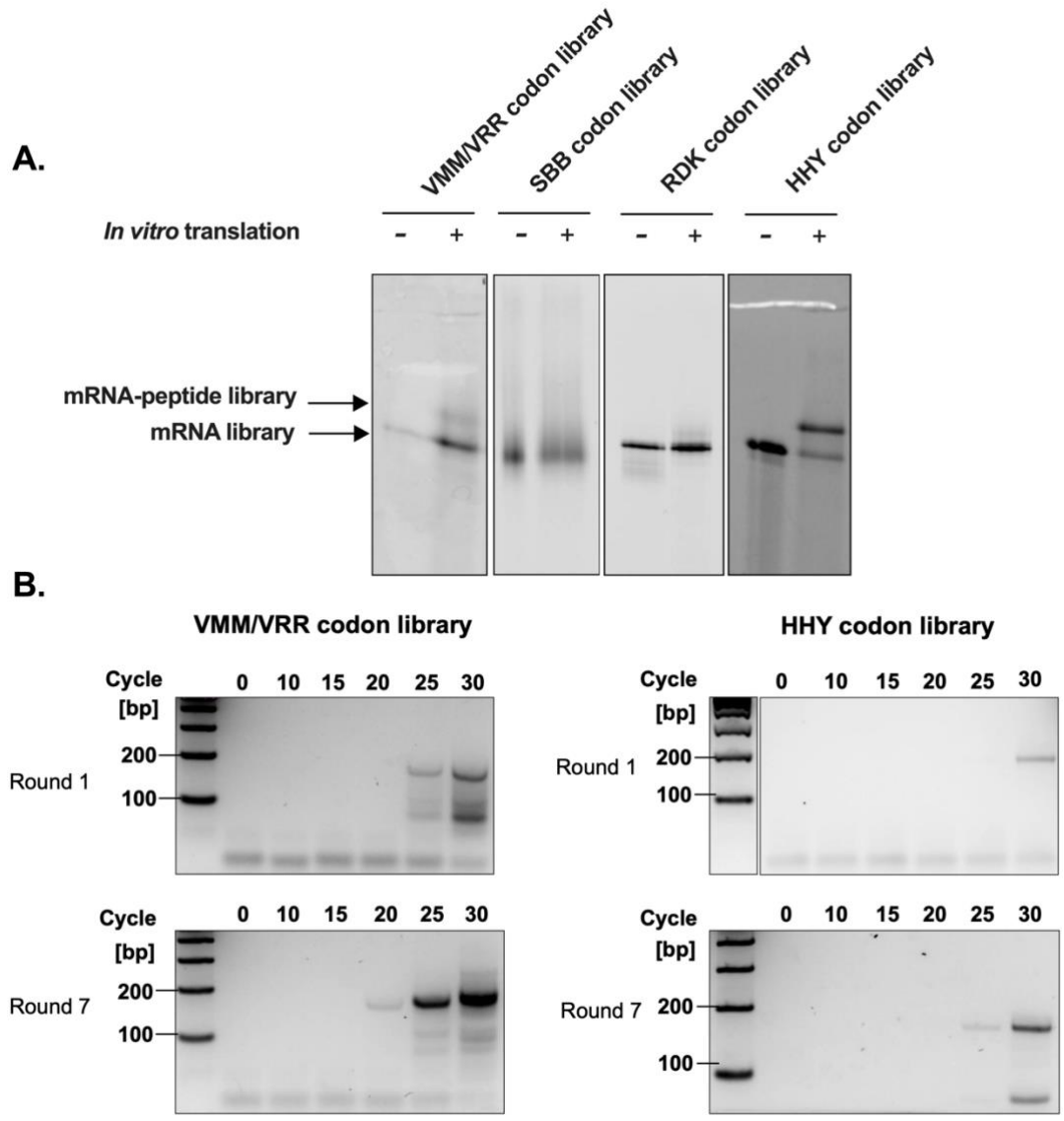


Figure 2-4. Synthesis of mRNA-peptide conjugates *via in vitro* translation. (A) *In vitro* translation reactions were conducted using puromycin-DNA-tagged mRNA libraries as templates and analyzed by SDS-PAGE. (B) cDNA libraries were synthesized from the mRNA-peptide conjugates selected by their targeted RNA binding capability *via* reverse transcription polymerase chain reaction (RT-PCR). The synthesized cDNA was electrophoresed on a tris-acetate-EDTA (TAE) agarose gel (3%, w/v) and visualized by 1x SYBR Gold.

2.3.3 Sequence analysis.

The peptide sequences obtained *via* deep sequencing were analyzed by FastAptamer software to count and rank the unique peptide sequences based on the RPM values (Alam, Chang, and Burke 2015). I first analyzed the top 100 enriched peptides from each round to compare the progression of enrichment throughout the selection. For the VMM/VRR codon library, the average RPM values of the top 100 enriched peptides constantly increased until round 6 (**Figure 2-5A**). In contrast, the HHY codon library has reached the RPM peak at round 3 and gradually decreases subsequently (**Figure 2-5B**). I further performed clustering analysis with the peptide sequences in round 7 from each library based on their sequence similarity. For the VMM/VRR codon library, the 30 most enriched sequence clusters account for only 12.1% of all peptide sequences with the most enriched cluster sequences sharing 1.5% of all peptide sequences (**Figure 2-5C**). In contrast, the HHY codon library was dominated by a single peptide cluster sequence covering 77.6% of all sequences, outcompeting the second-ranked cluster peptide sequences (**Figure 2-5D**).

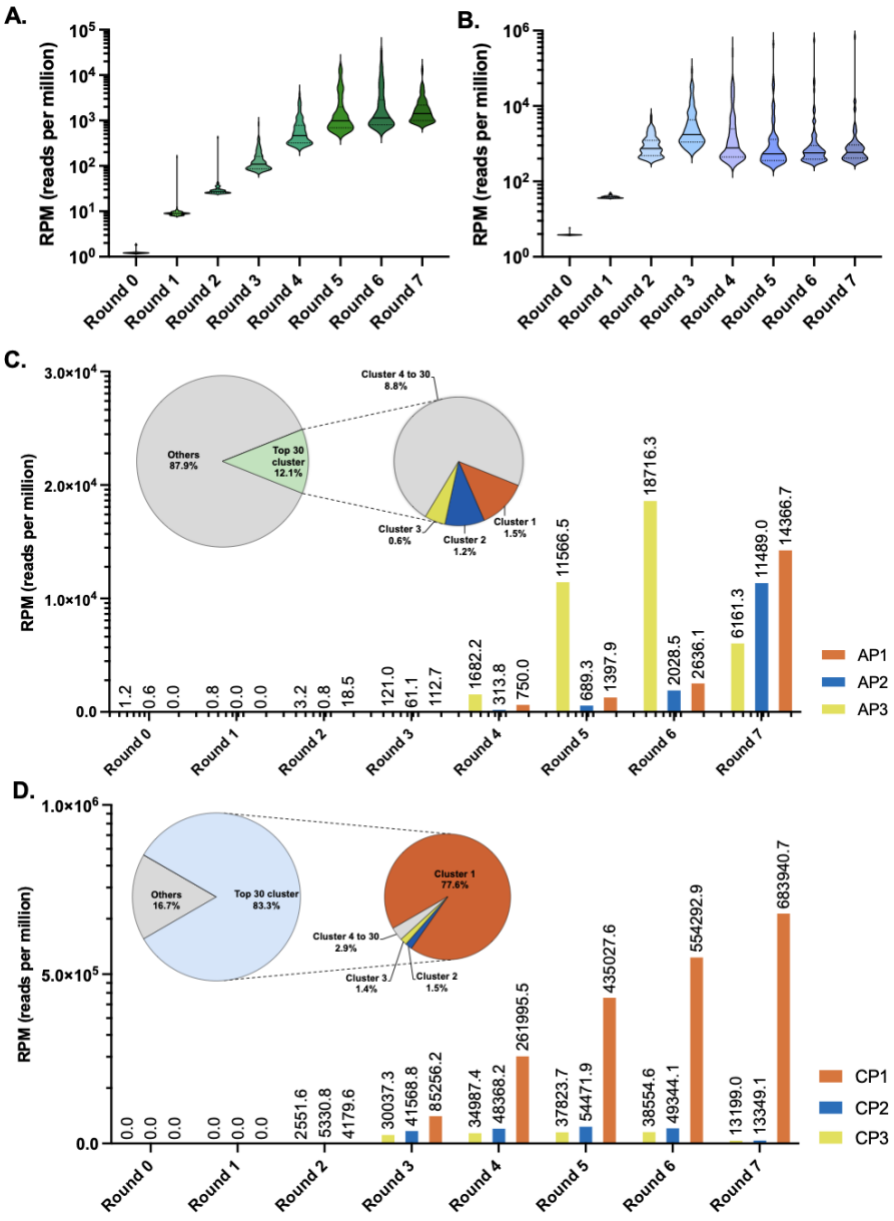


Figure 2-5. Sequence analysis of RNA-binding peptides. (A–B) Violin plots displaying reads per million (RPM) values for the top 100 enriched peptide sequences per selection round. The dashed lines indicate quartile values, while the solid line represents the mean. (C–D) Pie charts illustrating the proportions of the top 30 enriched peptide clusters from round 7 of screening. The RPM values of the top sequences within each cluster are depicted across the selection

rounds: (C) poly(A) RNA-targeting VMM/VRR codon library, (D) poly(C) RNA-targeting HHY codon library. AP1–3 and CP1–3 are the top sequences of the three most enriched clusters in round 7.

2.3.4 Amino acid composition analysis of enriched peptide sequences.

To further analyze the sequence profile of the enriched peptides, the consensus sequence logos were created from the top sequences of the most enriched 30 clusters (**Tables 2-3 and 2-4**). As a result, in poly(A) RNA-targeting VMM/VRR codon library, I observed a clear enrichment of glutamine (Q) appearing as the most frequently used amino acid at several positions (**Figure 2-6A**). Overall enrichment of asparagine (N) followed by threonine (T) and histidine (H) was observed throughout the poly(C) RNA-targeting HHY codon library (**Figure 2-6B**). Next, I calculated the log2-fold change of amino acid frequencies after seven rounds of selection. For the VMM/VRR codon library, the frequency of the positively charged amino acids Lys (K) and Arg (R), which are known to be common at RNA binding interfaces, decreased (log2-fold change = −0.46 and −0.33, respectively), while the frequency of the hydrophilic amino acids His (H), Gln (Q), Pro (P), and Thr (T) increased (log2-fold change = 0.94, 0.50, 0.32, and 0.20, respectively) (**Figure 2-7A**). Similarly, in the HHY codon library, the frequency of hydrophilic amino acids Asn (N), Thr (T), and His (H) increased (log2-fold change = 0.89, 0.56, and 0.29, respectively), and hydrophobic/aromatic amino acids Ile (I), Leu (L), Tyr (Y), and Phe (F) decreased (log2-fold change = −0.68, −0.79, −0.51, and −0.81, respectively) (**Figure 2-7B**). Also, the two-dimensional plot of net charge *vs.* hydrophobicity index has shown that the enriched peptides from the HHY codon library remained electrically neutral (at pH 7.4) while significantly shifting towards lower hydrophobicity within the peptide sequence space (**Figure 2-8**). Whereas the peptides from the initial VMM/VRR codon library possessed diverse net charge profiles (−5 to 15) along with overall low hydrophobicity profile (−2 to −3), enriched peptides converged toward a narrow net neutral to positive charge window

(0 to 5). The enrichment towards low hydrophobicity indicates the overall soluble feature of the selected peptide pool.

Table 2-3. The top peptide sequences of the most enriched 30 clusters in VMM/VRR codon library. The most dominant sequence within each cluster was listed as a representative sequence.

Cluster No.	Representative sequence	Unique sequences	RPM
1	MEEAKRTKTEEPKTQRHEEKQEQQKQPRHQATRP	19	15079.0
2	MPQPKQPQNQPERDQKNQNQAPQKQGTGARQKRP	10	12055.7
3	MDKEGGTQPRHTRAEKKENQTQQNARDRNQHQRP	4	6304.6
4	MPRNQQQKQKPARDQKTEHQNTRERGQRPETHRP	6	6644.2
5	MHKTEENQPKQHQHQEPKEQAHKTRGTRPRAQGN	4	5713.6
6	MNEDKQHRHKNTETKQPRTQTTRPQGPENETAKP	2	5182.2
7	MPEPKKAKQEKPEPEQTKPENKGAGQPRHQTARP	4	4862.4
8	MHQQKQTETETAKPRQHEPKTPQPERQRDRAKEH	3	4405.9
9	MPKNKQKRNQTKQDQKHEEQHQQTQKPQNQAHQN	4	4412.6
10	MARPRQHRPEPHRPEETGDETKQHQRDQNQQPGA	2	4156.8
11	MQKTQEPKDQEAQHQRQQKEQHRDQRPGTREHQP	2	3671.6
12	MTKHQQHQEQATRPQRNRNQKARPQGKRTEDKRP	6	3389.4
13	MTQEEKPQTRQNRHQKTQTQPPGEREHGAQHTEP	3	3155.6
14	MQEEEKTKKEEHEPKRPQQEPKQKRQHQPRDAGT	3	3144.6
15	MTREKKQKEQPAQQQKHQNQQPKTEGHQERHHKT	1	2968.2
16	MTQPERKRQETKEHKQTKHKETEHQGDRTKHTRD	5	3076.2
17	MAQEGRQEHQKTKAQQEKHRKDQQEETKHQAARP	1	2738.8
18	MTQNQRTKTQAPGDRENRPQKHENKQTETQTPQP	1	2652.8
19	MTKTKRTRNQDNKDQKPKDEAPQPKEHGPEAQGP	2	2661.7
20	MNEQQKQQTRTPETGQTKHENNQPKRQRTQNQRP	4	2641.8
21	MHKKEQPRDRTPENQQNQKQHTRNERARTKPPRP	2	2580.1
22	MHEDKKDEPQEQKHEKPRTQKNQTERKEDRAQEP	2	2549.2
23	MTRHQQHKTQTDEHEREEDETKRDRRPENRPART	2	2333.1
24	MTQPQKNQHRPNGTKRQEHGPNEHQQKRARAPQP	2	2293.4
25	MTQPKRTRPQNHQQKQNKPQNNRTRRKEPKEQEA	3	2255.9
26	MTEEERTQTKTERHQRPGDRTQQTKEHQHRDNKT	2	2220.6
27	MGKARRTKTEQQQAKRKENQHHQTEQQQHQHHRT	1	2123.6
28	MHEDQKQQKQHTREEQTKPEPKQTKQHRTRAQQD	4	2158.9
29	MEGAGRKGPQKTKTQETREKHAKHERTEHRPART	2	1991.3
30	MNQHQQTKTEDQRPGRTRAEQAQTREDRERHARP	1	1947.2

Table 2-4. The top peptide sequences of the most enriched 30 clusters in HHY codon library. The most dominant sequence within each cluster was listed as a representative sequence.

Cluster No.	Representative sequence	Unique sequences	RPM
1	MTHNHNYHHHITNNNINHTNPNNYTITNSPNNIY	362	775655.1
2	MTIFNTNHTNLYNNHTSNNNSTNSHTHHTPSSTT	37	14581.5
3	MHNNSNSNHTINNNTHHTTNNFILINYSYYHYHT	33	14426.1
4	MFNHYHTNNNTHNNHTNIHTTTHTINNPNHFYLL	21	9245.6
5	MGGPAAVLRGGAAVLVLVLLLVLLLVLALAPLAA	18	8783.1
6	MTFNYTTTHTTTTYNHYTTTNHHTNYHTSPYLYF	5	3349.5
7	MNYINHYNNHNNSNHNITNHNTLNHTPHLFFLFP	1	2597.2
8	MSYPYHTNTYNNIHNFPPFNHSNTYIYNTNHSLH	1	1131.1
9	MTSNFYHFYPHNNNHTNNTNHTILSNHNNYINHT	1	605.7
10	MHHTNSTSIYNIYNPTFYNSTNLHNTNNHYTNTN	1	549.8
11	MTNYTNTTHTNHNPPTNNNPYYTHFYTNLSPTNI	1	268.8
12	MYHINIIPINYTNNSNTNTNHYTLNNFINNTNHS	1	225.2
13	MINTFFNNSYTNTNTPYNNNTHNHTNNYTNPTHL	1	212.9
14	MHTNPTHPNNITSNSNHNNHINNNHSYYINSSYH	1	157.1
15	MTNTNTNHTNHNYNSYTTSNTNQHNYLLYILFNS	1	155.3
16	MYYIINNHPNNHYSNNTNNNHPSNNTINNLYYTT	1	141.4
17	MNNLNTIYPTTNHTYTNNHNHNTTNYTHHISHNN	1	129.2
18	MNHNIPIYNLTNNNTTTNFHNHNTNSNHNFIILH	1	120.4
19	MTFIHNHHHHHTNINHTNNSNIHNTTHSFFFPSL	1	113.5
20	MNNNTIFTHFNSHTNLNNNNNHNHNPISFFHTTH	1	106.5
21	MNYNSIIHNNNNHTTHFHTNNLNSTNNFIIHTNF	1	104.7
22	MLNNHINIYHHTTYIPNYNSNNNSHHTSIYFPST	1	97.7
23	MIHNSIHYTNTNPYNHTNSNITNYSNHTLFPTLH	1	90.8
24	MLNNHHTNHITTTLNHTNNNSNTIINSTYTIHNT	1	52.4
25	MTHNPIINSNYNYNYPFYNNNLTLNNHTNHNHTT	1	50.6
26	MHNSNNHTSYNHTHNILFNNNNSNNNNHTNLLHN	1	48.9
27	MNYFNYTNINNLNNHNNLTNHHPPNPPFTLNSFH	1	48.9
28	MTNTNTNHTNHNYNSYTTSNTNNTIIFFTFSSNS	1	41.9
29	MIFHYHTHNNYNTTSNLNYNHTNHIIHTSTLFHN	1	38.4
30	MTNYYINTHINNHTNYSNTTNPTSPLFHNIFHTS	1	34.9

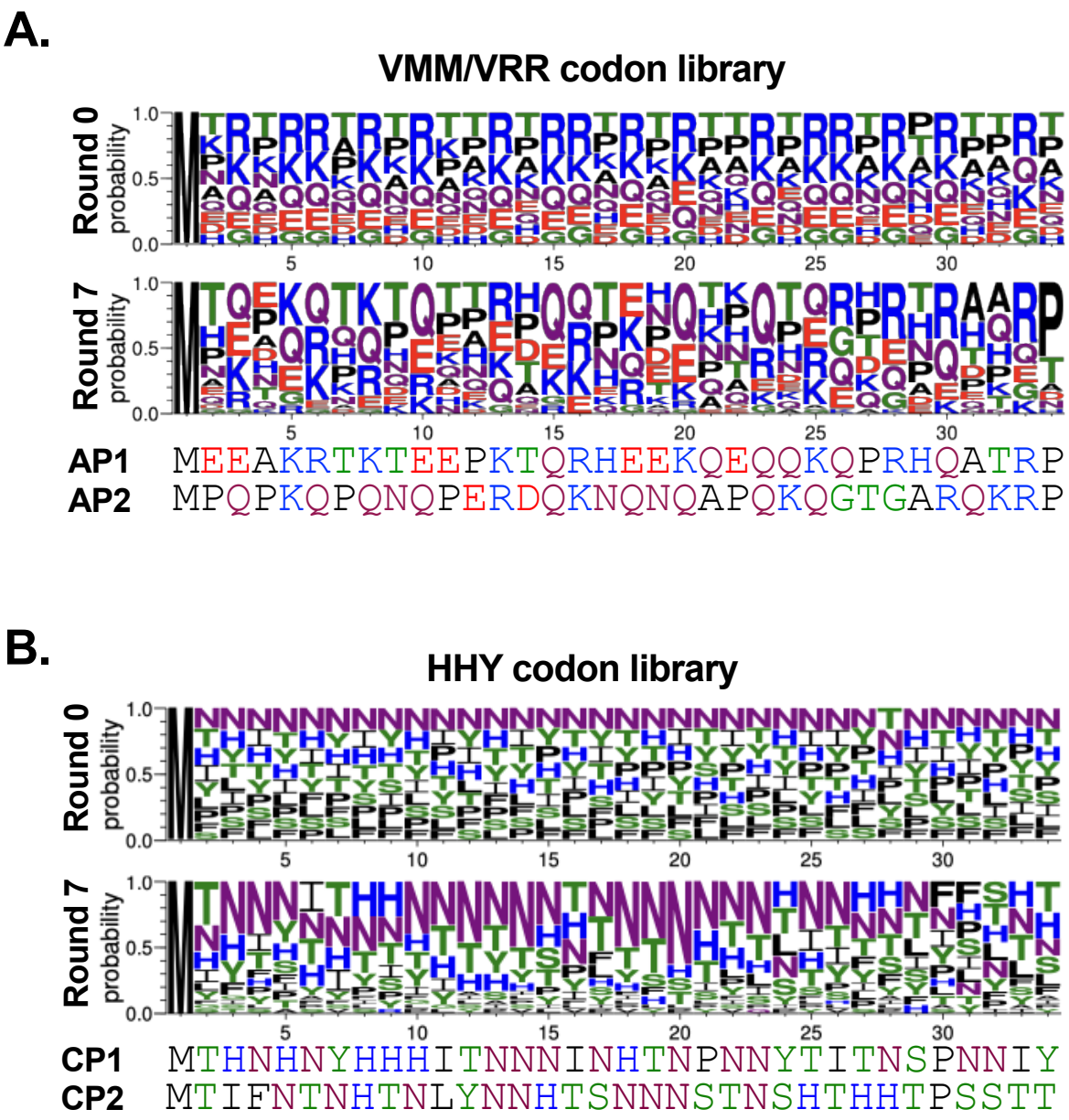


Figure 2-6. Amino acid composition feature of selected RNA-binding peptides. (A–B) Sequence logos were generated from peptide sequences in round 0 and round 7, respectively. For round 0 (before screening), 1,000 randomly selected peptide sequences were used. The top sequences of the two most enriched clusters are listed below: (A) VMM/VRR codon library, (B) HYY codon library.

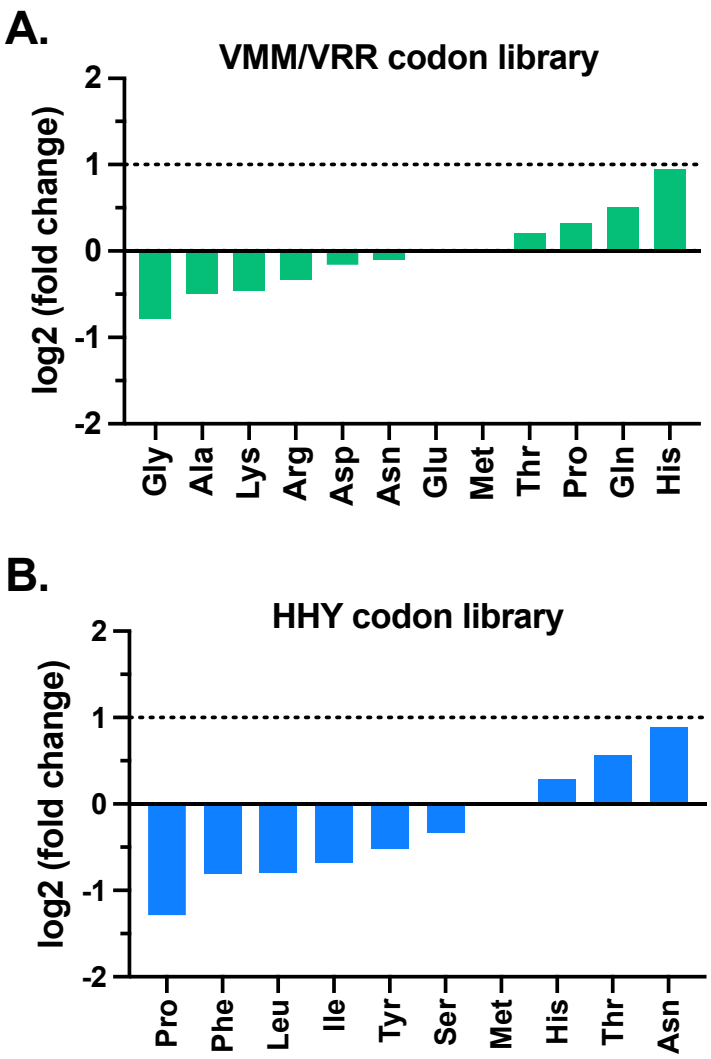


Figure 2-7. Log2-fold changes of amino acid frequencies after the whole selection. The dotted line indicates log2-fold change threshold of 1.0: (A) VMM/VRR codon library, (B) HHY codon library.

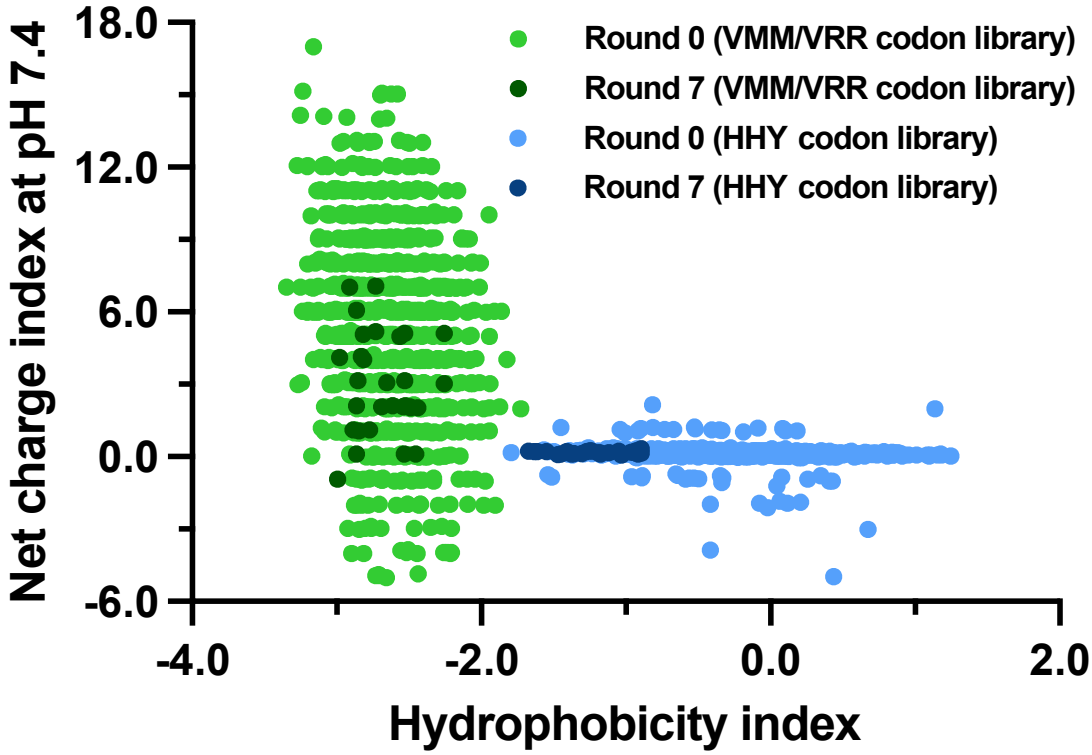


Figure 2-8. Two-dimensional plot representing hydrophobicity *vs.* net charge indexes of individual peptides. Each dot represents individual peptides from round 0 (randomly selected 1000 sequences) and round 7 (top 30 sequences) libraries.

