

# Development of mRNA-protein fusion at N-terminus for evolutionary protein engineering

(進化分子工学での使用を目的とした N 末端側 mRNA-protein 対応付け手法の開発)

2007 年 9 月

埼玉大学大学院理工学研究科（博士後期課程）

生物環境科学専攻（指導教員 伏見譲）

上野 真吾

## 目次

序論	3
Chapter I	
Development of mRNA-protein fusion at N-terminus for evolutionary protein engineering	10
1. Introduction	11
2. Materials and methods	
2.1 Synthesis of the DNA-linker having hydrazide group	14
2.2 Construction of the mRNA-templates	16
2.3 Ligation between the mRNA and the hydrazide-linker	17
2.4 Preparation of amino acid substrate	17
2.5 Preparation of aminoacyl-sup tRNA	18
2.6 Selective enrichment analysis	20
2.7 Detection of the mRNA-peptide fusion	23
3. Results and Discussions	
3.1 Ligation between the mRNA and the hydrazide-linker	24
3.2 Aminoacylation of tRNA and conjugation between hydrazide-linker and aminoacyl-tRNA	29
3.3 Selective enrichment analysis	31
3.4 Detection of the mRNA-peptide fusion	33
3.5 Conclusion	34
References	35

## Chapter II

### Development of a novel combinatorial method

to make high quality peptide libraries 39

1. Introduction 40

2. Materials and methods

2.1 MLSDS 合成 41

2.2 MLSDS 最適化アルゴリズム 44

2.3 配列解析 46

2.4 ライブラリーの長鎖化 46

3. Results and discussions

3.1 ライブラリーの設計及び合成

3.1.1 “MLSDS Natural” 49

3.1.2 “MLSDS NoCys” 56

3.1.3 “MLSDS Natural”, “MLSDS NoCys” 総合 63

3.1.4 他のライブラリー作成法との比較 65

3.2 ライブラリーの長鎖化 66

3.3 Conclusion 67

References 67

総括 70

謝辞 71

## 序論

### 進化分子工学

生物が有する多種多様な生体高分子は、40 億年にもわたる生物の進化に伴ってその高度な機能を獲得してきた。生命科学の誕生以来、多くの研究者がその機能や構造を解明し、目的に応じた機能を有する新規機能性高分子の創出に挑戦している。1980 年代前半、DNA 合成機の発明によって、タンパク質や RNA などの生体高分子の構造情報をコードする DNA を化学合成することが可能となり、新規生体高分子の合成が容易に行えるようになった。しかしながら研究者は、求める機能を有した生体高分子の設計図(DNA)を描くことが出来なかった。タンパク質の立体構造から機能を理解することは出来ても、機能からその立体構造を設計することは未だ難しい。そのような構造論的なアプローチとは別に、進化論的に新規生体高分子を得ようとするものがあり、それを進化分子工学 (Evolutionally Molecular Engineering) という。進化分子工学は、現存する生体高分子が進化によってその機能を得てきたように、実際にその進化を研究室レベルで超高速に行うことによって目的にかなった機能性高分子を得ようとするものである。

ダーウィン進化が起こる十分条件は次の 5 条件にまとめられる。

- ① 非平衡開放系
- ② 自己触媒反応系(自己増殖系)
- ③ 遺伝子型・表現型対応付け戦略を持つ系
- ④ 突然変異系
- ⑤ 配列空間上の適応度地形が適切な系

①と②の2条件が揃うと淘汰が発生する。①は、流通式反応器(flow reactor)や、アフィニティークロマトグラフィ、セルソーターなど分子の流動性がある系、②は細胞の2分裂増殖やPCR、3SR2)など分子数が指数関数的に増加する系で可能となる。

リボザイムやデオキシリボザイムなど、遺伝子型と表現型が同一分子であるものの進化の場合、③の条件は特別な戦略をせずとも満たされている。しかし、タンパク質の進化などのように遺伝子形と表現型が別分子の場合、進化のサイクルをまわす為には表現型による評価を遺伝子型の増幅にフィードバックする必要がある。それが③の条件である。この対応付け戦略は細胞型戦略、ウイルス型戦略、外部知性型戦略の3種に大別できる。

細胞型戦略：遺伝子型分子と表現型分子を1つの囲みに入れる戦略。単細胞生物の細胞膜、クローニング用試験管などがその囲みとなる。混合物として合成されたDNAライブラリー(遺伝子型)を1分子ずつ袋に入れ、その中でタンパク質(表現型)を発現させる。望む機能を有するタンパク質を発現しているものを袋ごとスクリーニングし、袋を壊して遺伝子を回収する。

ウイルス型戦略：天然に存在するウイルスの様に遺伝子型分子と表現型分子を物理的に直接結合させる戦略。遺伝子型と表現型が直接結合している為、各分子を混合した状態でおくことが可能であり空間上の制約が低い。結合様式の違いにより phage display, ribosome display, *in vitro* virus (mRNA display) などが開発されている。

外部知性型戦略：ペプチドライブラリーを、番地付けされたマイクロプレートやチップ、記憶用マイクロチップの入ったビーズ上で合成し、ハイスループットスクリーニング (HTS) などでスクリーニングを行い、スクリーニングされ

たペプチドに対応する番地や記憶情報を介して遺伝子を選択する。主にペプチドのコンビナトリアルケミストリーで採用されている。

①、②で淘汰は起こるが、これだけでは同一分子の濃縮が起こるだけで進化とはいえない、進化するためには④の突然変異が必要である。また、134 塩基長の DNA 配列空間のサイズは  $10^{80}$  となり、宇宙の全原子数と同じ大きさになる。このような配列空間を用意することは不可能であり、初期に存在しなかった配列を探索する為に、突然変異を行う必要がある。

進化分子工学には新規機能性分子の創生という純工学的な側面と同時に、進化そのものの原理の解明にも大きな貢献を果たしている。進化分子工学の実験によって得られたリボザイムやデオキシリボザイムは、生命の起源に「RNA ワールド」があった可能性を示唆している。

本研究は、ウイルス型戦略における、遺伝子型分子と表現型分子の連結手法の開発、及び、効率的な蛋白質進化を可能とする初期 DNA ライブラリー作製法を開発である。

#### 新規遺伝子型-表原型対応付け手法の開発

表現型分子である蛋白質と遺伝子型分子である RNA との対応付けを蛋白質の N 末端側で行うウイルス型戦略手法を新たに開発した。本題に入る前にウイルス型の対応化技術について概説する。ウイルス型の遺伝子型 - 表現型対応化技術には、phage display, ribosome display, *in vitro* virus (mRNA display) などがある。phage display は、遺伝子型である DNA をその表現型であるコートタ

ンパクが包み込んでいるファージの形態を利用したものであり、DNA ライブラリーをコートタンパク質遺伝子に挿入し、表現型をコートタンパクとしてファージ粒子表面に提示 (display) することによって遺伝子型と表現型の対応付けを行っている。この戦略はすでに多くの成果を上げているが、対応付けを *in vivo* で行う為、形質転換効率の限界によりライブラリーのサイズが制限される(〜10<sup>9</sup>)、宿主細胞にとって有害なタンパク質は提示できない、培養に時間が掛かるなどの欠点を有する。これらの問題を克服する為に *E.coli* S30 画分、小麦胚芽、ウサギ網状赤血球などから抽出された無細胞 (*in vitro*) 翻訳系を用いる対応化技術が開発されており、それが次に述べる ribosome display と *in vitro* virus である。*in vitro* では生態的制約を受けないため *in vivo* の場合と比べ、さまざまな利点がある。まず、細胞を使わないことによる空間的自由度からライブラリーサイズを大幅に増大することが出来る(〜10<sup>14</sup>)。他に、細胞内では発現しにくいタンパク質や細胞に有害なタンパク質の合成、特殊な環境状態の設定、非天然アミノ酸の使用が可能であり、細胞培養を伴わないことによって進化のサイクルを短縮(1 日単位)することも可能である。

ribosome display は、ストップコドンを除いた mRNA を用いて翻訳されたタンパク質をリボソーム上でトラップさせ、この結果生じる「タンパク質 - リボソーム - mRNA 複合体」という形で対応付けを行っている。対応付けがリボソームを介した結合の為不安定であり、リボソームという巨大分子が付随している為スクリーニング条件も制限される。

*in vitro* virus は、タンパク質とそれをコードする遺伝子を共有結合したものであり、遺伝子型と表現型のリンカーとして抗生物質のピューロマイシンを用いているものが現在の主流である。このピューロマイシンをリンカーとして用いる対応化技術は、そのウイルス粒子状の形態から、宿主を試験管(生体外)とす

るウイルスとして *in vitro virus* と名付けられた。ピューロマイシンは原核、及び真核細胞のタンパク質合成を阻害する抗生物質である。ピューロマイシンの構造は Tyr-tRNA の 3'末端と酷似しており、アミノアシル tRNA と競合しリボソーム A 部位に結合する。ついで、正常なペプチド転移でペプチジル - ピューロマイシンが生成する。しかし、ピューロマイシンのアミノ残基部位と tRNA 部位の結合がエステル結合ではなくアミド結合である為、リボソームは次のペプチド転移を行えず、ポリペプチド鎖合成を終結し、ペプチジル - ピューロマイシンを遊離する。ここで、ピューロマイシンと遺伝子型(mRNA)が結合されていると遺伝子型(mRNA)と表現型(タンパク質)の対応付けが為される。

これら ribosome display や *in vitro virus* 等、蛋白質の *in vitro* selection 手法の殆どは提示する蛋白質の C 末端側に遺伝子型分子を結合させる手法である。従って機能部位が C 末端側に存在する、若しくは、フリーの C 末端を機能上要求する蛋白質の進化には向いていない。

そこで sup tRNA を利用した蛋白質への非天然アミノ酸導入法にヒントを得て、蛋白質の N 末端側に遺伝子型分子を結合させる新しい手法を開発した。mRNA の開始コドンから数コドン下流に終止コドンを配置し、アミノアシル化した sup tRNA のアミノ酸側鎖を、リンカーを介して mRNA の 5'末端に結合させる。このコンストラクトを無細胞翻訳系に投入することにより、mRNA と連結したアミノアシル sup tRNA が 5'側の終止コドンで取り込まれ、蛋白質の N 末端側と mRNA の 5'末端が連結される。

His-Tag 配列を含む上記 mRNA コンストラクトを *E.coli* S30 extract に投入し、その翻訳産物を Ni-NTA でスクリーニングし、RT-PCR を行った結果、ネガティブコントロールとして Met 無し、及びアミノアシル sup tRNA を連結していない RNA 分子で翻訳したサンプルのスクリーニング産物と比較して有意に増



幅率が高い結果が得られ、スクリーニングサイクルを 1 サイクル回すことに成功した。また、6xHis-tag 配列を含む分子と、FLAG-tag 配列を含む分子の混合物から、Ni-NTA を用いて、6xHis-tag 配列を含む分子の濃縮に成功した。同様に、anti FLAG M2 抗体を用いて FLAG-tag 配列を含む分子の濃縮にも成功した。そして、FLAG-tag 配列を含む mRNA コンストラクトを PURESYSTEM 無細胞翻訳系に投入し、その翻訳産物をウェスタンブロッティングで解析したところ、mRNA-Peptide 複合体の形成を確認した。

本手法によって、フリーの C 末端を機能上要求する蛋白質の進化が可能となり、従来の C 末端側対応付け手法ではカバーしきれなかった配列空間の探索が可能となる。

#### 初期ランダム DNA ライブラリー作製法の開発

蛋白質の進化分子工学において、求める蛋白質ごとに最適な初期ペプチドライブラリーを用意することは、進化実験を効率的に行う上で重要である。そこで、各アミノ酸出現頻度の一様化やストップコドンの排除、二次構造の形成に関わるアミノ酸の含有率等を操作できる MLSDS (Multi-Line Split DNA Synthesis) 法を開発した。MLSDS 法は DNA 合成機を複数台使用し、3 塩基(1 コドン)伸長するごとに混合・分配を繰り返すコンビナトリアルケミストリー手法を用いて合成される。その際、各アミダイトの ATGC 混合比を多様に設定することによりアミノ酸(コドン)出現頻度を制御できる。

20 種アミノ酸出現頻度の一様化を行ったもの、天然に存在する球状蛋白質の平均アミノ酸組成と同じ組成を持つもの、および、Cys の出現率をゼロとし、

それ以外のアミノ酸を一様組成にしたものの、3 種のランダムライブラリー(ランダム部位 45base)を、遺伝的アルゴリズムを用いた ATGC 混合比最適化プログラムによって設計した。なお、これら 3 種はペプチド合成の早期終結を引き起こすストップコドンも排除されている。その結果、目標アミノ酸組成との相関係数が 0.9 以上のアミノ酸組成を持つライブラリーの設計に成功した。次に、それらライブラリーを実際に合成し、評価した。その結果、アミノ酸組成の実測値と設計値の相関係数が 0.8~0.9 程度となり、設計どおり合成できたことを確認した。

また、より高い機能を持つ蛋白質を得る為には、ある程度長いペプチド鎖を用意する必要がある。しかし、DNA の化学合成は 100~150base が合成限界である。そこで、合成したライブラリーユニット(87base)を PCR で二本鎖化し、制限酵素で切断・結合することにより 2~8 倍長 (ランダム部位 106~414base) の長鎖ランダムライブラリーを作成した。

## Chapter I

Development of mRNA-protein fusion  
at N-terminus for evolutionary protein engineering

## 1. Introduction

In evolutionary protein engineering, *in vitro* selection using a cell-free translation system carries advantages of large library size and applicability to cytotoxic proteins. Although various *in vitro* protein selection techniques such as *in vitro* virus [1], mRNA display [2] and ribosome display [3] have been developed, most of them are the techniques in which 3'-terminus of the genotype molecule is linked to the decoded peptide via its C-terminus. Thus, they are not adequate to the evolution of the protein, of which C-terminus is essential for its function (such as various peptide ligands, protein-protein interaction domain, membrane proteins, peptide hormones, etc). Therefore, it is worthwhile to devise the method in which the genotype molecule is linked to the decoded protein via its N-terminus, i.e. C-terminus of the protein is displayed freely.

Also in phage display method, the peptide is fused usually at the N-terminus of coat protein. However, the display method in which the peptide library was fused at the C-terminus of M13 phage major coat protein [4,5] and T7 phage D-coat protein [6] has been also reported. The carboxyl terminal end of PDZ domain-ligands, which is essential to be recognized by PDZ domain, was screened [7,8] and mapping of the C-terminal epitope of the Alzheimer's disease specific antibody [9] was carried out using such carboxyl-terminal phage display.

Several methods that link the protein to its mRNA (or DNA) via N-terminus have been reported. Sawata and Taira reported the binding procedure between MS2 coat protein dimer fused at N-terminus of the protein and the specific recognition sequence in its mRNA [10]. Yanagawa and coworkers developed the method named "STABLE" in

which the streptavidin fused at N-terminus of a peptide was bound to biotin labeled gene DNA in emulsion [11,12]. These methods are simple and effective; however, screening condition is limited because its linkage between the peptide and its genotype molecule is non-covalent. Baskerville and Bartel made the ribozyme that catalyzes a phosphoamide bond formation between the 5'-terminus of the encoding RNA and the N-terminus of the decoded polypeptide [13]. Nester and coworkers developed the method that employed the property of VirD2 protein, which covalently binds to the specific DNA sequence [14]. These methods employ a covalent bond to linkage the peptide and its genotype molecule; however, these have not been applied yet to *in vitro* selection.

To date, many studies on the non-natural amino acid incorporation into a protein using nonsense [15,16] or frame-shift [17] suppression have been reported. And recently, these have been applied to *in vitro* protein display methods to introduce the mRNA-peptide fusion molecule library containing non-natural amino acid [18,19].

In this article we report the novel method to link the N-terminus of the peptide to its mRNA. In our method, the mRNA is linked via ssDNA and spacer to phenylalanine derivative, which is acylated to amber sup tRNA. When this modified mRNA is subjected to a cell-free translation system, this phenylalanine derivative would be incorporated into the N-terminus region of the nascent peptide at the amber stop codon inserted at the beginning of the encoded region. Consequently, C-terminus of the peptide is displayed on its mRNA (Fig.1).

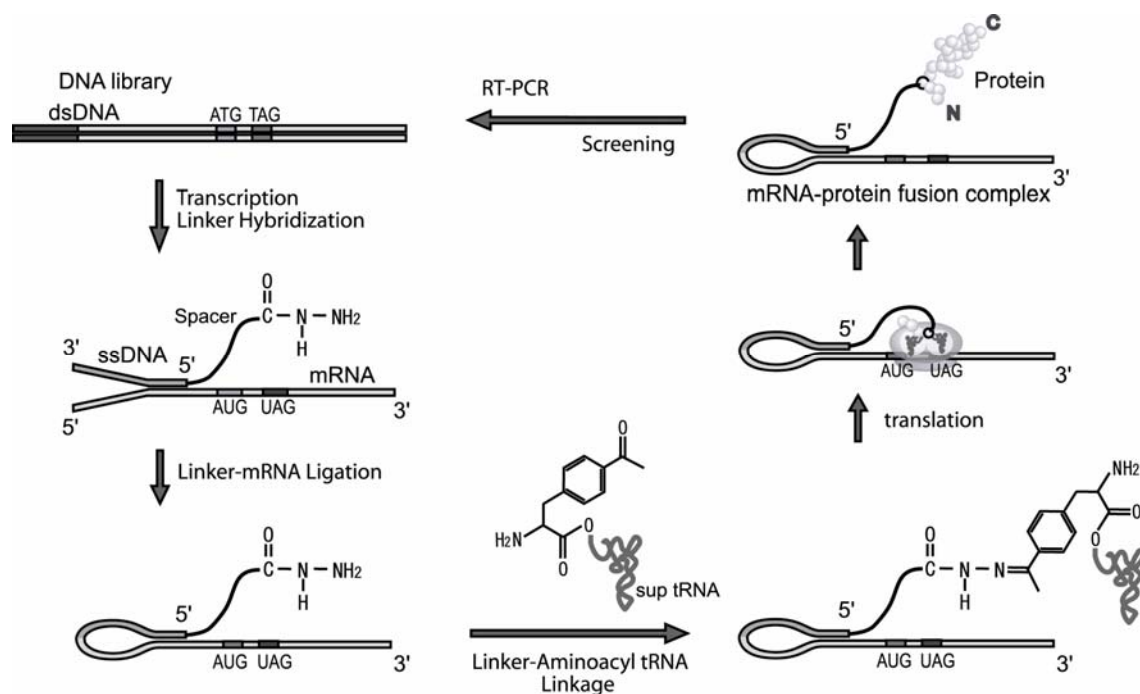


Fig. 1. Screening cycle of the mRNA-protein fusion in this study. The dsDNA library is transcribed to mRNA. The mRNA is hybridized to the DNA moiety of the linker having hydrazide group and ligated with T4 RNA ligase. Hydrazide group of the ligated product and acetyl group of the phenylalanine derivative that is acylated to sup tRNA are ligated chemically and the modified mRNA is translated. The modified aminoacyl sup tRNA tends to occupy the A-site of ribosome at UAG codon inserted near downstream of initiation codon and the phenylalanine is incorporated into the growing peptide. Thus, linkage between N-terminus of the nascent peptide and 5'-terminus of its mRNA is achieved. Screening of mRNA-peptide fusion library according to property of the displayed peptide and amplify the genotype molecules of the screened fusions by RT-PCR

## 2. Materials and methods

### 2.1 Synthesis of the DNA-linker having a hydrazide group

The DNA-linker having a hydrazide group at 5'-terminus (hydrazide-linker) was synthesized as follows. 5'-NH<sub>2</sub>-(CH<sub>2</sub>)<sub>6</sub>-(CH<sub>2</sub>CH<sub>2</sub>O)<sub>6</sub>-(dA)<sub>30</sub>-(CH<sub>2</sub>CH<sub>2</sub>O)<sub>6</sub>-GGG-(dT-FITC)-CGGGGGGCAAAA-3', where (dA)<sub>30</sub> is 30 nucleotides of deoxy adenosine, (dT-FITC) is fluorescein-modified deoxy thymidine, was purchased from Tsukuba Oligo Service Co., Ltd (Ibaraki, Japan). Single riboadenosine was added to 3'-terminus of the oligomer (1.0 nmol) using terminal deoxynucleotidyl transferase (24 U, Takara Bio) with 20 μmol ATP at 37 °C for overnight in 50 μl of supplied TdT buffer. The product was purified on P6 column (Bio Rad) to remove excess ATP, followed by phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation. The precipitate was dissolved in 65 μl of 0.5 M phosphate buffer (pH 9.0) and then mixed with 10 μl of 0.025 M succinic anhydride in DMSO for overnight at room temperature and purified on P6 column (Bio Rad) pre-equilibrated with H<sub>2</sub>O. Purified sample was mixed with 2 mg adipodihydrazide and 5 mg EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide) in 70 μl of 0.1 M phosphate buffer (pH 7.2) for 4 hour at room temperature and purified on P6 column pre-equilibrated with H<sub>2</sub>O.

