MACQUARIE
University

Simultaneous Yeast Display for Medical Biosensors

By

Rashika Sood

A thesis submitted to Department of
Molecular Sciences, Macquarie University
September 2022

Office of the Gene Technology Regulatory
(OGTR) Reference number: 5201401059

Declaration of Authorship and of thesis submission history

This is to certify that I, Rashika Sood, being a candidate for the degree of Master of Research, have not submitted this work to this or any other university or educational institution for the purposes of attaining a higher degree. I also certify that this work is my own, and was not taken, in whole or in part, from un-attributed sources. All work was conducted adhering to the biosafety and biosecurity regulations covered by 5831 - Paulsen - 5201401059 – Yeast 2.0 (2) – GMO Exempt.

1.09.2022

Candidate signature **Date**

Table of Contents

List of Figures

List of Tables

Abstract

Biosensors play an important role in medical diagnostics, industrial biotechnology, and environmental monitoring. A key challenge in biosensor design is developing complementary ligand binding domains which bind to the small molecule in a sandwich like fashion, causing fused biosensor output domains to co-localise and transduce a signal. We lay the foundation of designing an *in vivo* system utilising yeast mating for tackling the difficult task of generating novel pairs of complementary binding domains termed as Simultaneous Yeast Display (SYD). This system relies on strains of yeast that have their native sexual agglutination ability knocked out such that mating of haploid cells to form a diploid is dependent upon interactions between heterologous expressed surface proteins. While this system has been used previously to characterise and select for protein-protein interactions, its use in selecting protein-ligand-protein interactions has not been published. We present results showing that surface display of known ligand binding proteins causes an increase in mating efficiency upon introduction of the ligand in liquid culture. In the future, this technique can be further used to screen yeast display libraries of randomised peptides or proteins to discover novel complementary binding domains with high sensitivity and selectivity for a multitude of ligands.

1. Introduction

1.1 What are biosensors?

The ability of cells to sense and respond to their environment is a feature of life useful for survival in many conditions. For example, sense and respond systems underly the toxin sensing abilities of algae, electroreception in sharks and the extraordinary olfactory capabilities of canines[1,2,3]. Synthetic biologists have borrowed, engineered, and designed biosensors based on cellular sensing mechanisms existing in nature to harness the wealth of biological information around us.

Synthetic circuitry incorporates biological parts of the cell to mimic an electronic circuit. Biosensors engineered using synthetic biology design are built with a biorecognition domain, generally a protein or nucleic acid, coupled with a regulatory element such as transcription factor to produce a measurable output on interaction with the ligand of interest (Fig. 1). Such a synthetic circuit produces output signal in the form of fluorescence, colour change or generation of an electrical current when interacting with the target ligand. Target ligands can be diverse and range in size and complexity, thus biosensors have a variety of applications in drug discovery, point-of-care diagnostics, food safety and environmental monitoring.

The diversity of biosensor detection mechanisms lies in the broad range of biorecognition elements used. For example, transcriptions factors, nucleic acids, aptamers, antibodies, protein binding domains, biocatalytic enzymes and whole microorganisms have all been used to detect target ligands [4]. Each of the biorecognition elements has unique characteristics and binding mechanisms with the target ligand that contribute to the sensitivity, reusability, reliability, and selectivity of the biosensors. The mechanisms for engineering biosensor signal transduction have increased dramatically over the last decade ranging from electrochemical biosensors, optical sensors, acoustic-sensitive sensors, to field-effect transistors[5].

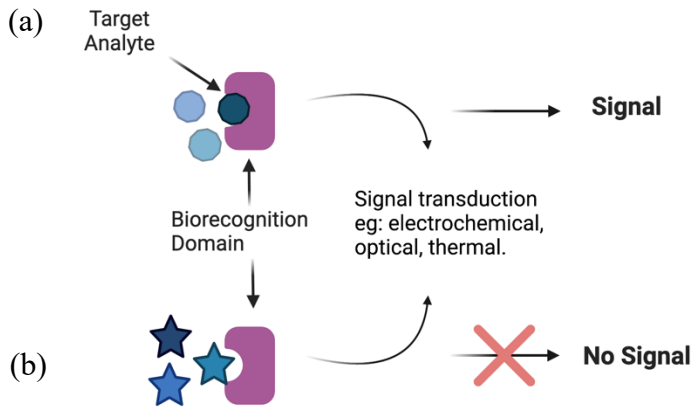

Figure 1: Schematic diagram of biosensor architecture. (a) Ligand of interest binds to the biorecognition domain, inducing a conformational/positional change and resulting in an observable output. (b) Ligands which do not have affinity for the biorecognition domain are not capable of inducing signal generation. Created with BioRender.com

1.2 Applications: Medical Diagnostics

Conventional diagnostics for pathology involve cell culturing and serology, which are labour intensive and take from two to fifteen days to produce results[6]. Additionally, some microorganisms can be extremely hard to cultivate[6]. This phenomenon is referred to as 'the great plate count anomaly'[16]. This has driven interest in developing culture-independent approaches such as reverse transcription-polymerase chain reaction RT-PCR and next generation sequencing (NGS). These modern techniques are high-throughput, fast, and sensitive. However, they utilise expensive machinery and have complex protocols which require specialised personnel. Other limitations are the time consuming nature of these approaches (NGS can take days and PCR can take hours), the requirement of a large sample volume (up to 1ml) and the production of false positives due to PCR errors and contamination. Label-free approaches such as Liquid and Gas Chromatography (LC and GC) coupled to tandem mass spectrometry (MS/MS) are also routinely used for detection of metabolites[7]. These techniques provide highly robust and accurate data; however, they are highly expensive, time-consuming and require skilled labour to execute whilst having low throughput.

