Title	分子進化工学を簡略化するための‘Head-to-Head' mRNA/cDNA ディスプレイとエレクトロスプレーマイクロアレイプラットフォームの開発
Author(s)	Sharma, Kirti
Citation	
Issue Date	2018-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/15535
Rights	
Description	Supervisor:高村　禅，マテリアルサイエンス研究科，博士

Development of 'Head-to-Head' mRNA/cDNA display and electrospray microarray platform for simplifying molecular evolutionary engineering

SHARMA Kirti

Japan Advanced Institute of Science and Technology

Doctoral Dissertation

Development of 'Head-to-Head' mRNA/cDNA display and electrospray microarray platform for simplifying molecular evolutionary engineering

SHARMA Kirti

Supervisor: Professor Yuzuru Takamura

School of Materials Science

Japan Advanced Institute of Science and Technology

September 2018

Development of 'Head-to-Head' mRNA/cDNA display and electrospray microarray platform for simplifying molecular evolutionary engineering

SHARMA Kirti

School of Materials Science, Japan Advanced Institute Of Science and Technology

In order to discover drugs for therapeutic and diagnostic purposes various affinity-based protein selection techniques has been developed till date such as phage display, ribosomal display, and mRNA display. With the recent advancements, a shift from affinity-based selection to function-based selection has been observed which includes single-molecule observation, genetic selection, and compartmentalization technologies. This work is a parallel study to develop two fundamental platforms to simplify molecular evolutionary engineering. On one side, 'Head-to-Head' (H2H) mRNA/cDNA display for high efficiency of genotype-phenotype yield is established whereas on the other side, electrospray microarray platform for larger library is developed.

H2H is an opposite link mode to unite genotype and phenotype. The key technique in H2H is the formation of a covalent bond between O^6-benzylguanine (BG) and O^6-alkylguanine-DNA alkyl transferase (AGT), where mRNA is linked to a nascent AGT via a BG-linker, resulting in a "(C-terminus) protein-BG-DNA linker-mRNA (5′-terminus)" conjugate. Thus, a head (N-terminus) to head (5′-terminus) linkage is formed (Fig. 1). Among the advantages of H2H, the generation of multi-copied proteins is the most promising and was proven to be possible owing to the restored stop codon, which had been intentionally removed in the conventional mRNA display. Another advantage is obviously having a free C-terminus of the protein, which can be used for modifications such as C-terminal methylation, α-amidation.

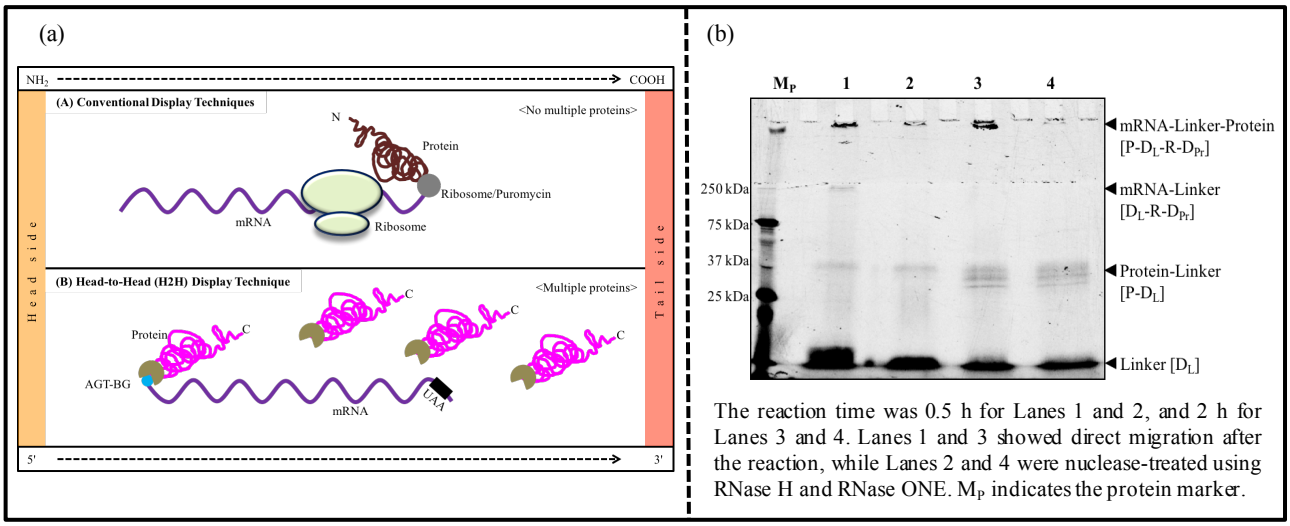


Figure 1: (a) Schematic drawing for the comparison between conventional display techniques (ribosome display and mRNA display) and novel 'Head-to-Head' display and (b) PAGE analysis for the formation of H2H conjugate. Lane 3 allows for an estimation based on the amount of fluorescent FITC, that around 20–30 % of the BG-linker (contained in the forms of Protein-Linker-mRNA plus Protein-Linker) could bind to AGT under these experimental conditions, since the addition of the bottom band of the Linker only accounts for the total amount of the BG linker input. This experiment indicates the generation of the H2H construct-holding phenotype-genotype conjugate at a yield of around 20% in a 2-h translation reaction.

Over the two decades, three main approaches have been widely used for IVC as i) Conventional methods using homogenizers and vortexing, ii) microfluidics-based approaches iii) use of single nozzle in electrospray. Each technique has its own disadvantages from polydispersity of droplet size to low-throughput generation to time consuming respectively, and thus, these approaches limit the application of IVC in the chemical and biological sciences. Electrospray incorporating single nozzle generates monodisperse and fL-sized droplets (B. Sharma et al., 2016) but it consumes hours-to-days for encapsulation of µl scale reaction with a limit of library size to 10^{8-9}. Hence, in this work I present a novel platform by integrating micro-hole chip with immersed electrospray for ultrahigh-throughput generation of highly monodisperse water-in-oil or agarose-in-oil compartments. The set-up of Micro-Hole Array Electrospray (µHAES) system is shown in Fig. 2(a) where a conductive aqueous solution is electrospray through an electrified micro-hole chip containing an array of 7x7 mm (24x16=384 holes) in an immiscible phase (mixture of oil and surfactant). A jet of water-in-oil droplets are obtained into the oil chamber when voltage of 1000 V was applied through the micro-hole array thereby generating bulk water-in-oil droplets as shown in Fig. 2(b). In this novel µHAES system, I not only can reduce the time of encapsulating cell-free reaction (10 µl) in 1fL IVC to just 5 minutes but can also increase library size to 10^{11-12} compared to 55 hours and 10^{8-9} library size of single nozzle electrospray therefore, establishing an ultrahigh-through w/o droplet generation platform in less time with high library size.

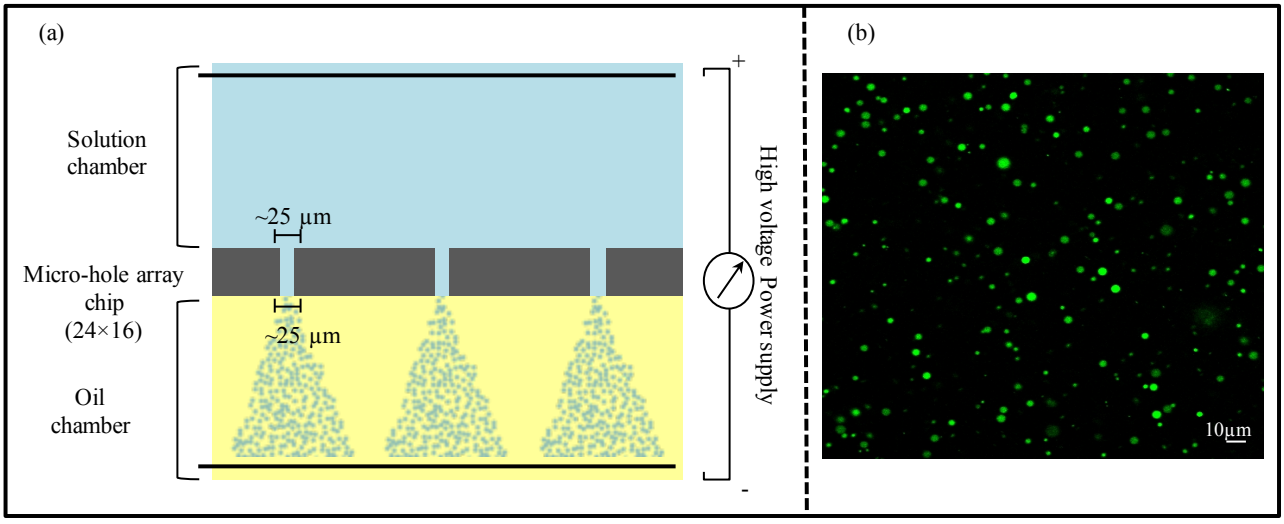


Figure 2: (a) The concept of Micro-Hole Array Electrospray (µHAES) system and (b) fluorescent image of agarose-in-oil droplets produced by µHAES.

In here, the two parallel fundamental techniques are successfully developed as 'Head-to-Head (H2H)' mRNA/cDNA display for high efficiency of genotype-phenotype yield and electrospray microarray platform for larger library size in the need to simplify molecular evolutionary engineering.

Keywords: 'Head-to-Head' (H2H) covalent linkage, Multi-copied proteins, C-terminal modifications, Micro-Hole Array Electrospray (µHAES) system, Library size.

PREFACE

The present dissertation is submitted for the Doctor of Materials Science at Japan Advanced Institute of Science and Technology. The dissertation is consolidation of results of the work on the topic "Development of 'Head-to-Head' mRNA/cDNA display and electrospray microarray platform for simplifying molecular evolutionary engineering" under the supervision of *Professor Yuzuru Takamura* at the School of Materials Sciences, Japan Advanced Institute of Science and Technology during October 2015 - September 2018.

Directed evolution is the mimic of natural evolution at small scale in laboratory (*in vitro*) to isolate the desired specific properties from the pool of "millions-to-quadrillion" molecules. It is basically based on the principles of diversification and selection of the evolved desired phenotype. The most suited screening or selection method are used to differentiate the best variant for desired traits amongst the millions of molecules. *In vitro* selection imitates Darwinian theory of the survival of the fittest at laboratory level. *In vitro* selection got a boost from the selection of nucleic acids (DNA and RNA) aptamers to the peptide based aptamer. But to select peptide based aptamers, peptide needs to be linked with its genotype as peptide itself cannot be amplified to get the genotypic information. At this stage, the role of display technology comes into the molecular evolutionary studies.

It has been more than 40 years, since the discovery of first display technique and yet no display method in itself enough to meet the demand of the growing need for drug discovery. The significance of directed molecular evolution has been so far well explained by the display technologies like phage display, ribosome display, mRNA display and *in vitro* compartmentalization (IVC). But all these technologies somewhere lack proper handling of "millions-to-quadrillion" molecules, resulting in search of new method. Phage display is a cell-based method, where library size is directly depended on the host cell. Weak physical bonding of the ternary complex (mRNA, linkage molecule and ribosome) and large size of ribosome leads to easy dissociation and uncontrolled inter molecular interactions resulting in misleading results in ribosome display. In mRNA display, the tether moiety, puromycin enters the P site of ribosome which results in pre-mature proteins and incorrect folding of proteins. In the past 10 years, IVC have been extensively used by entrapping gene and protein in one compartment to remove non-specific interaction between the surrounding biomolecules along with protection from degradation.

All the previous mentioned technologies are affinity-based selection. Affinity-based selection involves four main steps; the first, enrichment of protein or peptide from a pool of molecules; second, to recover the sequence information (DNA) linked to protein via cloning and sequencing; third, again convert DNA to its protein and fourthly, check the function of that enriched protein or peptide. Thus, making it long process and as multiple steps are involved, lot of information is lost. Why are we losing on the selected information? Instead, why not move towards developing function-based selection which can directly recover the enriched protein or peptide by checking its function on the first place thereby eliminating the need of above discussed second and third step. This will help to reduce the time involved in multiple step without much loss of the information obtained. Hence, giving a new direction to the molecular evolution studies.

If we specifically talk about IVC; yes, they are simple display techniques with lot of advantages over non-compartmentalized display techniques and multiple IVC technologies has been established with some or the other pros and cons by various groups of scientists all around the world but how to encapsulate the "million-to-quadrillion" library all together both in terms of time and capacity?

By studying the rise in the molecular evolution over the time, it has encouraged me to find solutions to the aforementioned questions. Let's shift from affinity-based selection to function-based selection in parallel dealing with the higher number of molecules together in short time. Thereafter, I started this research keeping in mind to answer these questions. By conducting this work I want to create a platform by developing not only function-based *in vitro* selection with improved Genotype-Phenotype (G-P) yield but also IVC-based *in vitro* selection for improved library size. I wish to play my part in filling the gap in between natural selection and *in vitro* selection to identify the desired properties for therapeutic and diagnostic purposes.

To the best of my knowledge, the work is original and no part of the thesis has been plagiarised.

Sharma Kirti

School of Materials Science
Japan Advanced Institute of Science and Technology
September 2018

GUIDE TO THE THESIS

My research work in doctoral course has led me to summarize the entire three years work in this thesis as five chapters including the background of the work in first chapter and final conclusions in last chapter. The three aspects of the research work has been presented in the main three chapters.

The results presented in this thesis have been published or under submission as:

Journal Paper-

'Head-to-Head' mRNA display for the translation of multi-copied proteins with a free C-terminus.

Kirti Sharma, Aya Hongo, Koichi Nishigaki, Yuzuru Takamura, Manish Biyani (*Analytical Biochemistry, Elsevier*; Volume 557, Pages 77-83, 2018).

Micro-hole array based electrospray in vitro compartmentalization for the bulk generation of functional beads.

Kirti Sharma, Yuzuru Takamura, Manish Biyani (To be submitted to APEX).

Patent-

1. 「略1fLカプセル化体の製造方法及びそれに用いる装置」:
   JP application number: 2017-202950
   Inventor: Manish Biyani, Yuzuru Takamura, Kirti Sharma, Phan Trong Tue, Noritaka Minami.
2. 略1fLカプセル化体の製造方法及びそれに用いる装置:
   PCT application (Approved by JST).

Conference Proceedings-

1. Kirti Sharma, Noritaka Minami, Phan Trong Tue, Yuzuru Takamura and Manish Biyani, "Electrospray microarray platform for bulk generation of highly monodisperse cell-like compartments" (Oral), The 79th JSAP Autumn Meeting-2018, September 18-21, 2018, Nagoya, Japan.

2. Kirti Sharma, Noritaka Minami, Phan Trong Tue, Yuzuru Takamura and Manish

Biyani, “Micro-hole array electrospray (μHAES) platform for ultrahigh-throughput generation of water-in-oil compartments” (Poster), JAIST Japan-India Symposium on Materials Science 2018 (JISMS-2018), March 5 and 6, JAIST, Japan.

3. Kirti Sharma, Koichi Nishigaki, Yuzuru Takamura, Manish Biyani, “Optimization of electrospray-based droplet display technology for compartmentalized *in vitro* selection of peptide aptamers” (Poster), The 30th International Microprocesses and Nanotechnology Conference (MNC 2017) November 6 - 9, 2017, JeJu, Korea.

4. Kirti Sharma, Koichi Nishigaki, Yuzuru Takamura, and Manish Biyani, “Encapsulated one-step genotype-phenotype covalent linkage in electrosprayed femtoliter compartments for on-demand *in vitro* selection” (Poster), The 21th International Conference on Miniaturized Systems for Chemistry and Life Sciences, μTAS2017 Conference - October 22-26, 2017 at Savannah, Georgia, USA.

5. Kirti Sharma, “On-site demonstration of Electrospray Technology” (Hands-on demonstration) in Workshop-9 “Functional droplets for femtoliter-scale synthetic biology” on 22 October, The 21th International Conference on Miniaturized Systems for Chemistry and Life Sciences, μTAS2017 Conference- October 22-26, 2017 at Savannah, Georgia, USA.

6. K. Sharma, K. Nishigaki, Y. Takamura and M. Biyani, “‘Head-to Head’ linkage in electrosprayed femtolitre droplets for systemic *in vitro* evolution of peptide aptamers” (Poster), JAIST Japan-India Symposium on Materials Science 2017, March 6 and 7, JAIST, Japan.

7. Kirti Sharma, Yuzuru Takamura, Koichi Nishigaki and Manish Biyani, “‘Head-to Head fusion’ and droplet display for simple and bias-free *in vitro* directed evolution” (Poster), The 39th Annual Meeting of the Molecular Biology Society of Japan, MBSJ 2016 Conference - 30 November - 2 December 2016 Pacifico Yokohama, Yokohama, Japan.

8. B. Sharma, K. Sharma, Y. Takamura and M. Biyani, “A bulk sub-femtoliter in vitro

compartmentalization system for minimal artificial cellular bioreactors" (Oral), The 20th International Conference on Miniaturized Systems for Chemistry and Life Sciences, μTAS2016 Conference - October 9-13, 2016 – Dublin, Ireland.

9. Kirti Sharma, Bineet Sharma, Yuzuru Takamura and Manish Biyani, "A simple and portable electrospray system for large-scale parallel in vitro compartmentalization" (4-minute Oral and Poster), 2016 Electrochemical Society Hokuriku Branch Conference - 15 September 2016 - University of Toyama, Toyama, Japan.

10. Sharma Kirti, Takamura Yuzuru, and Biyani Manish, "A novel technique of linking genotype and phenotype for directed molecular evolution" (Poster), Biowakate-2016, 第４回バイオ関連化学シンポジウム若手フォーラム 6 September 2016 – Kanazawa, Japan.

The key characteristics for *in vitro* evolution is to select the best functional candidate as per the tailor-demand. The various display technologies available are based on the coupling between the tail region of the gene with the tail region of the protein, freely exposing N-terminus without stop codon. Therefore, these techniques are not only time consuming but also cannot be used for C-terminus post-translational modification and generation of multiple functional unit. All these techniques are affinity-based selection. With time as the tremendous improvement has been seen from phage display to *in vitro* compartmentalization; we can say that by far, amongst the available display technologies, IVC has been well used and most highlighted. The conventional methods of IVC using homogenizers and microfluidics-based approaches have already been used to produce the cell like compartments, however, considerable variability in droplet size and low-yield of droplet numbers, respectively, with these approaches limit the application of IVC in the chemical and biological sciences. Hence, demanding a new and simple technique to preserve the initial encoding information with its nascent polypeptide along with multiple copies of proteins which can be directly used for checking their functions on the target.

This research is based on two fundamental works; the first, to establish platform for function-based *in vitro* selection with improved G-P yield and the second, IVC based selection for improved library size. In here, I report a novel 'Head-to-Head' (H2H) display

technology for multiple protein-recovery during *in vitro* protein evolution. H2H requires much shorter time due to novelty of the mechanism and is best suited for C-terminus post-translational modification. I am working on constructing a design for function-based *in vitro* selection. I expect this H2H covalent linkage will provide a more realistic approach to increase the efficiency and yield of genotype-phenotype linkage in molecular evolutionary engineering field. After developing H2H in tube, I applied the principles of electrospray (already established) to generate water-in-oil droplets and combined the concept of H2H in these droplets. Further, to accomplish the need of the hour of encapsulating large library, a system for multi electrospray has been established which I call as 'micro-hole array electrospray (μHAES) system' for ultralow-volume droplet generation. This function-based selection technique will open a new window in the field of evolutionary molecular engineering unlike that of the previous affinity-based selection.

In chapter I, I reviewed the common biomolecular display technologies and their recent advances in evolution based study. Advantages, disadvantages, and current issues of the display technique are carefully discussed. Based on such considerations, the objectives of this dissertation are proposed.

Chapter II is all the work performed to establish the function-based *in vitro* selection. For this purpose, I worked on covalently coupling the genotype with its phenotype in 'Head-to-Head' (H2H) direction i.e., fusing 5'-terminal of gene with N-terminus of protein. This construct is designed to make multiple copy number of the proteins making it a protein recoverable techniques. The yield of Genotype-Phenotype (G-P) was improved in H2H construct. H2H can be used as information-to-function unit.

Chapter III highlights the model experiments done to set a link between 'Head-to-Head' linkage and IVC. In this chapter, I have worked to extend the utility of H2H construct for electrospray-based *in vitro* compartmentalization demonstrating how H2H construct work as model experiments for one-step efficient coupling and hence, improve enrichment and recovery of proteins for *in vitro* selection.

Chapter IV proposes a novel IVC platform to maintain high number of library size using electrospray. I named this technique as 'micro-hole array electrospray (μHAES) system' for the generation of ultralow volume ultrahigh throughput water-in-oil droplets in few minutes.

Chapter V summarizes all the work done in this thesis and perspectives of the 'Head-to-Head' mRNA/cDNA display and electrospray microarray-based *in vitro* compartmentalization system for molecular evolutionary experiments.

ACKNOWLEDGEMENTS

Time flies in no time! Yes, I agree as it seems like yesterday when I started this adventure, which marks an important chapter in my professional and personal life. While, I alone, initiated this journey, my successful completion is far from being my own accomplishment, but rather for all the people that have encouraged, assisted and supported me along the way.

First and foremost, I would like to express my sincere gratitude towards my supervisor and committee chairman, *Professor Yuzuru Takamura*, who gave me the freedom and flexibility to be creative, in and outside of research, while also providing support as a great mental help. I am really horned to have been a member of his research group and to have worked with him. His advice and encouragement regarding this research are invaluable to me. It is a tremendous honour to work alongside such distinguished and dedicated scientist.

I would like to thank my internal committee members, *Professor Takahiro Hohsaka, Associate Professor Yuichi Hiratsuka* and *Associate Professor Kazuaki Matsumura* from School of Materials Science, JAIST. My special thanks go to *Associate Professor Miho Suzuki* from Saitama University as an external committee member. I thank all of them for their time and consideration in serving on my thesis committee.

I would like to extend my sincerest gratitude to minor research advisor *Professor Shinya Ohki* for giving permission to undergo practical training at his unit and also for his generous help and useful suggestions throughout the training period. I wish to recall my thankfulness to his lab membres for their warm support.

My special thank goes to *Research Associate Professor Manish Biyani* for his kind guidance, valuable suggestions and heartfelt encouragements throughout this work. He enlightened me from the very first day of this theme design and development. His guidance helped me in all the time of research and writing the thesis. His insightful comments and remarks has enhanced the quality of this dissertation from various perspectives. I know him from my Bachelors course as one of the directors of my previous institute but working under him as doctoral student helped me to grow both academically and personally. Through the past three years, I have gone through twists and turns as well as faced trials and tribulations. I was fortunate to have met him, who have listened, advised, tutored, supported and mentored me. His guidance has opened a new path in my life.

I take an opportunity to thank *Assistant Professor Phan Trong Tue* for welcoming, guiding me and for his fruitful comments. He has been supportive throughout my stay and for all the fun we had in lab.

Many thanks go to my lab members (especially, *Mr. Noritaka Minami* for fabricating micro-hole chip) for their camaraderie throughout my graduate school experience, especially for all of the hours of fun we have shared. I will always look back to my time in graduate school with fond memories because of them and to them I express my most sincere and heartfelt gratitude.

Finally, and most importantly, I express my deepest gratitude to my parents and dear ones, for their undying love and relentless encouragement at difficult times. If not for their help, I would never have ended up at JAIST.

Above all, I pay admiration to the almighty, who blessed me with patience and vigor to complete my doctoral degree successfully.

Sharma Kirti

School of Materials Science
Japan Advanced Institute of Science and Technology
September 2018

CONTENTS

Chapter IV: Development of electrospray microarray *in vitro* protein selection for improved library size

Chapter V: Conclusion

CHAPTER I

GENERAL INTRODUCTION AND DEMAND OF *IN VITRO* PROTEIN EVOLUTION

CHAPTER I

General introduction and demand of *in vitro* protein evolution

1.1. Directed Evolution

The change in the heritable characteristics in all forms of life (biological populations) over the several generations is called as the evolution[1, 2]. The theory of evolution is based on the idea that all species are related and gradually change over the time (Figure 1). Evolution gives rise to biodiversity at species level, at individual organism level and at the basic, molecular level[1]. It is basically repeated formation of new species, change within species and the loss of species over the time[3]. In the mid-19th century, the scientific theory of evolution was formulated by Charles Darwin called as natural selection. Individuals with characteristics which are best suited to their environment are more likely to survive and reproduce, thereby passing their genes on to their progeny whereas the individuals which are poorly adapted to their environment are less likely to survive and reproduce, thereby not passing their genes to the next generation. Thus, the individuals which survive, will gradually evolve over the time period.

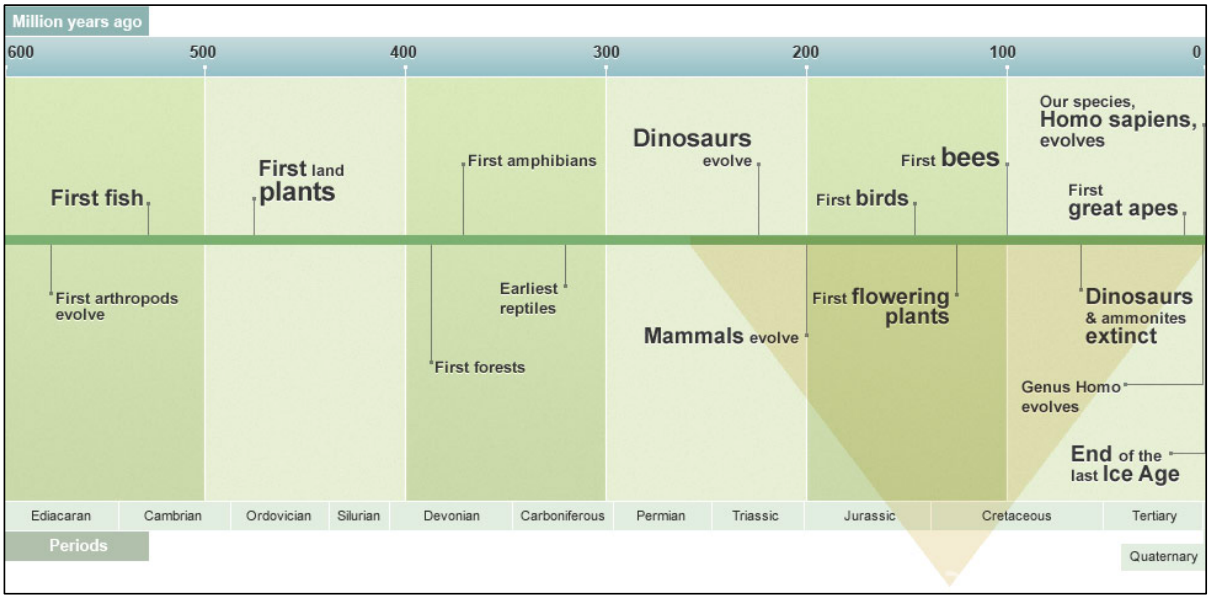


Figure 1: The timeline of the history of life on Earth. Picture credit- http://www.bbc.co.uk/nature/history_of_the_earth.

'Directed evolution' is biomolecular evolution for specific properties that mimics natural evolution at molecular level in laboratory as per tailor-demand[4]. The requirement of functional proteins for diagnostic or therapeutic purposes are increasing day by day and highly-functional biomolecules can be evolved *in vitro* by the process of directed molecular evolution in a matter of days[5]. *In vitro* evolution mimics repeated cycles of darwinian evolution (Figure 2) by subjecting a gene to iterative rounds of mutagenesis (diversification, creating a library of variants), selection (expressing the variants and isolating members with the desired function), and amplification (generating a template for the next round), on demand.

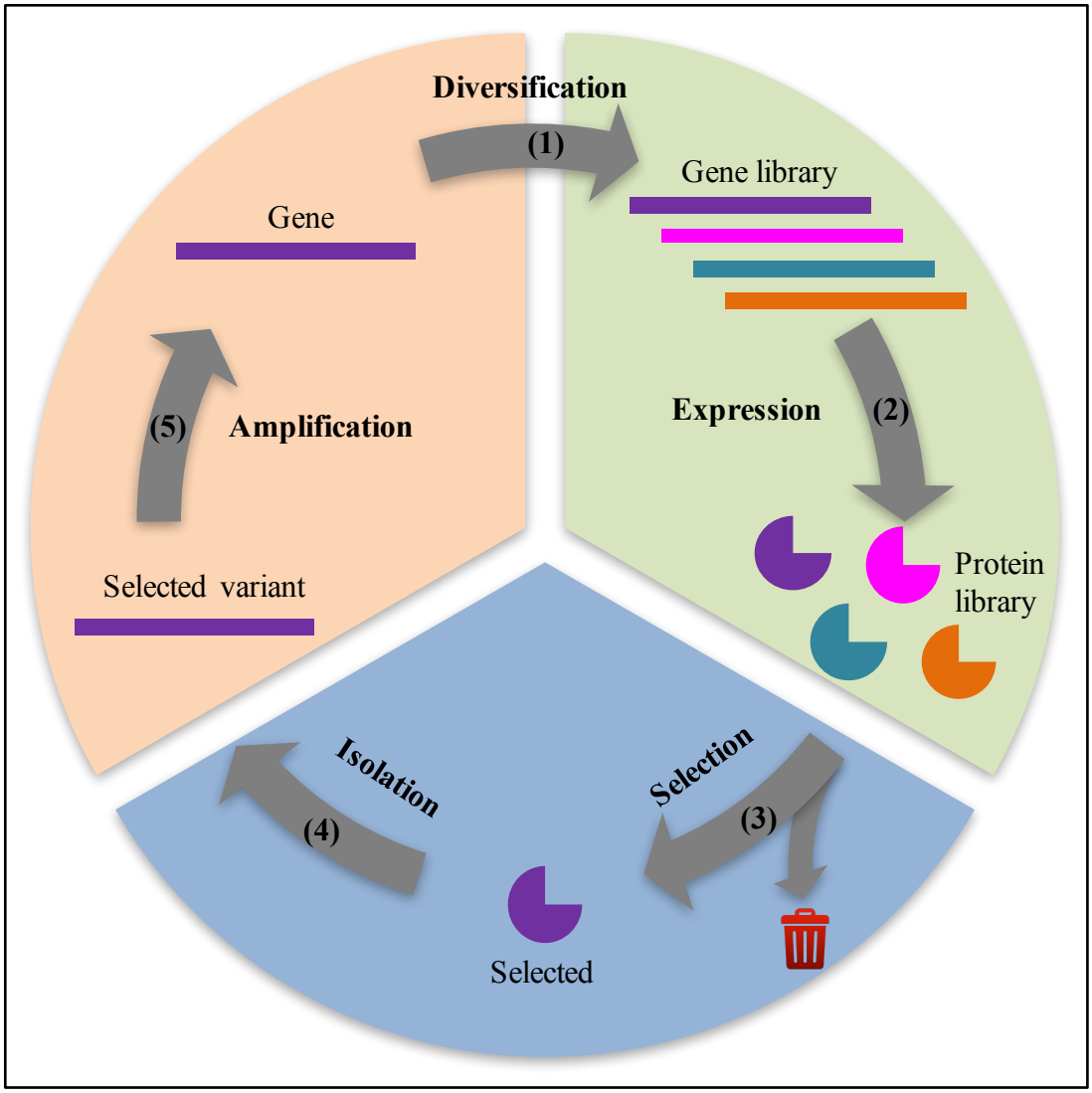


Figure 2: The cycle of directed evolution showing repeated cycles of diversification, selection and amplification.