The length of the DNA-linker used in this study was designed as follow. The local maximum concentration of two sites on a linear ssRNA chain (polyU) is described as:

$$j_{vM} = 0.13 \times v^{-\frac{3}{2}} \times \exp\left(\frac{-76}{v^2}\right) M, \quad \dots 1$$

where  $v$  is the distance-through-bond in nucleotides [20].

Probability distribution of local concentration  $j_v(r)$  is described as :

$$j_v(r) = j_{vM} \cdot \exp\left(\frac{-3r^2}{2nl^2}\right) \quad \dots 2$$

where  $r$  is the distance-through-space between the two sites,  $n$  is the distance-through-bond in the number of Kuhn segment and  $l$  is statistical segment length or Kuhn length. For ssRNA, the number of Kuhn segment and Kuhn length were reported for polyU as  $v/6$  and 4 nm, respectively.

The value of  $r$  for our case is 14 nm that is the distance-through-space between the exit position of mRNA on ribosome and the entrance of A-site of ribosome [21]. On the assumption that mRNA and the linker were composed of ssRNA, the probability distribution of local concentration of linked aminoacyl sup tRNA on the entrance of A-site of ribosome is given as:

$$j_v(14nm) = j_{vM} \cdot \exp\left(\frac{-110.4}{v}\right) = 0.13 \times v^{-\frac{3}{2}} \times \exp\left(\frac{-76}{v^2}\right) \times \exp\left(\frac{-110.4}{v}\right) M, \quad \dots 3$$

According to Equation 3, the maximum local concentration 45.5  $\mu$ M is obtained when the distance-through-bond  $v$  is 75 nt. Thus, the linker length was designed so that the chain length from the mRNA exit site to the 5'-terminus of the linker become equivalent to 75 nt.



## 2.2 Construction of the mRNA-templates

The His-tag encoding DNA (178 bp, Fig.2A) was constructed by three steps of PCR amplification. The first step PCR was performed using three oligomers; NL-temp (5'-ATGGGCTAGGGTGTGAGCAAGGGCGAGGAGCTGTTACCGGGGCAGCGCATCA C-3'), prNL+ (5'-CGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATA TATCCATGGGCTAGGGTGTGAGCA-3') and prNL- (5'-ACTCCTTAATGATGGTGATGG TGATGCGCTGCCCCGGT-3') and the resulting dsDNA was further amplified using two primers; prNL $\phi$ 2.5GC+ (5'- CAGTAATACGACTCACTATTAGGGCCCCCGACCCCGGT TTCCCTCTAGAAATAATTTTGT-3') and prNLHis- (5'-GCATCGACTCCTTAATGATGG TGATGGTGATGCGCTGCCCCGGTG-3'). The final step PCR was performed using the second step product dsDNA and two primers; prRT+ (5'-CAGTAATACGACTCACTATTA GGGCCCCCGAC-3') and prRT- (5'-GAATTCGCCCTTGCATCGACTC-3'). The FLAG encoding DNA (190 bp, Fig.2A) was constructed by two steps of PCR amplification. The first step PCR was performed using the His-tag encoding DNA and two primers; prRT+ and prNLFLAG- (5'-GCATCGACTCCTTATTACTTGTCATCGTCATCCTTGTAATCCGC CGCTGCCCCGGTGAACAG-3') and the resulting dsDNA was further amplified using two primers; prRT+ and prRT-. These PCR products were purified using QIAquick PCR purification kit (Qiagen) and transcribed in vitro using RiboMAX <sup>TM</sup> Large Scale RNA Production System - T7 (Promega) in the presence of 36 mM AMP and 3.6 mM each NTPs. The RNA transcript was purified on P6 column to remove excess AMP, followed by phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation.

### **2.3 Ligation between the mRNA and the hydrazide-linker**

The mRNA (2  $\mu$ M) was hybridized to the DNA moiety of the hydrazide-linker (4  $\mu$ M) and prRT<sup>-</sup> (4  $\mu$ M) by heating at 95 °C and cooling to 25 °C in 50  $\mu$ l of T4 RNA ligase buffer (Takara Bio) and ligation reaction was started by adding T4 RNA ligase (40 U, Takara Bio) and ribonuclease inhibitor (40 U, Takara Bio). Ligation reaction was performed at 25 °C for overnight and the ligated product was analyzed by 8 M urea 8 % PAGE using TBE running buffer at 65 °C and were visualized with fluorescence of FITC and then visualized again after staining by SYBR Green II (Cambrex) using a fluorescence imager (Pharos FX; Bio-Rad). Ligated product was purified using RNeasy Mini Kit (Qiagen).

### **2.4 Preparation of amino acid substrate**

4-acetyl-L-phenylalanine cyanomethyl ester was synthesized by as follows; 10 mg of Boc-4-acetyl-L-phenylalanine (Chem Impex) was dissolved in 1 ml of acetonitrile and mixed with 350  $\mu$ l of triethylamine and 100  $\mu$ l of chloroacetonitrile. This mixture was stirred at room temperature for over night and the reaction mixture was diluted with 10 ml of ethyl acetate and extracted with 20 ml of 0.5 N HCl. The organic layer was then washed with 20 ml of 4 % NaHCO<sub>3</sub> and 10 ml of saturated aqueous sodium chloride and dried using a rotary evaporator. The resulting residue was mixed with 250  $\mu$ l of TFA/anisole mixture (9:1) and stirred at room temperature for 30 min. The solvent was removed using a rotary evaporator and 1 ml of 4 M HCl in dioxane was added to form the hydrochloride salt. The solution was concentrated using a rotary evaporator and 2

ml of ether was added to precipitate the hydrochloride salt. These reactions were traced by Thin Layer Chromatography.

## 2.5 Preparation of aminoacyl-sup tRNA

Aminoacyl-sup tRNA was prepared by the same procedure previously reported [22] with minor modification using 4-acetyl-L-phenylalanine cyanomethyl ester as the amino acid substrate.

The template DNA encoding the aminoacyl ribozyme was constructed by two steps of PCR amplification. The first PCR was performed using three oligomers;

F<sub>x</sub> (5'-ACCTAACGCCAATACCCTTTCGGGCCTGCGGAAATCTTTCGATCC-3'), P5-1 (5'-ACGCATATGTAATACGACTCACTATAGGATCGAAAGATTTCCGC-3') and P5-2 (5'-GGTAACACGCATATGTAATACGACTC-3'). The resulting dsDNA was further

amplified using two primers; p5-2 and p3-2 (5'-T<sub>20</sub>-ACCTAACGCC-3'). The template DNA encoding the engineered tRNA<sup>Asn<sub>CUA</sub></sup> was constructed by two steps of PCR amplification. The first PCR was performed using three oligomers;

TR (5'-TGGTGCCTCTGACTGGACTCGAACCAGTGACATACGGATTTAGAGTCCGC CGTTCTACCGACTGAACTACAGAGGC-3'), p5-3 (5'-ACGCATATGTAATACGACTCACT ATAGCCTCTGTAGTTCAGTCGGT-3') and p3-3 (5'-TGGTGCCTCTGACTGGACTC-3').

The resulting dsDNA was further amplified using two primers; p5-2 and p3-3. Amplified these two dsDNAs, encoding aminoacyl ribozyme or tRNA<sup>Asn<sub>CUA</sub></sup>, were transcribed in vitro using RiboMAX™ Large Scale RNA transcription System – T7 in the presence of 7.5 mM GMP and 3.75 mM each NTPs. The RNA transcript was purified on P6 column to remove excess GMP, followed by phenol/chloroform/isoamyl alcohol extraction and

ethanol precipitation.

400  $\mu$ l of 0.1 M NaIO<sub>4</sub> was added to 1 ml of 30  $\mu$ M ribozyme, and the mixture was incubated on ice for 20 min. The oxidized RNA was then precipitated with 14 ml of 2 % LiClO<sub>4</sub> in acetone and washed with 1 ml acetone. The pellet was dissolved in 1.4 ml of 0.1 M sodium acetate (pH 5.0) and added to 0.7 ml of adipic acid dihydrazide-agarose (Sigma) that was prewashed with RNase-free water. The mixture was stirred at room temperature for three hour and then 300  $\mu$ l of 1 M NaBH<sub>3</sub>CN was added and the mixture was incubated for 30 min to stabilize the imine bond. The resin-immobilized ribozyme was washed with 1.4 ml of W1 (0.1 M sodium acetate, 300 mM NaCl, 7.5 M urea, 0.1% SDS, pH 5.0) and suspended in 2.1 ml of W1.

30  $\mu$ l of 50  $\mu$ M tRNA<sup>Asn</sup><sub>CUA</sub> was folded by heating at 95 °C and cooling to room temperature in a buffer (50 mM EPPS, 12.5 mM KCl, 12 mM MgCl<sub>2</sub>, pH 7.0) and then diluted in 15  $\mu$ l of a buffer (50 mM EPPS, 12.5 mM KCl, 3.6 M MgCl<sub>2</sub>, pH 7.0) and 20  $\mu$ l of a buffer (50 mM EPPS, 12.5 mM KCl, 1.2 M MgCl<sub>2</sub>, pH 7.0). The folded tRNA solution was mixed with 5 nmol of resin-immobilized ribozyme that was prewashed with RNase-free water, and then 12.5  $\mu$ l of 100 mM 4-acetyl-L-phenylalanine cyanomethyl ester in DMSO was added to this mixture followed by adding 1.6  $\mu$ l of 0.25 M KOH (to adjust pH to 7.0-7.2). After 2 hour incubation under rotation at 4 °C, EK buffer (50 mM EPPS, 12.5 mM KCl, 10 mM EDTA, pH 7.0) was added and the supernatant, containing aminoacyl-tRNA product, was recovered and ethanol precipitated. Confirmation of aminoacylation was performed as follows. 50 pmol of aminoacylated tRNA was dissolved in 5  $\mu$ l of EPPS (0.1 M, pH 5.9) containing 20 mM 6-(Biotinylamino)hexanoic acid N-hydroxy-sulfosuccinimide ester (Dojindo), and 1.0  $\mu$ l of EPPS-KOH (0.3 M, pH 9.1) was added to adjust pH to 8.0. The mixture was incubated for one hour on ice and

ethanol precipitated. The pellet of biotinylated aminoacyl-tRNA was dissolved in 10 µl of the loading buffer (167 pmol streptavidin, 0.1 M sodium acetate, 8 M urea, 0.05 % bromophenol blue, 0.05 % xylene cyanol, pH 5.0) and analyzed by 8 M urea 15 % PAGE using 50 mM sodium acetate (pH 5.0) as a running buffer at 30 °C and visualized after staining by SYBR Green II.

## 2.6 Selective enrichment analysis

10 pmol of the mRNA-linker ligation product encoding His-tag sequence was mixed with 100 pmol of 4-acetyl-phenylalanyl tRNA<sup>Asn</sup><sub>CUA</sub> in 10 µl of 1 mM sodium acetate (pH 5.2) for overnight at room temperature and translated in 16 µl volume of *E. coli* S30 Extract System for Linear Templates (Promega) for 5 min at 37 °C with SUPERase-In RNase inhibitor (20 U, Ambion). The translated sample was incubated in 100 µl of Ni-NTA agarose (Qiagen) suspended in wash buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20 mM imidazole, pH 8.0) for 1 hour at 4 °C using a rotary mixer. The Ni-NTA agarose was washed two times with 200 µl of wash buffer and eluted with 50 µl of elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluted sample was desalted with MicroSpin G-25 Column (GE Healthcare) and amplified by RT-PCR using *C. therm.* Polymerase One-Step RT-PCR System (Roche) using prRT<sup>+</sup> and prRT<sup>-</sup> as primers. As negative controls, the translation product without methionine and the translation product without modification of 4-acethyl-phenylalanine tRNA<sup>Asn</sup><sub>CUA</sub> to the linker before translation were screened with Ni-NTA agarose and amplified by RT-PCR. The amplified products were analyzed by 8 M urea 8 % PAGE using TBE running buffer at 65 °C and visualized after staining by SYBR Green I (Cambrex).

In selective enrichment analysis of the DNA encoding the His-tag sequence, Mixture of the mRNA-linker ligation products (1 pmol of the His-tag encoding molecule and 10 pmol of the FLAG-tag encoding molecule) was reacted with 100 pmol of 4-acetyl-phenylalanyl tRNA<sup>Asn</sup><sub>CUA</sub> in 10 µl of 1 mM sodium acetate (pH 5.2) for overnight at room temperature and translated in 16 µl volume of *E. coli* S30 Extract System for Linear Templates for 5 min at 37 °C with SUPERase-In RNase inhibitor (20U). The translated sample was incubated in 100 µl of Ni-NTA agarose suspended in wash buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20 mM imidazole, pH 8.0) for 1 hour at 4 °C using a rotary mixer. The Ni-NTA agarose was washed two times with 200 µl of wash buffer and eluted with 50 µl of elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluted sample was desalted with MicroSpin G-25 Column and amplified by RT-PCR using *C. therm.* Polymerase One-Step RT-PCR System using prRT<sup>+</sup> and prRT<sup>-</sup> as primers. The amplified product was transcribed and ligated with hydrazide-linker for the next round of screening. The amplified product of each screening round was analyzed by 8 M urea 8 % PAGE using TBE running buffer at 65 °C and visualized after staining by SYBR Green I.

In selective enrichment analysis of the DNA encoding the FLAG-tag sequence, Mixture of the mRNA-linker ligation products (1 pmol of FLAG-tag encoding molecule and 10 pmol of His-tag encoding molecule) was reacted with 100 pmol of 4-acetyl-phenylalanyl tRNA<sup>Asn</sup><sub>CUA</sub> in 10 µl of 1 mM sodium acetate (pH 5.2) for overnight at room temperature and translated in 16 µl reaction volume of *E. coli* S30 Extract System for Linear Templates (Promega) for 5 min at 37 °C with SUPERase-In RNase inhibitor (20U, Ambion). The translated sample was diluted in lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 % Triton X-100, pH 8.0) and added to 40 µl

of anti-FLAG M2 agarose (Sigma), followed by incubation for 1 hour at 4 °C using a rotary mixer. The anti-FLAG agarose was washed three times with 500 µl lysis buffer and incubated with 100 µl of TBS containing 0.1 mg / ml of FLAG peptide for 1 hour at 4 °C. The supernatant was desalted with MicroSpin G-25 Column and amplified by RT-PCR using *C.therm.* Polymerase One-Step RT-PCR System using prRT+ and prRT- as primers. The amplified product was transcribed and ligated with hydrazide-linker for the next round of screening. The amplified product of each screening round was analyzed by 8 M urea 8 % PAGE using TBE running buffer at 65 °C and visualized after staining by SYBR Green I.

Selective enrichment of the DNA encoding the His-tag sequence using PURESYSTEM [23] was performed as follows. The mixture of the mRNA-linker ligation products (0.5 pmol of His-tag encoding molecule and 5 pmol of FLAG-tag encoding molecule) was reacted with 50 pmol of 4-acetyl-phenylalanyl tRNA<sup>Asn</sup><sub>CUA</sub> in 5 µl of 1 mM sodium acetate (pH 5.2) for overnight at room temperature and the reaction product was ethanol precipitated. The precipitated sample was translated in 3.13 µl reaction volume of PURESYSTEM classic II (Post Genome Institute Co., Ltd.) for 15 min at 37 °C with SUPERase-In RNase inhibitor. The translated sample was incubated in 50 µl of Ni-NTA agarose suspended in wash buffer for 4 hour at 4 °C using a rotary mixer. The Ni-NTA agarose was washed five times with 400 µl of wash buffer and eluted with 100 µl of elution buffer. The eluted sample was desalted with MicroSpin G-25 Column and amplified by RT-PCR using *C.therm.* Polymerase Two-Step RT-PCR System (Roche) and KOD dash DNA polymerase (Toyobo) using prRT+ and prRT-FITC (fluorescein was modified on the 5'-terminus of prRT-) as primers. The amplified product was analyzed by 8 M urea 8 % PAGE using TBE running buffer at 65 °C and visualized with

fluorescence of FITC using a fluorescence imager.

Selective enrichment of the DNA encoding the FLAG-tag sequence using PURESYSTEM was performed as follows. The mixture of the mRNA-linker ligation products (0.5 pmol of FLAG-tag encoding molecule and 5 pmol of His-tag encoding molecule) was reacted with 50 pmol of 4-acetyl-phenylalanyl tRNA<sup>Asn</sup><sub>CUA</sub> in 5 µl of 1 mM sodium acetate (pH 5.2) for overnight at room temperature and the reaction product was ethanol precipitated. The precipitated sample was translated in 3.13 µl reaction volume of PURESYSTEM classic II for 15 min at 37 °C with SUPERase-In RNase inhibitor. The translated sample was diluted in TBS buffer and added to 40 µl of anti-FLAG M2 agarose, followed by incubation for 4 hour at 4 °C using a rotary mixer. The anti-FLAG M2 agarose was washed five times with 400 µl of TBS buffer and incubated with 100 µl of Elution buffer (0.1 M glycine-HCl, pH3.5) for 10 min at room temperature. The supernatant was desalted with MicroSpin G-25 Column and amplified by RT-PCR using *C.therm.* Polymerase Two-Step RT-PCR System and KOD dash DNA polymerase using prRT+ and prRT-FITC as primers. The amplified product was analyzed by 8 M urea 8 % PAGE using TBE running buffer at 65 °C and visualized with fluorescence of FITC using a fluorescence imager.