In comparison, biosensors provide an advantage in terms of ease of use, portability, accuracy, affordability, specificity and a faster response time of seconds[8]. They are an attractive tool to provide rapid and reliable information in medical diagnostics, like detection of infections, indication and progression of disease, real-time monitoring of biomolecules and therapeutics in the blood[9].

In 1956, Leland C. Clark Jr. developed the first glucose biosensor which works by using the enzyme glucose oxidase to convert glucose (the substrate) into gluconic acid and hydrogen peroxide. The hydrogen peroxide produced is then detected by an electrode, which sends a signal indicating the concentration of glucose present in a sample[10]. In 1975, the glucometer was commercialised and has been the most widely used biosensor in the world for monitoring diabetes patient health[10]. Today the global glucose biosensor market is projected to reach USD 31.0 billion by 2022[11].

Early glucose biosensors were expensive as they used gold or platinum electrodes and were prone to interference effects by other components of blood[12]. Some techniques which enhanced glucose biosensor specificity were immobilization of an auxiliary enzyme and/or antibody which reacts with the target analyte to enhance the output signal[13]. This led to the development of immunobiosensors such as enzyme-linked immunosorbent assay (ELISA) and Surface Plasmon Resonance (SPR) based biosensors which utilise recombinant antibodies or antibody fragments to sense target analytes (Fig. 2)[14]. Antibody-based systems are the gold standard in biosensors due to their high specificity to the target analyte. The pregnancy test strip is an example of a widely used antibody-based biosensor which recognizes the human growth hormone (hCG) with antibodies[15].

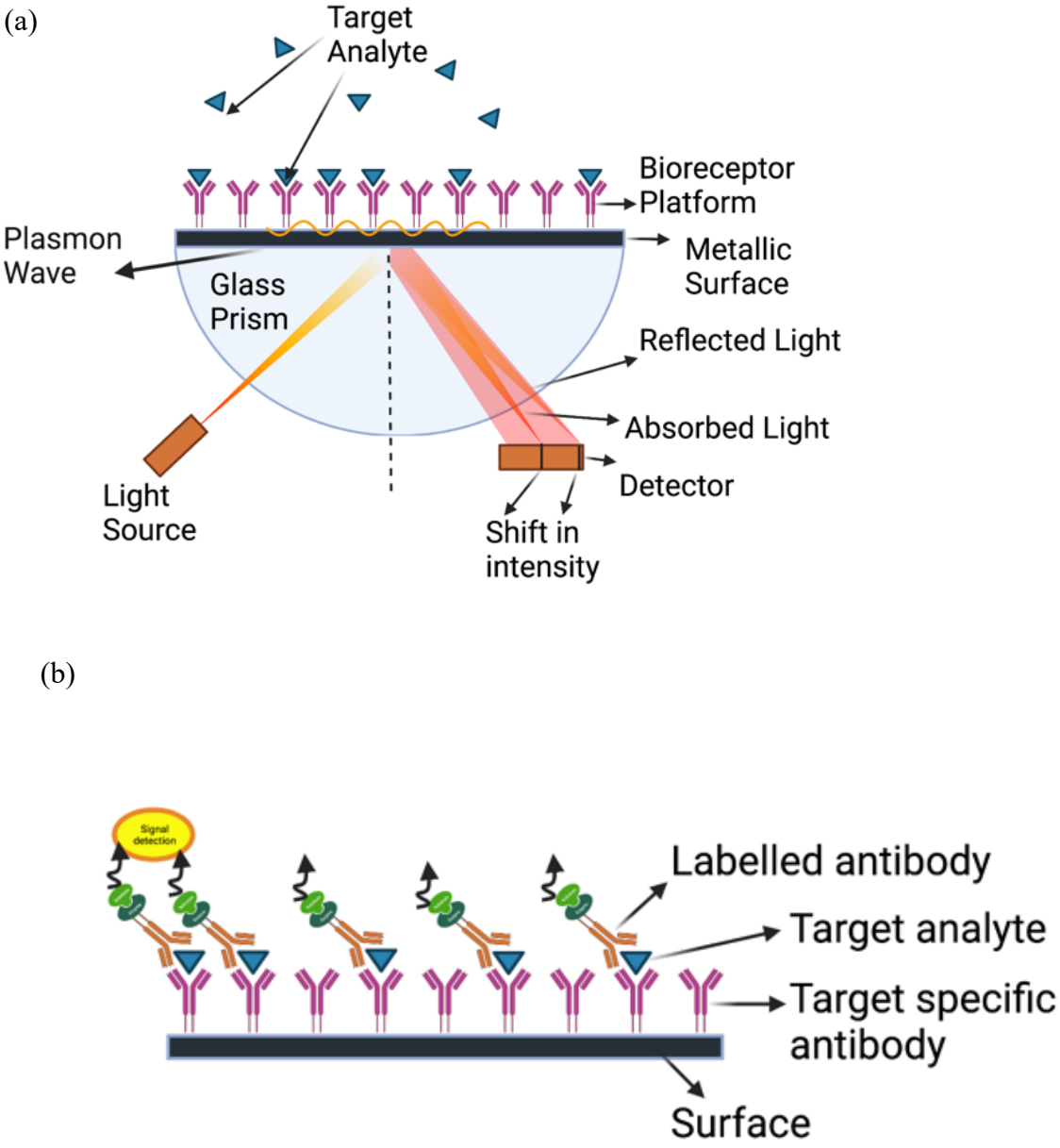

Figure 2: Schematic representation of the principal of Surface Plasmon Resonance (SPR) and Enzyme-Linked Immunosorbent Assay (ELISA). (a) When the target analyte binds to the biorecognition molecules on the metal detector surface, electrons in the metal absorb the light and cause a shift in the angle of the reflected light. This shift in the intensity minimum of the reflected light generates an SPR signal. (b) In ELISA, a capture antibody which binds with the target analyte is immobilised in a multi-well plate. A secondary antibody conjugated with an enzyme binds to the target analyte. The enzyme's substrate is added to produce a detectable signal, for example a colour change. Created with BioRender.com

Recent advancements in biosensor design such as lab-on-chip help in facilitating rapid detection and point-of-care diagnosis[16]. For example, the incorporation of enzymes with paper-based analytical devices (μPADs) has significantly improved analytical performance while exhibiting excellent chemical and storage stability[17]. Recently a hybrid paper-lab-on-a-chip platform was developed with four 3-D printed injectors to deliver 15 μl volumes to perform a multi-step protocol with an output signal which can be detected by a smartphone camera flash[18].

