MACQUARIE
University

# Optimising Yeast Surface Display For Identifying Cellular Targets of Natural Products

Student's Name- Mrs Ruchi Dhaval Mehta

Department- Molecular Sciences

Submission Date- 09/10/2017

I, Mrs. Ruchi Dhaval Mehta agree that the work has not been submitted for a higher degree to any other university or institution. Moreover, the experimental procedures involved in this project is approved as Exempt Dealing by The Institutional Biosafety Committee (File Reference Number- 5201600950).

Sign: _____  _____ Date: 09/10/2017

# 1. CONTENTS

## 2. ABSTRACT

While natural products have long been a valuable source of biologically active compounds, their cellular targets and modes of action are rarely identified. In this project, yeast surface display (YSD) was investigated as a platform technology to rapidly identify the cellular targets of natural products. YSD involves cloning a cDNA library into baker's yeast (*Saccharomyces cerevisiae*) such that the encoded foreign proteins are expressed fused to the yeast Aga2 surface receptor. Flow cytometry (FACS) is then used to separate yeast cells displaying proteins capable of binding to a fluorescently tagged probe, thereby allowing the cellular targets of the probe to be identified. In Part 1 of this project, a YSD clone displaying the well-studied human protein FKBP was constructed as a positive control. A fluorescently labelled analogue of FK506 (a known inhibitor of FKBP) was then used to optimise a range of FACS parameters for protein-small molecule interactions, including probe concentration, incubation time/temperature, detergent concentration, washing stringency and FACS binning stringency. In Part 2, a YSD cDNA library was constructed from the model nematode *Caenorhabditis elegans*. These preliminary studies have laid the groundwork for future YSD studies to identify the cellular targets of a range of biologically active natural products.

# 3. INTRODUCTION

## 3.1 PARASITIC DISEASES

Parasitic infections are one of the major causes of morbidity and mortality globally. A recent report from the World Health Organisation (WHO) revealed that approximately one-third of the deaths worldwide are due to parasitic and infectious diseases. Eleven of seventeen infectious diseases appearing on the WHO's list of neglected tropical diseases (NTDs) are caused by parasites. These diseases are referred to as "neglected" because they have been largely eliminated from the more developed parts of the world and persist only in the poorest, most marginalized communities and under-developed areas. Most of these diseases are transmitted through insects or animals and are more sensitive to climate variables. The burden of such diseases is increased due to the need for chronic and expensive observation strategies. During the years 2012 and 2013, leading pharmaceutical companies donated treatments worth 2.5 billion dollars for the treatments of such parasites in endemic countries[1]. Hence, there is a necessity to find cost-effective and targeted drugs to combat and eliminate these parasitic infections.

## 3.2 NATURAL PRODUCTS AND EMERGING DRUG RESISTANCE

Nature is a treasure trove of drugs used to prevent and treat parasitic infections. For example, quinine was extracted from cinchona tree bark and has been used for many years (up to mid-20$^{th}$ century) for the treatment of malaria. Artemisinin, which is recommended by the WHO as a first line of treatment for malaria, was first isolated from the sweet wormwood plant and has been used by Chinese traditionally for thousands of years. Avermectin is another anti-parasitic drug that was isolated from soil actinomyces and is used to treat river-blindness and lymphatic filariasis, primarily caused by roundworm parasites. Dr. William Campbell and Dr. Satoshi Omura as well as Dr. Youyou Tu shared the 2015 Nobel prize in Medicine and Physiology for developing avermectin and artemisinin respectively[2]. Apart from malaria, many other parasitic diseases are the cause of deaths of more than 200,000 people annually. These parasitic diseases include dracunculiasis, lymphatic filariais, schistosomiasis, and soil-transmitted helminths (STH). Very little is known about the clinical pharmacology of such diseases. The advent of mass drug administration programmes for the control of such diseases in humans requires better understanding and monitoring of drug resistance, and also to broaden the currently very narrow range of available anthelmintics[3].

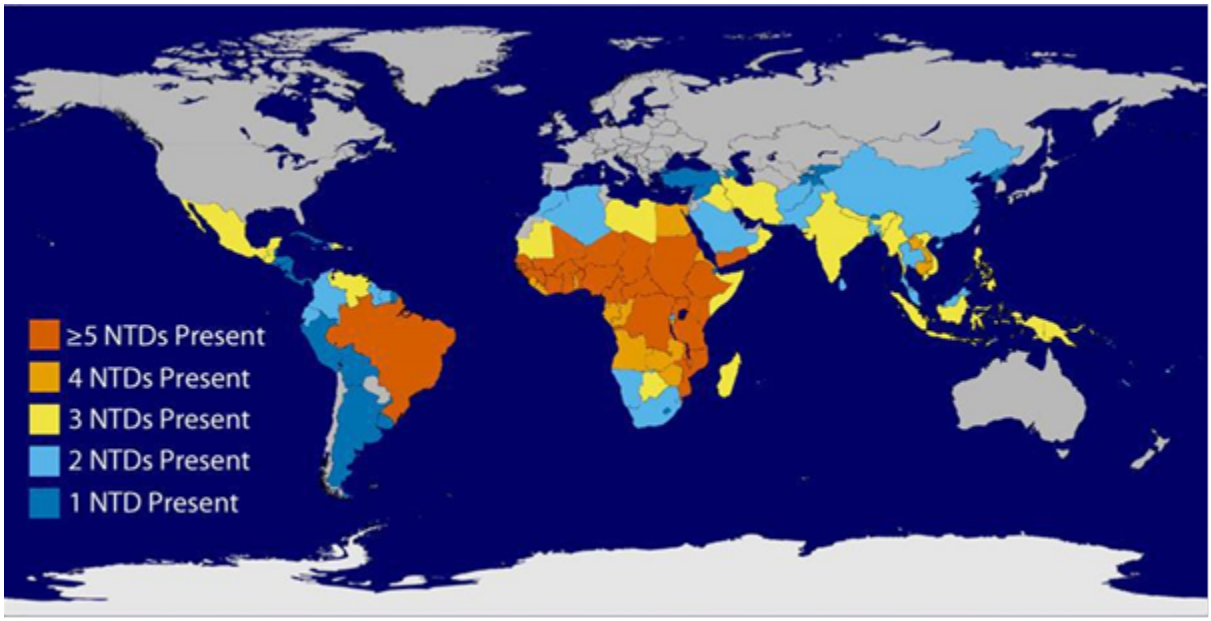



**Figure 1:** Global overlap of six of the common NTDs mostly caused by worms. These diseases are major constraints on the health and development of families living under the poverty line. (https://www.cdc.gov/globalhealth/ntd/diseases/index.html)

With the emergence of antibiotic drug resistance among pathogens, there is always a need to hunt for alternative treatment solutions. While diversity oriented synthesis can produce molecules that imitate natural products in terms of structural diversity and complexity, these molecules have not evolved to interact with biomolecules[4]. While Nature continues to provide potent and diverse anti-parasitic molecules, it is also equally important to discover their cellular targets and modes of action to design an appropriate pharmaceutical drug. Therefore, strategies need to be developed that can rapidly identify the cellular targets of natural products. This will make it possible to determine the biochemical pathway involved, thus facilitating the rational design and development of chemotherapeutic agents.

## 3.3 REVERSE CHEMICAL PROTEOMICS

Proteomics is a broad and systemic approach to understand functions, interactions, modifications and regulation of protein expression by a cell[5]. The conventional method of identifying the cellular targets of natural products is forward chemical proteomics. In this strategy, the natural product is radioactively or more recently fluorescently labelled and binding partners are identified on protein electrophoresis gels or by immobilisation of the natural product on a solid support to "pull down" binding partners from entire proteomes. This approach was first used to identify tubulin as a cellular target for the natural product colchicine, which results in the inhibition of mitosis in cultures of human cells[6]. Consequently, many other cellular targets and drug interactions have been identified using forward chemical proteomics

in combination with various labelling procedures. These include the detection of penicillin binding proteins in cell-lysates of bacteria using fluorescently or radioactively labelled penicillin analogs, kinase and the identification of non-kinase targets of BCR-ABL inhibitors, which are currently the frontline treatment for chronic myeloid leukemia[7, 8]. Such methods have subsequently been used to identify many interactions between natural products and their cellular targets[9].

However, protein concentrations in cells vary over many orders of magnitude and it is common to isolate weakly binding proteins that are quite abundant instead of rarer but more strongly binding proteins, which are likely to be responsible for the observed biological activities. Moreover, such techniques require significant amounts of proteins, which is not always feasible, particularly for proteins only present in low amounts. For example, didemnin B, a marine antiproliferative agent, was originally thought to inhibit of protein synthesis by binding with the abundant target elongation factor 1α. However, using affinity chromatography, additional binding proteins present in lower amounts were purified that showed high sequence similarity with palmitoyl protein thioesterase (PPT). At lower concentrations (three times less than required for inhibition of protein synthesis), the didemnin B is also responsible for cytostatic and immunosuppressive activities in tumour cells by binding to a PPT homolog[10, 11].

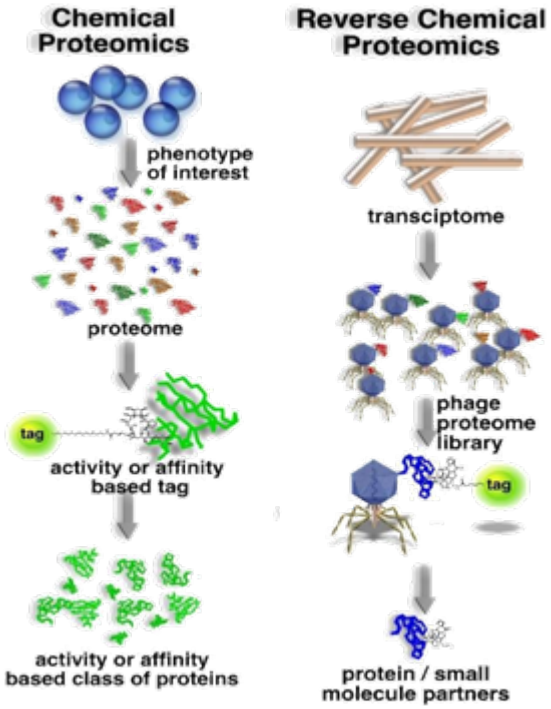



**Figure 2**: [A] Schematic diagram of forward chemical proteomics where a small molecule is used for affinity-based isolation of a single or group of proteins. [B] Schematic diagram of reverse chemical proteomics which starts with a transcriptome from desired sample which is expressed on the surface of compatible and reproducible vector and is probed with a small tagged molecule to identify its stronger binding target. (Figure adapted from Piggott and Karuso[4])

To overcome the limitations of forward chemical proteomics, reverse chemical proteomics was developed, which involves creating a physical link between each protein in a proteome under study and its encoding gene. This so-called "genotype-phenotype link" makes it possible to recover low copy number proteins using natural product affinity support, which can then be amplified using the attached gene and subjected to subsequent pull-down experiments[4]. One advantage of this technique is that the starting point is cDNA from the sample under study rather than a proteome, so there is no need to extract proteins or perform any manipulations as everything is carried out at the gene level. The most common approach for creating a phenotype-genotype link is display cloning[12], where the transcriptome is cloned on the surface of a compatible and amplifiable vector, which is later expressed as a fusion protein. This is followed by introducing either radioactively labelled or more recently fluorescently tagged, biologically active small molecule or other relevant epitope tags or probes that can isolate vectors displaying only those proteins capable of binding to the small molecule. The resulting vectors can then be sub-cultured, reamplified and procedure is repeated until only the strongest binding member remains[4]. This iterative approach helps to identify the most avid binding partner for a natural product and can be particularly useful where low copy number proteins are proposed to be the strongest binding target. A significant problem with both forward and reverse chemical proteomics is the derivatisation of small molecule can adversely affect its biological activity and hence this possibility needs to be excluded beforehand[13].

Many surface display cloning technologies have been used to date, including phage display, bacterial display, mammalian display, ribosomal display and mRNA display[14]. The most common display cloning strategy is phage display. This technique was first described by Prof. George Smith in 1985, whereby *E. coli* filamentous phages like M13 were used to express foreign proteins as a fusion with phage-coat protein[15, 16]. When such phages are transfected into *E. coli*, they take over the host machinery, resulting in reproduction and amplification of the phage particles. In order for phage particle to be secreted, they must be transported to the cell membrane of *E. coli*, the region of phage assembly after infection. This is fulfilled by a signal sequence encoded by the phage[17].

### 3.3.1 Phage Surface Display

Phage display has been used for variety of applications, including the identification of random peptide libraries to identify peptide mimics, antibody display as well as randomly fragmented polypeptide display[17-19]. It has also been used for the display of cDNA libraries[20]. Here, the cDNA library is cloned as a fusion with the gene encoding the phage coat protein, so

when such a phage infects its bacterial host and reproduces, the foreign protein is expressed as fusion with phage coat protein and each phage displays a different protein from cDNA library. The tagged natural product is then introduced and phages expressing foreign protein/s having affinity for the natural product can be separated by affinity selection procedure known as biopanning. Such phages expressing desired foreign protein from cDNA library can be enriched by subsequent sub-culturing and the process can be repeated for confirmation as well as identification of the strongest binding partners.

Phage display has been used to discover cellular targets and modes of action for large number of natural products to date. For example, ribosomal protein S25 was identified as a binding target for anticancer drug kahalalide F, which is under clinical trial phase II[13]. A major limitation with this technique is that phages are viruses and do not have their own biosynthetic machinery. *E. coli* machinery is prokaryotic and is not always compatible with the display of eukaryotic proteins. This is because of the lack of post-translational modifications and chaperone proteins in prokaryotes, which can lead to eukaryotic proteins being mis-folded and non-functional. Additionally, a relatively low copy number of fusion proteins are displayed on each phage particle[21].

### 3.3.2 Yeast Surface Display (YSD):

Yeast Surface Display (YSD) can complement phage display as yeast has the ability to successfully express most foreign eukaryotic proteins. Moreover, YSD is highly compatible as compared to other display systems for directed evolution of eukaryotic proteins. Yeast protein expression pathways are similar to those in mammalian cells, so foreign proteins displayed on the yeast cell surface have a higher chance of folding correctly and being functional. Besides being eukaryotic, *Saccharomyces cerevisiae* is a well-studied organism and also easy to handle due to its non-pathogenicity and small size. The genome of *S. cerevisiae* is fully sequenced and well-annotated and hence it is used widely for molecular biology applications like recombinant engineering, targeted proteomics and synthetic biology[22]. Yeast two-hybrid systems have been used to screen for larger expressed protein libraries with specific binding properties. However, this approach is largely dependent on internal co-expression of “bait” and “prey” fusion proteins and is not useful for identification of binding partners of natural compounds which are outside the cell[23]. Using yeast as a host combined with secretion-capture is another application based on label-free identification of target. Here, the yeast cell surface is chemically conjugated with small organic capture molecules. These molecules are characteristically specific but shows non-covalent interaction with target protein scaffolds that are secreted. This technique enables

study of protein-protein interactions without introducing labelled-target exogenously, which consequently decreases the probability of protein-scaffold modification[24].

YSD is similar to other strategies used for directed evolution such as phage display, ribosomal display, mRNA display, bacterial display and mammalian display in that it requires physical linkage between the gene encoding the protein (genotype) and protein expressed on the cell surface (phenotype), which was mentioned earlier as the genotype-phenotype link. A typical YSD system consists of three components (1) the target protein; (2) the anchor protein, and (3) the host strain.

One significant advantage of using YSD is that yeast contains a variety of cell surface anchor proteins, including Aga1p, Aga2p, Cwp1p, Cwp2p, Tip1p, Flo1p, Sed1p, YCR89w and Tir1[25]. A specific *N*-terminal hydrophobic sequence is present on all these surface receptor proteins that directs them through the secretory pathway of a yeast cell post-translational machinery. Additionally, an intramolecular spacer region rich in serine or threonine residues is also more likely to be present in these proteins. Their anchorage to the cell wall is dependent on the binding of glycosylphosphatidylinositol (GPI) to this *C*-terminal hydrophobic region[26]. Expression of foreign proteins fused at the *C*- or *N*- terminal of such surface receptor proteins can result in display of up to 10,000 copies on the surface of *S. cerevisiae*. The choice of anchor protein and fusion depends on the type of protein under study. Generally, the terminal that is farthest from the catalytic site/ functional portion of the protein is preferred for fusing the protein to avoid the loss of protein activity[27]. Moreover, steric hindrance may affect protein activity, so fusion orientations should be decided empirically[14]. The most common YSD system employs the *C*-terminus of the anchor protein a-agglutinin mating protein Aga2p subunit, which was first used by Boder and Wittrup[28]. The gene encoding the Aga1 protein is integrated into yeast chromosome, which also has Aga2 knockout mutation. The gene encoding the Aga2 receptor is provided by a yeast display vector, which allows fusion with foreign proteins. This Aga2 protein is linked with Aga1 through two disulfide bridges, resulting in a strong covalent complex (Figure-3) formed on the surface of yeast[14]. Expression of Aga1 and Aga2 is under the control of a galactose-inducible promoter, which helps in selection and specificity. Moreover, plasmids also possess auxotrophic selection markers, such as antibiotic resistance genes that imparts resistance to antibiotics like hygromycin and Zeocin to increase selectivity.

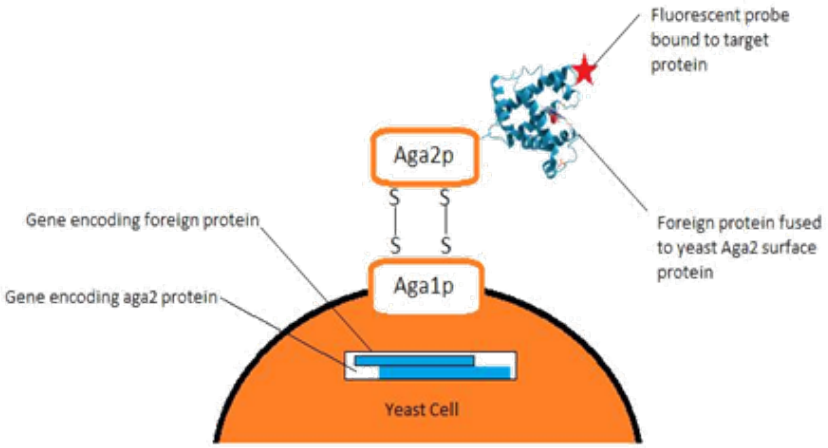



**Figure 3:** Schematic representation of YSD used for antibody/protein engineering, epitope mapping and other directed evolution strategies. The expression of foreign proteins fused with the Aga2 cell surface receptor take place under a galactose-inducible promoter on the surface of yeast cell[29].

YSD offers an additional advantage by allowing quantitative screening by flow cytometry using magnetic activated cell sorting (MACS) and more recently fluorescence activated cell sorting (FACS). FACS allows the observation of equilibrium activity and sample statistics during screening process, making it more efficient than MACS. Moreover, sorting via FACS provides more precise selection of cell populations and is less time consuming[30]. Also, this technique eliminates artifacts that may arise due to host expression bias. This is because FACS makes it possible to simultaneously screen combinatorial libraries binding to multiple targets with varying affinities. Moreover, biophysical characterisation of individual protein binders can also be performed as these proteins exist as fusion proteins on the cell surface and there is no need for sub-cloning, soluble expression and purification[14, 29]. YSD in conjunction with FACS is widely used for antibody engineering, allowing fine discrimination between mutants, epitope mapping and other directional evolution strategies[31]. For example, YSD has been successfully applied to engineer antibodies to T-cell receptors, huntingtin protein, carcinoembryonic antigen, botulinum neurotoxin and many more to increase antibody affinity, specificity and stability[27, 32].

## 3.4 APPLICATIONS OF YSD:

Due to all the advantages listed above (Section 3.3.2), YSD has been widely used for a variety of applications, some of which are described below:

### 3.4.1 Protein-Protein Interactions:

The yeast two-hybrid approach, which has been successfully used to analyse protein interactions in post-translational modifications as well as other small molecules, has some limitations. This is because the yeast two-hybrid approach relies on availability of assay-specific components which are not always feasible. Therefore, efficiency and specificity of the process varies, hence decreasing the reproducibility and reliability of the technique[14].

To overcome these limitations, YSD has been used to study protein-protein interactions involved in post-translational modifications. Here, target heterologous protein fragments are displayed in high copy numbers on the surface of *S. cerevisiae*, which provides the additional advantage of proper folding of eukaryotic proteins. This approach has proved to be useful in identifying high affinity protein-protein interactions of tyrosine-phosphorylated protein fragments derived from common auto-phosphorylation sites of epidermal growth factor receptor (EGFR) and focal adhesion kinase (FAK)[33]. For this, human testis cDNAs were synthesised followed by size selection (0.3-1.2 kb). These were cloned into a yeast display vector pYD1 plasmid to obtain high copy numbers of yeast surface displayed libraries. These human cDNA fragments were cloned as *C*-terminal fusions with the Aga2p gene, which were later expressed as a fusion of Aga2p + protein fragments translated from cDNA. Moreover, these cDNA library members were also attached with an Xpress tag at 5' end to monitor the level of expression and a V5 tag at 3' end to detect the amount of in-frame fusion. Both these epitope tags can bind to fluorescently tagged target only if there is no frame-shift mutation and the protein is fully expressed. For example, V5 epitope tag only gives fluorescence when it binds to monoclonal anti-V5 epitope antibody (Figure-4). Also, V5 epitope is only expressed if the cDNA open reading frame (ORF) insert is in the correct frame along its entire length both upstream with Aga2 coding region and downstream with V5 coding region.