## **2.7 Detection of the mRNA-peptide fusion**

5 pmol of mRNA-linker ligation product was reacted with 50 pmol of 4-acetyl-phenylalanyl tRNA<sup>Asn</sup><sub>CUA</sub> in 0.1 M sodium acetate (pH 5.2) for overnight at room temperature and the reaction product was ethanol precipitated. The precipitated pellet was translated in 6.25 µl reaction volume of PURESYSTEM classic II for 30 min



at 37 °C with SUPERase-In RNase inhibitor. The translation product was blotted to PVDF-membrane after 15 % SDS-PAGE and labeled with anti-FLAG M2 monoclonal antibody (Sigma) as primary antibody and Cy3-linked anti-mouse IgG (GE Healthcare) as secondary antibody. Fluorescence of Cy3 was detected using a fluorescence imager.

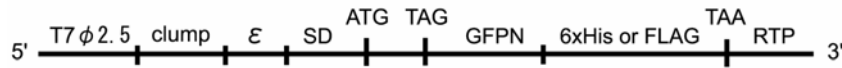
### **3. Results and Discussions**

#### **3.1 Ligation between the mRNA and the hydrazide-linker**

The connecting efficiency between the mRNA and the linker-molecule affects the diversity of the library. Thus, various sophisticated procedures have been devised [24-27]. In this report, we applied the Y-ligation method [25], which is the method of two single-strands ligation in the presence of a double-stranded stem to make a stem-loop structure in high efficiency. Because the transcripts were used as a phosphate donor substrate and the DNA moiety of hydrazide-linker were used as an acceptor substrate for T4 RNA ligase (Fig.2B), several modifications were required. First, transcription must be performed with excessive AMP (for T7 $\phi$ 2.5 promoter, GMP for normal T7 promoter and SP6 promoter) to generate the monophosphate at the 5'-terminus of the transcript and a riboadenosine must be attached to the 3'-terminus of the DNA moiety of hydrazide-linker by terminal deoxyribonucleotidyl transferase to be recognized by T4 RNA ligase (Fig.2B) [28]. Unless these modifications were performed, the ligated product was hardly observed (data not shown). Second, since 3'-OH of the transcript can be ligated with 5'-monophosphate of itself, self-ligation of the transcript was occurred

(Fig.3, lane 1-2). For blocking of the 3'-OH of the transcript, the DNA oligomer prRT<sup>+</sup> was hybridized to 3'-terminus region of the transcript before ligation reaction (Fig.2B). Self-ligation was effectively prevented (Fig.3, lane 4-5) and yield of desired product was increased from 10 % to 90 % (Fig.3B, lane 1 and 4). This blocking method is simple and effective compared with other methods (e.g., making 2', 3'-cyclic phosphate group by self-cleavage of ribozyme placed at 3'-terminus [29,30] or cleavage with deoxyribozyme [31,32].). Third, it is reported that T4 RNA ligase does not prefer guanine residue as a phosphate donor [33,34], but common promoters for *in vitro* transcription start the transcription with guanine residue, which is located at 5'-terminus of transcripts and act as phosphate donor in ligation reaction. Therefore, it is suspected that the transcripts generated by guanine-initiation type promoter are not suitable for donor substrate of T4 RNA ligase. Hence, We applied T7 $\phi$ 2.5 promoter, which initiates the transcription with adenine residue [35], because adenine is better than guanine as a donor substrate of T4 RNA ligase [33,34]. We compared the ligation efficiency of transcripts generated by three kind of promoters; T7, SP6 and T7 $\phi$ 2.5. These promoters generate the fixed triplet sequence at 5'-terminus of the transcripts; GGG, GAA and AGG respectively. The transcript by T7 $\phi$ 2.5 promoter gave the best efficiency of over 80 %. For SP6 and T7, the efficiency saturated at about 55 % and 20 %, respectively (Fig.4).

(A)



(B)

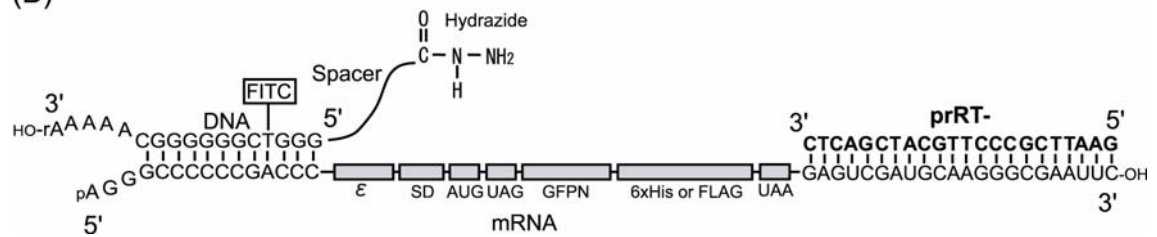


Fig. 2. Constructs of DNA and mRNA used in this study. (A) Construct of the model dsDNA. (B) Constructs of mRNA and DNA-linker having hydrazide group and the complex in the ligation reaction. T7 $\phi$ 2.5, T7 $\phi$ 2.5 promoter; clump, GC-rich sequence for hybridization to the DNA-linker;  $\epsilon$ , T7 gene 10 translation enhancer; SD, Shine-Dalgarno sequence; ATG or AUG, initiation codon; TAG or UAG, amber stop codon; GFPN, N-terminus 10 amino acids of Green Fluorescent Protein gene; 6x His, Histidine hexamer tag; FLAG, FLAG epitope tag, TAA or UAA, ocher stop codon; RTP, hybridization region for RT-primer prRT-; FITC, fluorescein isothiocyanate labeled to deoxythymidine.

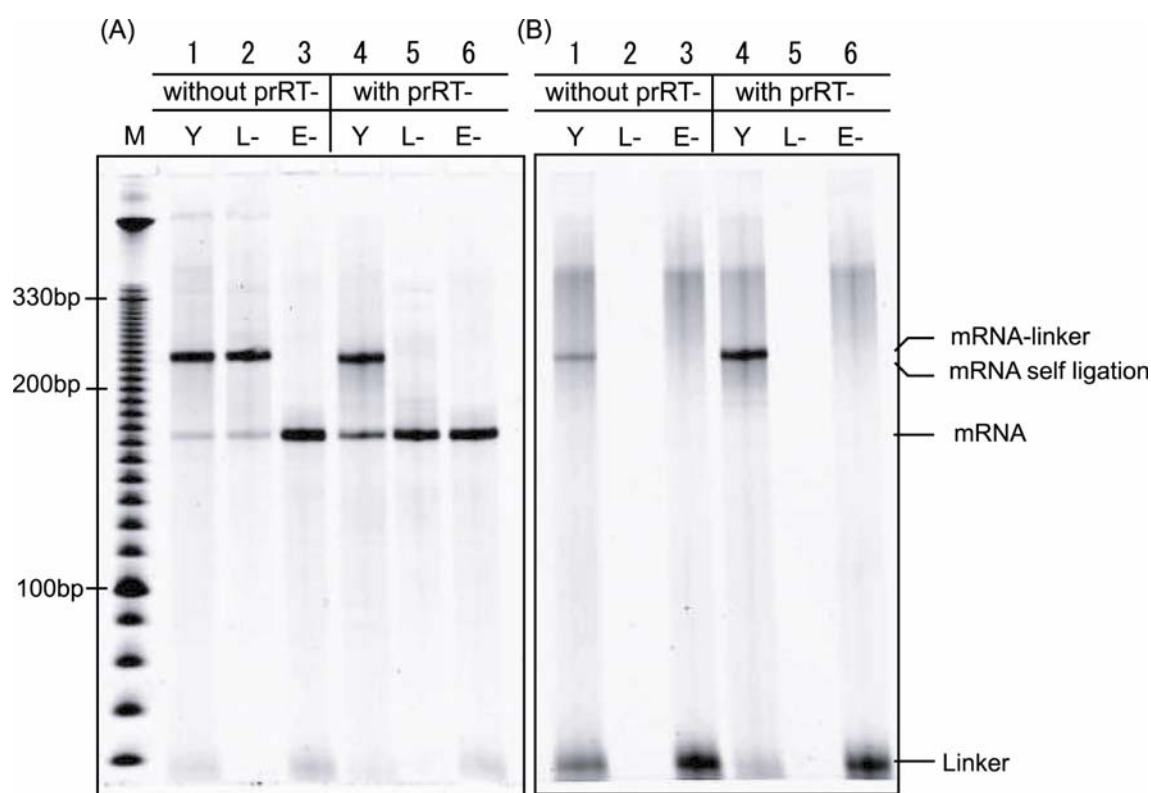


Fig. 3. Gel electrophoresis pattern of mRNA-linker ligation. The ligation products reacted with or without prRT<sup>-</sup> DNA oligomer used as a blocker of the 3'-end of mRNA were electrophoresis on 8 M urea 8 % PAGE at 65 °C and were visualized with fluorescence of (A) SYBR Green II and (B) FITC. Lane M: DNA ladder, Lane Y: ligation product, Lane L<sup>-</sup>: negative control, reaction product without DNA-linker, Lane E<sup>-</sup>: negative control, reaction product without T4 RNA ligase. Mobility of the mRNA-linker and the self-ligation product of mRNA are shown to be equivalent.

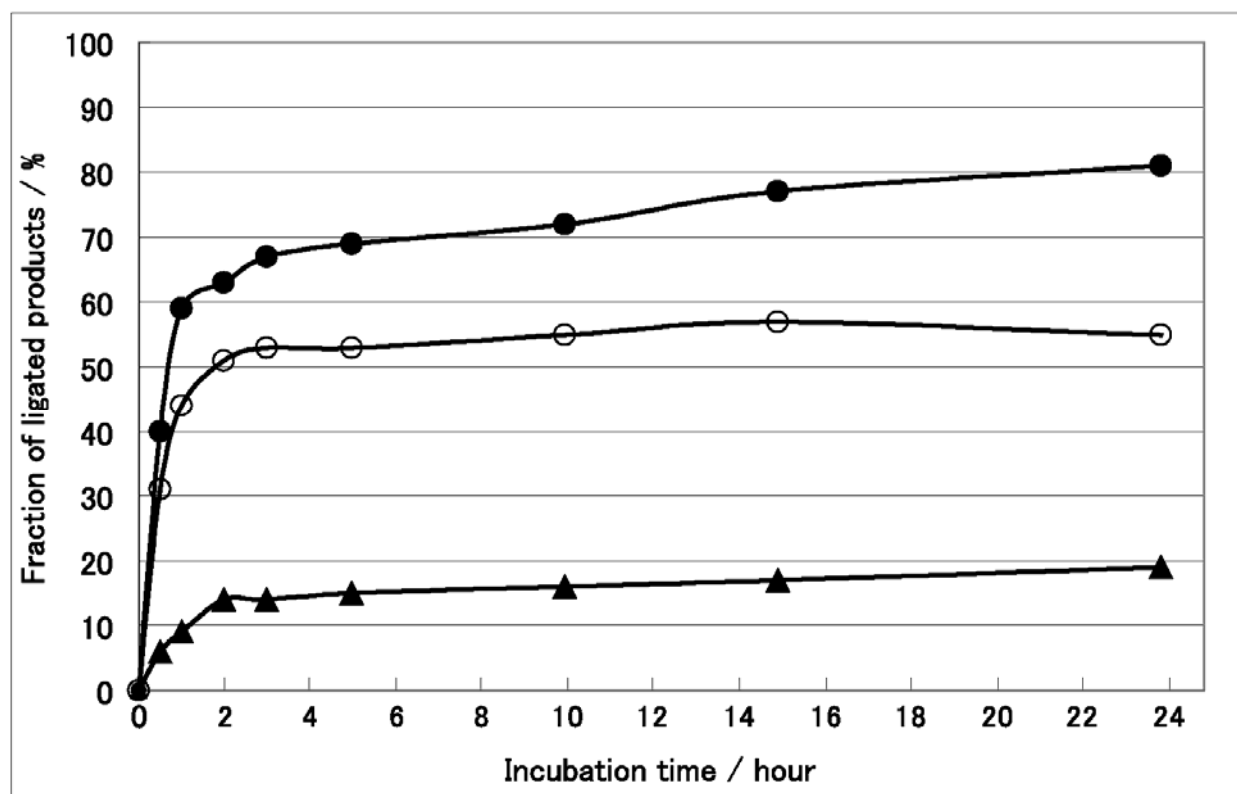


Fig. 4. Time course of ligation reactions using three mRNAs transcribed by different promoters. Filled circle, open circle and triangle are ligation product of mRNA transcribed by T7φ2.5 promoter, normal T7 promoter and SP6 promoter, respectively.

### 3.2 Aminoacylation of tRNA and conjugation between hydrazide-linker and aminoacyl-tRNA

tRNA<sup>Asn</sup><sub>CUA</sub> was acylated with 4-acetyl-L-phenylalanine using resin-immobilized ribozyme and aminoacylation efficiency was analyzed using a streptavidin-dependent gel-shift assay after biotinylation of the  $\alpha$ -amino group of 4-acetyl-L-phenylalanine. The efficiency was approximately 15 % in our procedure (Fig.5A).

The yield of conjugation between mRNA-linker-hydrazide and 4-acetyl-L-phenylalanyl tRNA<sup>Asn</sup><sub>CUA</sub> was estimated as follows. 7.5 pmol of hydrazide-linker and 75 pmol of 4-acetyl-L-phenylalanyl tRNA<sup>Asn</sup><sub>CUA</sub> were incubated in 5  $\mu$ l of 0.1 M sodium acetate (pH 5.2) for overnight at room temperature. The mixture was mixed with equal volume of loading buffer (0.1 M sodium acetate, 8 M urea, 0.05 % bromophenol blue, 0.05 % xylene cyanol, pH 5.0) and analyzed by 8 M urea 15 % PAGE using 50 mM sodium acetate (pH 5.0) as a running buffer at 30 °C and visualized with fluorescence of FITC and then visualized again after staining by SYBR Green II. The yield of the conjugation was approximately 10 % because of 90 % of hydrazide-linker was left without conjugation (Fig. 5B). This yield may be including the event of hydrolysis of aminoacyl-tRNA that has been conjugated to hydrazide-linker.

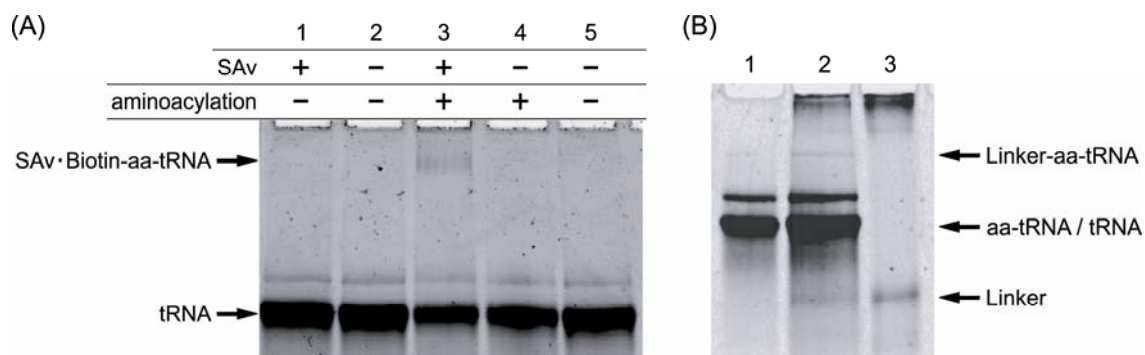


Fig.5 Aminoacylation of  $\text{tRNA}^{\text{Asn}}_{\text{CUA}}$  and conjugation between aminoacylated  $\text{tRNA}^{\text{Asn}}_{\text{CUA}}$  and hydrazide-linker. (A) Amino acylation of  $\text{tRNA}^{\text{Asn}}_{\text{CUA}}$  analyzed by streptavidin-dependent gel-shift assay. Approximately 15 % of tRNA was aminoacylated (Lane 3). Lane 1,  $\text{tRNA}^{\text{Asn}}_{\text{CUA}}$  in the presence of SAv; Lane 2 and 5,  $\text{tRNA}^{\text{Asn}}_{\text{CUA}}$  in the absence of SAv; Lane 3, aminoacylated  $\text{tRNA}^{\text{Asn}}_{\text{CUA}}$  in the presence of SAv; Lane 4, aminoacylated  $\text{tRNA}^{\text{Asn}}_{\text{CUA}}$  in the absence of SAv. (B) Conjugation between aminoacylated  $\text{tRNA}^{\text{Asn}}_{\text{CUA}}$  and hydrazide-linker analyzed by acidic denatured PAGE. Approximately 10 % of hydrazide-linker was conjugated with aminoacylated  $\text{tRNA}^{\text{Asn}}_{\text{CUA}}$  (Lane 2). Lane 1, aminoacylated  $\text{tRNA}^{\text{Asn}}_{\text{CUA}}$ ; Lane 2, conjugation product between aminoacylated  $\text{tRNA}^{\text{Asn}}_{\text{CUA}}$  and hydrazide-linker; Lane 3, hydrazide-linker

### 3.3 Selective enrichment analysis

We demonstrated a model selection experiment using a histidine-tag having high affinity to  $\text{Ni}^{2+}$ . The translation product displaying the His-tag peptide was screened using a Ni-NTA agarose column. The bound sample was eluted and used as a template for RT-PCR. As negative controls, the translation product without methionine that is necessary to initiate translation and the translation product without modification of 4-acetyl-phenylalanine sup tRNA to the linker before translation were screened with Ni-NTA agarose. The RT-PCR products were analyzed by electrophoresis and the desired product with higher amount was confirmed than one of negative controls generated by non-specific binding (Fig.6A).

Next, we performed enrichment experiments for the DNA mixture. It was achieved to enrich the DNA encoding the His-tag sequence from a mixture containing an order of magnitude excess of the other DNA encoding FLAG-tag sequence (Fig.2A) in three rounds screening with Ni-NTA (Fig. 6B). Furthermore, enrichment of the DNA encoding the FLAG-tag sequence from a mixture containing an order of magnitude excess of the other DNA encoding His-tag sequence was also accomplished in single round screening with anti-FLAG M2 antibody (Fig. 6B). Additionally, we also performed enrichment experiments using another translation system (PURESYSTEM) and achieved the enrichment of desired sequence (Fig. 6C). These results indicated that the enrichment was not an artefact by nonspecific binding of displayed peptide but a fact by specific binding of each peptide sequence.