1.3 Biosensor Architecture and Engineering

In the last five decades, the biosensor industry has seen an explosive growth and evolved into an inter-disciplinary field requiring expertise in engineering, nanotechnology, molecular biology, and chemistry. Today, advanced biosensors based on nanomaterials have been developed with the capability of sensing target analytes in plasma, serum, and urine. However, these fluids are composed of thousands of ions, proteins, nucleic acids, and cells, hence specific and strong binding of the target analyte becomes one of the most important challenges to increase the signal to noise ratio of biosensors[19]. Current biosensor designs are dependent on the ligand being detected, and the field of use, and there is no standardised structure. Broadly, biosensor engineering is based on nucleic acid or proteins to allow specific and sensitive detection of target ligands discussed in the next section.

1.3.1 Nucleic-Acid Based Biosensors

1.3.1.1 Aptamers

Aptamers are single stranded DNA or RNA molecules which can specifically bind to the target ligand[20]. They have a high reproducibility, long shelf-life of up to 2 years, higher stability over a wide range of temperature and pH, and a more sensitive detection limit due to their smaller size in comparison to the gold-standard antibodies[21]. Aptamers can be chemically synthesized using an iterative, high-throughput methodology, systematic evolution of ligands by exponential enrichment (SELEX), at a much lower cost than antibodies[22]. SELEX utilises a library of random oligonucleotide sequences which are exposed to the target ligand to identify binders[23]. The library contains 10^{15} unique members of 30-80 nucleotide base pairs which is incubated with an immobilised target of interest[24]. After non-binding oligonucleotides are washed away, the binders are amplified by PCR to create a new enriched library and the process is repeated for 8-15 rounds[22]. This allows generation of artificial aptamers which can be sequenced and individually assessed for

binding affinity and specificity against natural and synthetic ligands[25]. Aptasensors offer rapid testing (minutes to hours) and can be regenerated easily[20].

1.3.1.2 Riboswitches

Single stranded RNA based aptamers can be integrated in a class of biosensors called riboswitches[25]. Riboswitches are regulatory elements in ribonucleic acid (RNA) molecules that can bind to small molecules and change their conformation in response to the binding[26]. This allows riboswitches to act as sensors for the presence or absence of specific molecules and regulate gene expression in response. The response domain of a riboswitch is the region of the RNA molecule that is responsible for detecting and responding to the binding of a small molecule[26]. It is typically located within the riboswitch and is made up of specific sequences of nucleotides that are able to bind to the small molecule. When the small molecule binds to the response domain, it causes a conformational change in the riboswitch that can affect the way in which the RNA molecule functions. For example, the binding of a small molecule to the response domain of a riboswitch may activate or inhibit the expression of a gene by altering the way in which the RNA molecule interacts with other proteins or with other regions of the RNA molecule itself. Usually, the aptamer used in a riboswitch has a secondary structure which undergoes a conformational change to activate signal generation[27].

However, due to the sensitive nature of RNA and their complex secondary structures, the construction of functional riboswitches is a difficult process. The secondary structures are influenced by pH and temperature changes and we currently have limited knowledge to rationally design and synthesise aptamers[28]. Another limitation in the majority of aptasensors which have been developed for small molecules is that they have only been tested in ideal buffer system and not in complex samples such as plasma, urine, and serum[29]. Aptasensors lack robustness in less ideal conditions and can lead to low signal-to-noise ratios due to low target concentrations or cross reactivity and contamination due to nucleases in the sample[21]. In comparison, protein-based biosensors showcase more robust behaviour in complex clinical samples.

1.3.2 Protein-based biosensors

1.3.2.1 Transcription-factor based biosensors

Transcription-factor based biosensors are protein-based biosensors which undergo a conformational change upon binding with a metabolite to generate a signal by expressing the downstream reporter gene[30]. TFBs are easy to engineer due to their modular architecture consisting of transcription factor, promoter region and reporter genes. In addition, a wide range of ligand-regulated transcriptional factors existing in nature[31,32]. When the transcription factor corresponding to the target ligand are known in nature and can be identified in literature, the construction of biosensors is straight forward. However, when the transcription regulatory elements corresponding to the target ligand are not known, the construction of the biosensor can be a complicated process. Usually, the host is exposed to increasing concentrations of the target ligand to identify the genes and promoters which are upregulated in response to the ligand[30,33]. Despite being tested with structurally similar target analytes, such transcription factor-based biosensors run the risk of activation by unexpected ligands which could result in higher rates of false-positives.

Another limitation of TFBs is the unavailability of regulatory proteins and target promoters for every molecule of interest. For example, there is a finite number of regulatory mechanisms which control metabolic activity and compatible transcription factors don't exist for every metabolite. A strategy to circumvent this problem is the transplantation of transcription factors from another species, such as prokaryotes to eukaryotes[30]. However, this can give rise to issues in translation due to the incompatibility of the regulatory elements and the cellular transcription machinery[30]. When a native transcription regulator and ligand pair doesn't exist, a synthetic regulator can be constructed in which parts of the native transcriptional regulator are fused to a generic activation domain which transcribes the signal gene[30]. Synthetic TFs (synTFs) are synthetic switches with engineered DNA-binding proteins targeted to bind with a specific target sequence which is usually integrated with a strong promoter to control the downstream gene of interest[34,35]. This enhances the sensitivity to theoretically any metabolite, however, construction and fine-tuning of synTFs is a very difficult process, hampering the ability to use this strategy as a ubiquitous design approach. Another drawback which makes TFBs unsuitable for point-of-use diagnostic applications is the reliance on the cellular transcription and translation machinery to detect the output signal. This increases the response time of the biosensors and runs the risk of contamination issues which arise from using microorganisms.