2.3.5 Dissociation constant analysis of the identified RNA-binding peptides.

I chose the top two most enriched sequence clusters from each library (CP1/CP2 and AP1/AP2) to test their affinity against polyRNAs. As a result, AP1 only expressed very low affinity towards poly(A) RNA (**Figure 2-9A**), whereas the three peptides AP2, CP1, and CP2 have exhibited apparent affinity to their targeted polyRNAs with K_D of 18.8 μM, 0.2 μM, and 1.2 μM, respectively (**Figures 2-9BCD**). CP1 and CP2 exhibited slightly high affinity with the target poly(C) RNA over the non-target poly(A) RNA, although they retain the affinity with poly(A) RNA within a range of single micromolar (K_D) (**Figures 2-9CD**). On the contrary, AP2 revealed one to two orders of magnitude lower binding affinity towards its target RNA than CP1 and CP2; however, AP2 exhibited specific interaction with poly(A) RNA (**Figure 2-9B**).

Further, by using CP1 and AP2 as a model for *de novo* RNA-binding peptides, I tested the sugar recognition by switching the target sequence from polyRNA to polyDNA. As a result, both AP2 and CP1 still exhibited affinity toward their DNA targets but with approximately two to five-fold less affinity compared to that of RNA (**Figures 2-10AB**). Then, to examine the role of nucleobase recognition in RNA-peptide interactions, I performed additional affinity analysis using the RNA hybridized with antisense DNA. My findings revealed that AP2 lost its affinity for hybridized poly(A) RNA (**Figure 2-11A**). Similarly, CP1 exhibited a significantly lower affinity for poly(C) RNA, with approximately a 30-fold decrease compared to its affinity for non-hybridized poly(C) RNA (**Figure 2-11B**). This observation strongly suggests that nucleobase-peptide interactions play a significant role in determining the affinity.

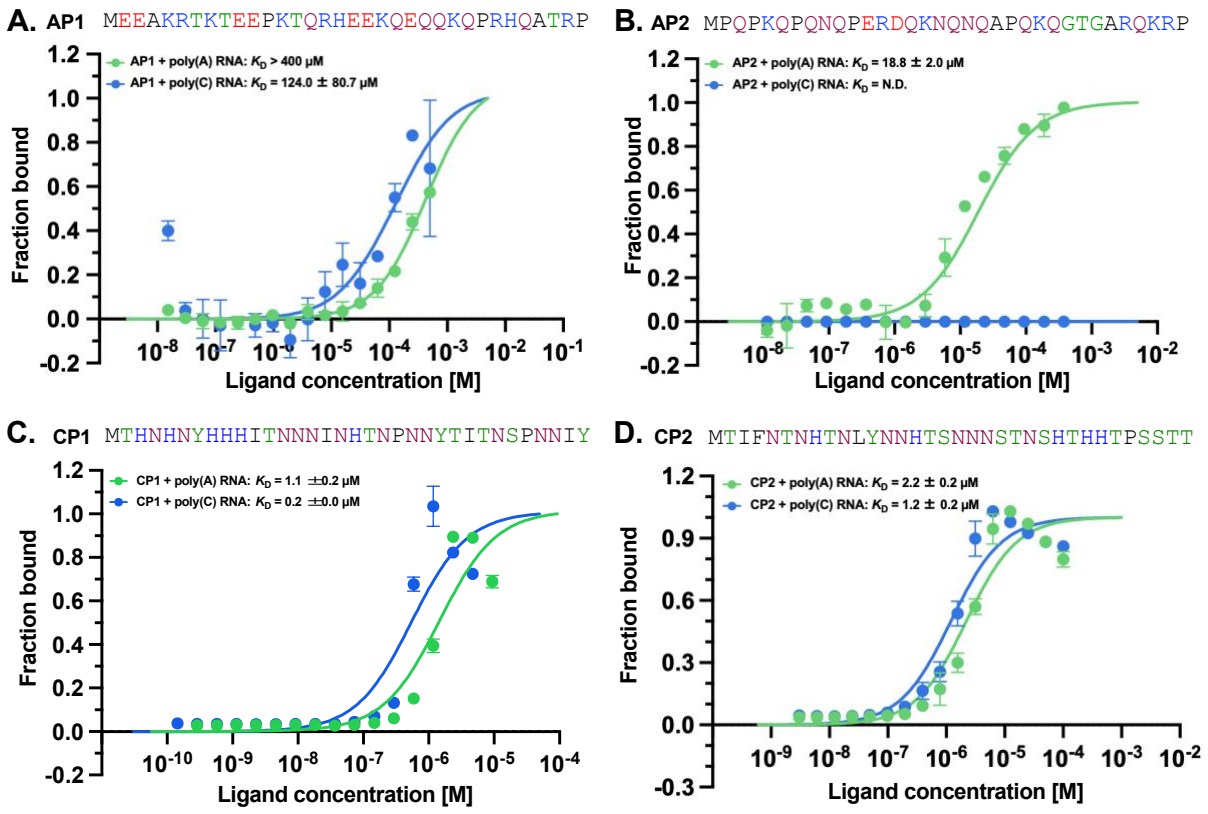


Figure 2-9. Thermodynamics plots of the identified RNA-binding peptides against polyRNAs. Thermodynamics plots were generated to determine K_D values between the identified peptides and polyRNAs. The K_D values were determined by spectral shift (SpS) experiments for (A) AP1, (B) AP2, (C) CP1, and (D) CP2. The bars represent the standard error, and the dots represent the mean from three experimental replicates.

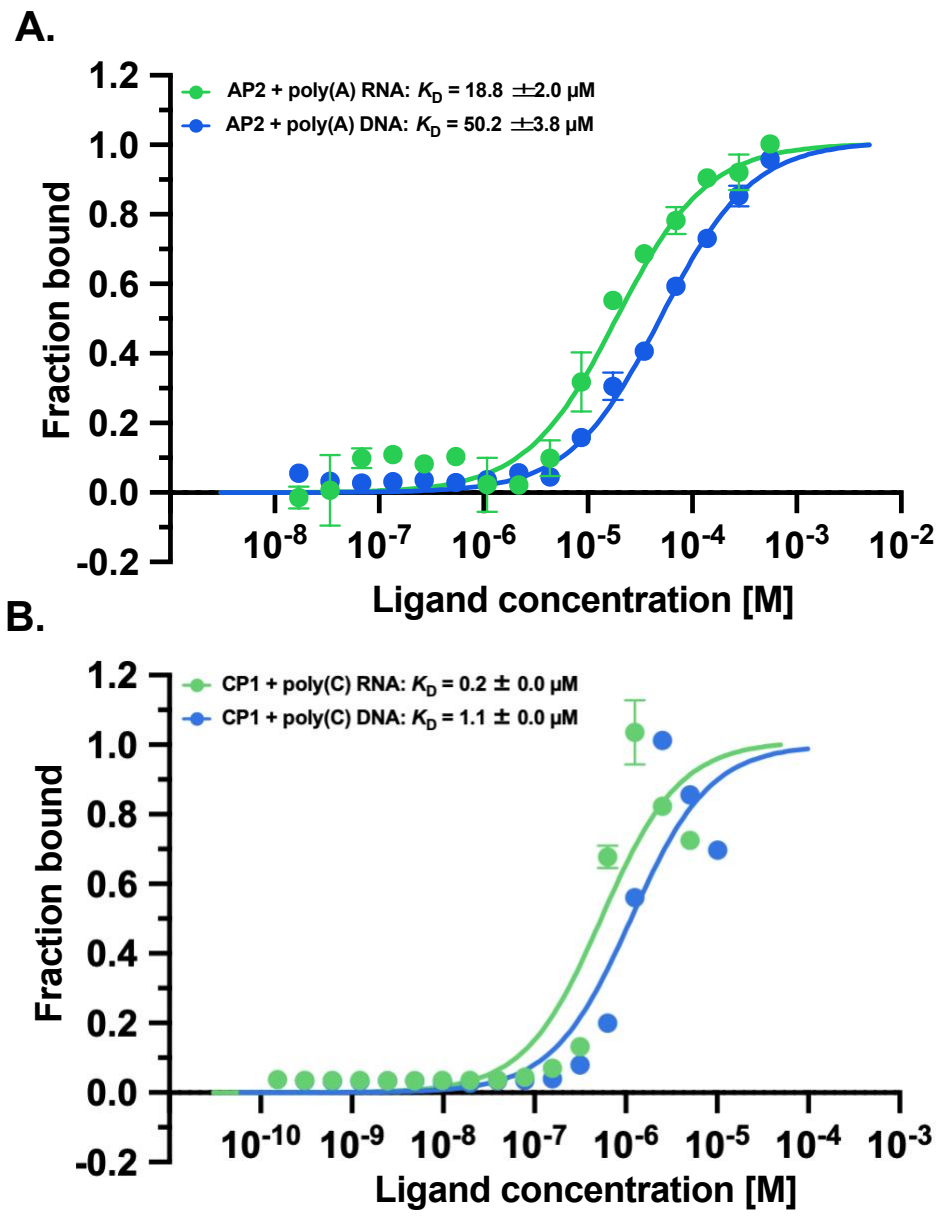


Figure 2-10. Thermodynamics plots of the identified RNA-binding peptides with polyRNA/DNA. The K_D values between polyRNA/DNA and AP2/CP1 were determined by spectral shift (SpS) experiments; (A) thermodynamics plots of AP2 against poly(A) RNA/DNA, (B) thermodynamics plots of CP1 with poly(C) RNA/DNA. The bars represent the standard error, and the dots represent the mean from three experimental replicates.

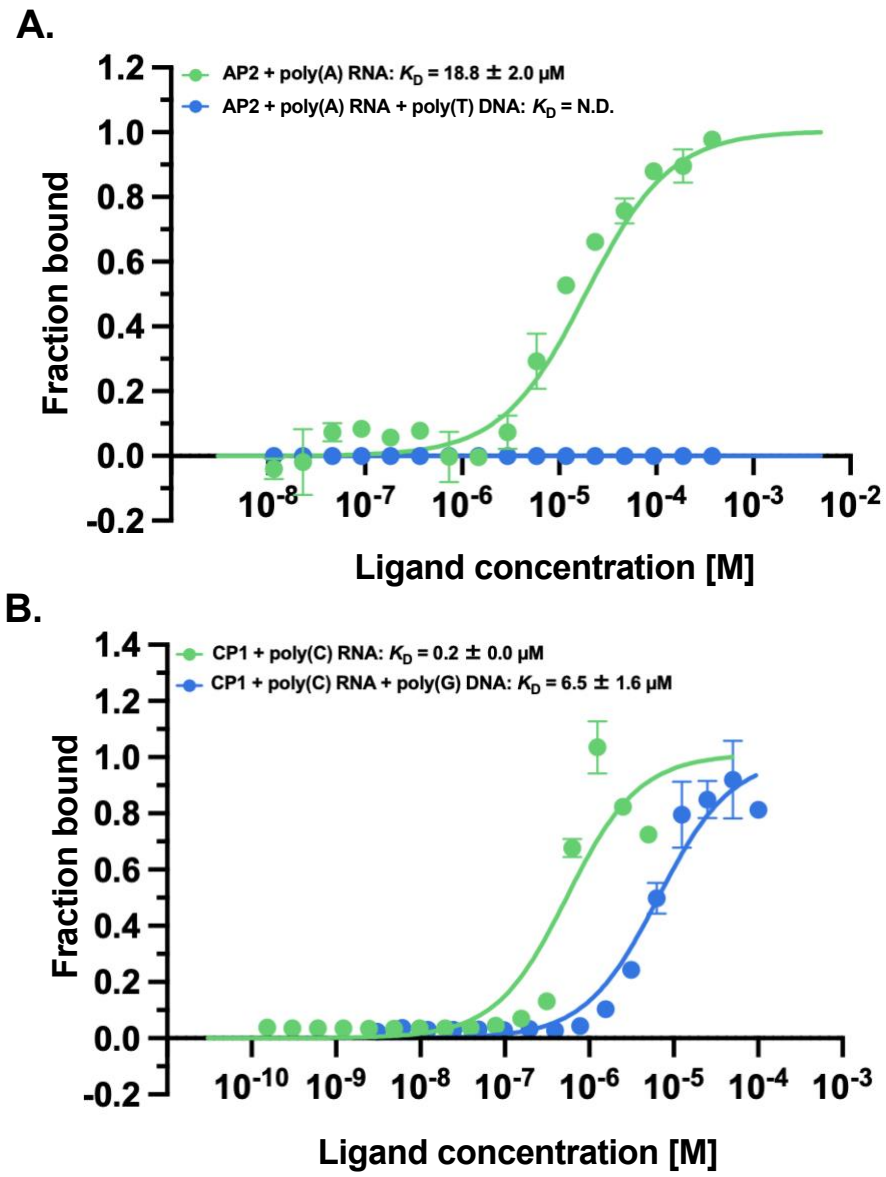


Figure 2-11. Thermodynamics plots of the identified RNA-binding peptides with hybridized RNAs. The K_D values between polyRNA and AP2/CP1 were determined in the presence of the antisense oligo DNA by spectral shift (SpS) experiments: thermodynamics plots of (A) AP2 with cyanine5 (cy5)-labeld poly(A) RNA in the presence of poly(T) DNA and (B) CP1 with cy5-labeled poly(C) RNA in the presence of poly(G) DNA. The bars represent the standard error, and the dots represent the mean from three experimental replicates.

2.3.6 The role of enriched dipeptide motifs found in the RNA-binding peptide sequences.

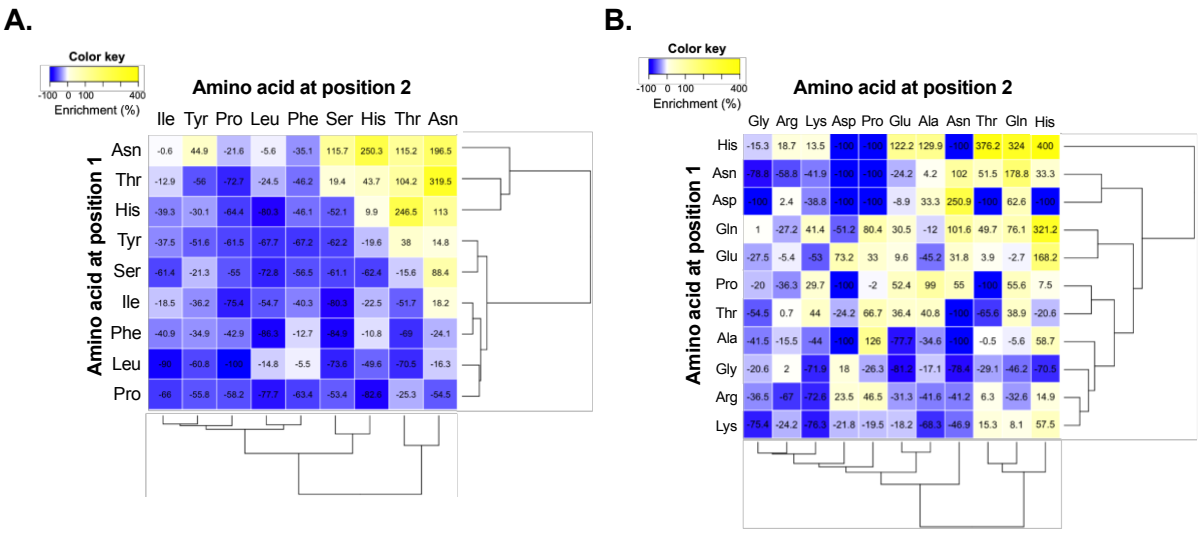
Next, I calculated the frequency of combinatorial amino acid pairs (*i.e.*, dipeptide motifs) enriched in the two libraries, expecting over-represented dipeptide motifs to play an essential role in RNA-binding. I found that the top 100 selected poly(C) RNA-binding peptides were strictly enriched in the dipeptide motifs comprising the (His/Asn/Thr) amino acid combinations (**Figure 2-12A)**. Whereas the top 100 selected poly(A) RNA-binding peptides still comprise a broader spectrum of the enriched dipeptide motifs that includes one of the six (His/Asn/Gln/Glu/Pro/Thr) amino acids (**Figure 2-12B**).

To experimentally validate the function of enriched dipeptide motif residues, I substituted QNQ with TTT in the AP2 peptide due to the significant enrichment of QN and NQ motifs, which both contain glutamine that is highly enriched in the VMM/VRR codon library. Nevertheless, the K_D and specificity toward adenine were unaffected by the replacement of QNQ with TTT (**Figure 2-13A**). Additionally, we created a truncated peptide called "AP2-short," consisting of 6–26 AP2 residues, to identify the specific residues responsible for adenine recognition. As a result, the affinity of AP2-short was approximately five times lower (K_D: 82.0 μM) than that of full-length AP2. However, the truncated peptide still recognized poly(A) RNA specifically (**Figure 2-13B**). Finally, we introduced a single point mutation, replacing glutamine (Q) with serine (S), in the middle of the sequence within AP2-short. This mutation resulted in a decreased dissociation constant (K_D: 114.7 μM), demonstrating that a single residue substitution can increase the overall affinity towards poly(A) RNA. The truncated region (1–5 and 27–33 positions) contains two of the most enriched amino acids in the top 30 clusters of the VMM/VRR library: glutamine (Q) and proline (P), as well as the two basic amino

acids, lysine (K) and arginine (R), which are well-known for their interaction with RNA and DNA. Additionally, five enriched dipeptide motifs (PQ/QP/PK/QK/RP) are found in the truncated region and could potentially serve as RNA-binding interfaces. However, additional mutagenesis is necessary to confirm the involvement of specific residues in the interaction between RNA and peptide.

As for the CP1 peptide, we designed two variants by replacing Asn with Ser within the two most enriched dipeptide motifs (NH and TN) to determine the residues responsible for the RNA-binding capability. Serine mutation was used instead of commonly used alanine to avoid decreasing the peptide solubility. Both mutants, NH to SH and TN to TS mutants lost the binding capability against polyRNAs to below detection limits (**Figures 2-14**), indicating that asparagine alongside histidine and threonine residues are involved in the recognition of polyRNAs. With an attempt to reveal the effect of the mutations on the peptide structure, CD spectroscopy was performed with the native CP1 and its mutant peptides (**Figure 2-15**). The CD spectrum of CP1 showed distinct maximum at 195 nm and minimum at 218 nm, indicating the formation of a beta-sheet structure. On the other hand, the TN to TS mutant peptide displayed a distinct random coil structure, with a single minimum at 198 nm, indicating TN to TS mutagenesis induced a structural transition from a beta-sheet into a random coil structure. Likewise, NH to SH mutagenesis resulted in a shift of the minimum from 218 nm in the native peptide to 208 nm, indicating a partial loss in beta-sheet structure to form a random coil structure. Altogether, the mutagenesis resulted in the structural transition from beta-sheet towards the random coil signal, emphasizing the importance of beta-sheet structure for the RNA-binding capability of CP1.

Finally, to observe the effect of TN and NH motifs, AP2 peptide sequence was modified by replacing the two QNQ motifs with the TNH motif (combination of TN and NH). This modification aimed to enhance AP2's poly(C) RNA-binding capability while retaining its inherent ability to bind poly(A) RNA. By replacing QNQ with TNH, AP2 recognized poly(C) RNA as expected (K_D: 120.3 μM) and displayed notably higher affinity for poly(A) RNA (**Figure 2-13C**). This observation is consistent with CP1's proficiency at binding poly(A) RNA. Thus, the TNH sequence acts as a versatile ssRNA-binding motif that prefers cytosine.



C.

Peptide name	Sequence	$K_D^{poly(A)\ RNA}$ (nM)	$K_D^{poly(C)\ RNA}$ (nM)
Native sequence (CP1)	MTHNHNYHHHITNNNINHTNPNNYTITNSPNNIY	1060 ± 220	218 ± 5.79
Mutant (CP1 NH→SH)	MTHSHNYHHHITNNNISHTNPNNYTITNSPNNIY	N.D.	N.D.
Mutant (CP1 TN→TS)	MTHNHNYHHHITSNNINHTSPNNYTITSSPNNIY	N.D.	N.D.

Figure 2-12. Identification of the enriched combinatorial amino acid pairs and dissociation constant analysis of the identified RNA-binding peptide variants. The enrichment scores of combinatorial amino acid pairs were calculated by comparing the frequency between the peptides from round 0 and round 7 of each library. Hierarchical clustering was employed to visualize the enrichment scores through heat mapping for (A) the HHY codon library and (B) the VMM/VRR codon library. (C) Table showing the dissociation constants between the peptide variants and polyRNAs. NH to SH substitution (pink) and TN to TS substitution (blue) are highlighted. The complete thermodynamics plots are available in Figure 2-14.

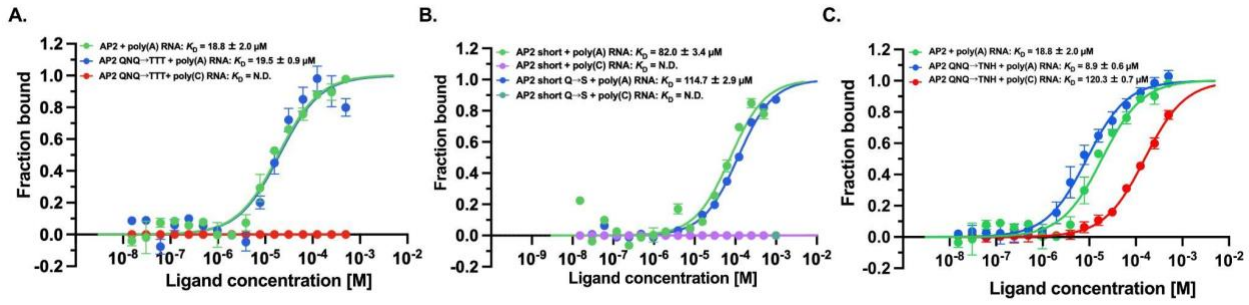


Figure 2-13. Thermodynamics plots of AP2 peptide and its mutated variants. The K_D values between AP2 variants and poly(A)/(C) RNA were determined by spectral shift (SpS) experiments. Native peptide (AP2) was compared to the mutated variants whose QNQ sequence was mutated to (A) TTT or (C) TNH, while (B) the truncated AP2 (position 6 to 26) was compared to the mutated variants whose Q was mutated to S. The bars represent the standard error, and the dots represent the mean from three experimental replicates.

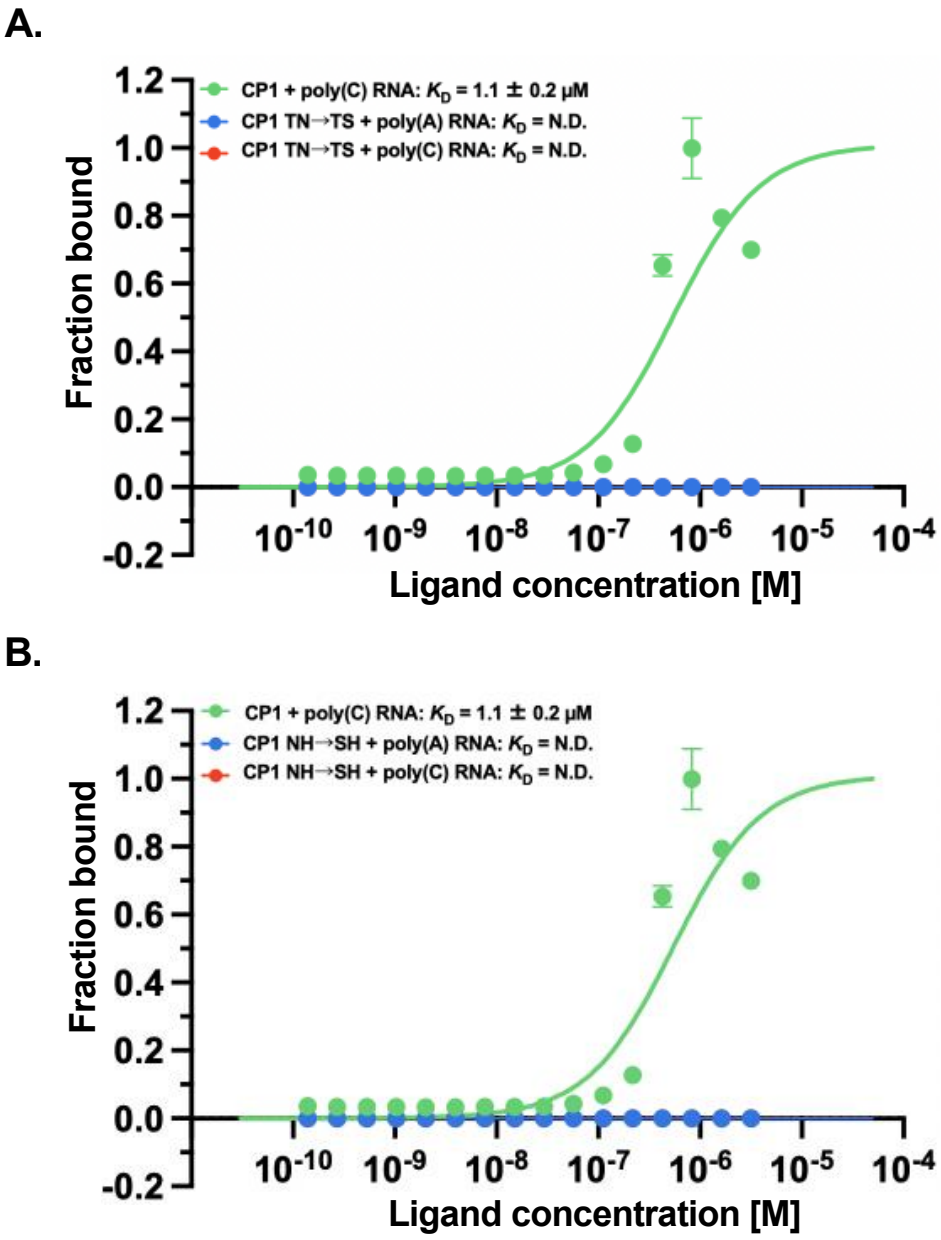


Figure 2-14. Thermodynamics plots of CP1 peptide and its mutated variants with poly(C) RNA. The K_D values between CP1 variants and poly(C) RNAs were determined by spectral shift (SpS) experiments. Native peptide (CP1) was compared to the mutated variants whose (A) TN sequence was mutated to TS or (B) NH sequence was mutated to SH. The bars represent the standard error, and the dots represent the mean from three experimental replicates.