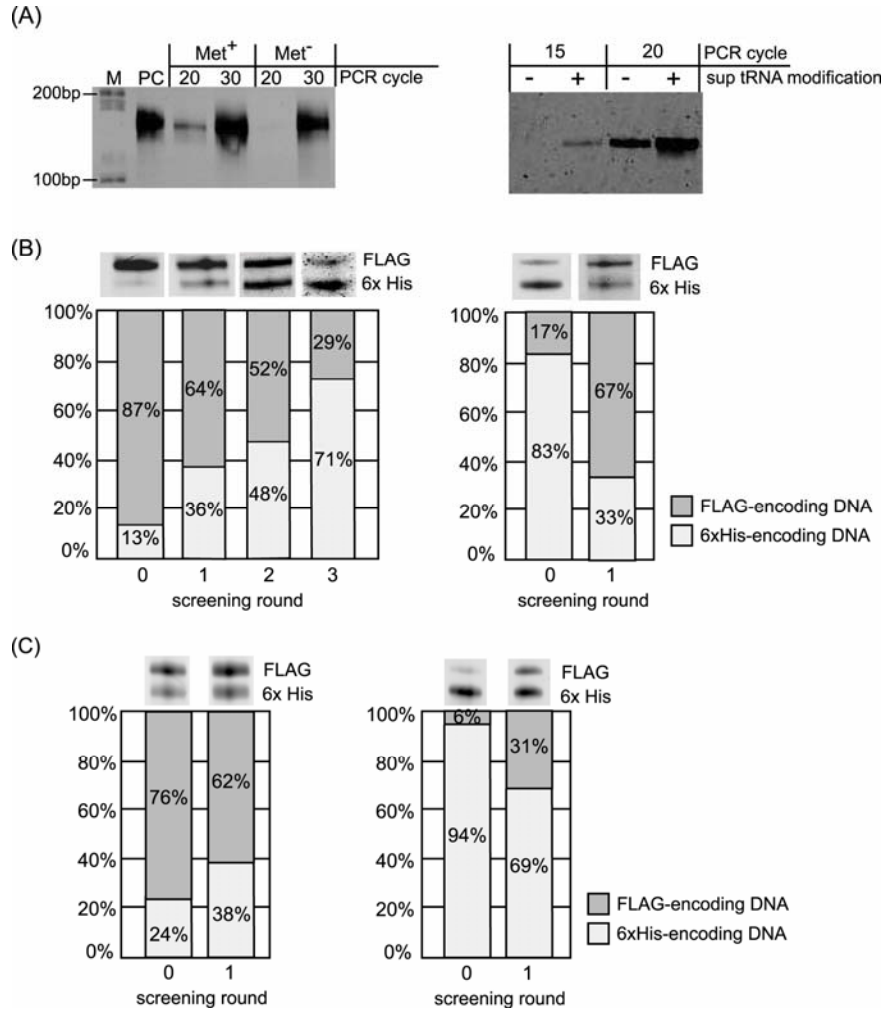


Fig. 6. Results of the selective enrichment experiment. (A) Result of RT-PCR after single-round Ni-NTA screening of the mRNA-peptide fusion displaying 6x His-tag sequence. Left panel, RT-PCR of the screened samples translated with or without methionine. Right panel, RT-PCR of the screened samples translated using mRNA with or without modification of 4-acetyl-phenylalanyl tRNA<sup>Asn</sup><sub>CUA</sub> before translation. Lane M: DNA ladder, Lane PC: positive control using mRNA before translation as template of RT-PCR. (B) Results of selective enrichment experiment for binary mixture: the FLAG-tag displaying fusion molecule and the 6x His-tag displaying fusion molecule. Initial molar ratio of the former to the latter is 87:13 for left panel and 17:83 for right panel. RT-PCR product at each screening round was analyzed by electrophoresis. Left panel; screening with Ni-NTA column. Right panel; screening with Anti-FLAG antibody column. (C) Results of selective enrichment experiment using PURESYSYSTEM for binary mixture: the FLAG-tag displaying fusion molecule and the 6x His-tag displaying fusion molecule. Initial molar ratio of the former to the latter is 76:24 for left panel and 6:94 for right panel. RT-PCR product at each screening round was analyzed by electrophoresis. Left panel; screening with Ni-NTA column. Right panel; screening with Anti-FLAG antibody column.

### 3.4 Detection of the mRNA-peptide fusion

The mRNA-linker ligation product encoding the FLAG peptide was translated and the mRNA-peptide fusion product was analyzed by SDS-PAGE and western blotting. Although the migration of the band caused by fusion formation was not detected on SDS-PAGE (Fig. 7A), the fusion molecule was detected by the western blotting (Fig. 7B). This low yield of the fusion formation was corresponding to relatively low efficiency of the selective enrichment (Fig. 6B, C). The low yield (~10 %) of the linker-aminoacyl-tRNA conjugate (Fig. 5B) affected directly the yield of the final product. In general, non-natural amino acid possessing large side group was incorporated into protein with relatively low efficiency. However, it is also reported that incorporation efficiency was not dependent on size of side group but on its shape, and the linear shape was appropriate [36]. In addition, non-natural amino acids possessing relatively long side chain (BODIPY FL-X-aminophenylalanine [37] or  $\epsilon$ -N-biotinyl-L-lysine [18,19,38]) were also successfully incorporated into the polypeptide. The side chain of the amino acid used in this study was linear and did not contain any large aromatic ring. However, since the side chain of the non-natural amino acid used in this study was extraordinarily long, this amino acid may be difficult to be incorporated into polypeptide or premature termination may be occur during peptide chain elongation if this amino acid was incorporated. The linker length also must affect the formation efficiency of the mRNA-peptide fusion. Therefore, experimental optimization of the linker length will give better efficiency although the length of the linker used in this study was optimized theoretically.

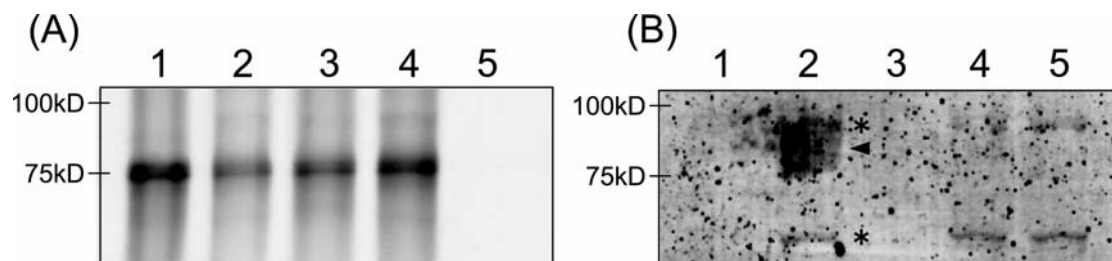


Fig. 7. Detection of mRNA-peptide fusion. The mRNA-linker ligation product that encoded FLAG peptide was translated and analyzed by the western blotting after SDS-PAGE separation. The fusion molecule was detected by the fluorescence of the Cy3 labeled on the secondary anti-body. (A) SDS-PAGE separation visualized by the fluorescence of FITC labeled on the hydrazide-linker. (B) Western blotting analysis visualized by the fluorescence of the Cy3 labeled on the secondary anti-body. The arrowhead and asterisk indicate the fusion molecule and nonspecific band, respectively. Lane 1: mRNA-linker-hydrazide. Lane 2: translated sample. Lane 3: proteinase K treated sample after translation. Lane 4: translated sample incubated with unaminoacylated sup tRNA in place of aminoacyl-sup tRNA before translation. Lane 5: translated sample without mRNA template.

### 3.5 Conclusion

We have developed a new method to prepare an mRNA-peptide fusion molecule covalently linked via N-terminus during *in vitro* translation. Because this fusion molecule displays the C-terminus of the peptide, it makes it possible to selection of the peptides for which function the C-terminus is essential. And also because this fused molecule has the free 3'-terminus of mRNA, it would make it possible to select an RNA replicase of de novo initiation-type.

And because a stop codon is inserted near downstream of the initiation codon, the translation reaction terminates just after initiation unless the sup tRNA is incorporated. Therefore unfused peptides that must compete with the fused peptides are never generated. Furthermore, this system is able to work even if the mRNA contains a termination codon. This method may be extended to the fusion formation linked at any position of protein by changing the position of amber codon that corresponds to the incorporation site of the linked aminoacyl sup tRNA. Thus, it may be possible that displaying the protein which having free terminus at both end and that arbitrary selection of the linking point on the protein according to its structure.

## REFERENCES

1. Nemoto, N., Miyamoto-Sato, E., Husimi, Y., Yanagawa, H. 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 Lett.* **414**, 405-408 (1997)
2. Roberts, R.W., Szostak, J.W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. USA* **94**, 12297-12302 (1997)
3. Hanes, J., Plückthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl. Acad. Sci. USA* **94**, 4937-4942 (1997)
4. Fuh, G., Sidhu, S.S. Efficient phage display of polypeptides fused to the carboxy-terminus of the M13 gene-3 minor coat protein. *FEBS Lett.* **480**, 231-234 (2000)
5. Weiss, G.A., Sidhu, S.S. Design and evolution of artificial M13 coat proteins. *J. Mol. Biol.* **300**, 213-219 (2000)
6. Sokoloff, A.V., Bock, I., Zhang, G., Sebestyen, M.G., Wolff, J.A. The interactions of peptides with

- the innate immune system studied with use of T7 phage peptide display. *Mol. Ther.* **2**, 131-139 (2000)
7. Fuh, G., Pisabarro, M.T., Li, Y., Quan, C., Lasky, L.A., Sidhu, S.S. Analysis of PDZ domain-ligand interactions using carboxyl-terminal phage display. *J. Biol. Chem.* **275**, 21486-21491 (2000)
  8. Laura, R.P., Witt, A.S., Held, H.A., Gerstner, R., Deshayes, K., Koehler, M.F., Kosik, K.S., Sidhu, S.S., Lasky, L.A. The Erbin PDZ domain binds with high affinity and specificity to carboxyl termini of  $\delta$ -catenin and ARVCF. *J. Biol. Chem.* **277**, 12906-12914 (2002)
  9. Khuebachova, M., Verzillo, V., Skrabana, R., Ovecká, M., Vaccaro, P., Panni, S., Bradbury, A., Novak, M. Mapping the C terminal epitope of the Alzheimer's disease specific antibody MN423. *J. Immunol. Methods.* **262**, 205-215 (2002)
  10. Sawata, S.Y., Taira K. Modified peptide selection in vitro by introduction of a protein-RNA interaction. *Protein. Eng.* **16**, 1115-1124 (2003)
  11. Doi, N., Yanagawa, H. STABLE: protein-DNA fusion system for screening of combinatorial protein libraries in vitro. *FEBS Lett.* **457**, 227-230 (1999)
  12. Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T., Yanagawa, H. DNA display for in vitro selection of diverse peptide libraries. *Nucleic. Acids. Res.* **31**, e118 (2003)
  13. Baskerville, S., Bartel, D.P. A ribozyme that ligates RNA to protein. *Proc. Natl. Acad. Sci. USA* **99**, 9154-9159 (2002)
  14. de Figueiredo, P., Roberts, R.L., Nester, E.W. DARTs: A DNA-based in vitro polypeptide display technology. *Proteomics* **4**, 3128-3140 (2004)
  15. Bain, J.D., Diala, E.S., Glabe, C.G., Dix, T.A., Chamberlin, A.R. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* **111**, 8013-8014 (1989)
  16. Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., Schultz, P.G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182-188 (1989)
  17. Hohsaka, T., Ashizuka, Y., Murakami, H., Sisido, M. Incorporation of nonnatural amino acids into streptavidin through in vitro frame-shift suppression. *J. Am. Chem. Soc.* **118**, 9778-9779 (1996)
  18. Li, S., Millward, S., Roberts, R. In vitro selection of mRNA display libraries containing an unnatural amino acid. *J. Am. Chem. Soc.* **124**, 9972-9973 (2002)

19. Muranaka, N., Hohsaka, T., Sisido, M. Four-base codon mediated mRNA display to construct peptide libraries that contain multiple nonnatural amino acids. *Nucleic. Acids. Res.* **34**, e4 (2006)
20. Rippe, K. Making contacts on a nucleic acid polymer. *Trends. Biochem. Sci.* **26**, 733-740 (2001)
21. Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H., Noller H.F. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883-896 (2001)
22. Murakami, H., Kourouklis, D., Suga, H. Using a solid-phase ribozyme aminoacylation system to reprogram the genetic code. *Chem. Biol.* **10**, 1077-1084 (2003)
23. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., Ueda, T. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751-755 (2001)
24. Kurz, M., Gu, K., Lohse, P.A. Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA-protein fusions. *Nucleic. Acids. Res.* **28**, e83 (2000)
25. Tabuchi, I., Soramoto, S., Suzuki, M., Nishigaki, K., Nemoto, N., Husimi, Y. An efficient ligation method in the making of an in vitro virus for in vitro protein evolution. *Biol. Proced. Online.* **4**, 49-54 (2002)
26. Tabuchi, I., Soramoto, S., Nemoto, N., Husimi, Y. An in vitro DNA virus for in vitro protein evolution. *FEBS Lett.* **508**, 309-312 (2001)
27. Miyamoto-Sato, E., Takashima, H., Fuse, S., Sue, K., Ishizaka, M., Tateyama, S., Horisawa, K., Sawasaki, T., Endo, Y., Yanagawa, H. Highly stable and efficient mRNA templates for mRNA-protein fusions and C-terminally labeled proteins. *Nucleic. Acids. Res.* **31**, e78 (2003)
28. Kinoshita, Y., Nishigaki, K., Husimi, Y. Enzymatic synthesis of sequencing primers based on a library of tetramers. *Chem. Express.* **7**, 149-152 (1992)
29. Kim, I., Lukavsky, P.J., Puglisi, J.D. NMR study of 100kDa HCV IRES RNA using segmental isotope labeling. *J. Am. Chem. Soc.* **124**, 9338-9339 (2002)
30. Schurer, H., Lang, K., Schuster, J., Morl, M. A universal method to produce *in vitro* transcripts with homogeneous 3' ends. *Nucleic. Acids. Res.* **30**, e56 (2002)
31. Santoro, S.W., Joyce, G.F. A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. USA* **94**, 4262-4266 (1997)
32. Pyle, A.M., Chu, V.T., Jankowsky, E., Boudvillain, M. Using DNzyme to cut, process, and map

- RNA molecules for structural studies or modification. *Methods Enzymol.* **317**, 140-146 (2000)
33. Romaniuk, E., McLaughlin, L.W., Neilson, T., Romaniuk, P.J. The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction. *Eur. J. Biochem.* **125**, 639-643 (1982)
34. England, T.E., Uhlenbeck, O.C. Enzymatic oligoribonucleotide synthesis with T4 RNA ligase. *Biochem.* **17**, 2069-2076 (1978)
35. Huang, F., Bugg, C.W., Yarus, M. RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. *Biochem.* **39**, 15548-15555 (2000)
36. Hohsaka, T., Kajihara, D., Ashizuka, Y., Murakami, H. and Sisido M. Efficient incorporation of nonnatural amino acids with large aromatic group into streptavidin in in vitro protein synthesizing systems. *J. Am. Chem. Soc.* **121**, 34-40 (1999)
37. Hohsaka, T., Abe, R., Shiraga, K. and Sisido, M. Incorporation of fluorescently labeled nonnatural amino acids into proteins in an E.coli in vitro translation system. *Nucleic. Acids. Research Supplement* **3**, 271-272 (2003)
38. Murakami, H., Ohta, A., Ashigai, H. and Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods* **3**, 357-359 (2006)

## Chapter II

Development of a novel combinatorial method  
to make high quality peptide libraries



## 1. Introduction

近年、ファージディスプレイ[1-4] やリボソームディスプレイ[5]、そして *in vitro* virus 法等の進化蛋白質工学を用いた新規機能性分子の実用性が示されてきている。新規機能性分子を獲得する効率は、採用するスクリーニング技術とライブラリーの品質に大きく依存する。進化蛋白質工学で用いられる初期 DNA ライブラリーに最も必要とされる性質はストップコドンの排除である。ライブラリー中にストップコドンが存在すると、翻訳の早期終結が引き起こされ、その結果、短いペプチドが多数発生し、長さの揃っていないペプチドライブラリーが生成される。次に必要とされるものは、ライブラリー中のアミノ酸組成の最適化である。アミノ酸配列空間を偏りなく探索する為には、ライブラリーのアミノ酸組成に、望まない偏りが発生しないようにする必要がある。さらに、目的とする機能分子ごとにライブラリーのアミノ酸組成を最適化することによって、広大なアミノ酸配列空間を効率よく探索することが可能となり、進化実験の時間の短縮につながる。単純なランダムライブラリー( $(NNN)_n$  ( $N=A/T/G/C$ ))の場合、30 アミノ酸長以上の長さになると確実にストップコドンが発生する。また、コドンの縮退を考慮していない為、アミノ酸組成に偏りが生じる。その為、それらの問題を克服する為に数々のライブラリー作成法が開発されてきている。従来、最も一般的に用いられていたライブラリーは $(NNK)_n$  若しくは $(NNS)_n$  ( $K=T/G$ ,  $S=G/C$ )である[4,6,7]。しかしこれらはストップコドン TAA を排除できておらず、アミノ酸組成設計の自由度も低い。

ストップコドンを排除するとともに、アミノ酸組成の制御も可能なライブラリー作成法が 2 例報告されている。その 1 つは、ホスフォロアミダイトのトリマーを 20 アミノ酸分 20 種合成したものを用いてライブラリーを合成するトリヌクレオチドホスフォロアミダイト (3 NPs)法である[8,9,10,16]。20 種のトリアミダイトを用いることでアミノ酸組成の制御はほぼ完全に可能であるが、トリアミダイトという特殊な試薬を用いる必要がある。トリアミダイト試薬は市販されているが、高価であり、反応効率も低いため汎用性にかける。2 つ目の方法は、コンピュータプログラムによって目的とするアミノ酸組成を実現するように、コドンの各塩基の ATGC 比を最適化し、翻訳後の C 末端にアフィニティタグを持つように合成したライブラリーを、mRNA display でプレセクションすることによってストップコドンを有するライブラリーを排除するという方法である[12]。アミノ酸組成の最適化の際、

ストップコドンの排除を考慮する必要が無いが、合成を 1 ラインで行うため、設計の自由度はそれほど高くない。

そこでライブラリーを作製する上で発生するこれらの問題を克服する為に、一般的な試薬と数台の DNA 合成機のみで高品質な DNA ライブラリーの作製が可能な手法を考案した。MLSDS(Multi-Line Split DNA Synthesis)法と名付けられたこの手法は、ペプチドのスプリット合成を DNA 合成に応用したもので、高い分子多様性を保ちつつ、コドンレベルでのアミノ酸組成の制御が可能である[11-15]。

高い機能を有する天然蛋白質の平均アミノ酸長は 300a.a.程度である。しかしながら DNA の化学合成は 100base 程度が品質保持限界である。より高い機能をもつ蛋白質を進化的に得るには、長い ORF を持つ DNA ライブラリーを用意する必要がある。そこで比較的短鎖の高品質 DNA ライブラリーを合成し、それを連結することによって長鎖ライブラリーを作製する。短鎖 DNA ライブラリーで最適化したアミノ酸組成を保つ為には、長鎖化によってライブラリーのコドンフレーム及び翻訳方向が換わらない実験手法を採用しなければならない。