Directed evolution can be performed *in vivo* i.e., in the living cells, or *in vitro* i.e., free in the solution or within the microdroplets.

1.2. Biomolecular Display Technology

The display or representation of molecules, mostly peptides and proteins, on the surface of a virus, a cell, or a molecular complex such as the ribosome or messenger RNA or complementary DNA or within the microdroplets is called as 'Biomolecular display technology'. For the enrichment of a molecule with the desired properties, biomolecular display techniques can be performed both in *in vivo* or *in vitro*. The biomolecular display technologies are a very valuable tool for the development of therapeutics with tissue-targeting and cell-targeting, for the isolation of affinity reagents with increased affinities and stabilities, and for the study of molecular interactions[6].

1.2.1. Need of Display Technology or Linkage

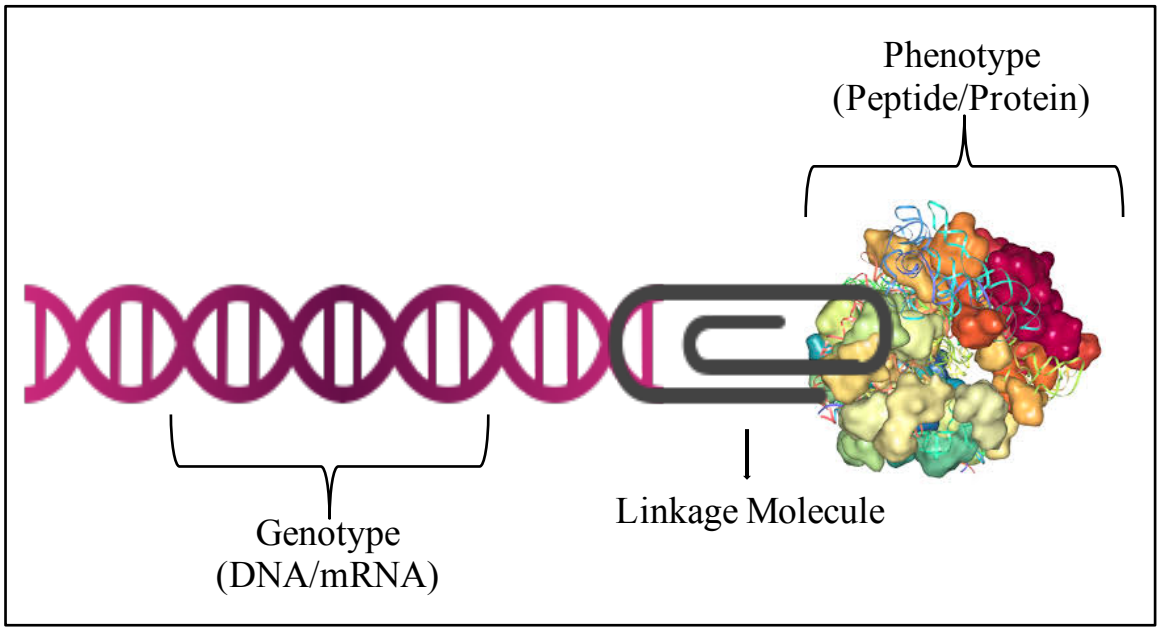


Figure 3: An illustration of a typical display module in the biomolecular display technologies. It consists of three major components as, the genetic code, linker molecule and displaying entity.

The display technologies established so far are broadly divided into two categories; firstly, based on synthetic generation of low-molecular weight compounds through combinatorial chemistry techniques and secondly, based on biological compounds such as nucleic acids, peptides or proteins[7]. Peptides are very useful as therapeutic and diagnostic

substances, their use is getting more popular, and display systems offer a useful way to engineer peptides and optimise their binding capabilities. The key practical aspect in directed evolution of functional biomolecules is the linkage of genotype and phenotype (Figure 3). 'Genotype' means the nucleic acid that can be replicated and 'phenotype' means the properties or functional unit, such as binding or catalytic activity. It is well understood that the selection of the nucleic acid libraries for DNA or RNA aptamers have advances much faster because nucleic acids themselves can be replicated and expressed than those for the functional units. Proteins cannot be amplified, therefore, to evolve proteins, it is necessary to have a linkage between the gene (nucleic acid) that encodes its phenotype (protein)[8]. The physical link between the gene and the gene product, i.e., the protein results in the retention of both target-specific polypeptide and its encoding DNA, thus, giving a united molecule. Biomolecular display technologies allow construction of a diverse and large pool of biomolecules. They allow screening of novel biomolecules from a pool of candidates as per the user-demand in laboratory for drug discovery.

1.2.2. Types of Molecular Display Technology

Over the past decades, many display technologies have been established, some varying in the coding molecule, some have different linking formats and some with types of displaying entity. Overall, the display technologies can be broadly classify into two categories as, cell-based or cell-free type on the basis of their expression system, but in here I would like to divide them as, open-type or closed-type as shown in Figure 4.

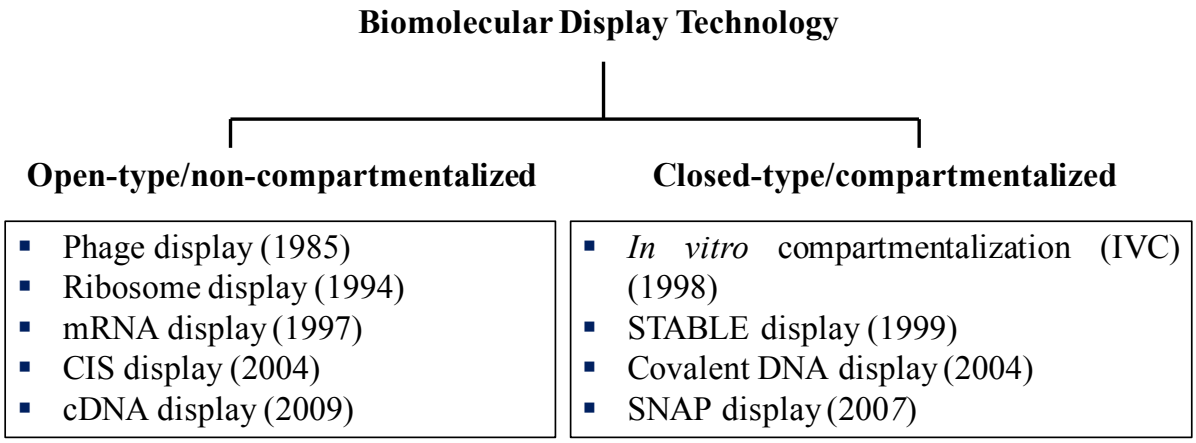


Figure 4: The different types of biomolecular display technologies established till date; divided as open-type and closed-type.

The biological reactions performed in the tube are named as Open-type/non-compartmentalized whereas the biological reactions performed in the droplets or compartments are named as Closed-type/compartmentalized.

Phage display, by George P. Smith in 1985 was one of the first display method for peptides on the coat of filamentous phage. Phage display was followed by describing bacterial surface display and yeast surface display; both based on cell based expression. However, cell-based technology suffered the limitation of library size due to transformation efficiency of host, long time duration of the process from few days to weeks and generation of toxic proteins leading to death of the host cell. Due to the above-mentioned limitations of cell-based type, various groups worked on developing cell-free display technologies. Ribosome display, in 1994 by Mattheakis et al. was the first cell-free display technique. It was based on *E. coli* lysate to display peptides on the ribosome. This technique linked the nascent peptide/protein to its gene via ribosome as the ribosome stalls on the mRNA template. This technique was further followed by mRNA display which coupled the mRNA to nascent polypeptide via puromycin. mRNA display was the first covalent linkage method. Later on, in 1998 Tawfik and Griffiths put-forward the generation of cell-like compartments using emulsion technology. They termed these water-in-oil compartments as *in vitro* compartmentalization (IVC), which kept the genotype and its phenotype in one cell like compartment. Another important advancement was made in IVC by the introduction of SNAP display in 2007 to covalently link the genotype with its phenotype within droplets or compartments. Cell-free type molecular display technologies win over the cell-based type molecular display as the entire process was done *in vitro* leading to the possibilities to generate large library size, reducing the time from weeks to days and easy production of toxic proteins. Regarding open-type or closed-type system, it is difficult to say which technique wins over which as both have their own advantages and disadvantages.

The next section briefly describes few of the display technologies established till date.

1.2.2.1. Open-type/non-compartmentalized display technique

Below display technique is based on the performance of biological reactions in tube which mainly includes:

1.2.2.1.1. Phage Display

With the aim to isolate the desired properties from a diverse pool of libraries, the oldest and most commonly established physical link between the genotype and phenotype was done in 1985 by Dr George P. Smith at the University of Missouri-Columbia. He inserted the coding regions for protein or peptide into the bacteriophage genome and allowed the expression of the desired properties (protein or peptide segment) from the virus. The gene encoding for the *Eco*RI restriction enzyme was inserted into the M13 bacteriophage genome and it was found that one of the minor coat protein fused with the *Eco*RI protein segment, the virus was found to be tolerant to the fusion as they were able to produce infectious particles[9]. The recombinant virus which have chimeric proteins can be easily distinguished from that of the nonrecombinant virus particles by the use of an antibody that recognized the expressed portion of the *Eco*RI protein through the process called as 'biopanning' (sequential rounds of affinity selection). The site chosen for display of protein was not interfering with either the morphogenesis or the infection process of the virus.

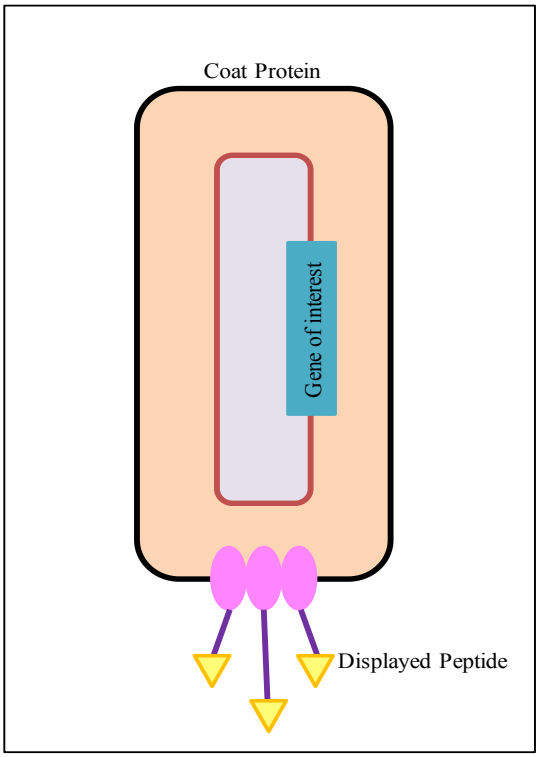


Figure 5: Schematic representation of filamentous phage display. The gene of interest is inserted into the vector of the phage and the desired peptide is displayed on the coat protein (outside).

The phage 'display' the protein on its outside while containing the gene encoding the protein on its inside thereby resulting in a physical connection between the genotype and its phenotype as shown in Figure 5. These displaying phages are then screened against other DNA sequences, peptides or proteins, to detect interaction between the displayed protein and the other molecules. Therefore, large libraries of proteins can be selected and amplified. Phage-display peptide or protein libraries are mostly fractioned by affinity selection. The target protein which is to be displayed is immobilized on a surface, incubated with an aliquot of viral particles. The nonbinding particles are washed away and bound particles are recovered from the population. Protease treatment (subtilase, trypsin) or exposer to extreme pH say pH 2 or 12 or denaturants of proteins (DTT, urea) are used to release the bound particles from unbound without destroying their infectivity. After neutralizing, the recovered phage are used to infect *E. coli* cells, resulting in amplifying the selected signal. Some of the examples of affinity maturation by phage display are, the development of human growth hormone (hGH) variants with high affinity to hGH binding protein (hGHbp) for treatments of acromegaly[10, 11], the development of applied bovine pancreatic trypsin inhibitor (BPTI) for inhibition of the serine protease human neutrophil elastase[12]. Some of the common bacteriophages used in phage display are bacteriophage M13, bacteriophage T4, bacteriophage T7, bacteriophage Lambda (λ), etc.

In vivo techniques like phage display as mentioned above, are hampered due to inability to select under conditions different from the cellular environment, selection of toxic peptides or proteins causing death of the host, host unable to withstand the selection pressure and transformation efficiency (library size of 10^8), thus demanding a new category called as cell-free or *in vitro* techniques which could by pass these problems[13]. *In vitro* display system are not limited by the biology of the host (virus or cell) as performed in cell free systems and thus allow much higher library size of the order 10^{12} to 10^{14}. The below mentioned Ribosome display and mRNA display are cell-free techniques and in here are classified under Open-type/non-compartmentalization.

1.2.2.1.2. Ribosome Display

Ribosome Display was the first cell-free based display technology established almost after a decade of the discovery of phage display. It is based on the similar concept of linking the genotype and phenotype as that of phage display, instead of introducing the gene of interest into the host vector, the ribosome itself becomes the link between the gene and the displayed protein.

The transfer of genetic information from gene to protein expression follows the principle of central dogma of biology[14], genetic information may be retained in the DNA or transferred to the next level to RNA by the process of transcription or may be reversed back to DNA from mRNA by the process of reverse transcription or at last, may be transferred to proteins from RNA by the process called as translation using ribosomes. The transcription of mRNA is performed into the nucleus whereas the translation is done in the cytoplasm of the cell.

The genetic information is transferred from DNA to mRNA, and this mRNA is then used as the template to produce the proteins. Normally, once the full-length protein is translated the ribosome dissociates from the mRNA but if the signals to dissociate the ribosome are missing, the ribosome stalls (stops) at the end of the mRNA thus linking the mRNA with the protein via ribosome. This principle is exploited in ribosome display to physically link the genotype with its phenotype. As previously mentioned, ribosome display was first developed by Mattheakis et al. for the selection of short peptides and further improved by Hanes and Plückthun and He and Taussig for the selection of folded proteins. It is based on the formation of ternary complex between the mRNA, ribosome and the nascent polypeptide which is non-covalent in nature[15]. This ternary complex ensures the physical link between gene and protein as shown in Figure 6.

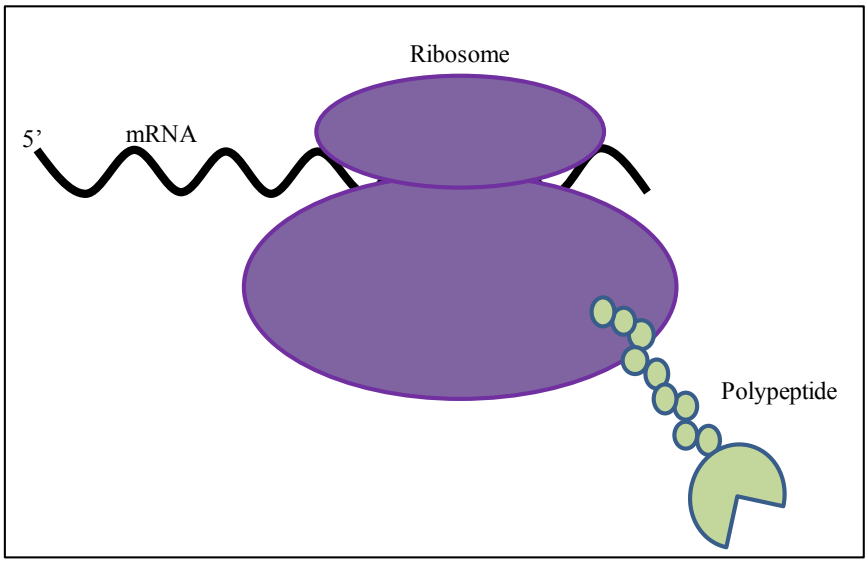


Figure 6: Schematic representation of ribosome display; the ribosome stalls at the end of the mRNA thus forming ternary complex between gene (mRNA), linking entity (ribosome) and protein (the nascent polypeptide).

The DNA construct for ribosome display consists of a promoter region recognized by T7 RNA polymerase, a ribosome binding site, the gene of interest sequence and a spacer region which act as tether at the 3'-terminal of mRNA without stop codon. The DNA is then transcribed and translated in such a way that gene of interest is fused to the C-terminal tether, which allow folding of the peptide while itself being into the ribosomal tunnel. As it lacks the stop codon in the mRNA, ribosome stops at the 3'-terminal thus preventing the release of both the mRNA template and the nascent polypeptide resulting in forming the ternary complex. This ternary complex can be further stabilized by low temperature and the addition of high concentration of magnesium. These complexes (mRNA, ribosome and the nascent polypeptide) formed during *in vitro* translation, can be used directly for selection on immobilized target. The mRNA is released from the ribosome by the addition of EDTA, as EDTA unstabilizes the ribosome complex and the sequences of interest are recovered by performing reverse transcription (RT) of the selected mRNA, and performing polymerase chain reaction (PCR) to amplify the resultant DNAs. This can be further continued to enrich and get the best candidate. Ribosome display has been used to select proteins with catalytic activities and mainly used for selecting and displaying scFV's antibodies[16-19]. This method has been used to select binders in the low

nanomolar range from ankyrin repeat libraries[20], peptides[21, 22], and folded proteins from random peptide libraries by selection against proteolytic cleavage or hydrophobicity[23, 24].

No doubt ribosome display is the first *in vitro* technique and allow large library size but the linkage is non-covalent in nature thus cannot withstand the various harsh selection conditions and therefore demands another technique with strong binding.

1.2.2.1.3. mRNA Display

After 3 years of the discovery of ribosome display, mRNA Display was established by Roberts and Szostak and Nemoto et al., independently in 1997. This was the first covalent linkage between the gene and protein[25]. It relies on the covalent coupling of the mRNA with its expressed protein via puromycin, refer to Figure 7. Puromycin is an antibiotic that mimics the aminoacyl end of tRNA, it enters into the ribosomal A-site and forms an amide bond with nascent polypeptide through the peptidyl transferase activity of the ribosome[26, 27].

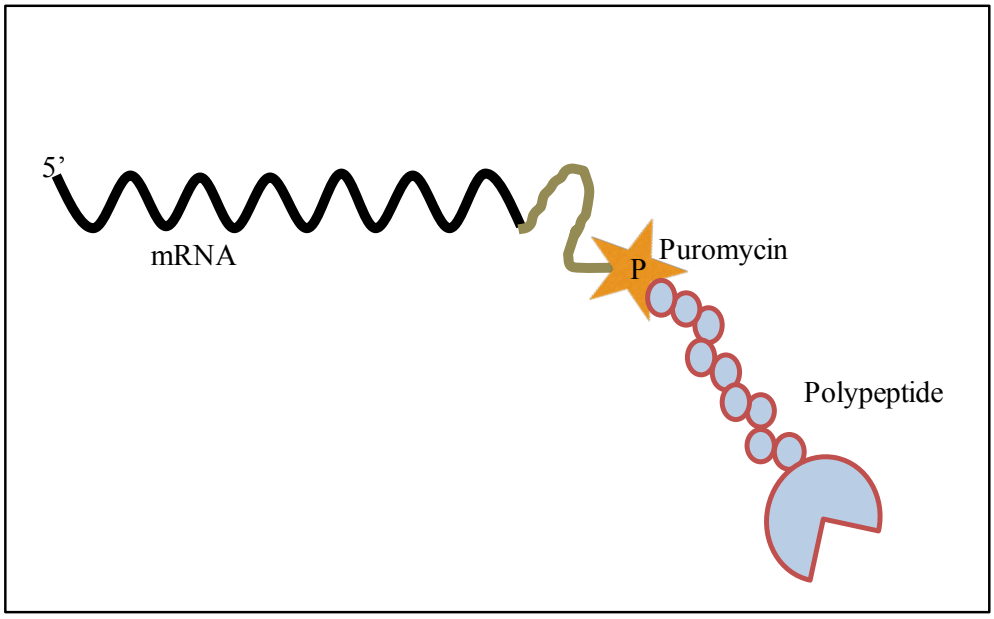


Figure 7: Schematic representation of mRNA display linking the mRNA to the polypeptide via puromycin.

A short DNA linker carrying a puromycin moiety is covalently linked to the mRNA. The gene of interest or the library is then *in vitro* translated same as of ribosome display. When the ribosome reaches the RNA–DNA junction the ribosome stalls as the end; the mRNA template here also lacks stop codon thus allowing the entry of puromycin to the peptidyltransferase site of the ribosome and thereby forming a covalent linkage between the nascent polypeptide and the mRNA template. The covalently coupled mRNA and the protein are subsequently isolated from the stalled ribosomes complex and purified for conversion into cDNA strand by reverse transcription (RT) to form stable RNA–DNA hybrid followed by performing selection using the complexes. Selection is performed by bounding the complex to immobilized ligands and bound complexes are eluted out. Elution step is followed by PCR to amplify the selected clone's genetic information. The finally selected clones are used for the next round of selection. Originally, puromycin was ligated to the 3'-end of the mRNA and more recently puromycin is hybridized to the 3'-end of the mRNA followed by covalent cross-linking via psoralen attached to the DNA by UV light[28, 29]. Same as ribosome display, mRNA display has also been improved from the generation of short or small peptides to full-length proteins over the years[30]. The mRNA display has been used to identify protein–protein interactions, for identifying binders based on the randomisation of exposed loops[31], successful selections of peptides[32-34], identification of kinase substrates[35] and evolution of an ATP binder from a pool of 80-mers[36].

However, as in both ribosome display and mRNA display, the genetic information is carried by the mRNA, they are susceptible to ribonucleases and difficult to manipulate. As in ribosome display the ternary complex is non-covalent, they suffer from potential dissociation of the triplet molecule either losing the genetic information or losing the expressed nascent polypeptide whereas in mRNA display, puromycin attachment requires complex steps thereby making them not so user-friendly techniques.

1.2.2.2. Closed-type/compartmentalized display technique

The biological reactions performed in the compartments made up of water-in-oil or oil-in-water droplets are called as closed-type/compartmentalized display

techniques. *In vitro* cell-free display technologies were further advanced by the formation of water-in-oil droplets using emulsion technologies to encapsulate one gene with its protein together in one tiny droplet. The advancement of IVC, explained in the next section, leads to the improvement of library size from 10^8 to 10^{10} in *in vitro* cell-free system.

1.2.2.2.1. *In vitro* compartmentalization

In vitro compartmentalization (IVC) enables to encapsulate biochemical reactions in extremely miniaturized aqueous micrometer-sized droplet reactors which acts as cell like compartments for the directed molecular evolution[37-39] (Figure 8). IVC provides an alternative way to link/couple genotype and phenotype to mimic the natural compartments of living organisms[40, 41]. It offers a means to parallelize biological and chemical assays. It was first introduced in 1998 by Tawfik and Griffiths. They utilized the emulsion technology for the generation of water-in-oil droplets (w/o). They called these water-in-oil droplets as man-made cell like compartments to unit gene and its protein[42-44]. Microdroplets, droplets in the range of pico to femtoliters scale, have been successfully used as single DNA molecule vessels. The water-in-oil droplet technology exploits many different selection pressures in a single experimental setup with high throughput.

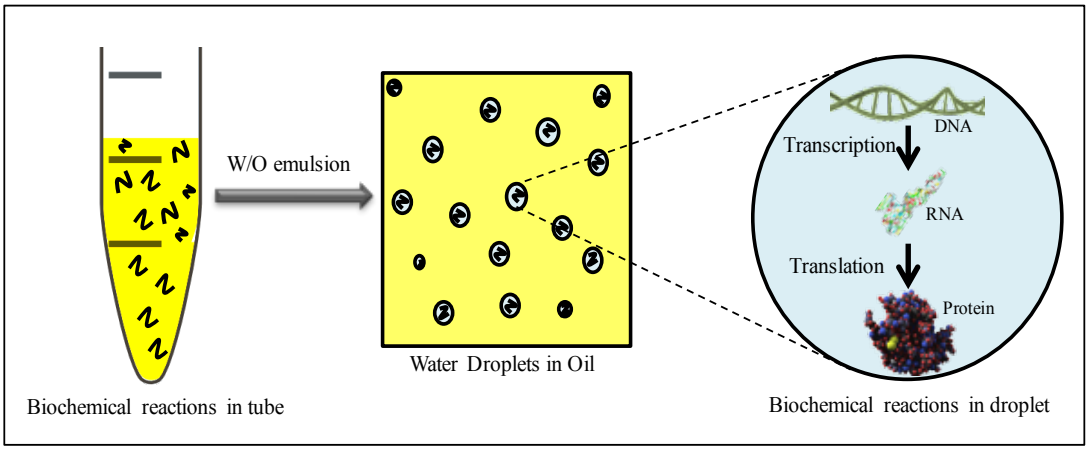


Figure 8: Schematic representation of In vitro compartmentalization (IVC).

To enhance the yield and reactivity of the reaction, water-in-oil emulsion has been extensively exploited as micrometer-sized reactors. Over the decades, there are

many different methods utilized for the generation of water-in-oil droplets with varying size like homogenizers, microfluidic based droplet generation, and extruder.

Homogenizer used shear stress or physical barrier to generate small size w/o droplets. This resulted in generation of large size difference or polydisperse droplets[45] causing loss in biological sample due to the high speed and pressure applied in homogenizer. The batch mode emulsion technique is based on generation of 1-100 μm size water-in-oil droplet by simple vortexing resulting in fL-to-nL difference in reaction volume[46, 47]. Vortexing method is quite similar to homogenizer with respect to generation of large size difference in the droplets. These leads to uneven distribution of reaction components causing loss of translation activity of the proteins.

To overcome the above discussed problems, microfluidic-based droplet generation[48, 49] was introduced, which used T-junction or flow focusing methods to generate similar size water-in-oil droplets or highly monodisperse w/o droplets. Microfluidic-based droplet generation, by far has been the best IVC technique in terms of uniformity of the droplets but they lack in speed. The generation speed is very slow (few thousand per minute)[50, 51] and thus the library size of "million-to-quadrillion" molecules cannot be completely encapsulated in short time.

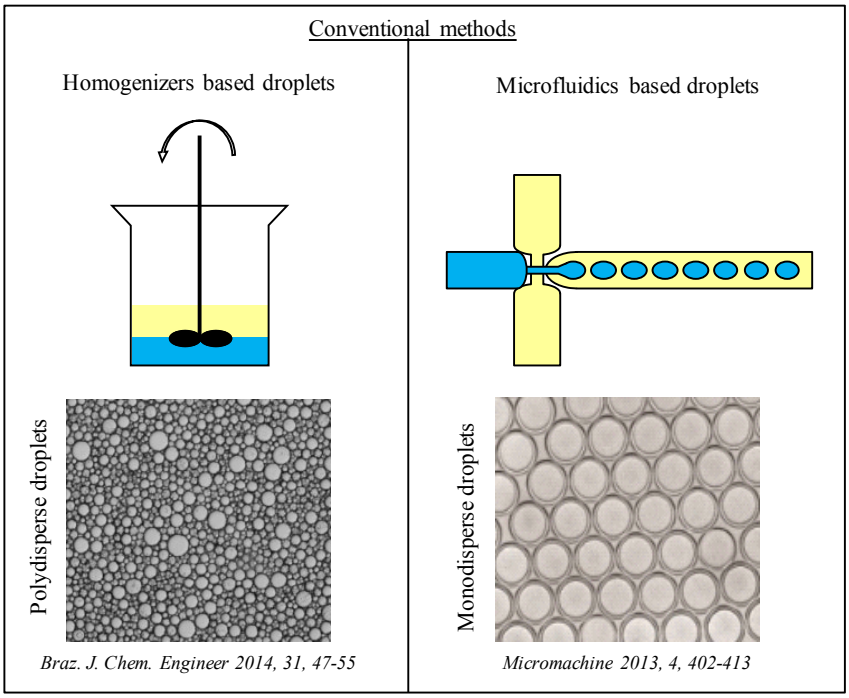


Figure 9: The conventional methods to produce water-in-oil droplets.

With some or the other merits and demerits, it can be concluded that homogenizers generate high speed droplets whereas microfluidic based droplet generation system produces monodisperse droplets[52] (Figure 9). But independently, these techniques do not fulfil the requirement of generating high throughput monodisperse water-in-oil droplets; refer to Table 1 for summary. Therefore, the need to find a new platform still continues.

Droplet generation methods	Diameter of the droplet	Generation speed	Uniformity
Homogenizer	~ 10 µm to mm	~ 10^9 droplets/min	Poor
Microfluidic based droplets	> 50 µm	< 10^5 droplets/min	High
Mini extruder	~ 10 µm	~ 10^9 droplets/min	Medium

Table 1: The comparison between three different water-in-oil droplet generation methods till date.