1.3.2.2 Synthetic Protein Switches (Transcription Independent Biosensors)

Synthetic protein switches are protein-based switches which contain the recognition and signal mechanisms which bypass the need for transcription[36]. On interaction with the target ligand, a conformational or localisation change occurs which produces an output signal[36]. Protein-based signalling in contrast to transcription-based signalling is orders of magnitude faster with a typical response time of minutes, instead of hours. Additionally, they have higher stability across different pH and temperatures. A variety of architectures have been commonly used in the past such as Bioluminescence Energy Transfer (BRET), Förster Resonance Energy Transfer (FRET), and Split Glucose Dehydrogenase.

1.3.2.2.1 Förster Resonance Energy Transfer (FRET) Based Biosensors

Genetically encoded Förster (or fluorescence) resonance energy transfer (FRET) (Fig. 4) based biosensors sandwich a ligand binding domain in between a pair of donor and acceptor fluorophores. Upon binding with the target ligand, the ligand binding domain undergoes a conformational change which brings the two fluorophores closer to produce a FRET signal[37]. FRET is a phenomenon in which energy is transferred from one molecule (the donor) to another molecule (the acceptor) through non-radiative energy transfer. This process occurs when the donor molecule is excited by absorbing light at a specific wavelength and then transferring its excess energy to the acceptor molecule through non-radiative dipole-dipole interactions. In order for FRET to occur, the donor and acceptor molecules must be within a certain distance of each other (typically less than 10 nm) and their electronic energy levels must be properly aligned. When the donor molecule is excited, it undergoes a conformational change that allows it to transfer energy to the acceptor molecule. FRET-based biosensors have been used in the detection of a wide variety of target analytes, such as ions, cofactors, amino acids, and real-time monitoring of intracellular protein concentration[38]. FRET biosensors have been employed in clinical diagnostics for imaging, monitoring disease progression, response to treatment and understanding intermolecular dynamics. For example, a FRET nanosensor for measuring NADPH concentration in cytosol and mitochondria of cancer cells revealed NADPH metabolism in cancer cells is impacted by the availability of glucose[39]. Whilst FRET-based biosensors are easy to construct and offer high orthogonality and resolution, their construction is dependent on finding the right ligand binding domain for the target analyte. They are efficient in reporting abundance of target analyte; however, they lack the ability to regulate downstream reactions in response to the signal.

1.3.2.2.2 Split Glucose Dehydrogenase

Another example of a synthetic protein switch is the split glucose dehydrogenase biosensor (Fig. 4) where the protein pyrroloquinoline quinone glucose dehydrogenase (PQQ-GDH) is bifurcated and the two halves of the protein are connected with a linker. The linker in a PQQ-GDH biosensor refers to the chemical compound or compounds that are used to attach or "link" the PQQ enzyme to the GDH enzyme. The linker is typically chosen based on its ability to efficiently and stably connect the two enzymes while still allowing them to function properly. In a biosensor, the two halves of PQQ-GDH are expressed as a fusion with a binding domain corresponding to the target ligand, thus upon binding, there is a conformational change that propagates the linker to reconstitute the PQQ-GDH protein to initiate electron transfer[40]. The electrochemical signal output can be easily integrated with electronic devices allowing development of easy, point-of-care diagnostics. PQQ-GDH based biosensors have been used for the detection of immunosuppressant drugs and enzyme activity of thrombin and Factor Xa[41].

The underlying challenge in developing synthetic protein switches is (1) inability to use antibodies and synthetic antibodies as they only display a minute conformational change on binding with the ligand which is not large enough to transfer to the switch module (2) time-consuming efforts required for engineering the binding domain protein(s) if a naturally occurring ligand binding proteins don’t exist. However, in comparison to nucleic acid biosensors, protein-based biosensors are more robust and industrially favourable for point-of-care diagnostics due to their stability and lower sensitivity to changes in pH or buffer composition.