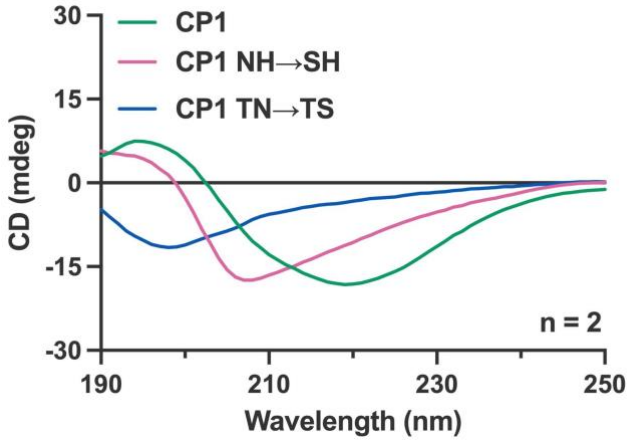


Figure 2-15. Circular dichroism spectra of CP1 peptide and its variants. Circular dichroism (CD) spectra were measured for each peptide with a concentration of 0.2 mg/mL. The CD spectra were acquired in duplicate and averaged: CP1 (green), CP1 NH→SH (pink), and CP1 TN→TS (blue).

2.3.7 Synthetic poly(C) RNA-binding peptide interacts with cytosine-rich oncogenic RNA motifs in a human plasma-like medium.

To evaluate the synthetic ssRNA-binding peptide's affinity for biologically derived RNA targets, I have chosen two cytosine-rich RNA motifs from the oncogenic Cdk6 3′UTR RNA and *MYU* lncRNA. These motifs have previously been identified as targets of hnRNPK, and they both consist of extended cytosine repeats (>4 nt) within a necessary single-stranded region for the progression of colon cancer cells (Nakamoto et al. 2020). I have found that the CP1 peptide can bind to both RNA motifs with submicromolar K_Ds Specifically, it can bind to both Cdk6 3′UTR RNA (103.2 nM) and *MYU* lncRNA (107.2 nM) in a medium that mimics human plasma (**Figure 2-16;** The predicted secondary structures of RNA ligands are available in **Figure 2-17** and the complete thermodynamics are available in **Figure 2-18**). Notably, CP1 showed a comparable K_D to that of hnRNPK with Cdk6 3′UTR RNA (K_D: 450 nM) (Nakamoto et al. 2020). On the contrary, CP1 did not demonstrate any affinity towards the established negative control for hnRNPK cytosine-poor Gas5 hp RNA, suggesting a preference for a single-stranded cytosine repeat. As CP1 exhibited a stronger inclination towards Cdk6 3′UTR RNA compared to hnRNPK, we proceeded to mutate the cytosine-rich, single-stranded regions of Cdk6 3′UTR RNA to confirm the contribution of these polyC regions to CP1 binding. The replacement of two extended cytosine repeats positioned at 4–8 and 20–22 with adenine repeats (C-repeat substituted mutant of Cdk6 3′UTR RNA) resulted in losing the binding ability (**Figure 2-16**), demonstrating that CP1 interacts with the poly(C) regions of Cdk6 3′UTR RNA.

RNA ligand	Sequence	Length (nt)	K_D^{CP1} (nM)	$^*K_D^{hnRNPK}$ (nM)
Cdk6 3′UTR RNA	GGUCCCCCGCCUCAUUCGCCCCUCUGCUCCC	31	103.2 ± 91.8	450 ± 60
MYU lncRNA	GGCUCCCCCGACCUCUGUGCUCCCCUCCCCCGACCUCUGUGC	42	107.2 ± 84.4	70 ± 7.0
Gas5 hp RNA	GGAGCCUCCCAGUGGUCUUUGUAGACUGCCUGAUGGAGUCUCC	43	N.D.	N.D.
C-repeat substituted mutant of Cdk6 3′UTR RNA	GGUAAAAAGCCUCAUUCGCAAAUCUGCUCCC	31	N.D.	—

Figure 2-16. Assessment of the affinity of poly(C) RNA-binding peptides to cytosine-rich oncogenic RNA motifs. The table displays the dissociation constants (K_D^{CP1}) between hnRNPK-associated RNAs and CP1 peptides. The K_D^{CP1}s were measured using microscale thermophoresis in a human plasma-like medium. Cytosines predicted to be single-stranded within RNA ligands are highlighted in red. Cytosine to adenine substitution is highlighted in blue. The asterisk denotes the K_D values between hnRNPK- associated RNAs and hnRNPK (K_D^{hnRNPK}) determined by fluorescence anisotropy binding assay reported in the previous study (Nakamoto et al. 2020).

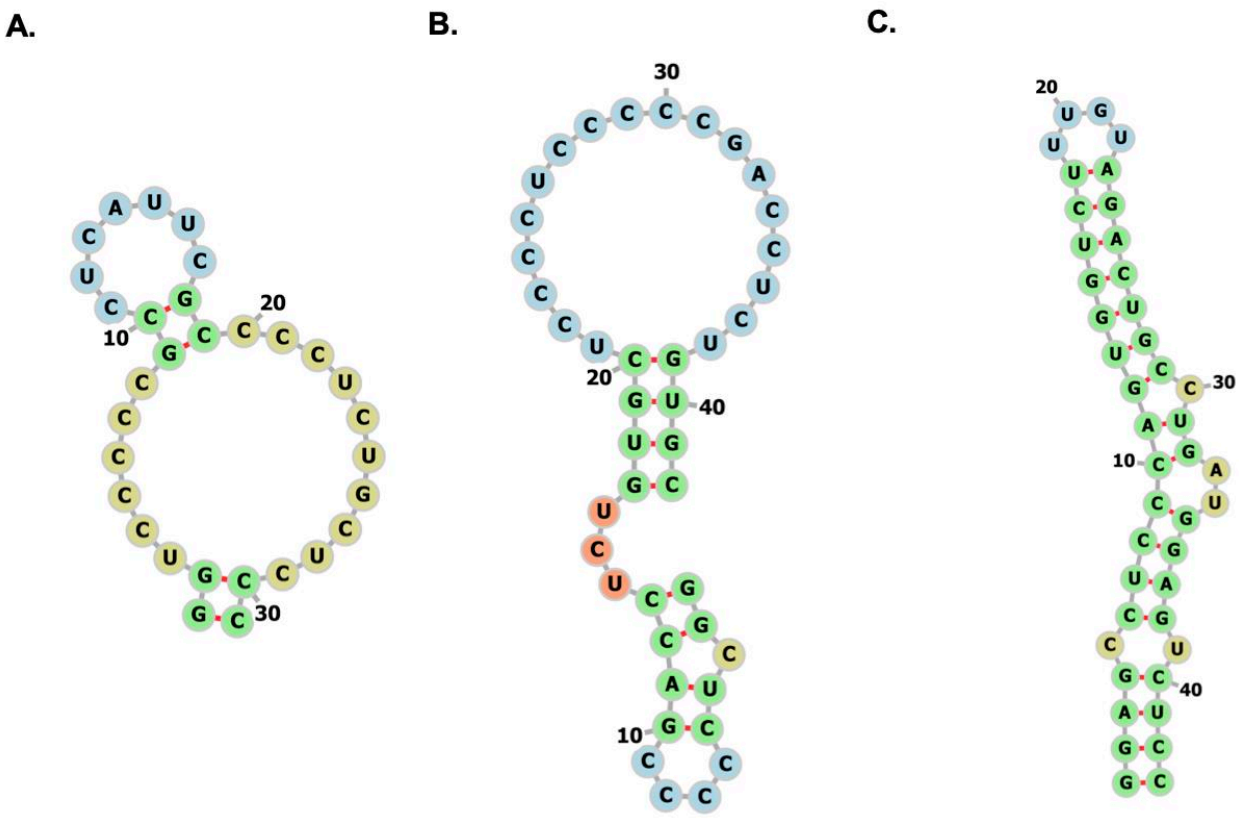


Figure 2-17. Lowest free energy structures for hnRNPK-associated RNAs. MXfold server (Sato, Akiyama, and Sakakibara 2021) was used to predict the lowest free energy structures for hnRNPK-associated RNAs used in this study: (A) Cdk6 3′UTR RNA, (B) MYU lncRNA, and (C) Gas5 hp RNA.

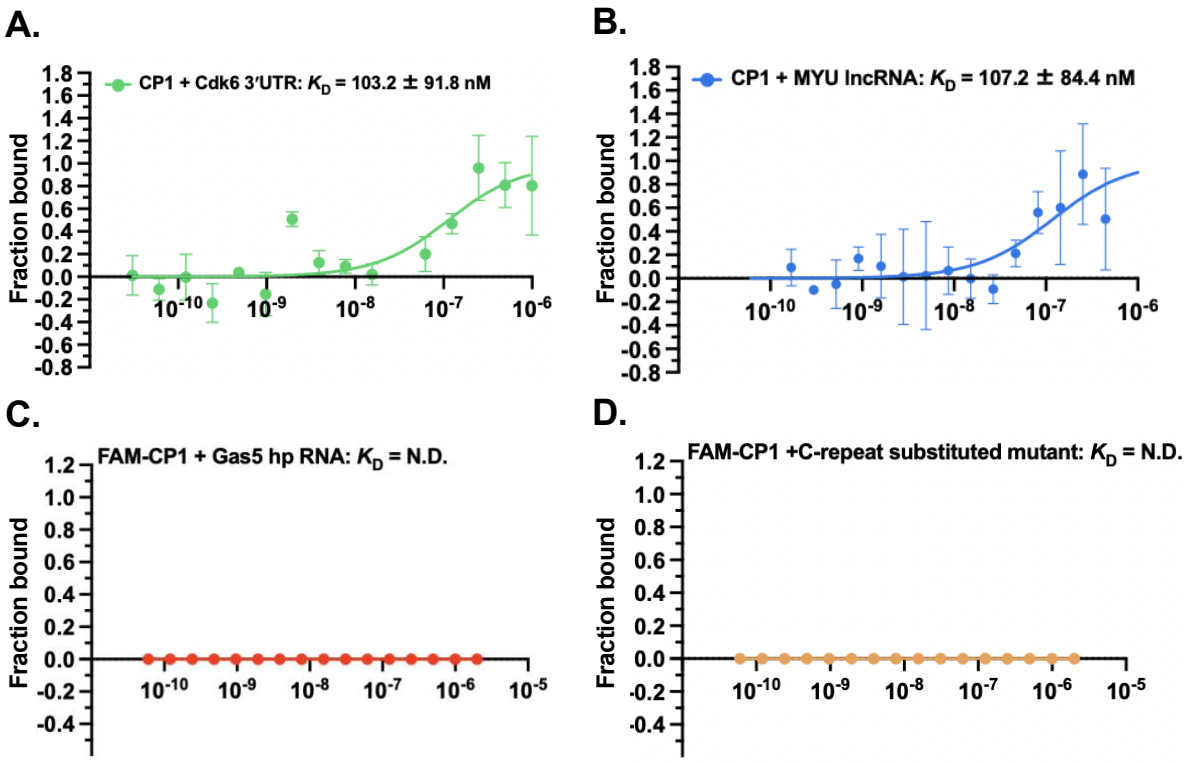


Figure 2-18. Thermodynamic plots of CP1 peptide with hnRNPK-associated RNAs. Microscale thermophoresis was employed to determine the K_D values of CP1 with a set of hnRNPK-associated oncogenic RNA motifs including hnRNPK-nonassociated RNA (Gas5 hp RNA) and C-repeat substituted mutant of Cdk6 3′UTR RNA. (A) Cdk6 3′UTR RNA, (B) MYU lncRNA, (C) Gas5 hp RNA, (D) C-repeat substituted mutant of Cdk6 3′UTR RNA. The bars represent the standard error, and the dots represent the mean from three experimental replicates.

2.4 Discussion

mRNA and cDNA display have been commonly used as efficient tools to explore large peptide sequence space against versatile binding targets (library size can reach up to approximately 10^{13} molecules). Previous *in vitro* selection studies have shown that various RNA and DNA-binding peptides can be selected (Kumachi, Husimi, and Nemoto 2016; Giacobelli et al. 2022; R. Liu et al. 2000; Barrick et al. 2001). However, screening for *de novo* peptides that can interact with low-complexity ssRNA remains a major challenge using the mRNA display due to unwanted base pairing between mRNA and target RNA. Therefore, I decided to take a renovated approach to combine codon-restricted DNA library with mRNA display to overcome the problem mentioned above.

Our *in vitro* selection successfully led to the identification of four novel *de novo* RNA-binding peptides (AP1/AP2/CP1/CP2) interacting with poly(A) or poly(C) RNA. Deep sequencing revealed significant glutamine (Q) and asparagine (N) composition enhancement in the top 30 enriched cluster sequences of each library. It has been known that Watson-Crick and Hoogsteen-type interaction can occur between the Q/N carboxamide group and purine/pyrimidine nucleobases (**Figure 2-19**) (Cheng et al. 2003; Kondo and Westhof 2011). It was, therefore, not surprising that all four newly identified RNA-binding peptides are enriched in Q/N residues (AP1: 18%, AP2: 39%, CP1: 36%, CP2: 27%). The most significant case was the CP1 peptide, relying solely on the asparagine residues within the top 2 most enriched TN and NH dipeptide motifs (enrichment > 250%, **Figure 2-12B**). The fact that CP1 binds to poly(C) RNA with a K_D of 200 nM, which is approximately five-fold stronger compared to binding poly(A) RNA and poly(C) DNA, strongly suggests that the asparagine carboxamide

group is in contact with both bases (adenine and cytosine) as well as 2' OH of the ribose. This is consistent with the evidence that the carboxamide group can form a hydrogen bond with 2' OH either directly or indirectly bridged by a water molecule (Kondo and Westhof 2011; Corley, Burns, and Yeo 2020). The decrease in K_D was not as prominent in the AP2-poly(A) DNA interaction, suggesting that 2' OH is less important for AP2 recognition. Indeed, AP2 affinity towards poly(A) RNA (K_D: 18.8 μM) was almost two orders of magnitude lower than that of CP1, and specificity towards adenine moiety may suggest the recognition *via* Hoogsteen pseudo pair. At least in the case of DNA duplex, Hoogsteen base pairing is approximately 3 kcal/mol higher in energy than the Watson–Crick base pairing (Nikolova et al. 2011), which may explain the relatively low affinity of AP2 and AP1. Because the configuration of cytosine–asparagine interaction in the HHY library is limited to the Watson–Crick pseudo pair (**Figure 2-19**), I speculate that limited structural configuration along with the lack of positively charged amino acids likely resulted in the steep and non-continuous fitness landscape leading to the dominance of CP1 cluster sequences occupying 77.6% of the entire population after seven rounds of selection.

As opposed to the HHY library, the variety of amino acids that are able to interact with polynucleotides is clearly diverse in the VMM/VRR library. I observed that the two QNQ sequences consisting of the highly enriched QN/NQ dipeptide motifs (enrichment > 100%, **Figure 2-13A**) can be replaced by TTT sequences, suggesting either (*i)* hydroxyl group in threonine can complement the hydrogen bonding of carboxamide group, *(ii)* peptide backbone is involved in the interaction, or *(iii)* these residues are not in direct contact with the poly(A) RNA. While truncation and point mutation led to a stepwise decrease of AP2 affinity but did

not entirely lose the binding, I consider the involvement of multiple amino acid residues that contain other enriched dipeptide motifs. Also, since I did not incorporate metal cations during the selection, previously reported cation-mediated interaction between acidic amino acids and RNA (Giacobelli et al. 2022) is unlikely to be involved. Therefore, I assume that net negative acidic peptides (net charge < 0) were unfavorable under our RNA-binding condition, resulting in a constrained chemical parameter space of the enriched peptides after seven rounds of selection. A clear trend toward decreasing hydrophobicity in both libraries may indicate that the peptide maintains the overall linear and disordered structure to maximize the interface with the unstructured linear RNA. The fact that RNA-binding capacity was affected by single amino acid (Q or N) substitution, as well as N- and C-terminal end truncation, further supports this hypothesis.

In nature, structured cellular PCBPs possess multiple KH domains that bind to cytosine triplet at the hydrophobic cleft formed by a small GXXG loop and variable loop (Du et al. 2005). The positively charged amino acid residues surrounding the cleft attract the oligonucleotide and specific recognition of cytosine is realized by a dense network of hydrogen bonding with conserved amino acid side chains (Asp, Arg, Glu, Gly, Ile, and Lys) (Yoga et al. 2012). Such coordination requires a stable tertiary structure that short peptides cannot provide. However, our novel poly(C) RNA-binding peptide (CP1), which is enriched with Asn, Thr, and His can interact with RNA that is rich in cytosine with a comparable affinity to that of hnRNPK with KH domain. This suggests that the RNA-binding mechanism differs significantly from typical RNA-binding domains. Since the CP1 peptide contains four motifs (NH and TN) that recognize cytosine within its sequence, it is possible that having multiple cytosine-recognizing

motifs contributes to the preference for cytosine repeats. Additionally, secondary structure analysis performed through CD spectroscopy confirmed a distinct beta-sheet structure in CP1, whereas the two unbound variants peptides led to complete/partial disruption of this beta-sheet structure (**Figure 2-15**). This result indicated that unlike the other poly(C) RBPs in nature, which utilize a loop structure as a binding surface, the CP1 peptide likely supports the interaction with polyRNAs *via* its beta-sheet structure. The fact that polyQ/polyN peptides do not exhibit strong affinity toward polyRNA (**Figure 2-20**) clearly indicates that a tandem Gln/Asn sequence alone is not enough to support affinity towards polyRNA. Although further characterization is necessary to uncover the precise binding mechanism, our findings showcase the effectiveness of our methodology in examining variations in peptide sequences that have not been investigated in natural settings. These *de novo* ssRNA-binding peptides exhibit significant promise as versatile RNA-binding tags with a wide array of potential applications. As demonstrated in the previous study, RBDs can precisely modulate RNA functions when fused to proteins of interest (Ozawa et al. 2007; Tilsner et al. 2009; Choudhury et al. 2012; Cooke et al. 2011; Puvvula et al. 2021; Puvvula, Buczkowski, and Moon 2021). Our methodology enables high-throughput discovery of *de novo* ssRNA-binding peptides with sequence specificity. Thus, the *de novo* RNA-binding peptides can serve as potent RNA-binding tags, enabling the modulation of translation and stability in targeted mRNA molecules and facilitating the labeling of specific RNA molecules within cells. It also offers the potential to further design RNA-targeted drugs and therapies targeting RNA viruses and neurodegenerative disorders. This advancement will facilitate the development of peptides and proteins targeting various RNA molecules, expanding the scope of our methodology.

In conclusion, our cell-free codon-restricted mRNA display approach identified a catalog of possible *de novo* ssRNA-binding peptides from a synthetic peptide library containing limited amino acid alphabets. For future work, the introduction of genetic code expansion and codon reassignment technology (Iwane et al. 2016) will enable us to expand our method to explore unnatural peptide sequence space with desired amino acid combinations, including non-canonical amino acids.

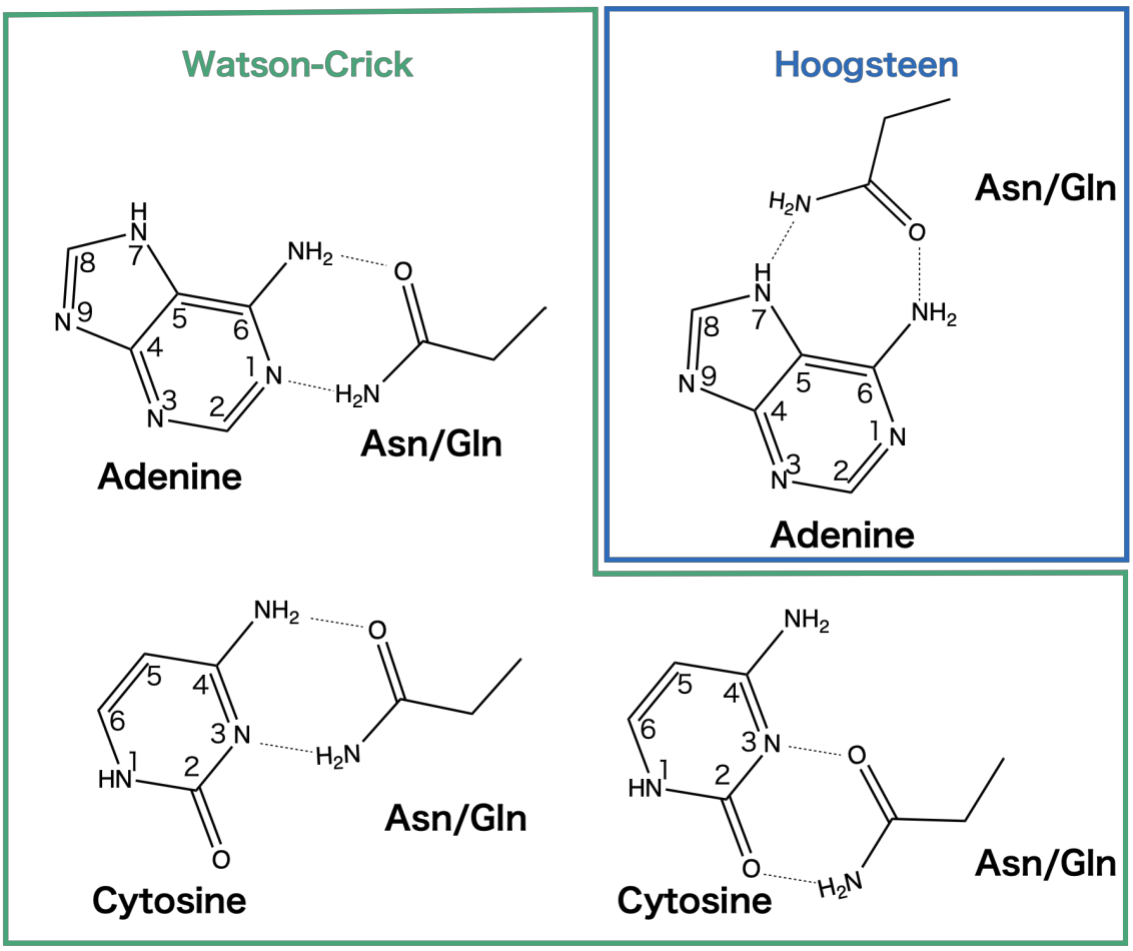


Figure 2-19. Proposed hydrogen bonding between Gln/Asn and adenine/cytosine. Green: Watson-Crick-type hydrogen bonding interaction between Gln/Asn and adenine/cytosine. Blue: Hoogsteen-type hydrogen bonding interaction between Gln/Asn and adenine. The common side chain structure of Gln and Asn is shown here. The dashed line indicates hydrogen bonding.

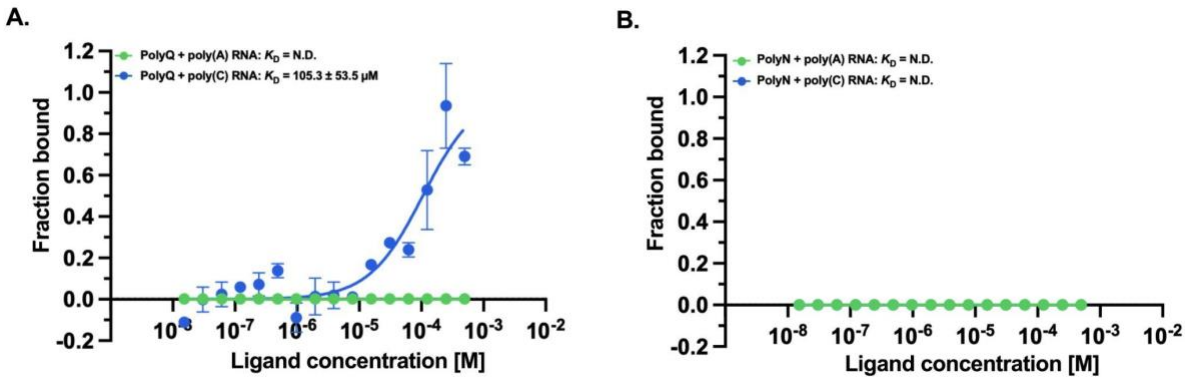


Figure 2-20. **Thermodynamic plots of polyQ/N peptide with polyRNAs.** Microscale thermophoresis was employed to determine the dissociation constants of polyQ/N peptide (6-mer) with poly(A)/(C) RNA. (A) polyQ peptide and (B) polyN peptide. The error bar represents the mean ± SE from three independent experimental replicates.

Reference

1. Alam, Khalid K., Jonathan L. Chang, and Donald H. Burke. 2015. “FASTAptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections.” *Molecular Therapy. Nucleic Acids* 4 (3): e230.

2. Anczuków, Olga, Avi Z. Rosenberg, Martin Akerman, Shipra Das, Lixing Zhan, Rotem Karni, Senthil K. Muthuswamy, and Adrian R. Krainer. 2012. “The Splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation.” *Nature Structural & Molecular Biology* 19 (2): 220–28.

3. Barrick, J. E., T. T. Takahashi, J. Ren, T. Xia, and R. W. Roberts. 2001. “Large libraries reveal diverse solutions to an RNA recognition problem.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (22): 12374–78.

4. Bashiruddin, Nasir K., and Hiroaki Suga. 2015. “Construction and screening of vast libraries of natural product-like macrocyclic peptides using *in vitro* display technologies.” *Current Opinion in Chemical Biology* 24 (February): 131–38.

5. Blobel, G. 1973. “A protein of molecular weight 78,000 bound to the polyadenylate region of eukaryotic messenger RNAs.” *Proceedings of the National Academy of Sciences of the United States of America* 70 (3): 924–28.