1.2.2.2.2. SNAP display

An important advancement was made in 2007 to covalently link genotype with its phenotype in emulsion droplets called as SNAP display. This technique utilized the AGT-BG chemistry which was based on covalent reaction between a DNA repair protein AGT (O^6-alkylguanine-DNA alkyltransferase) with its substrate BG (O^6-Benzylguanine)[53]. DNA encoding AGT was encapsulated with cell free protein synthesis system in water-in-oil droplets made using homogenizer of few micrometers size. They diluted the DNA to the level where only one copy of DNA is inserted into one droplet, which is then *in vitro* transcribed and translated within each droplet. As soon as the AGT protein attached to the protein of interest is expressed, it covalently binds to the BG bounded DNA hence, making a covalent bondage between the genotype and phenotype (Figure 10). Rounds of selection are performed by breaking the emulsion droplets followed by isolating the DNA-protein conjugated molecule. 'SNAP-tag' is the commercial name of the AGT, DNA repair protein. The covalent linkage via AGT-BG between protein and DNA makes easy handling during harsh

selection pressures. This allowed formation of library size ~10^9 compartments in 1 mL of emulsion[54].

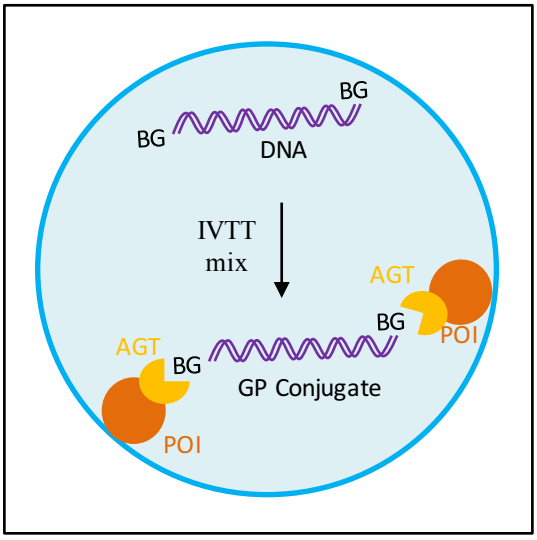


Figure 10: The SNAP display to produce water-in-oil droplets encapsulating covalently linked DNA with protein.

In 2016, B. Sharma et al.[55] utilized inkjet principles where upon the application of high voltage, a single glass nozzle immersed in oil phase is forced to burst into a jet of droplet leading to the uniform production of w/o droplets at a very high speed of more than 10^5 droplets per second. They generated water-in-oil droplets volumes ranging from 0.2 to 6.4 fL and agarose-in-oil droplets volume ranging from 0.3 to 15.6 fL. Both are sub-femtoliter compartments with average sizes in the range 1.3-1.5 µm. This technique no doubt fulfils the requirement of both monodisperse water-in-oil droplets with high speed but it still leaves scope for much improvement.

1.3. Objective of the research

As mentioned previously, till date many display technologies has been developed and it is important to mention here that in both; ribosome display and mRNA display, the 3'-terminus of mRNA is linked to C-terminal of protein thus these techniques cannot be used for C-terminal post translation modifications as the C-terminal of protein is engaged in linkage and the N-terminal is freely exposed. Also, as the linkage moiety is present at the end of the mRNA i.e., 3'-end of mRNA, the translation of the nascent polypeptide (phenotype) and its binding to the mRNA (genotype) takes time causing more chances of degradation. Also, all these display methods are affinity based selection with low Genotype-Phenotype (G-P) yield.

Therefore, a new, simple and rapid display technology to establish a strong linkage (via covalent bond) needs to be developed to overcome these challenges, which I present as 'Head-to-Head' genotype-phenotype display technology for improved genotype-phenotype yield and function-based *in vitro* protein selection, the details of which are described in chapter II. Chapter III utilizes the basic principle of electrospray to perform one step genotype-phenotype covalent linkage in water-in-oil droplets to integrate H2H in electrospray IVC for one-step *in vitro* protein selection.

After exploiting the principles of electrospray to generate uniform sized droplet and high throughput for IVC, it can further be expanded to generate ultra-high throughput water-in-oil droplets. Considering both monodispersity and generation of ultra-low volume ultra-rapid water-in-oil droplets, I further extended the principles of electrospray to accomplish the demand for improved library size generation using a micro-hole array. This easy and rapid micro-hole array electrospray technique to produce water-in-oil droplets is explained in chapter IV. Finally, the entire research work is concluded in chapter V.

CHAPTER II

DEVELOPMENT OF A NOVEL CONSTRUCT FOR HIGH EFFICIENCY OF GENOTYPE-PHENOTYPE CONJUGATE AND FUNCTION-BASED *IN VITRO* PROTEIN SELECTION

CHAPTER II

Development of a novel construct for high efficiency of genotype-phenotype conjugate and function-based *in vitro* protein selection

2.1. Introduction

The tremendous increase in the diseases has led to the search of the desired cure in short time span. The molecular evolutionary engineering and the development of *in vitro* selection methods enables us to generate functional molecular recognition candidate, such as aptamer, from a pool of artificial library for diagnostic or therapeutic purposes[56, 57]. In contrast to the reliability of DNA and RNA-based aptamers[58, 59], a peptide-based aptamer[60] which are comparatively smaller in molecular weight and exhibit a smaller binding footprint, binds to target proteins with a high specificity and strong affinity and thus has opened a new window as guides for small-molecule drug discovery[61-63]. As mentioned in the chapter-1, *in vitro* evolution of proteins and peptides is difficult compared to selection of single stranded DNA and RNA molecules as encoding information for a protein sequence is usually missing after its translation step[64] and thus comes the role of display technologies to allow physical association between peptide or proteins with its encoding information which can be amplified and characterised for new or altered function of peptide libraries[65]. Phenotype, the protein is the functional unit amalgamate to its genotype, the nucleic acid which can be easily replicated.

Over the period of time many different molecular display technologies have been utilized mainly, phage display, ribosome display and mRNA display. As described in chapter-1, phage display is a cell-based technique, ribosome display requires the ternary complex (the mRNA, ribosome and protein) to be stable enough to resist the selection pressure whereas the use of puromycin in mRNA display hinders the folding of proteins and pre-mature translation occurs as puromycin attaches to peptidyltransferase site and hence stops the expression of full length protein[66]. Ueno et. al. performed mRNA-protein fusion at N-terminus by inserting stop codon just downstream of initiation site. An ingenious method was devised to link an mRNA to the relevant nascent protein via an elaborate amino acid: the amino acid chemically links to mRNA, whereas it is incorporated into a peptide by being loaded onto an amber suppressor tRNA[67]. However, unfortunately, the efficiency of the

amber suppressor tRNA incorporation of a ribosome and/or the succeeding elongation step seemed low, providing only a PCR-based amount of ligation products, impeding general use.

Principally, the ribosome display or mRNA/cDNA display techniques rely on coupling the 3' terminus of gene (i.e., tail region of mRNA) with C-terminus of protein (i.e., tail region of nascent polypeptide), which results in free exposing of N-terminus of the nascent polypeptide chain. This kind of coupling between mRNA/cDNA and protein has often been performed for longer time (in hours) as it requires complete translation of the nascent polypeptide chain followed by fusion with the tether/linker molecule of encoding mRNA/cDNA sequence. Additionally, as the C-terminus is directly involved in the linkage, these techniques cannot be used for C-terminus post translational modifications like C-terminal α-amidation, C-methylation[68] which are one of the most important and essential requirements for biological activities as they improve the stability and receptor binding ability of many of the peptides and peptide-hormones. Moreover, the above described techniques lack stop codon[69], thus limiting the number of generated nascent protein to one copy only as the ribosome stalls due to unavailability of releasing factor; thereby the linked molecule needs to undergo additional multiple steps of PCR and transformation making them time consuming techniques. Therefore, in order to overcome aforementioned challenges, we need to design a construct for *in vitro* evolution which is rapid enough for generation of multiple copies of proteins for improved Genotype-Phenotype (G-P) yield and functional screening.

2.1.1. 'Head-to-Head' genotype-phenotype linkage

In here, I report a novel construct, termed 'Head-to-Head' (H2H), for multiple protein-recoverable display where the mRNA/cDNA part of library is physically linked to its encoded protein in 'Head-to-Head' direction, i.e., 5' of mRNA with N-terminal of nascent polypeptide chain (see Figure 11, next page).

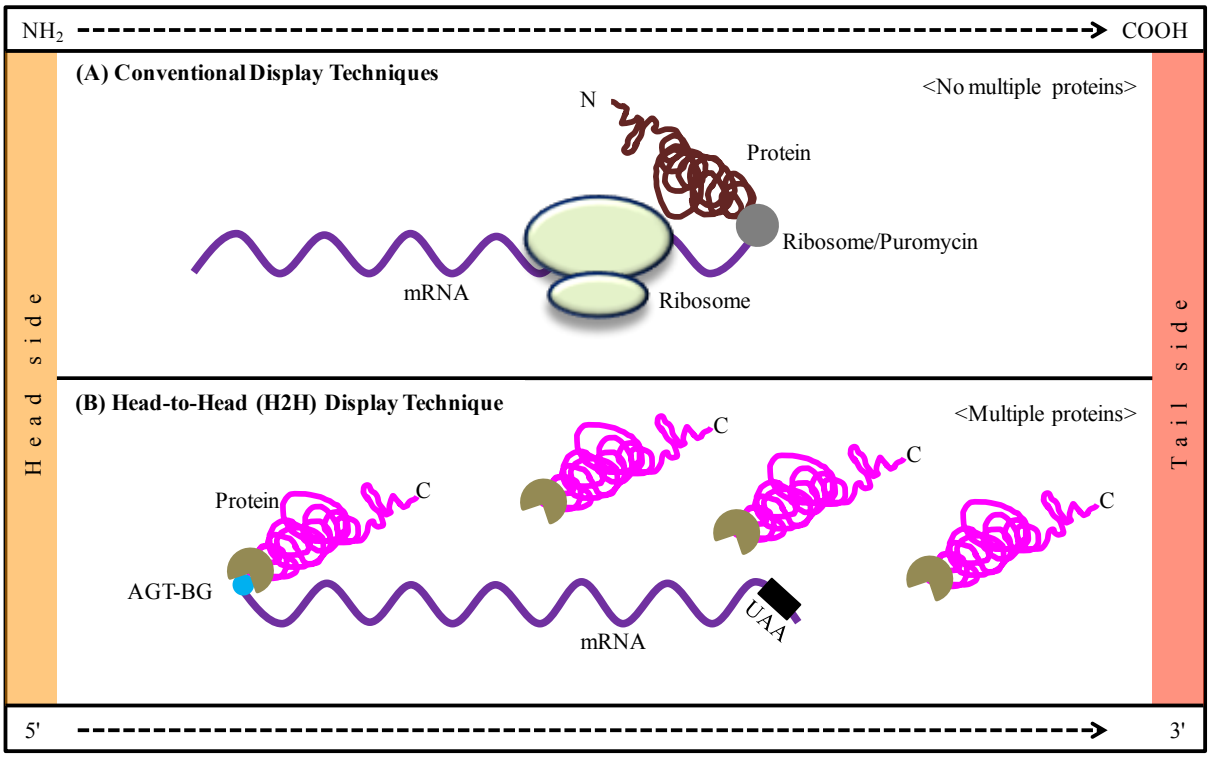


Figure 11: Schematic representation of comparison between conventional display techniques and our novel 'Head-to-Head' (H2H) concept for uniting genotype with its phenotype.
(A) Conventional display techniques like Ribosome display and mRNA display are based on tail to tail concept i.e., 3' of mRNA/DNA links with carboxyl terminal (C) of protein; thus giving free N-terminus of protein.
(B) H2H concept is based on 'Head-to-Head' covalent linkage/fusion i.e., 5' of mRNA/DNA with amino terminal (N) of protein; thus giving free C-terminus of protein and therefore united in a single molecule. H2H construct have stop codon thereby allowing generation of multiple-copies of functional unit from one molecule of gene.

Additionally, in H2H construct, the N-terminal is directly coupled with the mRNA/cDNA whereas the C-terminus of the nascent polypeptide is freely available for the C-terminal modifications. As a key feature of H2H, physical linkage between mRNA/cDNA and nascent polypeptide chain can take place rapidly and efficiently since both the head-ends remain in close proximity and so thus not requires waiting for the complete translation of mRNA transcript. Moreover, H2H construct is designed to include stop codon, thus facilitating the generation of multiple functional unit from one gene molecule thereby eliminating the need of additional multiple steps. The linked protein to mRNA via a covalent

linkage can be used to get the information of the selected protein whereas the multiple proteins formed from one gene can be used as functional unit; hence, developing a display technique from information-to-function. The binding and functionality of the peptide can be checked together from one gene copy. The detail illustration is given in Figure 12. Covalent bondage allow high Genotype-Phenotype (G-P) yield in H2H.

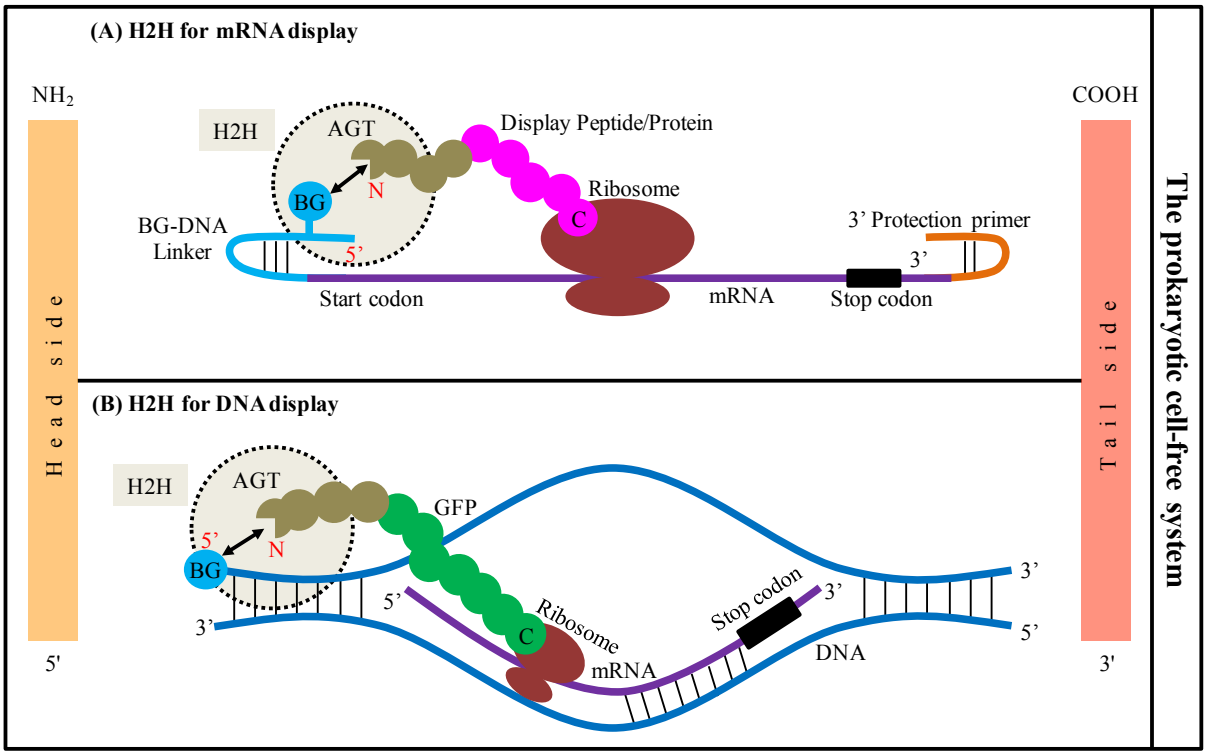


Figure 12: Detail illustration of 'Head-to-Head' linkage (H2H) concept uniting genotype with its phenotype. (A) H2H for mRNA display: Covalent linkage between mRNA and protein via AGT-BG. (B) H2H for DNA display: Covalent linkage between DNA and protein via AGT-BG.

A unique BG modified DNA linker is hybridized to the mRNA and a well know interaction of BG-AGT[70] is used for the covalent linkage between the head of the genotype (5'-terminus) and the head of the phenotype (N-terminus). O^6-Benzylguanine (O^6-BG) conjugates with O^6-alkylguanine-DNA alkyltransferase (also known as AGT, MGMT or AGAT) by transferring benzyl group to cysteine-145 in the AGT active site (as shown in Figure 13) and thereby forming a thioether covalent bond[71], this covalent linkage allows strong binding which results in efficient united G-P molecule.

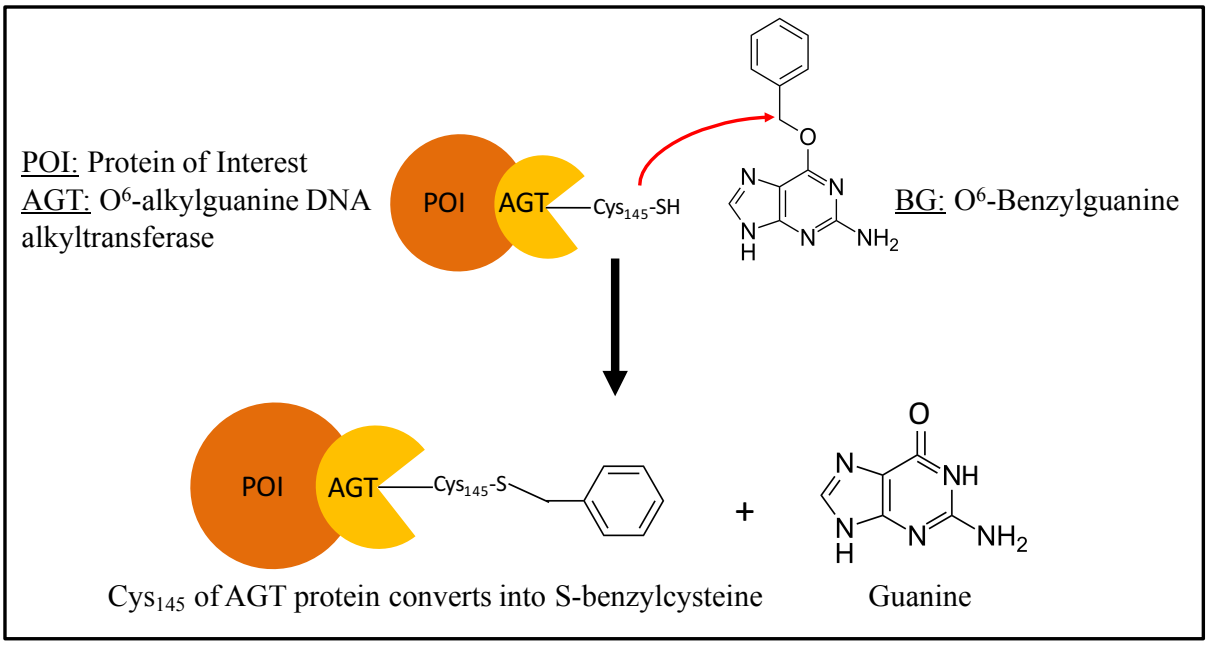


Figure 13: Detail illustration of the reaction of the AGT or SNAP-tag with BG resulting in the covalent thioether bond attachment to the active site cysteine.

Further, H2H construct also supports longer transcript, which in general is limited in other display techniques. I engaged both mRNA and DNA as template for improved protocol and high yield of linkage product. I successfully demonstrated the utility of H2H construct to display a transcript of about 1.5 kb with multiple protein domains. Our H2H strategy is new, simplified, rapid and efficient linkage for the recoverability of multiple protein copies for molecular evolution / protein engineering.

2.2. Experimental

2.2.1. H2H for mRNA/cDNA display

2.2.1.1. H2H DNA Construct

To obtain DNA construct for *in vitro* transcription, the gene encoding for AGT protein and random library were first cloned into pTAC-2 (purchased from EurofinsGenomics). Not I restriction enzyme (Promega) was used to cleave the plasmid and PCR amplified (PrimeSTAR HS DNA Polymerase, 2.5 U/µl, Takara Clontech) to get the desired sized DNA sequence followed by PCR clean-up purification (Macherey-Nagel). The purified DNAs concentration was assessed using NanoDrop 2000 UV-Vis spectrophotometer

and PAGE. The sequence of different constructs used are given in Figure 14 (without random library sequence) and Figure 15 (with random library sequence).

GATCCCGCGAAATTAATACGACTCACTATAGGGGCTCGCGAATACTGCGAAGGAGAGACAAC
GGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCAATGGACAAAGAT
TGCGAAATGAAACGTACCACCCTGGATAGCCCGCTGGGCAAACTGGAACTGAGCGGCTGCGA
ACAGGGCCTGCATGAAATTAAACTGCTGGGTAAAGGCACCAGCGCGGCCGATGCGGTTGAAG
TTCCGGCCCCGGCCGCCGTGCTGGGTGGTCCGGAACCGCTGATGCAGGCGACCGCGTGGCTG
AACGCGTATTTTCATCAGCCGGAAGCGATTGAAGAATTTCCGGTTCCGGCGCTGCATCATCC
GGTGTTTCAGCAGGAGAGCTTTACCCGTCAGGTGCTGTGGAAACTGCTGAAAGTGGTTAAAT
TTGGCGAAGTGATTAGCTATCAGCAGCTGGCGGCCCTGGCGGGTAATCCGGCGGCCACCGCC
GCCGTTAAAACCGCGCTGAGCGGTAACCCGGTGCCGATTCTGATTCCGTGCCATCGTGTGGT
TAGCTCTAGCGGTGCGGTTGGCGGTTATGAAGGTGGTCTGGCGGTGAAAGAGTGGCTGCTGG
CCCATGAAGGTCATCGTCTGGGTAAACCGGGTCTGGGTCTGGCTGGTTCTGGTGGCGGTGGC
AGCGAAGGCGAATGGCAGCAACAGCAACATCAGTGGGCACATCAGGAGCATCATCATCATCA
TCACGGATCC*ATCGAGGGCCGCTAAATCGGCA*GCTAGCATAACCCCTT

Spacer: 13 bp
T7 Promoter: 17 bp
Linker hybridization: 29 bp
Spacer: 39 bp
RBS: 6 bp
Spacer: 8 bp
pSNAP code: 543 bp
Spacer: 18 bp
His tag: 57 bp
His tag: 18 bp
Spacer + Xa: 18 bp
Stop codon: 3 bp
Spacer: 9 bp
3' Partial T7 Terminator: 14 bp

Figure 14: DNA Construct (792 mer) for H2H mRNA experiment without random library sequence.

GATCCCGCGAAATTAATACGACTCACTATAGGGGCTCGCGAATACTGCGAAGGAGAGACAAC
GGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCAATGGACAAAGAT
TGCGAAATGAAACGTACCACCCTGGATAGCCCGCTGGGCAAACTGGAACTGAGCGGCTGCGA
ACAGGGCCTGCATGAAATTAAACTGCTGGGTAAAGGCACCAGCGCGGCCGATGCGGTTGAAG
TTCCGGCCCCGGCCGCCGTGCTGGGTGGTCCGGAACCGCTGATGCAGGCGACCGCGTGGCTG
AACGCGTATTTTCATCAGCCGGAAGCGATTGAAGAATTTCCGGTTCCGGCGCTGCATCATCC
GGTGTTTCAGCAGGAGAGCTTTACCCGTCAGGTGCTGTGGAAACTGCTGAAAGTGGTTAAAT
TTGGCGAAGTGATTAGCTATCAGCAGCTGGCGGCCCTGGCGGGTAATCCGGCGGCCACCGCC
GCCGTTAAAACCGCGCTGAGCGGTAACCCGGTGCCGATTCTGATTCCGTGCCATCGTGTGGT
TAGCTCTAGCGGTGCGGTTGGCGGTTATGAAGGTGGTCTGGCGGTGAAAGAGTGGCTGCTGG
CCCATGAAGGTCATCGTCTGGGTAAACCGGGTCTGGGTCTGGCTGGTTCTGGTGGCGGTGGC
AGCGAAGGCGAATGGCAGCAACAGCAACATCAGTGGGCACATCAGGAGCATCATCATCATCA
TCACGGATCC*ATCGAGGGCCGC*NNNNNNNNNNNNNNNNNNNNNNNNTAAATCGGCAGCTAGC
ATAACCCCTT

Spacer: 13 bp
T7 Promoter: 17 bp
Linker hybridization: 29 bp
Spacer: 39 bp
RBS: 6 bp
Spacer: 8 bp
pSNAP code: 543 bp
Spacer: 18 bp
His tag: 57 bp
His tag: 18 bp
Spacer + Xa: 18 bp
Variable: 24 bp
Stop codon: 3 bp
Spacer: 9 bp
3' Partial T7 Terminator: 14 bp

Figure 15: DNA Construct (816 mer) for H2H mRNA experiment with 24 bp random library sequence.

2.2.1.2. *In vitro* transcription

PCR purified templates were used for cell-free *in vitro* transcription (RiboMAX Large Scale RNA Production System-T7, Promega) at 37 °C for 2 hours followed by removal of the DNA template by addition of RQ1 RNase-Free DNase (Promega) to a concentration of 1u/µg of template DNA and incubated for 15 minutes at 37 °C. The mRNAs were purified (RNeasy Mini Kit, Qiagen) and concentration was measured at 260 nm using Nanodrop.

2.2.1.3. Ligation of mRNA with linkers

2.2.1.3.1. Ligation of mRNA with BG-DNA linker

The 5'-terminal ends of the mRNA molecules were hybridized to the complementary strands of the BG-DNA linker under annealing conditions (lowering the temperature linearly from 90° to 25° C) in ligation buffer in the presence of BSA (see Figure 16). Following the addition of T4 RNA Ligase (40 U/µl, Takara Clontech), T4 Polynucleotide Kinase (10 U/µl, Takara Clontech) and SUPERase In RNase Inhibitor (Thermofisher Scientific), ligation reactions were performed in a dry incubator at 25° C for 60 min. Ligation optimization using 10-fold, 4-fold and 2-fold molar excess of linker DNA to mRNA were carried out and 2-fold molar excess was used for all the experiments. The products were analyzed in polyacrylamide gels followed by purification using RNeasy Mini kit. The ligated products were visualized by FITC fluorescence. All the fluorescent gel were scanned in Hitachi FMBIO III Multi View using a wavelength of 488 nm for excitation and 505 nm for emission.

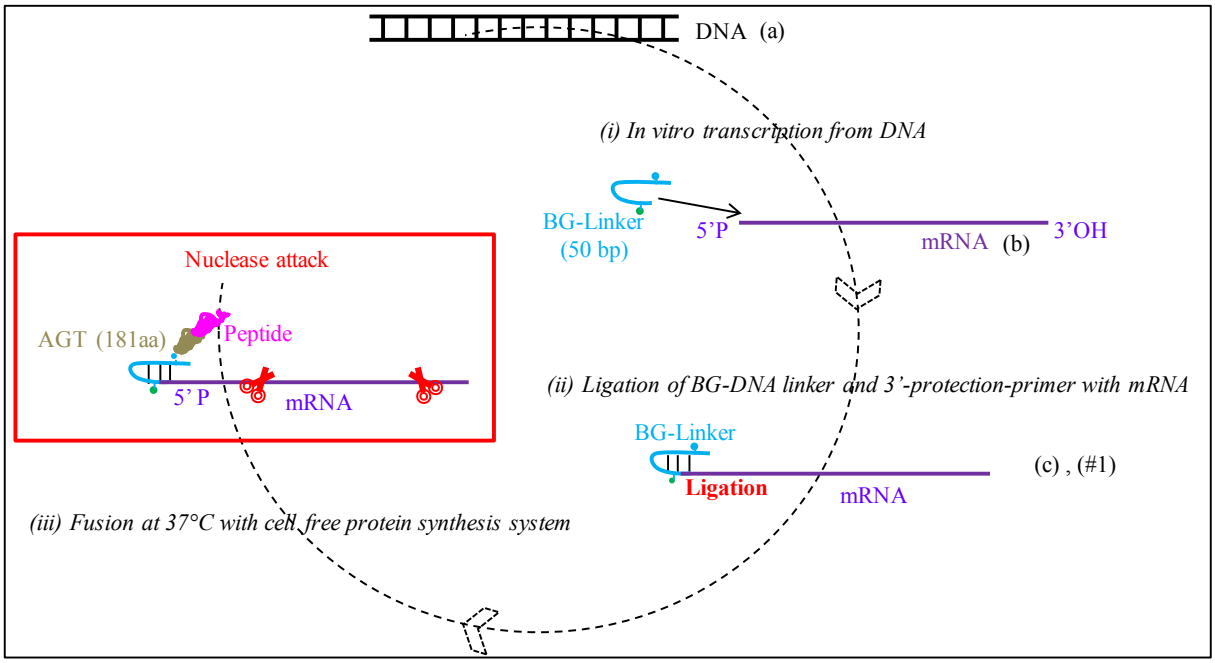


Figure 16: Experimental scheme of our novel H2H concept to link the genotype with the phenotype (before optimization).

(i) The conversion of DNA to mRNA in cell free in-vitro transcription system.

(ii) The transcribed mRNA was ligated with the BG-DNA linker at 5'-terminus.

(iii) 'Coupled in vitro protein synthesis and fusion' reaction at 37 °C using coupled in vitro cell free protein synthesis system.

2.2.1.3.2. Ligation of mRNA with BG-DNA linker at 5'-end and with 3'-protection-primer at 3'-end (Optimization)

The 5'-terminal ends of the mRNA molecules were hybridized to the complementary strands of the BG-DNA linker whereas 3'-terminal ends with 3'-protection-primer in the same manner as described in the previous section.