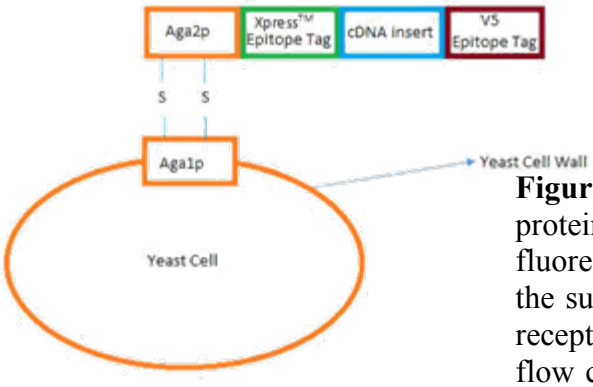



**Figure 4:** Schematic representation of YSD of proteins from human cDNA along with two fluorescent epitope tags, one at each terminal on the surface of yeast in fusion with Aga2p surface receptor protein. The fluorescence is detected by flow cytometric analysis and target cells showing positive fluorescence are sorted using FACS[34].

The cDNA library was induced by galactose and incubated with biotinylated tyrosine-phosphorylated peptides. These peptides were derived from auto-phosphorylation sites of EGFR or FAK. Non-phosphospecific binders were eliminated using corresponding non-phosphorylated, non-biotinylated peptides. During first round of FACS, only 0.5% of the population consisted of binding clones, which increased to 15% for EGFR diaplaying yeast clones and 40% for FAK displaying yeast clones after three rounds of selection. Besides this, human protein fragments with SH2 domains were also identified as these domains are known to interact with phosphotyrosine of EGFR[35]. Even though this human cDNA library was successfully displayed on yeast, there are potential issues for the display of cytoplasmic proteins. As the fusion proteins pass through the protein secretory pathway of yeast, this exposes them to an oxidising environment, thus increasing the probability of mis-folding. Additionally, some fusion proteins are glycosylated, thus altering their binding properties[28, 36].

### 3.4.2 Identification Small Molecule-binding Proteins:

In order to understand cellular-signalling pathways and molecular mechanisms, it is essential to identify interactions between participating proteins and small molecules[37]. However, this is often a challenging task. The activities of drugs, which are mostly small molecules, are frequently dependent on binding and interaction with the target proteins. Methods like affinity chromatography, protein microarrays, phage display, yeast hybrid and mammalian hybrid have helped to some extent to understand interactions between small molecules and their target proteins. However, due to some limitations mentioned previously and limited reproducibility, precision and specificity of these techniques, they have not been used extensively.

YSD was successfully used for the identification and characterisation of phosphatidylinositide-binding protein domains. Phosphatidylinositides are a class of lipids that play a vital role in a range of cellular processes, including signal transduction, cytoskeletal organisation and membrane trafficking. Construction of human cDNA libraries derived from human testes, liver, breast and brain, followed by display on the surface of *S. cerevisiae* has helped to identify proteins with affinity for phosphatidylinositides[38]. pYD1, which has been successfully used as a yeast display vector, was also exploited here. However, to increase the coverage of the entire human proteome, frameshift versions of pYD1 were constructed. Human cDNA fragments from several human tissues were cloned in these variants followed by its transformation in *S. cerevisiae* EBY100 creating a range of yeast surface-displayed libraries.

These libraries were expanded and induced for FACS based selection of yeast cells containing cDNA inserts for target proteins[38].

This approach helped identification of polypeptide fragments like (1) pleckstrin homology (PH) domains from SBF1, PDK1, β-spectrin, PSD and OSBP2; (2) phosphotyrosine-binding (PTB) domain and (3) apolipoprotein H (apoH) that had affinity for phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)$P_2$) and phosphatidylinositol-3,4,5-triphosphate (PtdIns(4, 5) $P_3$)[39]. For FACS, the PtdIns were attached to biotin followed by incubation with phycoerthrin-conjugated streptavidin (SA-PE). Also, the plasmid had a V5 epitope tag recognition sequence downstream of the cDNA cloning site and for expression of this tag, cDNA inserts must be in correct reading frame. The detection of this V5 epitope tag was achieved by treatment with Alexa-Fluor-647-conjugated anti-V5 monoclonal antibody followed by FACS based selection. This YSD method did not produce any major bias against intracellular proteins[38].

### 3.4.3 Epitope Mapping:

Another noteworthy application of YSD is for epitope mapping i.e. identifying binding of specific epitopes on the target protein. Yeast surface-displayed full length libraries or even single protein domain libraries can help to identify functionally important amino acid residues responsible for epitope binding[40]. Moreover, binding events are identified by labelling yeast cells with a target that is biotinylated along with a secondary anti-biotin reagent (in some cases antibody) that is attached to a fluorophore. This technique also involves a two-colour labelling procedure, where one fluorophore is specific for determining the expression levels of Aga2-protein scaffold fusion. For example, Aga2-protein scaffolds fusion are cloned in frame with well-studied epitope tags haemagglutinin (HA tag) and c-myc tags. So expression and normalisation of such proteins can be carried out by using corresponding fluorophores like anti-HA or anti-c-myc that can be further detected using flow cytometry. Apart from this, the use of a second fluorophore that is specific to target binding (in this case anti-biotin) has been shown to increase the specificity and also helps to select proteins with higher affinity and stability[14].

### 3.4.4 Antibody Engineering:

Besides epitope mapping, yeast surface display also has also been widely used for engineering antibodies and T-cell receptors. Using labelled antigens as targets, yeast surface display together with FACS has been used to select antibodies with high affinity for desired

antigen. This strategy allows more precise and quantitative selection process as compared to phage display. The selection process consequently results in clones enriched with the antibodies with desired binding properties[28]. Moreover, yeast surface display provides an additional advantage of cloning different genes from hybridomas or single chain variable fragments (scFvs) or antigen binding fragments (Fabs) can be directly inserted from phage display library into yeast display libraries by the use of improved vectors, site directed or random mutagenesis and ligation independent transformation procedures in yeast[29].

One of the most common applications of yeast surface display is engineering of scFvs and protocols are very well-established mainly for engineering protein affinity of single chain antibodies. The shuttle vector used (pCTCON) has single chain antibody fragments in fusion with Aga2p[25]. Only the ORF of the antibody fragment is inserted in frame with Aga2p to reduce chances of frame shift mutations. Mutant libraries of antibodies are generated using homologous recombination and transformed in yeast using electroporation. Homologous recombination reduces an extra step of transformation in *E. coli* before yeast transformation. The scFv insert DNA possesses random point mutations, which are introduced using error prone PCR or more recently DNA shuffling[41, 42].

After the generation of a large library diversity, the yeast cells displaying the antibody library are labelled with a fluorescent or biotinylated reagent, which allows quantification of binding affinity and those yeast cells displaying proteins with significant binding affinity for the target can be further sorted by FACS. An additional epitope tag at the *C*-terminal facilitates double specificity by eliminating non-displaying yeast and allows library normalisation. One of the major limitations that can be a problem when high-affinity antibodies are labelled is the depletion of antigen from the labelling mixture with time. This results in reduced concentration of freely available antigen leading to a lower detection signal. To avoid this, at least ten-fold higher concentration of antigen is used or the concentration of yeast cells displaying antibody libraries is reduced. After sufficient rounds of selection using FACS, increasing the binning stringency in each round, the final sorted cells are plated out for clonal analysis. The single clones are then sequenced to verify the identity of isolated proteins[41].

### 3.4.5 Use of Different Anchor Proteins:

Apart from a-agglutinin cell surface receptor protein (Aga protein) being used as an anchor protein to fuse foreign protein in most of the yeast surface displays stated above, other surface receptor proteins are also recently used for the display of small and large foreign protein

fragments as a fusion on the yeast surface on the yeast surface and its subsequent selection by FACS.

Proteins with internal repeats (Pir) are another class of anchor proteins which as a fusion partner, which provide an additional advantage of secreting target protein into the media. They also provide two binding sites for covalent attachment to the cell wall, (i) An ester linkage between *N*-terminal sequence repeats and β-1,3 glucose and (ii) *C*-terminal cysteine residues attach specifically to cell-wall by disulfide bonds. Moreover, Pir proteins allows three types of fusion of target proteins, (i) *C*-terminal fusion; (ii) *N*-terminal fusion and (iii) inserted fusion. Yeast surface display using Pir as anchor proteins has been used on proteins from a range of organisms including spike protein from rotavirus, human interleukin-1 β and endoglucanases of *Paenibacillus barcenonensis*. However, the efficiency and degree of secretion using Pir as anchor proteins in yeast display is higher with fusion of prokaryotic proteins as compared to eukaryotic proteins[43].

More recently, Spi1 has been used as an anchor protein for yeast surface display. Spi1 is a cell surface protein that is expressed in centromeric or episomal plasmids and is under the control of either its own promoter or phosphoglucokinase (*PGK1*) glycolytic gene, whose promoter is generally regulated under stress[26]. So Spi1 protein aids in increased resistance of cells in stationary phase to several stress conditions including herbicides, food preservatives and cell-wall lytic enzymes[44]. Spi1 protein was expressed in fusion with V5 epitope tag to detect its expression using fluorescence-based methods like FACS and fluorescence microscopy. After successful demonstration of Spi1 as an anchor protein, this system was successfully employed for the display of luciferase from *Photinus pyralis*. Luciferase is frequently used as a reporter protein as it catalyses the oxidation of luciferin in a reaction of ATP and oxygen to yield oxyluciferin, which emits light in visible range[45]. The display of this protein on the surface of yeast was confirmed by FACS, fluorescence microscopy as well as quantification of enzymatic activity. The results revealed that more than 70% of cells expressed Spi1 protein under starvation (sugar and nitrogen limiting conditions), which is similar to the other yeast surface display systems.

### 3.4.6 Dual Display of Foreign Proteins:

Another interesting application of YSD that was recently published used both the *N*- and *C*- terminal of Aga2p subunit for displaying two heterologous proteins as a fusion to Aga2p on each terminal. This was an addition to previous applications where foreign protein was fused

to either *N*- or *C*-terminal of the Aga2p subunit. The advantage of this approach was the antibody fragment, ligand or receptor can be directly fused to desired fluorescent protein, hence eliminating an additional tedious step of antibody staining of epitope tags. This method simplified the quantification of binding reactions between two proteins as this could be carried out directly on the yeast surface. Moreover, bioconjugation reactions between an enzyme and its substrate were also studied by co-expressing them on the same Aga2p construct. This allowed the study of enzyme expression and catalytic activity on the yeast surface itself[46].

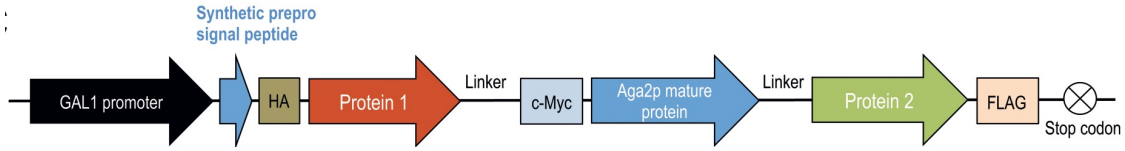



**Figure 5:** Schematic representation of dual protein yeast surface display (figure adapted from Lim *et al., 2017*[46]).

The study of protein-protein binding reactions was examplified by use of yeast codon optimised green fluorescent protein (yEGFP) and three model proteins (i) scFv D1.3, a murine antibody fragment that binds hen egg lysozyme, (ii) Ax1 Ig1, that binds its cognate ligand growth arrest specific 6 (Cas6) and (iii) NK1, a natural fragment first kringle domain of human hepatocyte growth factor ligand having strong binding for Met receptor[47, 48]. yEGFP was expressed at the *C*-terminal of Aga2p as the expression level was comparatively lower on *N*-terminal fusion. So all the model proteins were fused to the *C*-terminal of Aga2p followed by yEGFP to its *N*-terminal.

The bio-conjugation reactions on yeast surface were illustrated by well-studied enzyme system *Staphylococcus aureus* sortase A and its corresponding peptide substrate with various linkers. The bioconjugation reactions were confirmed by 3-azido-1-propanamine (Azp) induction, fluorescence studies and click reactions[46].

### 3.4.7 Analysis of Microecologic Therapy using YSD:

Microecologic therapy is the use of microorganisms like yeast and bacteria in a way that it benefits the gut microecology without harming the host[49]. The effects of enterotoxin fusion protein produced by enterotoxicogenic *Escherichia coli* (ETEC), one of the major cause of diarrhoea in animals has been recently studied using YSD. Heat labile and heat stable ETEC toxin *estA, estB* and *eltAB* were expressed on the surface of yeast and tested to study their effect on intestinal flora and mucosal immunity of rats. Techniques like ELISA, T-RFLP and real-time PCR were used to analyse and compare the levels of SIgA, IL-2, IL-4, IFN-γ as these

components have importance in intestinal immunogenicity. Results were compared with three different groups (i) rats injected with normal saline, (ii) rats injected with yeast only and (ii) rats injected with yeast displaying fusion proteins. It was confirmed that on feeding these groups with ETEC, the rats with yeast displaying enterotoxin fusion protein showed a significant increase in number and variety of gut microbiome as compared to other two groups[50].

## 3.5 SUMMARY:

In summary, yeast surface display can be used to identify target protein from whole proteomes if the probe can be fluorescently labelled. Full length of cDNA libraries can be screened using FACS. The equilibrium binding constant (Kd) and dissociation rate constants of proteins identified using yeast surface display are similar to those obtained using *in vitro* techniques like ELISA and SPR. Thus, yeast surface display can be used to study cellular mechanisms at molecular level thus broadening our understanding of a cell.

Despite such wide range of applications, yeast surface display has not yet been exploited to identify the cellular target and modes of action of small molecules such as anti-parasitic natural products. However, Parasites being, invariably eukaryotes, require a eukaryotic host for their protein expression. So, YSD is proposed to be highly compatible for identifying the cellular targets of anti-parasitic natural products. This project is primarily focused on optimising experimental parameters of YSD strategy for identification of a strong protein binders from an entire proteome of parasites for a natural product under study and construction of cDNA library of desired organism followed by its validation using well-studied fluorescently labelled natural product. The cDNA library of clinically important parasites will be cloned in fusion with Aga2 gene at C-terminal that later gets expressed as fusion protein. This yeast surface display library will then screened against fluorescently labelled anti-parasitic natural products to identify avid binding targets that are responsible for observed anti-parasitic activity.

# 4. CONSTRUCTION OF A YSD CLONE DISPLAYING HUMAN PROTEIN FKBP2 AS A POSITIVE CONTROL FOR FACS OPTIMISATION

## 4.1 OVERVIEW

As YSD technology has not been exploited for identifying cellular targets of natural products, it is essential to establish and optimise experimental procedures by constructing a positive control using a well-studied protein-target interaction. FK506, a well-known immunosuppressant isolated from *Streptomyces tsukubaensis*, strongly binds to FK506 binding proteins (FKBPs) and these interactions have been well-studied using phage display[51, 52]. Moreover, as FKBPs are ubiquitous in nature, FK506 can be used as a suitable probe to validate YSD libraries from a range of different organisms. To achieve this, a human FKBP2 (*h*FKBP2) gene previously rescued from a phage-displayed human brain cDNA library[52] was isolated from its T7 Select 10-3b vector and cloned in pCTZQ, a yeast display vector (Figure 8). On transformation into *Saccharomyces cerevisiae* EBY100, a model system EBY100 pCTZQ-FKBP2 was formed for further optimisation experiments.

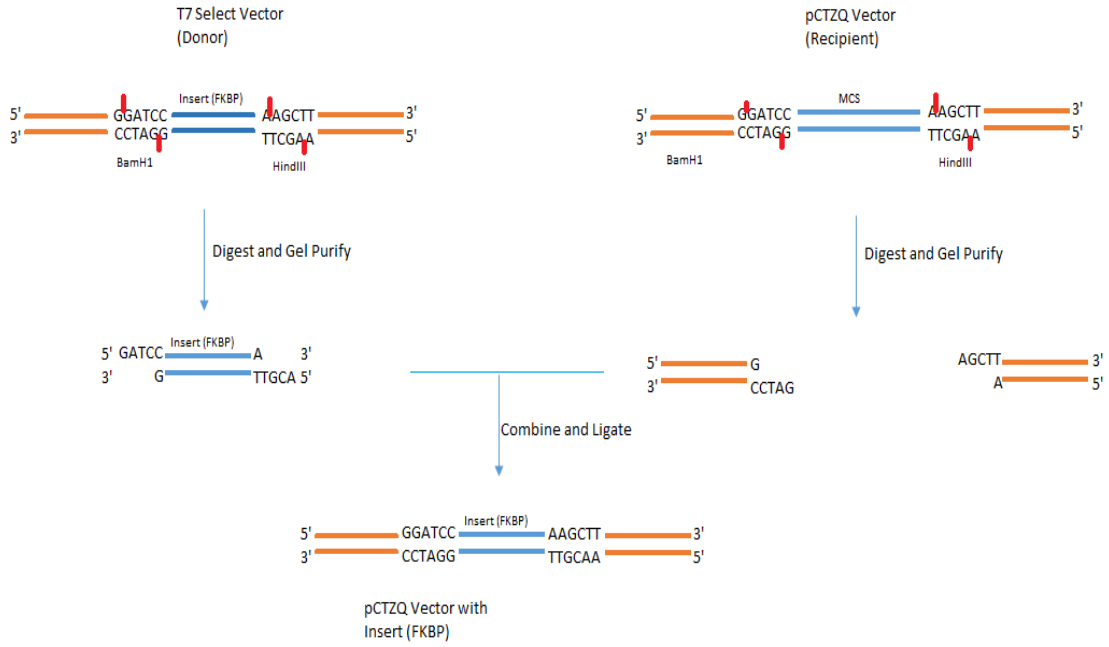



**Figure 6:** Schematic presentation of construction of pCTZQ plasmid containing a human FKBP2 gene. This gene was digested from T7 select vector using *Bam*H1 restriction site at 5' end and *Hin*dIII at 3'-end followed by insertion into pCTZQ plasmid, which was also digested using *Bam*H1 and *Hin*dIII at 5' and 3' end respectively.

### 4.1.1 T7 Select Vector with *h*FKBP2 Gene (Donor Vector)

The T7Select cloning system includes vectors with *Bam*H1/*Hin*dIII arms, ready for directional cloning of appropriately prepared inserts. The T7-Select 10-3b clone containing human FKBP2 was kindly provided by Dr. Andrew Piggott (Macquarie University)[52]. To provide compatibility with the yeast vector arms and obtain expression in-frame with the yeast Aga2p protein such that recombinant fusion proteins are successfully displayed on the yeast surface, inserts must contain correctly designed termini. The FKBP2 insert requires a 5'-GATC "sticky end" (*Bam*H1) on the top strand (amino-terminal side) and a 5'-AGCT "sticky end" (*Hin*dIII) on the bottom strand (carboxy-terminal side), either created with oligonucleotides using PCR or by *Bam*H1/*Hin*dIII double enzyme restriction digestion reaction.

5'...TGGTCTTCGCCCAGAAGCTGCAGGAGCTGTCGTATTCCAGTCAGGTGTGATGCTCGGGG
ATCCGAATTCAAGCACGGCCACGGGGGCCGAGGGCAAAAGGAAGCTGCAGATCGGGGTCAAG
AAGCGGGTGGACCACTGTCCCATCAAATCGCGCAAAGGGGATGTCCTGCACATGCACTACAC
GGGGAAGCTGGAAGATGGGACAGAGTTTGACAGCAGCCTGCCCCAGAACCAGCCCTTTGTCT
TCTCCCTTGGCACAGGCCAGGTCATCAAGGGCTGGGACCAGGGGCTGCTGGGGATGTGTGAG
GGGGAAAAGCGCAAGCTGGTGATCCCATCCGAGCTAGGGTATGGAGAGCGGGGAGCTCCCCC
AAAGATTCCAGGCGGTGCAACCCTGGTGTTCGAGTGGAGCTGCTCAAAATAGAGCGACGAAC
TGAGCTGTAA...AAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAAC
GGGTCTTGA...GGGGTTTTTTGCTGAAAGGAGG...3'

**Figure 7:** Important parts of T7 Select 10-3b vector used in this project including full-length coding sequence of *h*FKBP2 gene which is inserted between *Bam*H1 and *Hin*dIII vector arms. Key: **T7 Forward Primer**, ***Bam*H1**, ***h*FKBP2 Gene**, **T7 Reverse Primer**, ***Hin*dIII**

### 4.1.2 Plasmid pCTZQ- A Shuttle Vector used for YSD (Recipient Vector)

Plasmid pCTZQ (adapted from the unpublished work of Dr. Jenny Vo, Macquarie University) was derived from plasmid pCTZ (kindly provided by Dr. Prof. Lars Poulsen, Technical University of Denmark). One of the major advantages of this plasmid is its ability to act as a shuttle vector that can propagate in two different host species, namely *Escherichia coli*, which could act as a storage of large number of plasmid copies, and target species *S. cerevisiae*, which is the system under study. Additionally, the pCTZQ plasmid contains the Aga2p gene as *S. cerevisiae* strain EBY100 has an Aga2p knockout mutation. So, a complete cell-surface receptor protein is only expressed in presence of plasmid under galactose-inducible promoter provided by pCTZQ. Moreover, pCTZQ possesses two selection markers (a) an ampicillin marker, which confers resistance to ampicillin confirming successful transformation in *E. coli* and (b) a Zeocin marker, which confers resistance to Zeocin confirming successful transformation in *S. cerevisiae.* Plamid pCTZQ possesses a *Bam*H1 recognition sequence at the 5' end in top strand and a *Hin*dIII recognition sequence at the 3' end in bottom strand. This

provides sticky ends compatible for ligation with *h*FKBP2 gene purified from T7- select system having similar mapping of restriction enzyme recognition sequence.