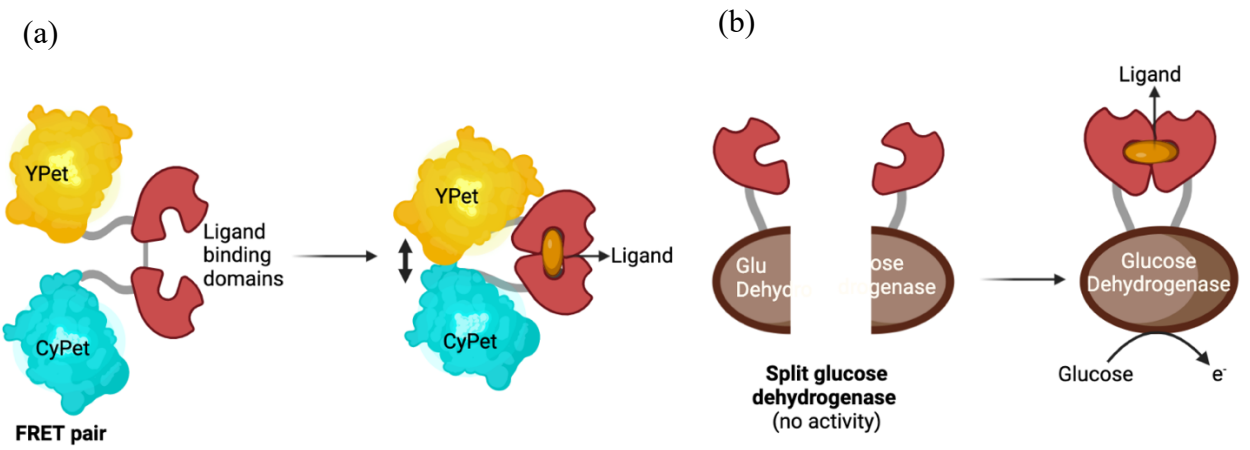

Figure 3: A general architecture of a Förster (or fluorescence) resonance energy transfer (FRET) pair and split glucose dehydrogenase synthetic protein switch. (a) The FRET pair consists of a ligand binding domain sandwiched between cyan fluorescent protein (CyPet) and yellow fluorescent protein (YPet) fluorophores as acceptor and donor molecules, respectively. When the ligand binding domains bind with the ligand, a conformational change between the donor and the acceptor molecule produces a measurable change in the emission spectrum of the acceptor molecule. (b) The pyrroloquinoline quinone glucose dehydrogenase PQQ-GDH is bifurcated and expressed as a fusion protein with the binding domain of the target ligand. On binding with the target, the enzyme reconstitutes and initiates electron transfer which can be detected as an electrochemical signal. Created with BioRender.com

1.4 Identifying and engineering biorecognition elements

Biosensors rely on a unique biorecognition element for the target analyte to be assessed in order to allow rapid, specific, and sensitive detection of the analyte of interest[42]. While a range of transduction and signal output mechanisms exist, the major factor dictating the performance of the biosensor is the specificity and affinity of the biorecognition element and its capacity to differentiate between the target analyte from other interfering substances[43].

Conventionally, biosensors utilised whole microorganisms integrated with a physical transducer to monitor specific metabolites. *Escherichia coli* and yeast-based biosensors have been used for detection of pathogens and carcinogens (Fig. 3.)[44]. However, a limitation with whole microorganism-based biosensors is the response time is slower as they are based on the transcription and translation machinery of the cell when the output is dependent on a reporter gene. Response time can even span multiple days when the output signal is growth.

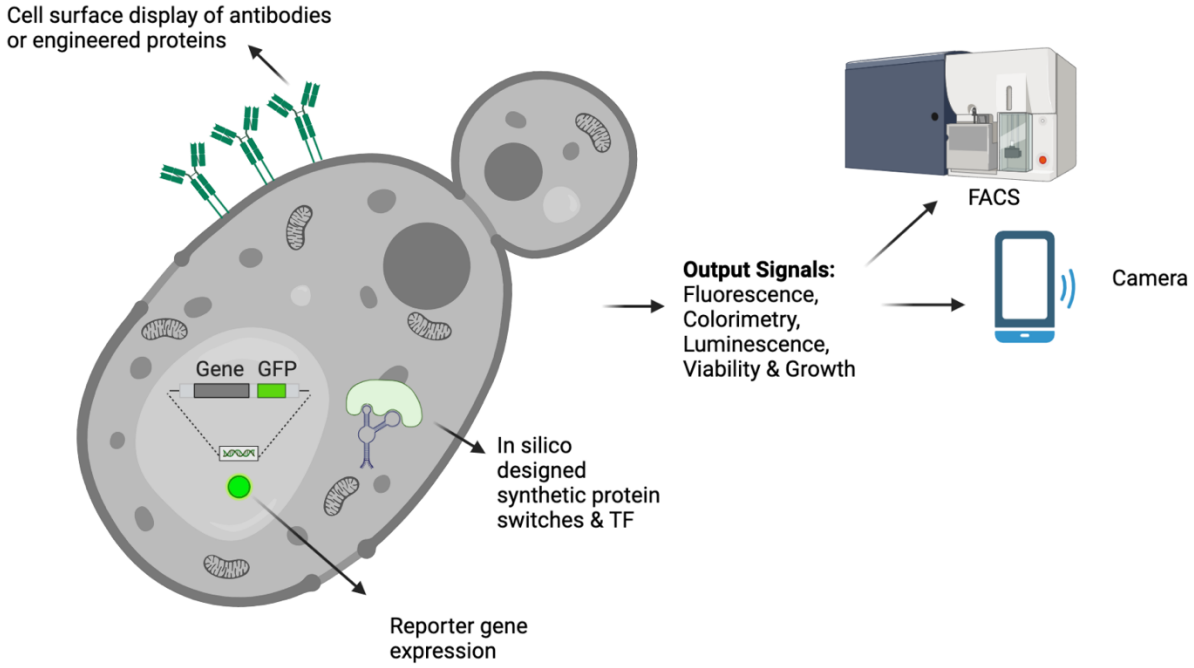

Figure 4: Role of yeast in biosensor design. Microorganisms such as yeast which can respond to an environmental change are used in a variety of ways for generating biosensors. Yeast display techniques involve expression of antibodies or engineered proteins on the surface which bind with the target analyte. Biosensors that use transcription factors as their signal transducing element rely on a promoter from a gene that has been fused to a protein that codes for fluorescence, bioluminescence, selective markers, or growth rate. This promoter initiates transcription when the target molecule is present. If the output signal is fluorescence, the signal can be detected via Fluorescence Activated Cell Sorting (FACS) or flow cytometry. Synthetic protein switches and riboswitches based on aptamers can be introduced in yeast cells where they can interact with the target ligand and produce a conformational or localisation change to produce an output signal. Created with BioRender.com

Immunobiosensors such as ELISA and SPR utilise recombinant antibodies or antibody fragments as a biorecognition element to sense target analytes[14]. Antibodies are highly specific and have a faster response time, however sample complexity and secondary antibody cross-reactivity can produce non-specific binding and give false positive results[45]. A major drawback of most immunosensors is the high cost and short shelf-life of antibodies. Moreover, antibodies do not exist for every molecule of interest, especially single atoms. Antibodies are difficult to generate utilising multiple screening of animal hybridomas with western blotting, while synthetic antibodies are challenging to isolate using phage and yeast display.