6. Blyn, L. B., K. M. Swiderek, O. Richards, D. C. Stahl, B. L. Semler, and E. Ehrenfeld. 1996. “Poly(rC) binding protein 2 binds to stemlLoop IV of the poliovirus RNA 5′ noncoding region: Identification by automated liquid chromatography-tandem mass spectrometry.” *Proceedings of the National Academy of Sciences of the United States of America* 93 (20): 11115–20.

7. Butterworth, P. J. 2005. “Lehninger: Principles of Biochemistry (4th Edn) D. L. Nelson and M. C. Cox, W. H. Freeman & Co., New York, 1119 Pp (plus 17 Pp Glossary), ISBN 0-7167-4339-6 (2004).” *Cell Biochemistry and Function* 23 (4): 293–94.

8. Cheng, Alan C., William W. Chen, Cynthia N. Fuhrmann, and Alan D. Frankel. 2003. “Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains.” *Journal of Molecular Biology* 327 (4): 781–96.

9. Cheong, Cheom-Gil, and Traci M. Tanaka Hall. 2006. “Engineering RNA sequence specificity of pumilio repeats.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (37): 13635–39.

10. Choudhury, Rajarshi, Yihsuan S. Tsai, Daniel Dominguez, Yang Wang, and Zefeng Wang. 2012. “Engineering RNA endonucleases with customized sequence specificities.” *Nature Communications* 3: 1147.

11. Cooke, Amy, Andrew Prigge, Laura Opperman, and Marvin Wickens. 2011. “Targeted translational regulation using the PUF protein family scaffold.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (38): 15870–75.

12. Corley, Meredith, Margaret C. Burns, and Gene W. Yeo. 2020. “How RNA-binding proteins interact with RNA: Molecules and mechanisms.” *Molecular Cell* 78 (1): 9–29.

13. Crooks, Gavin E., Gary Hon, John-Marc Chandonia, and Steven E. Brenner. 2004. “WebLogo: A sequence logo generator.” *Genome Research* 14 (6): 1188–90.

14. Du, Zhihua, John K. Lee, Richard Tjhen, Shang Li, Hu Pan, Robert M. Stroud, and Thomas L. James. 2005. “Crystal structure of the first KH domain of human poly(C)-binding protein-2 in complex with a C-rich strand of human telomeric DNA at 1.7 A.” *Journal of Biological Chemistry* 280 (46): 38823–30.

15. Filipovska, Aleksandra, Muhammad F. M. Razif, Karoline K. A. Nygård, and Oliver Rackham. 2011. “A universal code for RNA recognition by PUF proteins.” *Nature Chemical Biology* 7 (7): 425–27.

16. Fukuda, Isao, Kanehisa Kojoh, Noriko Tabata, Nobuhide Doi, Hideaki Takashima, Etsuko Miyamoto-Sato, and Hiroshi Yanagawa. 2006. “*In vitro* evolution of single-chain antibodies using mRNA display.” *Nucleic Acids Research* 34 (19): e127.

17. Gallie, D. R. 1991. “The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency.” *Genes & Development* 5 (11): 2108–16.

18. Ghanem, Louis R., Andrew Kromer, Ian M. Silverman, Priya Chatterji, Elizabeth Traxler, Alfredo Penzo-Mendez, Mitchell J. Weiss, Ben Z. Stanger, and Stephen A. Liebhaber. 2016. “The poly(C) binding protein Pcbp2 and its retrotransposed derivative Pcbp1 are independently essential to mouse development.” *Molecular and Cellular Biology* 36 (2): 304–19.

19. Giacobelli, Valerio G., Kosuke Fujishima, Martin Lepšík, Vyacheslav Tretyachenko, Tereza Kadavá, Mikhail Makarov, Lucie Bednárová, Petr Novák, and Klára Hlouchová. 2022. “*In vitro* evolution reveals noncationic protein-RNA interaction mediated by metal ions.” *Molecular Biology and Evolution* 39 (3).

20. Hanes, J., and A. Plückthun. 1997. “*In vitro* selection and evolution of functional proteins by using ribosome display.” *Proceedings of the National Academy of Sciences of the United States of America* 94 (10): 4937–42.

21. Hentze, Matthias W., Alfredo Castello, Thomas Schwarzl, and Thomas Preiss. 2018. “A brave new world of RNA-binding proteins.” *Nature Reviews. Molecular Cell Biology* 19 (5): 327–41.

22. Iwane, Yoshihiko, Azusa Hitomi, Hiroshi Murakami, Takayuki Katoh, Yuki Goto, and Hiroaki Suga. 2016. “Expanding the amino acid repertoire of ribosomal polypeptide synthesis *via* the artificial division of codon boxes.” *Nature Chemistry* 8 (4): 317–25.

23. Kawasaki, Yoshihiro, Mimon Komiya, Kosuke Matsumura, Lumi Negishi, Sakiko Suda, Masumi Okuno, Naoko Yokota, et al. 2016. “*MYU*, a target lncRNA for Wnt/c-Myc signaling, mediates induction of CDK6 to promote cell cycle progression.” *Cell Reports* 16 (10): 2554–64.

24. Kechavarzi, Bobak, and Sarath Chandra Janga. 2014. “Dissecting the expression landscape of RNA-binding proteins in human cancers.” *Genome Biology* 15 (1): R14.

25. Keefe, A. D., and J. W. Szostak. 2001. “Functional proteins from a random-sequence library.” *Nature* 410 (6829): 715–18.

26. Kim, Hong Joo, Nam Chul Kim, Yong-Dong Wang, Emily A. Scarborough, Jennifer Moore, Zamia Diaz, Kyle S. MacLea, et al. 2013. “Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS.” *Nature* 495 (7442): 467–73.

27. Kondo, Jiro, and Eric Westhof. 2011. “Classification of pseudo pairs between nucleotide bases and amino acids by analysis of nucleotide-protein complexes.” *Nucleic Acids Research* 39 (19): 8628–37.

28. Kumachi, Shigefumi, Yuzuru Husimi, and Naoto Nemoto. 2016. “An RNA binding peptide consisting of four types of amino acid by *in vitro* selection using cDNA display.” *ACS Omega* 1 (1): 52–57.

29. Kyte, J., and R. F. Doolittle. 1982. “A simple method for displaying the hydropathic character of a protein.” *Journal of Molecular Biology* 157 (1): 105–32.

30. Langer, Andreas, Tanja Bartoschik, Ondrej Cehlar, Stefan Duhr, Philipp Baaske, and Werner Streicher. 2022. “A new spectral shift-based method to characterize molecular interactions.” *Assay and Drug Development Technologies* 20 (2): 83–94.

31. Li, Peiying, Philipp Holliger, and Shunsuke Tagami. 2022. “Hydrophobic-cationic peptides modulate RNA polymerase ribozyme activity by accretion.” *Nature Communications* 13 (1): 3050.

32. Li, Ziwei, Xiaole Wang, and Rong Jia. 2022. “Poly(rC) binding protein 1 represses the translation of STAT3 through 5′ UTR.” *Current Gene Therapy* 22 (5): 397–405.

33. Liu, R., J. E. Barrick, J. W. Szostak, and R. W. Roberts. 2000. “Optimized synthesis of RNA-protein fusions for *in vitro* protein selection.” *Methods in Enzymology* 318: 268–93.

34. Lunde, Bradley M., Claire Moore, and Gabriele Varani. 2007. “RNA-binding proteins: Modular design for efficient function.” *Nature Reviews. Molecular Cell Biology* 8 (6): 479–90.

35. Luo, Zhen, Xingchen Dong, Youxing Li, Qi Zhang, Cholho Kim, Yu Song, Lei Kang, Yingle Liu, Kailang Wu, and Jianguo Wu. 2014. “PolyC-binding protein 1 interacts with 5′-untranslated region of enterovirus 71 RNA in membrane-associated complex to facilitate viral replication.” *PloS One* 9 (1): e87491.

36. Mattheakis, L. C., R. R. Bhatt, and W. J. Dower. 1994. “An *in vitro* polysome display system for identifying ligands from very large peptide libraries.” *Proceedings of the National Academy of Sciences of the United States of America* 91 (19): 9022–26.

37. Minami, Shintaro, Naohiro Kobayashi, Toshihiko Sugiki, Toshio Nagashima, Toshimichi Fujiwara, Rie Tatsumi-Koga, George Chikenji, and Nobuyasu Koga. 2023. “Exploration of novel Aβ-protein folds through *de novo* design.” *Nature Structural & Molecular Biology* 30 (8): 1132–40.

38. Miyamoto-Sato, Etsuko, Masamichi Ishizaka, Kenichi Horisawa, Seiji Tateyama, Hideaki Takashima, Shinichiro Fuse, Kaori Sue, Naoya Hirai, Kazuyo Masuoka, and Hiroshi Yanagawa.

2005. “Cell-free cotranslation and selection using *in vitro* virus for high-throughput analysis of protein–protein interactions and complexes.” *Genome Research* 15 (5): 710–17.

39. Mochizuki, Yuki, Manish Biyani, Sachika Tsuji-Ueno, Miho Suzuki, Koichi Nishigaki, Yuzuru Husimi, and Naoto Nemoto. 2011. “One-pot preparation of mRNA/cDNA display by a novel and versatile puromycin-linker DNA.” *ACS Combinatorial Science* 13 (5): 478–85.

40. Nagumo, Yu, Kei Fujiwara, Kenichi Horisawa, Hiroshi Yanagawa, and Nobuhide Doi. 2016. “PURE mRNA display for *in vitro* selection of single-chain antibodies.” *Journal of Biochemistry* 159 (5): 519–26.

41. Nakamoto, Meagan Y., Nickolaus C. Lammer, Robert T. Batey, and Deborah S. Wuttke. 2020. “hnRNPK recognition of the B motif of Xist and other biological RNAs.” *Nucleic Acids Research* 48 (16): 9320–35.

42. Nemoto, N., E. Miyamoto-Sato, Y. Husimi, and H. Yanagawa. 1997. “*In vitro* virus: Bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome *in vitro*.” *FEBS Letters* 414 (2): 405–8.

43. Nikolova, Evgenia N., Eunae Kim, Abigail A. Wise, Patrick J. O’Brien, Ioan Andricioaei, and Hashim M. Al-Hashimi. 2011. “Transient Hoogsteen base pairs in canonical duplex DNA.” *Nature* 470 (7335): 498–502.

44. Nishigaki, K., K. Taguchi, Y. Kinoshita, T. Aita, and Y. Husimi. 1998. “Y-ligation: An efficient method for ligating single-stranded DNAs and RNAs with T4 RNA ligase.” *Molecular Diversity* 4 (3): 187–90.

45. Oikonomou, Panos, Roberto Salatino, and Saeed Tavazoie. 2020. “*In vivo* mRNA display enables large-scale proteomics by next generation sequencing.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (43): 26710–18.

46. Osorio, Daniel, Paola Rondón-Villarreal, and Rodrigo Torres. 2015. “Peptides: A package for data mining of antimicrobial peptides.” *Small* 12: 44–444.

47. Ozawa, Takeaki, Yutaka Natori, Moritoshi Sato, and Yoshio Umezawa. 2007. “Imaging dynamics of endogenous mitochondrial RNA in single living cells.” *Nature Methods* 4 (5): 413–19.

48. Puvvula, Pavan Kumar, Stephanie Buczkowski, and Anne M. Moon. 2021. “hnRNPK-derived cell-penetrating peptide inhibits cancer cell survival.” *Molecular Therapy - Oncolytics* 23 (December): 342–54.

49. Puvvula, Pavan Kumar, Yao Yu, Kaelan Renaldo Sullivan, Henok Eyob, Julian Rosenberg, Alana Welm, Chad Huff, and Anne M. Moon. 2021. “Inhibiting an RBM39/MLL1 epigenomic regulatory complex with dominant-negative peptides disrupts cancer cell transcription and proliferation.” *Cell Reports* 35 (9): 109156.

50. Reyes, Sabrina Galiñanes, Yutetsu Kuruma, Mai Fujimi, Masako Yamazaki, Sumie Eto, Shota Nishikawa, Satoshi Tamaki, et al. 2021. “PURE mRNA display and cDNA display provide rapid

detection of core epitope motif *via* high-throughput sequencing." *Biotechnology and Bioengineering* 118 (4): 1736–49.

51. Roberts, R. W., and J. W. Szostak. 1997. "RNA-peptide fusions for the *in vitro* selection of peptides and proteins." *Proceedings of the National Academy of Sciences of the United States of America* 94 (23): 12297–302.

52. Sato, Kengo, Manato Akiyama, and Yasubumi Sakakibara. 2021. "RNA secondary structure prediction using deep learning with thermodynamic integration." *Nature Communications* 12 (1): 941.

53. Seelig, Burckhard, and Jack W. Szostak. 2007. "Selection and evolution of enzymes from a partially randomized non-catalytic scaffold." *Nature* 448 (7155): 828–31.

54. Shimizu, Y., A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda. 2001. "Cell-free translation reconstituted with purified components." *Nature Biotechnology* 19 (8): 751–55.

55. Tilsner, Jens, Olga Linnik, Nynne M. Christensen, Karen Bell, Ian M. Roberts, Christophe Lacomme, and Karl J. Oparka. 2009. "Live-cell imaging of viral RNA genomes using a pumilio-based reporter." *The Plant Journal: For Cell and Molecular Biology* 57 (4): 758–70.

56. Wang, Linya, King-Song Jeng, and Michael M. C. Lai. 2011. "Poly(C)-binding protein 2 interacts with sequences required for viral replication in the hepatitis C virus (HCV) 5′ untranslated region and directs HCV RNA replication through circularizing the viral genome." *Journal of Virology* 85 (16): 7954–64.

57. Wells, S. E., P. E. Hillner, R. D. Vale, and A. B. Sachs. 1998. "Circularization of mRNA by eukaryotic translation initiation factors." *Molecular Cell* 2 (1): 135–40.

58. Yamagishi, Yusuke, Ikuo Shoji, Shoji Miyagawa, Takashi Kawakami, Takayuki Katoh, Yuki Goto, and Hiroaki Suga. 2011. "Natural product-like macrocyclic N-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed *de novo* Library." *Chemistry & Biology* 18 (12): 1562–70.

59. Yamaguchi, Junichi, Mohammed Naimuddin, Manish Biyani, Toru Sasaki, Masayuki Machida, Tai Kubo, Takashi Funatsu, Yuzuru Husimi, and Naoto Nemoto. 2009. "cDNA display: A novel screening method for functional disulfide-rich peptides by solid-phase synthesis and stabilization of mRNA-protein fusions." *Nucleic Acids Research* 37 (16): e108.

60. Yoga, Yano M. K., Daouda A. K. Traore, Mahjooba Sidiqi, Chris Szeto, Nicole R. Pendini, Andrew Barker, Peter J. Leedman, Jacqueline A. Wilce, and Matthew C. J. Wilce. 2012. "Contribution of the first K-homology domain of poly(C)-binding protein 1 to its affinity and specificity for C-rich oligonucleotides." *Nucleic Acids Research* 40 (11): 5101–14.

Chapter 3 – One-pot *de novo* synthesis of [4Fe-4S] proteins using recombinant SUF system under aerobic conditions

3.1 Introduction

In **Chapter 2**, a novel approach called codon-restricted mRNA display was developed and applied to identify ssRNA-binding peptides from a random peptide sequence space. This method works by preventing base-pairing interactions between the target RNA and mRNAs. However, codon restriction in this technique exhibits the drawback of limiting the diversity of amino acids that can be incorporated into the peptides, which in turn restricts the range of chemical properties that can be tested. As described in **Chapter 2**, one solution to address this limitation is to utilize codon-reassignment technology. This involves reassigning synonymous codons to code for other canonical or non-canonical amino acids. However, the efficiency of incorporating multiple amino acids remains not high enough to generate a peptide library. Alternative strategy to diversify the chemical properties of the peptide library is to employ RiPP enzymes. As mentioned in **Chapter 1**, rSAM enzymes, which contain a O_2-labile [4Fe-4S] cluster and SAM at their catalytic sites, are capable of catalyzing various types of modifications on peptides, including methylation, epimerization, incorporation of non-canonical amino acids, and macrocyclization reactions. These diverse catalytic activities offer promising applications in *in vitro* evolution campaigns. In this chapter, I have thus developed a method to produce Fe-S cluster proteins using a PURE system under aerobic conditions. This advancement offers new opportunities for harnessing the catalytic capabilities of rSAM enzymes in diverse applications.

Fe-S clusters are essential prosthetic groups of many redox-active proteins and are employed by organisms across all three domains of life, serving as electron shuttles in a variety of metabolic and respiratory networks (Evans, Buchanan, and Arnon 1966; Zhang et al. 1998; Yankovskaya et al. 2003; H. Beinert, Holm, and Münck 1997). Among multiple Fe-S cluster

stoichiometries and coordination, the cubane-type [4Fe-4S] cluster is associated with the lowest reported standard reduction potential down to −645 mV, depending on pH (Stephens, Jollie, and Warshel 1996; Bak and Elliott 2014). [4Fe-4S] proteins play essential roles in central carbon metabolism, N_2 fixation, H_2 respiration, organohalide respiration, the biogeochemical sulfur cycle, and one carbon metabolism (Helmut Beinert, Kennedy, and David Stout 1996; Peters et al. 1995; Odom and Peck 1984; Holliger, Wohlfarth, and Diekert 1998; Jørgensen and Kasten 2006; Ragsdale 2003). Biogenesis of [4Fe-4S] clusters is a complex process involving multiple maturases with long evolutionary histories (Garcia et al. 2022). In bacteria, the assembly process follows a scheme of (***i***) the extraction of the sulfide group from the precursor cysteine; (***ii***) the assembly of [4Fe-4S] cluster on a scaffold; (***iii***) the transfer of the [4Fe-4S] cluster to recipient proteins (Esquilin-Lebron et al. 2021). This scheme can be accomplished by three types of helper protein systems: NIF (**NI**trogen **F**ixation), ISC (**I**ron-**S**ulfur **C**luster), and SUF (mobilization of **SU**l**F**ur). The NIF system is responsible for the cubane-type Fe-S cluster assembly of nitrogenase for N_2 fixation (Jacobson et al. 1989), while the ISC system is the primary system for Fe-S cluster assembly in bacteria (Zheng et al. 1998). The third Fe-S cluster assembly system, the SUF system, is more O_2-tolerant and is activated when cells suffer from iron starvation and oxidative stress (Takahashi and Tokumoto 2002; Dai and Outten 2012; Boyd et al. 2014). The SUF system is encoded by the *suf* operon (*sufABCDSE*). The encoded SUF subunits are assembled into two complexes: SufSE and $SufBC_2D$. The SufS-SufE pair extracts S^{2-} from cysteine; the $SufBC_2D$ complex acts as a scaffold, receiving the S^{2-} group from the SufES complex and incorporating $Fe^{2+/3+}$ for Fe-S cluster assembly, which is an ATP- and $FADH_2$-consuming process (Wollers et al. 2010b); the SufA is a Fe-S cluster carrier protein that transfers the Fe-S clusters to the recipient proteins (**Figure 3-1**).

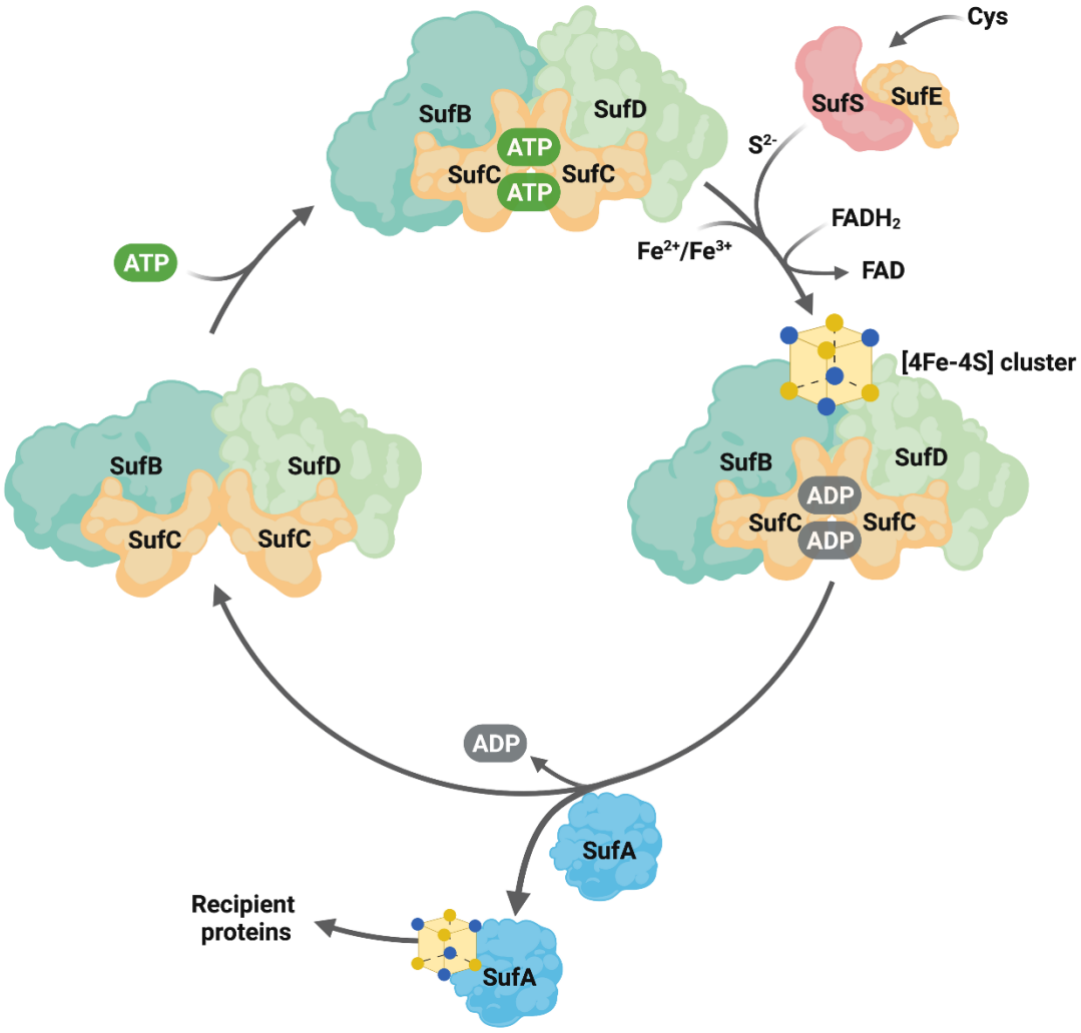


Figure 3-1. Schematic diagram of the SUF pathway. The $SufBC_2D$ complex acts as a scaffold, receiving the sulfide (S^{2-}) group from the SufES complex and incorporating Fe^{2+} and/or Fe^{3+} depending on the type of cluster for Fe-S cluster assembly. $FADH_2$ is used for Fe^{3+} reduction and [4Fe-4S] cluster assembly is driven by ATP hydrolysis. The assembled [4Fe-4S] cluster are then transferred to the SufA which further transfers the cluster to the recipient proteins.

Mature [4Fe-4S] proteins have been purified from cells across the tree of life (Yakunin and Hallenbeck 1998; Kennedy et al. 1992). However, due to the O_2 lability of [4Fe-4S] clusters, the expression and purification of mature [4Fe-4S] proteins often needs to be performed under strictly anaerobic and highly reduced conditions (Khoroshilova et al. 1997). Moreover, purified [4Fe-4S] proteins are often contaminated by Fd due to the high affinity of sub micromolar K_D concentrations (Wan and Jarrett 2002). An alternative method for [4Fe-4S] protein maturation is to reconstitute [4Fe-4S] clusters chemically. For example, the Clostridial apo-Fd was reconstituted into holo-form by incubating apo-Fd with 2-mercaptoethanol, S^{2-}, and $Fe^{2+/3+}$ under anaerobic and alkaline conditions (Hong and Rabinowitz 1967). The traditional chemical reconstitution method has been applied to reconstitute many Fe-S cluster proteins *in vitro* and has reached a level of refinement (Tsibris et al. 1968; Petering, Fee, and Palmer 1971).

Cell-free protein synthesis systems have emerged as a powerful tool for high-throughput protein expression (Shimizu et al. 2001; D.-M. Kim and Swartz 2001; Jewett and Swartz 2004; Swartz, Jewett, and Woodrow 2004). These cell-free systems are flexible, enabling the adjustment of the environmental conditions to produce correctly folded proteins with cofactors (Niwa et al. 2012), and allow the functional expression of enzymes, membrane structural proteins, toxic proteins, and *de novo*-designed polypeptides with high homogeneity. Furthermore, cell-free systems enable protein expression from linear DNA, bypassing the procedures of plasmid cloning and microbe culturing that often cause sequence mutations. Cell extracts-based cell-free systems are easily prepared, robust, and economical, suitable for large-scale, high-yield protein synthesis and high-throughput mRNA translation (Swartz, Jewett, and Woodrow 2004), while the reconstituted cell-free systems (*i.e.*, the PURE systems) allow

protein synthesis in an environment free of proteases, nucleases, and any undefined contaminated components, suitable for biochemical characterization of complex metabolic networks and the bottom-up reconstruction of synthetic cells (Shimizu et al. 2001; Lavickova and Maerkl 2019; Lavickova, Laohakunakorn, and Maerkl 2020).

Previously, the Swartz group established the protocols for synthesizing mature [2Fe-2S] Fd in a cell extract-based cell-free system (*i.e.*, S30 system) with a high yield of 0.45 mg per mL and a maturation efficacy of ~85% (Boyer, Wang, and Swartz 2006). Interestingly, co-expression of the ISC helper proteins in the cells used for cell-extract preparation did not facilitate mature [2Fe-2S] Fd assembly, likely due to the limited ISC expression and activity in host cells under aerobic conditions. Cell-free synthesis of [FeFe] hydrogenase that contains the cubane-type [4Fe-4S] cluster was also attempted by the same group; nevertheless, the maturation efficacy of holo-[FeFe] hydrogenase (44%) was much lower than the [2Fe-2S] Fd even in the presence of the HydEFG helper proteins (Boyer et al. 2008). Due to the O_2 lability of [4Fe-4S] clusters, the cell-free hydrogenase expression and purification were performed under anaerobic conditions, along with the addition of strong reductants to prevent loss of function upon O_2 exposure.

3.2 Materials and Methods

All experiments requiring the anaerobic environment were performed in a Coy Lab type B vinyl anaerobic chamber with an atmosphere of 5:95% (v/v) H_2:N_2 gas mixture.