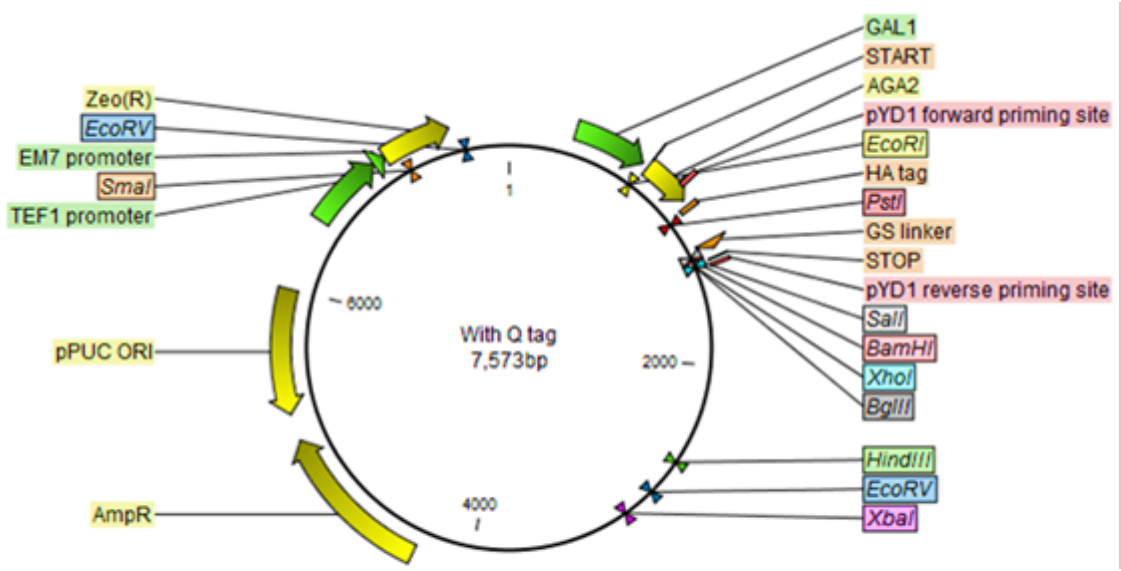



**Figure 8:** Plasmid map of pCTZQ (adapted from unpublished work of Dr. Jenny Vo, Macquarie University). The promoter is induced in presence of galactose which allows the use of selective media.

## 4.2 RESULTS AND DISCUSSION

### 4.2.1 Construction of pCTZQ-FKBP2

The FKBP2 gene from the T7 Select vector was amplified using PCR (Section 4.3.2). The gel image of the PCR products obtained after agarose gel electrophoresis (1.5%) (Figure 9) confirmed that the PCR product was ~500-600 bp. The concentration of the purified PCR product (*h*FKBP2) was 10.4 ng/µL.

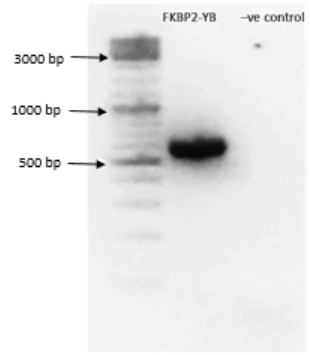
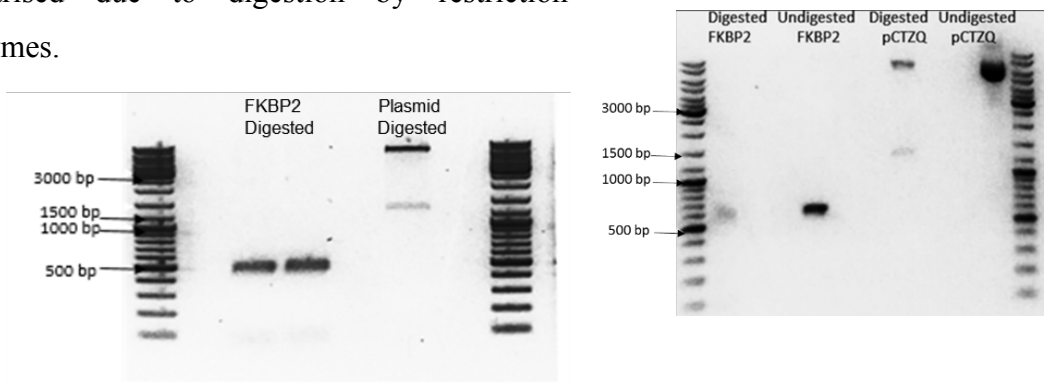



**Figure 9:** Gel image obtained for 1.5% agarose gel electrophoresis using UV-transilluminator. The size of PCR product is approximately 500-600 bp obtained with T7 forward and reverse primers which supports the amplification of FKBP2 gene. Negative control was the PCR reaction without any template DNA.

pCTZQ stored in *E. coli* DH5α strain was grown overnight at 37 °C to obtain a fresh exponentially growing culture. Moreover, ampicillin present in LB-broth provided good selection in a way that the resulting fresh culture was comprised of only the cells containing

pCTZQ. So, pCTZQ was isolated from cells (as described in Section 4.3.2) and the resultant plasmid miniprep (46.2 ng/µL) was used for restriction enzyme double digestion.

Restriction enzyme double digestion was carried out for 4 h to ensure nearly 100% digestion. Agarose gel electrophoresis (1%) confirmed double digestion of FKBP2 and pCTZQ using *Bam*H1 and *Hin*dIII (**Figure 10** A). For FKBP2, the desired digested fragments are around 500 bp (507 bp as per sequence). For pCTZQ, the desired fragments are approximately 6000-7000 bp, while undesired fragments are approximately 1200-1500 bp (1274 bp as per sequence). Agarose gel electrophoresis (0.7%) was also performed (Figure 10 B). Uncut pCTZQ plasmid migrates further as compared to digested pCTZQ plasmid which gets linearised due to digestion by restriction enzymes.




**(A)** **(B)**

**Figure 10: (A)** Gel image of 1% agarose gel electrophoresis to visualise double digestion of FKBP2 and pCTZQ using *Bam*H1 and *Hin*dIII. **(B)** Gel image of 0.7% agarose gel electrophoresis of FKBP2 and pCTZQ for comparing migration behaviour of cut DNA fragments with uncut DNA. The digested FKBP2 band was faint due to decrease in size of the band after digestion by restriction enzyme.

From the data of DNA profiling obtained from restriction double digestion and the plasmid sequence, the length of desired vector was 6299 bp and the length of insert DNA was 507 bp. As *Bam*H1 could not be heat inactivated, the restriction digestion reaction needed to be purified by extracting the desired fragments from gel followed by purification silica membrane on which DNA binds in presence of salts and at pH equal to lower than that of surface silanol groups (~7.9 at RT). Moreover, this also reduced the probability of self-annealing of DNA fragments from the same vector. However, the purification and gel extraction decreased the yield of both pCTZQ (13.6 ng/µL) and *h*FKBP2 gene fragment (1.1 ng/µL) which were then processed for test ligations.

The ligation reactions were setup after calculating the volume of insert for desired molar ratio of vector: insert ($x$) obtained using following formula:

$$insert\ mass\ (ng) = x \times \frac{insert\ length\ (bp)}{vector\ length\ (bp)} \times vector\ mass\ (ng)$$

As both vector and insert DNA had compatible sticky ends, lower insert : vector ratios were required as compared to insert : vector ratios required to ligate blunt ends. However, taking into account the low DNA yield obtained from gel extraction and purification, the following three combinations of insert : vector ratios were tested (i) 4:1 (ii) 7:1 (iii) 9:1. The ligation reactions were purified (as described in Section 4.3.2) to eliminate salts present in ligase buffer which may cause arcing during electroporation.

### 4.2.2 Transformation of pCTZQ-FKBP2 in *Escherichia coli* using Electroporation

Electro-transformation procedure predominantly depends on two characteristics of electric pulse (i) electric field strength and (ii) pulse length. Moreover, the transformation frequency is directly proportional to DNA concentration while transformation efficiency is directly proportional to cell concentration subjected to electroporation[53]. So, thawed *E. coli* Electromax DH10B cells were mixed with test ligations as well as pUC19, a reference plasmid as positive control and subjected to optimum exponential decay pulse generated at 2.0 kV, 200 Ω and 25 μF. The effective resistance in parallel with electrodes, which determines the time constant of the pulse, was lower than that of the sample, giving time constant of 5 ms.

Two types of positive control were used to confirm both cell viability and transformation efficiency. On LB-plate without ampicillin, 107 colonies were obtained with a 1:10 dilution which confirmed cell viability. When a 1:10 dilution of cells transformed with pUC19 was plated on LB-plate containing ampicillin, 91 colonies were obtained, giving 85% transformation efficiency. However, with a 1:10 dilution of the cells transformed with test ligations, only 1 colony was obtained with a 4:1 ratio, 1 colony with a 7:1 ratio and no colonies with 9:1 ratio giving 9% transformation efficiency. The low transformation efficiency with test ligations can be explained by the low DNA concentration obtained after purification of ligation reactions, thereby decreasing the amount of plasmid available for the competent cells.

However, as the aim of this study was to obtain a cell/colony with plasmid containing only one type of insert (pCTZQ-FKBP2), no further optimisation of ligation reaction or transformation procedures was performed. Instead, the plasmids were isolated from the colonies that were obtained to confirm the identity of the insert within the plasmid and were quantified for further analysis.

Consequently, both the colonies were re-grown (Section 4.3.3) followed by plasmid isolation. As the plasmid miniprep of this transformants was highly concentrated [(i) Transformant-1 = 173 ng/µL (ii) Transformant-2 = 236 ng/µL], each of the solutions was diluted 100-fold to prevent non-specific amplification due to high template concentration. T7 forward and reverse primers used previously (Section 4.1.1) could not be used at this stage because after digestion by *Bam*H1 and *Hin*dIII, the primer-binding site for those primers were no longer present. So, FKBP2 forward and reverse primers, with annealing sites within the FKBP2 gene making a product of 235 bp (as per sequence) were used here. Agarose gel electrophoresis (1.5%) confirmed ~250 bp product with 35 and 40 PCR cycles (Figure 11). Moreover, optimum results were obtained with 40 PCR cycles with an annealing temperature of 55-56 °C. These PCR products were sequenced for BLAST analysis to confirm their identities.

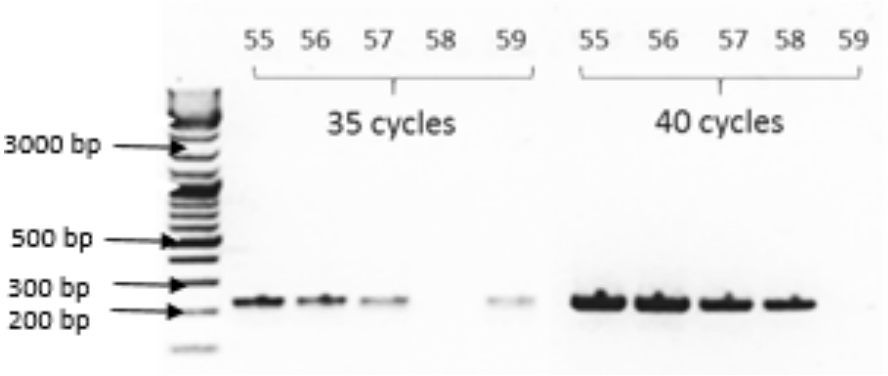



**Figure 11:** Gel image obtained from agarose gel electrophoresis (1.5%) of temperature gradient PCR carried out at 5 different annealing temperatures from 55 °C to 59 °C (title of each well) and two different PCR cycles (35 and 40) to obtain optimised PCR conditions. At 58 °C with 35 cycles as well as 59 °C with 40 cycles, no PCR product was obtained.

BLAST analysis of the sequence confirmed that 12-205 bp had 99% similarity with FKBP2 variant-1 gene at location of 409 bp to 604 bp within the gene from *Homosapiens*. This was followed by its annotation to locate reverse primer binding site. These sequencing analysis and BLAST results confirmed that the pCTZQ-FKBP2 plasmid was successfully transformed into *E. coli*.

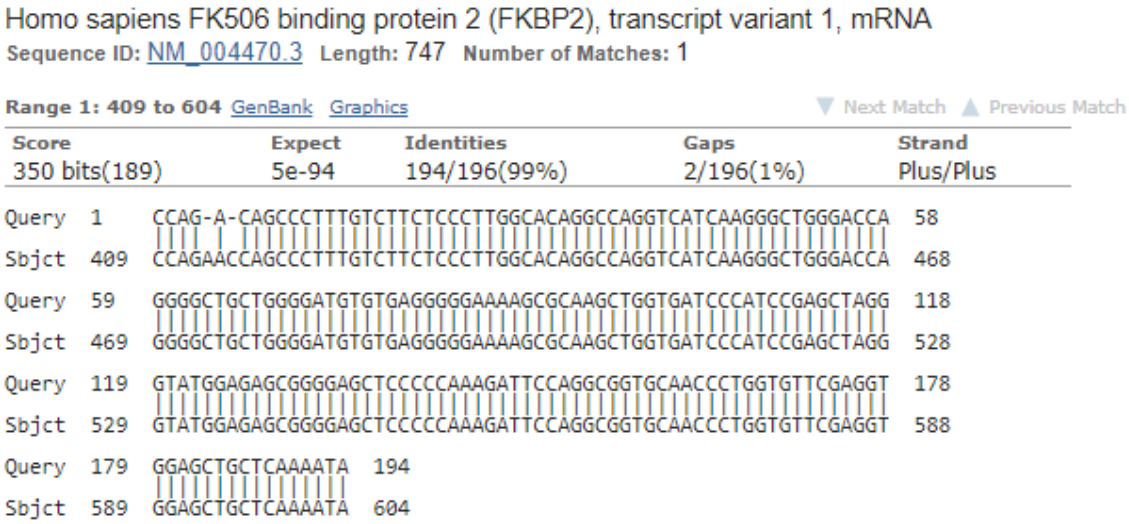



**Figure 12:** BLAST analysis of sequencing data for the PCR product obtained from temperature gradient PCR (Figure 11) which confirmed identity of *h*FKBP2 gene as well as a successful pCTZQ-FKBP2 construct and its transformation in *E. coli*.

### 4.2.3 Transformation of pCTZQ-FKBP2 in *S. cerevisiae* EBY100

After sufficient evidence was obtained to confirm successful transformation into *E. coli*, the derived plasmid construct pCTZQ-FKBP2 was subjected to final step of transformation in *S. cerevisiae* EBY100. This step also did not demand high transformation efficiency as transformation of only one kind of plasmid (unlike a library) was required. So, yeast transformation in *S. cerevisiae* EBY100 was carried out by lithium acetate (LiOAc)- mediated yeast transformation (described in Section 4.3.5) using single-stranded DNA (ss-DNA). A high concentration of plasmid miniprep was used to ensure availability of plasmid DNA to the maximum number of exponentially growing yeast cells leading to successful transformation.

The transformation efficiency was compared on two kinds of selection media (i) YPD + Zeo$^{50}$ and (ii) SD-CAA + Zeo$^{50}$. On a YPD plate without Zeocin, in 48 h, $10 \times 10^4$ colonies were obtained and on YPD plate with Zeocin, an average of $9 \times 10^3$ giving 88% transformation efficiency. However, on SD-CAA plates containing Zeocin, transformants were obtained after 72 h with transformation efficiency of only 27%. This can be explained by Zeocin being sensitive to high salt concentration as well as SD-CAA media having some inorganic salts making it less nutrient-rich as compared to YPD, which mainly comprises of all organic constituents. Success of the transformation was confirmed by yeast colony PCR.

### 4.2.4 Yeast Colony PCR and DNA Profiling of PCR Products

Yeast cells cannot generally be lysed simply by physical or mechanical cell-disruption methods due to presence β-glucans on yeast cell-wall. β-glucans are alkali-insoluble making the cell-wall strong and rigid[54]. So, enzymatic cell-lysis using 2U Zymolase was used to release plasmid from the yeast cells before PCR. pCTZQ-F primer, FKBP2F primer and FKBP2-R primer were used to confirm presence of plasmid pCTZQ-FKBP2 in *S. cerevisiae* EBY100 as well to confirm that the FKBP2 gene was in-frame with Aga2p subunit. Agarose gel electrophoresis (1.5%) confirmed a ~700 bp product with pCTZQ-F primer and FKBP2R primer and ~250 bp product with FKBP2F primer and FKBP2R primer (Figure 13). The latter confirmed the presence of FKBP2 gene (Section- 2.2.2) while ~700 bp product was sequenced for further analysis.

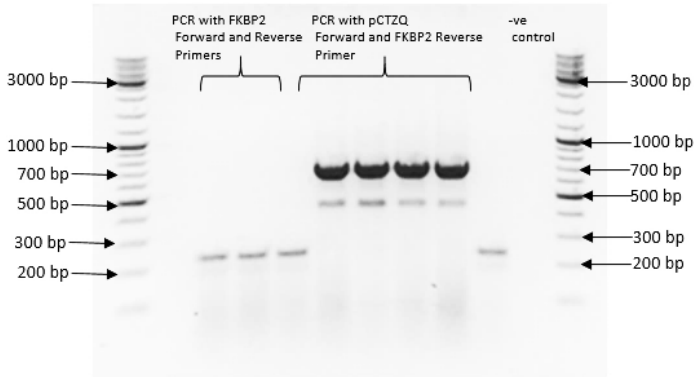



**Figure 13:** Gel image from agarose gel electrophoresis (1.5%) of yeast colony PCR with two different forward primers (pCTZQ-F and FKBP2-F in separate tubes) and FKBP2-R primer to confirm if *h*FKBP2 gene is in-frame with Aga2P subunit gene. The ~400 bp product formed with pCTZQ-F and FKBP2-R primers could be incomplete *h*FKBP2 being cloned in the plasmid pCTZQ. The PCR product in the well before negative control is the positive control for FKBP2 F and R primers (Section 4.2.2).

A chromatogram of 719 bases was obtained from sequencing and was annotated as for the confirmation of in-frame pCTZQ-FKBP2 construct. The final plasmid map of the entire pCTZQ-FKBP2 construct is shown in the Figure 14.

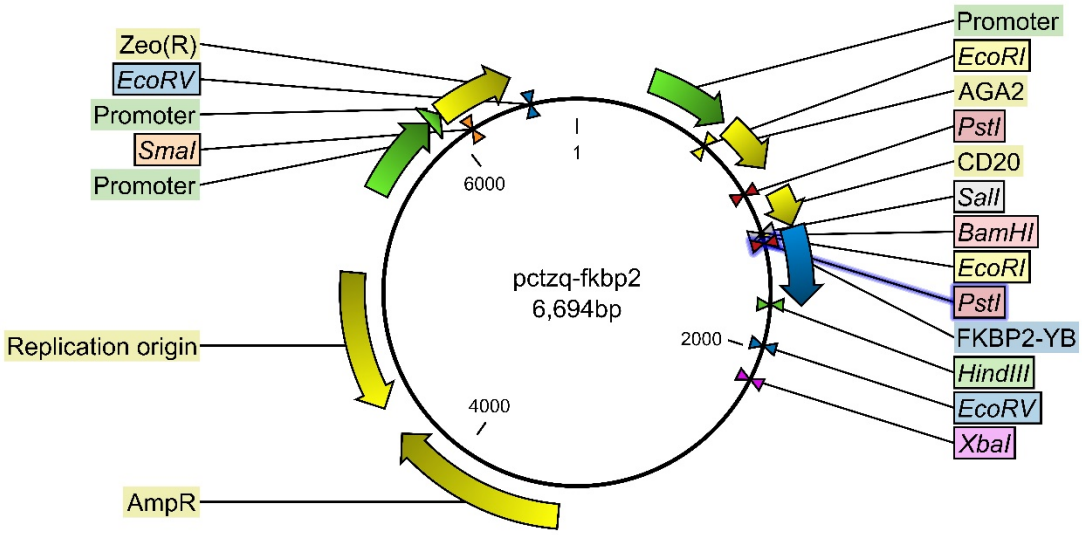



**Figure 14:** Schematic representation of pCTZQ-FKBP2 where *h*FKBP2 gene is inserted in-frame with Aga2p subunit.

## 4.3 EXPERIMENTAL PROCEDURES

### 4.3.1 Growth, Strains and Media

For fresh plasmid miniprep, 50 µL of −80 °C frozen stock of *E. coli* DH5α strain containing pCTZQ plasmid was aliquoted in 5 mL LB broth (1% tryptone, 1% NaCl, 0.5% yeast extract) supplemented with 100 µg/mL of Ampicillin as an antibiotic marker and incubated at 37 °C overnight. *E. coli* ElectroMax DH10B cells (catalogue number 18290015) were used for transformation of test ligations using SOC broth (2% w/v tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, 20 mM glucose) and LB-pates containing 100 µg of Ampicillin and an overnight incubation at 37 °C. *S. cerevisiae* EBY100 strain was used for yeast transformation using YPD media (1% yeast extract, 2% bactopeptone, 2% glucose/dextrose) as well as SD-CAA media as described previously[29] both supplemented with 50 µg of Zeocin, and 2-3 days incubation at 30 °C.

### 4.3.2 Construction of YSD Vector pCTZQ-FKBP2

FKBP2 gene (NCBI Reference Sequence- NM_004470.3, Uniprot-ID- P26885) was amplified from T7 Select 10-3b vector using T7 Select forward (Length- 20 Tm- 70.2 °C) and reverse primers (Length- 22 Tm- 68.4 °C) as mentioned earlier (Section 2.1.1). The thermal cycling conditions were initial denaturation at 94 °C for 3 min, 35 cycles of (i) denaturation (94 °C, 45 s), (ii) annealing (52 °C, 45 s), (iii) extension (72 °C, 45 s) and final extension at 72 °C for 6 min followed by agarose gel electrophoresis (1.5%) for confirmation of desired PCR

product. The desired PCR product was purified for further downstream processing using Sigma-Aldrich GenElute PCR clean-up kit (catalogue no- NA1020) according manufacturer's instructions. pCTZQ plasmid was isolated from the freshly grown overnight culture of *E. coli* DH5α strain containing pCTZQ plasmid (adapted from the unpublished work of Dr. Jenny Vo) using NEB Monarch Plasmid Isolation Kit (catalogue number- T1010S). All the quantifications were carried out using Invitrogen Qubit Fluorometer 1.0 according to manufacturer's instructions.