A critical element in the designing of novel biosensors is the development of highly specific recognition molecules to produce a specific signal. Today antibodies are being replaced by synthetic receptors such as peptide and oligonucleotide aptamers. They are sensitive at nanomolar or lower levels and have a fast response time of milliseconds[23]. However, a significant drawback of detecting aptamers for small molecule binding domains by using SELEX is the immobilization of the small molecule targets on agarose, magnetic beads or sepharose. Immobilisation usually recruits a functional group for conjugation which could contain the binding site residues for facilitating binding with the aptamer or many small molecules might not possess residues for immobilisation. Due to the small size of the aptamers, it is also common for the binder aptamers to possess binding affinity to the immobilisation matrix. A bottleneck in the SELEX technique is also the multiple selection and counter selection (with similar targets) required to find specific binding domains.

Both transcription-based factor and synthetic protein switches require specific recognition domains and are limited by difficulty of engineering proteins. It would be ideal to create peptide/protein pairs to interface with the target analyte in a non-competitive manner. The peptide/protein pair could then be integrated in the transcription-factor based or synthetic protein switch architecture.

1.5 Methods for identifying biorecognition elements

Nature has evolved many protein domains that bind metabolite or other proteins, such as calmodulin, PDZ domain, galectins, zinc finger domains, FKBP, however there are not enough known domains to generate biosensors for all desirable target molecules. Biosensors have a range of transduction and signal outputs but the primary limiting factor with the use of modular biosensors is the small number of analytes that can be detected using these systems. Hence, there is a need to discover binding domains which can expand the repertoire of target analytes which can be detected.

1.5.1 Small molecule and protein microarrays

Microarray technology allows screening of a broad range of biomolecules such as aptamers, proteins, small molecules, and oligonucleotides to identify protein-protein and protein-ligand interactions and determine intramolecular dynamics[23]. The microarrayed molecules are immobilised on a solid support that can be imaged when other molecules are washed over the array. For example protein microarrays immobilize protein spots to detect molecules which bind to the sample[46]. Identifying small molecule specific protein binding domains is mostly achieved by fluorescence-based detection

methods where the protein is either labelled by fluorescent tags such as GFP or it is labelled with a secondary probe which binds with the protein[47]. However, small molecule microarray fluorescence-based detection has some drawbacks. For example, the efficiency of labelling varies depending on the protein and addition of the fluorescent tags require extra steps, directly labelling the protein can alter the protein structure and its binding affinity, specific conditions such as incubation time, protein concentration, and buffer composition need to be optimised and can result in incomplete characterisation of the protein-ligand interaction. Possibly most importantly, microarray methods can only generate one binding domain for the target analyte. Unless the binding domain undergoes a conformational shift upon ligand binding, significant further engineering is required to establish signal transduction in the context of a biosensor.

1.5.2 Yeast-two hybrid

A system which exploits the modular nature of the transcription factor is the yeast two hybrid system (Y2H) which has been widely utilised for identifying protein-protein interactions. Yeast two hybrid depends on splitting the DNA binding domain (DBD) and activator domain (AD) of the Gal4 transcription factor[49]. The protein that is fused to the DBD is referred to as the "bait" protein and the protein that is fused to the AD is referred to as the "prey" protein[50]. The bait protein is typically he protein that is being tested for interactions with other proteins, while they prey protein is used as a tool to identify these interactions. On interaction of the two proteins, the transcription factor reconstitutes and recruits the RNA polymerase to translate the reporter gene present downstream[49]. Y2H can screen a single prey at a time or a library of prey peptides/proteins. Additionally, it is low throughput, missing out on 75% of the known interactions and can only detect interactions *in vivo*, not in vitro[51]. Y2H screens are also prone to false negatives and false positive and lack accuracy and reproducibility[52].

1.5.3 Display techniques: Yeast Display and Phage Display

Combinatorial phage and yeast display are powerful, high-throughput tools for researchers for screening random peptide libraries to identify novel binding domains for target molecules[53, 54]. *S. cerevisiae* can express multiple proteins targeted to the cell surface with glycosylphosphatidylinositol (GPI) anchors. Some cell surface anchor proteins such as Cwp1p, Cwp2p, Tip1p, Flo1p, Sed1p are involved in flocculation, stabilisation of the cell wall and stress response such as pH change[55]. In yeast display technology, the cell surface proteins of the yeast are genetically engineered to express fusion proteins. Fusion proteins can produce up to 100,000 copies on the surface of the yeast[53].

To discover novel binding domains for biosensors, a library of peptides or proteins is usually displayed on the surface of phage or yeast and screened against the target analyte[56]. In phage display, proteins are fused with the coat protein of the bacteriophage and screened against the target via a process called biopanning[56] (Fig. 5). The phage display library is washed over a target analyte which is immobilized on a solid surface[57]. Consequently, washing steps are conducted to wash away the non-binders[57]. High-affinity phages are purified and sequenced to find the binding domains. This method has been successfully utilised for finding novel peptides, binding domains, and antibody fragments for developing biorecognition components of biosensors. In yeast display, the target analyte is conjugated with a fluorescent marker and incubated with the yeast population. Yeast cells which possess proteins with affinity to the target ligand then bind and can be enriched and separated using FACS (Fluorescence Activated Cell Sorting). The population is then sorted utilising FACS (Fluorescence Activated Cell Sorting) to identify binders of the target analyte[58].