## 2. Materials and method

### 2-1 MLSDS 合成

まず各合成ラインで対応するコドンユニットを 1 塩基ずつ合成する。次に全合成ラインの合成産物を混合し、再び各ラインに分配、その後各合成ラインでコドンユニットを 1 塩基ずつ合成し、合成産物を再び混合・分配する。以上の操作を繰り返すことによって長鎖で大きな分子多様性を有するランダム DNA ライブラリーを作製する。各コドンユニットの第 1、第 2、第 3 塩基の合成の際、其々独立した ATGC アミダイトミックスを用いるが、各アミダイトミックスの ATGC 混合比を操作することによってライブラリー中のアミノ酸出現頻度をコドン単位でコントロールすることが出来る。従って、ライブラリーから終止コドンを排除することが可能である上、アミノ酸の含有率を自由に操作することが可能である。また、コドンユニットの各塩基の ATGC 混合比を等比として上記合成サイクルを 3 ライン

で 10 回行った場合の分子多様性は、 $(4^3)^{10} \times 3^{10} = 6.8 \times 10^{22}$  となり十分なダイバーシティも得ることが出来る。合成ライン 3 で MLSDS 合成を行った場合の概略図を Fig.2-1 に示す。

天然に存在する球状蛋白質の平均アミノ酸組成をもつライブラリー”MLSDS Natural”、システインの含有率を 0%としたライブラリー”MLSDS NoCys”を合成した。実際の MLSDS 合成は DNA 合成会社に委託した。3 ラインの合成サイクルを 15 回行った。各ラインの合成スケールは  $1 \mu\text{M}$  で行った。配列は以下の通りである。5'-GAT GAG GCG AAG ACG NAC TGS ( $\alpha_1 \alpha_2 \alpha_3 / \beta_1 \beta_2 \beta_3 / \gamma_1 \gamma_2 \gamma_3$ ) $\times 15$  NAC TGS GAG GCT GGC TGC CAC-3'。ここで N(A/T/G/C), S(G/C),  $\alpha_{1,2,3}$ ,  $\beta_{1,2,3}$ ,  $\gamma_{1,2,3}$  は Fig.2.1 に順ずる。

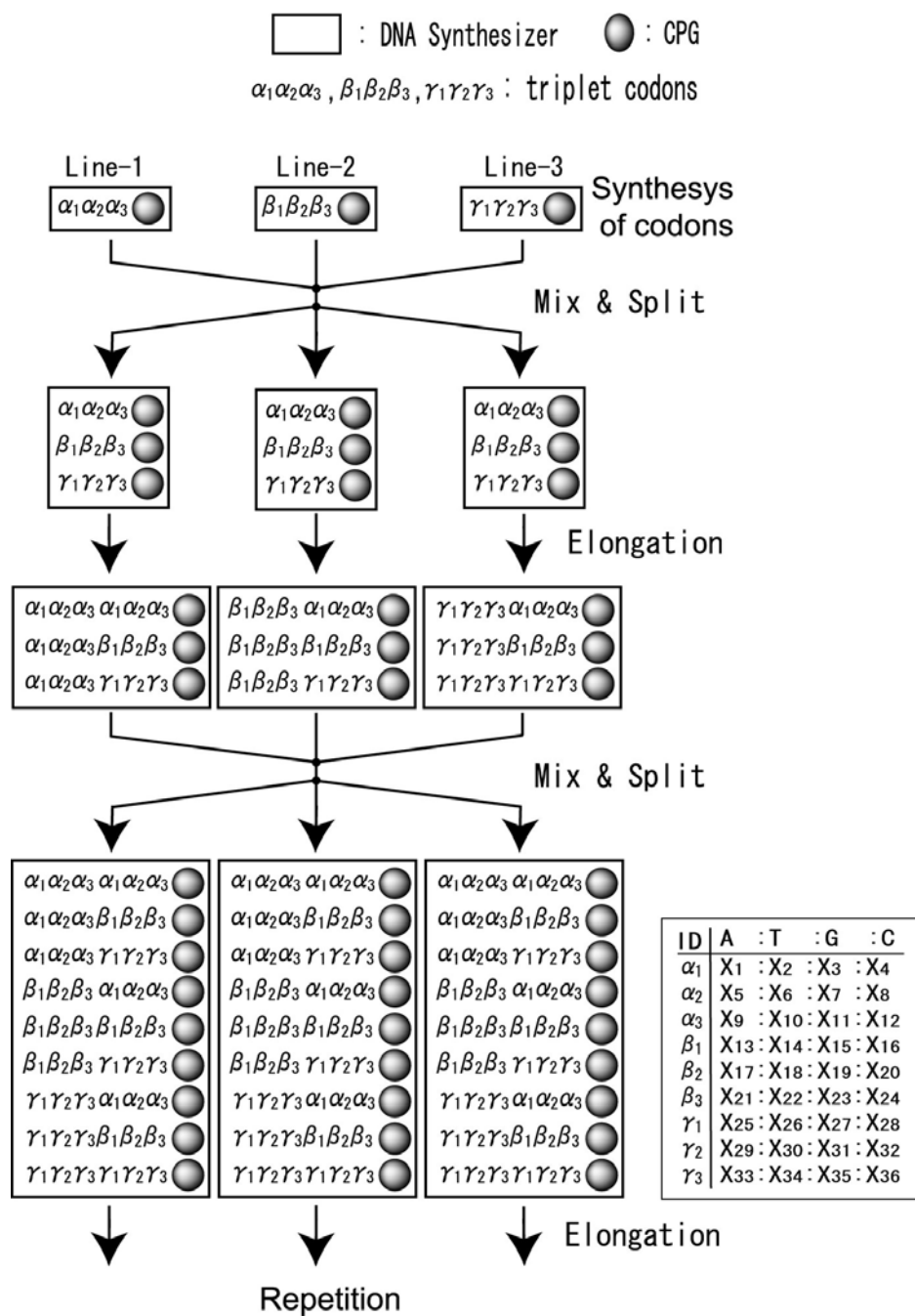


Fig.2-1 合成ライン 3 での MLSDS 法

$\alpha, \beta, \gamma$  は各々 DNA1 塩基を示しており、各々独立した ATGC 混合比を設定する。

## 2-2 MLSDS 最適化アルゴリズム

各ライブラリーの ATGC アミダイトミックスの塩基混合比  $X_i (i=1,2,3,\dots,n)$  は遺伝的アルゴリズム(GA)を用いて最適化する。ここで、 $n=4(A,T,G,C) \times 3(\text{nucleotides composed single codon}) \times L(\text{synthesis line})$  である。 $X_i$  の値域は実験操作を考え簡単のため  $\{0,1,2,3,4,5\}$  の整数とし、目標アミノ酸組成  $\mathbf{p}_t$  と設計アミノ酸組成  $\mathbf{p}$  間の相関係数  $F[\mathbf{x}]$  を評価関数とする。

$$F[\mathbf{x}] = \frac{21\mathbf{p}_T \cdot \mathbf{p} - \sum_{i=1}^{21} p_{T,i} \cdot \sum_{i=1}^{21} p_i}{\sqrt{(21p_T^2 - 1)(21p^2 - 1)}}$$

$$\mathbf{p} = (p_1, p_2, p_3, \dots, p_{21}) = (p_{Ala}, p_{Arg}, p_{Asn}, \dots, p_{Term})$$

$$\mathbf{p}_T = (p_{T1}, p_{T2}, p_{T3}, \dots, p_{T21}) = (p_{TAla}, p_{TArg}, p_{TAsn}, \dots, p_{TTerm})$$

$L=3$  のとき、 $p_1 (=Ala/GCN)$  は次で表される。

$$p_1 = \frac{1}{3} \left\{ \frac{x_3 x_8}{(x_1 + x_2 + x_3 + x_4)(x_5 + x_6 + x_7 + x_8)} + \frac{x_{13} x_{20}}{(x_{13} + x_{14} + x_{15} + x_{16})(x_{17} + x_{18} + x_{19} + x_{20})} + \frac{x_{27} x_{32}}{(x_{25} + x_{26} + x_{27} + x_{28})(x_{29} + x_{30} + x_{31} + x_{32})} \right\}$$

この  $X_n$  の最適化を行うために、遺伝的アルゴリズムを用いる。アルゴリズムの主な流れは、  
 1.  $X_n$  をランダムに生成する。 2. 目標組成との相関係数を算出する。 3. 相関係数の高い  $X_n$  に交差・変異を加える。最も相関係数を高いものはそのまま次世代へ移す。2に戻る。  
 この操作を繰り返すことにより  $X_n$  を最適化する(Fig.2-2)。DNA合成におけるフォスフォロアミダイトの縮合反応1回につき 0.5%の確率で生じる塩基欠失(rational deletion)によるフレームシフトも考慮に入れる。プログラミングソフトはMathematica<sup>®</sup> ver4.0 (Wolfram Reserch)を使用した。

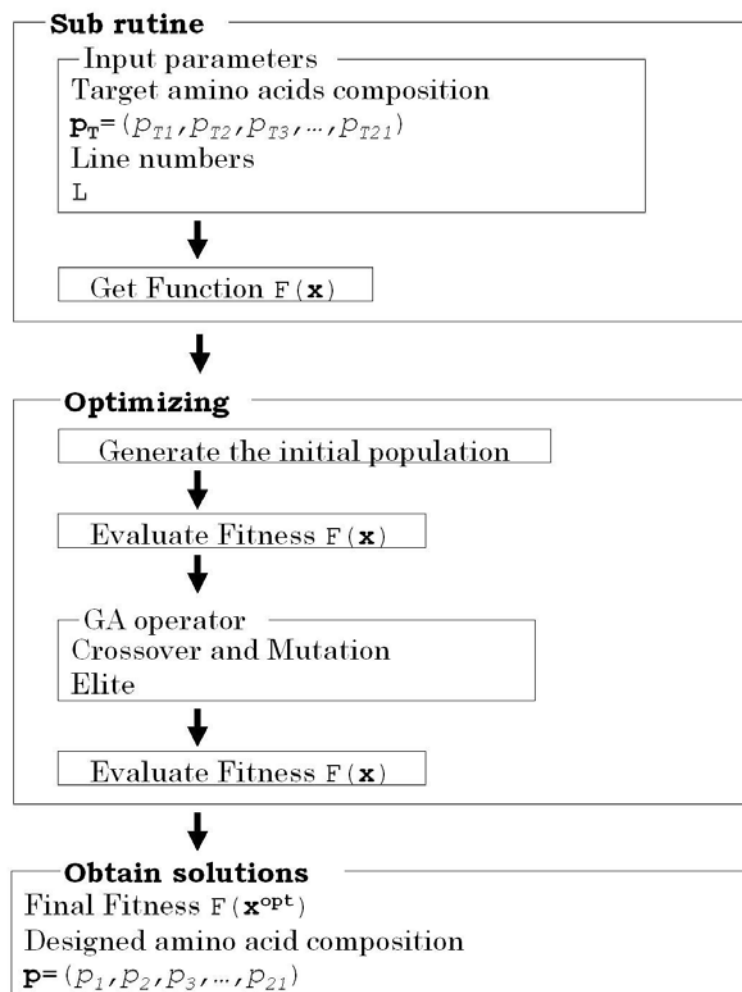


Fig.2-2 Xn 最適化アルゴリズムのフローチャート

## 2-3 配列解析

“MLSDS Natural” “MLSDS No Cys” を プ ラ イ マ ー prMLSDSSTOP+(5'-TAAATAGATGAGGCGAAGACG-3'), prMLSDSSTOP-(5'-TGATTAATTAGTGGCAGCCAGCCTC-3')、及び KOD Dash DNA Polymerase (TOYOBO) を用いて PCR した。

PCR 産物を TOPO TA Cloning® Kit (Invitrogen)を用いてクローニングし、得られたコロニーからプラスミドを抽出、M13 Forward(-20)プライマーを用いて配列解析した。プラスミド抽出、配列解析はシーケンス受託会社に依頼した。

## 2-4 ライブラリーの長鎖化

短鎖 DNA ライブラリーを合成後、DNA ポリメラーゼを用いて二本鎖化する。二本鎖 DNA を制限酵素で切断し、切断断片を T4 DNA Ligase を用いて連結する。以上の操作を繰り返すことにより長鎖 DNA を作製する。この長鎖化手法は Jack W. Szostak からも採用している [12]。短鎖 DNA ライブラリー合成の際、ランダム領域の上下流に 2 種類の class II S 制限酵素の認識部位と切断部位、及びポリメラーゼ伸長反応の為にプライミング配列を配置する。各々の制限酵素による切断面は同塩基数だけ突出し、突出面同士が Watson-Crick 対合する方向でなければならない。突出末端で対合し、連結することで、コドンフレームと翻訳方向を制御できる。また、ポリメラーゼ伸長反応と制限酵素反応を用いることにより、プライミング配列及び制限酵素認識部位を含まない DNA 合成不完全長産物を排除することが出来る。切断断片は 2 種類の制限酵素による突出末端が Watson-Crick 対合する配列でなければならないが、混合塩基を含む 2 つのコドン配列の中に突出配列を配置することによって、連結後にある特定のアミノ酸をコードするコドンが発生しないようにする。class II S 制限酵素として *Bbs* I と *Bbv* I を採用した場合の配列例を Fig2-2 に示す。先の切断断片の制約をクリアする class II S 制限酵素の対なら他の酵素を用いて構わない(例、*Bsm* A I と *Bsm* B I)。上記条件を満たす短鎖 DNA ライブラリーを合成後、ポリメラーゼ伸長反応で二本鎖化する。二本鎖化した DNA ライブラリーを 2 種類の制限酵素で別々に反応させ、

各々の切断断片同士を T4 DNA Ligase で連結する。連結産物を再び制限酵素で処理し、T4 DNA Ligase で連結する。この反応サイクルを繰り返すことによって、長鎖 DNA ライブラリーを作製する(Fig.2-3)。

各酵素反応及び反応後の精製における収率が 100%ではないので、反応サイクルを繰り返すことにより分子多様性が減少する。しかし、T4 DNA Ligase 反応後に減少分だけ PCR で増幅することにより、理論上は分子多様性の減少を相殺することが可能である。

“Natural”, “No Cys” を プ ラ イ マ ー prRD15+(5'-GAGGCGAAGACGNACTG-3', N=A/T/G/C), prRD15-(5'-GTGGCAGCCAGCCTC-3')、KOD Dash DNA Polymerase (TOYOBO) を用いて PCR した。PCR 産物を制限酵素 *Bbs* I (TAKARA)若しくは *Bbv* I (TAKARA)で 37℃で 2~8 時間処理する。*Bbs* I および *Bbv* I で切断した DNA 断片を T4 DNA Ligase(TAKARA)で 16℃で一晩インキュベートしてライゲーションした。ライゲーション産物を 10%ポリアクリルアミドゲルで精製し、プライマーとして prRD15+, prRD15-, PCR 酵素は KOD Dash DNA Polymerase (TOYOBO) を用いて PCR で増幅した。増幅産物を再び、制限酵素 *Bbs* I、*Bbv* I で処理し、切断断片を T4 DNA Ligase で連結する操作を繰り返すことによりライブラリーの長鎖化を行った。



<i>Bbs</i> I 認識部位	<i>Bbs</i> I 切断部位	<i>Bbv</i> I 切断部位	<i>Bbv</i> I 認識部位
5'-GAT GAG GC	G AAG AC	N AC TG S	GAG GCT G GC TGC
3'-CTA CTC CG	C TTC TG	N TG AC	S CTC CGA C CG ACG
(123/456/789)×15 ← ランダム領域 →			
			CAC-3'
			GTG-5'

Fig.2-2 短鎖 DNA ライブラリー配列例

連結後の配列”NAC TGS”(N=A/T/G/C, S=G/C)は Asn,Asp,His,Tyr 及び Cys,Trp をコードしている。

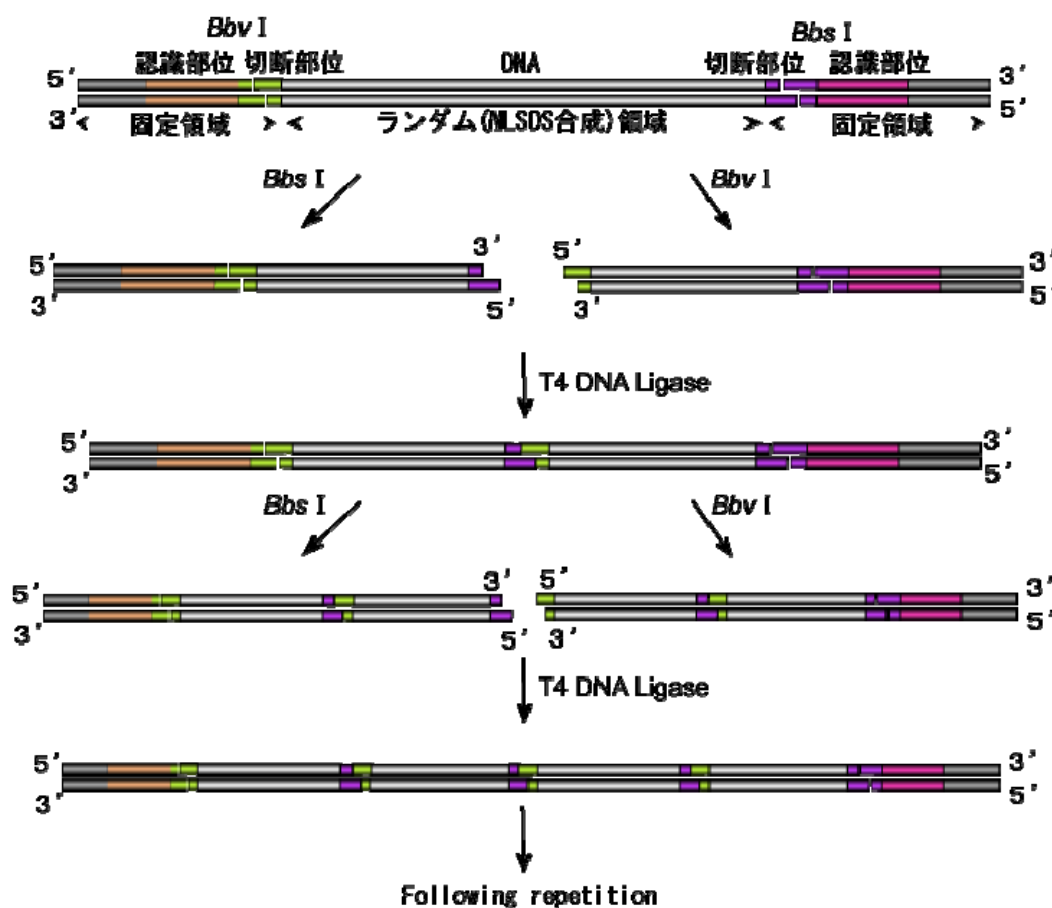


Fig.2-3 制限酵素処理法による DNA ライブラリーの長鎖化  
class II S 制限酵素として *Bbs*I と *Bbv*I を用いた場合。

### 3. Results and discussions

#### 3-1 ライブラリーの設計及び合成

##### 3-1-1 “MLSDS Natural”

立体構造をとりやすいアミノ酸組成をもつライブラリーから進化を開始することで、機能を有する蛋白質を効率よく得ることが出来ると考えられる。そこで、天然に存在する球状蛋白質の平均アミノ酸組成を持つライブラリーを作製した。ATGC 混合比の最適化条件として 1.ストップコドンの排除、2.天然に存在する球状蛋白質の平均アミノ酸組成[17] に近似、3.rational deletion によるフレームシフト後のアミノ酸組成の影響を最小化の 3 条件を満たすように合成ライン 3 の MLSDS 法に基づき塩基混合比  $X_n$  を最適化した(Table.3-1)。最適化は、天然の球状蛋白質の平均アミノ酸組成(Table.3-2, “Target composition”)に対する相関係数を評価関数に適用した遺伝的アルゴリズムを用いて行った。最終的に得られた組成の目標組成との相関係数は、rational deletion による point deletion を考慮しない場合 0.93、考慮した場合 0.99 となり非常に高い近似配列を得ることが出来た。最適化アミノ酸組成と目標アミノ酸組成を Table.3-2 に、それらのヒストグラムを Fig.3-1 に示す。Point deletion を考慮しない場合はストップコドン(TERM)の出現を 0%に抑えることに成功している。Point deletion が発生した場合も組成全体の傾向を大きくずらすことなく、ストップコドンの発生率を最小限に抑えている。