3.2.1 Experimental materials.

All chemicals were purchased from Sigma-Aldrich (St Louis, MO, USA) unless specified otherwise. The PURExpress reagents were provided by GeneFrontier Corp. (Kashiwa, Chiba, Japan). The pUC57 plasmids carrying the recombinant *E. coli* SUF proteins, *Thermococcus* Fd, and the pET-21d(+) plasmid and the pETDuet-1 plasmid carrying the recombinant *E. coli* catalase (CAT) and superfolder green fluorescent protein (sfGFP), respectively, were purchased from GenScript Biotech Corporation (Nanjing, Jiangsu, China). The pET-23b(+) plasmid carrying the recombinant *Pseudomonas* formate dehydrogenase (FDH) gene is a gift from Prof. Alexander F. Yakunin, Bangor University, United Kingdom (Batyrova et al. 2020). The ASKA *E. coli* clones for aconitase (AcnA) and NADH-dependent flavin reductase (FRE) were purchased from the National Institute of Genetics (SHIGEN) (Mishima, Shizuoka, Japan) (Kitagawa et al. 2005). The detailed information for corresponding recombinant proteins is shown in **Table 3-1**.

3.2.2 *In vitro* transcription.

In vitro transcription was performed using ScriptMax Thermo T7 Transcription Kit (TOYOBO, Osaka, Japan) following the manufacturer's standard protocols. Briefly, DNA templates (0.7 μg) were added to the RNA transcription reaction mixtures (40 μL) and were

incubated at 40°C for 4 h. The RNA transcripts were then purified using NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, Germany) and the concentrations were measured by NanoDrop^{TM}2000c spectrophotometer (Thermo Fisher Scientific, MA, USA).

3.2.3 Gene cloning.

The *suf* genes (*sufA/sufB/sufC/sufD/sufE/sufS*) were cloned into a pET-26b(+) expression vector by seamless ligation cloning extract (SLiCE) cloning (Motohashi 2015), while *acnA* gene was cloned into pET-23a(+) for *in vitro* transcription. The linearized vector was amplified from pET-26b(+)/pET-23a(+) plasmid DNA using the appropriate forward and reverse primers. The *suf* genes were amplified from pUC57 plasmid DNA, while *acnA* gene was amplified from the corresponding ASKA *E. coli* clone. Note that the human rhinovirus (HRV) 3C protease-cutting site was inserted between the gene of N-His_6-tag and *acnA* by PCR. All primer sequences are available in **Table 3-1**. The SLiCE cloning reaction was performed as described previously (Motohashi 2017). Briefly, SLiCE solution (1 μL) and 10× SLiCE buffer (1 μL) were mixed with 60 ng of the insert DNA and 60 ng of the linearized pET-26b(+) vector (C-His_6 tag removed by PCR), and the final volume of the SLiCE reaction was adjusted to 10 μL using sterilized ddH_2O. The reaction mixture was incubated at 37°C for 15 min to produce the pET-26b(+)-SufA/SufB/SufC/SufD/SufE/SufS and pET-23a(+)-AcnA plasmids, respectively. The pET-26b(+)-SufA/SufB/SufC/SufD/SufE/SufS plasmids were transformed into *E. coli* BL21(DE3) cells (New England Biolabs, Inc., Ipswich, MA, USA) for heterologous protein overexpression, respectively, while pET-23a(+)-AcnA plasmid was subsequently amplified by PCR to make a linear DNA harboring the T7 promoter for the following *in vitro* transcription.

3.2.4 Heterologous protein overexpression and purification.

Heterologous protein overexpression and purification were performed based on the established protocols published previously with some modifications (P.-H. Wang et al. 2020). Briefly, the cultures were grown under aerobic conditions at 37℃ in a shaker using Terrific Broth medium at 200 rpm until a specific optical density was reached ($OD_{600\ nm}$ 0.6), followed by the induction using isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.4 mM) and overnight incubation at 16℃. For *Thermococcus profundus* Fd and *Escherichia coli* AcnA expression, after 2 h of the IPTG induction, the cultures were transferred to 1 L screw-capped glass bottles and were supplemented with cysteine (0.2 mM) and $FeCl_3$ (0.1 mM), followed by overnight incubation at 37℃ for Fd and at 16℃ for AcnA, respectively, with constant stirring. Recombinant proteins were purified to at least 95% homogeneity using Ni-sepharose-based immobilized affinity chromatography as previously described (Kuznetsova et al. 2006), and protein purity was analyzed by SDS-PAGE (SDS-PAGE gels (5–20%, w/v); Supercep™ Ace; Fujifilm Wako Pure Chemical, Osaka, Japan) followed by Coomassie staining (**Figure 3-2**). Briefly, cells were collected by centrifugation at 8,000 × g for 10 min at 4℃, suspended in the binding buffer: HEPES-Na (pH 7.5; 25 mM), NaCl (0.5 M), glycerol (5%, v/v), and imidazole (5 mM), and the cells were homogenized *via* sonication in an ice bath for 10 min (1 s on and 2 s off). For Fd, the lysate was subsequently boiled for 20 min at 92℃ to denature the endogenous proteins. The lysate was centrifuged at 20,000 × g for 30 min at 4℃ and the supernatant was passed through a glass gravity column containing approximately 0.5 mL Ni Sepharose™ High Performance resin (GE Healthcare, IL, USA). After washing with the wash buffer: HEPES-Na (pH 7.5; 25 mM), NaCl (0.5 M), glycerol (5%, v/v), and imidazole (25 mM), recombinant

proteins were eluted using the elution buffer: HEPES-Na (pH 7.5; 25 mM), NaCl (0.5 M), glycerol (5%, v/v), and imidazole (250 mM). For Fd and holo-AcnA, $FeSO_4$ (0.1 mM) and Na_2S (0.8 mM) were supplemented to the lysate and wash buffer to chemically reconstitute Fe-S cluster on the proteins. The protein stock solutions were flash-frozen in liquid N_2 and stored at −80°C. The N-terminal His_6-tag was retained for all experiments except the experiment for purification of the holo-AcnA synthesized by the PUREfrex.

3.2.5 Preparation of apo-Fd and apo-AcnA.

Apo-Fd and apo-AcnA devoid of the [4Fe-4S] clusters were synthesized using a PUREfrex2.0 following the manufacturer's standard protocols. Briefly, in an ice bath, solution I (amino acids, NTPs, tRNAs and substrates for enzymes, etc.; 10 μL), solution II (enzyme mixtures; 1 μL), and solution III (ribosome; 2 μL) were mixed with the mRNAs (0.4–0.7 μM) to a final volume of 20 μL. Cell-free protein synthesis was performed at 37°C for up to 6 h and the reaction mixture was kept on ice after incubation before further SDS-PAGE analysis or reconstitution experiments.

3.2.6 *In vitro* assembly of the [4Fe-4S] cluster on the $SufBC_2D$ complex.

(*i*) Anaerobic conditions.

The [4Fe-4S]-charged $SufBC_2D$ complex was anaerobically produced inside the anoxic chamber. Reaction mixture containing HEPES-Na (pH 7.6; 50 mM), NaCl (100 mM), $MgCl_2$ (5 mM), dithiothreitol (DTT) (1 mM), pyridoxal 5′-phosphate (10 μM), ATP (1 mM), SufE (0.5 μM), SufS (0.5 μM), SufB (10 μM), SufC (20 μM), SufD (10 μM), $FeSO_4$ (0.1 mM), cysteine (0.5 mM), and glycerol (5%, v/v). The reaction was initiated by the addition of cysteine

and incubated at 37°C for 1 h. Optical analysis of the $SufBC_2D$ complex was performed using a microplate reader (Epoch 2 microplate reader; Agilent, CA, USA). The [4Fe-4S]-charged $SufBC_2D$ was later purified using Ni-sepharose-based immobilized affinity chromatography as described in **section 3.2.4**. The protein stock solutions were flash-frozen by liquid N_2 and stored at -80°C.

(*ii*) Aerobic conditions.

The [4Fe-4S]-charged $SufBC_2D$ complex was aerobically produced on the bench top. Reaction mixture containing HEPES-Na^+ (pH 7.6; 50 mM), NaCl (100 mM), $MgCl_2$ (5 mM), DTT (1 mM), pyridoxal 5′-phosphate (10 μM), ATP (1 mM), sodium formate (10 mM), NADH (0.1 mM), FAD (20 μM), FDH (0.05 mg/mL), FRE (0.05 mg/mL), CAT (0.01 mg/mL), SufE (0.5 μM), SufS (0.5 μM), SufB (10 μM), SufC (20 μM), SufD (10 μM), $FeSO_4$ (0.1 mM), cysteine (0.1 mM), and glycerol (5%, v/v). The reaction mixture amended with the O_2-scavenging enzyme cascade was first assembled and incubated at 37°C for 10 min to scavenge the dissolved O_2. The SUF proteins were then added to the reaction mixture. The reaction mixture was incubated at 37°C for 60 min. Optical analysis of the [4Fe-4S]-charged $SufBC_2D$ complex was performed using a microplate reader (Epoch 2 microplate reader; Agilent, CA, USA). The reaction mixture was flash-frozen in liquid N_2 and stored at -80°C for the following *in vitro* maturation of [4Fe-4S] proteins.

3.2.7 *In vitro* maturation of apo-Fd and apo-AcnA using SUF system under anaerobic conditions.

Apo-Fd and apo-AcnA were synthesized by the PUREfrex2.0 as described in **section 3.2.5**. The apo-proteins were then mixed with mature $SufBC_2D$ complex inside the anoxic chamber. The reaction mixture contained HEPES-K (pH 7.6; 50 mM), DTT (0.5 mM), 2-mercaptoethanol (0.1%, v/v), $FeCl_3$ (40 μM), anaerobically/aerobically mature $SufBC_2D$ (2 μM), SufA (1.5 μM), apo-Fd/AcnA (0.01–0.1 mg/mL), and glycerol (5%, v/v). Note that for control experiments, heat-denature SUF was prepared by incubating the mature $SufBC_2D$ and SufA at 95℃ for 10 min and was added to the reaction as well as S^{2-} (8 μM) was added to the reaction instead of the SUF system. The reaction mixture was then incubated at 37℃ for 10 min and dialyzed using the Amicon Ultra-0.5 centrifugal filter (molecular weight cut-off: 3 kDa) (Merck KGaA, Darmstadt, Germany) inside the anaerobic chamber.

3.2.8 The bifunctional enzyme cascade for $FADH_2$ regeneration and O_2 scavenge.

The O_2-scavenging enzyme cascade reaction was performed in the reaction mixture (100 μL) with HEPES-Na (pH 7.6; 50 mM), NaCl (100 mM), $MgCl_2$ (5 mM), sodium formate (10 mM), NADH (0.1 mM), FAD (20 μM), FDH (0.05 mg/mL), FRE (0.05 mg/mL), CAT (0.01 mg/mL), and glycerol (5%, v/v). The reaction was initiated by the addition of FDH. Assays were performed under aerobic conditions using an optical 96-well microplate at 37℃ for 10 min with a UV-vis scanning at $\lambda_{445\ nm}$ every 5 s on a microplate reader. The stock solution of sodium formate (1 M) was adjusted to pH 7.6 in HEPES-Na buffer (50 mM) to prevent a drastic pH shift in the reaction mixture.

3.2.9 Time-course [4Fe-4S] cluster degradation under aerobic conditions.

The mature SufBC$_2$D produced under anaerobic conditions was incubated under aerobic conditions with and without the O_2-scavenging enzyme cascade. The reaction mixture contained HEPES-Na (pH 7.6; 50 mM), NaCl (100 mM), $MgCl_2$ (5 mM), DTT (1 mM), sodium formate (10 mM), NADH (0.1 mM), FAD (20 μM), FDH (0.05 mg/mL), FRE (0.05 mg/mL), CAT (0.01 mg/mL), mature SufBC$_2$D (10 μM), and glycerol (5%, v/v). The reaction mixture was assembled to a final volume of 200 μL in the 300 μL PCR tubes under anaerobic conditions, and the reaction was initiated by opening the lid under aerobic conditions. The reaction mixture was incubated at 37°C for 0, 30, 60, and 90 min. Time-course [4Fe-4S] cluster degradation ($\lambda_{420\ nm}$) in the reaction mixtures was monitored using an Epoch 2 microplate reader (Agilent, CA, USA) under anaerobic conditions.

3.2.10 sfGFP synthesis using a cell-free synthesis system amended with the bifunctional $FADH_2$ regeneration/O_2-scavenging enzyme cascade.

sfGFP was synthesized using a customized PUREfrex2.0 system following the manufacturer's standard protocols. Briefly, in an ice bath, solution I (amino acids, NTPs, tRNAs and substrates for enzymes, etc.; 100 μL), solution II (enzyme mixtures; 10 μL), and solution III (ribosome; 20 μL) were mixed with the components to final concentrations of FDH (0.05 mg/mL), FRE (0.05 mg/mL), and CAT (0.025 mg/mL), sodium formate (10 mM), pyridoxal 5′-phosphate (10 μM), NADH (0.2 mM), FAD (20 μM), and sfGFP mRNA (0.4–0.7 μM) to a final volume of 200 μL. Note that the reaction mixture was assembled in a 0.2 mL tube to reduce the headspace of the reaction. Cell-free protein synthesis was performed at 37°C for 6 h, and the

reaction mixtures were kept on ice before fluorescence analysis. The production of sfGFP was estimated using fluorescence emission at 518 nm (excitation, 494 nm) on a microplate reader (Epoch 2 microplate reader, Agilent, CA, USA), with standard curves generated by purified sfGFP (50–350 μg/mL).

3.2.11 One-pot cell-free synthesis of holo-AcnA under aerobic conditions.

Apo-AcnA was synthesized using a customized PURExpress2.0 system following the manufacturer's standard protocols on a bench top. Briefly, in an ice bath, solution I (amino acids, NTPs, tRNAs and substrates for enzymes, etc.; 100 μL), solution II (enzyme mixtures; 10 μL), and solution III (ribosome; 20 μL) were mixed with the components to final concentrations of FDH (0.05 mg/mL), FRE (0.05 mg/mL), CAT (0.025 mg/mL), sodium formate (10 mM), pyridoxal 5′-phosphate (10 μM), NADH (0.2 mM), FAD (20 μM), and AcnA mRNA (0.4–0.7 μM) to a final volume of 200 μL. Cell-free protein synthesis was performed at 37°C for 6 h, and the reaction mixtures were kept on ice after incubation before *in vitro* maturation of AcnA using the SUF system. The PURExpress reaction containing apo-AcnA was then supplemented with $FeSO_4$ (0.1 mM) and individually purified SUF subunits: SufA (1.5 μM), SufB (8 μM), SufC (16 μM), SufD (8 μM), SufE (0.4 μM), and SufS (0.4 μM), and was incubated at 37°C for 10 min. The holo-AcnA was purified using Ni-sepharose-based immobilized affinity chromatography as described in **section 3.2.4**.

3.2.12 Cysteine desulfurase activity assays.

Total sulfide was measured based on a previously published protocol (Urbina et al. 2001). Reactions were carried out anaerobically in a buffer containing HEPES-Na (pH 7.5; 50 mM),

NaCl (0.1 M), SufE (3 μM), SufS (0.5 μM), and glycerol (5%, v/v). Pyridoxal 5′-phosphate was added to 10 μM, and reactions were initiated by dilution of an L-cysteine/DTT stock (both 40 mM) to a 1 mM final concentration in a total reaction volume of 1 mL. Reactions were allowed to proceed for 15 min at 37°C and then were quenched by mixing 100 μL of the reaction with 100 μL of N, N-dimethyl-p-phenylenediamine (17 mM), and $FeCl_3$ (22.2 mM) in HCl (6 M). Subsequently, the quenched reactions were incubated for 20 min, leading to the formation of methylene blue. Precipitated protein was removed by 30 s centrifugation at 20,000 × g, and methylene blue was measured with a UV-vis scanning at $\lambda_{670\ nm}$ on a microplate reader (Epoch 2 microplate reader, Agilent, CA, USA).

3.2.13 Cytochrome C reduction optical assays.

Oxidized Fd is reduced enzymatically by spinach Fd-$NADP^+$ reductase (FPR; Sigma-Aldrich, MO, USA) using NADPH. Reduced Fd was used to reduce cytochrome C (Cyt C) from the equine heart (Sigma-Aldrich, MO, USA) in a non-enzymatic manner. The formation of reduced Cyt C is monitored by the time-dependent increase in $\lambda_{550\ nm}$. The reaction mixtures (300 μL) contained HEPES-Na (pH 8.1; 50 mM), NaCl (0.1 M), glycerol (5%, v/v), NADPH (1 mM), Cyt C (0.1 mM), FPR (0.25 μg/mL; Sigma-Aldrich, MO, USA), and reconstituted holo-Fd. The holo-Fd was diluted and assayed at concentrations ranging from 0.5 to 4.0 μM. The reaction was initiated by the addition of holo-Fd. Under aerobic conditions, assays were performed in an optical 96-well microplate at 30°C for 25 min with a UV-vis scanning at $\lambda_{550\ nm}$ every 5 s on a microplate reader (EnSpire; PerkinElmer, MA, USA).

3.2.14 AcnA and isocitrate dehydrogenase-coupled activity assays.

AcnA activity was assayed using the procedure previously described (Gardner and Fridovich 1992) by monitoring the formation of NADPH through the increase in $\lambda_{340\ nm}$. The reaction mixture (100 µL) contained HEPES-K (pH 7.6; 50 mM), $MnCl_2$ (0.5 mM), citrate (10 mM), $NADP^+$ (1 mM), 2-mercaptoethanol (0.1%, v/v), and isocitrate dehydrogenase (IDH) (0.2 U; Sigma-Aldrich, MO, USA) with different concentrations of recombinant holo-AcnA. Note that the stock solution for the citrate (1 M) was adjusted to pH 7.6 in HEPES-K (50 mM) buffer to prevent a pH shift in the reaction mixture. The reaction was initiated by the addition of holo-AcnA. Under anaerobic conditions, assays were performed in an optical 96-well microplate at 37°C for 10 min with a UV-vis scanning at $\lambda_{340\ nm}$ every 10 s on a microplate reader (Epoch 2 microplate reader, Agilent, CA, USA). The $\lambda_{340\ nm}$ was later converted to NADPH concentration based on the standard curves derived from the different concentrations of NADPH. All enzyme assays were conducted in duplicates or triplicates as indicated in the figures. The concentration of reconstituted holo-AcnA was estimated using a standard curve generated by serial diluted holo-AcnA overexpressed and purified from *E. coli*. Concentration of AcnA in the enzyme assays were estimated from the band intensity in the SDS-PAGE by using AcnA purified from *E. coli* as a standard.

3.2.15 ESI-LC-MS analysis.

The completed reactions of the AcnA and IDH-coupled activity assays were collected and filtered through the Amicon® Ultra-0.5 centrifugal filter (molecular weight cut-off: 3 kDa) (Merck KGaA, Darmstadt, Germany), and the flowthrough fractions were used for an ESI-LC-

MS system consisting of an Acquity UPLC plus system (Waters, Milford, MA, USA) connected to XEVO G2-XS QTof Mass Spectrometry (Waters, Milford, MA, USA). The citrate and alpha-ketoglutarate were separated and detected as described previously (Hodek et al. 2022). Briefly, the separation of underivatized carboxylates was achieved by injecting 5 μL of a sample onto an Atlantis Premier BEH C18 AX column (100×2.1 mm, 1.7 mm; Waters, Milford, MA, USA). The mobile phase was composed of (A) 10 mM ammonium formate in the pH 2.6 water and (B) 10 mM ammonium formate in 50/50 water/methanol (v/v) of pH 9.4 (pH of aqueous buffer before mixing with methanol). The mobile phase was delivered to the column at a flow rate of 0.35 mL/min with the following gradient: 0.0 min (0% B), 0.5 min (0% B), 3.0 min (70% B), 6.0 min (70% B), 7.0 min (0% B), and 10.0 min (0% B). The column was thermostated at 30°C. Analytes were ionized in an electrospray ion source operated in the negative mode. The source and gas parameters were set as follows: capillary 2.50 kV, source temperature 100 °C, cone gas 100 L/h, desolvation temperature 300°C, and desolvation gas 490 L/h.

Table 3-1. Complete list of recombinant proteins, expression vectors, and primers for gene cloning used in this work.

Protein name	Gene name	Organism	Uniprot ID	Expression plasmid	Original plasmid	Molecular weight (kDa)	Forward primer sequence (5′ to 3′)	Reverse primer sequence (5′ to 3′)
Fd	*frd*	*Thermococcus profundus*	Q9HHD4	pET-23a(+)	pET-23a(+)	7.8	TAATACGACTCACTATAGGG	CAATCCGGATATAGTTCCTCC
SufA	*sufA*	*Escherichia coli*	P77667	pET-26b(+)	pUC57	14.6	ATGCATCACCACCACCACCATTCACAAGATCCAGATATGCACTCAGGAACATTTAAT	CCAACTCAGCTTCCTTTCGGGCTTTAACGATATCGCGCTATTTGCC
SufB	*SufB*	*Escherichia coli*	P77522	pET-26b(+)	pUC57	56.0	ATGCATCACCACCACCACCATTCACAAGATCCATCAAGGAATACAGAAGCTACTG	CCAACTCAGCTTCCTTTCGGGCTTTAACGATATCGCGCTATTTGCC
SufC	*sufC*	*Escherichia coli*	P77499	pET-26b(+)	pUC57	29.9	ATGCATCACCACCACCACCATTCACAAGATCCACTATCAATAAAAGATTTACACGTAAGT	CCAACTCAGCTTCCTTTCGGGCTTTAACGATATCGCGCTATTTGCC
SufD	*sufD*	*Escherichia coli*	P77689	pET-26b(+)	pUC57	48.1	ATGCATCACCACCACCACCATTCACAAGATCCAGCAGGACTACCCAACAGT	CCAACTCAGCTTCCTTTCGGGCTTTAACGATATCGCGCTATTTGCC
SufE	*sufE*	*Escherichia coli*	P76194	pET-26b(+)	pUC57	17.1	ATGCATCACCACCACCACCATTCACAAGATCCAGCTTTACTACCCGATAAAGAAAAG	CCAACTCAGCTTCCTTTCGGGCTTTAACGATATCGCGCTATTTGCC
SufS	*sufS*	*Escherichia coli*	P77444	pET-26b(+)	pUC57	45.7	ATGCATCACCACCACCACCATTCACAAGATCCAATATTTTCAGTTGATAAAGTAAGGGCT	CCAACTCAGCTTCCTTTCGGGCTTTAACGATATCGCGCTATTTGCC
AcnA	*acnA*	*Escherichia coli*	P25516	pET-23a(+)	ASKA clone	99.8	CACCATCTGGAAGTTCTGTTCCAGGGGCCCTCACAAGATCCATCGTCAACCCTACGAGAA	CTTTGTTAGCAGCCGGATTCATTATTTCAACATATTACGAATGAC
FRE	*fre*	*Escherichia coli*	P0AEN1	pET-23a(+)	pET-23a(+)	27.5		
FDH	*fdh*	*Pseudomonas sp. (strain 101)*	P33160	pET-26b(+)	pET-26b(+)	45.4		
CAT	*katE*	*Escherichia coli*	P21179	pET-21d(+)	pET-21d(+)	84.1		

Expression vector	Forward primer sequence (5′ to 3′)	Reverse primer sequence (5′ to 3′)
pET-23a(+)	ATCCGGCTGCTAACAAAGC	TGGATCTTGTGAATGGTGGTG / GGGCCCCTGGAACAGAACTTCCAGATGGTGGTGGTGGTGATGCAT (to insert HRV 3C cutting site sequence between N-His6 tag and *acnA*)
pET-26b(+)	AAAGCCCGAAAGGAAGCTGAGTTGG	TGGATCTTGTGAATGGTGGTGGTGGTGATGCATCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTC

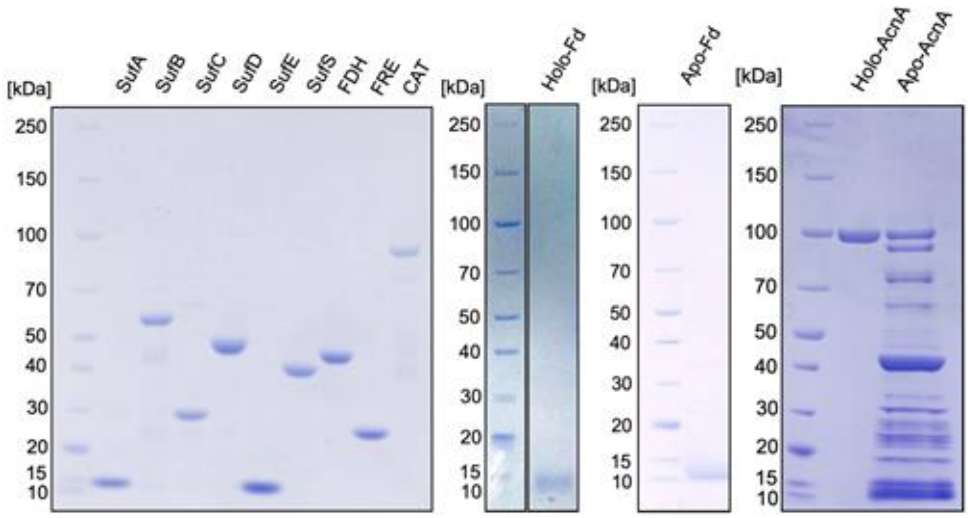


Figure 3-2. SDS-PAGE gel images of purified recombinant proteins and apo-ferredoxin/aconitase A synthesized in the PURE system. All recombinant proteins (SufABCDSE, formatae dehydrogenase (FDH), flavin reductase (FRE), catalase (CAT), holo-ferredoxin (Fd) and holo-aconitase A (AcnA)) are overexpressed and purified from *E. coli.* apo-Fd was expressed in the PURE system and purified by boiling and centrifugation. apo-AcnA was ran on the gel along with the protein components in the PURExpress reaction mixture. We did not purify apo-AcnA due to its application for the one-pot *de novo* synthesis of holo-AcnA.