Despite their long history of success in allowing identification of novel ligand binding domains and peptides, phage and yeast display techniques have significant limitations. For example, target molecules must either be immobilised on a surface for phage display, or linked to a fluorescent molecule for yeast display, consuming functional groups that could otherwise be bound by library peptides. In addition, while these techniques allow for finding binding peptides from a mixed population, they are only able to assay the affinity of 1 peptide/protein at a time (one-dimensional) (Fig. 5). Binding peptides generated one-dimensionally can hinder the ability of a second binding pair to access the ligand. Display techniques are not optimized for generating compatible binding pairs of two simultaneously binding peptides to the target ligand, which is ideal for biosensor creation. Biosensor designs such as FRET pairs and split protein-switches are more readily engineered with colocalization domains as they require two binding domains that bind to different moieties on the target molecule. Thus, it would be ideal to develop a technique to find two binding peptides to the target analyte free in the solution to overcome these limitations.

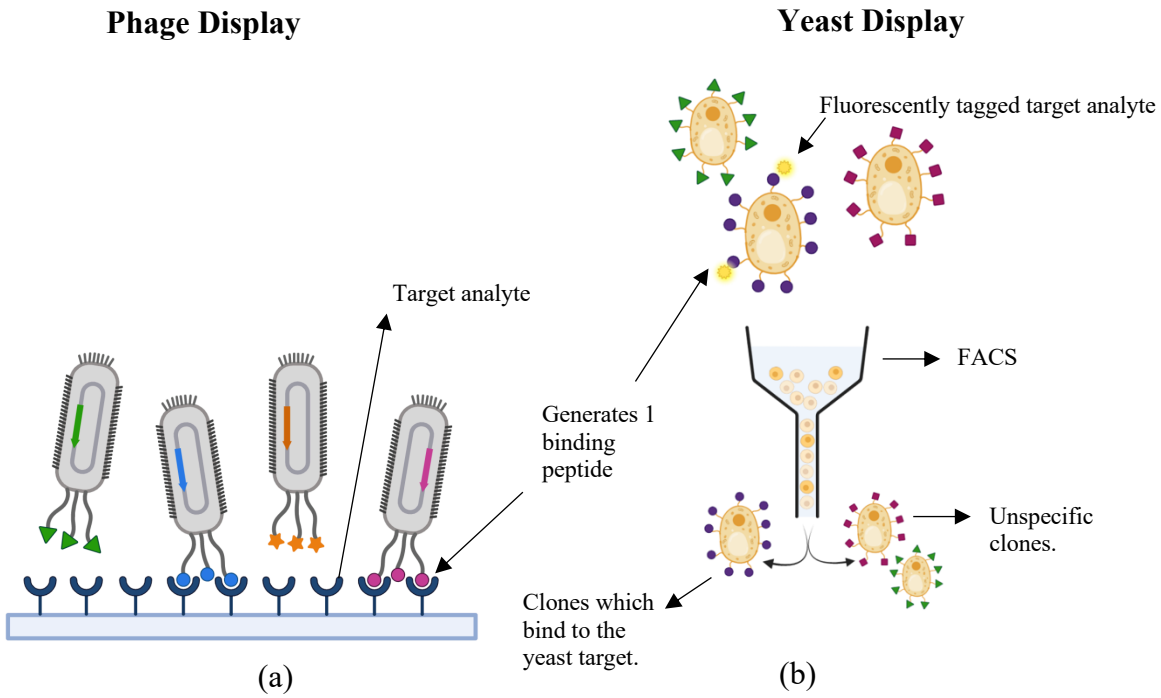

Figure 5: One dimensional identification of binding proteins/peptides utilising phage display and yeast display techniques for assessing the affinity to the target ligand immobilized on a matrix. (a) In phage display, a combinatorial library of proteins is expressed as a fusion to capsid coat proteins. (b) In yeast display, the library is expressed as a fusion to Aga2p sexual agglutination protein and incubated with the target of interest conjugated with a fluorescent marker. FACS (Fluorescence Activated Cell Sorting) is used to separate yeast cells which possess proteins with affinity to the target ligand to identify binders. Created with BioRender.com

1.6 Synthetic Yeast Agglutination

There is an existing yeast display technique that could potentially be applied to generate simultaneous ligand binding proteins, providing a solution to one of the critical limitations in biosensor development. The system, developed by David Younger, utilises sexual agglutination of *Saccharomyces cerevisiae* to identify protein binders for screening protein-protein interactions[59]. This technique could hypothetically be used for screening protein-ligand-protein interactions, however that has not been reported yet.

Yeast mating is a biological phenomenon which has been utilised for many applications such as protein-protein interaction, screening drug candidates, and detecting extracellular targets. The mating type of *S. cerevisiae* is determined by the genetic composition of the *MAT* locus[60]. *MAT*a haploids express the genes *MATA1* and *MATA2* and *MATα* haploids express *MATALPHA1* and *MATALPHA2* from the *MAT* locus[61]. Yeast mating is initiated by a "shmooing" behaviour where cells in proximity

make direct contact with one another[62]. The cells can also release mating factor *a* and *α* which induces haploid cells to express specific sexual agglutinin proteins. The mating factor increases the abundance of sexual agglutinin proteins from 0-10^4 molecules per cell to 10^4-10^5 molecules per cell to initiate the shmoo formation between the opposing mating types[55]. This allows for strong interaction and binding between many pairs of agglutinins between two haploid cells to initiate mating. In native sexual agglutination of yeast, *AGA2* expressed by *MAT*a cells and *SAG1* is expressed by *MAT*α extend outside the cell with GPI anchors and initiate binding of the cells. In *MAT*a, the sexual agglutinin protein consists of two subunits: Aga1p and Aga2p; Aga2p forms a disulfide bond with Aga1p cell surface protein[59]. Aga1p is only responsible for anchoring Aga2p but doesn't have a direct role in the binding. However, Sag1p and Aga2p interaction is essential in sexual agglutination as Aga2p carries the adhesive domain for the binding with the α-agglutinin[59](Fig. 6a).