The restriction digests for FKBP2 gene purified from T7 Select system (donor) and plasmid pCTZQ (recipient) were set up using NEB *Bam*H1 High Fidelity (HF) (catalogue number- R3136S) and NEB *Hin*dIII HF (catalogue number- R3104S) restriction enzymes according to manufacturer's instructions. A large quantity of DNA was used as a starting material (vector-92.4 ng, insert- 104 ng) due to probability of losing some DNA during the gel purification step. Agarose gel electrophoresis (1%) was carried out to confirm sufficient digestion and to prepare for gel extraction of desired digested products. Moreover, agarose gel electrophoresis (0.7%) was carried out to study the difference of migration behaviour between digested and undigested products. Both the desired digested fragments (~1200 bp from pCTZQ and ~500 bp *h*FKBP2 were gel extracted using GenElute™ Gel Extraction Kit (catalogue number- NA1111-1KT) as per manufacturer's instructions.

The gel-extracted purified digested products were subjected to ligation reactions with different molar ratios of vector : insert. These ratios were decided considering the ligations of sticky ends and loss of some DNA during gel extraction which lowers the DNA yield. T4 DNA ligase (catalogue number- M0202) was used to set up test ligations according to manufacturer's protocol followed by overnight incubation at 16 °C as recommended for ligation of cohesive ends. The ligation reaction was heat inactivated at 65 °C for 10 min followed by purification of test ligations from other components of reaction for electro-transformation in *E. coli*.

### 4.3.3 Transformation in *E. coli*

Each of the purified test ligations was mixed with thawed *E. coli* Electromax DH10B cells (catalogue number- 18290015, ThermoFisher Scientific). Electroporation (BioRad GenePulser Xcell™) and heatshock treatment (42 °C) were carried out according to manufacturer's instructions. Different dilutions of electroporated samples and positive controls were spreaded on LB plate containing 100 µg of ampicillin and incubated overnight at 37 °C. These plates were used to determine the transformation efficiency. The transformant colonies

were expanded in 5 mL LB- broth + $Amp^{100}$ and plasmids were isolated and quantified using Qubit Fluorometer as described previously (Section 2.3.2). Temperature gradient PCR was carried out for confirmation of desired transformation and other optimisations.

### 4.3.4 Temperature Gradient PCR

Temperature gradient PCR was used to optimise the temperature for annealing conditions for primers at its binding site within template based on its Tm as well as optimum number of PCR cycles which enables maximum amplification of the desired product. So, PCR was carried out using Dream *Taq* Polymerase (catalogue number- EP0711, ThermoFisher Scientific) as well as FKBP2-F (5' GAAGCTGGAAGATGGGACAG 3' - Tm 63.7 °C) and FKBP2-R (5' ATTTTGAGCAGCTCCACCAC 3' - Tm 63.3 °C) primers. The thermal cycling conditions used were initial denaturation at 95 °C, 2.5 min; 35 and 40 cycles each for (i) denaturation 95 °C, 30 s, (ii) annealing 55 °C-59 °C gradient, 30 s, (iii) extension 72 °C, 60 s; followed by final extension at 72 °C for 10 min. Agarose gel electrophoresis (1.5%) was carried out for DNA profiling and PCR products were purified (Section 2.3.2) and sequenced (Macrogen Sequencing Services) for further analysis.

### 4.3.5 Transformation in *S.* cerevisiae

Transformation in *S. cerevisiae* was carried out by LiOAc mediated yeast transformation with some modifications using Salmon sperm ss-DNA (catalogue number- 15632011, ThermoFisher Scientific) as a carrier. An overnight expanded culture of *S. cerevisiae* EBY100 (500 µL) was further inoculated in YPD media (5 mL) until $OD_{600}$ reaches 1.0 (~ $1\times 10^7$ cells/mL) or 6 h whichever was early. This exponentially growing yeast culture (1 mL) was centrifuged twice at 4000 × *g* for 1 min to ensure that no traces of older media remains by carefully aspirating supernatant at each step.

Pre-made transformation solution (0.5 mL; 10% $PEG_{400}$, 0.1 M LiOAc, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA) was mixed with yeast pellet along with 4 µL ss-DNA (10 mg/mL) and 10 µL plasmid DNA (pCTZQ-FKBP2, ~1 µg) and incubated overnight at RT. This was followed by heatshock at 42 °C precisely for 20 min and pelleted at 4000 × *g* for 1 min. The pellet was resuspended in YPD media (1 mL) for 4 h to allow the plasmid to expresss followed by centrifugation at 4000 × *g* for 1 min and resuspension in YPD media. Different dilutions of this inoculum (1,000 fold and 10,000 fold) was spreaded on YPD plates and SD-CAA plates both supplemented with 50 µg of Zeocin and incubated at 30 °C for 48-72 h followed by analysing transformants and transformation efficiency.

### 4.3.6 Yeast Colony PCR and Sequencing Analysis

Yeast cell lysis was performed by dissolving an individual yeast colony in 2 U Zymolase (50 µL in 1× TE buffer) followed by incubation at 37 °C for 30 min and stopping enzyme activity by heat inactivation at 95 °C for 10 min. An aliquot (2 µL) of this lysed cell solution was used as template for PCR which was carried out using pCTZQ-F Primer (5' AGTAACGTTTGTCAGTAATTGC 3' - length 22, Tm 57.5 °C), which binds to the Aga2p subunit gene in pCTZQ and FKBP2-F as well as FKBP2-R primers used previously (Section 2.3.4), Platinum *Taq* polymerase (catalogue number- 11966018, Invitrogen) according to manufacturer's instructions. The thermal cycling conditions were set at 94 °C , 2 min initial denaturation, 40 cycles for (i) denaturation, 94 °C, 30 s, (ii) annealing, 56 °C, 30 s, (iii) extension, 72 °C, 40 s followed by final extension at 72 °C for 3-5 min. Agarose gel electrophoresis (1.5%) was carried out to confirm PCR product which were purified and sequenced for further analysis. The final plasmid map of entire construct was derived from the sequencing results.

# 5. OPTIMISATION OF FLOW CYTOMETRY PARAMETERS

## 5.1 OVERVIEW

Based on sequencing analysis, the *h*FKBP2 gene was successfully cloned into the pCTZQ plasmid, which will serve as an ideal positive control for further studies. Fluorescein-$PEG_{12}$-FK506 (kindly obtained from Dr. Andrew Piggott, Macquarie University), a fluorescently labelled probe (mol wt- 1839.20) was employed for further optimisation of experimental procedures. The FK506 probe should only bind to *h*FKBP2 protein that is expressed in fusion with Aga2 cell surface receptor protein and hence present on the surface of the cell and not yeast FKBP2, which is present within the cell and acts as a chaperone for endoplasmic reticulum (ER). So, the fluorescence data obtained from flow cytometric analysis should be specific to the interaction between *h*FKBP2 and FK506. FK506 and fluorescein molecules are connected to each other by $PEG_{12}$, a long hydrophilic linker which increases the conformational flexibility of FK506 as bulky fluorophores have previously shown to interfere with efficient protein binding and cell permeability[55]. Moreover, it also provides sufficient space for binding of larger proteins.

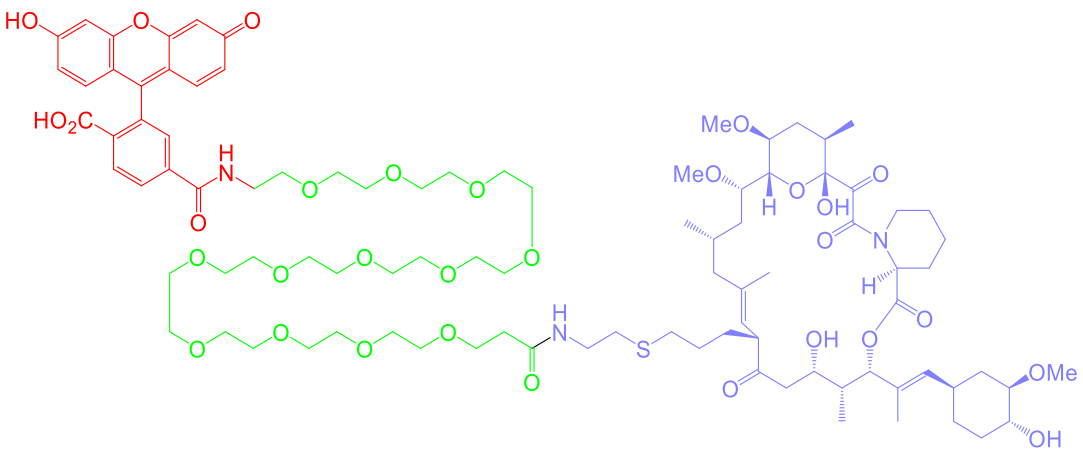


**Figure 15:** Structure of fluorescein-$PEG_{12}$-FK506 probe (mol wt- 1839.20). FK506 (blue) is linked to a fluorescent dye fluorescein (red) via a long, hydrophilic $PEG_{12}$ linker (green).

## 5.2 RESULTS AND DISCUSSION

### 5.2.1 Staining Profile

Both the pCTZQ and pCTZQ-FKBP2 containing *S. cerevisiae* EBY100 were incubated with fluorescein-$PEG_{12}$-FK506 probe with varying concentrations from 100 pM to 10 µM. 0.01% Tween-20 was not used at this stage. The aim of this experiment was to confirm if there was any increase in fluorescent intensity with increasing availability of probe that results in increased interaction between FK506 and *h*FKBP2 which gives the staining profile. The flow cytometric analysis confirmed that the staining was specific to the *h*FKBP2 that was displayed on the surface of yeast cell as there was considerably high fluorescence intensity obtained with *S. cerevisiae* EBY100 containing pCTZQ-FKBP2 (Figure 15). However, the amount of fluorescence obtained was very low as the blocking buffer consisted of only 1× PBS + 0.01% BSA and the incubation period was kept up to 3 h as used previously[52]. Moreover, as the experiment was carried out in triplicates (n=3), the fluorescence intensity was calculated as average of the data collected referred as Mean Fluorescence Intensity (MFI)). Also, the flow cytometric data were collected after 1 washing step which eliminates all non-specific staining and unwanted dye to study the staining profile.

Therefore, for further analysis fluorescence spectra using 1× PBS and 1× PBS + 0.01% Tween-20 was studied to study if there was any adsorption of dye with time on the polystyrene 96-well plates (all flow cytometric analysis were carried out on Grenier 96-well plates, flat bottom) and the effect of detergent on the fluorescent intensity. Tween-20 is a non-ionic detergent used for a variety of biochemical applications and as a blocking agent for immunoassays but at a very low concentration like 0.01% to prevent its negative effect.

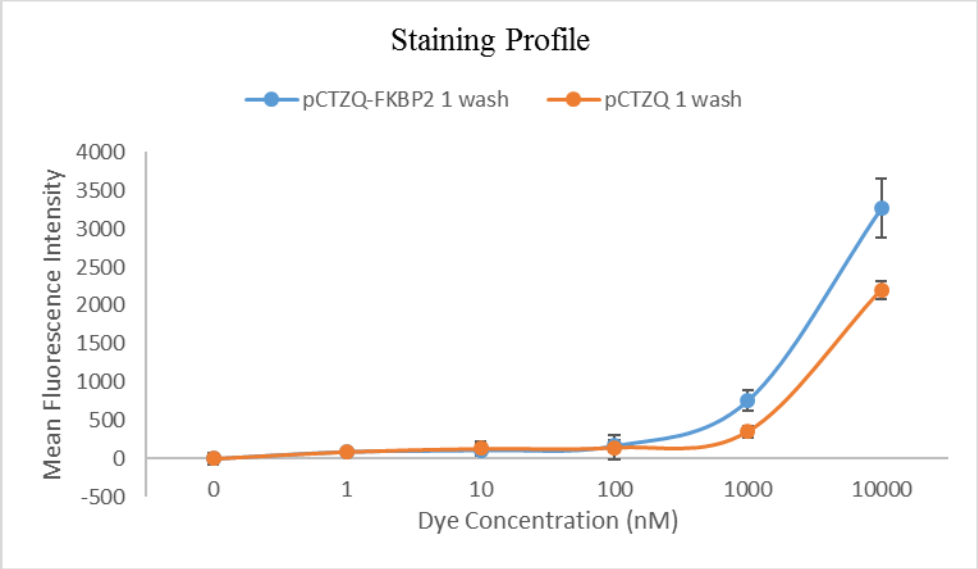
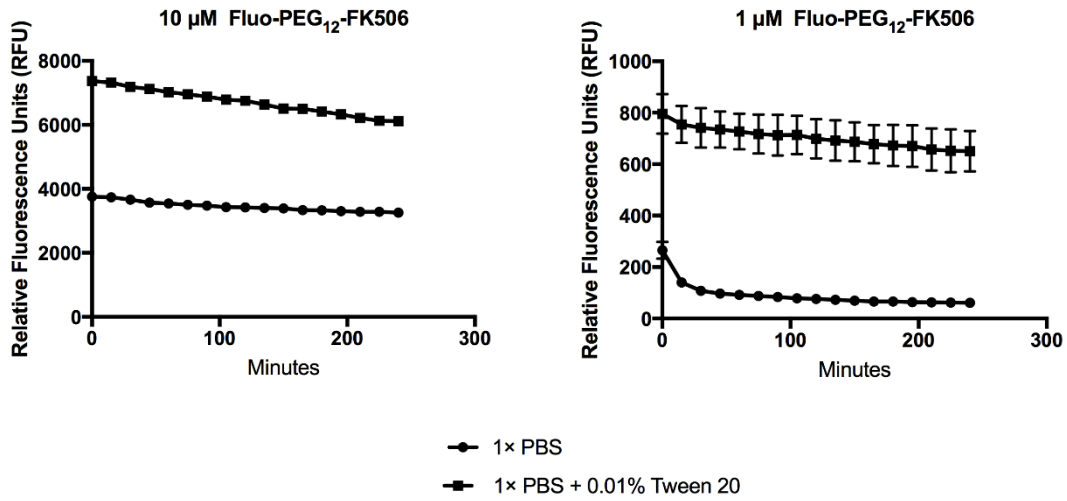



**Figure 16:** Staining profile of *S. cerevisiae* EBY100 containing pCTZQ and pCTZQ-FKBP2 at varying dye concentration (fluorescein-$PEG_{12}$-FK506) using flow cytometric analysis. More fluorescence intensity is obtained with yeast cells displaying *h*FKBP2 on the cell surface. The increase in fluorescence in yeast cells with empty plasmid was due probable non-specific staining due to high concentration of probe due to which dye might have entered inside the cells giving false positive.

### 5.2.2 Fluorescein-$PEG_{12}$-FK506 Fluorescence Intensity Studies:

The studies of fluorescence intensity was necessary to analyse if there was any decrease in fluorescence with time and concentration. Moreover, two types of buffer were used to see if use of detergent like 0.01% Tween-20 reduced the adsorption of dye on 96-well plates (BD-black clear plates, flat bottom). The fluorescence intensity was measured in terms of Relative Fluorescence Intensity (RFU) over 4 h time as previous YSD experiments used incubation time of 1-3 h using emission and excitation data of fluorescein.



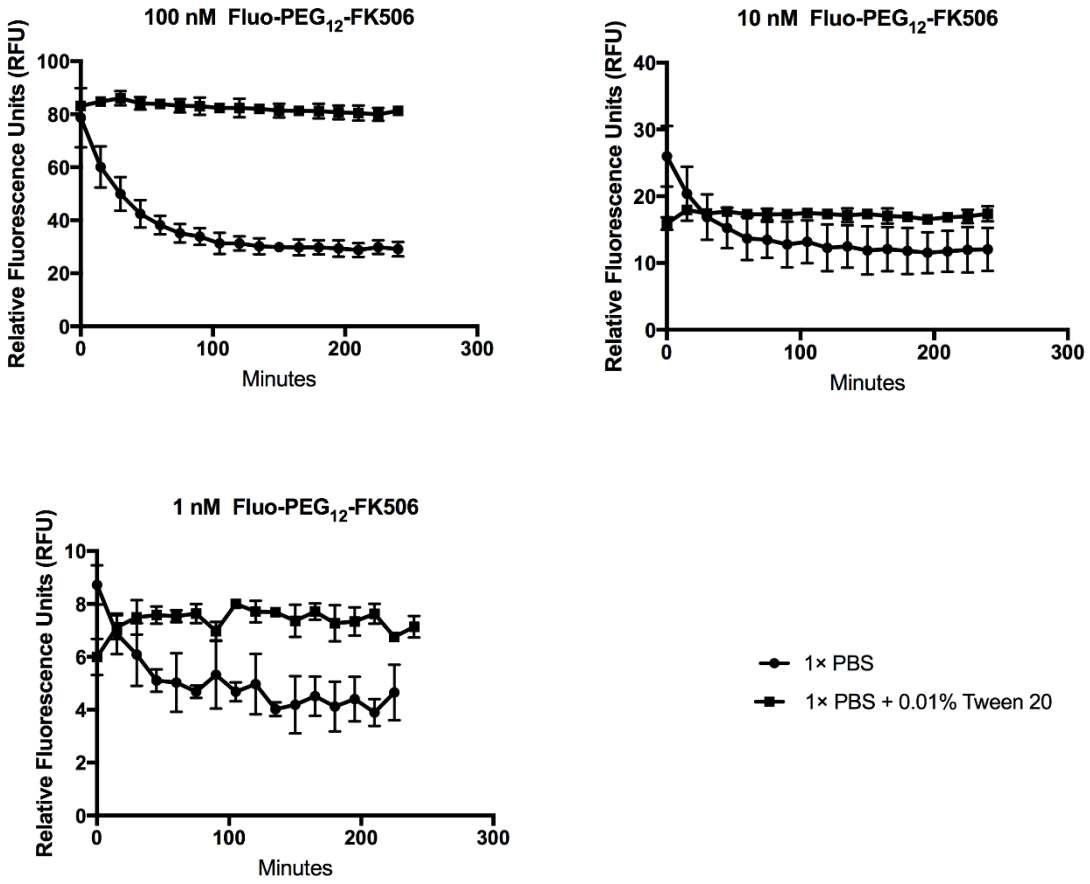



**Figure 17:** Comparative Fluorescence data of Fluorescein-$PEG_{12}$-FK506 in 1× PBS and 0.01% Tween-20 for 240 min (4 h) for concentrations ranging from 10 µM to 1 nM. Data collected at the interval of 15 min (17 cycles). The data confirms that with the addition of detergent like Tween-20, the amount of RFU increases nearly two-fold.

The data obtained confirmed that detergent like Tween-20 in minor concentration was required to stop the decrease in RFU with time which is due to the adsorption of dye on plate. There was also a possibility of dye sticking to the plate which may prevent the probe from binding to its target protein (in this case *h*FKBP2). Therefore, for all the flow cytometric experiments 10 µM concentration in 1× PBS + 0.01% Tween-20 was used where the decrease in fluorescence intensity was not really significant (Figure 17). Moreover, as positive control is a pure clone of YSD system expressing *h*FKBP2 on the cell surface, a higher concentration of probe is required than that for a library as ideally, all the cells would get stained.

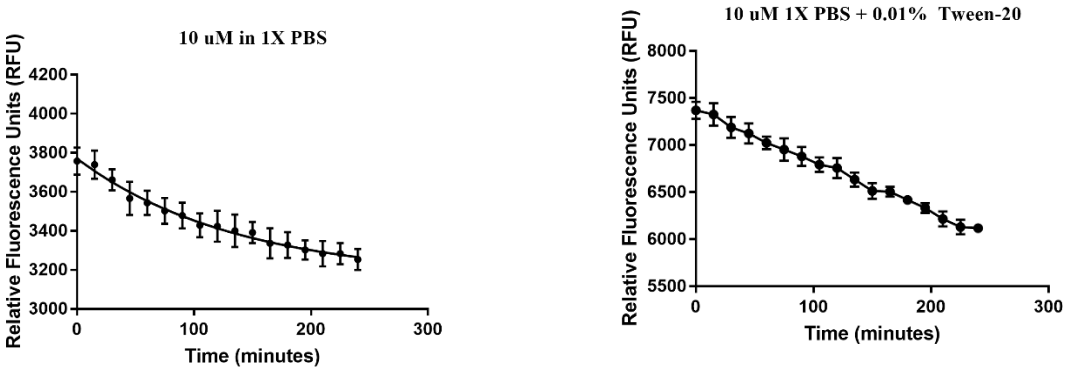



**Figure 18:** Fluorescence data of 10 μM Fluorescein-$PEG_{12}$-FK506 (excitation at 485 nm and emission at 515 nm in 1× PBS) and 0.01% Tween-20 which confirms that no significant decrease in fluorescence intensity with time is observed. So, this concentration was used for subsequent flow cytometric experiments.

### 5.2.3 Yeast Cell Concentration

The *S. cerevisiae* EBY100 cells containing pCTZQ-FKBP2 were induced at $OD_{600} \sim 1$, which corresponds to approximately $1 \times 10^7$ cells/mL and the induced yeast cells were taken for flow cytometric analysis aiming at a range of optimisation experiments by maintaining $OD_{600} \sim 1$ as at this concentration, the yeast cells are considered to be in exponentially growing phase. Moreover, plasmids are mostly expressed in the log phase of *S. cerevisiae*. A higher $OD_{600}$ before induction may decrease protein expression on the cell surface which may lead to experimental failure or lack of precision. The cells and all reagents were maintained on ice at all times (ideally 4 °C and even the washes were carried out at 4 °C during flow cytometric analysis as higher temperature may lead to loss of protein activity or protein getting inactive before the actual experiment[33].

### 5.2.4 Incubation Time Profile

As per the data obtained from spectra of RFU of Fluorescein-$PEG_{12}$-FK506, fluorescence intensity decreased with time at all concentrations (Figure 17 and Figure 18). However, it was necessary to determine the optimum time required for efficient binding between FK506-*h*FKBP2 to take place. So, another profile of incubation time over 4 h was carried out to optimise the time for efficient binding interaction (Figure 19). An incubation time of 90 min (1.5 h) was determined to be optimal incubation time as the fluorescence intensity initiates to decrease with time, however MFI remained almost constant from 1.5 h to 3 h which was the result of binding interaction.