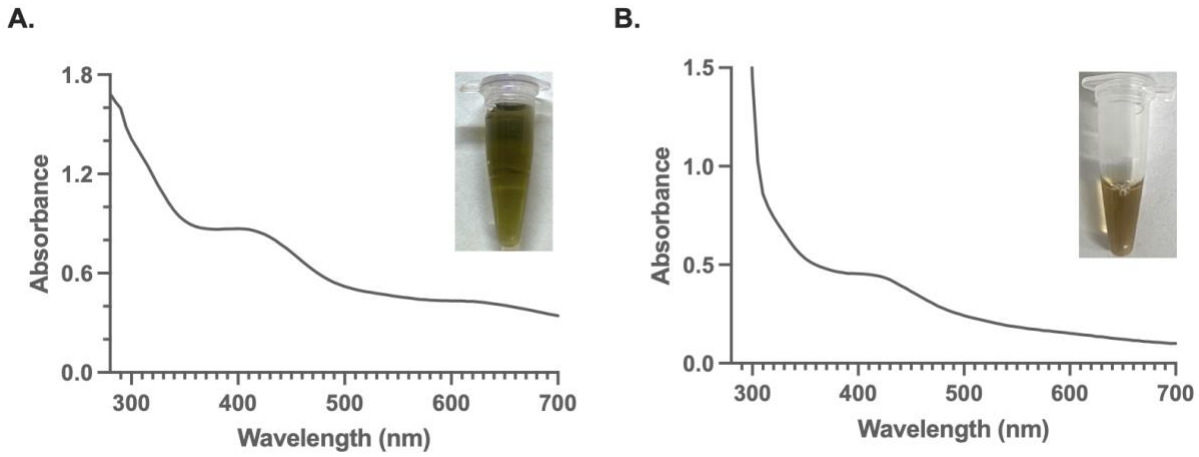


Figure 3-3. UV-vis analysis of the recombinant proteins purified from the anaerobically cultivated *E. coli* BL21 cells. The two recombinant proteins (A) holo-ferredoxin (Fd) and (B) aconitase A(AcnA) were purified in an anaerobic chamber under strictly anaerobic conditions. The UV-vis analysis was also performed in an anaerobic chamber. Visual image of the purified protein in a test tube are also presented.

3.3 Results and Discussion

In this study, I investigated the expression of mature [4Fe-4S] proteins using the PURE system (PURExpress) and the SUF helper protein system. Ideally, the synthesis of [4Fe-4S] proteins would be possible under aerobic conditions, alleviating the need for complex anaerobic systems. The PURExpress already includes an efficient ATP regeneration system which can also be used to drive SUF-mediated [4Fe-4S] cluster assembly (P.-H. Wang et al. 2020). To supply $FADH_2$ to the SUF system and to overcome the O_2 lability of [4Fe-4S] clusters, I designed an O_2-scavenging cascade including FDH, FRE, and CAT to (***i***) sequentially regenerate $FADH_2$ from formate *via* NADH and (***ii***) scavenge the dissolved O_2.

3.3.1 Stepwise *in vitro* reconstitution of the SUF system from individually purified SUF subunits.

I first sought to reconstitute the SUF system *in vitro via* a stepwise manner using the recombinant SUF subunits individually purified from *E. coli* (**Figure 3-2**). According to the literature, the first step in the SUF-mediated Fe-S cluster assembly is S^{2-} extraction from cysteine by the SufES complex (Layer et al. 2007). Indeed, I observed S^{2-} accumulation in reactions when both cysteine and recombinant SufES were added (**Figure 3-4A**). Next, I tried to validate the activities of the recombinant $SufBC_2D$ complex *in vitro,* because previous works generally co-expressed the *sufBCD* and directly purified the mature $SufBC_2D$ complex from the host *via* multi-step liquid chromatography (Outten et al. 2003). [4Fe-4S] cluster assembly on the $SufBC_2D$ complex can be inferred from the increase in the characteristic optical absorbance of [4Fe-4S] clusters at 420 nm (Johnson 1998). I first reconstituted the $SufBC_2D$

complex under anaerobic conditions. Notably, a small initial amount of [4Fe-4S] cluster was observed (**Figure 3-4B**) and the $\lambda_{420\ nm}$ of the reconstituted $SufBC_2D$ complex decreased after aeration. I then incubated the oxidized $SufBC_2D$ complex with cysteine, $Fe^{2+/3+}$, ATP, $FADH_2$ (reduced by dithionite), and the recombinant SufES under anaerobic conditions, and observed a time-course increase in the $\lambda_{420\ nm}$ (**Figure 3-4C**). Moreover, the increase in $\lambda_{420\ nm}$ were not observed in the assays devoid of the recombinant SufES (**Figure 3-4D**), suggesting that the recombinant $SufBC_2D$ complex, along with SufES, is capable of assembling [4Fe-4S] clusters *in vitro* under anaerobic conditions. It is worthwhile mentioning that adding a low $Fe^{2+/3+}$ concentration (< 0.1 mM) in the assays prevents the formation of blackish FeS (Sanden et al. 2020), which is required for optical measurement in the absence of turbidity (**Figure 3-5**).

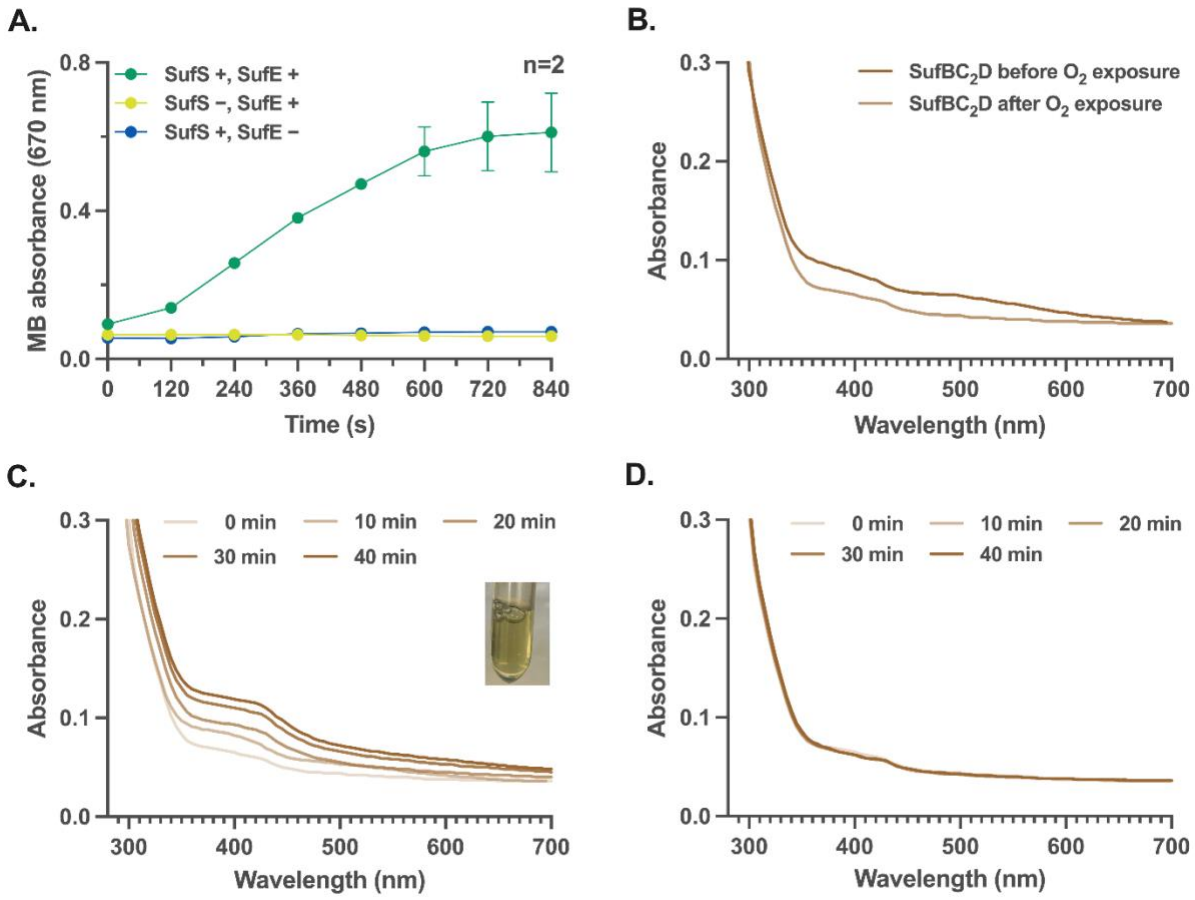


Figure 3-4. Anaerobic [4Fe-4S] cluster reconstitution using the SUF system. (A) Cysteine desulfurase activity of the SufES complex was measured with and without SufS or SufE. Free S^{2-} production was measured based on the time-dependent production of methylene blue ($\lambda_{670\ nm}$). The bars represent the range and the data points represent the average from two independent experimental replicates, respectively. (B) Observation of [4Fe-4S] clusters ($\lambda_{420\ nm}$) degradation on the reconstituted $SufBC_2D$ complex after a 10-min aeration. (C–D) SUF-mediated [4Fe-4S] assembly was conducted with the oxidized $SufBC_2D$ complex in the (C) presence or (D) absence of SufES under anaerobic conditions.

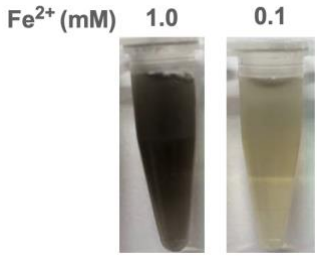


Figure 3-5. Images of the mature SufBC$_2$D at different concentrations of Fe^{2+}.

Next, I validated whether the recombinant SufA is capable of transferring [4Fe-4S] clusters from the $SufBC_2D$ complex to recipient proteins *in vitro*. A thermotolerant [4Fe-4S] Fd from the hyperthermophilic archaeon *Thermococcus profundus* was used as [4Fe-4S] cluster recipient protein (Imai et al. 2001). Cyt C reduction by Fd was followed at $\lambda_{550\,nm}$, with electrons supplied by a FPR system for Fd reduction (**Figure 3-6A**) (Boyer, Wang, and Swartz 2006). Often purified mesophilic [4Fe-4S] Fd results in contaminated Fd-binding proteins. The thermotolerance of *Thermococcus* Fd allows a single-step purification by heating (at 92℃ for 10 minutes) (**Figure 3-2**). Accordingly, I heterologously expressed and purified the holo-Fd from anaerobically cultivated *E. coli* cultures. Following a chemical reconstitution to fully mature the recombinant Fd (**Figure 3-3A)**, the brown-colored holo-Fd preparations with a defined gradient of concentrations were used to establish the optical Cyt C reduction assay (**Figure 3-6B**).

Subsequently, I synthesized and purified the recombinant apo-Fd using the PURExpress free of the SUF system and $Fe^{2+/3+}$ under aerobic conditions (**Figure 3-2**). The recombinant apo-Fd was incubated with the recombinant $SufBC_2D$ complex charged with [4Fe-4S] clusters (from **Figure 3-3C**) with and without either Fe^{2+}, cysteine, or recombinant apo-Fd (**Figure 3-6C**). The background activities in the negative controls could result from (***i***) the trace Fd contaminated in the commercial spinach FPR (K_m of Fd < 3 μM) used for Fd reduction in the Cyt C reduction assays (Aliverti et al. 2008); (***ii***) the less efficient [4Fe-4S] cluster transfer by the $SufBC_2D$ complex; and (***iii***) the chemical reduction by reductants (*e.g.*, Na_2S and DTT). These data suggested that the reconstituted SUF system (SufABCDSE) likely can mature apo-[4Fe-4S] proteins *in vitro*.

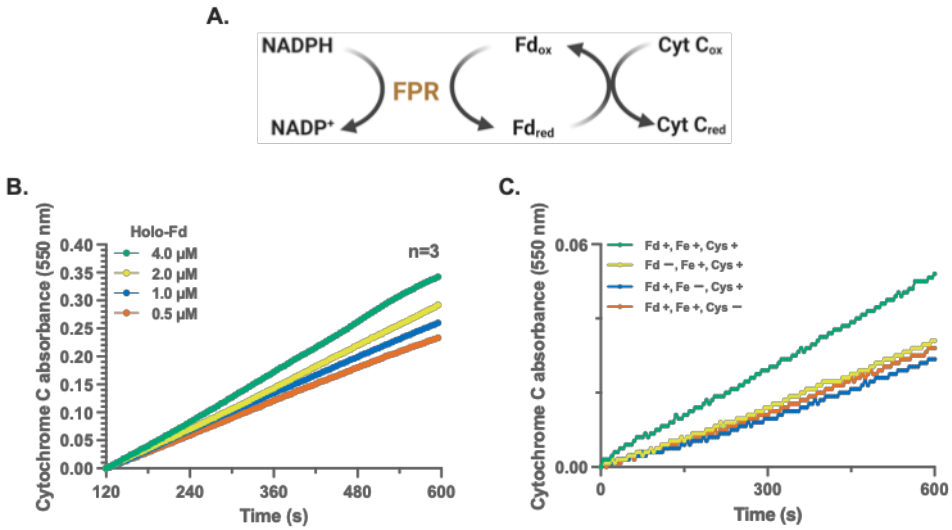


Figure 3-6. Standard curve to estimate the holo-ferredoxin concentration using the cytochrome C-coupled optical assays. (A) Schematic diagram showing the enzyme cascade of the cytochrome C (Cyt C)-coupled optical assays. (B) The activity of the recombinant holo-ferredoxin (Fd) was measured by the Cyt C-coupled optical assays. The bars represent the standard error, and the dots represent the mean from three experimental replicates. (C) The activity of the anaerobically reconstituted Fd was measured by the Cyt C-coupled optical assays. Apo-Fd was matured by the [4Fe-4S]-charged $SufBC_2D$ complex with and without Fd, Fe^{2+}, and cysteine. Cyt C_{ox}, oxidized Cyt C; Cyt C_{red}, reduced Cyt C; Fd_{ox}, oxidized Fd; Fd_{red}, reduced Fd; FPR, Fd-$NADP^+$ reductase.

To gain a more definite answer, I also used PURExpress to produce apo-form of *E. coli* AcnA. The maturation of the SUF system can be validated by an NADP-dependent optical assay (λ_{340nm}) (**Figure 3-7A**) where isocitrate, the product of AcnA, is converted to alpha-ketoglutarate by the NADP-dependent IDH, with holo-AcnA purified from anaerobically cultivated *E. coli* cultures as the standard (fully reconstituted *via* the chemical method) (**Figures 3-3B and 3-7BC**). Apo-AcnA incubated with the SUF system for 15 min under anaerobic conditions demonstrated an enzyme activity of ~90% of holo-AcnA activity; no AcnA activity was observed in the reactions without apo-AcnA nor the matured SUF system (**Figure 3-8A**). Moreover, the apo-AcnA incubated with the $SufBC_2D$ complex but without SufA demonstrated a reduced enzyme activity (~70%). Furthermore, no AcnA activity was observed when the apo-AcnA was incubated with either heat-denatured $SufBC_2D$ complex or S^{2-} for 15 min, suggesting that the [4Fe-4S] cluster of holo-AcnA was transferred from the recombinant SUF system but not assembled abiotically under the reduced conditions (**Figure 3-8B**). Together, my data revealed that the recombinant $SufBC_2D$ complex is capable of assembling and subsequently transferring the [4Fe-4S] clusters to the recipient proteins with a limited efficiency *in vitro* under anaerobic conditions, and the efficiency can be enhanced by adding the recombinant SufA.

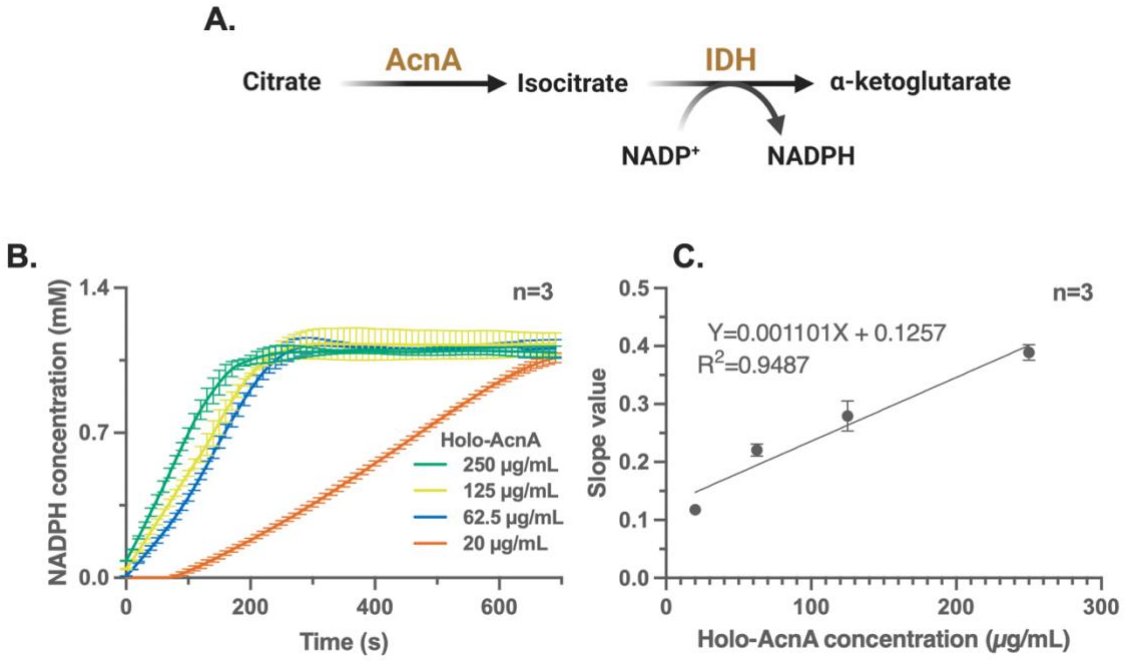


Figure 3-7. Standard curve to estimate the holo-aconitase A concentration using the aconitase A and isocitrate dehydrogenase-coupled activity assays. (A) Schematic diagram showing the enzyme cascade of the aconitase A (AcnA) and isocitrate dehydrogenase (IDH)-coupled activity assays. (B) The activity of the recombinant holo-AcnA purified from *E. coli* was measured by the AcnA and IDH-coupled activity assays. (C) Standard curves of holo-AcnA derived from the slope value (calculated from the linear range (100–130 s)) of the activity measured in (B). The bars represent the standard error, and the dots represent the mean from three experimental replicates.

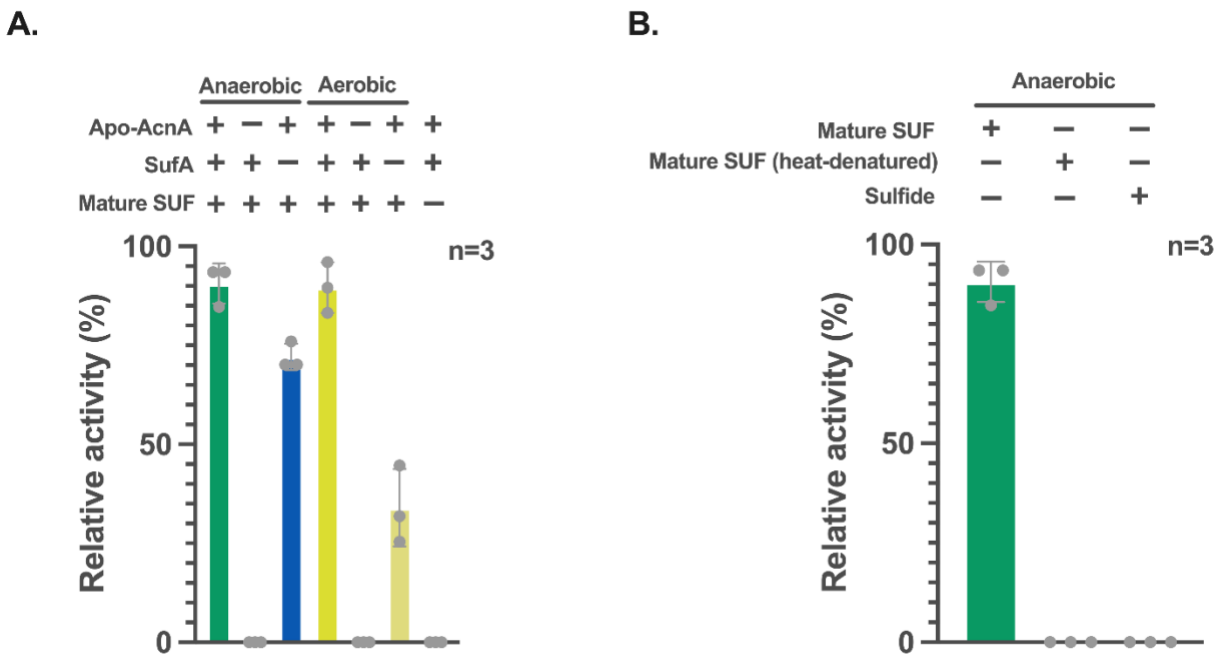


Figure 3-8. [4Fe-4S] cluster transfer from mature SufBC$_2$D complex to apo-aconitase A. (A) Apo-aconitase A (AcnA) maturation by the SUF system under aerobic or anaerobic conditions. (B) Anaerobic holo-AcnA synthesis using mature SUF system, heat-denatured SUF system or S^{2-}. The reconstitution efficiency of holo-AcnA is evaluated as a function of the relative enzymatic activity to that of the fully chemically reconstituted AcnA using AcnA and isocitrate dehydrogenase (IDH) activity assay (See section 3.2.12). The bars represent the standard deviations. The dots and columns represent the individual values and mean from three independent experimental replicates, respectively.

3.3.2 A bifunctional O_2-scavenging enzyme cascade enabling aerobic [4Fe-4S] cluster assembly.

Following *in vitro* reconstitution of the SUF system under anaerobic conditions, I then worked to overcome the O_2 lability of [4Fe-4S] clusters by amendment with an O_2-scavenging system. Inherent from the O_2-lability of Fe-S cluster proteins, several enzymatic O_2-scavenging approaches have been developed during the past few decades. CAT, glucose oxidase, and protocatechuate dioxygenase represent the most common O_2-scavenging enzymes(Patil and Ballou 2000; Aitken, Marshall, and Puglisi 2008). However, in my case, their substrates/products (*i.e.*, gluconolactone and protocatechuate, respectively) are UV-vis active, which affects the optical analysis for monitoring [4Fe-4S] cluster assembly. Recently, bilirubin oxidase has been used as a single-enzyme O_2-scavenger without H_2O_2 generation (Monteiro et al. 2022). Unfortunately, bilirubin (red) and the oxidized product biliverdin (green) also affect the optical analysis of the [4Fe-4S] clusters. Interestingly, the NADH-dependent FRE required for $FADH_2$ regeneration in the SUF system can also scavenge O_2 to form H_2O_2 (Gaudu et al. 1994). Working with CAT, the translucent $FADH_2$ can sequentially reduce O_2 to H_2O *via* H_2O_2 (O_2 + 2 $FADH_2$ → 2 H_2O + 2 FAD). Therefore, I employed FRE and CAT for O_2-scavenging. The NADH required for $FADH_2$ regeneration was supplied by the NAD^+-dependent FDH using formate as the donor of reducing equivalent (Batyrova et al. 2020) (**Figure 3-9A**). Accordingly, I tested if this bifunctional three-enzyme cascade can function in the buffer system of the PURExpress and observed time-course NAD^+ reduction ($\lambda_{340\ nm}$) when both formate and FDH were added to the PURExpress buffer (**Figure 3-9BC**). The addition of both FRE and FAD to the PURExpress buffer resulted in a time-course FAD reduction ($\lambda_{445\ nm}$) (**Figures 3-9DE**), and the

yellow color of the PURExpress buffer gradually faded away to become translucent. The same reaction mixture with CAT addition demonstrated a delayed FAD reduction compared to the CAT-free mixture, suggesting $FADH_2$ was recycled back to FAD. Likely, the produced $FADH_2$ was mainly consumed for O_2 degradation in the CAT-added PURExpress buffer in the first 8 minutes. Together, these data suggested that the O_2-scavenging enzyme cascade can sustain continuous $FADH_2$ regeneration in the PURExpress buffer system under aerobic conditions.

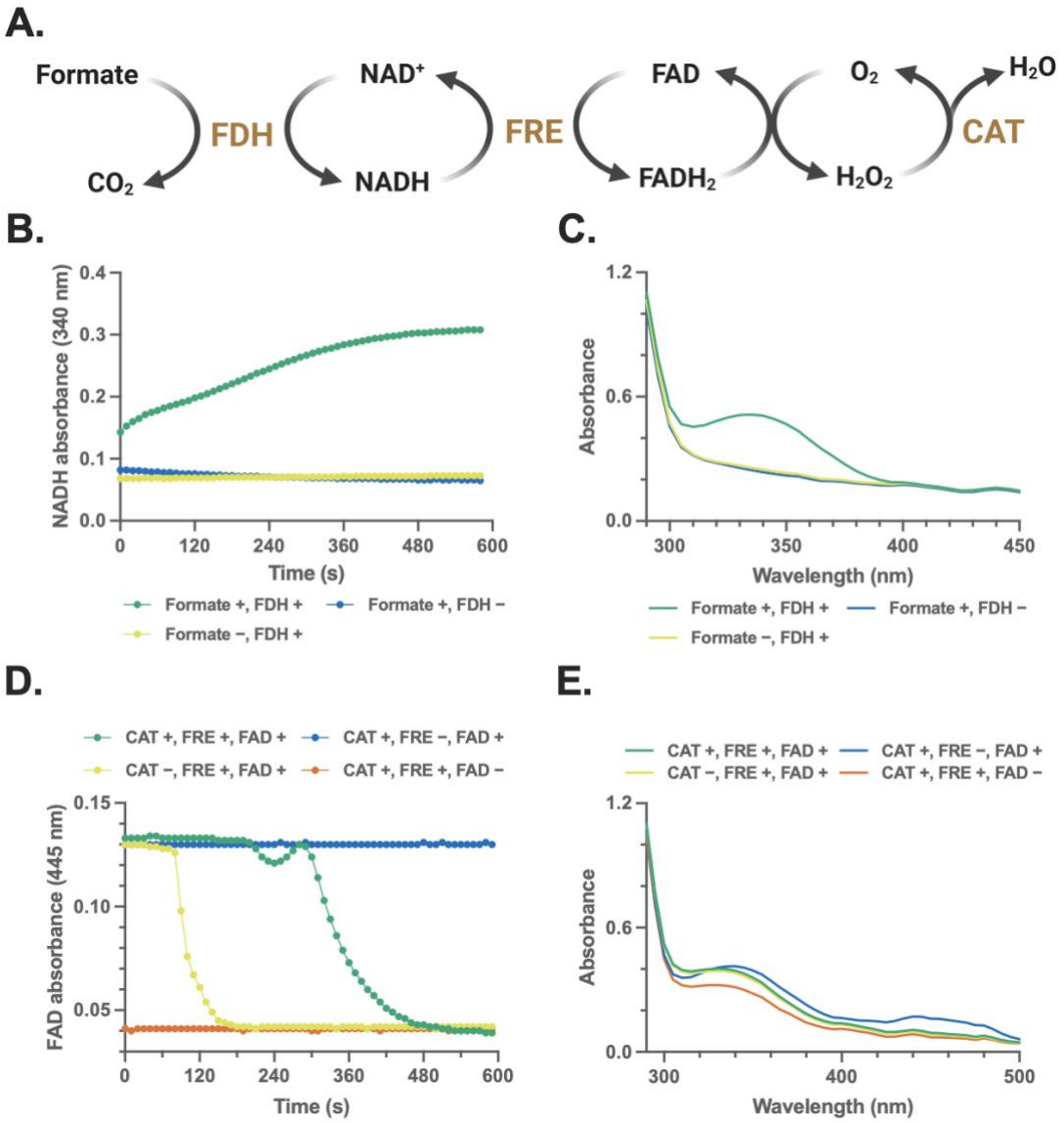


Figure 3-9. The bifunctional enzyme cascade for $FADH_2$ regeneration and O_2 scavenge. (A) Schematic diagram showing the bifunctional O_2-scavenging enzyme cascade. (B) The production of NADH was monitored for 10 min using the absorbance at $\lambda_{340\ nm}$ with formate and formate dehydrogenase (FDH) supplemented PURExfrex buffer. (C) The eaction mixtures

were subjected to UV-vis analysis (300–450 nm). (D) $FADH_2$ production was monitored by a time-dependent decrease in $\lambda_{445\ nm}$ (0–10 min). The reactions were conducted with and without catalase (CAT), flavin reductase (FRE), or FAD. (E) The reaction mixtures were subjected to UV-vis analysis (300–450 nm). The absorbances of NADH $\lambda_{340\ nm}$ and $FADH_2$ $\lambda_{445\ nm}$ were measured after 10 min incubation.