**Table.3-1**  
“MLSDS Natural”の  
ATGC 混合比

ID	A:T:G:C
$\alpha_1$	5:0:5:3
$\alpha_2$	4:1:0:1
$\alpha_3$	2:1:2:0
$\beta_1$	2:3:5:2
$\beta_2$	0:2:5:2
$\beta_3$	0:3:2:1
$\gamma_1$	3:4:4:2
$\gamma_2$	4:2:0:4
$\gamma_3$	0:5:0:0

**Table.3-2** Amino acids composition designed for “MLSDS Natural”

	Ala	Gly	Leu	Ser	Lys	Val	Glu	Thr	Asp	Arg	Pro	Ile	Asn	Gln	Tyr	Phe	Cys	His	Met	Trp	TERM
Designed composition neglecting point deletions [%]	9.33	7.72	4.16	8.01	6.84	7.27	6.84	6.45	5.81	4.12	4.57	3.64	4.79	4.10	4.10	3.29	3.09	3.08	1.27	1.54	0.00
Designed composition regarding point deletion [%]	8.56	7.55	6.70	7.59	6.67	7.57	6.02	5.67	5.12	4.86	4.16	4.55	4.60	3.78	3.69	3.41	2.92	2.81	1.89	1.56	0.33
Target composition [%]	8.90	7.80	7.50	7.10	7.00	6.90	6.20	6.00	5.50	4.70	4.60	4.60	4.40	3.90	3.50	3.50	2.80	2.00	1.70	1.10	0.00

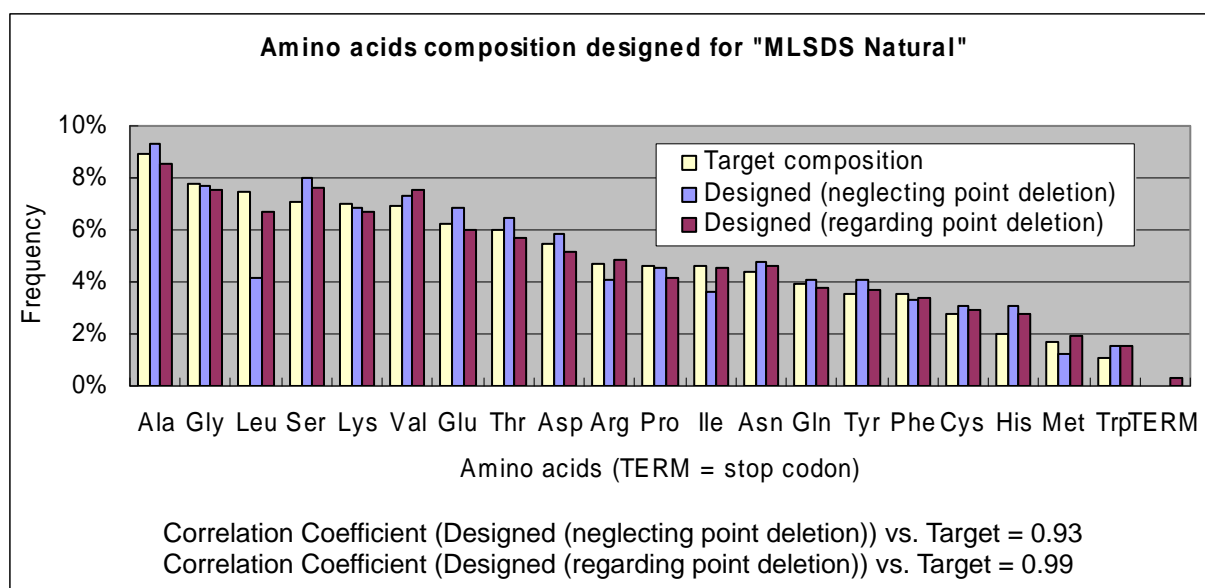


Fig.3-1 “MLSDS Natural”における目標アミノ酸組成及び最適化された ATGC 混合比によって導かれるアミノ酸組成

Point deletion を考慮しない最適化アミノ酸組成の目標組成に対する相関係数 0.93。

Point deletion を計算に入れた最適化アミノ酸組成の目標組成に対する相関係数 0.99。

## 合成結果

Table.3-1 の組成にて、MLSDS の合成サイクルを 15 回行い、両末端に制限酵素部位と PCR プライミング領域を含むランダム領域 45base(15a.a.)、準ランダム領域 6base(2a.a.)、全長 87base のライブラリーを合成した。

合成産物 51 サンプルをシーケンシングした。74.5%が完全長配列で、数 base 以上まとまって欠失が起こる cluster deletion を含むサンプルは 2.0%であった(Table.3-3)。完全長配列、point deletion を含む配列、シーケンシングした全配列のアミノ酸組成を Table.3-4、Fig.3-2 に示す。

cluster deletion や insertion 等、ATGC 比の最適化の際に計算に入れていない合成ミスが発生しているにも関わらず、シーケンシングした全配列によるアミノ酸組成に大きなズレは発生していなかった。point deletion を含む配列のアミノ酸組成、及びシーケンシングした全配列のアミノ酸組成の、完全長配列のアミノ酸組成に対する相関係数は各々 0.998 と 0.966 であり、数値の上でもフレームシフトによるアミノ酸組成への影響を抑えることに成功していることが確認できる

Table.3-3  
“MLSDS Natural”シーケンス結果

	Number of samples
Complete sequence	38 (74.5%)
1mer deletion	7 (14%)
2mer deletion	3 (5.9%)
3mer deletion	2 (3.9%)
cluster deletion	1 (2.0%)
1mer insertion	0 (0.0%)
total	51

**Table.3-4** Amino acids composition of synthesized “MLSDS Natural”

	Ala	Gly	Leu	Ser	Lys	Val	Glu	Thr	Asp	Arg	Pro	Ile	Asn	Gln	Tyr	Phe	Cys	His	Met	Trp	TERM
Only complete sequences [%]	11.44	11.97	3.70	8.10	6.16	6.87	6.87	5.81	8.63	3.52	2.46	3.87	3.52	2.11	4.40	4.05	2.46	1.23	1.06	1.76	0.00
Complete and single deleted sequences [%]	10.81	11.56	5.56	7.81	6.61	6.76	6.16	4.95	7.51	4.35	2.70	5.11	3.75	2.25	3.90	3.75	2.25	1.20	1.50	1.50	0.00
All sequences [%]	10.62	11.29	6.05	7.39	6.32	6.32	6.45	4.70	6.99	5.11	2.42	4.84	3.76	2.28	3.76	4.03	2.69	1.21	1.61	2.02	0.13

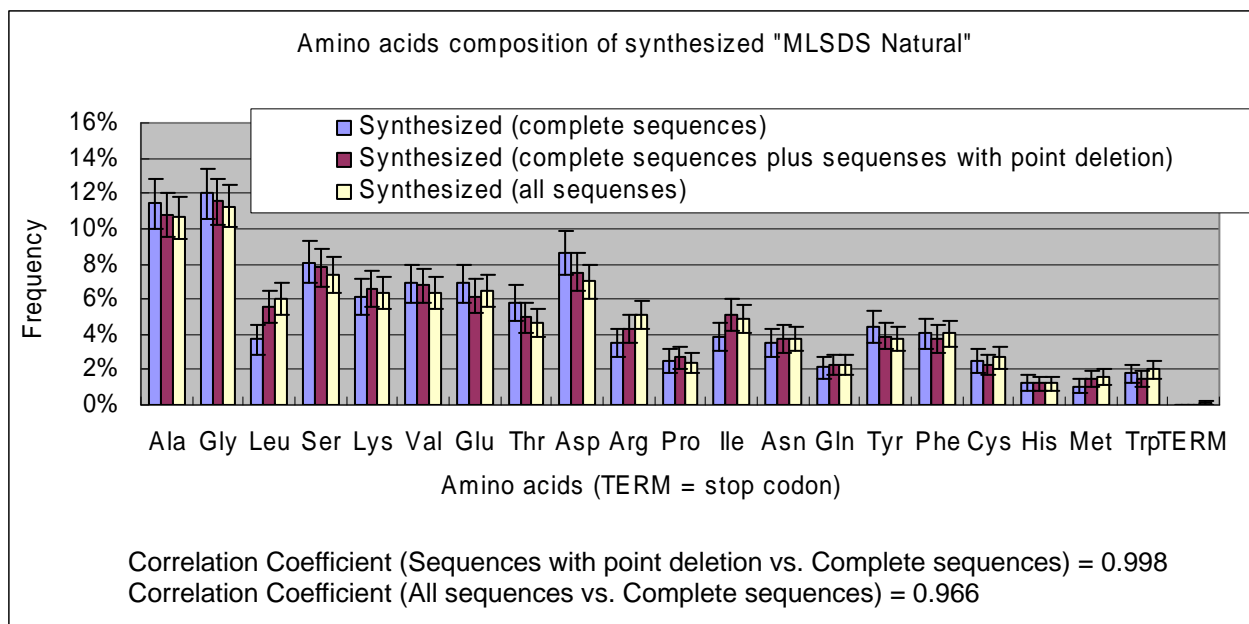


Fig.3-2 合成“MLSDS Natural”における合成エラーの影響

合成ライブラリーのアミノ酸組成、最適化アミノ酸組成、目標アミノ酸組成を並記したヒストグラムを Fig.3-3 に示す。完全長配列、point deletion を含む配列、シーケンシングした全配列のアミノ酸組成全てにおいて、設計組成との相関係数が 0.9~0.91 と高い値を示した。目標組成に対する相関係数も 0.85~0.91 と高い値を示した。

MLSDS はコドン単位での制御を行っている。従って、合成産物の評価には最適化したコドン組成と合成産物のコドン組成を比較すべきである。合成ライブラリーのコドン組成と、最適化コドン組成を並記したヒストグラムを Fig.3-4 に示す。完全長配列、point deletion を含む配列、シーケンシングした全配列のコドン組成全てにおいて、設計組成との相関係数が 0.92 と高い値を示し、コドン単位での制御に成功していることがわかる。

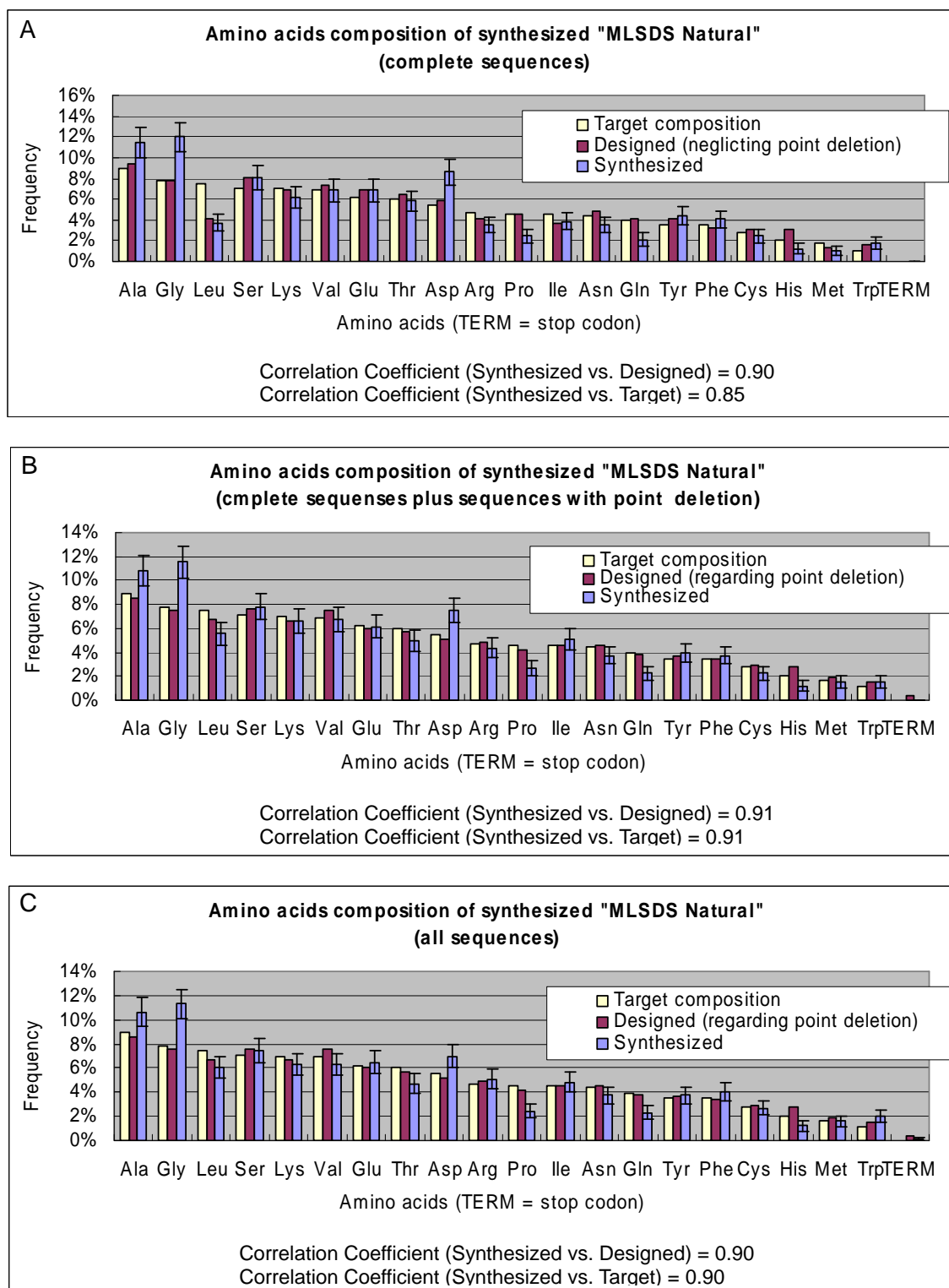


Fig.3-3 合成"MLSDS Natural"ライブラリーのアミノ酸組成、及び最適化、目標アミノ酸組成

- A. Point deletion を含まないライブラリーのみの実測アミノ酸組成、及び Point deletion を考慮しない最適化アミノ酸組成と目標アミノ酸組成。最適化組成及び目標組成に対する実測アミノ酸組成は各々0.90、0.85。
- B. Point deletion によりフレームシフトした結果得られるアミノ酸組成を含む実測組成、及び Point deletion を考慮した最適化アミノ酸組成と目標アミノ酸組成。最適化組成及び目標組成に対する実測アミノ酸組成は各々0.91、0.91。
- C. 配列決定した全ての配列から算出したアミノ酸組成、及び Point deletion を考慮した最適化アミノ酸組成と目標アミノ酸組成。最適化組成及び目標組成に対する実測アミノ酸組成は各々0.90、0.90。

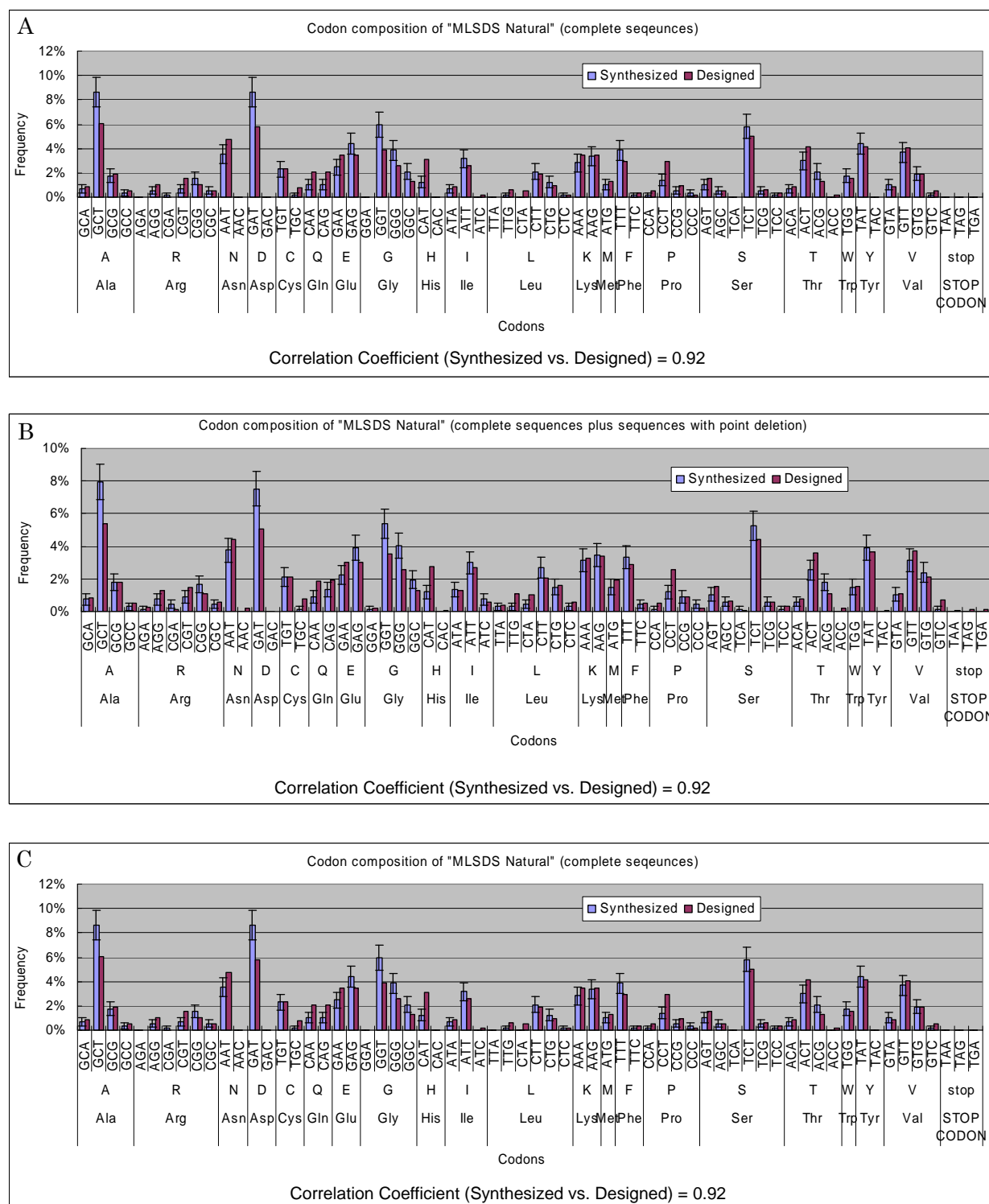


Fig.3-4 合成“MLSDS Natural”ライブラリーのコードン組成、及び最適化コードン組成

- A. point deletion を含まないライブラリーのみの実測コードン組成と、point deletion を考慮しない最適化コードン組成。最適化組成に対する実測コードン組成の相関係数は 0.92。
- B. point deletion によりフレームシフトした結果得られるコードン組成を含む実測組成、及び point deletion を考慮した最適化コードン組成。最適化組成に対する実測コードン組成の相関係数は 0.92。
- C. 配列決定した全ての配列から算出したコードン組成と、point deletion を考慮した最適化コードン組成。最適化組成に対する実測アミノ酸組成の相関係数は 0.92。