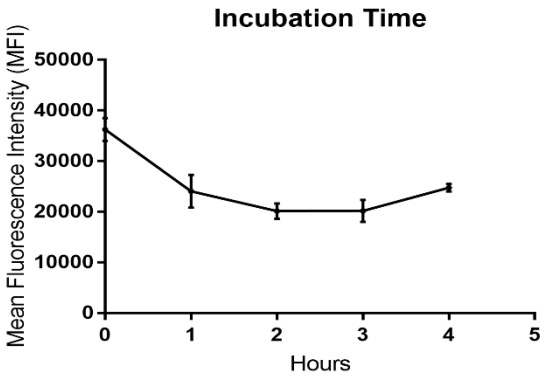



**Figure 19:** Incubation time profile of *S. cerevisiae* EBY100 containing pCTZQ-FKBP2 incubated with Fluorescein-$PEG_{12}$-FK506 from 0 h to 4 h. From the graph, 90 min (1.5 h) can be taken as an optimum incubation time for efficient binding to occur. 4 h data seems to be like an outlier as the MFI increased with time unlike fluorescence spectra obtained earlier (Figure 17 and Figure 18).

### 5.2.5 Incubation Temperature Profile

Incubation temperatures of 4 °C, 20 °C, 25 °C, 30 °C, 37 °C and 50 °C were investigated to select the best optimum temperature for efficient binding to take place. All the experiments were carried out in triplicates (n=3) and the MFI was the result of its average. The histogram (Figure 20) obtained from the temperature profile co-ordinated with data obtained previously with individual FKBP2-FK506 interaction carried out by differential scanning calorimetric (DSC) analysis[56]. According to both the data, the fluorescence intensity and hence the binding increases with increasing temperature. The thermal stability of binding interaction increased with increasing temperature due to binding between FK506 and *h*FKBP2 which results in high fluorescence and cell viability even at 50 °C. Moreover, the yeast cell-wall being tough due to presence of β-glucans makes it viable at high temperature of 50 °C that aids in strength of binding interactions. However, as yeast cells grow best at 30 °C, further all optimisation experiments were carried at 30 °C.

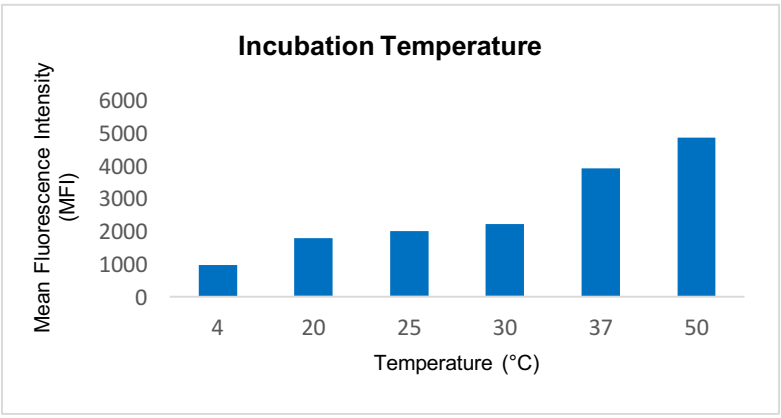



**Figure 20:** Incubation temperature profile for *S. cerevisiae* EBY100 containing pCTZQ+FKBP2. The fluorescence intensity and hence the binding interaction between FKBP2 and FK506 increases with increasing temperature within 90 min.

### 5.2.6 Number of Washes

Washes after incubation of YSD displaying *h*FKBP2 and fluorescent FK506 probe are essential to eliminate unbound probe and all non-specific bindings. All the experiments were carried out in triplicates (n=3) and the MFI was the result of its average. However, in positive control, as ideally all cells should be stained, the only purpose of washes was to eliminate unbound dye. So, the graph of fluorescence intensity with increasing number of washes was plotted and the data confirmed that after 3 washes even the binding interaction gets washed away as the washes are carried out by pelleting the cells by centrifugation at 1800 × *g* for 3 min and aspirating supernatant hence taking average of 5 min per wash. So, 2 washes were taken as optimum number of washes for further experiments.

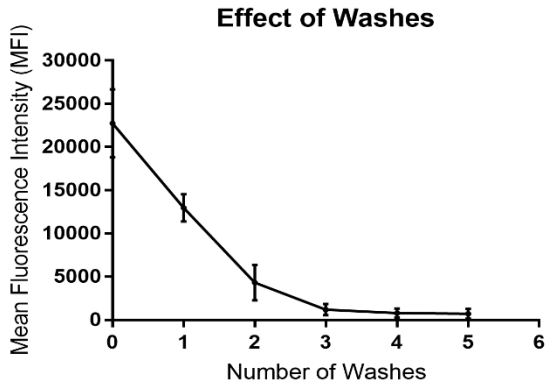



**Figure 21:** Graphical representation of effect of washes on binding interactions between *h*FKBP2 and FK506 which confirms that the unbound dye and non-specific interactions are washed away with 3 washes and hence 2 washes were carried out for further experiments.

### 5.2.7 Predicting Ratio of Desired YSD Cells Required by Mimicing Library Conditions

After optimising all the required conditions for experimental procedures, library like conditions were created where *S. cerevisiae* EBY100 containing empty plasmid were spiked with different ratios of *S. cerevisiae* EBY100 containing pCTZQ-FKBP2. The top 7-8% of positive staining cells were sorted via FACS in each of the sample followed by its enrichment on YPD + Zeo[50] plates and carrying out yeast colony PCR as well as agarose gel electrophoresis (1.5%).

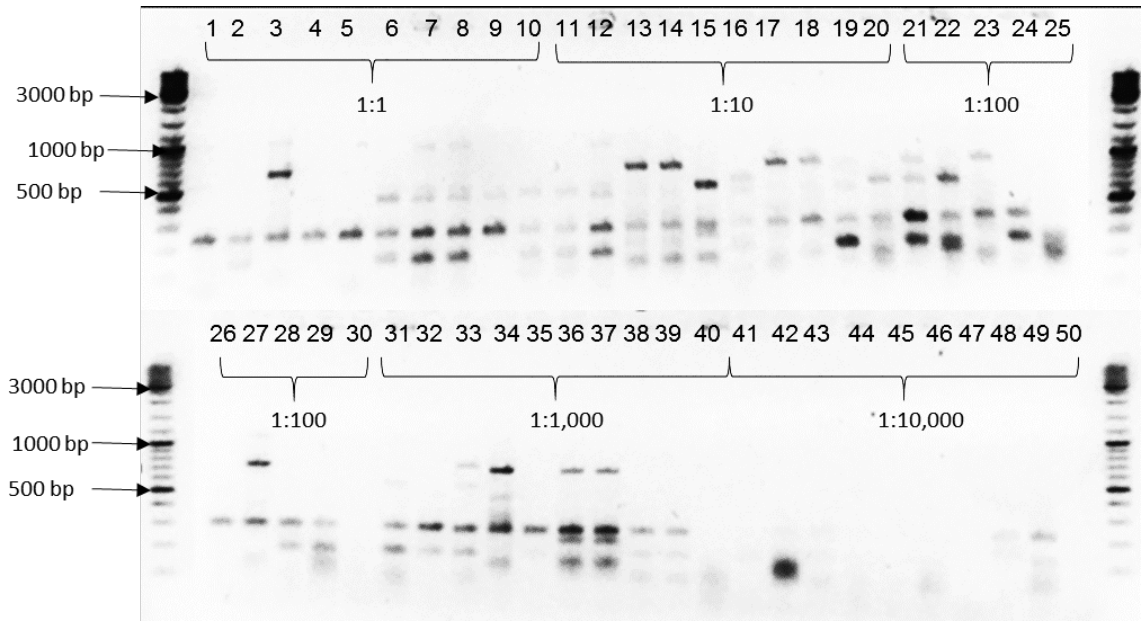



**Figure 22:** Agarose gel electrophoresis (1.5%) of enriched sorted cells of *S. cerevisiae* EBY100 containing pCTZQ+FKBP2 from *S. cerevisiae* EBY100 containing pCTZQ. Cells were stained with Fluorescein-$PEG_{12}$-FK506 probe for 90 min and subjected to FACS. After first round of selection by sorting 7-8% of top positive staining population using BD-Influx II Flow Cytometer, the population was mostly enriched with *S. cerevisiae* EBY100 containing pCTZQ-FKBP2 (250 bp product) up to 1:1,000 ratio but not with 1:10,000 ratio. The 500 bp and 750 bp product seems to be the dimer of trimer of *h*FKBP2 gene which would have been multiple cloning in the ligation step (Section 2.1.1) of construction of positive control.

The results confirmed that a 250 bp product was obtained after FACS with samples 1:1, 1:10 and 1:100 ratio of *S. cerevisiae* EBY100 containing pCTZQ+FKBP2. While, hardly any bands were obtained with 1:10,000 ratio sample. YSD cells with desired cDNA insert can be sorted via FACS with maximum of 1:1,000 ratio of desired YSD cells to library. Therefore, when FACS is used with original library, there should be at least 1 cell out of 1,000 cells that display desired cDNA insert on yeast cell wall. However, if stronger binding protein target is present in low amounts, more round of selection are required to enrich the cells with desired ratio of at least 1:1,000. So, this data was used to validate the Q-fly library using Fluorescein-$PEG_{12}$-FK506 probe.

## 5.3 EXPERIMENTAL PROCEDURES

### 5.3.1 Fluorescence Spectra Analysis

Fluorescence spectra of Fluorescein as well as Fluorescein-$PEG_{12}$-FK506 probe was analysed using Spectromax M5 instrument. Two kinds to buffers were used to see if there was an adsorption on a 96-well plate (i) 1× PBS and (ii) 1× PBS + 0.01% Tween-20. Fluorescence spectra was studied for 4 h (240 min) with excitation at 485 nm and emission at 515 nm which

is normally used for fluorescein dye. The experiment was carried out in triplicates (n=3) and graphs were plotted using Graphpad Prism.

### 5.3.2 Expansion and Induction of Yeast Cell Cultures

*S. cerevisiae* EBY100 with pCTZQ and *S. cerevisiae* EBY100 with pCTZQ-FKBP2 were freshly grown by aliquoting (10 µL) from −80 °C frozen stock in YPD + Zeo$^{50}$ liquid media (YPD liquid media for wild type. Both the inoculated media were incubated at 30 °C for 16-18 h such that $OD_{600}$ ~ 1. Both the strains of *S. cerevisiae* EBY100 were then induced in galactose media SG-CAA by repeated washing at 1800 × *g* for 3 min up to 3 washes. The yeast cultures were incubated in SG-CAA media at final $OD_{600}$ ~ 1 at 25 °C for 48 h shaking at 220 rpm. These induced cultures were used for flow cytometric analysis.

### 5.3.3 Staining Cultures for Flow Cytometry

An aliquot (200 µL) of both wild type and induced pCTZQ-FKBP2 containing *S. cerevisiae* EBY100, both with $OD_{600}$ ~ 1 were pelleted at 1800 × *g* for 3 min at 4 °C[30] and washed twice with blocking buffer (1× PBS + 0.01% BSA + 0.01% Tween-20). The washed pellet was incubated with Fluorescein-$PEG_{12}$-FK506 (200 µL) with variable dye concentrations (100 pM to 10 µM), variable incubation time (1 h to 4 h) and variable temperature (4 °C, 20 °C, 25 °C, 30 °C, 37 °C and 50 °C). Care was taken the cells were in suspension at all times using Intelli-Mixer RM-2M, POCD Scientific. Following incubation, cells were washed (0-5 washes) with blocking buffer and the washed pellet was suspended in blocking buffer (400 µL). This cells were read using flow cytometer (Cytoflex, Beckman Coulter) using FITC filter which is competent with fluorescein.

### 5.3.4 FACS and Enrichment of Sorted Cells

*S. cerevisiae* EBY100 containing pCTZQ + FKBP2 cells were mixed with *S. cerevisiae* EBY100 containing only pCTZQ in the ratio of 1:1, 1:10, 1:100, 1:1,000 and 1:10,000 and incubated with 10 µM Fluorescein-$PEG_{12}$-FK506 dye (250 µL) followed by incubation of 1.5 h (90 min) at 30 °C and 2 washes with blocking buffer. The cells were suspended in blocking buffer (1 mL) and $10^5$ cells top positive 7-8% population from each sample were sorted via FACS using BD-Influx II Flow Cytometer. An aliquot of these sorted cells (10 µL) was plated on SD-CAA + Zeo$^{50}$ plates and remaining was enriched in SD-CAA + Zeo$^{50}$ liquid media. Yeast colony PCR was carried out (Section 2.2.4) using FKBP2 forward and reverse primers was carried out to see if there is any enrichment after FACS and with which maximum ratio, enrichment starts and this was followed by agarose gel electrophoresis (1.5%).

# 6. VALIDATION OF *Bactrocera tryoni* cDNA LIBRARY

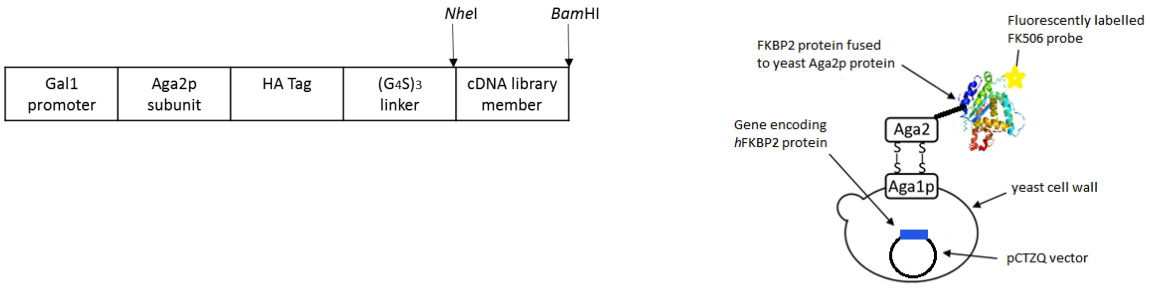
## 6.1 OVERVIEW

### 6.1.1 *B. tryoni* (Q-fly) as a model organism

As the overarching aim of developing YSD is to accelerate anti-parasitic drug discovery, *Bactrocera. tryoni* was selected as a model organism for establishing a proof-of-concept. *B. tryoni*, a fruit fly commonly referred as Queensland fly (Q-fly) is one of the most damaging horticultural insect pest in Australia[57]. Moreover, these fruit-flies are ideally suited as a model organism due to its (i) fully sequenced small genome (ii) ease of handling and (iii) short life cycle. Besides this, the genomes of eukaryotic organisms such as *D. melanogaster, C. elegans, A. thaliana, S. pombe and S. cerevisiae* encode different numbers of the FKBP paralogues, some of which are orthologues to the *h*FKBP2 described previously (Section 2.1.1). So, construction of cDNA library of Q-fly and validating with YSD technology in combination with FACS exploiting known protein-target interactions between FKBP2 and FK506 can provide very useful information to identify unknown cellular target for natural products thus expanding the area of anti-parasitic drug discovery.

### 6.1.2 Construction and Validation of cDNA library

A cDNA library of Q-fly cloned in YSD vector pCTZQ between *Nhe*I and *Bam*HI as a fusion protein with yeast Aga2p subunit (Figure 23 B) and displayed on the surface of *S. cerevisiae* EBY100 was kindly provided by Dr. Jenny Vo (unpublished data, Macquarie University). Three variants of pCTZQ were constructed (i) pCTZQ (ii) pCTZQ+1and (iii) pCTZQ–1 assuming that one of the variants will form in-frame constructs of the cDNA library[38] and transformed in *S. cerevisiae* EBY100 using infusion cloning method. This Q-fly cDNA library was validated with the same fluorescently labelled FK506 probe (as used in chapter-2) where the aim was to pull-out and enrich yeast cells expressing FKBPs of Q-fly (Figure 23 B)




**[A]** **[B]**

**Figure 23:** (A) cDNA library cloned in pCTZQ vector between *Nhe*I and *Bam*HI vector arms in fusion with yeast Aga2 cell surface receptor proteins. (B) Schematic representation of yeast cell displaying FKBP2 protein form Q-fly which emits fluorescence on labelling with fluorescent FK506 probe that can further be separated by FACS.

## 6.2 RESULTS AND DISCUSSION

### 6.2.1 Titering Q-fly Library

The YSD Q-fly libraries were titered to estimate the library diversity. Various dilutions were spread on SD-CAA plates and the average library diversity of $1.52 \times 10^7$ CFU/mL was obtained. Moreover, wild type *S. cerevisiae* was spread on SD-CAA plates as a negative control to test for the possibility avoid of contamination. This library diversity was well-suited for FACS selection as FACS is amenable to library diversities of $1 \times 10^7$ or less. For higher library diversities, magnetic bead selections are recommended before FACS to decrease library size or the volume of reagents, the number of washes also has to be increased accordingly for scale-up[30].

### 6.2.2 Induction of Q-fly Library and Preparing for FACS

Q-fly library was cloned in pCTZQ vector where each member of the library was in-frame with Aga2p gene. The translation of this gene into a fusion protein is regulated by a galactose promoter, which requires the presence of galactose as a carbon source. For induction purpose, the yeast cells were expanded freshly before use (within one week) to ensure a higher number of plasmids in cell. Moreover, for FACS analysis, the yeast cells were maintained in exponential phase ($OD_{600}$=1 = $1 \times 10^7$ cells/ mL). Cell overgrowth ($OD_{600}$ greater than 3 before induction) reduces protein expression as well as surface display efficiency. Additionally, $OD_{600}$ of induced cells was kept below 1 but higher than 0.5 to obtain optimum growth and protein expression after inductions[30]. The incubation temperature was kept lower (20 °C) to maintain stability of the displayed proteins. Besides this, during staining, cells were kept in suspension

by continuous mild shaking and inversion on a rotating wheel to ensure equal probability of staining for each cells[14].

### 6.2.3 First Round of Selection

The initial library of *B. tryoni* was induced, stained with fluorescently labelled with FK506 probe (fluorescent dye- Fluorescein) and subjected to FACS analysis. The FACS parameters on instrument were set in accordance to fluorescein isothiocyanate (FITC) as the FK506 probe was fluorescently labelled with fluorescein and FITC is the analogue of fluorescein with wide applications in flow cytometry. In terms of validation, our aim was to validate the library with a fluorescent to identify FK506 probe Q-fly FKBPs that are displayed as fusion proteins on the yeast surface. The probability of each yeast cell expressing Q-fly FKBP2 as a fusion protein is very low. Therefore, a lower concentration (100nM) of fluorescently labelled FK506 probe was used, unlike 10 µM in optimisation experiments, and top 8-10% of cells were sorted in the first round of selection. Typically, the initial number of cells taken for FACS analysis was ten times the library diversity to ensure the identification of desired yeast clones that may be present in low amounts. Sorting 8-10% of the most fluorescent cells will further reduce the probability of missing desired yeast cells even if they are present in low amounts.

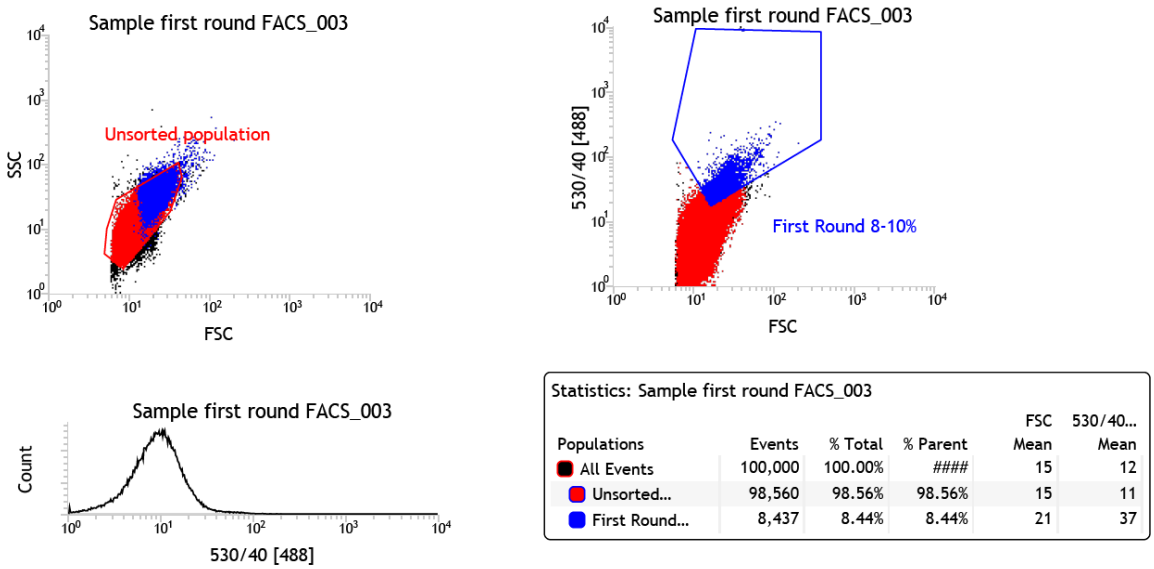



**Figure 24:** Statistics and sorting parameters of first round of selection of Q-fly library where the entire population of yeast cells were gated as unsorted population (red) and the 8-10% sorted cells were gated as First Round Cells (blue). The histogram showed where most of the population was lying. Polygonal gating was used for this purpose as a starting point for sorting.

### 6.2.4 Second Round of Selection

The sorted population from first round was enriched in SD-CAA + Zeo$^{50}$ liquid media and again subjected to second round of FACS (Section 4.3.3). The 7-9% most fluorescent cells were again sorted in second round.