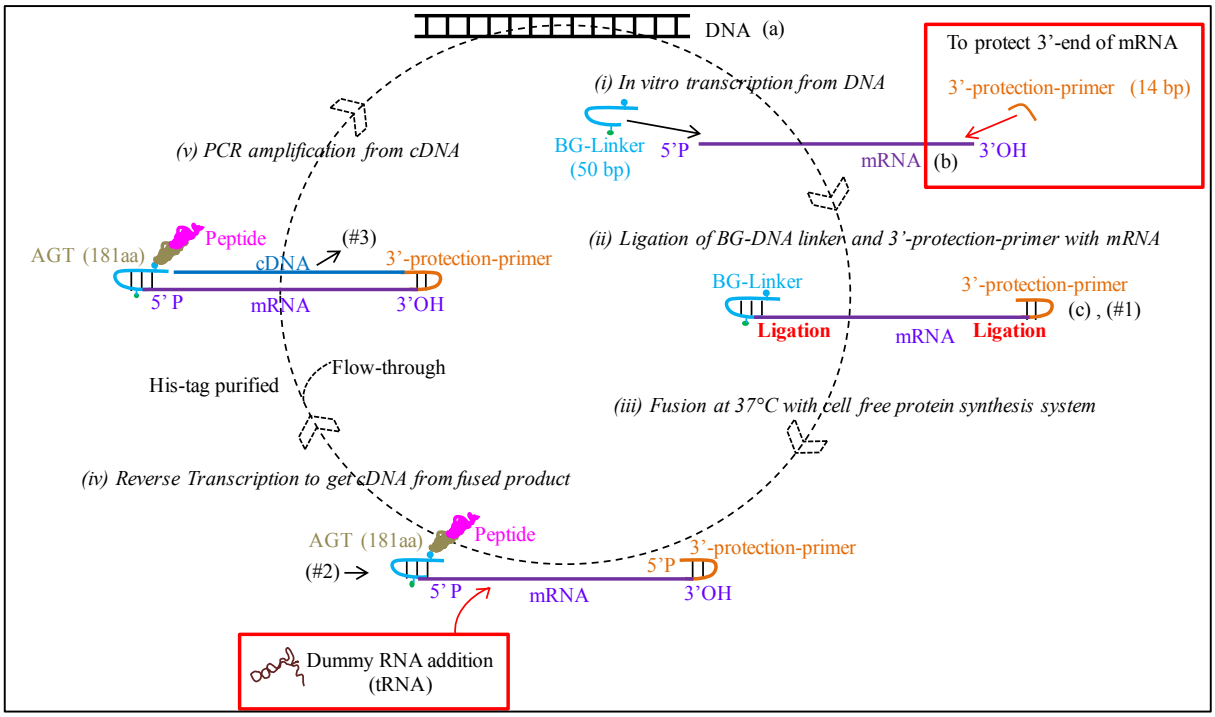


Figure 17: Experimental scheme of our novel H2H concept to link the genotype with the phenotype (after optimization).

(i) The conversion of DNA to mRNA in cell free in-vitro transcription system.

(ii) The transcribed mRNA was ligated with the BG-DNA linker at 5' whereas with 3'-protection-primer at 3' terminus.

(iii) 'Coupled in vitro protein synthesis and fusion' reaction at 37 °C using coupled in vitro cell free protein synthesis system.

(iv) The formation of cDNA by reverse transcription process using mRNA as template.

(v) The amplification of DNA from cDNA obtained.

2.2.1.4. Coupled *in vitro* translation and linkage

The mRNA-BG DNA linker conjugates were added to 17.5 µL of the cell-free translation system (PURExpress *In Vitro* Protein Synthesis Kit, New England Biolabs) and

the mixture was incubated at 37 °C. I tried different fusion time and finally reduced the time from 4 hours[72] to only 15/30/120 minutes as and when stated. To increase the yield and protect mRNA-protein fusions in 15 min for RT-PCR, the post translation fusion reaction was incubated for an additional 15 min at 37 °C in the presence of a high concentration of salts (KCl and $MgCl_2$ at final concentrations of 600 and 85 mM, respectively). The fused products in SDS-PAGE were visualized by FITC fluorescence. I faced problems to protect the mRNA from degradation therefore I added dummy RNA (Yeast tRNA, 10 mg/mL, from Ambion) to preserve our H2H construct (Optimization) (refer Figure 17, previous page).

2.2.1.5. Immobilization of mRNA-protein conjugates on magnetic beads

Ni-NTA magnetic agarose beads (Qiagen) with an average diameter of 50 µm and a range of 20-70 µm diameter were washed thrice with binding buffer according to the manufacturer's instructions. The expressed protein-mRNA and Ni-NTA magnetic agarose beads were then incubated in binding buffer for 30 min at 4 °C. The beads were subsequently washed once each with binding buffer and 1X RT buffer, prior to one step RT-PCR.

2.2.1.6. One Step RT-PCR

His-tag purified mRNA-protein conjugate was used as the template for one step RT-PCR (SuperScript III One-Step RT-PCR system with Platinum *Taq* DNA Polymerase, Invitrogen). Reverse Transcription was performed at 50 °C for 30 min followed by PCR reaction at 98 °C denaturing temperature, 58 °C annealing temperature and 72 °C extension temperature for 20 cycles. The samples were analyzed in polyacrylamide gels.

2.2.2. H2H to validate multiple-domain functionality

2.2.2.1. Multiple protein DNA construct

1509 bp size DNA consisting of AGT gene, 6-His tag sequence, Xa factor and GFP sequence was amplified at 58 °C annealing temperature for 20 cycles followed by PCR clean-up purification. The purified DNAs concentration was assessed using Nanodrop and band size was checked by PAGE. The sequence of the constructs is given in Figure 18.

GATCCCGCGAAATTAATACGACTCACTATAGGGGCTCGCGAATACTGCGAAGGAGAGACAAC
GGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCAATGGACAAAGAT
TGCGAAATGAAACGTACCACCCTGGATAGCCCGCTGGGCAAACTGGAACTGAGCGGCTGCGA
ACAGGGCCTGCATGAAATTAAACTGCTGGGTAAAGGCACCAGCGCGGCCGATGCGGTTGAAG
TTCCGGCCCCGGCCGCCGTGCTGGGTGGTCCGGAACCGCTGATGCAGGCGACCGCGTGGCTG
AACGCGTATTTTCATCAGCCGGAAGCGATTGAAGAATTTCCGGTTCCGGCGCTGCATCATCC
GGTGTTTCAGCAGGAGAGCTTTACCCGTCAGGTGCTGTGGAAACTGCTGAAAGTGGTTAAAT
TTGGCGAAGTGATTAGCTATCAGCAGCTGGCGGCCCTGGCGGGTAATCCGGCGGCCACCGCC
GCCGTTAAAACCGCGCTGAGCGGTAACCCGGTGCCGATTCTGATTCCGTGCCATCGTGTGGT
TAGCTCTAGCGGTGCGGTTGGCGGTTATGAAGGTGGTCTGGCGGTGAAAGAGTGGCTGCTGG
CCCATGAAGGTCATCGTCTGGGTAAACCGGGTCTGGGTCTGGCTGGTTCTGGTGGCGGTGGC
AGCGAAGGCGAATGGCAGCAACAGCAACATCAGTGGGCACATCAGGAGCATCATCATCATCA
TCACGGATCC*ATCGAGGGCCGC*ATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAA
TTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGCGGAGAGGGTGAA
GGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGT
TCCATGGCCAACACTTGTCACTACTCTGACGTATGGTGTTCAATGCTTTTCCCGTTATCCGG
ATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGC
ACTATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGA
TACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCG
GACACAAACTCGAGTACAACTATAACTCACACAATGTATACATCACGGCAGACAAACAAAAG
AATGGAATCAAAGCTAACTTCAAAACTCGCCACAACATTGAAGATGGCTCCGTTCAACTAGC
AGACCATTATCAGCAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATT
ACCTGTCGACACAATCTGCCCTTTTGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTT
CTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAA**TAATAA**AT
CGGCAGCTAGCATAACCCCTT

Spacer: 13 bp
T7 Promoter: 17 bp
Linker hybridization: 29 bp
Spacer: 39 bp
RBS: 6 bp
Spacer: 8 bp
pSNAP code: 543 bp
Spacer: 18 bp
His tag: 57 bp
His tag: 18 bp
Spacer + Xa: 18 bp
GFPuv4+Stop Codon: 714+6 bp
Spacer: 9 bp
3' Partial T7 Terminator: 14 bp

Figure 18: DNA Construct (1509 mer) for H2H DNA experiment mainly with AGT, His-tag, Xa factor and GFP sequence.

2.2.2.2. Coupled *in vitro* translation and linkage

The cDNA construct were added to 17.5 µL of the cell-free translation system and the mixture was incubated at 37 °C for 30 min. BG-DNA linker was further modified by replacing FITC to biotin molecule called as 'BG-biotin linker' for capturing on magnetic beads via streptavidin-biotin interaction. The translated/expressed protein sample was incubated with BG-biotin linker (10 picomole) at 37 °C for 15 min to allow conjugation of BG with expressed AGT-GFP. The fused sample was incubated at 25 °C for 10 min with 100 µmole of streptavidin for biotin-streptavidin interaction to get BG-biotin linker coupled with AGT-GFP protein. The samples with and without streptavidin were analysed in SDS-PAGE and green fluorescence was obtained from the linked and free GFP.

2.3. Results and Discussion

For the rapid *in vitro* protein evolution, I here present a display technique called H2H display, which is the linkage between head of mRNA (5') with head of protein/display protein (N) via AGT-BG covalent interaction as depicted in Figure 11. I checked H2H construct for both mRNA template and DNA template (refer Figure 12).

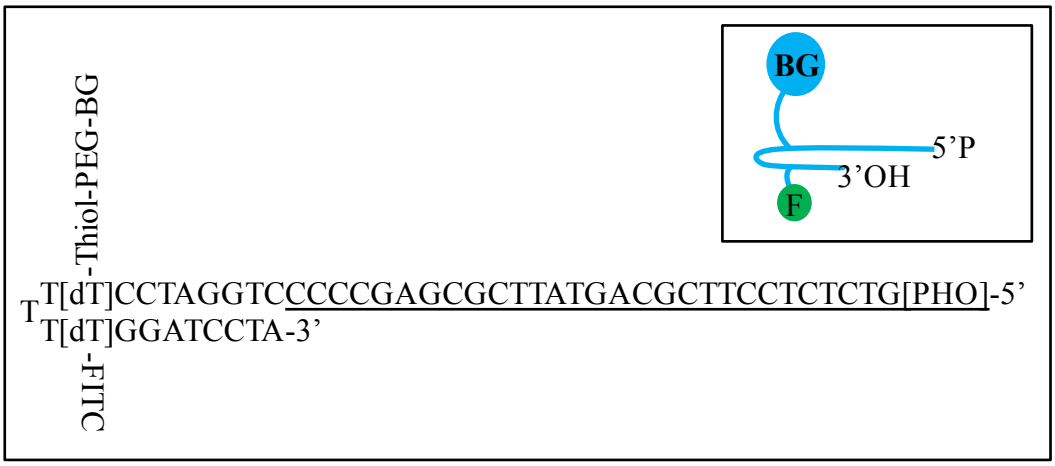


Figure 19: Sequence of BG-DNA-Linker (50 bp) used for the fusion of peptide with AGT. Inset showing schematic representation of the linker. The underlined sequence (29 bp) is the complimentary sequence which hybridises with the mRNA.

For the availability of BG, a unique modified BG- DNA linker was constructed to ligate at 5'-terminus of mRNA, the sequence of which is given in Figure 19.

The principal features of the linker are:

(i) a hybridization region, a short 29 bp sequence at the 5'-teminus which is complementary to the 5'- terminal end of the mRNA of interest and is used to hybridize the linker-DNA to mRNA molecule in order to maintain a free 5'-phosphate.

(ii) a PEG region, to protrude BG for easy binding with AGT.

(iii) BG, which will bind with the AGT expressed protein.

(iv) loop region, 3 thymine (T) bases for refold and easy hybridization.

(v) FITC moiety, for visualization of linked product.

(vi) a 3'-ligation site, a free 3'-hydroxyl-terminated end which forms a phosphodiester bond with the free 5'-phospahte-terminated end of the mRNA.

DNA sequence of 792 bp (without random library sequence) and 816 bp (with random library sequence) were constructed comprising a T7 promoter (17 bp) followed by a 29 bp of linker hybridization, we used AGT gene (543 bp) which encodes O^6-alkylguanine DNA alkyltransferase protein (AGT). Polyhistidine-tags (His tag) downstream AGT are used for affinity purification of polyhistidine-tagged recombinant proteins expressed followed by terminator region to eliminate the transcription process.

The final steps involved in H2H for mRNA display are depicted in Figure 17. The mRNA was first transcribed in cell-free transcription system from the desired H2H DNA sequence. The 2-fold molar excess of BG-DNA linker is hybridized and ligated to the 5' end of the 1-fold molar excess of transcribed mRNA whereas the 3'-terminus of mRNA was hybridized and ligated with 2-fold molar excess of 14 bp 3'-protection-primer. The ligated sample which have BG-linker at 5'-terminus and 3'-protection-primer at 3'-terminus is placed in coupled *in vitro* transcription and translation mixture for protein expression followed by covalent fusion of the expressed protein with BG of the linker at 37 °C and named this as 'Co-translation and fusion' reaction. As the fusion protein AGT is present at the 5'-terminal, the moment it is translated it is fused to BG and thereby giving a rapid mechanism for selection technology *in vitro*. I incorporated stop codon in order to get multiple copies of the linked protein. After fusion, the his-tag purified mRNA/protein fused sample is reverse transcribed to get the cDNA which can be amplified for further rounds of selection.

2.3.1. From DNA to linker-mRNA for H2H display

2.3.1.1. Before optimization

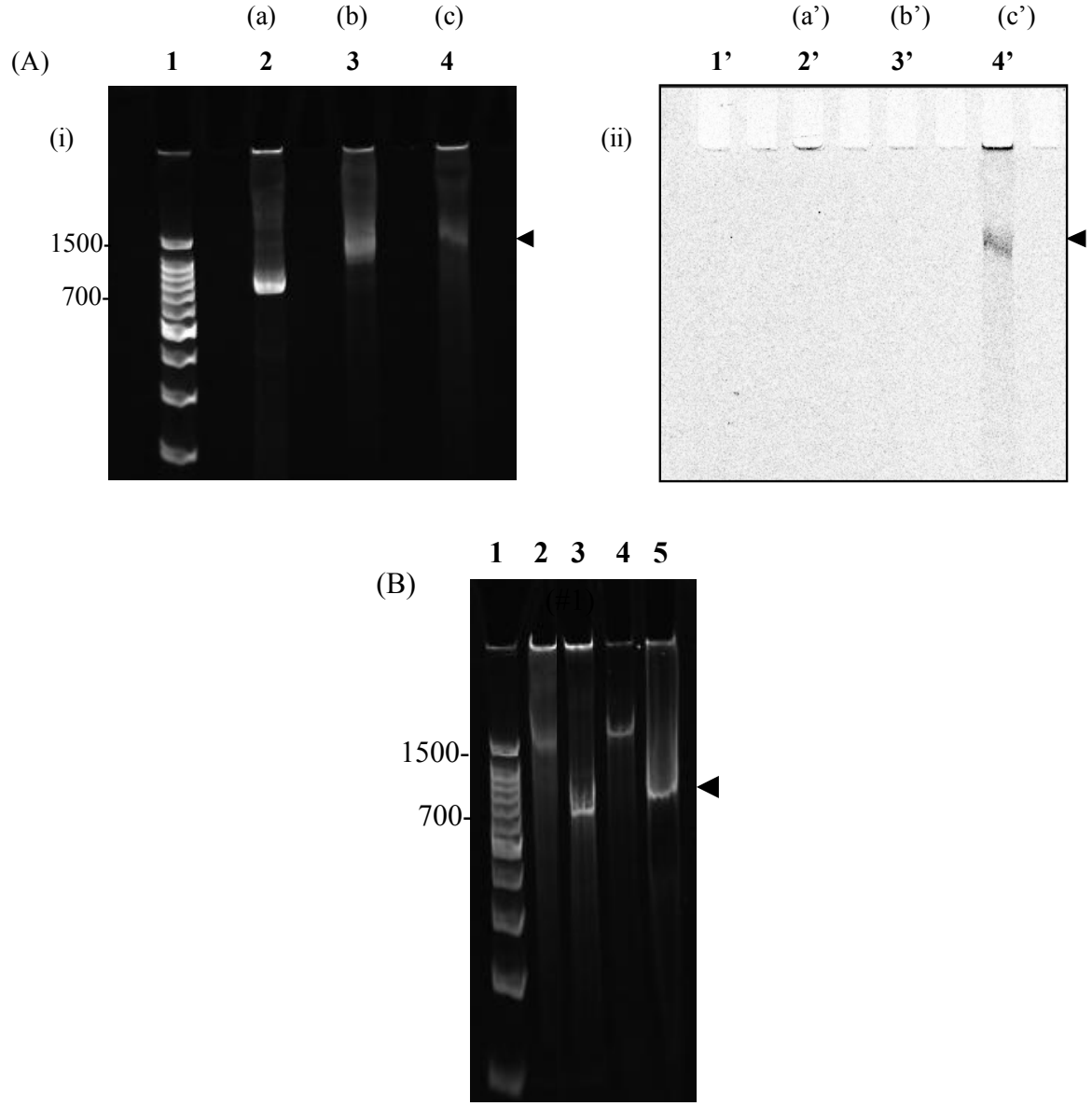


Figure 20: (A) 4% PAGE analysis of ligated sample with DNA and RNA. L1 and L1', 100 bp DNA ladder; L2 and L2', DNA; L3 and L3', RNA and L4 and L4', Ligated sample. (i) SYBR Gold stain (ii) Gel scanner result [Before selection].

(B) 4% PAGE analysis of ligated sample and RT sample treated with RNase. L1, 100 bp DNA ladder; L2, Ligated sample; L3, RT product of ligated sample; L4, RT sample treated with RNase H and L5, RT sample treated with RNase ONE. Successful RT (lane-3) with RT Primer was performed to confirm the formation of cDNA from RNA. The sample in lane-4 migrates slower than RT thus confirming formation of cDNA as single stranded DNA migrates slower

than RNA:DNA hybrid. Lane-5 have almost same band as lane-3 after digestion by RNase ONE indicates that it is DNA not RNA.

The desired DNA band from plasmid construct is shown in the 4% Native PAGE in Figure 20A(i). The RNA band is above DNA as RNA in native form migrates slower than DNA due to the formation of secondary structure[73-75]. The BG-DNA linker is hybridized and ligated to the 5' end of the transcribed mRNA. I observed a shift in ligated sample band (lane-4) compared to the RNA band as expected due to the addition of 50 bp of BG-DNA Linker. Figure 20A(ii) is the FITC fluorescence from the ligated sample (mRNA + linker DNA) containing BG-DNA Linker.

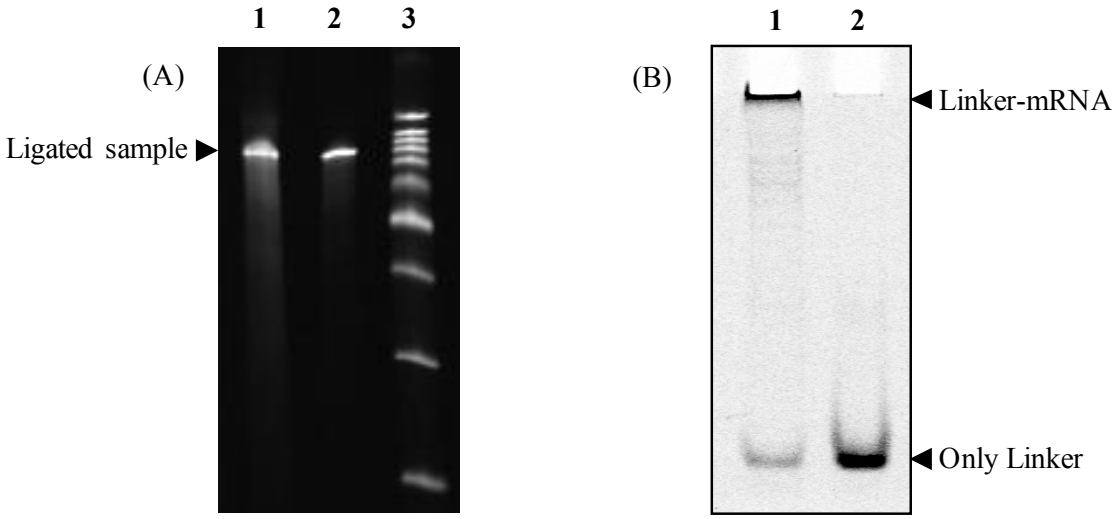


Figure 21: (A) 4% Urea PAGE (55°C) analysis of ligation sample and RNA with marker. The shift in ligation band (lane 1) indicates addition of linker DNA. L1, Ligated sample; L2, RNA and L3, 100 bp DNA ladder.
(B) 8% PAGE analysis of ligated sample treated with and without RNase H. L1, Ligated sample and L2, Ligated sample treated with RNase. Lane 1 is the ligated sample whereas lane 2 is only linker at 50 bp as RNA was digested by the RNase.

The denaturing urea gel was also performed to check the samples by denaturing the RNA and ligated sample (Figure 21A) and the same shift in band was observed as that of native gel. Also, to further confirm the ligation of BG-DNA Linker at 5'-terminus of mRNA, the ligated sample was digested with RNase H (Thermofisher Scientific) (Figure 21B) and only free linker was observed.

After successful ligation, I performed Reverse Transcription using RT-primer and the desired cDNA band between 700-800 bp was observed (Figure 20B). To further confirm the

generation of cDNA, RT product was digested by RNase H and the sample migrated slower than only RT product as single stranded cDNA migrates slower than RNA/cDNA hybrid and also, digested by RNase ONE, it was seen that the band still persist which means that it is DNA not RNA.

2.3.1.2. Addition of 3'-protection primer at 3'-end during ligation (Optimization)

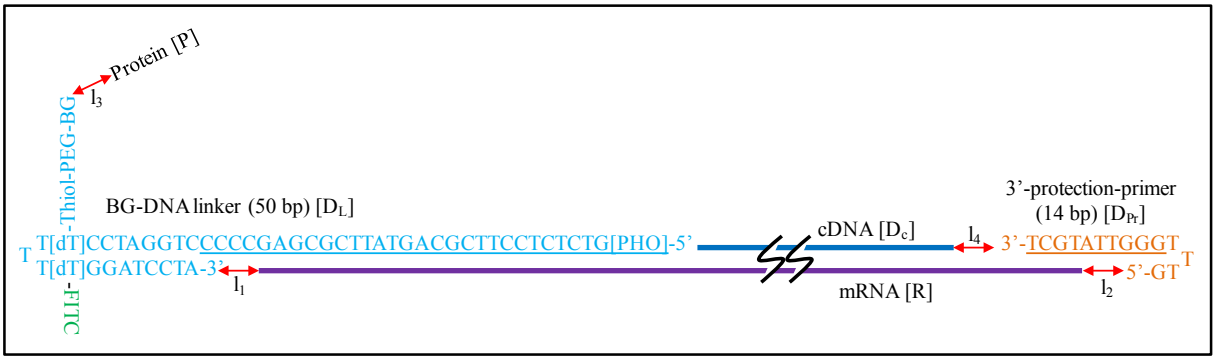


Figure 22: The nomenclatures used in H2H; the sequence of BG-DNA-Linker (50 bp) is shown in blue color whereas the sequence of 3'-protection-primer (14 bp) is shown in orange color. The underlined sequences are the hybridizing regions in both the linkers. R, RNA; D_L, BG-DNA-Linker; D_{pr}, 3'-protection-primer; P, Protein; l_1, Ligation between 5'-mRNA and BG-DNA-Linker; l_2, Ligation between 3'-mRNA and 3'-protection-primer; l_3, Linkage between 5'-mRNA and N-terminus of protein via AGT-BG; l_4, Link of 3'-protected mRNA to produce cDNA and D_c, cDNA formed using 3'-protection-primer as primer.

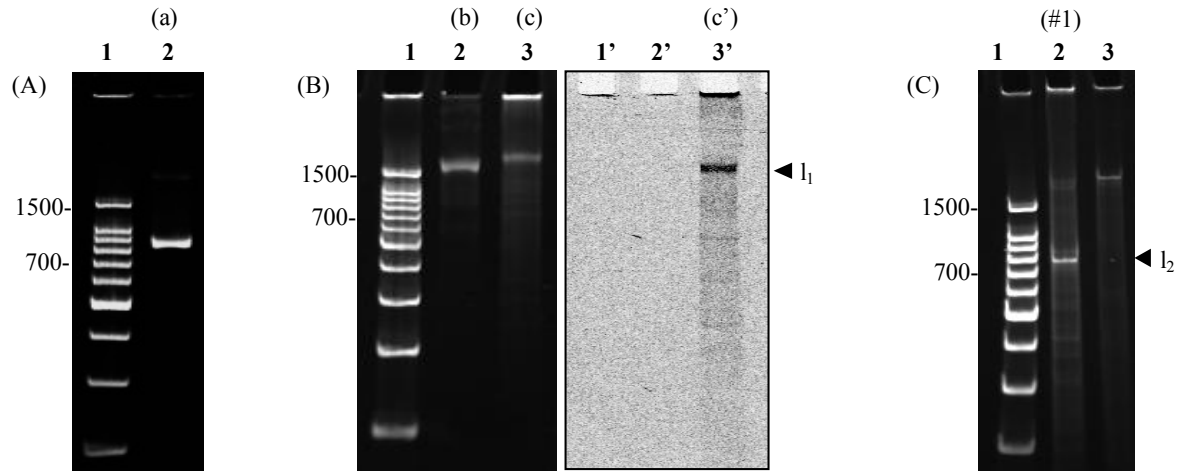


Figure 23: 4% PAGE analysis of different samples. (A) SYBR Gold stain of 100 bp DNA ladder with DNA. L1, 100 bp DNA ladder and L2, DNA. (B) SYBR Gold stain of 100 bp DNA

ladder with RNA and the ligated product (left). L1 and L1', 100 bp DNA ladder; L2 and L2', RNA and L3 and L3', Ligated sample. Gel scanner result of ligated sample [Before selection] (right). L1, 100 bp DNA ladder; L2, RNA and L3, Ligated sample. The fluorescence from L3 indicates successful ligation between 5'-BG DNA Linker [D_L] and mRNA; D_L have FITC moiety for fluorescence. (C) 4% PAGE analysis of RT sample. L1, 100 bp DNA ladder; L2, RT product of ligated sample and L3, RT sample treated with RNase H. Successful RT (lane-2) was performed to confirm 3' ligation of 3'-protection-primer with RNA. The band in lane-3 migrates slower than lane-2 as single stranded cDNA migrates slower than RNA/cDNA hybrid.

Figure 22 shows the nomenclature used in H2H experiment. The desired DNA band from plasmid construct is shown in Figure 23A. The 4% Native PAGE in Figure 23B, left image shows RNA band above 1500 bp, RNA in native form migrates slower than DNA due to the formation of secondary structure as described in the previous section. The BG-DNA linker is hybridized and ligated to the 5' end of the transcribed mRNA whereas to protect the 3'-terminus of mRNA, the 3'-terminus of mRNA was hybridized and ligated to 14 bp 3'-protection-primer. I observed a shift in ligated sample band in Figure 23B, left image compared to the RNA band as expected due to the addition of 50 bp of BG-DNA Linker and 14 bp of 3'-protection primer. Right image in Figure 23B, is the FITC fluorescence from the ligated sample (mRNA + linker DNAs) containing BG-DNA Linker. The ligation of 3'-protection-primer to mRNA was checked by performing Reverse Transcription and the desired cDNA band between 700-800 bp was observed (Figure 23C). To further confirm the generation of cDNA, RT product was digested by RNase H and the sample migrated slower than only RT product as single stranded cDNA migrates slower than RNA/cDNA hybrid.

2.3.2. H2H mRNA-protein fusion

2.3.2.1. Before optimization

After successful ligation, I performed fusion for 30 minutes as well as 120 minutes. The fusion sample were run in 8% Resolving and 4% Stacking SDS-PAGE. I observed that fusion efficiency increased from 30 min incubation (Lane 1) to 120 min incubation (Lane 2) in Figure 24, left image. When only mRNA fusion (acts as negative control here) was run with 120 min fusion products (acts as positive control), no band was observed in negative control Figure 24, right image; which shows that mRNA is not fused due to unavailability of

BG-Linker which further supports fusion but I was unable to detect the presence of mRNA in the sample as fused sample with and without RNase showed same band pattern.

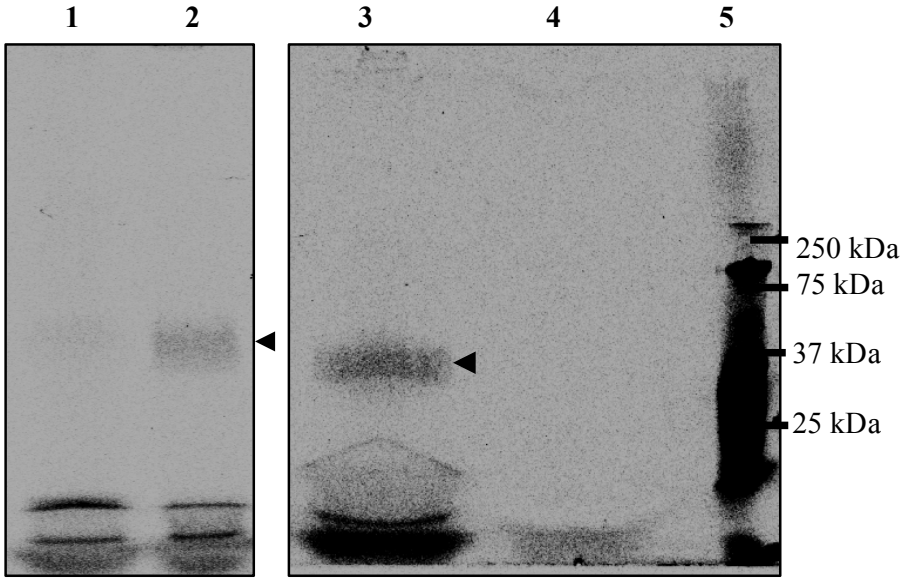


Figure 24: 8% Resolving and 4% Stacking SDS-PAGE analysis of 30 min and 120 min fused product. L1, 30 min fusion; L2, 120 min fusion; L3, 120 min fusion followed by treating with RNase H; L4, only mRNA expression and L5, Protein marker (Time comparison in left image & comparison with control sample in right image). The top line of both the gels are well-position where sample are loaded.