Next, I employed resorufin, a sensitive fluorescent O_2 indicator, to assess the O_2-scavenging capability of the O_2-scavenging cascade. Under reduced conditions (*i.e.*, dissolved O_2 removed), the pink-colored and highly fluorescent resorufin would undergo a reversible reduction to form the non-fluorescent dihydroresorufin (standard reduction potential approximately −51 mV) (Guerin et al. 2001). Consistent with O_2 consumption, the PURExpress buffer with the O_2-scavenging enzyme cascade demonstrated a fluorescence emission ($\lambda_{590\ nm}$) much lower than reactions devoid of either FRE or FAD (**Figure 3-10**). Moreover, the fluorescence emission of the PURExpress buffer amended with the O_2-scavenging enzyme cascade is comparable to those of the titanium citrate-reduced reactions. These data suggested a faster O_2 scavenge rate by the enzyme cascade than the O_2 diffusion rate into the PURExpress buffer. Therefore, I amended this O_2-scavenging enzyme cascade to the SUF-incorporated PURExpress buffer and performed the SUF-mediated [4Fe-4S] assembly *in vitro* under aerobic conditions. After 30 minutes of incubation, the color of SUF-incorporated PURExpress buffer turned brown, along with an obvious increase in the characteristic optical absorbance of [4Fe-4S] clusters at $\lambda_{420\ nm}$, comparable to that of the SUF-incorporated PURExpress buffer incubated anaerobically (**Figure 3-11A**). Furthermore, the O_2-scavenging enzyme cascade also protected the [4Fe-4S] cluster from oxidation under aerobic conditions for at least 1.5 hours, while the absence of the enzyme cascade resulted in a rapid oxidation of the [4Fe-4S] cluster in the first 30 min (**Figure 3-11B**).

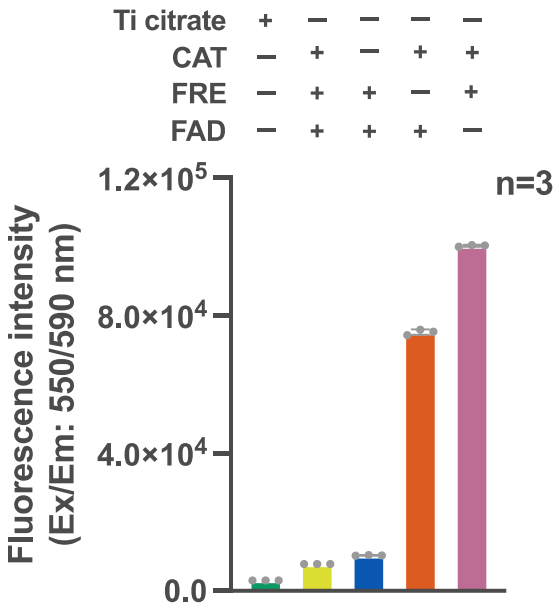


Figure 3-10. The evaluation of anaerobicity in the simulated PURExpress buffer containing the O_2-scavenging system. The anaerobicity of the reaction mixture was evaluated by an oxidation-reduction indicator, resorufin. The bars represent the standard deviation. The dots and columns represent the individual value and mean from three independent experimental replicates, respectively. CAT, catalase; FRE, flavin reductase; Ti citrate, titanium citrate.

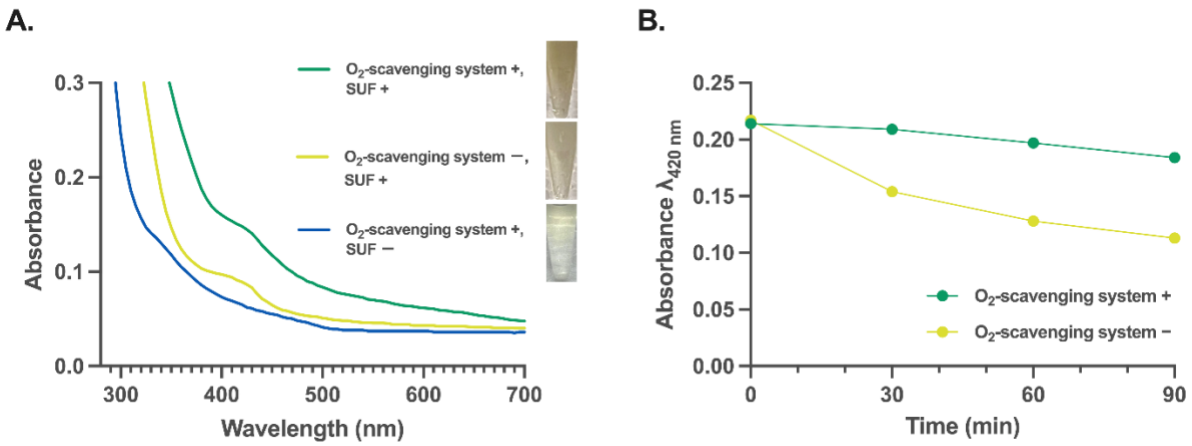


Figure 3-11. The bifunctional O_2-scavenging enzyme cascade enables SUF-mediated [4Fe-4S] cluster assembly under aerobic conditions. (A) The SUF-mediated [4Fe-4S] assembly (based on the increase in $\lambda_{420\ nm}$) with and without the O_2-scavenging system under aerobic conditions. (B) The bifunctional O_2-scavenging enzyme cascade protects the [4Fe-4S] cluster from oxidation (based on the decrease in $\lambda_{420\ nm}$) under aerobic conditions. The degradation and formation of [4Fe-4S] clusters were monitored by the UV-vis spectrophotometric analysis (280–700 nm).

The combination of the O_2-scavenging enzyme cascade and the SUF system enabled *in vitro* [4Fe-4S] cluster reconstitution under aerobic conditions. In *E. coli*, FRE is the main enzyme responsible for $FMNH_2$/$FADH_2$ regeneration, *i.e.*, the native enzyme supporting the [4Fe-4S] cluster assembly by the $SufBC_2D$ complex (Wollers et al. 2010b; Gaudu et al. 1994). Moreover, my O_2-scavenging system only employs a low FAD concentration (~20 μM) comparable to the intracellular FMN/FAD concentration in *E. coli* (Kimata et al. 2018). One can envisage that such a biomimetic system for [4Fe-4S] protein maturation would be suitable and compatible to reconstruct the cell-wide metabolic networks in both cell extract-based and reconstituted cell-free systems. Going further, the required O_2-scavenging capacity, which is proportional to the formate concentration, can be estimated based on the ideal gas law (pv = nRT), atmospheric O_2 content (20%), and Henry's law constant of O_2 ($H_{aq/gas}$ = 0.032). For example, since my PURExpress reactions were performed in a capped 0.3-mL reaction tube with 0.1 mL of the headspace of the atmosphere, the amount of O_2 in the PURExpress reactions would be ~0.9 μmol/tube. Given that the O_2-scavenging enzyme cascade reduces each O_2 molecule at the cost of two $FADH_2$, complete O_2 removal would require ~1.8 μmol (9 mM) of $FADH_2$ (*i.e.*, formate). Excessive formate addition to the system, while securing the anaerobicity, would result in a drastic pH shift.

3.3.3 One-pot, two-step cell-free synthesis of active aconitase under aerobic conditions.

Finally, I tested the possibility of combining the processes of mRNA translation, O_2 scavenge, and [4Fe-4S] cluster maturation into a one-pot process. Due to the adverse effects of S^{2-} (metal precipitation) and Fe^{2+} (in-line cleavage of RNA and generation of reactive oxygen species (Guth-Metzler et al. 2020)) during cell-free translation, I separated the *in vitro*

translation and cluster reconstruction process into two steps (**Figure 3-12A**). In step 1, AcnA mRNA was added to the PUREfrex along with the three enzymes for O_2-scavenging (FDH/FRE/CAT) to maintain anoxic conditions while producing apo-AcnA. In step 2, $FeSO_4$ and SufABCDSE were added to the reactions for the maturation of the apo-AcnA. This two-step process can prevent continuous cysteine desulfurization by SufES that would result in (***i***) a decrease of available cysteine for mRNA translation and (***ii***) metal (*e.g.*, Fe^{2+}) precipitation by the S^{2-}.

For a successful one-pot synthesis, a prerequisite is the compatibility of the PUREfrex with the O_2-scavenging enzyme cascade. The incompatibility of the two components can lead to a significant reduction of the protein yield. Therefore, I first used the mRNA of sfGFP as an indicator of the protein yield in the PUREfrex and compared the sfGFP yield of the PUREfrex reactions with and without the O_2-scavenging enzyme cascade, respectively. My data showed that the incorporation of the O_2-scavenging enzyme cascade to the PUREfrex only resulted in a reduction of sfGFP yield <5% (**Figure 3-12B**), suggesting that the two components are compatible. Another prerequisite is the feasibility of the SUF system to mature the apo-[4Fe-4S] proteins in the O_2-scavenged PUREfrex under aerobic conditions. After establishing successful protein synthesis in the presence of the O_2-scavenging enzyme cascade, I incubated purified apo-AcnA (produced by the PUREfrex) with both the SUF system and the O_2-scavenging enzyme cascade under aerobic conditions. Use of an HRV 3C protease for N-His_6-tag removal allowed me to solely purify holo-AcnA from other recombinant proteins in the reaction mixture (**Figure 3-12C**). The purified holo-AcnA demonstrated a relative enzyme activity of ~95% to that of the fully chemically matured holo-aconitase activity (assuming that

the N-His_6-tag did not alter the AcnA activity) (**Figure 3-12D**). Moreover, ESI-LC-MS analysis revealed concurring citrate consumption and alpha-ketoglutarate production in the assays with the AcnA purified from the one-pot, two-step system but not in the assays with the AcnA purified from the PURExpress reactions devoid of the SUF system (**Figures 3-13 and 3-14**). Interestingly, the AcnA produced from the SUF-amended PURExpress reactions without the O_2-scavenging enzyme cascade demonstrated a relative enzyme activity of ~30% of the AcnA activity (**Figure 3-12D**), suggesting that, consistent with the observation of previous studies, the SUF system is relatively O_2-tolerant. Altogether, these data suggested the successful production of holo-AcnA from the one-pot, two-step process.

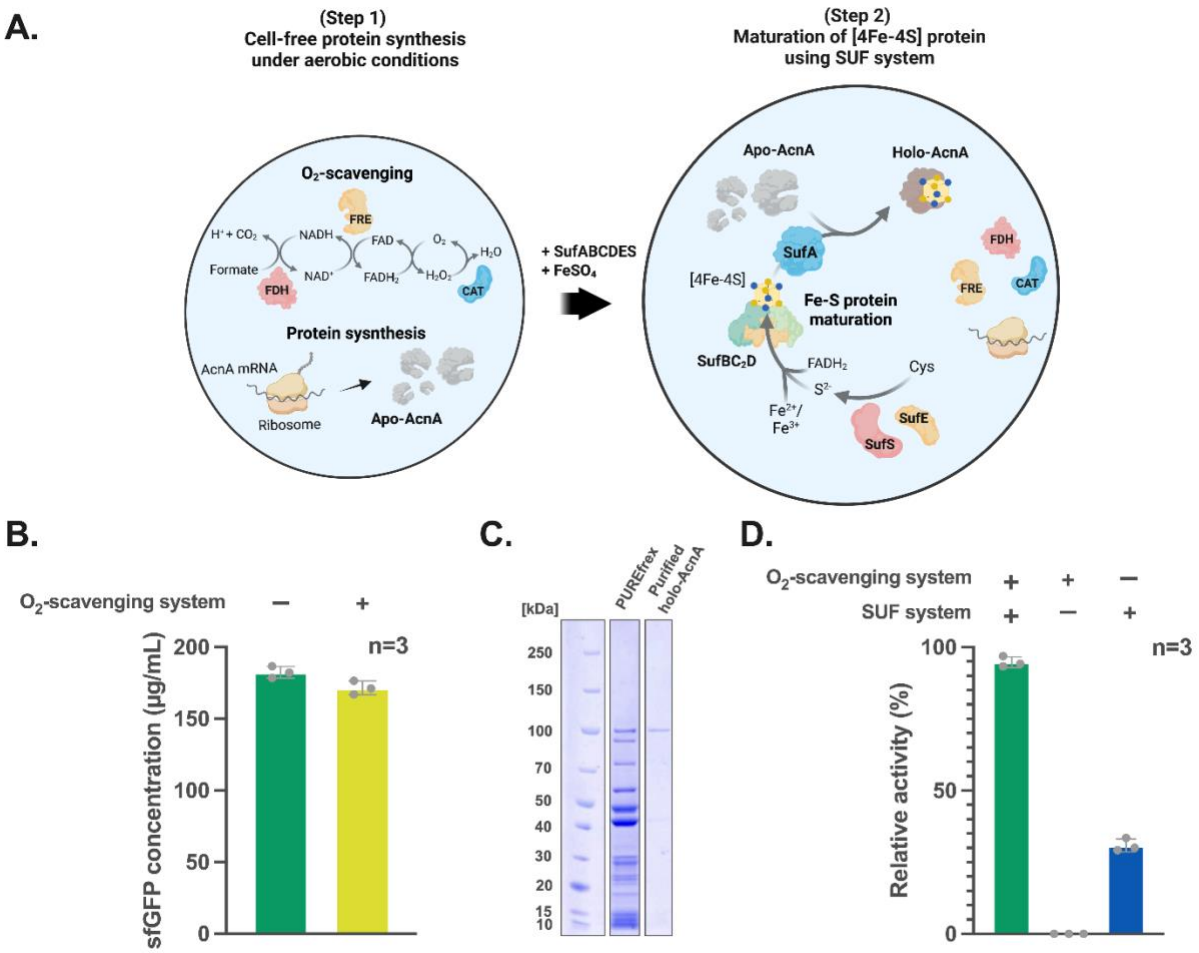
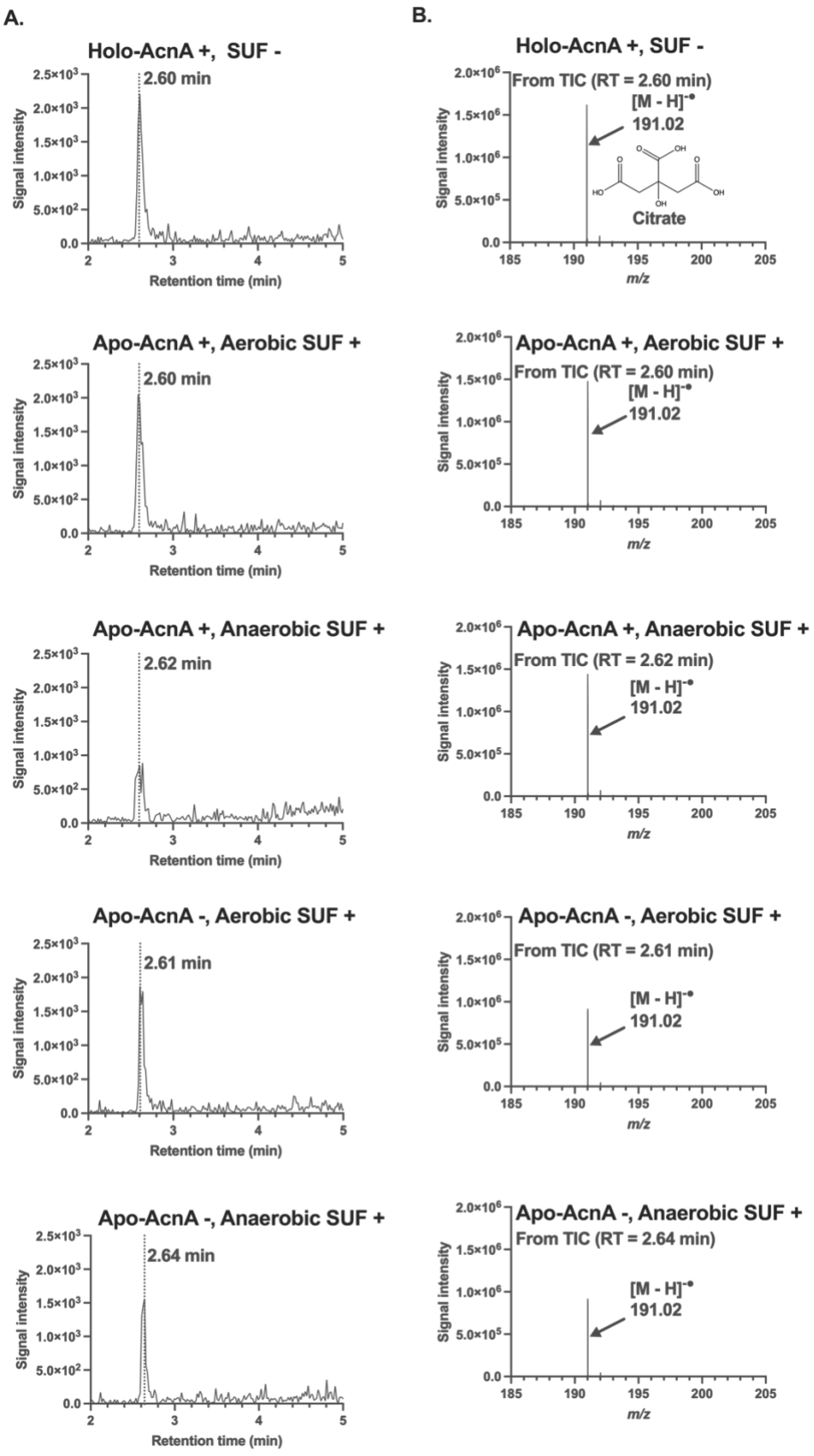


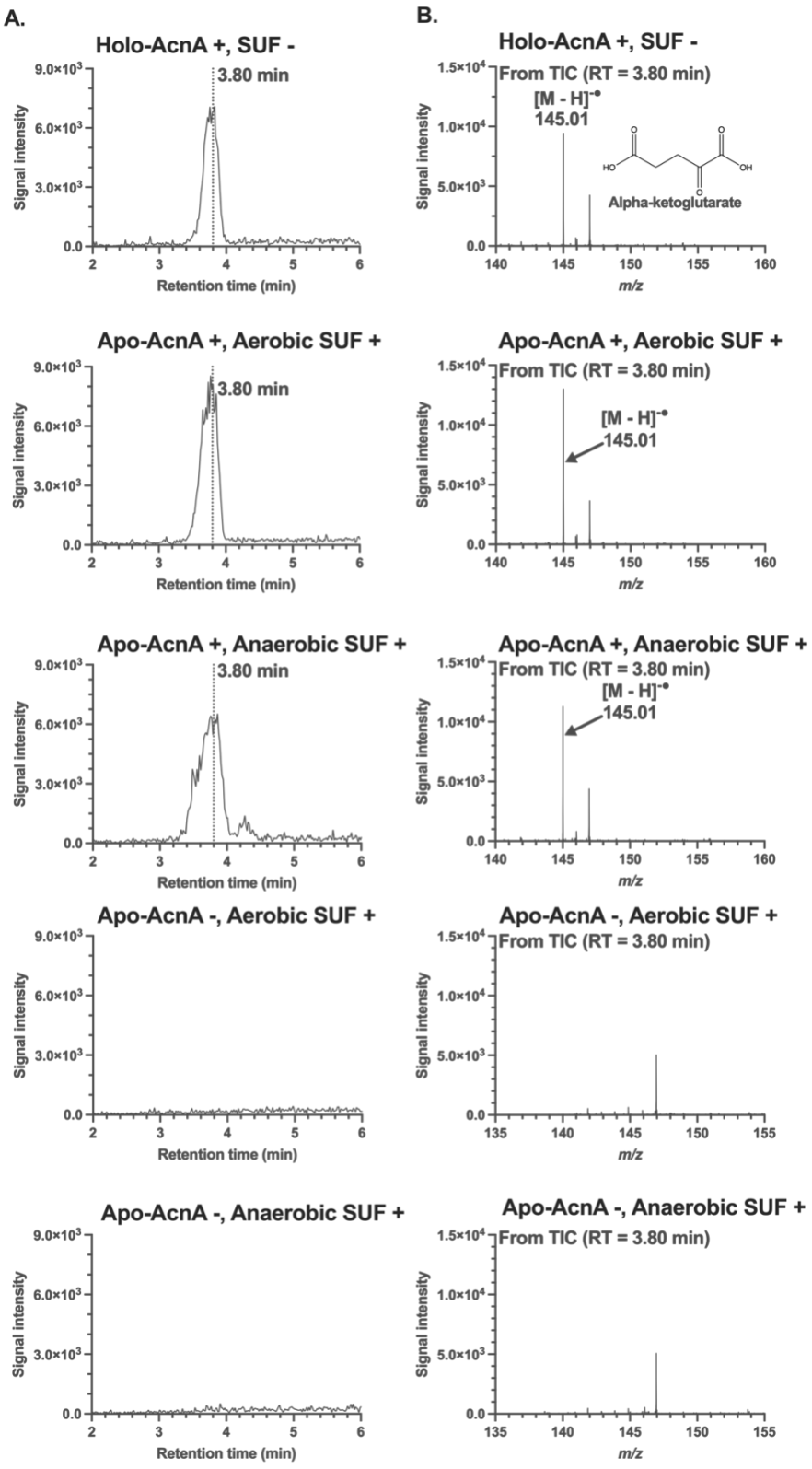
Figure 3-12. One-Pot *de novo* synthesis of mature [4Fe-4S] protein under aerobic conditions. (A) Schematic diagram showing one-pot two-step cell-free synthesis of mature [4Fe-4S] cluster bearing aconitase A(AcnA) under aerobic conditions. *In vitro* translation of AcnA mRNA was carried out in a test tube along with the O_2-scavanging enzyme cascade (step 1). $FeSO_4$ and SUF components are further added to drive maturation of the apo-AcnA to its holo-form (step 2). Other required components such as cysteine (Cys) and ATP are carried over from the PURExpress reaction. Catalase, CAT; FDH, formate dehydrogenase. (B) Superfolder green fluorescent protein (sfGFP) production in the PURExpress reaction mixtures with and without the O_2-scavenging enzyme cascade after 6 h of incubation. (C) Purification of holo-

AcnA free of N-His_6-tag verified by SDS-PAGE analysis. (D) Relative activity of AcnA in the one-pot PURE system with and without the SUF and O_2-scavanging system. The bars represent the standard deviations. The dots and columns represent the individual values and mean from three independent experimental replicates, respectively.



(Legend on next page)

Figure 3-13. ESI-LC-MS spectra of citrate. (A) ESI-LC-MS extracted ion chromatograms and (B) MS spectrum of the completed reaction obtained from the aconitase A (AcnA) and isocitrate dehydrogenase (IDH)-coupled assay. The expected mass of citrate detected in negative ion mode ($[M - H]^{-\bullet}$) is 191.027. Apo-AcnA was produced in the PURExpress and then the reactions were combined with the [4Fe-4S]-charged SUF under aerobic (aerobic SUF) or anaerobic (anaerobic SUF) conditions. The experiment was conducted with and without apo-AcnA. RT, retention time; TIC, total ion current spectrum.



(Legend on next page)

Figure 3-14. ESI-LC-MS spectra of alpha-ketoglutarate. (A) ESI-LC-MS extracted ion chromatograms and (B) MS spectrum of the completed reaction obtained from the aconitase A (AcnA) and isocitrate dehydrogenase (IDH)-coupled assay. The expected mass of alpha-ketoglutarate detected in negative ion mode ($[M - H]^{-\bullet}$) is 145.021. Apo-AcnA was produced in the PURExpress and then the reactions were combined with the [4Fe-4S]-charged SUF under aerobic (aerobic SUF) or anaerobic (anaerobic SUF) conditions. The experiment was conducted with and without apo-AcnA. RT, retention time; TIC, total ion current spectrum.

3.4 Conclusions

Functional expression and purification of [4Fe-4S] proteins have been a long-term challenge in biochemical research. In this study, I demonstrated *in vitro* assembly and transfer of [4Fe-4S] clusters from $Fe^{2+/3+}$ and S^{2-} from cysteine using the SUF system reconstituted from individually purified SUF subunits. The O_2 tolerance of the SUF helper proteins allows facile heterologous expression and purification of each subunit under aerobic conditions and later reconstitution. Interestingly, based on the results of the *Thermococcus* Fd-mediated Cyt C reduction assays, the *E. coli* SUF system seems to be capable of transferring the [4Fe-4S] clusters to [4Fe-4S] proteins from other organisms, suggesting broad applicability of the reconstituted SUF system. In conclusion, the PURExpress supplied with both the SUF system and the bifunctional O_2-scavenging enzyme cascade serves as a facile and efficient platform for synthesizing O_2-labile [4Fe-4S] proteins. I anticipate that this renovated PURE system, with the flexibility to be paired with other helper proteins, would facilitate future studies on more challenging [4Fe-4S] proteins, such as FeMoco-dependent nitrogenase, cobalamin-dependent dehalogenases, and NiFeS acetyl-CoA synthase/CO dehydrogenase. Furthermore, although not fully demonstrated in this study, the amendment of the O_2-scavenging enzyme cascade also makes the PURE system redox-active. If paired with FPR, the four-enzyme cascade can support the regeneration of all common redox equivalents in cells (Fd, NADPH, Cyt C, ubiquinone, and $FMNH_2$/$FADH_2$). Therefore, I also foresee the compatibility of this renovated PURE system with *in vitro* reconstitution of complex biosynthesis pathways and respiratory networks, likely to the extent of *de novo* synthetic cell reconstruction.