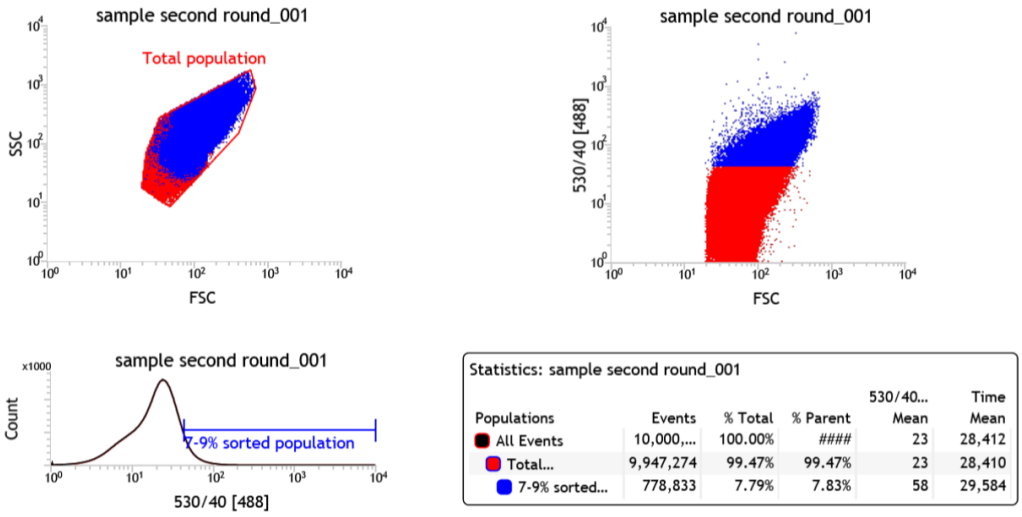



**Figure 25:** Statistics and Sorting parameters of second round of selection of previously sorted library (Figure 24) where the entire population of yeast cells were gated as unsorted population (red) and the 7-9% of most fluorescent cells were gated and sorted (blue). The histogram showed where most of the population was lying. Line segment gate was used for sorting most fluorescent cells as this gate enables precise sorting of cells from desired area.

The histogram of fluorescence at 530/540[488] vs count (Figure 25) described where most of the population was lying for both the rounds of selection and clearly demonstrated a population shift if any in the direction of increasing fluorescence from the first to the second round of selection. Both the histogram data from first and second round of selection were overlaid to study the population shift (Figure 26). The overlay confirmed that the number of topmost positive staining cells increased with second round. The sorted cells from the second round of selection were expanded and spread on SD-CAA + Zeo$^{50}$ plates for DNA profiling and diversity study of the library.

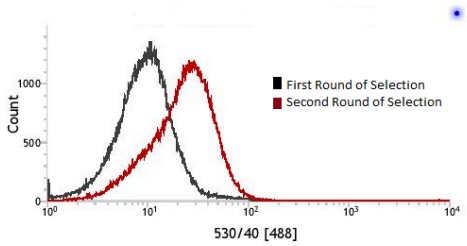



**Figure 26:** Overlay histogram of fluorescence shift between first and second round of selection. The result confirms that the number of cells that are highly fluorescent increased in second round of selection which were further analysed.

### 6.2.5 DNA Profiling from Second Round of Selection

After two rounds of selection, yeast colony PCR was carried out (Section 4.3.5) to understand the pattern of library and find out if there was any enrichment of the yeast cells displaying FKBPs from the cDNA library.

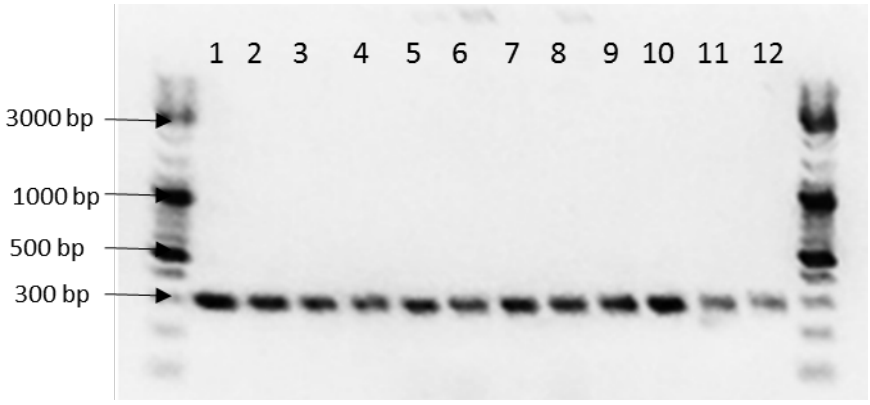



**Figure 27:** Gel image of agarose gel electrophoresis (1.5%) of yeast colony PCR for 12 random clones from second round of selection. All the bands from each clone are converged approximately 300 bp.

Agarose gel electrophoresis (1.5%) (Figure 27) clearly confirmed that all the sorted cells were clones from either same population or the plasmid contained no insert from cDNA library. DNA fingerprinting using *Hin*f digestion could not be carried out as the size of PCR product was too small (~300 bp) for the study. So, sequencing was carried to confirm the identity of obtained PCR product and data interpretation.

### 6.2.6 Sequencing Analysis

As shown in Figure 23, Q-fly cDNA library was cloned in between the *Nhe*1 and *Bam*H1 restriction sites. However, from the sequencing data, it was confirmed that the PCR products obtained after the two rounds of selection with fluorescent FK506 did not contain any insert in between the preferred restriction sites, which further confirmed that the plasmid pCTZQ did not contain any cDNA insert in either of the random yeast clones selected. So, Q-fly cDNA library could not be used for any further analysis.

### 6.2.7 Conclusion and Future Directions

Based on sequencing analysis, it was confirmed *S.cerevisiae* EBY100 cells did not contain Q-fly library which means there was no cDNA insert between desired restriction sites. The high fluorescence obtained during FACS could be possibly due to high probe concentration containing fluorescein that was unable to bind to its protein target due to lack of any cDNA insert due to which the dye would have entered the cell resulting in fluorescence detected in

FACS. So, it is recommended to check the diversity of original cDNA library that was constructed.

However, the library is present in pSMART vector provided in the kit used for library construction (Clonetech, Catalogue Number- 634933). So, primers needs to be designed as per the sequence of this vector to check the library diversity. If the diversity is as per desired, then whole cDNA library has to be cloned in YSD vector pCTZQ between *Nhe*1 and *Bam*H1 restriction sites followed by transformation in *E. coli*. Before, final transformation in *S. cerevisiae* EBY100 to get the final YSD Q-fly library, it is recommended to check the library diversity which also estimates the transformation efficiency. In case of library diversity being not desirable, then the process of library construction has to be repeated from total RNA isolation. More resources and information is required to proceed of infusion cloning used to prepare this library. However, due to time constraint, this task could not be continued.

## 6.3 EXPERIMENTAL PROCEDURES

### 6.3.1 Growth, Media and Titering Libraries

*S. cerevisiae* EBY100 cells containing Q-fly library were stored at −80 °C and an aliquot of these cells (10-50 µL) was thawed and expanded in SD-CAA + Zeo$^{50}$ liquid media (3 mL) for 18-24 h at 30 °C and shaking at 220 rpm. To test the frozen stock and titer libraries to calculate library diversity, different dilutions (thousand fold and ten thousand fold) of the frozen cells were plated on 150 mM SD-CAA plates and incubated at 30 °C for 48-72 h. Wild type *S. cerevisiae* EBY100 was used as negative control and spread on SD-CAA + Zeo$^{50}$ plates. Colony Forming Units (CFUs/mL) provided an estimate of the library diversity before proceeding for induction.

### 6.3.2 Induction of Expanded Q-fly Library

The freshly expanded library ($OD_{600}$=1; $1 \times 10^7$ cells/mL) were inducted in SG-CAA liquid media (composition described previously[29]) as mentioned earlier (Section- 3.3.2). Typically yeast cells were induced to a final concentration of $1 \times 10^7$ cells/ mL by repeated washes (2-3) at 1800 × *g* for 3 min at 4 °C with SG-CAA media. Finally, SG-CAA media was added to washed pellet such that $OD_{600}$~1 and incubated at 25 °C for 48 h and shaking at 220 rpm**.**

### 6.3.3 Staining and FACS

$OD_{600}$ of the cells was checked post-induction to determine the concentration of induced yeast cells. The volume of the sample required for FACS (10 mL) was determined such that it covered ten times the diversity of the library. The cells were pelleted out by centrifugation at 1800 × *g* for 3 min at 4 °C and washed twice with blocking buffer (1× PBS, pH 7.4 +0.01% BSA). The pellet was dissolved in 100 nM fluorescein-$PEG_{12}$-FK506 probe solution (1 mL of 1× PBS, excitation- 485 nm, emission- 520 nm, molecular weight- 1839.20 g/mol) and incubated for 1.5 h at 30 °C with constant rotation and inversion to avoid pellet formation. After incubation, cells were washed twice with blocking buffer by centrifugation at 1800 × *g* for 3 min at 4 °C and final pellet was dissolved in blocking buffer (10 mL) followed by FACS analysis (BD Influx II).

### 6.3.4 Enrichment of Sorted Cells

The sorted cells (approximately $10^8$ cells in each round) were centrifuged at 4000 × *g* for 1 min and expanded in SD-CAA media (10 mL) and again induced with SG-CAA media for the next round of selection (as described in Section 4.3.2). These cells were stained and subjected to the second round of selection with FACS (Section 4.3.3) and sorted cells were expanded in SD-CAA media. An aliquot was also plated on SD-CAA plates for yeast colony PCR and incubated at 30 °C for 2-3 days.

### 6.3.5 Yeast Colony PCR and Sequencing

Yeast colony PCR was carried out as described in Section 2.3.6 with few modifications. pCTZQ-F (5' AGTAACGTTTGTCAGTAATTGC 3' - length 22, Tm 57.5 °C and pCTZQ-R (5' GTCGATTTTGTTACATCTACAC 3'- length 22, Tm 55.3 °C) primers which covers Aga2p gene from the 5' end and foreign gene from the 3' end. The thermal cycling conditions were set to initial denaturation at 94 °C, 30 cycles of (i) denaturation at 94 °C, 30 s, (ii) annealing at 56 °C, 30 s, (iii) extension at 72 °C, 30 s followed by final extension at 72 °C for 3 min. Agarose gel electrophoresis (1.5%) was carried out for observing library diversity and desired PCR products were sent for sequencing for further data analysis.

# 7. CONSTRUCTION OF *Caenorhabditis elegans* CDNA LIBRARY

## 7.1 OVERVIEW

*C. elegans*, a soil nematode has been used as a powerful tool and model organism in indrug discovery and target identification for more than a decade. Moreover, *C. elegans* is easy

to cultivate under laboratory conditions, is non-pathogenic, has a short life cycle and its biology as well as its genome have been thoroughly studied. Besides being transparent, this hermaphrodite worm is 1 mm long and 80 µm wide in adult stage making it easy to handle.

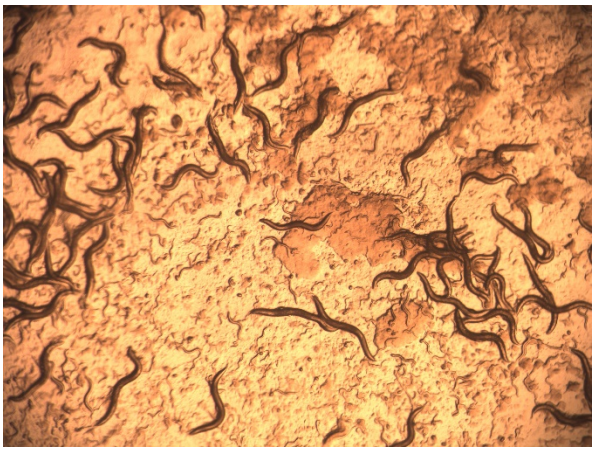
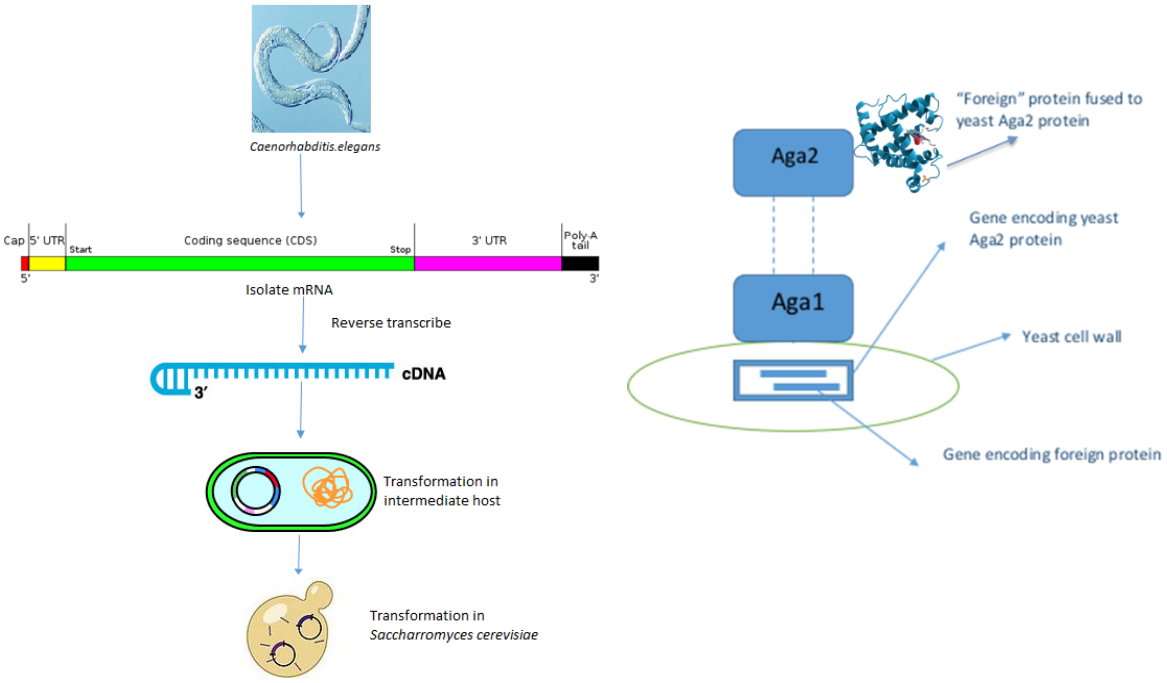


**Figure 28:** Mixed population of *C. elegans* sampled for the construction of cDNA library (kindly obtained from Microbial Screening Technologies, Sydney, Australia).

Previous studies have revealed highly conserved homology in cellular pathways between nematodes and mammals. Moreover, the mutations can be easily identified and studied[58]. A large number of parasitic disease is caused by worms (Section 1.2) which are sometimes difficult to treat. However, due to limited knowledge of pharmacology and mode of actions of some potentially bioactive natural products, they cannot be effectively developed into a drug[3]. *C. elegans* therefore can be used as a model organism to decipher the biochemical pathway and modes of action of antiparasitic natural products. So, this section describing the construction of a cDNA library of *C. elegans* followed by displaying this library on the surface of *S. cerevisiae* EBY100 to yield a *C. elegans* yeast surface display library that covers the entire transcriptome of *C. elegans.*




**[A]** **[B]**

**Figure 29:** (A) Overall steps involved in construction of YSD cDNA library of *C. elegans*. The cDNA library was stored in the intermediate host *E. coli* before final transformation in *S. cerevisiae EBY100*. (B) Schematic representation of YSD cDNA library of *C. elegans* where a cDNA library member is expressed in fusion with Aga2p subunit.

## 7.2 RESULTS AND DISCUSSION

### 7.2.1 Total RNA Isolation and mRNA Purification

Total RNA isolation from *C. elegans* is challenging particularly due to its tough cuticle, which necessitates use of a tissue homogeniser as a cuticle disruption method to release the contents of cell and additional β-mercaptoethanol to irreversibly denature RNases by reduction of disulfide bonds besides destroying the native conformation required for optimum enzyme function. Moreover, a large mixed population of *C. elegans* is required for efficient yield of total RNA (7.41 µg/mL)[59]. As the aim was to obtain cDNA which is only a transcriptome of an organism, mRNA was purified using OligodT beads present in the Oligotex suspension provided in the kit to eliminate other types of RNA which may possibly interfere with first strand cDNA synthesis reducing the yield. *C. elegans* being eukaryotic, has polyadenylation at the 3'- end of its mRNA, which was exploited for first strand cDNA synthesis.

### 7.2.2 Primer Design

The 5'- end of cDNA library of *C. elegans* must be cloned in-frame with the 3'- end of the Aga2p subunit in the pCTZQ plasmid, which would then be translated as a fusion protein

on the surface of a cell. So, the 5'- end of cDNA library must not contain untranslated regions (5'-UTR) as this may contain pre-mature stop codons resulting in premature termination of translation. Kozak sequences, more commonly known as Kozak consensus sequences are known to be a sites recognised by ribosomes as translational start sites on mRNA[60]. These sequences are highly conserved in mammals and was used to design primer for first strand cDNA synthesis. However, as *C. elegans* have low %(G+C) content of ~36%, its Kozak sequence and was different to that from vertebrate consensus. This strategy was aimed at removing almost 90% of 5'-UTR[61]. Moreover, an additional primer consisting of invertebrate Kozak consensus[62] (Section- 5.3.2) was designed to cover all possible diversity. Both these Kozak sequences were not as conserved as those in vertebrates leading to the use of equal proportion of some most probable bases on each place.

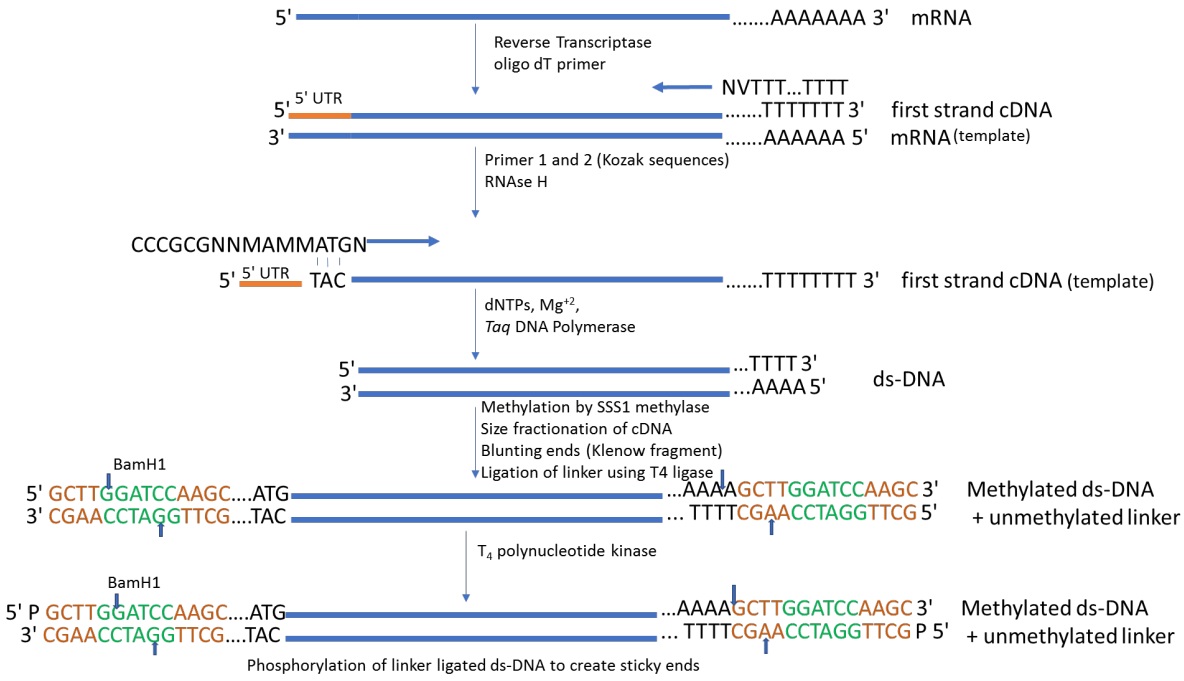



**Figure 30:** Schematic representation of overall strategy for construction of YSD cDNA library of *C. elegans*.

### 7.2.3 First Strand cDNA Synthesis

An indication of good library is that it should be large enough to be a representative of transcriptome of an organism and should contain all desired sequences. Moreover, the cDNA inserts should be full length copies of original mRNA containing all relevant information required for gene identification and expression. Construction of cDNA from mRNA provides additional advantage of obtaining low copy number DNA transcripts as mRNA is used as a template. Superscript IV Reverse Transcriptase is an enzyme with increased thermal stability

which is derived from commonly used moloney murine leukaemia virus (MMLV) Reverse Transcriptase which produces cDNA from mRNA transcripts in 3'-5' direction. The poly-A tailing of eukaryotic mRNA was primed with oligod$T_{20}$ primer and Kozak sequence were used from 5'-end to eliminate 5'-UTR in initial stages. RNase H was employed after completion of first strand cDNA synthesis as contamination from RNases may damage the quality of cDNA as the mRNA template will be degraded leading to shorter cDNA product. This obtained cDNA product with $A_{260}/A_{280} = 1.9$ (1.8 represents pure DNA) was then subjected to different methods of RT-PCR.

### 7.2.4 Optimisation of RT-PCR to Obtain ds-cDNA Library

ds-cDNA synthesis was carried out by RT-PCR using Platinum *Taq* Polymerase. The initial PCR and reaction was tried with annealing temperature of 52 °C and visualised using agarose gel electrophoresis (1.5%) which showed no bands (Figure 31).

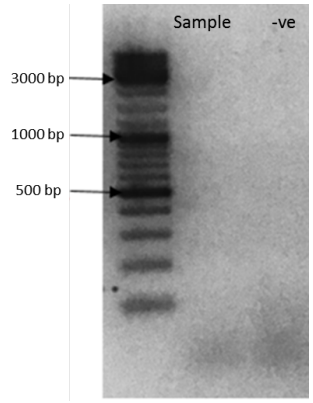



**Figure 31:** RT-PCR at annealing temperature of 52 °C. As no bands were obtained with this temperature, gradient-PCR was performed. -ve control is the PCR reaction without template.