2.3.2.2. Addition of dummy RNA during coupled *in vitro* translation and linkage (Optimization)

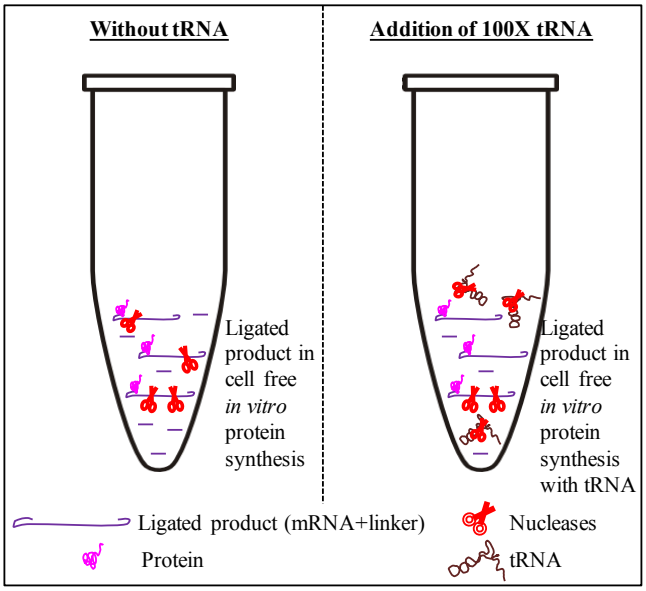


Figure 25: Schematic illustration of addition of dummy RNA into the fused sample.

To protect the mRNA from degradation of the ribonucleases during the fusion reaction, I first protected 3'-end of mRNA by ligating it to 3'-protection-primer and then second, I added yeast tRNA into the sample while performing coupled *in vitro* translation and linkage. Here, tRNA act as dummy RNA molecule to fool the ribonuclease present in the system and preserve the desired RNA molecule. Ribonuclease now have lots of food (and also variety) to eat and thus increases more chances of preserving the desired H2H construct (see Figure 25).

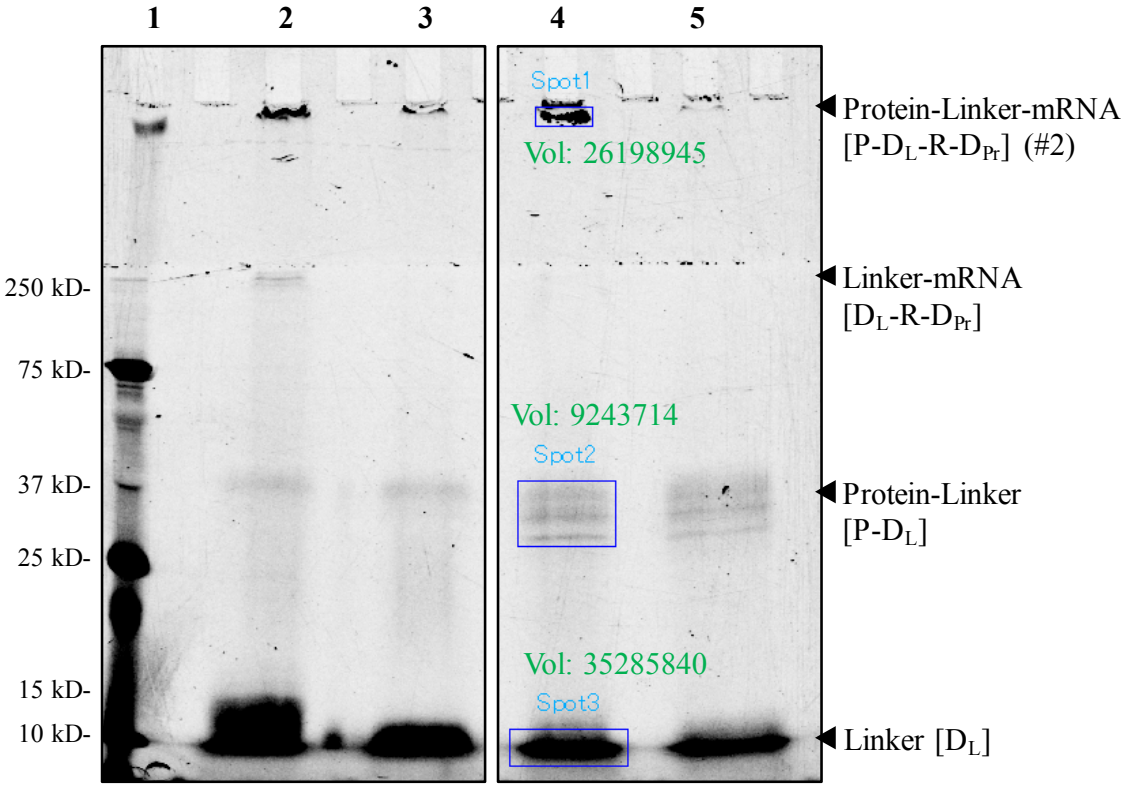


Figure 26: 12.5% Resolving and 4% Stacking SDS-PAGE analysis of 30 min and 120 min fused product. L1, Protein marker; L2, 30 min fusion; L3, 30 min fusion followed by treating with RNase H and RNase-ONE; L4, 120 min fusion and L5, 120 min fusion followed by treating with RNase H and RNase-ONE. Here, in lane-2, numbered (#2) is the 'Head-to-Head' linked protein with RNA via BG-linker whereas lane-3 is the protein with BG-linker without RNA; as the RNA is digested by RNase. I observed the same patter in lane-4 and 5 as that of lane-2 and 3, respectively. For the presence of protein in the united G-P molecule, we performed 'co-translation and fusion' for 120 minutes and observed the increase in intensity of band in lane-4 thus indicating the presence and increased amount of proteins in fused sample. This confirms the presence of genotype (mRNA) covalently linked to the phenotype (protein) in the fused product.

I confirmed the G-P covalent linkage (mRNA with protein) generated by H2H approach using SDS-PAGE analysis for 30 minutes as well as 120 minutes fusion (Figure 26) in the absence and presence of RNase. The fusion product in SDS-PAGE was detected by the fluorescence visualized from the FITC moiety attached to the DNA linker. For 30 minutes 'Co-translation and fusion' reaction, mRNA-linked-protein fusions were observed as a shifted band by gel electrophoresis due to an increase in the molecular weight caused by the covalent linkage between BG of BG-linker and the nascent protein. The shifted band was confirmed to contain mRNA by digestion of the matured fused product with RNase, a ribonuclease that cleaves RNA in smaller components. When mRNA/protein fused sample was digested by RNase, only protein with linker DNA was observed around 37 kDa. Also, when the fusion time was increased from 30 minutes to 120 minutes, we observed the same band pattern as that for short time i.e., 30 minutes. Moreover, increase in the intensity of band was observed due to more protein synthesis during 'co-translation and fusion' from 30 to 120 minutes.

Band name	Width	Height	Area	Mean	Std	Vol
Protein-Linker-mRNA [P-D_L-R-D_{Pr}]	5.96	2.08	12.41	5598.07	8995.13	26198945
Protein-Linker [P-D_L]	10.13	6.80	68.83	351.02	98.72	9243714
Linker [D_L]	12.21	3.88	47.41	1934.53	1921.86	35285840

$$\textbf{Fluorescence Intensity} = \frac{[\text{P-D}_L\text{-R-D}_{Pr}] + [\text{P-D}_L]}{[\text{P-D}_L\text{-R-D}_{Pr}] + [\text{P-D}_L] + [\text{D}_L]} \text{ X } 100$$

$$= \frac{[26198945] + [9243714]}{[26198945] + [9243714] + [35285840]} \text{ X } 100$$

$$\textbf{Fluorescence Intensity} = 50.11\ \%$$

Calculation 1: The fluorescence intensity calculation using Image Analysis software by Hitachi FMBIO III Multi View of the three bands in lane 4 for linking efficiencies. Firstly, the fluorescence of individual bands was calculated using the software followed by dividing the Protein-Linker-mRNA plus Protein-Linker (linked molecule) to Protein-Linker-mRNA plus Protein-Linker plus Linker (total linker input as ligated sample). Note: Each top of lanes is

imaged to be faintly dark due to the possible reflection of light used for fluorescence monitoring. Therefore, we need to consider this effect when we calculate the exact yield. In this study, we tried to obtain semi-quantitative results and neglected this faint band effect which might have affected by several % in the final yield (estimated to be around 20-50 %).

Lane 4 in Figure 26 allows for an estimation based on the amount of fluorescent FITC, that around 20–50 % of the BG-linker (contained in the forms of Protein-Linker-mRNA plus Protein-Linker) could bind to AGT under these experimental conditions, since the addition of the bottom band of the Linker only accounts for the total amount of the BG linker input. This experiment indicates the generation of the H2H construct-holding phenotype-genotype conjugate at a yield of around 20-50 % in a 2-h translation reaction. I tried to protect the mRNA from exonucleases at 5'-terminus with 5' ligation of BG-DNA linker and 3'-terminus with 3'-protection-primer ligation but I am not sure about endonucleases digestion site in the mRNA; thus I believe the three bands in lane 4 and 5 at P-D_L are due to mRNA degradation in 120 minute incubation (the more the time for incubation, the more the chances for mRNA decay). These three bands precisely named as P-D_L are in the range 30-40 kDa and I call them as a (first band), b (second band) and c (third band). I believe, a is full length protein-BG-DNA Linker-around 29 bp hybridized region of RNA to BG-DNA linker (P_F-D_L-R_{29}), b is partially degraded protein due to mRNA degradation in 120 minutes-BG-DNA Linker (P_{P1}-D_L), and c can be another partially degraded protein-BG-DNA Linker (P_{P2}-D_L); where P_{P1}> P_{P2}. The c band is sharp compared to a and b, possibly because of degradation of library region.

This confirms the presence of genotype (mRNA) covalently linked to the phenotype (protein) in the fused product in H2H construct.

2.3.3. Full-length translation by H2H

To confirm the full-length translation, His-tag purification was performed. The protein expressed using cDNA as template was his-tag purified and run in SDS-PAGE (refer to Figure 27), single band in the purified sample was observed as compared to the multiple bands observed in non-purified sample. This confirmed the correct translation of the full construct as his-tag domain perform correct function of binding to Ni-NTA beads and pulled down giving single band..

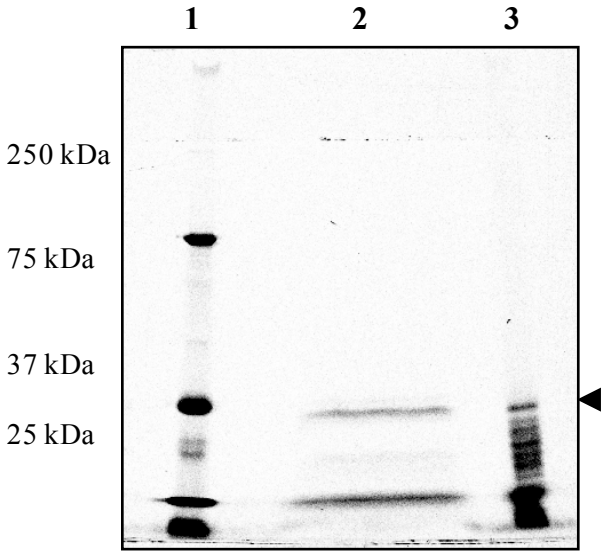


Figure 27: Gel scanner result of comparison between his-tag purified and non-purified protein in 12.5% Resolving and 4% Stacking SDS-PAGE. L1, Protein marker; L2, His purified sample and L3, Non-purified sample. A clear band was observed around 37 kDa (above then expected- 23.32 kDa) in purified sample (lane 2) as compared to non-purified sample indicating correct folding of His protein and complete (full-length) translation.

I then performed, his-tag purification to extract fused G-P molecule followed by one-step RT-PCR to amplify the cDNA for further use.

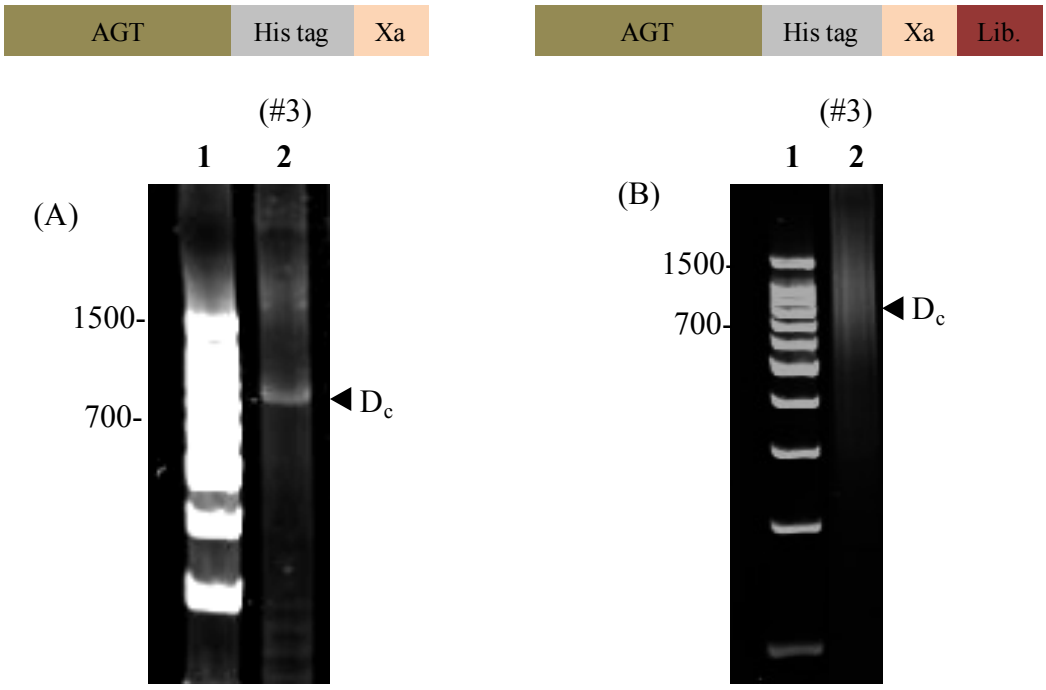


Figure 28: (A) 4% PAGE analysis of RT product after fusion. L1, 100 bp DNA ladder; L2, RT product of fused sample. The band in lane-2 is around 700 bp indicating formation of cDNA after fusion reaction [After selection].
(B) 4% PAGE analysis of RT product after fusion with 24bp. L1, 100 bp DNA ladder; L2, RT product of fused sample. The band in lane-2 in the form of a smear around 700 bp indicating formation of cDNA after fusion reaction (with 24bp) [After selection].

A band between 700-800 bp was observed in RT product after fusion indicating formation of cDNA after fusion reaction (Figure 28A). When the sequence containing random library was fused and RT was performed, a smear around 700 bp was observed indicating formation of cDNA after fusion reaction (Figure 28B). This cDNA band shows that our construct is suitable for multi-domain functionality.

2.3.4. H2H based multiple protein recoverability

H2H construct contains the stop codon UAA, so that from one mRNA multiple copies of the proteins can be generated; the linked protein to mRNA can be used to get the information of the selected protein whereas the multiple proteins formed from one gene can be used as functional unit to check the function/binding activity of the evolved protein without the need to further perform PCR or transformation. For this, I performed fusion

reaction for 120 minutes and observed not only linked proteins but also free protein (confirmed by FluoroTect Green lysine-incorporation experiment) which can act as functional unit for function based selection (Figure 29). I observed band for linked proteins at 37 kDa along with the generation of free proteins below 25 kDa. The free proteins were obtained at lower band position compared to linked proteins due to low molecular weight as they lack the linkage moiety. Under the consideration of the residue number of fluorescent lysines incorporated in the protein (maximally 11 residues; the expectation value is much lower due to the existence of ordinary lysine), the amount of freely translated proteins is sufficiently larger than that of the protein-mRNA conjugate. Intriguingly, an immature protein ladder is seen below the full length of the protein (24.9 kDa), probably indicating the existence of halts on the way of translation and/or truncated mRNA-dependent translation due to long time incubation (120 min).

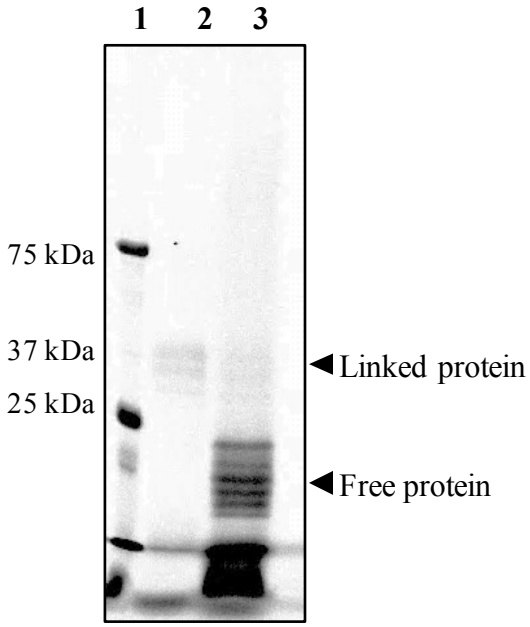


Figure 29: 12% Resolving and 4% Stacking SDS-PAGE analysis of 120 min fused product to show multiple protein recoverability of 5'-linker bound mRNA. L1, Protein marker; L2, 120 min fusion sample and L3, 120 min fusion sample FluoroTect lysine labelled. The bands around 37 kDa are linked protein (Protein-mRNA) whereas the bands below 25 kDa are the free protein which will act as free functional unit produced by H2H.

2.3.5. H2H to validate multiple-domain functionality

To check the multi-domain functionality of the expressed protein through H2H, we designed a new constructs containing AGT gene, His tag gene, Xa factor and Green Fluorescent Protein making it 1509 bp long construct. We used expressed cDNA to check H2H for DNA template.

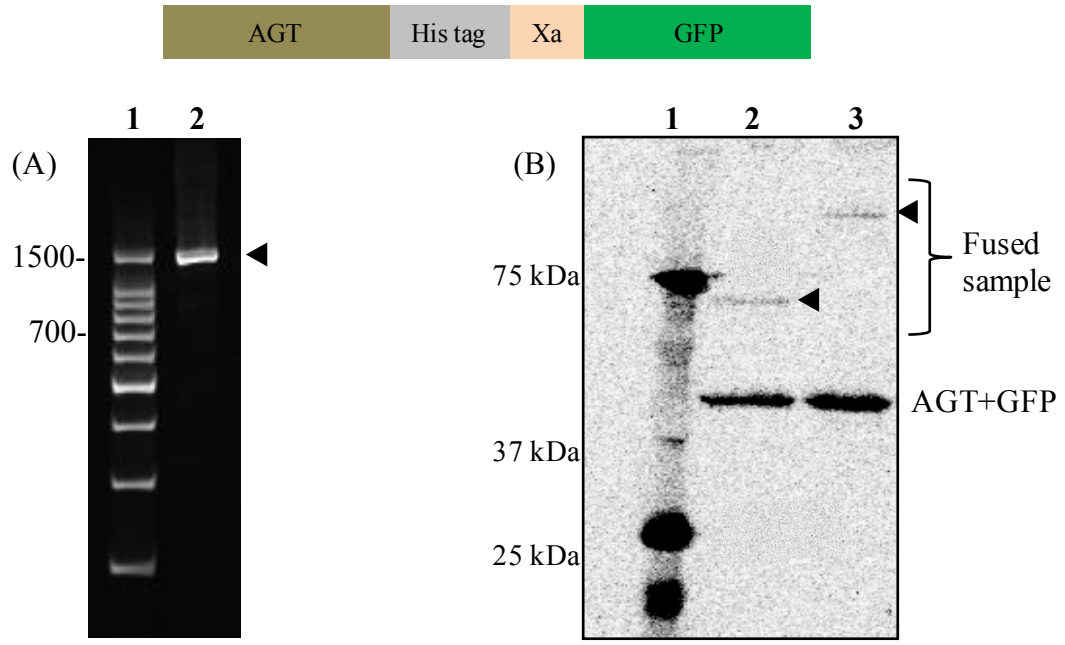


Figure 30: (A)4% PAGE analysis of 100 bp DNA ladder with DNA. L1, 100 bp DNA ladder; L2, DNA.
(B)12.5% Resolving and 4% Stacking SDS-PAGE analysis of 120 min cDNA expression with and without streptavidin. L1, Protein marker; L2, 120 min cDNA expression without streptavidin and L3, 120 min cDNA expression with streptavidin. The shift in the band in lane-3 was observed as streptavidin conjugated with biotin on BG-biotin linker capturing AGT and GFP.

GFP was used as model protein to check correct folding of the protein for multi-domain functionality validation. The desired DNA band migrated to 1500 bp as shown in Figure 30A. The amplified DNA and BG-biotin linker was placed in cell free coupled *in vitro* transcription and translation system without and with streptavidin. Each sample gave two bands (Figure 30B), one above 37 kDa and other just below 75 kDa in without streptavidin sample whereas above 75 kDa in with streptavidin sample. The common band above 37 kDa in both the samples is the free expression of GFP with AGT protein due to the native folding.

The band just below or above 75 kDa in both samples should be the conjugation of expressed AGT and GFP on BG. The shift in lane-3 indicates that streptavidin binding to BG-biotin linker capturing AGT and GFP makes it heavy molecule thus slow migration in the gel. The fluorescence comes from GFP, which indicates that our H2H concept gives functional folding of the nascent protein and can be used for longer transcript.

Conventional techniques like ribosome display, mRNA display involves the linking of 3'-terminus of mRNA with C-terminal of protein. Here, I report a novel H2H concept for multiple protein-recoverable display for rapid *in vitro* protein selection. 'Head-to-Head' display is the covalent conjugation of 5'-terminus of mRNA with N-terminal of protein. I obtained covalent genotype-phenotype mRNA-protein fusion in head-to-head direction. The H2H approach possesses a number of unique features. Firstly, the ease and short time period with which the entire process is done. Secondly, the H2H construct has been designed to consist of stop codon unlike that of conventional display techniques so that one mRNA template can be directly linked with one protein and generating multiple copies of the protein thus allowing information-to-function linkage. Thirdly, using H2H construct, post-translation modification can be easily done utilizing free C-terminus (C-terminal modification of proteins). Fourthly, generally, the C-terminus of protein is exposed and N-terminus is inside the protein structure thus our H2H would be most applicable technique as in H2H molecule C-terminus of protein is exposed to the surroundings. Thereby, our H2H technique provides significant importance for rapid characterisation of peptide hormones, peptide-ligands, membrane proteins, protein-protein interaction domain, C-terminal PDZ domain-ligand (PDZ-domain), C-terminal epitope of Alzheimer's disease specific Ab, F_c site of Ab to recognize Ag.

2.4. Conclusion

In conclusion, I have developed a simplified, easy and rapid display technology with high efficiency of G-P yield (~20-50%). Unlike the conventional molecular display technology, 'Head-to-Head' is based on linking the 5' of gene with N-terminal of protein. The optimizing parameters for H2H are the protection of 3'-end of mRNA via 3'-protection-primer ligation and addition of dummy RNA to protect it from ribonucleases. H2H construct can be used not only for C-terminus post translational modifications but also for recoverability of multiple copies of the proteins from a single gene for simple and faster *in vitro* directed evolution. From single molecule, we can get both the sequence information via linkage and multiple copies of proteins generated can be directly used to check their function without the need of additional steps. Thus, it can be used as easy and simple technique to find/isolate peptide-based aptamer for function-based screening with improved G-P yield.

CHAPTER III

DEMONSTRATION OF ELECTROSPRAY-BASED *IN VITRO* COMPARTMENTALIZATION FOR ONE-STEP *IN VITRO* PROTEIN SELECTION

CHAPTER III

Demonstration of electrospray-based *in vitro* compartmentalization for one-step *in vitro* protein selection

3.1. Introduction

The efficiency and realization of the whole *in vitro* selection system or molecular evolution depends to a large extent on the intactness of given starting library size, prior to selection, and then of the recovery of selected candidates, after the selection. Nevertheless, this is largely compromised by nucleases and proteases contaminations, originated either by internal and/or external sources, which can often not be controlled effectively during the process of *in vitro* selection. First, utilization of cell free expression system, a prerequisite for devising *in vitro* selection, which often contain endogenous nucleases and proteases (internal source) leads to undesirable activities. For example, a delay of few minute incubation in crude lysate-based cell free expression system leads to degradation of the library; for instance, the majority of mRNA molecules have lifetimes between 3 to 8 minutes in *E. coli* crude extract-based system[76]. As an alternative to the crude lysate-based cell free expression system, PURE system has been introduced which is prepared by individually overexpressing and purifying all essential factors required for *in vitro* protein expression with minimal detrimentals[77]. Although attractive, PURE system suffers with a low translation efficiency and misfolding of the nascent polypeptides[78]. Secondly, external source of detrimentals via an uncontrolled DNA/RNA handling in the laboratory can lead to the degradation of starting library and so thus can reduce the number of identifiable hits and void output of a selection round. Consequently, as the rounds of selection are performed, the starting information of library may not completely retained leading to reduced number of candidates for *in vitro* selection than of the size of starting library. Therefore, a full realization of *in vitro* selection of peptide/proteins using display technologies is highly depended to a large extent on protecting all the variations of starting library without loss, followed by preserving nascent polypeptide chain with the encoding sequence information (i.e., efficient coupling of phenotype and genotype) thoroughly during the process of *in vitro* selection.

As described in the previous chapter, I established a covalent genotype-phenotype linkage for simplified and rapid molecular evolution or *in vitro* selection. But, the covalent G-P molecule was largely suffered by the ribonuclease attack thus causing degradation of the

mRNA which leads to the loss of the initial information. This is mainly because of the noise or background created while performing the reactions.

We very well know that as we scale down the biomolecular reaction from 'micro-to-nano-to-femto' the chances of degradation reduce thus, opening a new door to protect the initial information. In this regard, I have focused on compartmentalizing the DNA and its encoding protein by performing the 'Head-to-Head' covalent genotype-phenotype linkage in water-in-oil droplets generated by electrospray for one-step *in vitro* protein selection.

3.2. Experimental

3.2.1. Electrospray setup and procedure

A 90-mm glass capillary was used to make ~40 µm glass nozzle by Puller machine PC-10 (Narishige, Japan) using two step modes, keeping heater level 60 and 50. The ~40 µm glass nozzle was dipped in a small hole with an approximately 6-mm diameter punched into commercially available silicone rubber (3 cm × 3 cm × 3 mm) and placed on a glass slide (35 × 55 mm) to construct a chamber for droplet collection. Tungsten wire was used as electrode. The water-in-oil droplets were generated in 100 µl of an oil/surfactant mixture containing 50% ABIL EM 90 (Evonik Industries), 36% Tegosoft DEC (Evonik Industries), and 14% mineral oil (Sigma Aldrich) at voltage of 1000 V and frequency of 100 Hz. The oil mixture used in the experiments were freshly prepared by vortexing at 2,500 rpm for 5 minutes followed by incubation at 30 °C for 30 minutes.

3.2.2. Coupled *in vitro* translation and linkage in electrospray water-in-oil droplet

The AGT-GFP template was prepared by PCR amplification of plasmid pEX-K4J1-GeneMGMT-GFPuv4 encoding AGT and GFPuv4 followed by purification using a PCR clean-up kit. The DNA concentrations were determined by absorbance at 260 nm using a NanoDrop. The sequence of which is given in Figure 31.

2.5 µL of purified cDNA template encoding AGT and green fluorescent protein (GFPuv4, the brightest variant of GFPuv) was gently mixed with 8.75 µL of the cell-free translation system and the mixture was incubated at 37 °C for 120 min to permit cell-free coupled expression and linkage within droplets. The glass nozzle was pre-rinsed with an aliquot of the reaction mixture followed by pipetting fresh solution to avoid the loss of template DNA or other components of the cell-free system by adsorption onto the inner glass surface of the nozzle. The 5'-thiol modified forward primer was coupled to BG using BG-

Maleimide (NEB) according to manufactures instruction whereas reverse primer was attached with dual biotin molecule. Biotin labeled DNA is used to immobilize on beads which can then be easily isolated using magnetic separation. A dual biotin with two biotin molecules in sequence can increase binding strength with streptavidin. This helps to keep biotinylated DNA on the beads during heating at higher temperatures. It has been seen that dual biotin prevents or effectively reduces leakage of biotinylated DNA from beads during heating[79]. Biotinylated cDNAs were immobilized to streptavidin dynabeads myone magnetic beads followed by spitting through nozzle in electrospray set-up. We produced both kinds of droplets with and without beads and observed under confocal laser scanning microscope (Olympus FluoView 1000 spectral-based) to capture brightfield and fluorescence microscopic images of the droplets. A laser light with an Ar 488-nm wavelength was used to excite the GFP, which was observed through an Alexa Fluor 488 green dye filter. All images were captured using a 60x lens that was focused on the equatorial section.

GAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTT
AACTTTAAGAAGGAGATATACCAATGGACAAAGATTGCGAAATGAAACGTACCACCCTGGAT
AGCCCGCTGGGCAAACTGGAACTGAGCGGCTGCGAACAGGGCCTGCATGAAATTAAACTGCT
GGGTAAAGGCACCAGCGCGGCCGATGCGGTTGAAGTTCCGGCCCCGGCCGCCGTGCTGGGTG
GTCCGGAACCGCTGATGCAGGCGACCGCGTGGCTGAACGCGTATTTTCATCAGCCGGAAGCG
ATTGAAGAATTTCCGGTTCCGGCGCTGCATCATCCGGTGTTTCAGCAGGAGAGCTTTACCCG
TCAGGTGCTGTGGAAACTGCTGAAAGTGGTTAAATTTGGCGAAGTGATTAGCTATCAGCAGC
TGGCGGCCCTGGCGGGTAATCCGGCGGCCACCGCCGCCGTTAAAACCGCGCTGAGCGGTAAC
CCGGTGCCGATTCTGATTCCGTGCCATCGTGTGGTTAGCTCTAGCGGTGCGGTTGGCGGTTA
TGAAGGTGGTCTGGCGGTGAAAGAGTGGCTGCTGGCCCATGAAGGTCATCGTCTGGGTAAAC
CGGGTCTGGGATCCATCGAGGGCCGCATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTC
CCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGCGGAGAGGG
TGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTAC
CTGTTCCATGGCCAACACTTGTCACTACTCTGACGTATGGTGTTCAATGCTTTTCCCGTTAT
CCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGA
ACGCACTATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAG
GTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATT
CTCGGACACAAACTCGAGTACAACTATAACTCACACAATGTATACATCACGGCAGACAAACA
AAAGAATGGAATCAAAGCTAACTTCAAAACTCGCCACAACATTGAAGATGGCTCCGTTCAAC
TAGCAGACCATTATCAGCAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAAC
CATTACCTGTCGACACAATCTGCCCTTTTGAAAGATCCCAACGAAAAGCGTGACCACATGGT
CCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAATAAT
AAATCGGCAGCTAGCATAACCCCTT

T7 Promoter (From PURE): 22 bp
RBS (From PURE): 63 bp
pSNAP code: 543 bp
Spacer + Xa: 18 bp
GFPuv4+Stop Codon: 714+6 bp
Spacer: 9 bp
3' Partial T7 Terminator: 14 bp

Figure 31: DNA Construct (1389 mer) mainly consisting of gene for AGT and GFP for 'Head-to-Head' covalent genotype-phenotype linkage in water-in-oil droplet experiment.

3.3. Results and Discussion

With the aim to protect the genetic information, 'Head-to-Head' covalent genotype-phenotype linkage was performed in electrospray-based femtoliter water-in-oil compartments as depicted in Figure 32.

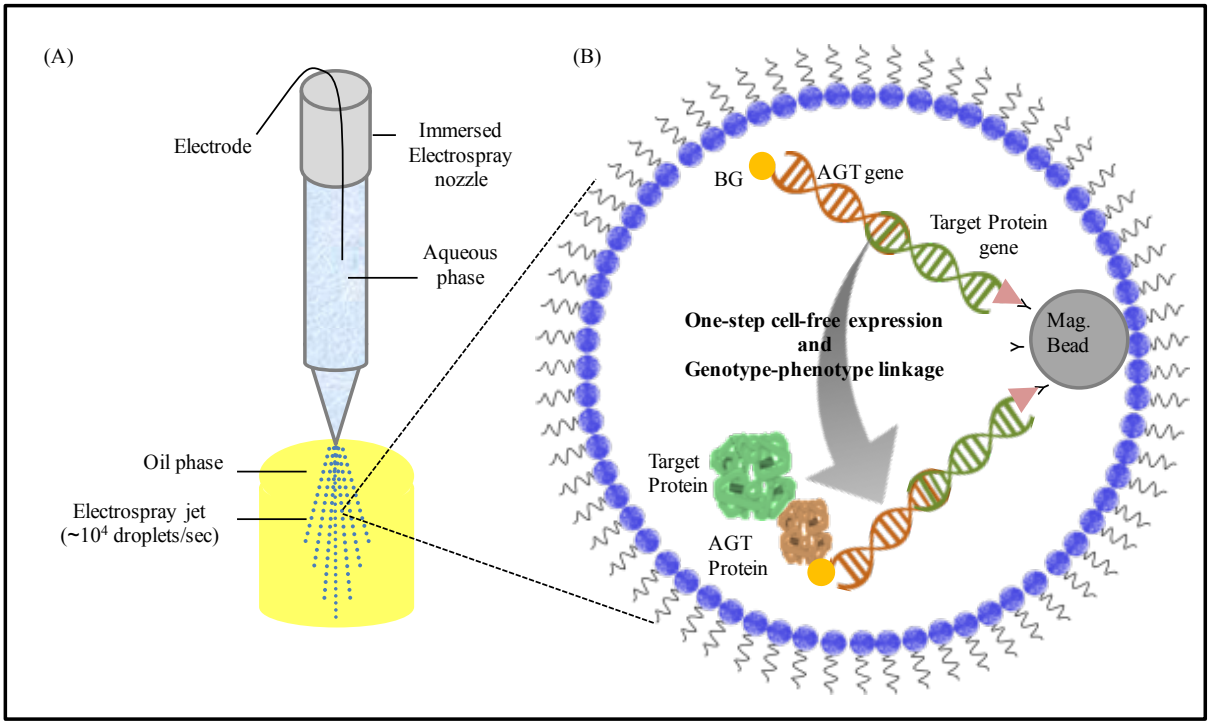


Figure 32: (A) Schematic representation of electrospray-droplet set-up and (B) Enlarged view of single droplet showing 'Head-to-Head' genotype-phenotype linkage on bead.

The glass nozzle of ~40 µm was immersed in oil phase in a chamber on a glass slide to collect water-in-oil droplets as shown in figure 32A. The mechanism of one step covalent genotype-phenotype linkage is shown in figure 32B, where DNA molecule is encapsulated with cell-free *in vitro* translation system. As soon as the protein is translated having both AGT and GFP, the AGT covalently bonds to the BG bound to the primer if available in the droplet to undergo one-step cell-free expression and genotype-phenotype linkage. The green fluorescence comes from the expressed GFP. It is important to mention here that, droplets will show fluorescence as not only BG bound GFP but also freely expressed GFP. Therefore, in the following microscopic images we can see that two kinds of fluorescence is obtained (positive sample).

3.3.1. One step H2H covalent G-P linkage in electrospray-based *in vitro* compartments (without beads/off beads)

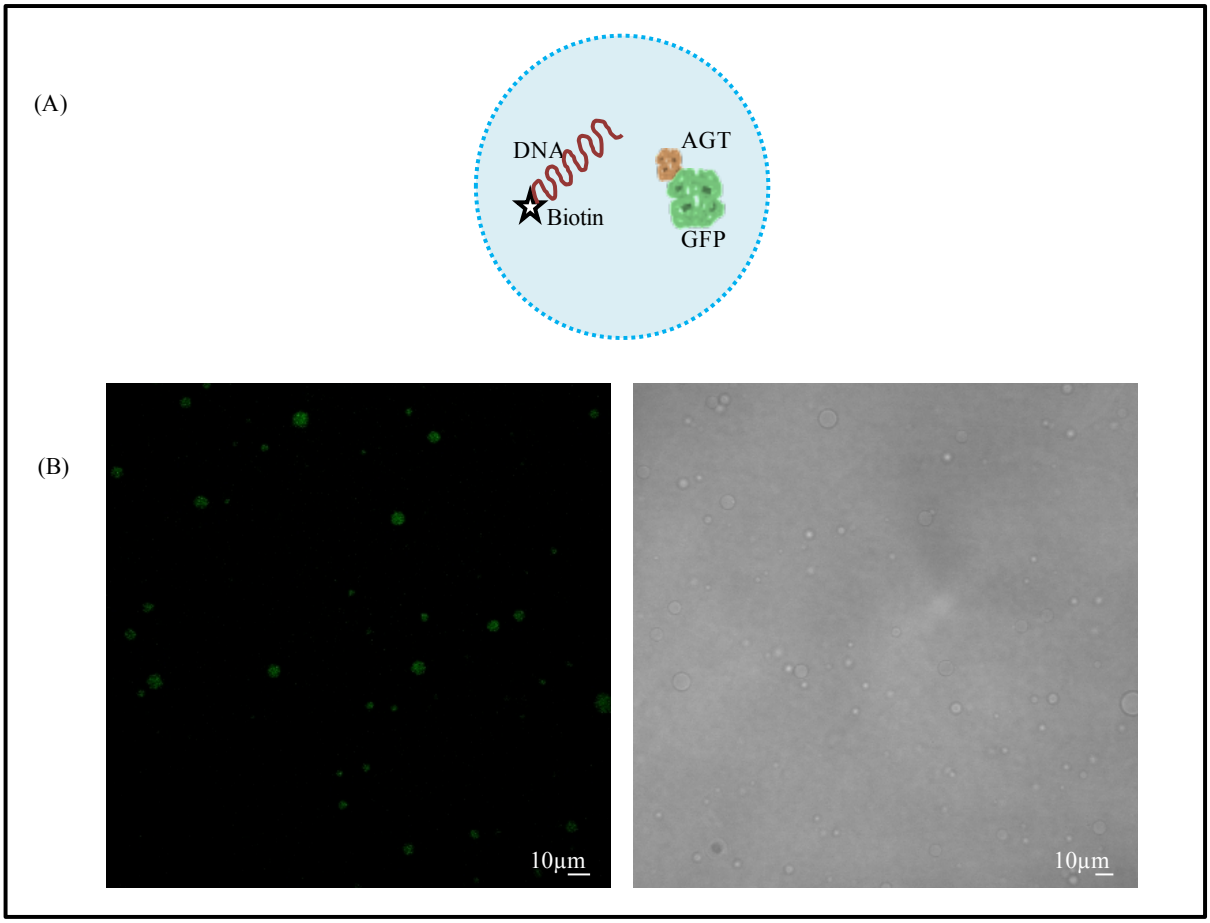


Figure 33: (A) Schematic representation of water-in-oil droplet showing only DNA. Free expression of GFP in the absence of BG (without beads).
(B) Fluorescence (left) and Brightfield image (right) of droplets generated in oil mixture showing free expression of GFP [Control sample - without BG].

When DNA without BG primer (acts as control sample here) was encapsulated in droplets only one kind of fluorescence was observed Figure 33B i.e., spread all over, as per my understanding it was due to the free GFP expression, in this case no concentrated fluorescence was observed (as in case of positive sample) due to the absence of BG leading to no covalent G-P fusion.

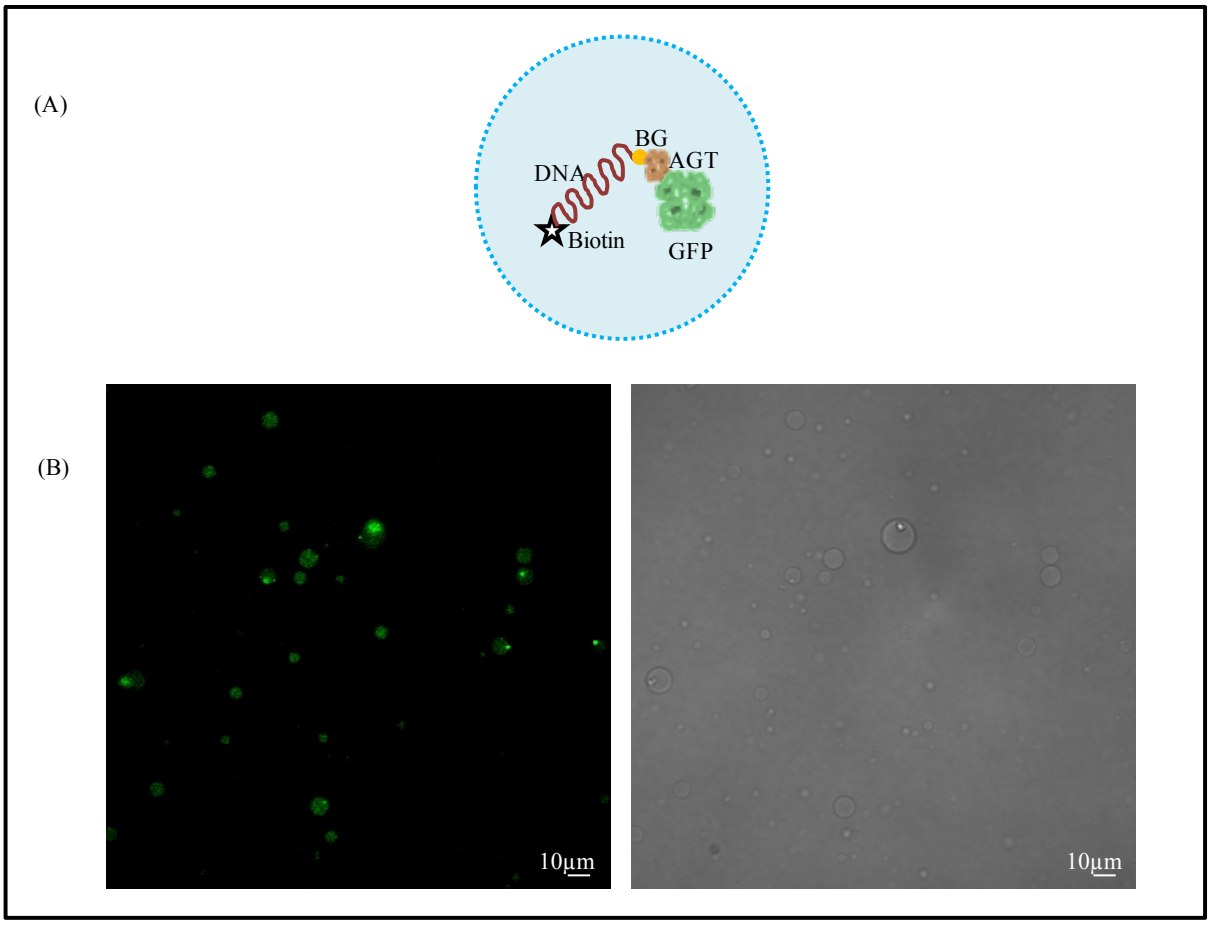


Figure 34: (A) Schematic representation of water-in-oil droplet showing only DNA. Concentrated fluorescence indicates the G-P covalent linkage in the presence of BG (without beads).
(B) Fluorescence (left) and Brightfield image (right) of droplets generated in oil mixture showing one step genotype-phenotype linkage [Positive sample - with BG].

On the contrary, when the DNA containing BG primer was encapsulated in w/o droplets: two kinds of fluorescence was observed, one concentrated at some places within the droplet and the other spread all over in the droplets as shown in Figure 34B. The concentrated fluorescence is believed to be because of the one step genotype-phenotype covalent linkage due to the presence of BG whereas fluorescence spread in all around the droplet is because of free expression of GFP. Fusion process between protein and mRNA via BG might cause not only bright spots inside droplets but also whole bright droplet due to the gathering of ribosomes, template molecules and fused complex at the periphery/surface of the droplet (droplet surface-enhanced reaction)[80, 81].

3.3.2. One step H2H covalent G-P linkage in electrospray-based *in vitro* compartments (with beads/on beads)

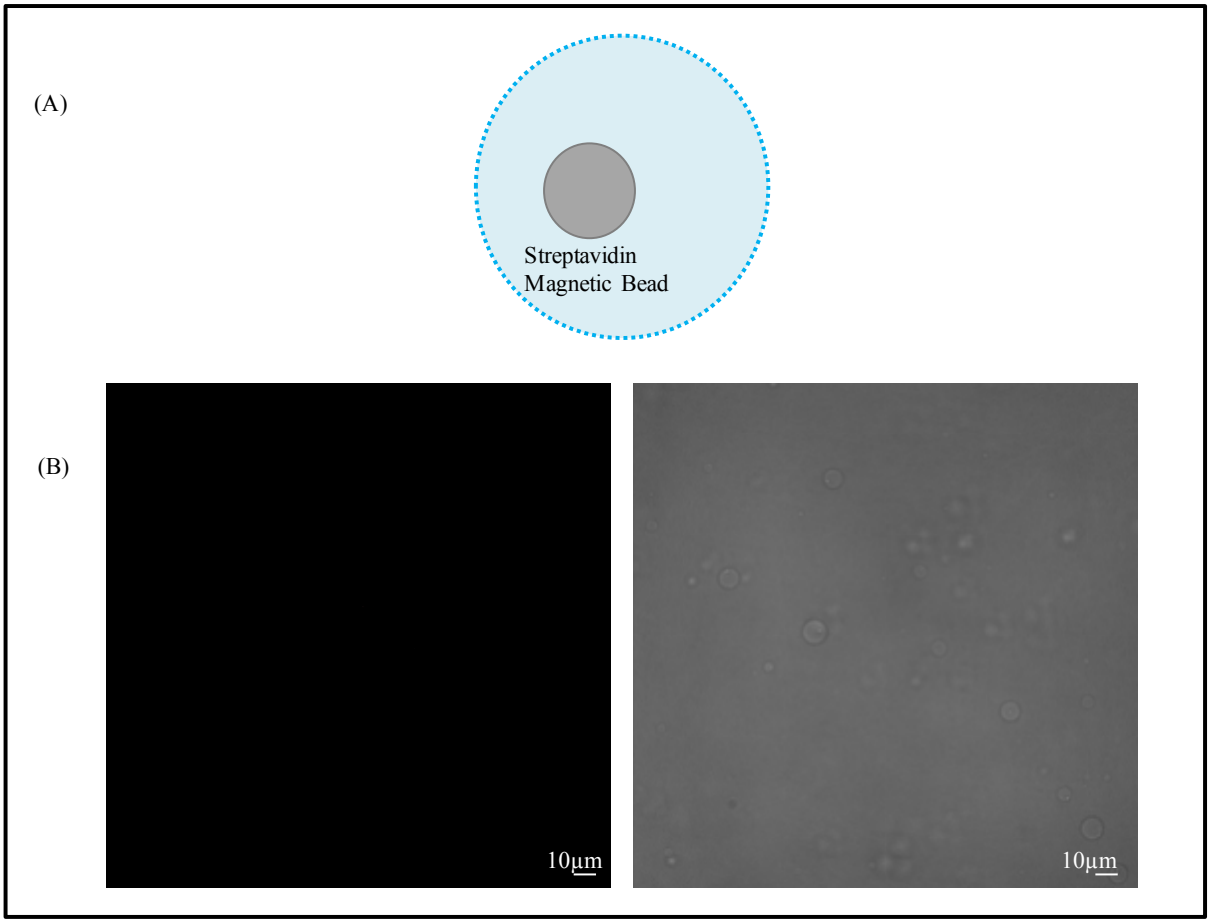


Figure 35: (A) Schematic representation of water-in-oil droplet showing magnetic bead. (B) Fluorescence (left) and Brightfield image (right) of droplets generated in oil mixture encapsulating only magnetic beads. No fluorescence from beads / no background in droplets was observed.

The H2H electrospray droplet display can also be used for the solid-phase translation and fusion[82]. I used streptavidin magnetic beads to immobilize biotinylated DNA and performed one step covalent G-P linkage. But first, to check for the autofluorescence from the magnetic beads, the only beads sample was encapsulated in water-in-oil droplets and no fluorescence was observed (Figure 35).

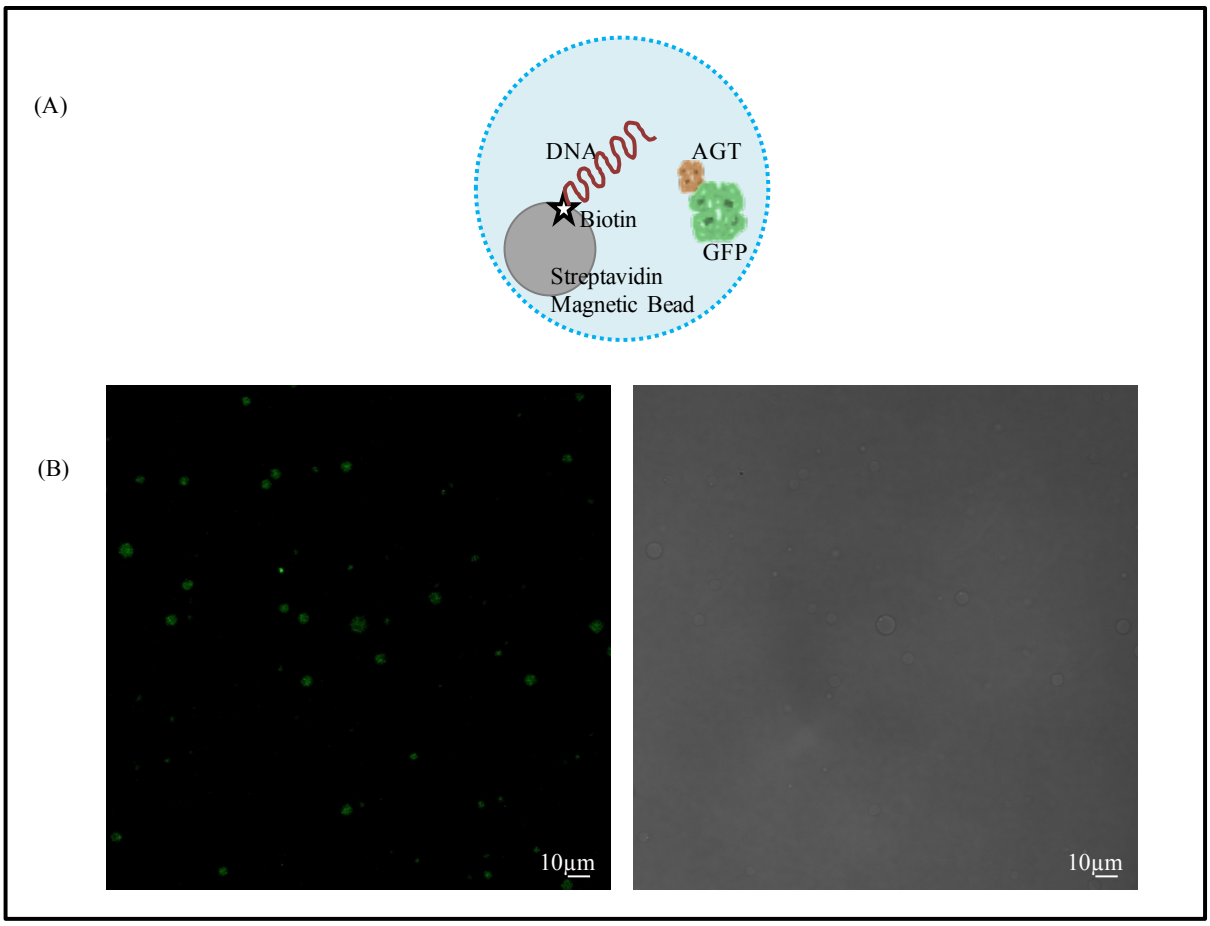


Figure 36: (A) Schematic representation of water-in-oil droplet showing DNA immobilized on magnetic bead. Free expression of GFP in the absence of BG (with beads).
(B) Fluorescence (left) and Brightfield image (right) of droplets generated in oil mixture showing free expression of GFP [Control sample - without BG].

In this case also, only all-over fluorescence was observed in DNA sample without BG (Figure 36B) whereas, I observed two kinds of fluorescence, one concentrated at some places in the droplet and the other spread all over in the droplets (Figure 37B) containing BG primer DNA. These results obtained are matching with GFP expression in the gel (Chapter-2, Figure 30B).

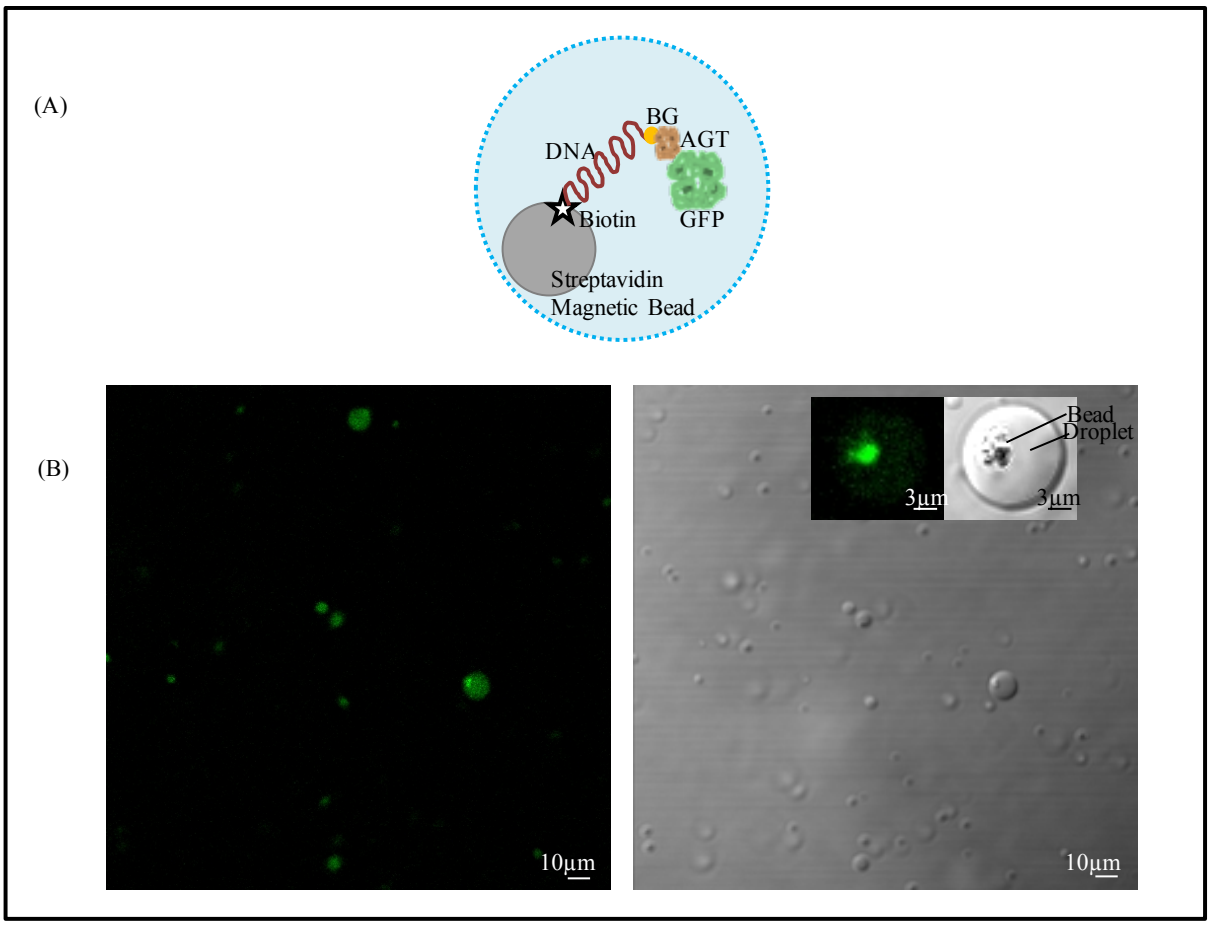


Figure 37: (A) Schematic representation of water-in-oil droplet showing DNA immobilized on magnetic bead. Concentrated fluorescence indicates the G-P covalent linkage in the presence of BG (on beads).
(B) Fluorescence (left) and Brightfield image (right) of droplets generated in oil mixture showing one step genotype-phenotype linkage on beads [Positive sample - with BG]; inset showing the enlarged view of single droplet with bead.

In here, I tried to encapsulate DNA with its expressed protein in H2H within water-in-oil droplets produced by electrospray. These cells like compartments enables in one-to-one linkage of individual gene with its exactly same phenotype with increased yield and efficiency for one step selection-on-demand.

3.4. Conclusion

With the aim to produce 'highly-selective' functional molecules for molecular evolution engineering, in this chapter, I tried to integrate the 'Head-to-Head' genotype-phenotype covalent linkage in the femtoliter sized water-in-oil droplets generated by electrospray technology. These cell like compartments enables in one-to-one linking of genotype with its encoding phenotype for efficient and high yield selection-on-demand by producing H2H beads in the future. As in one step the DNA is converted and covalently linked to the protein within small compartments hence, this can be used for one-step *in vitro* protein selection.

CHAPTER IV

DEVELOPMENT OF ELECTROSPRAY MICROARRAY *IN VITRO* PROTEIN SELECTION FOR IMPROVED LIBRARY SIZE

CHAPTER IV

Development of electrospray microarray *in vitro* protein selection for improved library size

4.1. Introduction

The efficiency of cell-free system has been redefined by the use of IVC technology. The more the size has been reduced from micro-to-femto, the more is the rate of reaction[41]. Time and again, IVC has been widely developed from vortexing and homogenizers for bulk emulsification to microfluidic-based water-in-oil generation to recently using electrospray for sub-fL scale monodisperse droplets. Homogenizer and vortexing face the problem of variation in size, microfluidic-based droplet generation struggles with the speed whereas the use of single nozzle in electrospray can produce library size of 10^{8-9} and still consumes lot of time. A single nozzle in electrospray using frequency of 1 kHz can generate near 50000 droplets per second (B. Sharma et al., 2016) and thus this setup will requires hours-to-days for encapsulation of µl-scale cell-free reaction into fL-scale IVC, e.g., the encapsulation of 10 µl cell-free reaction in 1 fL IVC will requires 55 hours.

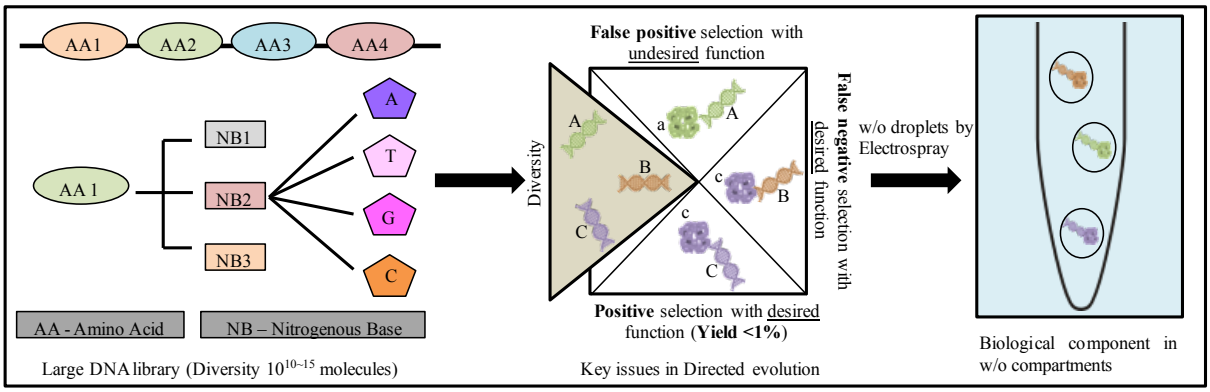


Figure 38: Overall scheme of our rapid ultra-low volume in vitro compartmentalization.