### 3-1-2 “MLSDS-No Cys”

蛋白質内、蛋白質間でジスルフィド結合を形成する Cys の含有率を操作することは、蛋白質をデザインする上で興味深い。ジスルフィド結合を持たない機能性分子も進化工学的に得られている[18]。そこで、Cys の含有率を 0%としたライブラリーを作製した。ATGC 混合比の最適化条件として、1.ストップコドンの排除、2. Cys の出現率 0%、3. Cys を除く 19 種アミノ酸の出現率を均等化、4.rational deletion によるフレームシフト後のアミノ酸組成の影響を最小化の 4 条件を満たすように合成ライン 3 の MLSDS 法に基づき塩基混合比 Xn を最適化した(Table.3-5)。最適化は、Table.3-6, “Target composition”に対する相関係数を評価関数に適用した遺伝的アルゴリズムを用いて行った。最終的に得られた組成の目標組成との相関係数は、rational deletion による point deletion を考慮しない場合 0.93、考慮した場合 0.99 となり非常に高い近似配列を得ることが出来た。最適化アミノ酸組成と目標アミノ酸組成を Table.3-6 に、それらのヒストグラムを Fig.3-5 に示す。Point deletion を考慮しない場合はストップコドン(TERM)の出現を 0%に抑えることに成功している。Point deletion が発生した場合も組成全体の傾向を大きくずらすことなく、ストップコドンの発生率を最小限に抑えている

Table.3-5  
“MLSDS No Cys”の  
ATGC 混合比

ID	A:T:G:C
$\alpha_1$	2:0:3:3
$\alpha_2$	3:0:0:1
$\alpha_3$	5:3:0:0
$\beta_1$	5:4:4:0
$\beta_2$	0:3:3:1
$\beta_3$	0:0:1:0
$\gamma_1$	2:3:1:1
$\gamma_2$	5:5:0:3
$\gamma_3$	0:2:0:5

**Table.3-6** Amino acids composition designed for “MLSDS No Cys”

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	TERM
Designed composition neglecting point deletions [%]	5.69	5.49	6.01	5.35	0.00	5.86	5.86	4.40	5.35	3.66	6.23	3.91	5.49	5.49	4.22	4.76	6.11	4.40	5.49	6.23	0.00
Designed composition regarding point deletion [%]	5.02	5.40	6.04	4.71	0.73	5.52	5.17	5.58	4.97	4.36	6.38	4.51	5.24	5.01	4.37	5.34	6.47	4.32	4.81	5.50	0.54
Target composition [%]	5.26	5.26	5.26	5.26	0.00	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	0.00

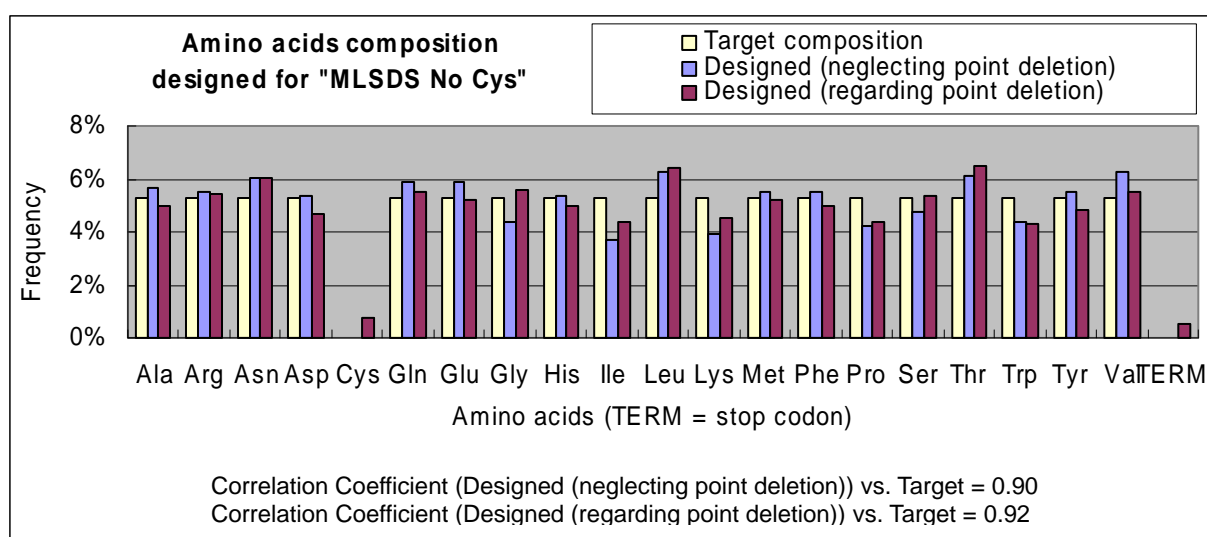


Fig.3-5 “MLSDS No Cys”における目標アミノ酸組成及び最適化された ATGC 混合比によって導かれるアミノ酸組成

Point deletion を考慮しない最適化アミノ酸組成の目標組成に対する相関係数 0.90。

Point deletion を計算に入れた最適化アミノ酸組成の目標組成に対する相関係数 0.92。

## 合成結果

Table.3-5 の組成にて、MLSDS の合成サイクルを 15 回行い、両末端に制限酵素部位と PCR プライミング領域を含むランダム領域 45base(15a.a.)、準ランダム領域 6base(2a.a.)、全長 87base のライブラリーを合成した（配列の詳細は”材料・DNA オリゴマー”を参照）。

合成産物 37 サンプルをシーケンシングした。45.9%のサンプルが完全長配列で、cluster deletion を含む配列は 27.0%だった(Table.3-7)。完全長配列、point deletion を含む配列、シーケンシングした全配列のアミノ酸組成を Table.3-8、Fig.3-6 に示す。

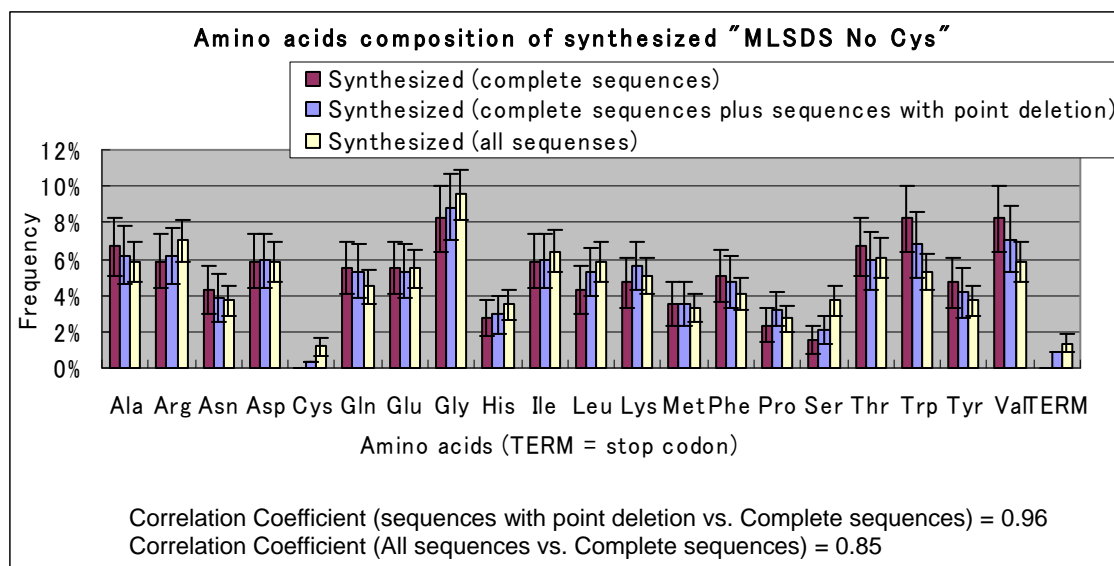
Cluster deletion や insertion 等、ATGC 比の最適化の際に計算に入れていない合成ミスが発生しているにも関わらず、シーケンシングした全配列によるアミノ酸組成に大きなズレは発生していなかった。Point deletion を含む配列のアミノ酸組成、及びシーケンシングした全配列のアミノ酸組成の、完全長配列のアミノ酸組成に対する相関係数は各々0.96 と 0.85 であり、数値の上でもフレームシフトによるアミノ酸組成への影響を抑えることに成功していることが確認できる。

Table.3-7  
“MLSDS No Cys”シーケンス結果

	Number of samples
Complete sequences	17 (45.9%)
1mer deletion	6 (16.2%)
2mer deletion	3 (8.1%)
3mer deletion	0 (0.0%)
cluster deletion	10 (27.0%)
1mer insertion	1 (2.7%)
total	37

**Table.3-8** Amino acids composition of synthesized “MLSDS No Cys”

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	TERM
Only complete sequences [%]	6.67	5.88	4.31	5.88	0.00	5.49	5.49	8.24	2.75	5.88	4.31	4.71	3.53	5.10	2.35	1.57	6.67	8.24	4.71	8.24	0.00
Complete and single deleted sequences [%]	6.19	6.19	3.83	5.90	0.29	5.31	5.31	8.85	2.95	5.90	5.31	5.60	3.54	4.72	3.24	2.06	5.90	6.78	4.13	7.08	0.88
All sequences [%]	5.85	7.02	3.70	5.85	1.17	4.48	5.46	9.55	3.51	6.43	5.85	5.07	3.31	4.09	2.73	3.70	6.04	5.26	3.70	5.85	1.36



**Fig.3-6** 合成“MLSDS No Cys”における合成エラーの影響

合成ライブラリーのアミノ酸組成、最適化アミノ酸組成、目標アミノ酸組成を並記したヒストグラムを Fig.3-7 に示す。完全長配列、point deletion を含む配列、シーケンシングした全配列のアミノ酸組成全てにおいて、設計組成との相関係数が 0.62~0.63 と低い値を示した。目標組成に対する相関係数も 0.60~0.67 と低い値を示した。

合成ライブラリーのコドン組成と、最適化コドン組成を並記したヒストグラムを Fig.4-8 に示す。完全長配列、point deletion を含む配列、シーケンシングした全配列のコドン組成全てにおいて、設計組成との相関係数が 0.87~0.88 と比較的高い値を示した。

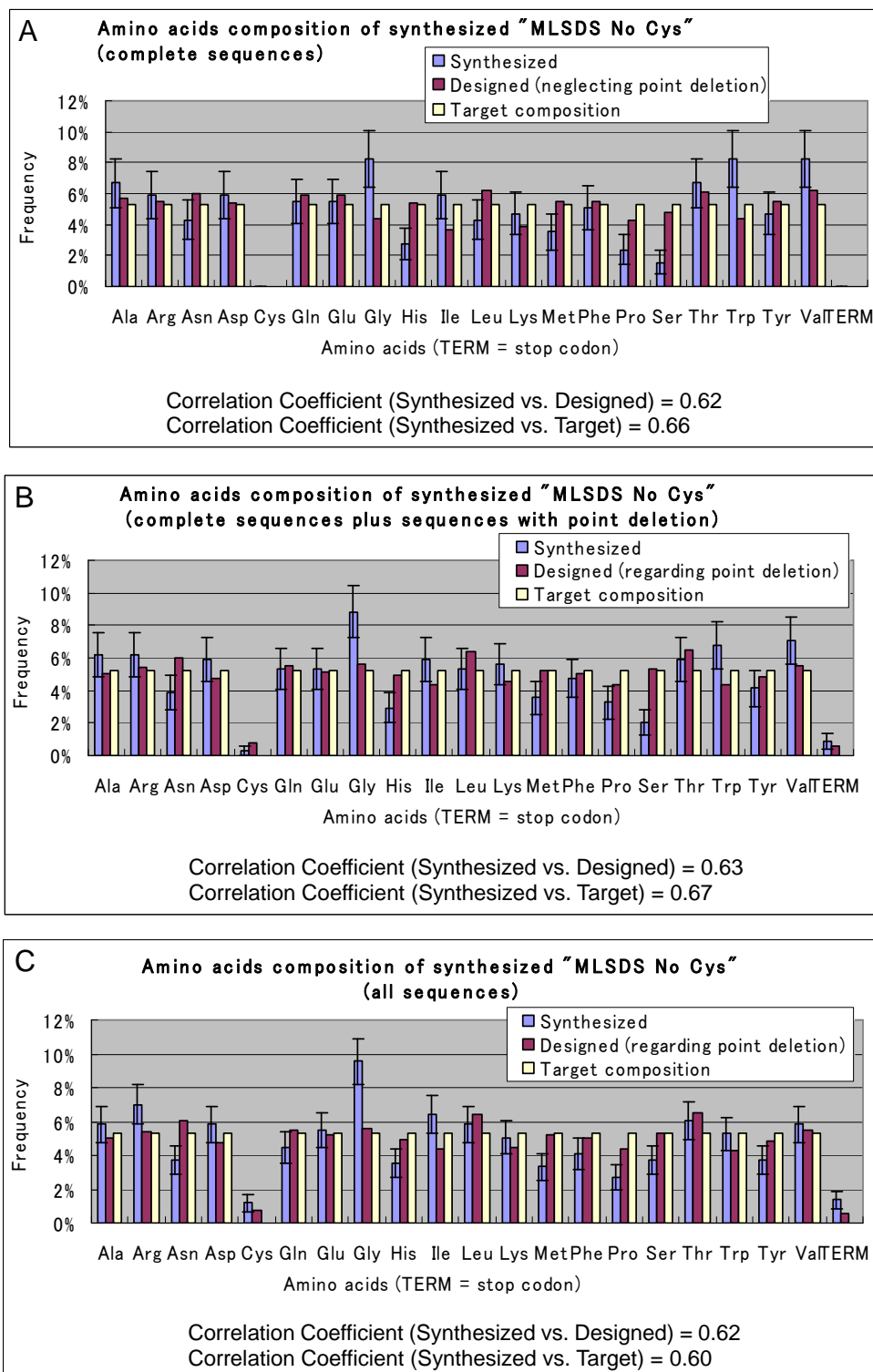


Fig.3-7 合成"MLSDS No Cys"ライブラリーのアミノ酸組成、及び最適化、目標アミノ酸組成

- A. Point deletion を含まないライブラリーのみの実測アミノ酸組成、及び point deletion を考慮しない最適化アミノ酸組成と目標アミノ酸組成。最適化組成及び目標組成に対する実測アミノ酸組成は各々0.62、0.66。
- B. point deletion によりフレームシフトした結果得られるアミノ酸組成を含む実測組成、及び point deletion を考慮した最適化アミノ酸組成と目標アミノ酸組成。最適化組成及び目標組成に対する実測アミノ酸組成は各々0.63、0.67。
- C. 配列決定した全ての配列から算出したアミノ酸組成、及び point deletion を考慮した最適化アミノ酸組成と目標アミノ酸組成。最適化組成及び目標組成に対する実測アミノ酸組成は各々0.62、0.60。

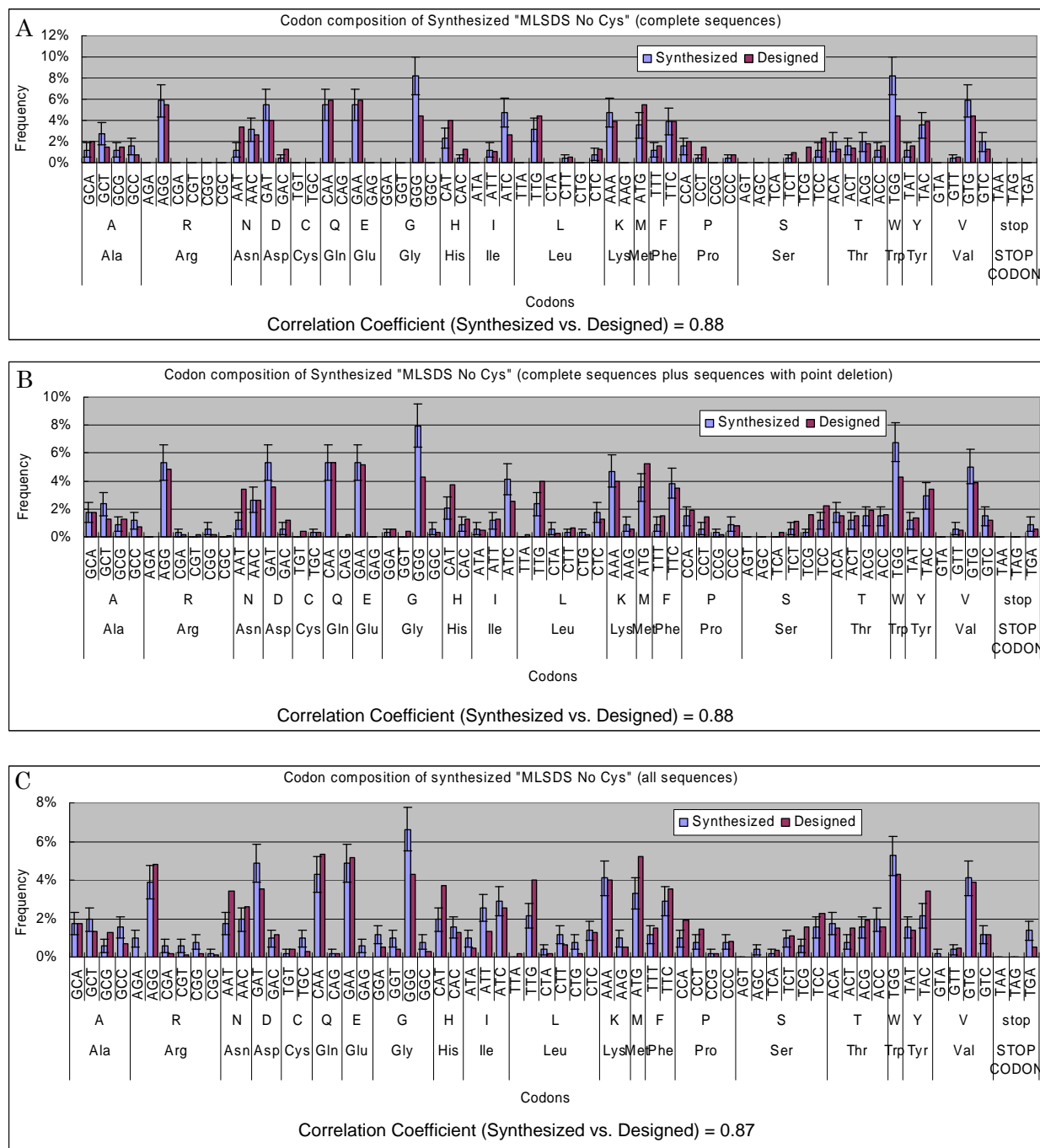


Fig.3-8 合成"MLSDS No Cys"ライブラリーのコドン組成、及び最適化コドン組成

- point deletion を含まないライブラリーのみの実測コドン組成と、point deletion を考慮しない最適化コドン組成。最適化組成に対する実測コドン組成の相関係数は 0.88。
- point deletion によりフレームシフトした結果得られるコドン組成を含む実測組成、及び point deletion を考慮した最適化コドン組成。最適化組成に対する実測コドン組成の相関係数は 0.88。
- 配列決定した全ての配列から算出したコドン組成と、point deletion を考慮した最適化コドン組成。最適化組成に対する実測アミノ酸組成の相関係数は 0.87。

### 3-1-3 “MLSDS Natural”, “MLSDS No Cys” 総合

合成ライブラリーのアミノ酸組成と、目標アミノ酸組成及び最適化アミノ酸組成の相関係数は、“MLSDS Natural”の場合は 0.85～0.91 と高い値を示したが、“MLSDS No Cys”は 0.60～0.67 と低い値になった。これは相関係数が比較する数列の分散が低いと高い値を出しにくいという性質によるものと思われる。実際“MLSDS Natural”の最適化アミノ酸組成と合成産物の完全長アミノ酸組成の標準偏差は各々 2.4%と 3.3%なのに対し、“MLSDS No Cys”は、1.8%と 2.4%と低い値をとる。そこで、偏りの無い評価を行うため、相関係数とは別の 2 種類の方法で評価を行った。即ち、各アミノ酸(コドン)ごとに、合成産物の実測値から最適化値を引いたものの絶対値を取り、平均を取った値。もうひとつは、各アミノ酸(コドン)ごとに、合成産物の実測値を最適化値で割ったものの絶対値を取り、平均を取った値の 2 つである。前者は“0”に近い程、後者は“1”に近い程、品質の高いライブラリーだといえる。この 2 種の評価値を、“MLSDS Natural”、“MLSDS No Cys”の完全長産物のみに対して算出した。アミノ酸組成に対して Table.3-9 に、コドン組成に対して Table. 3-10 に示す。



Table.3-9 The other values to assessment synthesized amino acids composition.