As the primers specific to Kozak sequence were not highly conserved, temperature gradient PCR was carried out with annealing temperatures from 39-49 °C to obtain optimum annealing conditions. The results from agarose gel electrophoresis (1.5%) showed no ds-DNA synthesised in any temperature range (Figure 32). This suggested that some procedures with RNA isolation and first strand cDNA synthesis needs to be further improvised and optimised. However, time constraint prevented this from being preferred in this project.

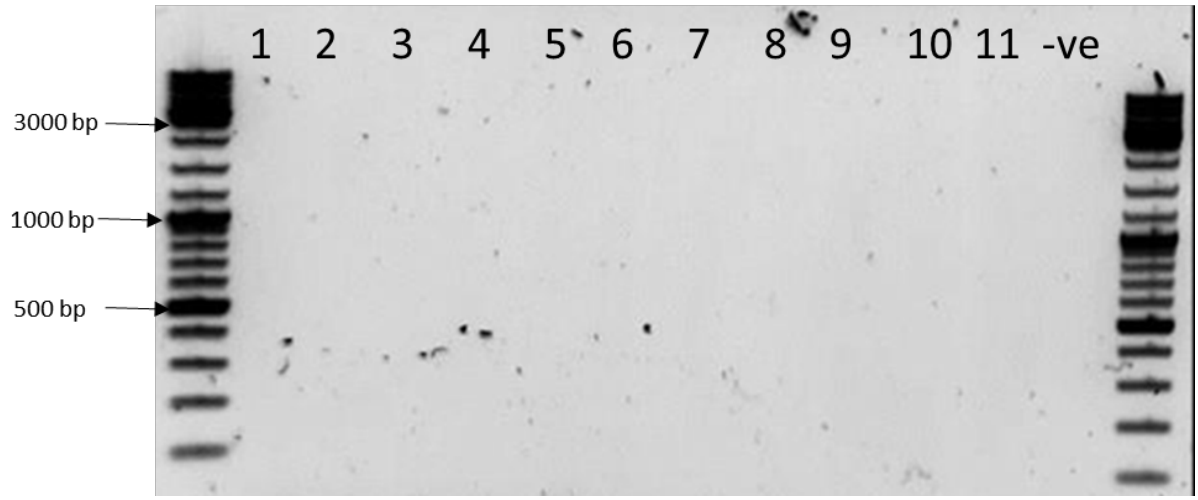



**Figure 32**: Temperature gradient RT-PCR for second strand cDNA synthesis with annealing temperature from 39 °C to 49 °C (wells 1-11). As no bands were obtained in any of these conditions, more optimisations are needed for the success of this step. Negative control is the PCR reaction without any template.

### 7.2.5 Future Directions

Some more insights, primer design and analysis is required for the construction of cDNA library without 5'-UTR and a library that covers entire transcriptome of *C. elegans* at all of its developmental stages. Many procedures for synthesizing cDNA are known which mostly aims at maximising the amount of cDNA from a limited amount of mRNA. However mostly, the completeness of the cDNA synthesis is variable as well as unpredictable. This variation depends mainly on mRNA isolation, first strand cDNA synthesis, and second strand cDNA synthesis. So, conditions need to be optimised by going back to total RNA isolation to can obtain optimum library diversity. This will be followed by linking methylated cDNA library to a bifunctional linker[63] which provides both *Bam*H1 restriction site at 5'-end and *Hind*III site at 3'-end on linking to poly-A site. cDNA library will need to be methylated to protect it from digestion by restriction enzymes.

## 7.3 EXPERIMENTAL PROCEDURES

### 7.3.1 Nucleic acid Isolation and Purification

Mixed population of *C. elegans* was stored in RNAlater stabilisation solution (Thermofisher Scientific, Catalogue Number- AM7020). Total RNA was isolated from mixed population of *C. elegans* (30 mg) using RNAeasy mini kit (Qiagen, Catalogue number- 74104) according to manufacturer's protocol with few modifications. The tissues were disrupted and homogenised in RLT buffer provided in the kit with additional β-mercaptoethanol (10 µL). The

eluted total RNA was quantified using Qubit fluorometer and subjected to mRNA purification using Oligotex mRNA mini kit (Qiagen, Catalogue number- 70022) according to manufacturer's instructions. This isolated mRNA was then employed for first strand cDNA synthesis using Reverse Transcriptase.

### 7.3.2 First Strand cDNA Synthesis

First strand cDNA synthesis was carried out using Superscript IV Reverse Transcriptase enzyme (Invitrogen, Catalogue Number- 18090010) according to manufacturer's instructions. The reaction was set up using 2 primers specific to Kozak sequences of *C. elegans* (5' BVVBVVBVNNMAMMATGN 3'- Tm 53.5 °C) and invertebrates (5' BVVBVVBVAAATMAACATGRC 3'- Tm 53.5 °C) as forward primers and Oligo $dT_{20}$ primer as reverse primer where, N= equal proportion of A,T,G or C, V= equal proportion of A,G or C, M= equal proportion of A and C, B= equal proportion of C, G or T and R= equal proportion of A or G. After the completion of reaction, the reaction was inactivated by incubation at 80 °C for 10 min. Moreover, the residual RNA was removed by treating the reaction with RNAse H (1 µL) and incubating it at 37 °C for 230 min. ds-cDNA synthesis was carried out using Reverse Transcription PCR (RT-PCR) immediately after this step. The remaining template was stored at –20 °C indefinitely.

### 7.3.3 ds-cDNA Synthesis by RT-PCR

ds-CDNA synthesis was carried using same Kozak primers as forward primers and oligo $dA_{20}$ as reverse primers using Platinum *Taq* Polymerase and thermal cycling conditions of initial denaturation at 94 °C for 3 min, 35 cycles of (i) denaturation at 94 °C for 45 s, (ii) annealing at 52 °C for 45 s and (iii) extension at 72 °C for 45 s followed by final extension at 72 °C for 6 min. Moreover, an additional temperature gradient PCR was carried out with Dream *Taq* Polymerase to obtain optimum annealing temperature. The thermal cycling conditions used here were initial denaturation at 95 °C for 3 min, 40 cycles of (i) denaturation at 95 °C for 30 s, (ii) temperature gradient of 39-49 °C for 30 s and (iii) extension at 72 °C for 1 min followed by final extension at 72 °C for 3 min. The results were analysed with agarose gel electrophoresis (1.5%) for further processing.

# 8. PROJECT OUTPUTS AND FUTURE DIRECTONS

## 8.1 SUMMARY

The project was aimed at developing YSD to identify cellular targets of natural products for advancement of anti-parasitic drug discovery. For this purpose, *h*FKBP2 was cloned in YSD vector pCTZQ and displayed on the surface of *S. cerevisiae* EBY100 as a *C*- terminal fusion protein with Aga2p cell surface receptor protein. This served as an ideal positive control for the optimisation of experimental procedures and parameters as YSD was not previously used for identification of cellular targets of natural products.

For the optimisation experiments, only a single clone of *S. cerevisiae* EBY100 containing pCTZQ-FKBP2 was used. The optimisation experiments produced valuable information regarding the experimental conditions optimum for effective binding of *h*FKBP2 and FK506. Based on data obtained, all the further experiments on Q-fly cDNA library were carried out with 100 nM fluorescently labelled FK506 at 30 °C for 90 min incubation time followed by 2 washes with 1× PBS + 0.01% BSA + 0.01% Tween-20 as blocking buffer.

The obtained data from optimisation experiments were employed for validation of previously constructed Q-fly library (Section 6.1.2) using fluorescein-$PEG_{12}$-FK506 to pull out Q-fly FKBPs from cDNA library by enrichment using FACS with each round of selection. However, after two rounds of selection, the gel image obtained from yeast colony PCR revealed that the yeast cells did not contain any cDNA insert between the restriction sites used (Section 6.2.7) for cloning Q-fly cDNA library in *S. cerevisiae* EBY100 suggesting a problem with the quality of the pre-made library. No further analysis or validation of this library was carried out due to time constraints.

Simultaneously, construction of a cDNA library of *C. elegans* was also attempted starting from total RNA isolation. Kozak consensus sequences were used for primer design for ds-cDNA synthesis from purified mRNA. As the Kozak sequence of *C. elegans* is not highly conserved, optimisation of annealing temperature during PCR was attempted using temperature gradient PCR (Section 7.2.4). However, as no bands were observed in this experiment, further analysis and designing new strategy was required, which was not possible due to time constraints.

## 8.2 FUTURE DIRECTIONS

The data obtained from optimisation of experimental parameters were employed for validation of the Q-fly library. However, due to lack of cDNA insert, it will be necessary to check the cDNA library diversity present in plasmid pSMART that was used for transformation in *S. cerevisiae* EBY100. If the library diversity looks sufficient to cover entire proteome of *B. tryoni*, then the library can be transferred to YSD vector pCTZQ using Infusion cloning technology. Hence, detailed analysis of primers and adapters for infusion cloning is needed to proceed. However, if library diversity in pSMART vector is not as desired, then original cDNA library sample used for transformation in pSMART vector will be needed to be analysed. Furthermore, attempts should be made to eliminate the need to use intermediate vector pSMART that is available from the kit before transferring the cDNA library into pCTZQ vector. This will simplify the process of construction of cDNA library as transformation in two vectors will not be required. For this, more analysis of pCTZQ vector and its components will be required. Moreover, addition of restriction sites to cDNA library and primer design should be carried out as per the sequence and restriction sites employed for cDNA synthesis and its insertion in pCTZQ vector between compatible sites.

Simultaneously, the construction of cDNA library of *C. elegans* using Kozak sequence was also unsuccessful. So, a new strategy will be needed to be designed by either changing Kozak priming strategy or direct construction of cDNA library including 5'-UTR and eliminating this sites using vector variant strategy as used earlier with Q-fly cDNA library (Section 6.1.2).

Once cDNA library is successfully constructed and transformed in *S. cerevisiae* EBY100 representing YSD cDNA library of an organism it will be validated using Fluorescein-$PEG_{12}$-FK506 probe in FACS analysis to study FKBP2-FK506 interactions. Followed by validation, the cDNA library will be screened against potentially bioactive natural product to identify its potential protein targets using FACS analysis. Once protein targets are identified, biochemical pathway modelling will be carried out to expand the possibility to develop this natural product as a potential antiparasitic drug.

# 9. REFERENCES


1. Organization, W.H. Investing to Overcome the Global Impact of Neglected Tropical Diseases: Third WHO Report on Neglected Tropical Diseases 2015, Vol. 3. (World Health Organization, 2015).
2. Tambo, E., Khater, E.I., Chen, J.-H., Bergquist, R. & Zhou, X.-N. Nobel prize for the artemisinin and ivermectin discoveries: a great boost towards elimination of the global infectious diseases of poverty. *Infectious Diseases of Poverty* **4**, 58 (2015).
3. Geary, T.G. et al. Unresolved issues in anthelmintic pharmacology for helminthiases of humans. *International Journal for Parasitology* **40**, 1-13 (2010).
4. Piggott, A.M. & Karuso, P. Quality, not Quantity: The Role of Natural Products and Chemical Proteomics in Modern Drug Discovery. *Combinatorial Chemistry & High Throughput Screening* **7**, 607-630 (2004).
5. Wilkins, M., Williams, K.L., Appel, R.D. & Hochstrasser, D. Proteome research: new frontiers in functional genomics. (Springer Science & Business Media, 2013).
6. Taylor, E.W. The mechanism of colchicine inhibition of mitosis : I. Kinetics of Inhibition and the Binding of H(3)-Colchicine. *The Journal of Cell Biology* **25**, 145-160 (**1965**).
7. Rix, U. et al. Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib reveal novel kinase and nonkinase targets. *Blood* **110**, 4055-4063 (2007).
8. Gee, K.R., Kang, H.C., Meier, T.I., Zhao, G. & Blaszcak, L.C. Fluorescent Bocillins: Synthesis and application in the detection of penicillin-binding proteins. *Electrophoresis* **22**, 960-965 (2001).
9. Schenone, M., Dančík, V., Wagner, B.K. & Clemons, P.A. Target identification and mechanism of action in chemical biology and drug discovery. *Nature chemical biology* **9**, 232-240 (2013).
10. Crews, C., Collins, J., Lane, W., Snapper, M. & Schreiber, S. GTP-dependent binding of the antiproliferative agent didemnin to elongation factor 1 alpha. *Journal of Biological Chemistry* **269**, 15411-15414 (1994).
11. Crews, C.M., Lane, W.S. & Schreiber, S.L. Didemnin binds to the protein palmitoyl thioesterase responsible for infantile neuronal ceroid lipofuscinosis. *Proceedings of the National Academy of Sciences* **93**, 4316-4319 (1996).
12. Sche, P.P., McKenzie, K.M., White, J.D. & Austin, D.J. Display cloning: functional identification of natural product receptors using cDNA-phage display. *Chemistry & Biology* **6**, 707-716 (1999).
13. Piggott, A.M. & Karuso, P. Rapid identification of a protein binding partner for the marine natural product kahalalide F by using reverse chemical proteomics. *ChemBioChem* **9**, 524-530 (2008).
14. Angelini, A. et al. in Yeast Surface Display: Methods, Protocols, and Applications. (ed. B. Liu) 3-36 (Springer New York, New York, NY; 2015).

15. Jacobsson, K., Rosander, A., Bjerketorp, J. & Frykberg, L. Shotgun phage display-selection for bacterial receptins or other exported proteins. *Biol Proced Online* **5**, 123-135 (2003).

16. Smith, G.P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315-1318 (1985).

17. Rosander, A., Bjerketorp, J., Frykberg, L. & Jacobsson, K. Phage display as a novel screening method to identify extracellular proteins. *Journal of microbiological methods* **51**, 43-55 (2002).

18. Li, W. & Caberoy, N.B. New perspective for phage display as an efficient and versatile technology of functional proteomics. *Applied microbiology and biotechnology* **85**, 909-919 (2010).

19. Li, J.W.-H. & Vederas, J.C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **325**, 161-165 (2009).

20. Caberoy, N.B., Zhou, Y., Jiang, X., Alvarado, G. & Li, W. Efficient identification of tubby-binding proteins by an improved system of T7 phage display. *Journal of Molecular Recognition* **23**, 74-83 (2010).

21. Bidlingmaier, S. & Liu, B. in cDNA Libraries: Methods and Applications. (eds. C. Lu, J. Browse & J.G. Wallis) 199-210 (Humana Press, Totowa, NJ; 2011).

22. Cherry, J.M. et al. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic acids research*, gkr1029 (2011).

23. Guo, D. et al. A tethered catalysis, two-hybrid system to identify protein-protein interactions requiring post-translational modifications. *Nature biotechnology* **22**, 888-892 (2004).

24. Holler, P.D. et al. In vitro evolution of a T cell receptor with high affinity for peptide/MHC. *Proceedings of the National Academy of Sciences* **97**, 5387-5392 (2000).

25. Cherf, G.M. & Cochran, J.R. in Yeast Surface Display: Methods, Protocols, and Applications. (ed. B. Liu) 155-175 (Springer New York, New York, NY; 2015).

26. Andreu, C. & del Olmo, M.l. Development of a new yeast surface display system based on Spi1 as an anchor protein. *Applied Microbiology and Biotechnology* **101**, 287-299 (2017).

27. Boder, E.T. & Wittrup, K.D. Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol* **328**, 430-444 (2000).

28. Boder, E.T. & Wittrup, K.D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* **15**, 553-557 (1997).

29. Chao, G. et al. Isolating and engineering human antibodies using yeast surface display. *Nat. Protocols* **1**, 755-768 (2006).

30. Smith, S.N., Harris, D.T. & Kranz, D.M. in Yeast Surface Display: Methods, Protocols, and Applications. (ed. B. Liu) 95-141 (Springer New York, New York, NY; 2015).

31. Harding, M.W., Galat, A., Uehling, D.E. & Schreiber, S.L. A receptor for the immuno-suppressant FK 506 is a cis–trans peptidyl-prolyl isomerase. *Nature* **341**, 758-760 (1989).

32. Kieke, M.C., Cho, B.K., Boder, E.T., Kranz, D.M. & Wittrup, K.D. Isolation of anti-T cell receptor scFv mutants by yeast surface display. *Protein Engineering, Design and Selection* **10**, 1303-1310 (1997).

33. Bidlingmaier, S. & Liu, B. Construction and application of a yeast surface-displayed human cDNA library to identify post-translational modification-dependent protein-protein interactions. *Molecular & cellular proteomics : MCP* **5**, 533 (2006).

34. Bidlingmaier, S. & Liu, B. in Yeast Surface Display: Methods, Protocols, and Applications. (ed. B. Liu) 193-202 (Springer New York, New York, NY; 2015).

35. Hu, J., Liu, J., Ghirlando, R., Saltiel, A.R. & Hubbard, S.R. Structural Basis for Recruitment of the Adaptor Protein APS to the Activated Insulin Receptor. *Molecular Cell* **12**, 1379-1389 (2003).

36. Shusta, E.V., Holler, P.D., Kieke, M.C., Kranz, D.M. & Wittrup, K.D. Directed evolution of a stable scaffold for T-cell receptor engineering. *Nature biotechnology* **18**, 754-759 (2000).

37. Bidlingmaier, S. & Liu, B. in Yeast Surface Display: Methods, Protocols, and Applications. (ed. B. Liu) 203-214 (Springer New York, New York, NY; 2015).

38. Bidlingmaier, S. & Liu, B. Interrogating yeast surface-displayed human proteome to identify small molecule-binding proteins. *Molecular and Cellular Proteomics* **6**, 2012-2020 (2007).

39. Harlan, J.E., Hajduk, P.J., Yoon, H.S. & Fesik, S.W. Pleckstrin homology domains bind to phosphatidylinositol-4,5-bisphosphate. *Nature* **371**, 168-170 (1994).

40. Miura, N., Kuroda, K. & Ueda, M. in Yeast Surface Display: Methods, Protocols, and Applications. (ed. B. Liu) 217-232 (Springer New York, New York, NY; 2015).

41. Colby, D.W. et al. in Methods in Enzymology, Vol. Volume 388 348-358 (Academic Press, 2004).

42. Bidlingmaier, S., Su, Y. & Liu, B. in Yeast Surface Display: Methods, Protocols, and Applications. (ed. B. Liu) 51-63 (Springer New York, New York, NY; 2015).

43. Yang, N. et al. The contribution of Pir protein family to yeast cell surface display. *Applied Microbiology and Biotechnology* **98**, 2897-2905 (2014).

44. Simoes, T., Teixeira, M.C., Fernandes, A.R. & Sa-Correia, I. Adaptation of Saccharomyces cerevisiae to the herbicide 2,4-dichlorophenoxyacetic acid, mediated by Msn2p- and Msn4p-regulated genes: important role of SPI1. *Applied and environmental microbiology* **69**, 4019-4028 (2003).

45. De Wet, J.R., Wood, K., DeLuca, M., Helinski, D.R. & Subramani, S. Firefly luciferase gene: structure and expression in mammalian cells. *Molecular and cellular biology* **7**, 725-737 (1987).

46. Lim, S., Glasgow, J.E., Filsinger Interrante, M., Storm, E.M. & Cochran, J.R. Dual display of proteins on the yeast cell surface simplifies quantification of binding interactions and enzymatic bioconjugation reactions. *Biotechnology Journal* **12**, 1600696-n/a (2017).

47. Jones, D.S., Tsai, P.-C. & Cochran, J.R. Engineering hepatocyte growth factor fragments with high stability and activity as Met receptor agonists and antagonists. *Proceedings of the National Academy of Sciences* **108**, 13035-13040 (2011).

48. Van Antwerp, J.J. & Wittrup, K.D. Fine Affinity Discrimination by Yeast Surface Display and Flow Cytometry. *Biotechnology Progress* **16**, 31-37 (2000).

49. Wang, Y. & Wang, Y. Advance in the mechanisms and application of microecologics in aquaculture. *Prog. Vet. Med* **29**, 72-75 (2008).

50. Li, S. et al. Yeast Surface Display of Escherichia coli Enterotoxin and Its Effects of Intestinal Microflora and Mucosal Immunity. *Current Microbiology* **74**, 854-862 (2017).

51. Kino, T. et al. FK-506, a novel immunosuppressant isolated from a Streptomyces. II. Immunosuppressive effect of FK-506 in vitro. *The Journal of antibiotics* **40**, 1256-1265 (1987).

52. Piggott, A.M., Kriegel, A.M., Willows, R.D. & Karuso, P. Rapid isolation of novel FK506 binding proteins from multiple organisms using gDNA and cDNA T7 phage display. *Bioorganic & Medicinal Chemistry* **17**, 6841-6850 (2009).

53. Dower, W.J., Miller, J.F. & Ragsdale, C.W. High efficiency transformation of E.coli by high voltage electroporation. *Nucleic Acids Research* **16**, 6127-6145 (1988).

54. Phaff, H.J. in Food Proteins, Vol. 160 244-282 (American Chemical Society, 1977).

55. Yang, Y., Hahne, H., Kuster, B. & Verhelst, S.H.L. A Simple and Effective Cleavable Linker for Chemical Proteomics Applications. *Molecular & Cellular Proteomics : MCP* **12**, 237-244 (2013).

56. Sekiguchi, M. et al. An evaluation tool for FKBP12-dependent and -independent mTOR inhibitors using a combination of FKBP-mTOR fusion protein, DSC and NMR. *Protein Engineering, Design and Selection* **24**, 811-817 (2011).

57. Moadeli, T., Taylor, P.W. & Ponton, F. High productivity gel diets for rearing of Queensland fruit fly, Bactrocera tryoni. *Journal of Pest Science* **90**, 507-520 (2017).

58. Artal-Sanz, M., de Jong, L. & Tavernarakis, N. Caenorhabditis elegans: A versatile platform for drug discovery. *Biotechnology Journal* **1**, 1405-1418 (2006).