Reference

1. Aitken, Colin Echeverría, R. Andrew Marshall, and Joseph D. Puglisi. 2008. “An oxygen scavenging system for improvement of dye stability in single-,olecule fluorescence experiments.” *Biophysical Journal* 94 (5): 1826–35.

2. Aliverti, Alessandro, Vittorio Pandini, Andrea Pennati, Matteo de Rosa, and Giuliana Zanetti. 2008. “Structural and functional diversity of ferredoxin-NADP(+) reductases.” *Archives of Biochemistry and Biophysics* 474 (2): 283–91.

3. Bak, Daniel W., and Sean J. Elliott. 2014. “Alternative FeS cluster ligands: tuning redox potentials and chemistry.” *Current Opinion in Chemical Biology* 19 (April): 50–58.

4. Batyrova, Khorcheska A., Anna N. Khusnutdinova, Po-Hsiang Wang, Rosa Di Leo, Robert Flick, Elizabeth A. Edwards, Alexei Savchenko, and Alexander F. Yakunin. 2020. “Biocatalytic *in vitro* and *in vivo* FMN prenylation and (de)carboxylase activation.” *ACS Chemical Biology* 15 (7): 1874–82.

5. Beinert, H., R. H. Holm, and E. Münck. 1997. “Iron-sulfur clusters: Nature’s modular, multipurpose structures.” *Science (New York, N.Y.)* 277 (5326): 653–59.

6. Beinert, Helmut, Mary Claire Kennedy, and C. David Stout. 1996. “Aconitase as iron−sulfur protein, enzyme, and iron-regulatory protein.” *Chemical Reviews* 96 (7): 2335–2374.

7. Boyd, Eric S., Khaleh M. Thomas, Yuyuan Dai, Jeff M. Boyd, and F. Wayne Outten. 2014. “Interplay between oxygen and Fe-S cluster biogenesis: Insights from the Suf pathway.” *Biochemistry* 53 (37): 5834–47.

8. Boyer, Marcus E., James A. Stapleton, Jon M. Kuchenreuther, Chia-Wei Wang, and James R. Swartz. 2008. “Cell-free synthesis and maturation of [FeFe] hydrogenases.” *Biotechnology and Bioengineering* 99 (1): 59–67.

9. Boyer, Marcus E., Chia-Wei Wang, and James R. Swartz. 2006. “Simultaneous expression and maturation of the iron-sulfur protein ferredoxin in a cell-free system.” *Biotechnology and Bioengineering* 94(1):128-38.

10. Dai, Yuyuan, and F. Wayne Outten. 2012. “The *E. coli* SufS–SufE sulfur transfer system is more resistant to oxidative stress than IscS–IscU.” *FEBS Letters* 586 (22): 4016–22.

11. Esquilin-Lebron, Karla, Sarah Dubrac, Frédéric Barras, and Jeffrey M. Boyd. 2021. “Bacterial qpproaches for qssembling iron-sulfur proteins.” *MBio* 12 (6): e0242521.

12. Evans, M. C., B. B. Buchanan, and D. I. Arnon. 1966. “A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium.” *Proceedings of the National Academy of Sciences of the United States of America* 55 (4): 928–34.

13. Garcia, Pierre Simon, Francesca D’Angelo, Sandrine Ollagnier de Choudens, Macha Dussouchaud, Emmanuelle Bouveret, Simonetta Gribaldo, and Frédéric Barras. 2022. “An

early origin of iron-sulfur cluster biosynthesis machineries before Earth oxygenation." *Nature Ecology & Evolution* 6 (10): 1564–72.

14. Gardner, P. R., and I. Fridovich. 1992. "Inactivation-reactivation of aconitase in *Escherichia coli*. A sensitive measure of superoxide radical." *Journal of Biological Chemistry* 267 (13): 8757–63.

15. Gaudu, P., D. Touati, V. Nivière, and M. Fontecave. 1994. "The NAD(P)H:flavin oxidoreductase from *Escherichia coli* as a source of superoxide radicals." *Journal of Biological Chemistry* 269 (11): 8182–88.

16. Guerin, T. F., M. Mondido, B. McClenn, and B. Peasley. 2001. "Application of resazurin for estimating abundance of contaminant-degrading micro-organisms." *Letters in Applied Microbiology* 32 (5): 340–45.

17. Guth-Metzler, Rebecca, Marcus S. Bray, Moran Frenkel-Pinter, Suttipong Suttapitugsakul, Claudia Montllor-Albalate, Jessica C. Bowman, Ronghu Wu, et al. 2020. "Cutting in-line with iron: Ribosomal function and non-oxidative RNA cleavage." *Nucleic Acids Research* 48 (15): 8663–74.

18. Hodek, Ondřej, Lidia Argemi-Muntadas, Adnan Khan, and Thomas Moritz. 2022. "Mixed-mode chromatography-mass spectrometry enables targeted and untargeted screening of carboxylic acids in biological samples." *Analytical Methods* 14 (10): 1015–22.

19. Holliger, Christof, Gert Wohlfarth, and Gabriele Diekert. 1998. "Reductive dechlorination in the energy metabolism of anaerobic bacteria." *FEMS Microbiology Reviews* 22 (5): 383–98.

20. Hong, J., and J. C. Rabinowitz. 1967. "Preparation and properties of clostridial apoferredoxins." *Biochemical and Biophysical Research Communications* 29 (2): 246–52.

21. Imai, T., K. Taguchi, Y. Ogawara, D. Ohmori, F. Yamakura, H. Ikezawa, and A. Urushiyama. 2001. "Characterization and cloning of an extremely thermostable, *Pyrococcus furiosus*-type 4Fe ferredoxin from *Thermococcus profundus*." *Journal of Biochemistry* 130 (5): 649–55.

22. Jacobson, M. R., V. L. Cash, M. C. Weiss, N. F. Laird, W. E. Newton, and D. R. Dean. 1989. "Biochemical and genetic analysis of the NifUSVWZM cluster from *Azotobacter vinelandii*." *Molecular & General Genetics: MGG* 219 (1–2): 49–57.

23. Jewett, Michael C., and James R. Swartz. 2004. "Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis." *Biotechnology and Bioengineering* 86 (1): 19–26.

24. Johnson, Michael K. 1998. "Iron—Sulfur proteins: New roles for old clusters." *Current Opinion in Chemical Biology* 2 (2): 173–81.

25. Jørgensen, Bo Barker, and Sabine Kasten. 2006. "Sulfur cycling and methane oxidation." in *Marine Geochemistry*, edited by Horst D. Schulz and Matthias Zabel, 271–309. Berlin, Heidelberg: Springer Berlin Heidelberg.

26. Kennedy, M. C., L. Mende-Mueller, G. A. Blondin, and H. Beinert. 1992. “Purification and characterization of cytosolic aconitase from beef liver and its relationship to the iron-responsive element binding protein.” *Proceedings of the National Academy of Sciences of the United States of America* 89 (24): 11730–34.

27. Khoroshilova, N., C. Popescu, E. Münck, H. Beinert, and P. J. Kiley. 1997. “Iron-sulfur cluster disassembly in the FNR protein of *Escherichia coli* by O_2: [4Fe-4S] to [2Fe-2S] conversion with loss of biological activity.” *Proceedings of the National Academy of Sciences of the United States of America* 94 (12): 6087–92.

28. Kim, Dong-Myung, and James R. Swartz. 2001. “Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis.” *Biotechnology and Bioengineering* 74 (4): 309-16.

29. Kimata, Shinya, Daichi Mochizuki, Junichi Satoh, Ken Kitano, Yu Kanesaki, Kouji Takeda, Akira Abe, Shinji Kawasaki, and Youichi Niimura. 2018. “Intracellular free flavin and its associated enzymes participate in oxygen and iron metabolism in *Amphibacillus xylanus* lacking a respiratory chain.” *FEBS Open Bio* 8 (6): 947–61.

30. Kitagawa, Masanari, Takeshi Ara, Mohammad Arifuzzaman, Tomoko Ioka-Nakamichi, Eiji Inamoto, Hiromi Toyonaga, and Hirotada Mori. 2005. “Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): Unique resources for biological research.” *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes* 12 (5): 291–99.

31. Kuznetsova, Ekaterina, Michael Proudfoot, Claudio F. Gonzalez, Greg Brown, Marina V. Omelchenko, Ivan Borozan, Liran Carmel, et al. 2006. “Genome-wide analysis of substrate specificities of the *Escherichia coli* haloacid dehalogenase-like phosphatase family.” *Journal of Biological Chemistry* 281 (47): 36149–61.

32. Lavickova, Barbora, Nadanai Laohakunakorn, and Sebastian J. Maerkl. 2020. “A partially self-regenerating synthetic cell.” *Nature Communications* 11 (1): 6340.

33. Lavickova, Barbora, and Sebastian J. Maerkl. 2019. “A simple, robust, and low-cost method to produce the PURE cell-free system.” *ACS Synthetic Biology* 8 (2): 455–62.

34. Layer, Gunhild, S. Aparna Gaddam, Carla N. Ayala-Castro, Sandrine Ollagnier-de Choudens, David Lascoux, Marc Fontecave, and F. Wayne Outten. 2007. “SufE transfers sulfur from SufS to SufB for iron-sulfur cluster assembly.” *Journal of Biological Chemistry* 282 (18): 13342–50.

35. Monteiro, Tiago, Miguel Moreira, Sara B. R. Gaspar, and Maria Gabriela Almeida. 2022. “Bilirubin oxidase as a single enzymatic oxygen scavenger for the development of reductase-based biosensors in the open air and its application on a nitrite biosensor.” *Biosensors & Bioelectronics* 217 (December): 114720.

36. Motohashi, Ken. 2015. “A simple and efficient seamless DNA cloning method using SLiCE from *Escherichia coli* laboratory strains and its application to SLiP site-directed mutagenesis.” *BMC Biotechnology* 15 (June): 47.

37. ———. 2017. “Seamless ligation cloning extract (SLiCE) method using cell lysates from laboratory *Escherichia coli* strains and its application to SLiP site-directed mutagenesis.” *Methods in Molecular Biology* 1498: 349–57.

38. Niwa, Tatsuya, Takashi Kanamori, Takuya Ueda, and Hideki Taguchi. 2012. “Global analysis of chaperone effects using a reconstituted cell-free translation system.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (23): 8937–42.

39. Odom, J. M., and H. D. Peck Jr. 1984. “Hydrogenase, electron-transfer proteins, and energy coupling in the sulfate-reducing bacteria Desulfovibrio.” *Annual Review of Microbiology* 38: 551–92.

40. Outten, F. Wayne, Matthew J. Wood, F. Michael Munoz, and Gisela Storz. 2003. “The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in *Escherichia coli*.” *Journal of Biological Chemistry* 278 (46): 45713–19.

41. Patil, P. V., and D. P. Ballou. 2000. “The use of protocatechuate dioxygenase for maintaining anaerobic conditions in biochemical experiments.” *Analytical Biochemistry* 286 (2): 187–92.

42. Petering, D., J. A. Fee, and G. Palmer. 1971. “The oxygen sensitivity of spinach ferredoxin and other iron-sulfur proteins. The formation of protein-bound sulfur-zero.” *Journal of Biological Chemistry* 246 (3): 643–53.

43. Peters, J. W., K. Fisher, W. E. Newton, and D. R. Dean. 1995. “Involvement of the P cluster in intramolecular electron transfer within the nitrogenase MoFe protein.” *Journal of Biological Chemistry* 270 (45): 27007–13.

44. Ragsdale, Stephen W. 2003. “Pyruvate ferredoxin oxidoreductase and its radical intermediate.” *Chemical Reviews* 103 (6): 2333–46.

45. Sanden, Sebastian A., Ruiqin Yi, Masahiko Hara, and Shawn E. McGlynn. 2020. “Simultaneous synthesis of thioesters and iron-sulfur clusters in water: Two universal components of energy metabolism.” *Chemical Communications (Cambridge, England)* 56 (80): 11989–92.

46. Shimizu, Y., A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda. 2001. “Cell-free translation reconstituted with purified components.” *Nature Biotechnology* 19 (8): 751–55.

47. Stephens, P. J., D. R. Jollie, and A. Warshel. 1996. “Protein control of redox potentials of iron−sulfur proteins.” *Chemical Reviews* 96 (7): 2491–2514.

48. Swartz, James R., Michael C. Jewett, and Kim A. Woodrow. 2004. “Cell-free protein synthesis with prokaryotic combined transcription-translation.” *Methods in Molecular Biology* 267: 169–82.

49. Takahashi, Yasuhiro, and Umechiyo Tokumoto. 2002. “A third bacterial system for the assembly of iron-sulfur clusters with homologs in archaea and plastids.” *Journal of Biological Chemistry* 277 (32): 28380–83.

50. Tsibris, J. C., R. L. Tsai, I. C. Gunsalus, W. H. Orme-Johnson, R. E. Hansen, and H. Beinert. 1968. “The number of iron atoms in the paramagnetic center (G =1.94) of reduced putidaredoxin, a nonheme iron protein.” *Proceedings of the National Academy of Sciences of the United States of America* 59 (3): 959–65.

51. Urbina, H. D., J. J. Silberg, K. G. Hoff, and L. E. Vickery. 2001. “Transfer of sulfur from IscS to IscU during Fe/S cluster assembly.” *Journal of Biological Chemistry* 276 (48): 44521–26.

52. Wan, Jason T., and Joseph T. Jarrett. 2002. “Electron acceptor specificity of ferredoxin (flavodoxin):$NADP^+$ oxidoreductase from *Escherichia coli*.” *Archives of Biochemistry and Biophysics* 406 (1): 116–26.

53. Wang, Po-Hsiang, Kosuke Fujishima, Samuel Berhanu, Yutetsu Kuruma, Tony Z. Jia, Anna N. Khusnutdinova, Alexander F. Yakunin, and Shawn E. McGlynn. 2020. “A bifunctional polyphosphate kinase driving the regeneration of nucleoside triphosphate and reconstituted cell-free protein synthesis.” *ACS Synthetic Biology* 9 (1): 36–42.

54. Wollers, Silke, Gunhild Layer, Ricardo Garcia-Serres, Luca Signor, Martin Clemancey, Jean-Marc Latour, Marc Fontecave, and Sandrine Ollagnier de Choudens. 2010. “Iron-sulfur (Fe-S) cluster assembly: The SufBCD complex is a new type of Fe-S scaffold with a flavin redox cofactor.” *Journal of Biological Chemistry* 285 (30): 23331–41.

55. Yakunin, A. F., and P. C. Hallenbeck. 1998. “Purification and characterization of pyruvate oxidoreductase from the photosynthetic bacterium *Rhodobacter capsulatus*.” *Biochimica et Biophysica Acta* 1409 (1): 39–49.

56. Yankovskaya, Victoria, Rob Horsefield, Susanna Törnroth, César Luna-Chavez, Hideto Miyoshi, Christophe Léger, Bernadette Byrne, Gary Cecchini, and So Iwata. 2003. “Architecture of succinate dehydrogenase and reactive oxygen species generation.” *Science* 299 (5607): 700–704.

57. Zhang, Z., L. Huang, V. M. Shulmeister, Y. I. Chi, K. K. Kim, L. W. Hung, A. R. Crofts, E. A. Berry, and S. H. Kim. 1998. “Electron transfer by domain movement in cytochrome Bc1.” *Nature* 392 (6677): 677–84.

58. Zheng, L., V. L. Cash, D. H. Flint, and D. R. Dean. 1998. “Assembly of iron-sulfur clusters. Identification of an IscSUA-HscBA-Fdx gene cluster from *Azotobacter vinelandii*.” *Journal of Biological Chemistry* 273 (21): 13264–72.

Chapter 4 – Summary, significance, and future perspectives

4.1 Summary

The major goal of this thesis was to ***(i)*** develop a renovated mRNA display method for a high throughput identification of *de novo* RNA-binding peptides and ***(ii)*** open a new field in the cell-free system to handle post-translational modification enzymes, especially those employing [4Fe-4S] cluster for the catalysis. In addressing these goals, two objectives were proposed and pursued (reported in **Chapter 2 and 3**). The major conclusions from each Chapter are summarized below.

Objective 1: Experimentally demonstrating the validity of codon-restricted mRNA display for the identification of *de novo* single-stranded RNA-binding peptides.

mRNA display is an established method for high throughput identification of *de novo* functional peptides. However, mRNA display cannot be applied to the screening of RNA-binding peptides due to the formation of base-pairing between a target RNA and mRNA portion originating from mRNA-peptide conjugates. In this thesis, a renovated mRNA display was developed, which employs codon-restricted DNA libraries to mitigate the base-pairing problem, thus enabling the identification of the *de novo* ssRNA-binding peptides, one of which exhibits affinity to oncogenic RNA motifs at submicromolar concentration range. The renovated mRNA display can serve as a versatile tool for screening RNA-binding peptides with numerous RNA targets.

Objective 2: Developing a method to produce O_2-labile [4Fe-4S] cluster proteins using the PURE system under aerobic conditions, especially for future applications in *in vitro* evolution techniques.

A downside of Objective 1 was the limited chemical diversity of the peptide library due to the codon-restriction. One solution to expand the chemical diversity is to incorporate non-canonical amino acids into the peptides encoded in the DNA library. However, traditional genetic code engineering techniques require pre-aminoacylated tRNA molecules, thus the variety of non-canonical amino acids for the incorporation highly depends on the compatibility with ribosome. Alternatively, I found a possibility to introduce non-proteinogenic structure into peptide library *via* post-translational modification enzymes, especially, [4Fe-4S] cluster-containing rSAM enzymes. The metallocofactor, [4Fe-4S] cluster serves as an electron shuttle to catalyze numerous radical reactions along with SAM, which leads to the various types of peptide modifications. However, the O_2-lability of the cluster has hindered its application in *in vitro* evolution experiments. In this thesis, the Fe-S cluster assembly system (*i.e.*, SUF system) was reconstituted *in vitro* and employed to produce [4Fe-4S] cluster proteins (Fd and AcnA) in the PURE system. To produce [4Fe-4S] cluster proteins under aerobic conditions, a synthetic O_2-scavenging enzyme cascade was developed, enabling one-pot, two-step synthesis of [4Fe-4S] cluster proteins under aerobic conditions.

4.2 Significance and future perspectives

The two projects present in this thesis aim to expand the application of the PURE system. The two methods developed in this thesis are particularly relevant to ***(i)*** the development of *de novo* RNA-binding peptide, ***(ii)*** high-throughput functional characterization of redox-active proteins, and ***(iii)*** the future applications in the generation of a RiPP library, which allows exploration of the natural product chemical space in a cell-free context.

Many bioactive natural products are produced by microorganisms that thrive in anaerobic environments. The development of cell-free systems capable of functioning anaerobically and producing O_2-sensitive enzymes is an emerging area of exploration. Access to these enzymes has the potential to unlock a wide range of unique and valuable chemical reactions. Particularly, [4Fe-4S] cluster-containing rSAM enzymes are versatile post-translational modification enzymes. As described in **Chapter 1**, rSAM enzymes catalyze various types of post-translational modification while the majority of predicted rSAM enzymes remain functionally uncharacterized. By harnessing the method developed in **Chapter 3**, one can produce rSAM enzymes within four hours under aerobic conditions (unpublished data). This enables a high-throughput characterization of rSAM enzymes with unknown functions, leading to the discovery of a novel type of post-translational modifications and further expanding the repertoire of cell-free biochemical reactions. Furthermore, by combining the characterized rSAM enzymes with the codon-restricted mRNA display developed in **Chapter 2**, one can generate a RiPP library as well as perform peptide screening against rRNA, which most of the antibiotics are known to target.

My techniques developed in this thesis can also contribute to improving cell-free system by introducing tRNA modification. Previously, Hibi et al. demonstrated the *in vitro* transcribed tRNA can be used for cell-free translation, but they found that the protein yield is lower than with the modified tRNA (Hibi et al. 2020). In *E. coli*, 30 modifications have been found and reported to play a pivotal role in decoding fidelity. Most of the modifications are catalyzed by rSAM enzymes harboring [4Fe-4S] clusters, thus our system allows us to study the relationship between post-transcriptional modification and protein synthesis *in vitro*. In summary, my technology allows a rapid exploration of the vast chemical diversity of peptides resembling natural products *in vitro*. The field holds great promise for expanding access to redox-active enzymes and discovering novel antibiotics in a cell-free context.

Building on my research during the Ph.D. program, my focus has primarily been on leveraging molecular biology, bioinformatics, biochemistry, and enzymology to elucidate RNA-peptide interactions as well as to reconstitute biological systems *in vitro*. I foresee that these techniques and insights will empower me to reconstruct cell-wide metabolic pathways *in vitro*, ultimately leading to the creation of synthetic biological systems (*i.e.*, synthetic cells).

Synthetic cells, often constructed by replicating specific attributes of cellular life, offer a valuable platform for investigating the universal principles governing biological processes at the molecular level. These synthetic systems are simpler than their natural counterparts since these systems are often constructed *via* the assembly of minimal required components, serving as tools to study the cellular metabolic pathways and signal transduction pathway as well as microscale biofoundries for producing value-added products. Synthetic cells comprise various components, including DNA, proteins, and lipids, which are sourced and engineered from

different organisms to attain the desired functionalities. However, a fundamental function of a cellular life, self-regeneration, has yet to be achieved.

As described in **Chapter 1**, although the PURE system enables the operation of the central dogma *in vitro*, unlike cellular life, which couples the central dogma to metabolism-expressing biosynthetic enzymes to regenerate cellular building blocks *via* enzyme-driven metabolism, any synthetic biological system designed with the PURE system entirely relies on an external supply of the internal building blocks (*i.e.*, amino acids and nucleic acids) rather than any coupled metabolic pathways. Therefore, to embed self-regenerating property with the PURE system, *in vitro* reconstitution of the biosynthetic metabolic pathway is necessary. *In vitro* reconstitution of the biosynthetic metabolic pathway will provide significant insights into understanding the behavior of intricate cellular metabolic networks, and when integrated with the PURE system, self-regeneration can be achieved.

As a first step to create a self-regenerating biological system, I aim to integrate the amino acid biosynthesis pathways with the PURE system, allowing it to self-regenerate the building block of protein (*i.e.*, amino acids). So far, I have successfully produced proteins by coupling protein synthesis with an amino acid-regenerating multi-enzyme cascade, generating four amino acids (Asn, Asp, Gln, and Glu) from lactate as the amino acid precursor (unpublished data). To make the PURE system more "life-like" system, I plan to further integrate other amino acid biosynthesis pathways into the PURE system, ultimately enabling the production of polypeptides/enzymes from 20 proteinogenic amino acids *in vitro*. This amino acid-self-regenerating PURE system will serve as a robust platform to synthesize valuable enzymes, therapeutic peptides, and even RiPPs.

In addition to producing protein building blocks, the amino acid synthesis pathways can also be integrated into cell-wide metabolic networks. For example, serine is involved in folate cycles, C1 metabolism, and nucleic acid biosynthesis (Anguera et al. 2006); cysteine is required for the biosynthesis of the Fe-S clusters involved in cellular respiration (Garcia et al. 2022). Reconstituting all 20 amino acid synthesis pathways in the PURE system would enable PURE systems to emulate more complex cellular life forms and allow us to analyze cellular metabolism in a simplified context. In summary, I anticipate that bottom-up reconstruction of self-regenerating synthetic biological systems will allow us to elucidate the fundamental principle of the current biological system as well as serve as an alternative workhorse for protein/peptide production.

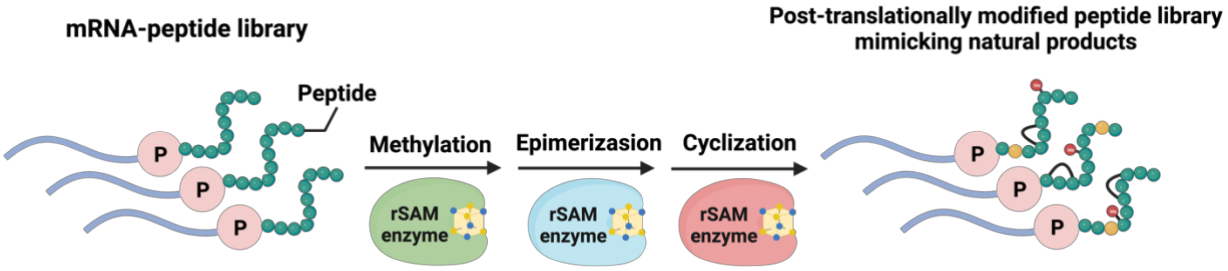


Figure 4-1. Radical S-adenosyl-l-methionine (rSAM) enzyme-mediated post-translationally modified peptide library generation *via* mRNA display.

Reference

1. Anguera, Montserrat C., Martha S. Field, Cheryll Perry, Haifa Ghandour, En-Pei Chiang, Jacob Selhub, Barry Shane, and Patrick J. Stover. 2006. “Regulation of folate-mediated one-carbon metabolism by 10-formyltetrahydrofolate dehydrogenase.” *Journal of Biological Chemistry* 281 (27): 18335–42.

2. Garcia, Pierre Simon, Francesca D’Angelo, Sandrine Ollagnier de Choudens, Macha Dussouchaud, Emmanuelle Bouveret, Simonetta Gribaldo, and Frédéric Barras. 2022. “An early origin of iron-sulfur cluster biosynthesis machineries before Earth oxygenation.” *Nature Ecology & Evolution* 6 (10): 1564–72.

3. Hibi, Keita, Kazuaki Amikura, Naoki Sugiura, Keiko Masuda, Satoshi Ohno, Takashi Yokogawa, Takuya Ueda, and Yoshihiro Shimizu. 2020. “Reconstituted cell-free protein synthesis using *in vitro* transcribed tRNAs.” *Communications Biology* 3 (1): 350.