	“MLSDS Natural” (complete sequences)	“MLSDS No Cys” (complete sequences)
Average of   Synthesyzed – Target   [%]	1.28	1.34
Average of   Synthesyzed – Designed   [%]	1.03	1.44
Average of   Synthesized/Target	0.91	0.93
Average of   Synthesized/Designed	1.09	1.00

Table.3-10 The other values to assessment synthesized codon composition.

	“MLSDS Natural” (complete sequences)	“MLSDS No Cys” (complete sequences)
Average of   Synthesyzed – Designed   [%]	0.50	0.57
Average of   Synthesized/Designed	0.90	0.97

アミノ酸組成、コドン組成ともに、“差の絶対値の平均” に関しては、“MLSDS Natural”のほうが優れた値を示したが、“比の絶対値の平均”では、“MLSDS No Cys”の方が優れた値を示した。従って、“MLSDS Natural”も、“MLSDS No Cys”もともに望んだアミノ酸(コドン)組成を持つ、優れたライブラリーであるといえる。

### 3-1-4 他のライブラリー作成法との比較

3NPs 法、mRNA display 法との比較を Table.3-11 に示す。

Table.3-11 The difference between of the target amino acid composition and the actual composition for various libraries.

Method	Uniform Type		
	3NPs	mRNA display	MLSDS
Reference, Library	[16], “96T λ 6”	[12], “Random”	This work, “No Cys”
Target value	Equimolar 19 amino acids	Equimolar 20 amino acids	Equimolar 19 amino acids
Correlation coefficient <sup>a</sup>	0.43	0.56	0.66
Average of   Synthesized – Target   [%]	2.6	1.1	1.34
Average of   Synthesized/Target	0.99	1.00	0.93
percentage of stop codon	No Data	0.46	0.00
percentage of cassettes containing stop codon	No Data	8.33	0.00

The values shown were calculated using the context of full-length libraries. The values of 3NPs and RNA display method were estimated with some assumptions. In mRNA display method, the actual composition including stop codon’s contribution were estimated from data provided in the paper and some assumptions, first assumption is that one cassette couldn’t have two or more stop codons, second is that stop codons appeared equally in every positions, third is that stop codons were not shown in the flanking region. And it is assumed that stop codons were not appeared in full-length libraries of 3NPs method.

<sup>a</sup> Correlation coefficient is calculated using target and actual amino acid composition including stop codon’s contribution.

相関係数は”MLSDS No Cys”が最も優れている。“差の絶対値の平均”は MLSDS 方と mRNA display 法の差は僅かだが、3NPs 法に対して優れた値を示している。“比の絶対値の平均”は全ての方法でほぼ差が無い。ストップコドンの排除では、MLSDS 法は mRNA display 法と比較して非常に優れていることがわかる。

### 3-2 ライブラリーの長鎖化

2 倍長(ランダム領域 102 ( $= (45+6) \times 2$ ) bp、固定領域 36bp)、4 倍長(ランダム領域 204( $51 \times 4$ )bp、固定領域 36bp)、8 倍長(ランダム領域 408( $51 \times 8$ )bp、固定領域 36bp)の長鎖化に成功した(fig.4-11)。配列多様性  $10^{12}$  分子の 1 倍長ライブラリーから開始し、8 倍長ライブラリーを作製する過程で配列多様性は  $10^9$  分子まで減少した。この原因は主にライゲーション反応、及びライゲーション後の精製過程でのロスによるものである。しかし、制限酵素処理前の PCR による増幅率を制御することによって、このロスを補填することが出来るため、理論的には配列多様性の減少を抑えることは可能である。長鎖化に用いる分子の多様性は反応スケールに依存している。現実的に可能なスタートライブラリーの最大分子多様性は  $10^{14}$  分子程度であると考えられる。

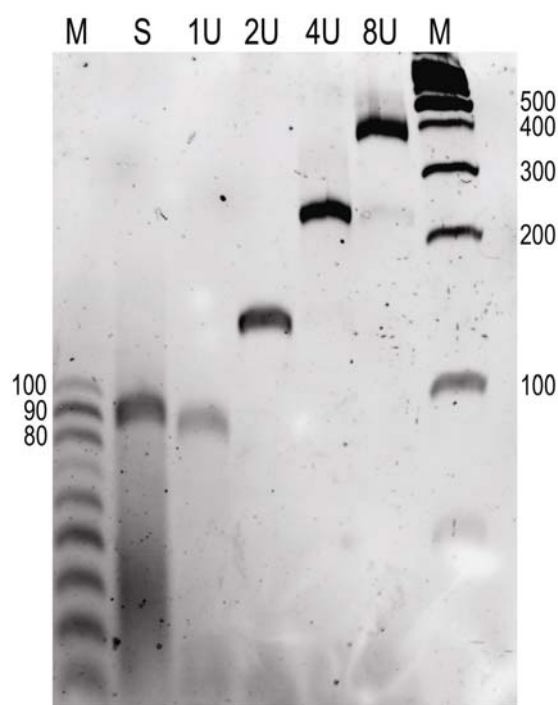


Fig.3-11 2,4,8倍長ライブラリの作製

M:サイズマーカー

S:合成ライブラリー (87base)

1U: 1 倍長ライブラリー (84bp)

2U: 2 倍長ライブラリー (138bp)

4U: 4 倍長ライブラリー (240bp)

8U: 8 倍長ライブラリー (444bp)

### 3.3 Conclusion

MLSDS 法によってアミノ酸組成を高度に制御した DNA ライブラリーの作製が可能であることを実際に示した。これにより、配列空間の戦略的探索が可能となり、スクリーニング効率を格段に上昇させることができる。

また、短鎖ライブラリーを連結することによって高い品質を保ったまま長鎖ライブラリーを作製することに成功した。即ち、高品質長鎖ライブラリーを用意することが可能となり、より高機能な蛋白質を探索する可能性が開けた。

本研究の成果は、進化蛋白質工学の発展に大きく寄与するだろう。

### References

1. Hoogenboom, H. R.: Designing and optimizing library selection strategies for generation high-affinity antibodies. *Trends. Biotechnol.*, **15**, 62-70 (1997)
2. McCafferty, J., Johnson K. S.: Construction and screening of antibody display libraries. in *Phage Display of Peptides and Proteins* (eds. Kay, B. K., Winter J. et al.), pp79-110, Academic Press (1996),
3. Cwirla, S. E., Peters, E. A., Barrett, R. W. and Dower, W.J.: Peptides on phage: a vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. USA* **87**, 6378-6382 (1990)
4. Kay, B. K., Adey, N. B., He, Y. S., Manfredi, J. P., Mataragnon, A. H. and Fowlkes, K. M.: An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene* **128**, 59-65 (1993)
5. Larry C. Mattheakis, Ramesh R, Bhatt, and William J. Dower: An in vitro polysome display system for identifying ligands from very large peptide libraries, *Proc. Natl. Acad. Sci. USA*, **91**, 9022-9026 (1994)
6. Cwirla, S. E., Peters, E. A., Barrett, R. W. and Dower, W.J.: Peptides on phage: a

- vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. USA* **87**, 6378-6382 (1990),
7. Scott, J. K. and Smith, G. P.: Searching for peptide ligands with an epitope library. *Science* **249**, 386-390 (1990),
  8. Sondek, J., Shortle, D.: A general strategy for random insertion and substitution mutagenesis: substoichiometric coupling of trinucleotide phosphoramidites. *Proc Natl Acad Sci U S A.* **89**, 3581-3585 (1992)
  9. Virnekas, B., Ge, L., Pluckthun, A., Schneider, K.C., Wellnhofer, G., Moroney, S.E.: Trinucleotide phosphoramidites: Ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res.* **22**, 5600-5607 (1994)
  10. Kayushin, A.L., Korosteleva, M.D., Miroshnikov, A.I., Kosch, W., Zubov, D., Piel N.: A convenient approach to the synthesis of trinucleotide phosphoramidites--synthons for the generation of oligonucleotide/peptide libraries. *Nucleic Acids Res* **24**, 3748-3755 (1996)
  11. Ono, A., Matsuda, J., Zhao, J., Santi, D.V.: The synthesis of blocked triplet-phosphoramidites and their use in mutagenesis. *Nucleic Acids Res.* **23**, 4677-4682 (1995)
  12. Cho, G., Keefe, A.D., Liu, R., Wilson, D.S., Szostak, J.W.: Constructing high complexity synthetic libraries of long ORFs using in vitro selection. *J. Mol. Biol.* **297**, 309-319 (2000)
  13. 上野真吾 ランダム DNA ライブラリーの作成とその in vitro virus への応用 卒業論文 (2002)
  14. 上野真吾、田淵一郎、空本清香、深石圭、新井秀直、伏見譲 生物物理 **43**(suppl.1), B219(2003)
  15. Soramoto, S., Ueno, S., Tabuchi, I., Husimi, Y., *Genome Informatics* **13**, 527-528( 2002)
  16. Achi Knappik, Liming Ge , Annemarie Honegger, Peter Pack, Melanie Fischer., Gunter Wellnhofer , Adolf Hoess, Joachim Wolle, Andreas Pluckthun and Bernhard Virneka: Fully Synthetic Human Combinatorial Antibody Libraries (HuCAL) Based on Modular Consensus Frameworks and CDRs Randomized with Trinucleoties. *J. Mol. Biol.* **296**, 57-86 (1999)
  17. Klapper, M.H.: The independent distribution of amino acid near neighbor pairs into

polypeptides. *Biochem. Biophys. Res. Commun.* **78**, 1018-1024 (1977)

18. Proba, K., Worn, A., Honegger, A., Pluckthun, A.: Antibody scFv fragments without disulfide bonds made by molecular evolution. *J. Mol. Biol.* **275**, 245-253 (1998)

## 総括

本研究で開発したフリーのカルボキシル末端を提示する新しい遺伝子型-表現型対応付け手法は、混合状態から配列特異的に目的のペプチドを濃縮することに成功し、本手法を用いたペプチドのスクリーニングが可能であることが実証された。

この手法は、開始コドンの直ぐ下流にアンバーコドン配している為、sup tRNA が取り込まれない限り、蛋白質は生成されないため、スクリーニング時における競合タンパク質が存在しないことになる。これは将来、ワンポッドでの自律型自己増殖系を構築する際に非常に重要な役割を果たすと考えられる。

また、自律型自己増殖系を構築する際、核酸分子の増幅法として現在の PCR に換わり主として等温増幅が可能である転写反応が採用されることが考えられる。その際、蛋白質が mRNA の 5'末端側で連結されていることにより 3'末端がフリーになることから、DNA 依存 RNA ポリメラーゼによる run off 転写を可能とするだけでなく、Q $\beta$  レプリカーゼ等の 3'末端から複製を開始する RNA 依存 RNA ポリメラーゼを適用することも可能になることから、自律型自己増殖系を構築する際本手法が有効であると考えられる。

本手法は、ペプチドの N 末端側で mRNA と連結するようにデザインされている。しかしながら、sup tRNA が取り込まれるアンバーコドンの位置を変えることにより、mRNA との連結位置を自由に選択することが可能かもしれない。従って、N 末端、C 末端ともにフリーとすることなども可能となる。

また、ランダム DNA ライブラリー中のアミノ酸出現頻度を制御できる MLSDS 法の開発にも成功した。本手法により、目的とする機能分子ごとにライブラリーのアミノ酸組成を最適化することによって、広大なアミノ酸配列空間を効率よく探索することが可能となり、進化実験の時間の短縮につながる。

以上のように、本研究で開発した、新しい遺伝子型-表現型対応付け手法とランダム DNA ライブラリー作製法は、今後の進化蛋白質工学の発展に少なからず寄与するものと考えられる。

## 謝辞

本研究を進めるにあたり、直接御指導を賜りました埼玉大学工学部教授 伏見譲博士に深く感謝いたします。そして、研究活動を行う上で貴重な御指導、御助言を頂きました、埼玉大学工学部教授 西垣功一博士、埼玉大学工学部助教 鈴木美穂博士に深く感謝いたします。

また、共同研究者として御指導頂きました、田淵一郎 博士、空本清香氏、共に研究を進めてきた深石圭、新井秀直 両氏、そして、伏見・西垣研究室の関係者皆様に心より感謝いたします。



## 要旨

進化蛋白質工学において無細胞翻訳系を用いる *in vitro* selection は、大きな多様性が確保でき、細胞毒性を持つ蛋白質の進化も可能である。しかしながら、ribosome display や *in vitro* virus (mRNA display) 法等、蛋白質の *in vitro* selection 手法の殆どは提示する蛋白質の C 末端側に遺伝子型分子を結合させる手法である。従って機能部位が C 末端側に存在したり、フリーの C 末端を機能上要求する蛋白質の進化には向いていない。

そこで sup tRNA を利用した蛋白質への非天然アミノ酸導入法にヒントを得て、新しい手法を開発した。mRNA の開始コドンから数コドン下流に終止コドンを配置し、アミノアシル化した sup tRNA のアミノ酸側鎖を、リンカーを介して mRNA の 5'末端に結合させる。このコンストラクトを無細胞翻訳系に投入することにより、mRNA と連結したアミノアシル sup tRNA が 5'側の終止コドンで取り込まれ、蛋白質の N 末端側と mRNA の 5'末端が連結される。

His-Tag 配列を含む上記 mRNA コンストラクトを *E.coli* S30 extract に投入し、その翻訳産物を Ni-NTA でスクリーニングし、RT-PCR を行った結果、ネガティブコントロールとして Met 無し、及びアミノアシル sup tRNA を連結していない RNA 分子で翻訳したサンプルのスクリーニング産物と比較して有意に増幅率が高い結果が得られ、スクリーニングサイクルを 1 サイクル回すことに成功した。また、6xHis-tag 配列を含む分子と、FLAG-tag 配列を含む分子の混合物から、Ni-NTA を用いて、6xHis-tag 配列を含む分子の濃縮に成功した。同様に、anti FLAG M2 抗体を用いて FLAG-tag 配列を含む分子の濃縮にも成功した。そして、FLAG-tag 配列を含む mRNA コンストラクトを PURESYSTEM 無細胞翻訳系に投入し、その翻訳産物をウェスタンブロッティングで解析したところ、mRNA-Peptide 複合体の形成を確認した。

本手法によって、フリーの C 末端を機能上要求する蛋白質の進化が可能となり、従来の C 末端側対応付け手法ではカバーしきれなかった配列空間の探索が可能となる。

蛋白質の進化分子工学において、求める蛋白質ごとに最適な初期ペプチドライブラリを用意することは、進化実験を効率的に行う上で重要である。そこで、各アミノ酸出現頻度の一様化やストップコドンの排除、二次構造の形成に関わるアミノ酸の含有率等を操作できる MLSDS (Multi-Line Split DNA Synthesis) 法を開発した。MLSDS 法は DNA 合成機を複数台使用し、3 塩基(1 コドン)伸長することにより混合・分配を繰り返すコンビナトリアルケミストリー手法を用いて合成される。その際、各アミダイトの ATGC 混合比を多様に設定することによりアミノ酸(コドン)出現頻度を制御できる。

天然に存在する球状蛋白質の平均アミノ酸組成と同じ組成を持つもの、および、Cys の出現率をゼロとし、それ以外のアミノ酸を一様組成にしたものの、2 種のランダムライブラリー(ランダム部位 45base)を、遺伝的アルゴリズムを用いた ATGC 混合比最適化プログラムによって設計した。なお、これら 3 種はペプチド合成の早期終結を引き起こすストップコドンも排除されている。その結果、目標アミノ酸組成との相関係数が 0.9 以上のアミノ酸組成を持つライブラリーの設計に成功した。次に、それらライブラリーを実際に合成し、評価した。その結果、アミノ酸組成の実測値と設計値の相関係数が 0.8~0.9 程度となり、設計どおり合成できたことを確認した。

また、より高い機能を持つ蛋白質を得る為には、ある程度長いペプチド鎖を用意する必要がある。しかし、DNA の化学合成は 100~150base が合成限界である。そこで、合成したライブラリーユニット(87base)を PCR で二本鎖化し、制限酵素で切断・結合することにより 2~8 倍長 (ランダム部位 106~414base) の長鎖ランダムライブラリーを作成した。

本研究で開発した、新しい遺伝子型-表現型対応付け手法とランダム DNA ライブラリー作製法は、今後の進化蛋白質工学の発展に少なからず寄与するものと考えられる。