59. Ly, K., Reid, S.J. & Snell, R.G. Rapid RNA analysis of individual Caenorhabditis elegans(). *MethodsX* **2**, 59-63 (2015).

60. Kozak, M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *Journal of Molecular Biology* **196**, 947-950 (1987).

61. Lee, S., Lee, I., Jung, Y., McConkey, D. & Czerniak, B. In-Frame cDNA Library Combined with Protein Complementation Assay Identifies ARL11-Binding Partners. *PLoS ONE* **7**, e52290 (2012).

62. Mankad, R.V., Gimelbrant, A.A. & Mcclintock, T.S. Consensus Translational Initiation Sites of Marine Invertebrate Phyla. *The Biological Bulletin* **195**, 251-254 (1998).

63. Meissner, P.S., Sisk, W.P. & Berman, M.L. Bacteriophage lambda cloning system for the construction of directional cDNA libraries. *Proceedings of the National Academy of Sciences* **84**, 4171-4175 (1987).

# 10. SUPPLEMENTARY DATA

## 10.1 MEDIA COMPOSITION

### 10.1.1 Media used for transformation in *E. coli*

1. Luria-Bertani (LB)- broth (1L):

   Tryptone- 10 g

   NaCl- 10 g

   Yeast Extract- 5 g

   Agar- 30 g (for plates)

   Adjust final volume to 1 L using milli-Q water. Autoclave and allow it to cool to 50 °C. Add ampicillin to the final concentration of 100 µg/mL. Store at RT upto 6 months for LB-broth and plates at 4 °C upto 6 months

2. SOC Broth (1 L):

   Tryptone- 20 g

   Yeast Extract- 5 g

   10 mM NaCl- 0.584 g

   2.5 mM KCl- 0.186 g

   10 mM $MgCl_2$- 0.952 g

   10 mM $MgSO_4$- 1.204 g

   Adjust final volume to 1 L using milli-Q- water and autoclave. Store at RT for 6 months.

### 10.1.2 Media used for *S. cerevisiae*:

1. YPD Broth/ Media (1 L):

   Yeast Extract- 10 g

   Bacto-peptone- 20 g

   Dextrose- 20 g

   Agar- 30 g (for plates)

Bring volume to 1 L using milli-Q water and autoclave. Autoclave and allow it to cool to 50 °C. Add Zeocin (if required) to the final concentration of 50 µg/mL. Add dextrose after autoclaving to avoid caramelisation and mix properly. Store broth at RT and plates at 4 °C for 1 month.

2. SD-CAA Media (1 L):

   Sodium Citrate- 14.8 g

   Citric acid monohydrate- 4.2 g

   Casaminoacids- 5 g

   Yeast nitrogen base (without amino acids)- 6.7 g

   Dextrose- 20 g

   Zeocin- 50 µg/mL

   Bring volume to 1 L using milli-Q water, sterile filter and store at 4 °C for up to 6 months.

3. SD-CAA plates (500 mL):

   Sorbitol- 91.1 g

   Agar- 15 g

   Sodium Citrate- 7.4 g

   Citric acid monohydrate- 2.1 g

   Dissolve in 400 mL milli-Q water, auto clave and cool to 50 °C.

   In separate container, combine

   Dextrose- 10 g

   Yeast nitrogen base (without amino acids)- 3.35 g

   Zeocin- 50 µg/mL

   Dissolve in 100 mL milli-Q water, sterile filter and add to cooled autoclave solution. Store the plates at 4 °C for up to 6 months.

4. SG-CAA media (1 L):

   Sodium Citrate- 14.8 g

Citric acid monohydrate- 4.2 g

Casaminoacids- 5 g

Yeast nitrogen base (without amino acids)- 6.7 g

Galactose- 20 g

Zeocin- 50 µg/mL

Bring volume to 1 L using milli-Q water, sterile filter and store at 4 °C for up to 6 months.

## 10.2 BUFFER COMPOSITION

### 10.2.1 Buffer used for Flow Cytometry

1. 1× PBS Buffer, pH 7.4:

Start with 800 ml of distilled water.

NaCl- 8 g

KCl- 0.2 g

$Na_2HPO_4$- 1.44 g

$KH_2PO_4$- 0.24 g

Adjust the pH to 7.4 with 0.1 M HCl.

Add milli-Q water to a total volume of 1 L and autoclave. Store at room temperature for 1 month. Prepare fresh 1X PBS buffer each time it is used for flow cytometry.

### 10.2.2 Buffer used for Agarose Gel Electrophoresis

1. 50× Tris-acetate EDTA (TAE) Buffer Stock (1 L):

Dissolve 242g Tris base in milli-Q water. Add 57.1mL glacial acetic acid and 100mL of 500mM EDTA (pH 8.0) solution. Bring the final volume to 1 L using milli-Q water. This stock solution is diluted 50 fold with water to make a 1X working solution which is used for agarose gel electrophoresis. Alternatively 0.5× Tris-borate EDTA (TBE) buffer can also be used as higher concentrations of TBE can lead to poor band due to excess heat generated during elctrophoresis.

### 10.2.3 Buffer used for Yeast Colony PCR

1. 1× TE Buffer, pH 8 (100 mL):

1× TE Buffer is used for making 2 U Zymolase solution from 10,000 U stock solution.

Tris Base (pH-8)- 1 mL

EDTA- 0.2 mL

Make final volume to 100 mL using milli-Q water. Adjust pH using 0.1 M HCl. Sometimes stock solution of 100× TE is also prepared which can be diluted further using milli-Q water according to requirement.

## 10.3 COMPOSITION OF PCR REACTIONS

### 10.3.1 Concentration of components in PCR reaction (25 µL):

| Component | Volume (µL) | Final Concentration |
|---|---|---|
| 10× PCR Buffer | 2.5 | 1× |
| 10 mM dNTPs | 0.5 | 0.2 mM |
| 10 µM Forward Primer | 0.5 | 0.2 µM |
| 10 µM Reverse Primer | 0.5 | 0.2 µM |
| 5 U DNA Polymerase | 0.125 | 0.25 U |
| Nuclease- free $H_2O$ | 19.875 | - |
| Template DNA | 1 | - |
| **Total Volume** | **25** | **-** |

For yeast colony PCR, 10% Triton-X 100 (1 µL) is used to avoid formation of secondary structures. Moreover if PCR buffer does not contain $MgCl_2$, 50 mM of $MgCl_2$ is added separately to a final concentration of 1.5 mM.

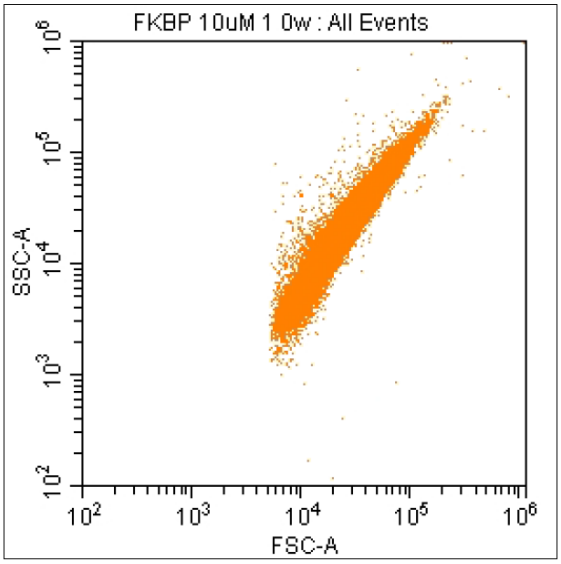
## 10.4 LONGTERM STORAGE OF YEAST

### 10.4.1 Storage at -80 °C:

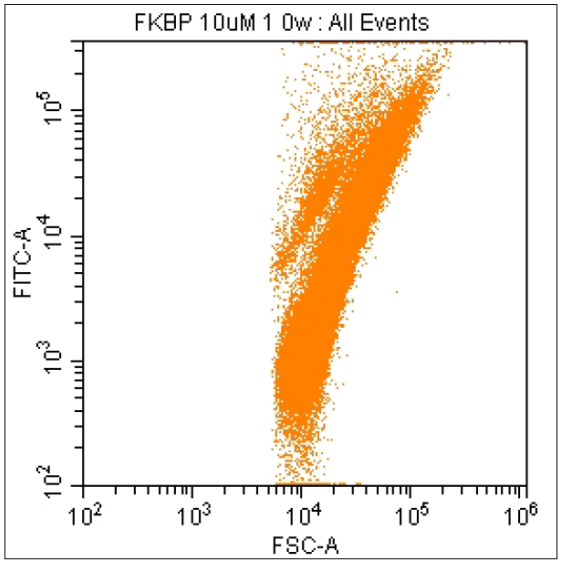
Pick a single yeast colony from a plate. Inoculate into 3 ml YPD liquid media and grow overnight at 30 °C with shaking at 220 rpm. Aliquot about 250 µl culture into a sterile microfuge tube and add 1 volume of 30% glycerol. Mix well and store at –80 °C.

## 10.5 FLOW CYTOMETRIC DATA

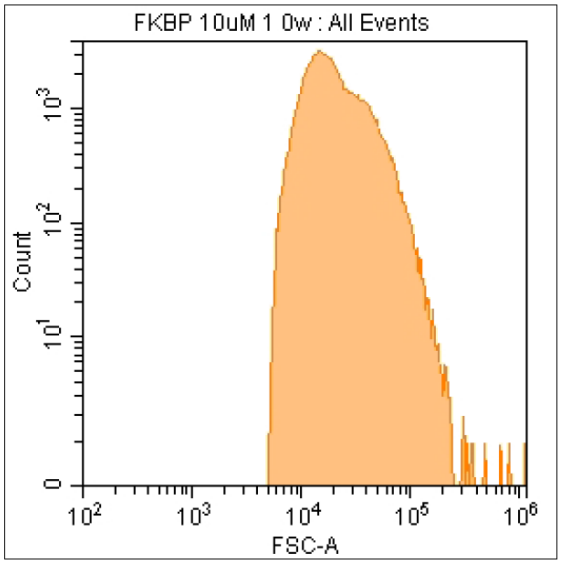
### 10.5.1 Plots obtained with Beckman Coulter Flow Cytometer Used for Optimisation Experiments (Section 5):



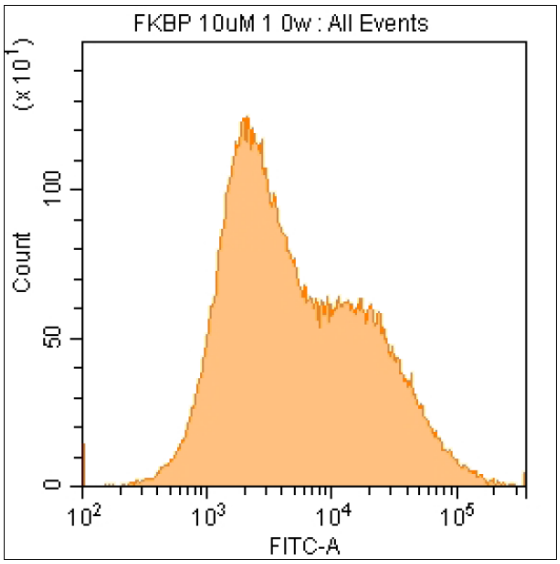

[A] Forward Scatter vs Side Scatter



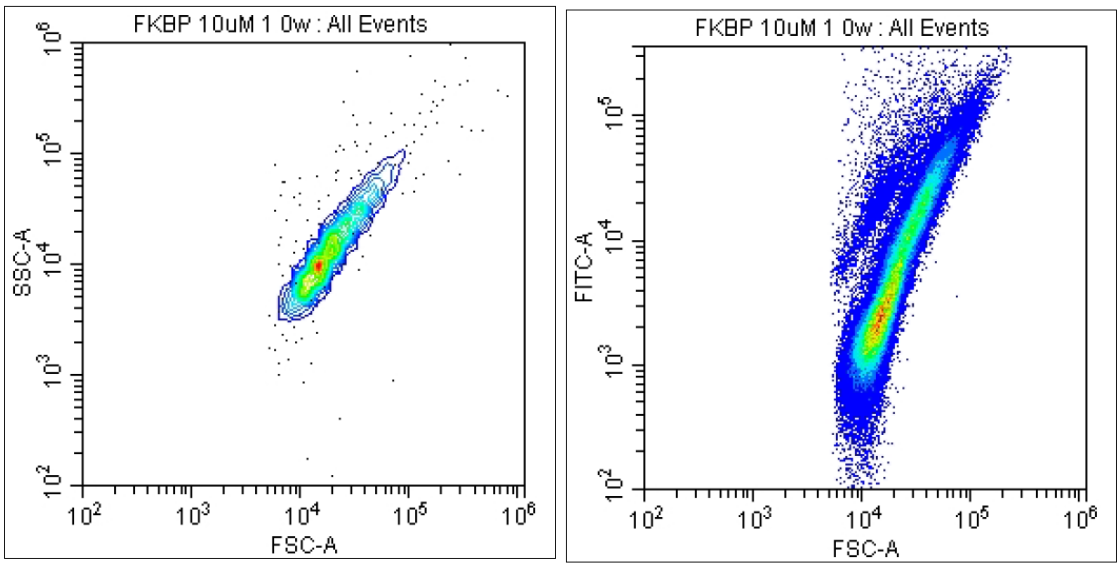

[B] Forward Scatter vs Fluorescence Intensity




[C] Forward Scatter vs Count




[D] Fluorescence Intensity vs Count

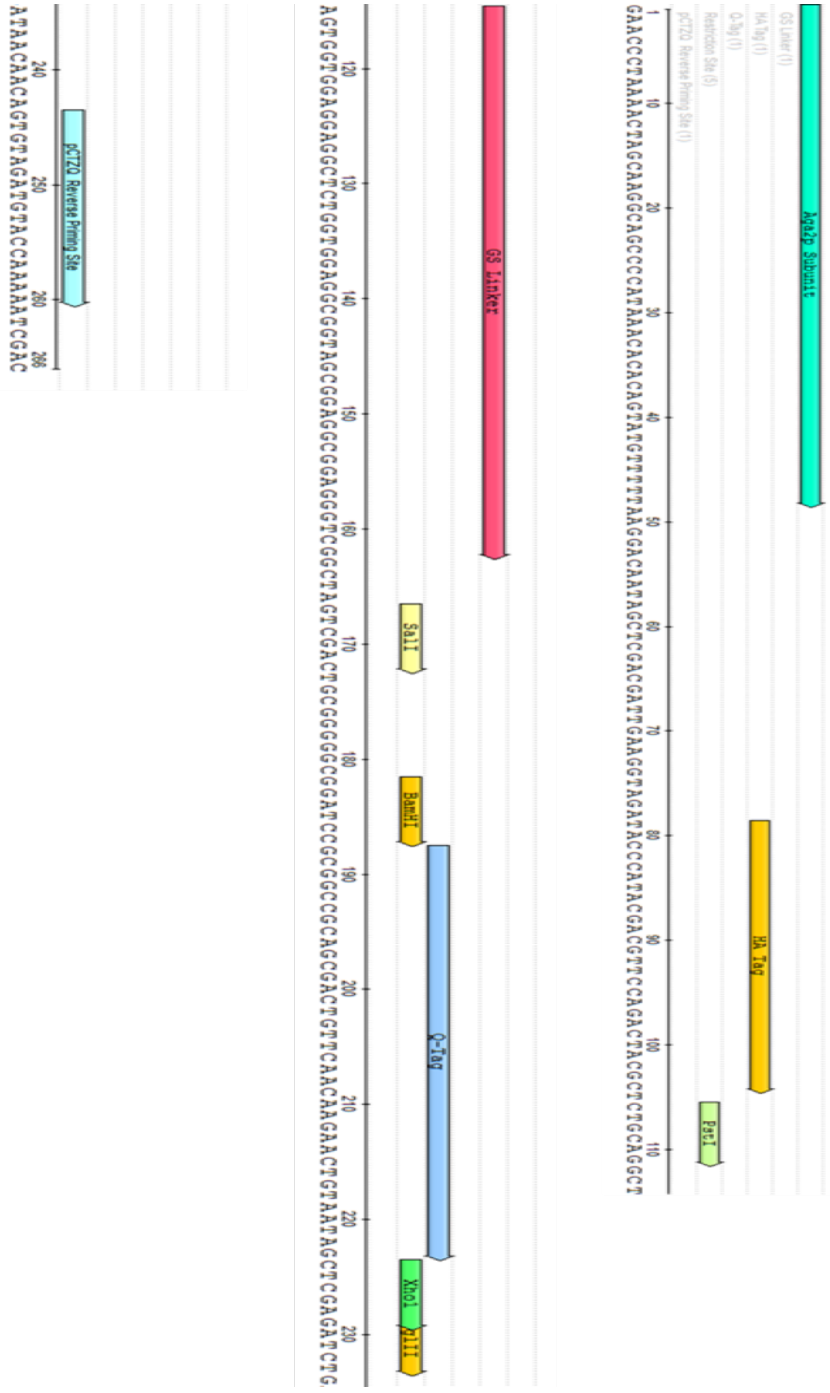



[E] Contour Plot [F] Density Plot

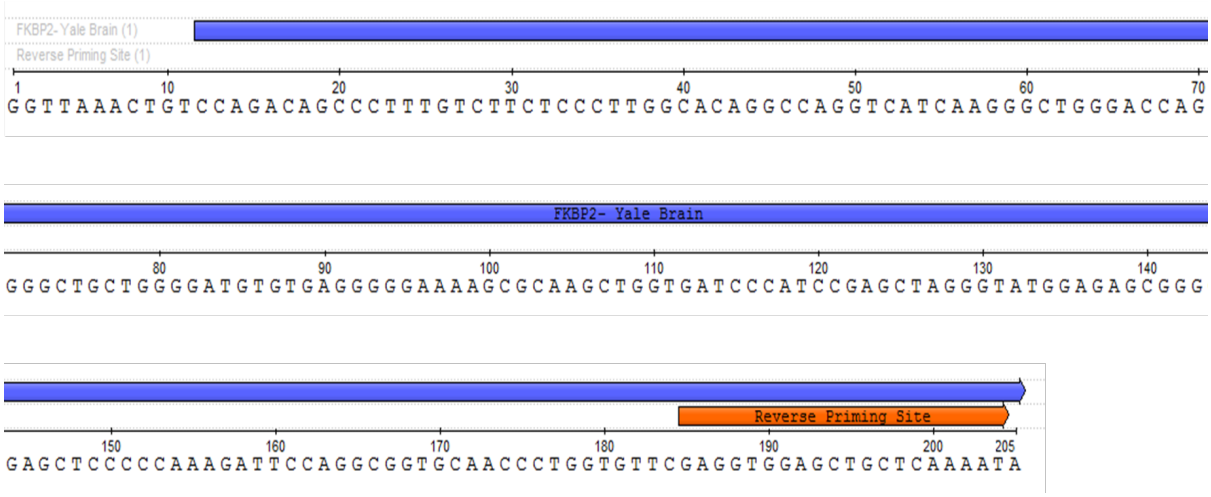
The scatter plots, histograms, contour plots and density plots in the above figures are the one obtained with flow cytometric analysis of *S. cerevisiae* EBY100 containing pCTZQ-FKBP2 stained with 10 µM Fluorescein-$PEG_{12}$-FK506. 100,000 events were noted and analysed here and similar analysis was carried out with other concentrations as well as range of other parameters.

## 10.6 SEQUENCING ANALYSIS

### 10.6.1 Annotated Sequence of empty plasmid pCTZQ between pCTZQ Forward and Reverse Primers:



### 10.6.2 Annotated Sequence of hFKBP2 between FKBP2 Forward and Reverse Primers:



MACQUARIE UNIVERSITY

**Office of the Deputy Vice-Chancellor (Research)**
Research Office

C5C Research HUB East, Level 3, Room 324

MACQUARIE UNIVERSITY NSW 2109 AUSTRALIA

**Phone** **+61 (0)2 9850 4063**
Fax +61 (0)2 9850 4465

Email biosafety@mq.edu.au

7th December 2016

Dear Dr Piggott

**Re: Exempt Dealing "Yeast display as a platform technology to accelerate antiparasitic drug discovery and development" [5201600950]**

Thank you for your application for the above Exempt Dealing. The Institutional Biosafety Committee (IBC) has reviewed your application and approval has been granted effective December 7th 2016.

Approval has been granted subject to your compliance with the Office of the Gene Technology Regulator's standard conditions for exempt work listed below:

1. The project must be conducted in accordance with the OGTR Guidance Notes for the Containment of Exempt Dealings (http://www.ogtr.gov.au/internet/ogtr/publishing.nsf/Content/ExemptDealGuideSept11_2-htm
2. The Guidance Notes are only applicable to exempt dealings conducted under the *Gene Technology Act 2000*. They do not provide guidance for laboratory safety, good laboratory practice or work health and safety issues. For these purposes, refer to AS/NZS 2243.3:2010.
3. You must inform the Institutional Biosafety Committee if you complete or abandon the exempt dealings with GMOs.

Please note the following standard requirements of approval:

1. Approval will be for a period of 5 years *subject to the provision of annual reports* (http://www.research.mq.edu.au/current_research_staff/gene_technology_and_biosafety/submitting_a_new_application). If, at the end of this period the project has been completed, abandoned, discontinued or not commenced for any reason, you are required to submit a Final Report. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. Please contact the Committee Secretary at biosafety@mq.edu.au in order to obtain a report.

   A Progress/Final Report for this project will be due on: December 7th 2017

2. If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have final approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of Final Approval to an external organisation as evidence that you have Final Approval, please do not hesitate to contact the Committee Secretary at biosafety@mq.edu.au or by phone 9850 4063.
Please retain a copy of this email as this is your formal notification of final Biosafety approval.

Kind regards,

**Biosafety Secretariat**
Research Office
Level 3, Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia

T: +61 2 9850 4063
F: +61 2 9850 4465
http://www.mq.edu.au/research