Many groups have worked on multi-electrospray to produce aerosol, space propulsion, nanofibers, nanoparticles and emitters[83, 84]. In order to encapsulate "million-to-quadrillion" molecules (see Figure 38) into 1 fL IVC, a system which can encapsulate the entire library in minutes is required to be developed. In this regard, I focussed on using multi electrospray[85, 86], the concept of multiple nozzle[87, 88] to multiple the number of droplets generated by single nozzle. To produce massive amount of droplet, I integrated the use of micro-hole array chip

which converts the µ-scale drop into subfemtoliter-scale droplets (see Figure 39) to fulfil the demand of larger library size (more than 10^{8-9}) for *in vitro* protein selection.

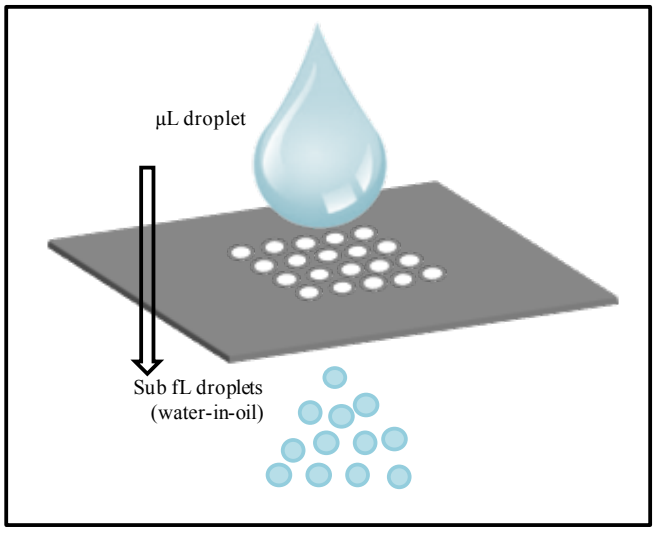


Figure 39: The conversion of µ-scale drop to subfemtoliter-scale droplets via micro-array.

A micro-hole array chip with immersed electrospray technology is designed and demonstrated to realize ultrahigh-throughput generation of highly monodisperse IVC with volume in subfemtoliter-scale (see Figure 40B). A conductive liquid solution is electrospray through an electrified micro-hole array chip in an immiscible-phase (a mixture of oil and surfactant). The electric field is generated by applying a high-voltage drop from the solution-phase, and the liquid is extruded and sprayed through micro-hole array into the oil-phase in the form of monodisperse-droplet jets with the diameters in the sub-micrometer range.

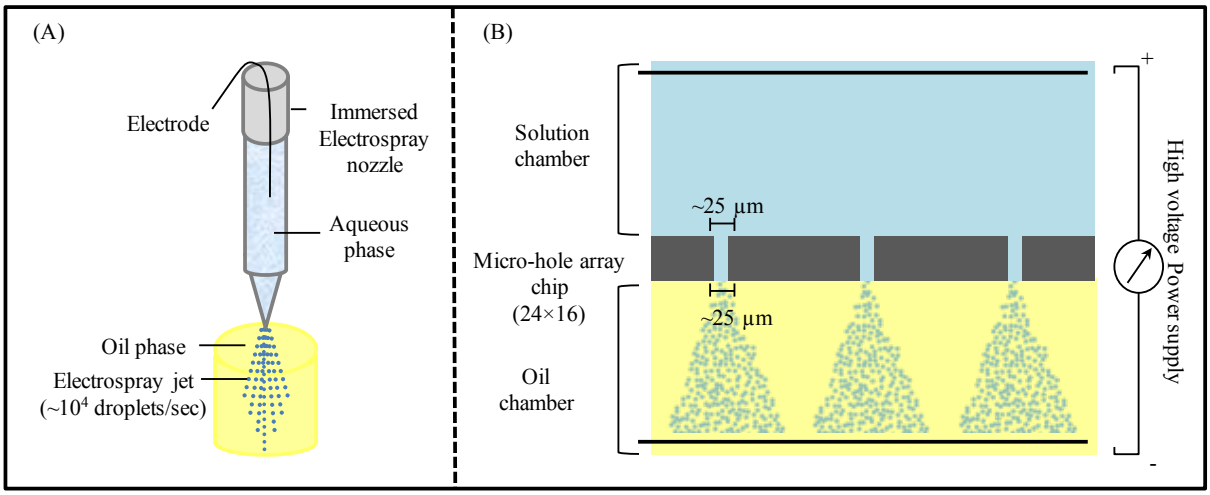


Figure 40: The comparison between single nozzle electrospray and micro-hole array electrospray (µHAES).

(A) Single nozzle electrospray-droplet set-up.

(B) Overall concept of μHAES to produce subfemtoliter-scale droplets.

4.2. Experimental

4.2.1. Fabrication of micro-hole array chip

Micro-hole chip was fabricated in six basic steps as follows:

Step-1: Prepare glass substrate.

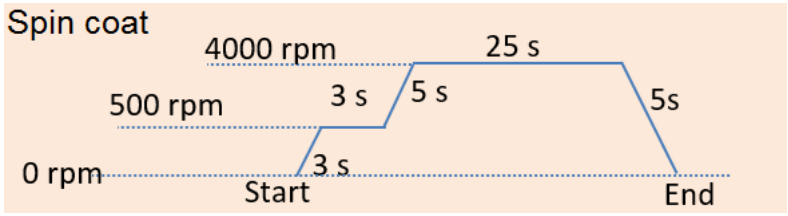
The glass substrate was dipped in acetone for 5 minutes, ethanol for 2 minutes, ultrasonic cleaning with DI water for 1 minute for cleansing. It was N_2 dry followed by O_2 ashing at 15 W, 180 sec, 30 sccm.

Step-2: Fabrication of thin Cr film

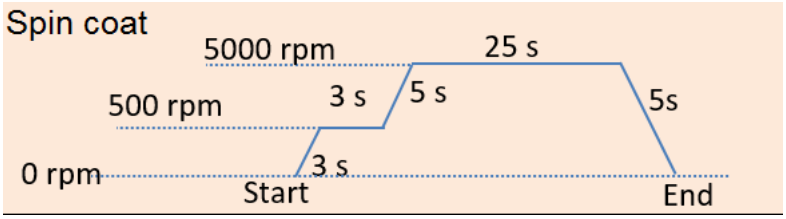
Cr film of 700-800 nm was coated by sputtering on to the substrate.

Step-3: Photolithography

OAP (adhesive) coating was performed using the following parameters.



OFPR-800 (photoresist) coating was performed following the below parameters.



Spin coating was followed by baking on hot plate at 90 ℃ for 90 secs followed by exposure of 5.3 sec in photomask aligner (MEP-800, Union). This was followed by development with NMD-w for 90 secs and rinsed with DI water.

Once again O_2 ashing at 15 W, 180 secs, 30 sccm and baking on hot plate at 120 ℃ for 5 minutes was performed followed by cool to room temperature (RT).

Step-4: Cr and SiO_2 wet etching

Cr wet etching was done for 10 min (KANTO CHEMICAL CO., INC. 混酸クロムエッチング液) whereas SiO_2 wet etching was done for 30 min (BHF) [front side] followed by rinsing with water.

Step-5: Reactive Ion Etching (RIE)

Deep RIE (200-iPB, SAMCO) was done for 6 cycles with 900 loop each time meaning 5400 loops in total to get the through-holes into the substrate.

Step-6: Final step

Finally, SiO_2 wet etching was done for 30 min (BHF) [back side] followed by cleansing the chip as described in step-1.

The sequential steps are given in Figure 41, below.

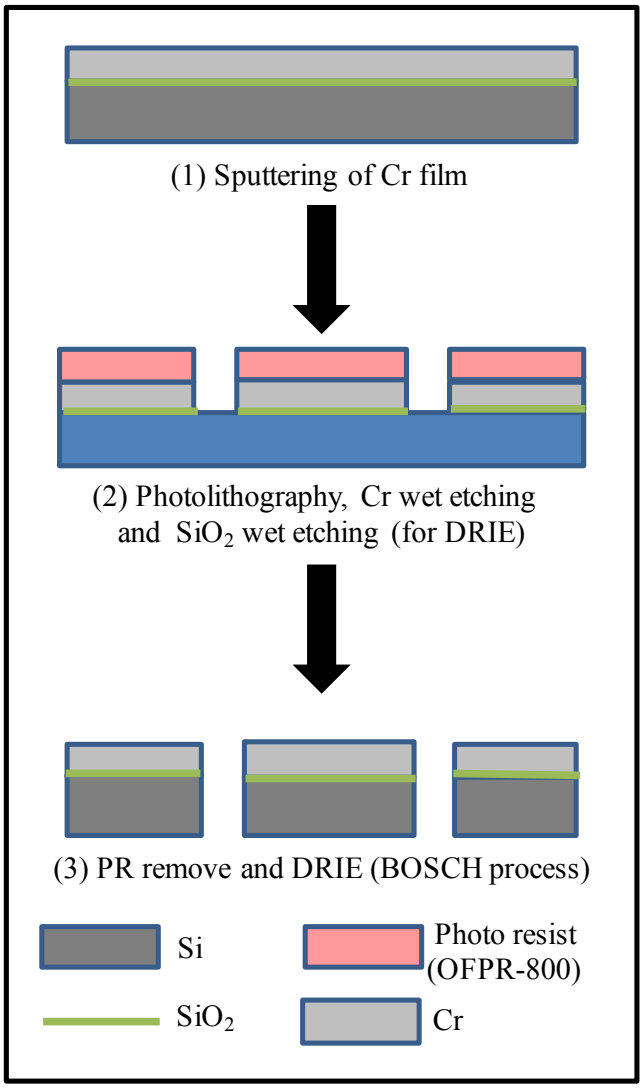


Figure 41: The sequential steps involved in fabrication of micro-hole array.

4.2.2. Micro-hole array electrospray setup and procedure

The water-in-oil droplets generation system was created using a micro-hole chip of 2.5 cm x 2.5 cm containing array of 7×7 mm (24 x 16 = 384 holes). A chamber was formed on chip using silicone rubber of 1.5 cm x 1.5cm x 2mm with square shaped cavity of 1 cm x 1 cm to act as aqueous phase. A 3 cm x 3 cm x 3 mm silicone rubber with square shaped cavity of 1.5 cm x 1.5 cm was placed on glass slide (35 x 55 mm) to act as oil chamber. ~ 700 µl oil mixture was poured in oil chamber electrified with a tungsten wire (act as earthing wire) followed by ~ 160 µl of aqueous solution in chip chamber electrified with a tungsten wire (to provide high voltage). The oil mixture used in these experiments were same as that of single nozzle electrospray. A voltage of 1000 V and frequency of 100 Hz was applied through the micro-hole array causing the aqueous solution to jet into the cavity of oil chamber thus producing a large number of water-in-oil droplets through the holes (about 5 minutes). The experimental step-up is shown in Figure 42.

To generate water-in-oil droplets only two parameters are required; first, voltage and the second, frequency unit. Therefore, to compress the bulky power source a portable compact power supply source was designed.

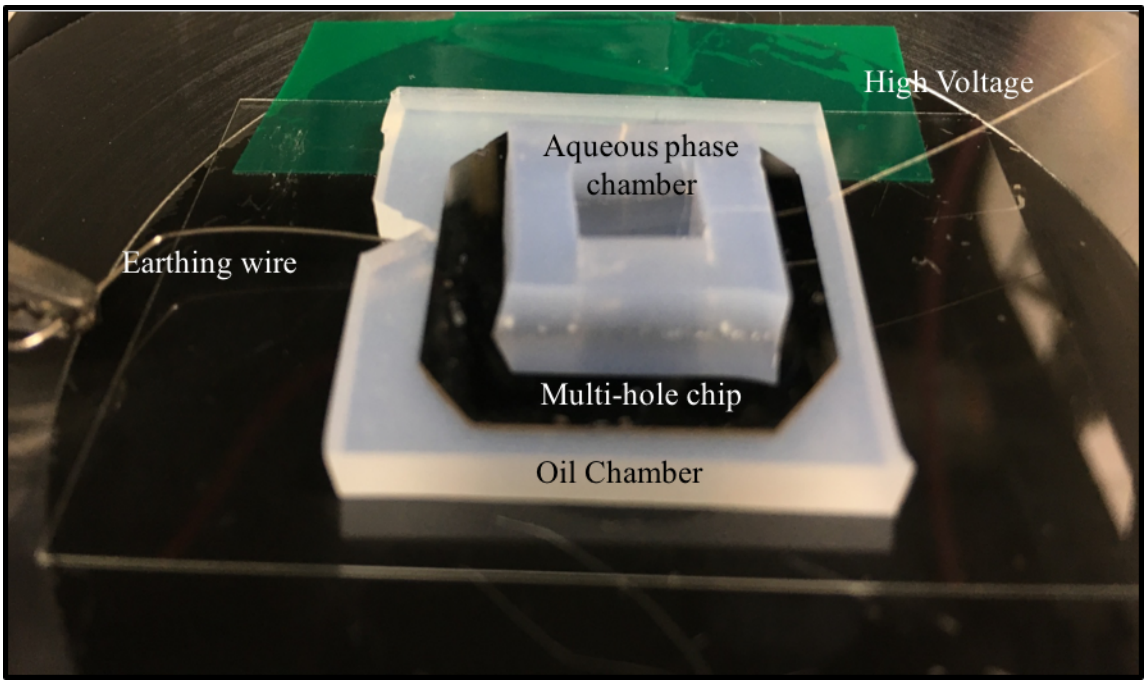


Figure 42: Experimental photograph of micro-hole array electrospray (µHAES) set-up to produce water-in-oil droplets.

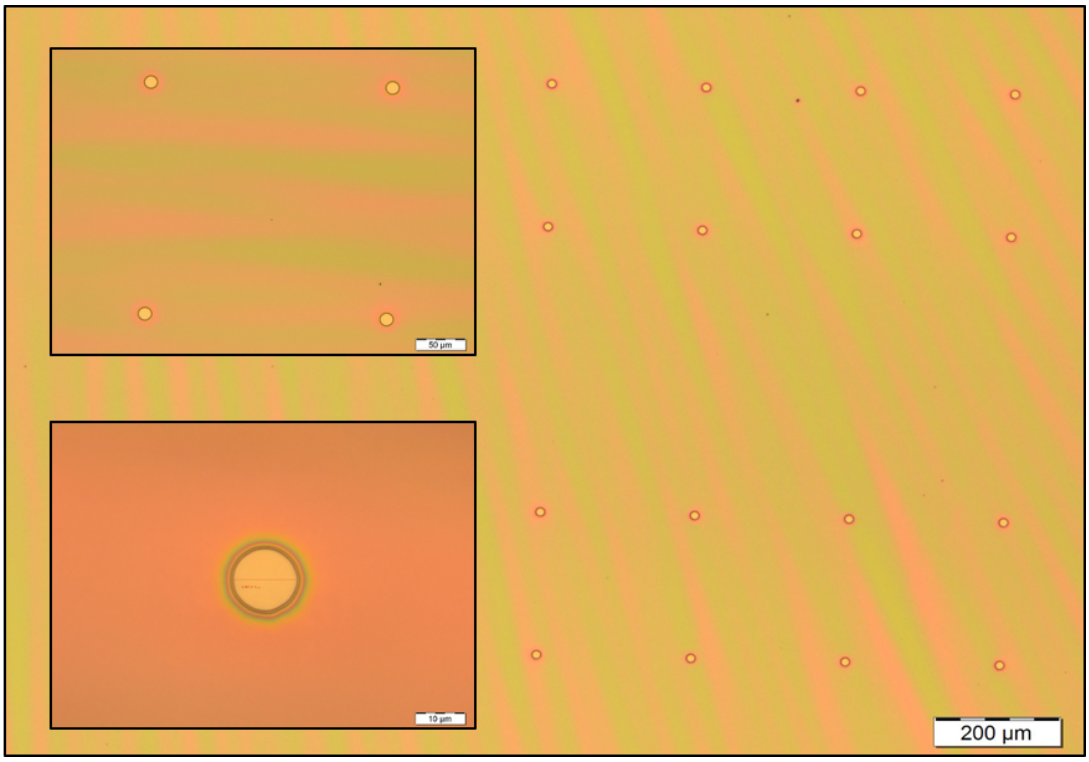
4.2.3. Agarose-in-oil droplet generation

Agarose-in-oil droplet were generated in oil mixture as described in the above section. 1% B-4-F (Biotin-4-fluorescein) in 1% agarose solution was used for fluorescence from the droplets. The produced water-in-oil droplets were observed under confocal laser scanning microscope (CLSM) to capture brightfield, fluorescence and overlap images of the droplets (Olympus FluoView 1000 spectral-based). A laser light with an Ar 488-nm wavelength was used to excite the fluorescein, which was observed through an Alexa Fluor 488 green dye filter. All images were captured using a 60x lens that was focused on the equatorial section same as single nozzle experiments. Droplet average size was measured using CLSM with the help of ImageJ software.

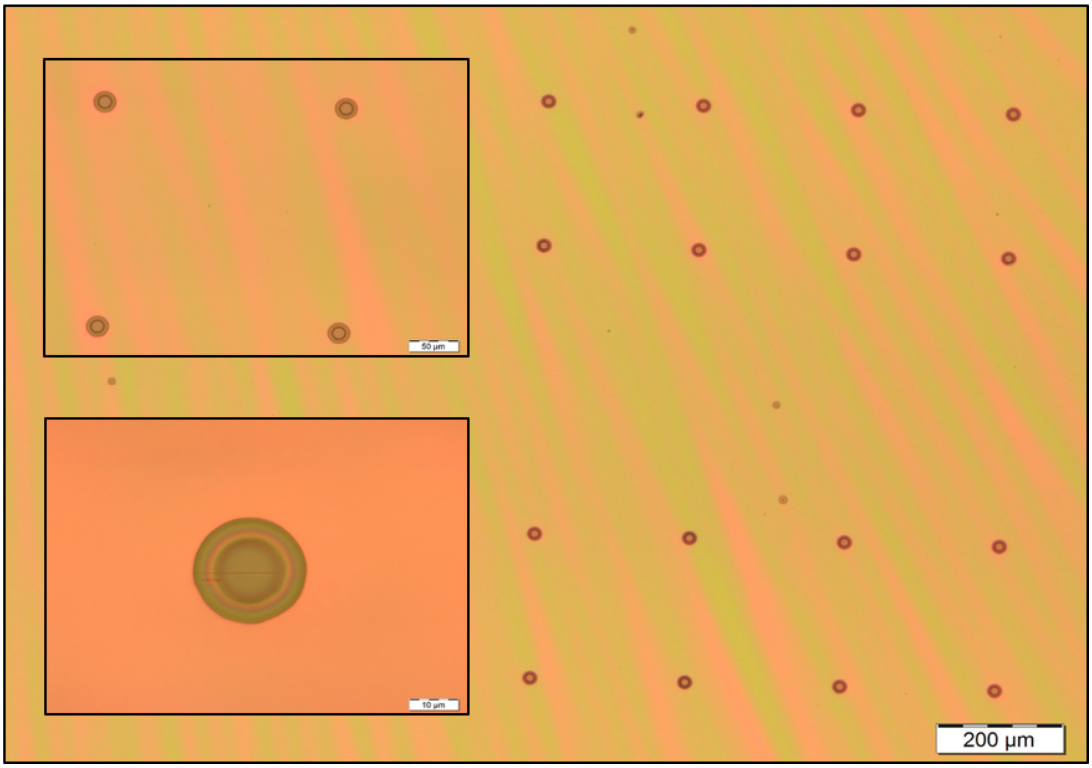
4.3. Results and Discussion

4.3.1. Fabrication of micro-hole array chip

(A)



(B)



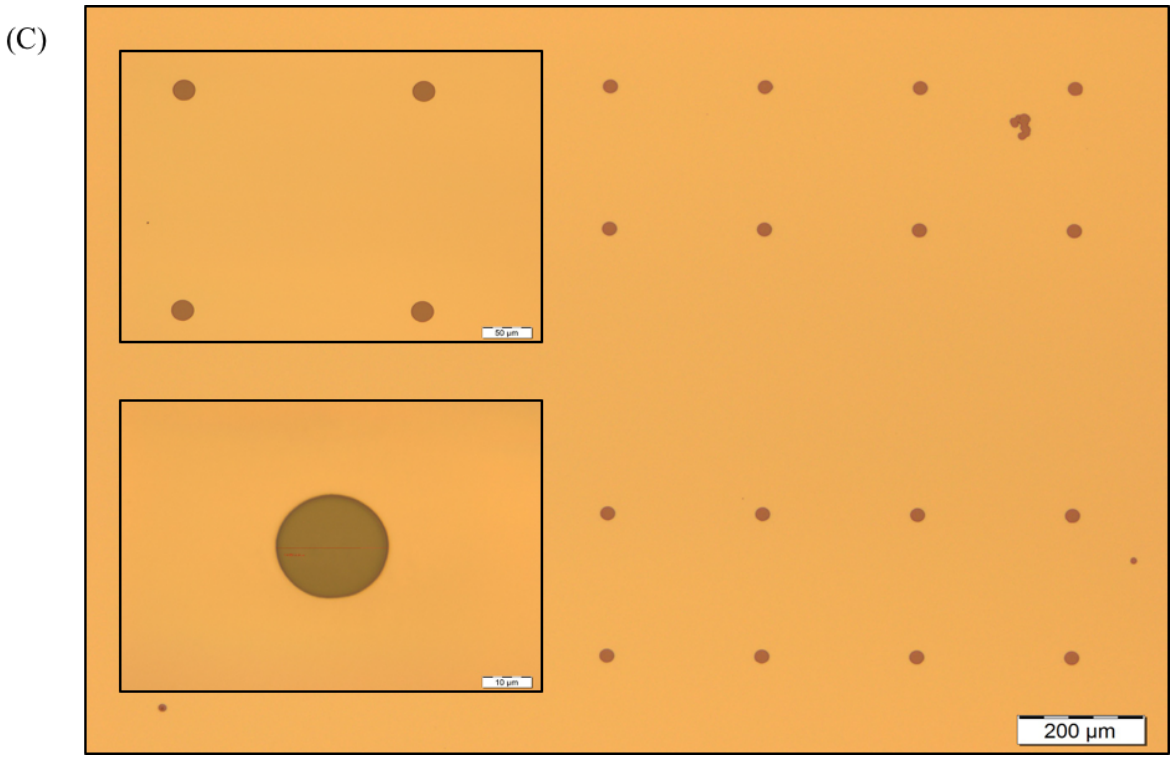


Figure 43: The microscopic images of SiO_2 substrate (A) Before Cr etching (B) After Cr etching and (C) After PR removal.

The micro-hole chip was observed under microscope and the size of the holes were measured. Before Cr etching the size of holes were ~12 µm, after Cr etching the size of holes were ~20-22 µm and finally after PR removal the size of holes measured as ~22-24 µm (see Figure 43).

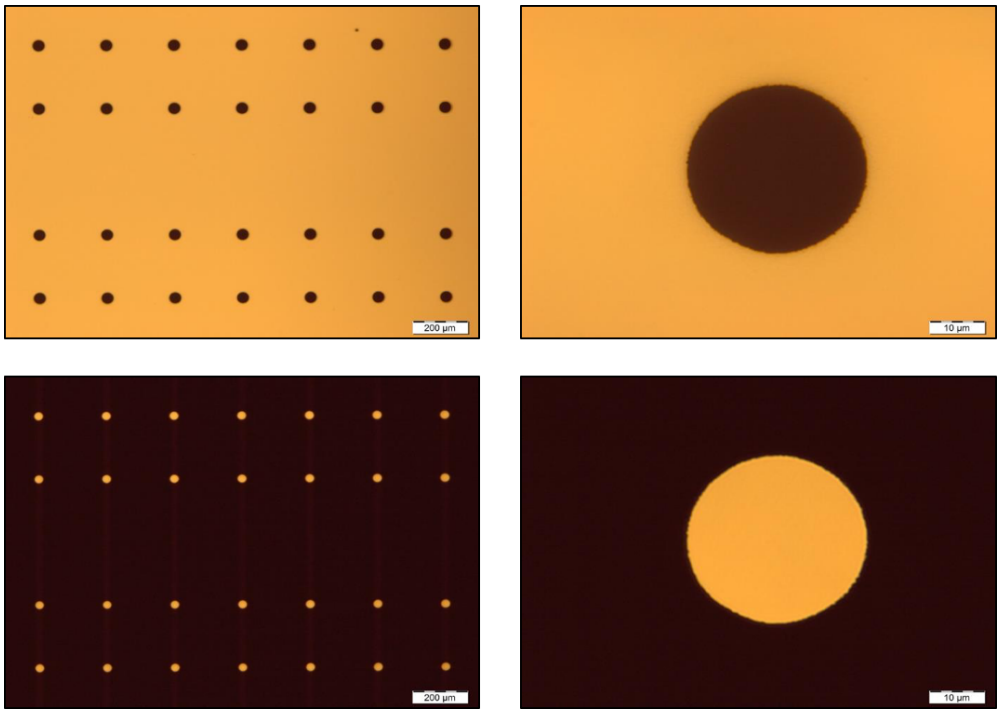


Figure 44: The microscopic image of micro-hole chip after BHF etching.

In Figure 44, after final BHF etching from RIE through-holes were obtained and the size of the through-holes were measured as ~24 µm.

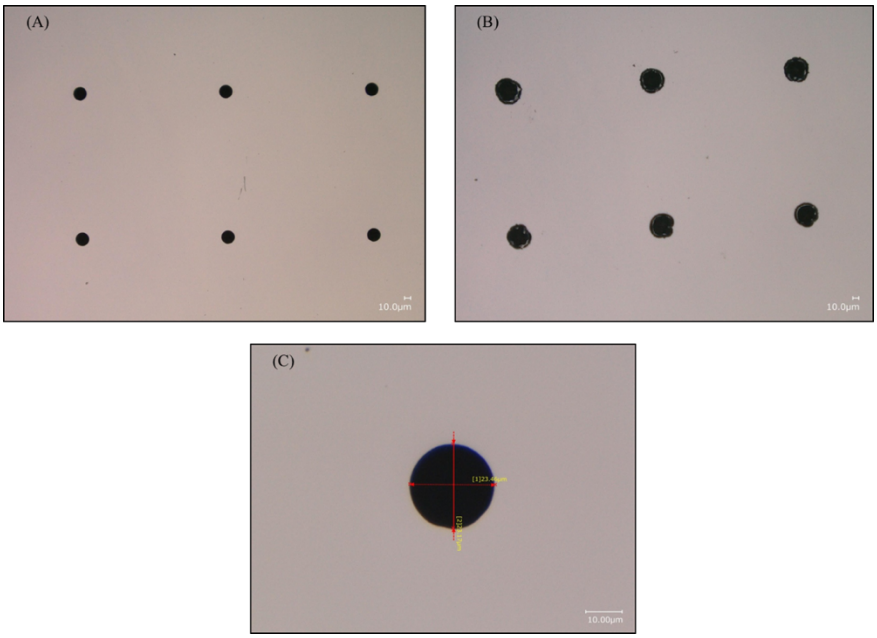


Figure 45: Microscopic images of micro-hole chip. (A) Chromium film side (B) SiO_2 side *and (C) Single hole with the dimensions are shown.*

Figure 45 are the microscopic images obtained and the size of the through-hole were measured as 23.4 µm (Figure 45C).

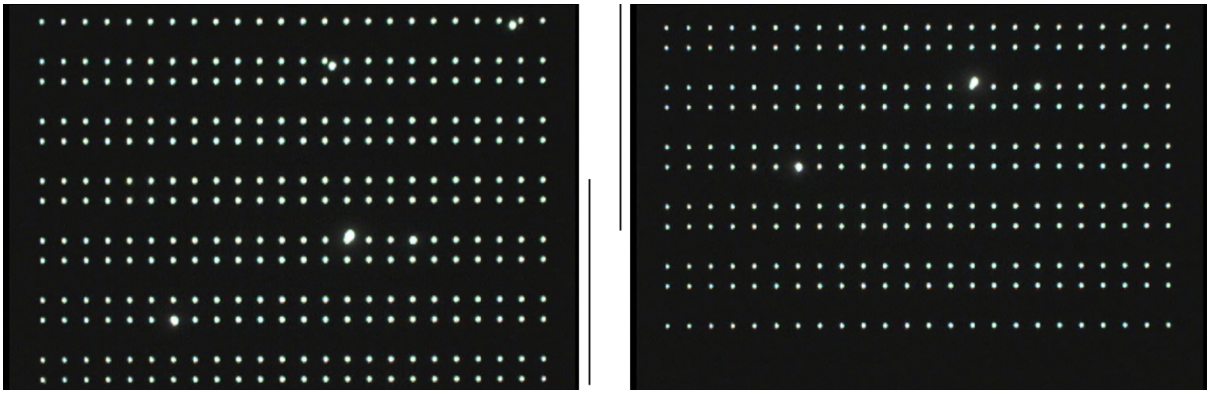


Figure 46: The microscopic image of Multi-hole chip at 2X (24x16 = 384 holes) to visualize total holes. Black side line indicates repeating rows.

At last, before using the multi-hole chip for the generation of water-in-oil droplets, the chip was observed under the microscope to completely see the through-holes. In total, 16 rows were observed, each containing 24 holes resulting in total 384 holes (24x16 = 384 holes) which can be easily calculated/estimated from Figure 46. Using 384 holes, library size can be easily increased to $10^{11\text{-}12}$ in the current electrospray system. The library size can further be increased by increasing the number of the holes and electrospray time.

4.3.2. Portable power supply device

A small 105 mm x 75 mm x 40 mm power box was designed to regulate voltage and frequency. It mainly consists of DC convertor to produce 1000 V, microcomputer to generate time and two switch boxes to switch between time and voltage. It is capable of supplying voltage of 1000 V and frequency of 100 Hz which is required to generate droplets. This small power supply device is so small and light weight that it can be easily carried from one place to other, thereby being portable enough to carry and generate water-in-oil droplets in molecular biology lab to handle the cell-free reaction in extremely miniaturized scale. The photograph of the device is shown in Figure 47.

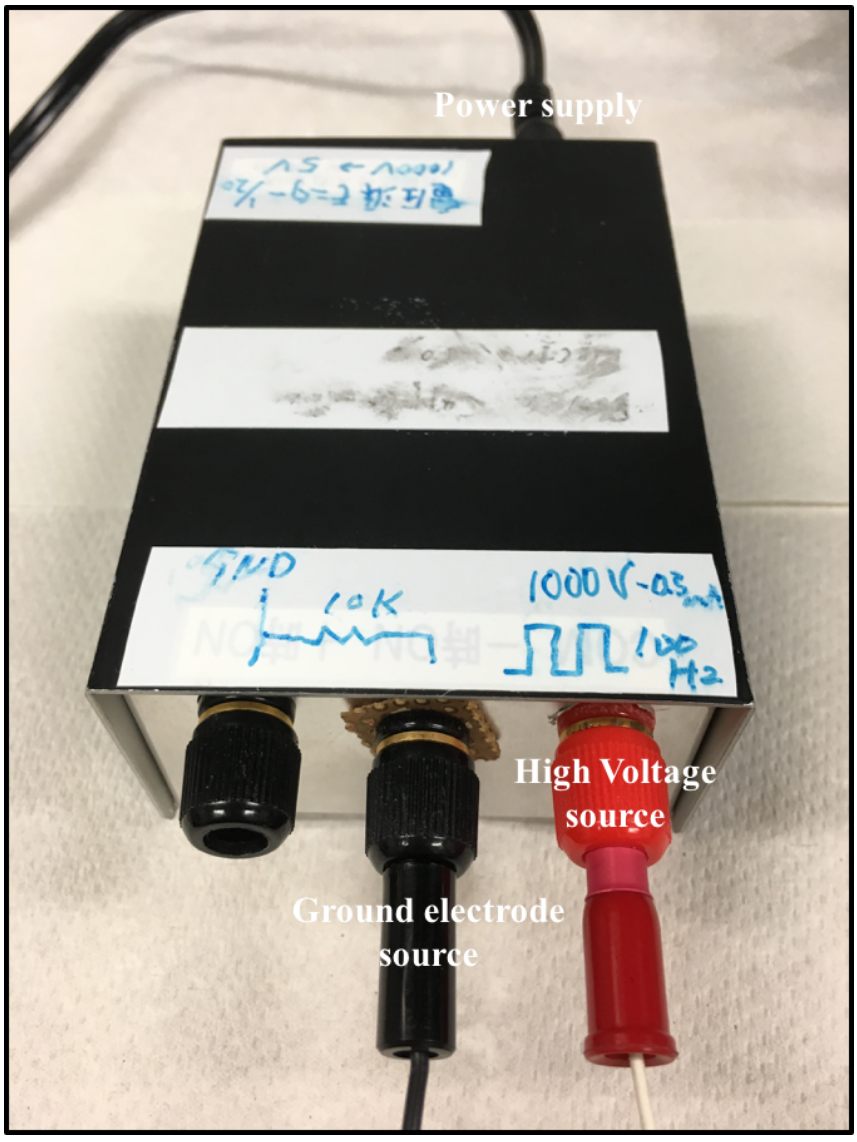


Figure 47: Experimental photograph of our portable-compact power supply device for micro-hole array electrospray (μHAES) set-up

4.3.3. Agarose-in-oil droplet generation using micro-hole array electrospray (μHAES)

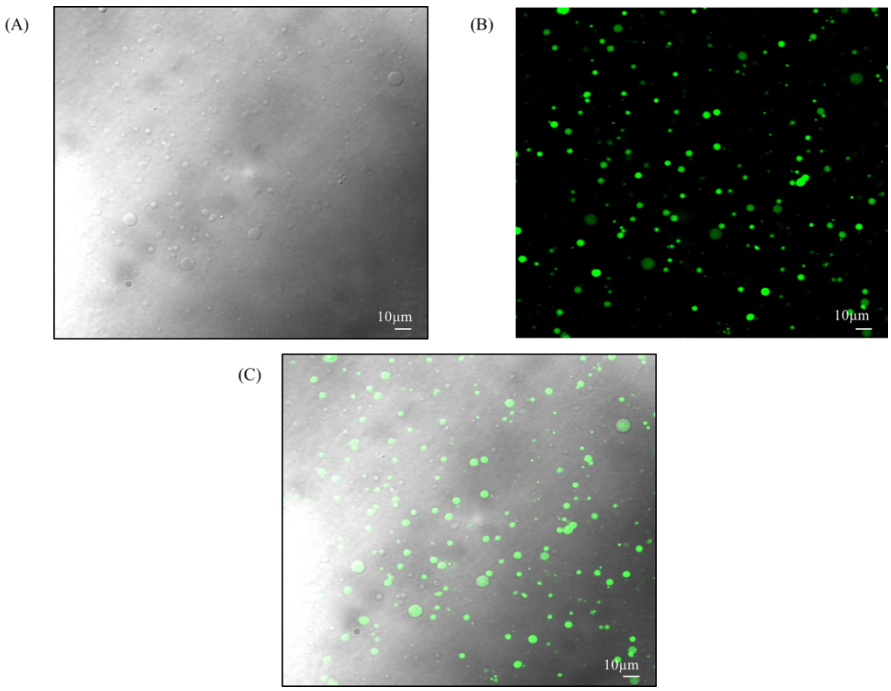


Figure 48: (A) Brightfield image (B) Fluorescence image and (C) Overlap image of agarose-in-oil droplets generated in oil mixture using micro-hole array in one plan under CLSM.

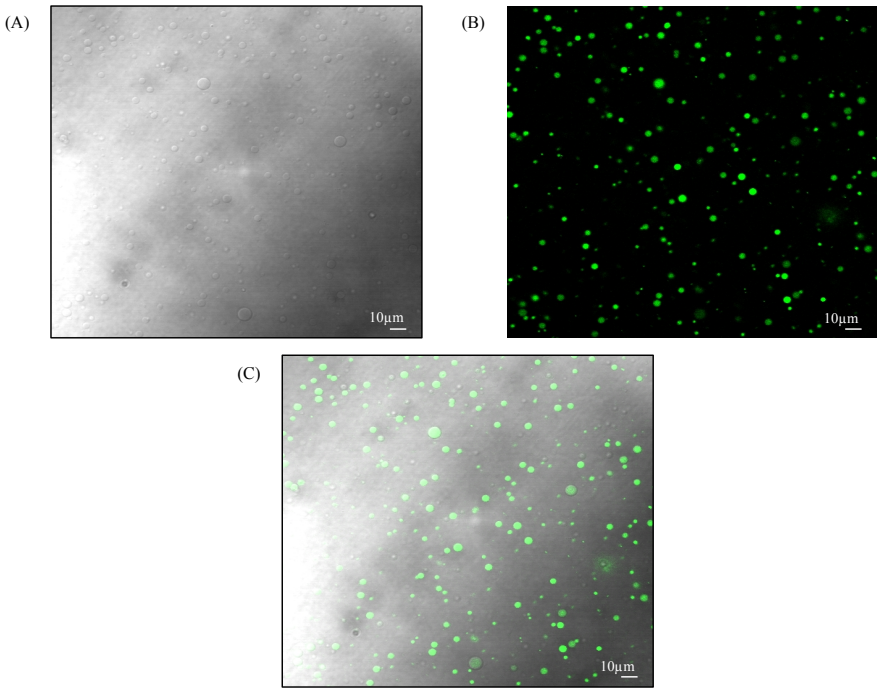


Figure 49: (A) Brightfield image (B) Fluorescence image and (C) Overlap image of agarose-in-oil droplets generated in oil mixture using micro-hole array in another plain under CLSM.

Ultrahigh-throughput water-in-oil droplets were observed under CLSM. The green colour was observed from the droplets due to the encapsulation of Biotin-4-fluorescein in agarose solution (Figure 48 & 49).

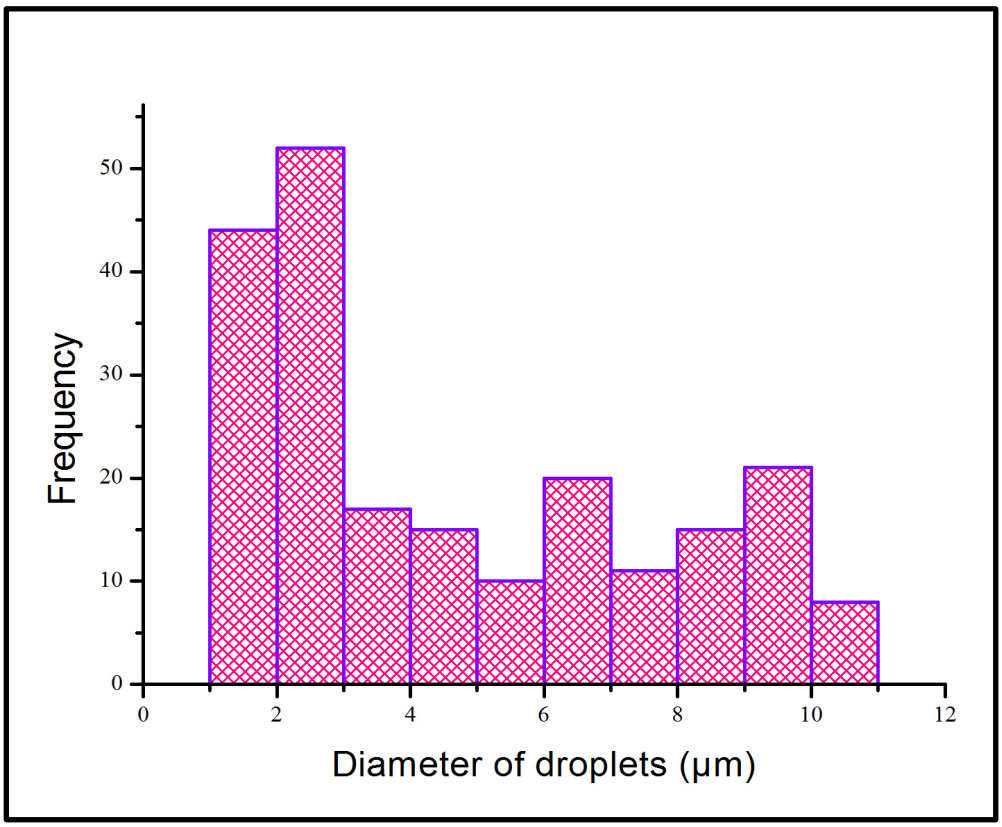


Figure 50: Graphical representation of size distribution of water-in-oil droplets by micro-hole array electrospray (µHAES).

From the graph (Figure 50) calculated by ImageJ using CLSM image, average size was found to be 1-3 µm with a range of 1-11 µm indicating the generation of ultra-low volume and monodisperse agarose-in-oil beads.

4.4. Conclusion

With this new μHAES platform, I can not only reduce the time of encapsulating cell-free reaction (10 μl) in 1fL IVC to just 5 minutes but can also increase library size to $10^{11\text{-}12}$ compared to 55 hours and $10^{8\text{-}9}$ library size of single nozzle electrospray therefore, establishing an ultrahigh-through w/o droplet generation platform in less time with improved library size for synthetic biology. It can be used for the purpose of large scale production of functional polymer bead with diameters in the sub-micrometer range, such as Ni-NTA agarose beads which are currently limited to produce in several tens of micrometer range, application. Also, a compact portable power supply source was designed.

From the study in this chapter, it can be concluded that our micro-hole array electrospray (μHAES) system can be used for ultra-rapid generation of water-in-oil or agarose-in-oil droplets to encapsulate "millions-to-quadrillions" of libraries to reach the superior library size demand in *in vitro* protein selection. We are planning to produce ultra-low volume and monodisperse beads to perform evolutionary molecular engineering experiments.

CHAPTER V

CONCLUSION

CHAPTER V

Conclusion

5.1. Conclusion

In this work, 'Head-to-Head' mRNA/cDNA display and electrospray microarray based *in vitro* compartmentalized system has been developed for simplifying the molecular evolutionary engineering. I have worked to develop two platforms simultaneously, one for function-based *in vitro* selection with improved Genotype-Phenotype (G-P) yield and the other for IVC based selection for larger library size.

In order to design a construct for functional based *in vitro* selection unlike that of previous affinity based selection, I introduced 'Head-to-Head' construct for covalent genotype-phenotype linkage. To protect the mRNA from degradation two main parameters were studied. H2H allows multiple protein formation from a single gene hence, it is an information-to-function based biomolecular display method for simplified and rapid *in vitro* molecular evolution. Using H2H construct the G-P efficiency was increased to ~ 20%.

Further, I have worked to simplify the creation of 'highly-selective lead molecules' as the next step of evolutionary molecular engineering for on-demand *in vitro* selection of highly-functional biomolecules by performing H2H covalent linkage in electrospray sub-femtoliter droplet compartments. I was successful in performing model experiments of H2H in water-in-oil droplets.

Also, to encapsulate large library size (millions-to-quadrillions) all together in short time period, μHAES (Micro-hole array electrospray) platform was created and successful generation of agarose-in-oil droplets were performed using micro-hole chip. To the best of my knowledge it is for the first time to use 384 holes to produce uniform droplets. This system is capable of encapsulating 10 μl cell-free reaction in 1fL IVC in just 5 minutes with library size of $10^{11\text{-}12}$ as compared to 55 hours and $10^{8\text{-}9}$ library size of previous single nozzle electrospray.

5.2. Summary

In summary, I have developed a simplified, easy and rapid display technology for function-based selection with improved G-P yield and IVC-based selection with larger library size. I started with both mRNA and DNA as template for improved protocol and high yield of linkage product. The H2H design is also applicable for longer transcript as shown by the complete functional expression of AGT and GFP protein. H2H should enable simple and faster *in vitro* directed evolution. I also tried to integrate novel H2H covalent genotype-phenotype linkage in femtoliter in water-in-droplets generated by electrospray technology for one-step *in vitro* protein evolution. Later, μHAES platform to generate ultralow volume droplets to compartmentalize larger library size in less time was established. Therefore, I would like to conclude that the novel H2H construct and electrospray microarray based IVC would be a key step to simplify molecular evolutionary engineering.

5.3. Future Prospective

This H2H construct can be used to find allosteric binding site based peptide aptamer to work as function-based selection. The integration of H2H in water-in-oil droplets by electrospray will enable to produce H2H agarose beads for selection on demand. μHAES can be used to produce ultralow Ni-NTA beads and commercialized in the near future, application.

REFERENCES

1. Hall, B.K. & Hallgrimsson, B. Strickberger's Evolution (4th ed.). *Sudbury, Massachusetts: Jones and Bralett Publishers*, 4-6 (2008).
2. Evolution Resources. *National Academies of Sciences, Engineering and Medicine* (2016).
3. Panno, J. The Cell: Evolution of the First Organism. *Facts on File science library, New York*, xv-16 (2005).
4. Lutz, S. Beyond directed evolution - semi-rational protein engineering and design. *Current opinion in biotechnology* **21**, 734-743 (2010).
5. Packer, M.S. & Liu, D.R. Methods for the directed evolution of proteins. *Nature Reviews Genetics* **16**, 379 (2015).
6. Karatan, E., Han, Z. & Kay, B. Molecular Display Technologies. *Encyclopedia of Molecular Cell Biology and Molecular Medicine* **8**, 409-430 (2005).
7. FitzGerald, K. In vitro display technologies - new tools for drug discovery. *Drug discovery today* **5**, 253-258 (2000).
8. Leemhuis, H., Stein, V., Griffiths, A.D. & Hollfelder, F. New genotype–phenotype linkages for directed evolution of functional proteins. *Current Opinion in Structural Biology* **15**, 472-478 (2005).
9. Smith, G.P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315 (1985).
10. Lowman, H.B. & Wells, J.A. Affinity Maturation of Human Growth Hormone by Monovalent Phage Display. *Journal of Molecular Biology* **234**, 564-578 (1993).
11. Lowman, H.B., Bass, S.H., Simpson, N. & Wells, J.A. Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry* **30**, 10832-10838 (1991).
12. Blaise, L. et al. Construction and diversification of yeast cell surface displayed libraries by yeast mating: application to the affinity maturation of Fab antibody fragments. *Gene* **342**, 211-218 (2004).
13. Aharoni, A., Griffiths, A.D. & Tawfik, D.S. High-throughput screens and selections of enzyme-encoding genes. *Current Opinion in Chemical Biology* **9**, 210-216 (2005).
14. Harris, L.F., Sullivan, M.R. & Hatfield, D.L. Directed Molecular Evolution. *Origins of life and evolution of the biosphere* **29**, 425-435 (1999).

15. Mattheakis, L.C., Bhatt, R.R. & Dower, W.J. 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**, 9022-9026 (1994).
16. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H.R. & Plückthun, A. Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries. *Proceedings of the National Academy of Sciences* **95**, 14130 (1998).
17. Hanes, J., Schaffitzel, C., Knappik, A. & Plückthun, A. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nature Biotechnology* **18**, 1287 (2000).
18. Zahnd, C. et al. Directed in vitro evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity. *J Biol Chem* **279**, 18870-18877 (2004).
19. Yau, K.Y.F. et al. Affinity maturation of a VHH by mutational hotspot randomization. *Journal of Immunological Methods* **297**, 213-224 (2005).
20. Binz, H.K. et al. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nature Biotechnology* **22**, 575 (2004).
21. Sawata, S.Y. & Taira, K. Modified peptide selection in vitro by introduction of a protein–RNA interaction. *Protein Engineering, Design and Selection* **16**, 1115-1124 (2003).
22. Lamla, T. & Erdmann, V.A. Searching Sequence Space for High-affinity Binding Peptides using Ribosome Display. *Journal of Molecular Biology* **329**, 381-388 (2003).
23. Matsuura, T. & Plückthun, A. Selection based on the folding properties of proteins with ribosome display. *FEBS Letters* **539**, 24-28 (2003).
24. Matsuura, T. & Pluckthun, A. Strategies for selection from protein libraries composed of de novo designed secondary structure modules. *Orig Life Evol Biosph* **34**, 151-157 (2004).
25. Roberts, R.W. & Szostak, J.W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences* **94**, 12297 (1997).
26. Monro, R.E. & Vazquez, D. Ribosome-catalysed peptidyl transfer: Effects of some inhibitors of protein synthesis. *Journal of Molecular Biology* **28**, 161-165 (1967).

27. Monro, R.E. & Marcker, K.A. Ribosome-catalysed reaction of puromycin with a formylmethionine-containing oligonucleotide. *Journal of Molecular Biology* **25**, 347-350 (1967).
28. 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 Research* **28**, e83-e83 (2000).
29. Kurz, M., Gu, K., Al-Gawari, A. & Lohse, P.A. cDNA-protein fusions: covalent protein-gene conjugates for the in vitro selection of peptides and proteins. *ChemBioChem* **2**, 666-672 (2001).
30. Liu, R., Barrick, J.E., Szostak, J.W. & Roberts, R.W. Optimized synthesis of RNA-protein fusions for in vitro protein selection. *Methods in enzymology* **318**, 268-293 (2000).
31. Xu, L. et al. Directed Evolution of High-Affinity Antibody Mimics Using mRNA Display. *Chemistry & Biology* **9**, 933-942 (2002).
32. Wilson, D.S., Keefe, A.D. & Szostak, J.W. The use of mRNA display to select high-affinity protein-binding peptides. *Proceedings of the National Academy of Sciences* **98**, 3750 (2001).
33. Raffler, N.A., Schneider-Mergener, J. & Famulok, M. A Novel Class of Small Functional Peptides that Bind and Inhibit Human α-Thrombin Isolated by mRNA Display. *Chemistry & Biology* **10**, 69-79 (2003).
34. Ja, W.W. & Roberts, R.W. In Vitro Selection of State-Specific Peptide Modulators of G Protein Signaling Using mRNA Display. *Biochemistry* **43**, 9265-9275 (2004).
35. Cujec, T.P., Medeiros, P.F., Hammond, P., Rise, C. & Kreider, B.L. Selection of v-Abl Tyrosine Kinase Substrate Sequences from Randomized Peptide and Cellular Proteomic Libraries Using mRNA Display. *Chemistry & Biology* **9**, 253-264 (2002).
36. Keefe, A.D. & Szostak, J.W. Functional proteins from a random-sequence library. *Nature* **410**, 715 (2001).
37. Kelly, B.T., Baret, J.-C., Taly, V. & Griffiths, A.D. Miniaturizing chemistry and biology in microdroplets. *Chemical Communications*, 1773-1788 (2007).
38. Souza, T.P.D., Stano, P. & Luisi, P.L. The minimal size of liposome-based model cells brings about a remarkably enhanced entrapment and protein synthesis. *ChemBioChem* **10**, 1056-1063 (2009).

39. Luisi, P.L. & Stano, P. Synthetic biology minimal cell mimicry. *Nat Chem* **3**, 755-756 (2011).

40. Lu, W.-C. & Ellington, A.D. In vitro selection of proteins via emulsion compartments. *Methods* **60**, 75-80 (2013).

41. Okano, T., Matsuura, T., Suzuki, H. & Yomo, T. Cell-free Protein Synthesis in a Microchamber Revealed the Presence of an Optimum Compartment Volume for High-order Reactions. *ACS Synthetic Biology* **3**, 347-352 (2014).

42. Tawfik, D.S. & Griffiths, A.D. Man-made cell-like compartments for molecular evolution. *Nature Biotechnology* **16**, 652 (1998).

43. Griffiths, A.D. & Tawfik, D.S. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *The EMBO Journal* **22**, 24-35 (2003).

44. Griffiths, A.D. & Tawfik, D.S. Miniaturising the laboratory in emulsion droplets. *Trends in Biotechnology* **24**, 395-402 (2006).

45. Ró aDska, S., Ró aDski, J., Ochowiak, M. & Mitkowski, P.T. Extensional viscosity measurements of concentrated emulsions with the use of the opposed nozzles device. *Brazilian Journal of Chemical Engineering* **31**, 47-55 (2014).

46. Katepalli, H. & Bose, A. Response of Surfactant Stabilized Oil-in-Water Emulsions to the Addition of Particles in an Aqueous Suspension. *Langmuir* **30**, 12736-12742 (2014).

47. Miller, O.J. et al. Directed evolution by in vitro compartmentalization. *Nature Methods* **3**, 561 (2006).

48. Abate, A.R., Thiele, J. & Weitz, D.A. One-step formation of multiple emulsions in microfluidics. *Lab on a Chip* **11**, 253-258 (2011).

49. Umbanhowar, P.B., Prasad, V. & Weitz, D.A. Monodisperse Emulsion Generation via Drop Break Off in a Coflowing Stream. *Langmuir* **16**, 347-351 (2000).

50. Guo, M.T., Rotem, A., Heyman, J.A. & Weitz, D.A. Droplet microfluidics for high-throughput biological assays. *Lab on a Chip* **12**, 2146-2155 (2012).

51. Ma, F., Xie, Y., Huang, C., Feng, Y. & Yang, G. An Improved Single Cell Ultrahigh Throughput Screening Method Based on In Vitro Compartmentalization. *PLOS ONE* **9**, e89785 (2014).

52. Yan, J. et al. Monodisperse Water-in-Oil-in-Water (W/O/W) Double Emulsion Droplets as Uniform Compartments for High-Throughput Analysis via Flow Cytometry. *Micromachines* **4** (2013).
53. Kaltenbach, M. & Hollfelder, F. in Ribosome Display and Related Technologies: Methods and Protocols. (eds. J.A. Douthwaite & R.H. Jackson) 101-111 (Springer New York, New York, NY; 2012).
54. Gillian, H., David, L. & Florian, H. SNAP Display - an In Vitro Method for the Selection of Protein Binders. *Current Pharmaceutical Design* **19**, 5421-5428 (2013).
55. Sharma, B., Takamura, Y., Shimoda, T. & Biyani, M. A bulk sub-femtoliter in vitro compartmentalization system using super-fine electrosprays. *Scientific Reports* **6**, 26257 (2016).
56. Ma, H. et al. Nucleic acid aptamers in cancer research, diagnosis and therapy. *Chemical Society Reviews* **44**, 1240-1256 (2015).
57. Li, J., Tan, S., Chen, X., Zhang, C.Y. & Zhang, Y. Peptide Aptamers with Biological and Therapeutic Applications. *Current Medicinal Chemistry* **18**, 4215-4222 (2011).
58. Zhou, J. & Rossi, J. Aptamers as targeted therapeutics: current potential and challenges. *Nature Reviews Drug Discovery* **16**, 181 (2016).
59. Röthlisberger, P. & Hollenstein, M. Aptamer chemistry. *Advanced Drug Delivery Reviews* (2018).
60. Colas, P. et al. Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* **380**, 548 (1996).
61. Reverdatto, S., Burz, D.S. & Shekhtman, A. Peptide Aptamers: Development and Applications. *Current topics in medicinal chemistry* **15**, 1082-1101 (2015).
62. Baines, I.C. & Colas, P. Peptide aptamers as guides for small-molecule drug discovery. *Drug Discovery Today* **11**, 334-341 (2006).
63. Röthlisberger, P., Gasse, C. & Hollenstein, M. Nucleic Acid Aptamers: Emerging Applications in Medical Imaging, Nanotechnology, Neurosciences, and Drug Delivery. *International Journal of Molecular Sciences* **18**, 2430 (2017).
64. Jijakli, K. et al. The in vitro selection world. *Methods* **106**, 3-13 (2016).
65. Speight, R.E., Hart, D.J., Sutherland, J.D. & Blackburn, J.M. A new plasmid display technology for the in vitro selection of functional phenotype–genotype linked proteins. *Chemistry & Biology* **8**, 951-965 (2001).

66. Litovchick, A. & Szostak, J.W. Selection of cyclic peptide aptamers to HCV IRES RNA using mRNA display. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 15293-15298 (2008).

67. Ueno, S., Arai, H., Suzuki, M. & Husimi, Y. An mRNA-protein Fusion at N-terminus for Evolutionary Protein Engineering. *International Journal of Biological Sciences* **3**, 365-374 (2007).

68. Marino, G., Eckhard, U. & Overall, C.M. Protein Termini and Their Modifications Revealed by Positional Proteomics. *ACS Chemical Biology* **10**, 1754-1764 (2015).

69. Lipovsek, D. & Plückthun, A. In-vitro protein evolution by ribosome display and mRNA display. *Journal of Immunological Methods* **290**, 51-67 (2004).

70. Mollwitz, B. et al. Directed Evolution of the Suicide Protein O6-Alkylguanine-DNA Alkyltransferase for Increased Reactivity Results in an Alkylated Protein with Exceptional Stability. *Biochemistry* **51**, 986-994 (2012).

71. Loktionova, N.A. & Pegg, A.E. Interaction of mammalian O(6)-alkylguanine-DNA alkyltransferases with O(6)-benzylguanine. *Biochemical pharmacology* **63**, 1431-1442 (2002).

72. Stein, V., Sielaff, I., Johnsson, K. & Hollfelder, F. A Covalent Chemical Genotype–Phenotype Linkage for in vitro Protein Evolution. *ChemBioChem* **8**, 2191 – 2194 (2007).

73. Woodson, S.A. & Koculi, E. in Methods in Enzymology, Vol. 469 189-208 (Academic Press, 2009).

74. Zhang, L. & Tripathi, A. Archaeal RNA ligase from thermoccocus kodakarensis for template dependent ligation. *International Journal of Biological Sciences* **14**, 365-374 (2017).

75. Nishigaki, K., Husimi, Y. & Tsubota, M. Detection of differences in higher order structure between highly homologous single-stranded DNAs by low-temperature denaturant gradient gel electrophoresis. *J Biochem.* **99**, 663-671 (1986).

76. Bernstein, J.A., Khodursky, A.B., Lin, P.-H., Lin-Chao, S. & Cohen, S.N. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences* **99**, 9697 (2002).

77. Shimizu, Y. et al. Cell-free translation reconstituted with purified components. *Nature Biotechnology* **19**, 751 (2001).

78. Li, J. et al. Dissecting limiting factors of the Protein synthesis Using Recombinant Elements (PURE) system. *Translation* **5**, e1327006 (2017).

79. Dressman, D., Yan, H., Traverso, G., Kinzler, K.W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences* **100**, 8817 (2003).

80. Kato, A., Yanagisawa, M., Sato, Y.T., Fujiwara, K. & Yoshikawa, K. Cell-Sized confinement in microspheres accelerates the reaction of gene expression. *Scientific Reports* **2**, 283 (2012).

81. Fiordemondo, D. & Stano, P. Lecithin-Based Water-In-Oil Compartments as Dividing Bioreactors. *ChemBioChem* **8**, 1965-1973 (2007).

82. Biyani, M., Husimi, Y. & Nemoto, N. Solid-phase translation and RNA–protein fusion: a novel approach for folding quality control and direct immobilization of proteins using anchored mRNA. *Nucleic Acids Research* **34**, e140-e140 (2006).

83. Quang Tran, S.B. et al. Polymer-based electrospray device with multiple nozzles to minimize end effect phenomenon. *Journal of Electrostatics* **68**, 138-144 (2010).

84. Su, S., Gibson, G.T.T., Mugo, S.M., Marecak, D.M. & Oleschuk, R.D. Microstructured Photonic Fibers as Multichannel Electrospray Emitters. *Analytical Chemistry* **81**, 7281-7287 (2009).

85. Parhizkar, M. et al. Performance of novel high throughput multi electrospray systems for forming of polymeric micro/nanoparticles. *Materials & Design* **126**, 73-84 (2017).

86. Fu, Y. et al. Polymer Micronozzle Array for Multiple Electrosprays Produced by Templated Synthesis and Etching of Microstructured Fibers. *Analytical Chemistry* **87**, 747-753 (2015).

87. Lhernould, M.S. & Lambert, P. Compact polymer multi-nozzles electrospray device with integrated microfluidic feeding system. *Journal of Electrostatics* **69**, 313-319 (2011).

88. Choi, K.-H. et al. Effects of nozzles array configuration on cross-talk in multi-nozzle electrohydrodynamic inkjet printing head. *Journal of Electrostatics* **69**, 380-387 (2011).