Younger et. al, knocked out *SAG1* from *MAT*α to cease native mating[59]. Next, complementary synthetic proteins fused to *AGA2* were expressed in both the mating types (Fig. 6b). This only allowed cells with complementary synthetic proteins to mate again due to successful cell adhesion. Non-complementary synthetic proteins were unable to recover mating. Hypothetically, direct agglutination maybe be possible by expressing compatible binders to a target analyte, such that the agglutination and mating only occurs in the presence of the target, however this has not been reported to date. In such a system, mating efficiency would be low in the absence of target and high when the target is present. Yeast display conducted in this way is optimized for searching co-binding peptides/proteins ideal for biosensor production (Fig. 6c).

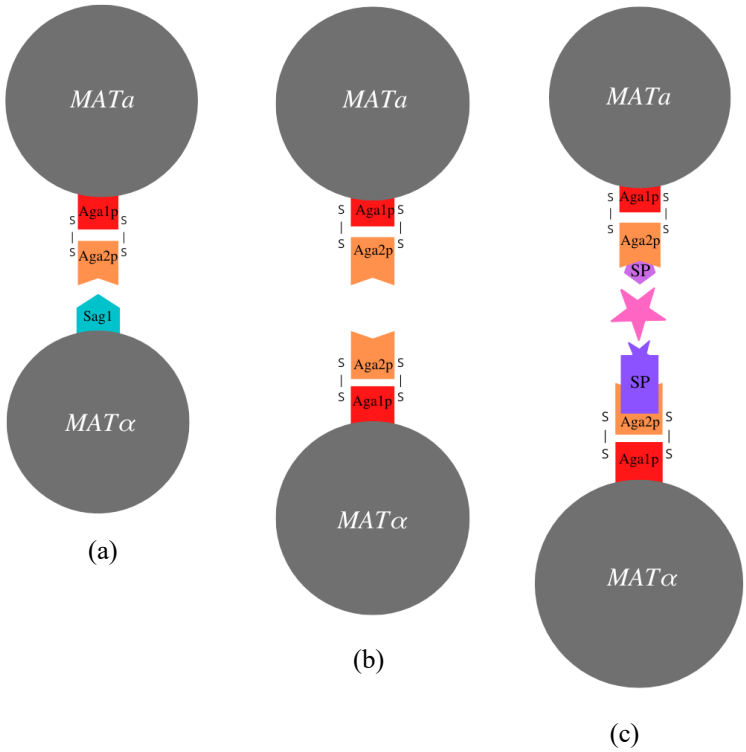

Figure 6: Adaptation of synthetic yeast agglutination method to find binding domains for a target analyte. (a) shows the native sexual agglutination in yeast requiring Aga2p linked to Aga1p in *MATa* and Sag1p in *MATα* for cell adhesion. (b) shows unsuccessful mating of the yeast cells as Sag1p is knocked out from *MATα* and Aga1p and Aga2p are expressed on the surface in *MATα* (c) shows synthetic proteins fusions to Aga2p, non-native to yeast by replacing the endogenous agglutination proteins in both the mating types. Only complementary synthetic proteins to the target ligand will be able to mate.

1.7 Simultaneous Yeast Display (SYD): A platform for generating novel biosensor binding domains

Here we propose a system termed Simultaneous Yeast Display (SYD) that overcomes the limitation of generating high-quality pairs of binding domains for target analytes in *S. cerevisiae*. SYD aims to restrict mating in opposing mating type yeast strains, forcing cell mating to only occur upon introduction of the ligand and cell-cell fusion via displayed protein domains or peptides. We hypothesize that this will allow identification of complementary binding domains expressed on the surface of each mating type, which bind to the ligand. The binding of the peptide/proteins to the ligand will create a sandwich-like configuration, bringing the two yeast strains in close proximity. This will initiate cell agglutination which induces the shmooing process for diploid formation. Diploids can then be isolated or enriched via co-selection for different marker genes that were

present in haploid cells. The identity of pairs of ligand binding peptides can then be identified by sequencing the surface display expression constructs from isolated or enriched diploids.

To confirm this hypothesis, we explored an existing method utilising synthetic yeast agglutination described above. We tested the methodology for detecting protein-ligand-protein interactions for biosensor development. This method is based on validated research on yeast surface display which has been used for decades for engineering peptides and proteins[59]. We constructed yeast strains to test combinations of known protein-ligand-protein interactions. We utilised binding domains Pro1a and Pro1b which bind to progesterone and FKBP-rapamycin domain (FRB) and a 12k-Da FK506 binding protein (FKBP12 or FKBP) which bind to rapamycin in a sandwich like fashion. Pro1a and Pro1b are computationally designed progesterone-dependent heterodimers[63]. Both domains have been used in a transcription-based biosensor previously. FKBP12-rapamycin-FRB complex ($K_d \approx$ 2.5 nm) is a strong ligand dependent protein heterodimerization[64]. Previous enzymatic studies have shown that FKBP and FRB do not interact in the absence of rapamycin. Thus, it is a perfect candidate for cell surface localisation fusion protein to test the binding affinity in the absence and presence of rapamycin[64]. The ultimate aim of SYD is to screen diverse haploid cell libraries of displayed peptides or proteins such as NNK libraries (where N = A/C/G/T, B = C/G/T, S = C/G, and K = G/T) to generate compatible binders for target ligands (Fig. 7). An NNK library is a fourteen amino acid peptide library which can encode all 20 amino acids at each position, but still encodes a stop codon 3% of the time. An advantage of using smaller random amino acid peptides is higher suitability to simultaneous binding due to the smaller size of the peptides[,65]

1.7.1 Advantages of SYD

SYD is theoretically superior in comparison to existing techniques for identifying ligand binders as it allows simultaneous binding of two peptides to a ligand. Most biosensor designs such as FRET pair or split glucose dehydrogenase system require two ligand binding domains. Most existing tools employed to design novel binding domains require in-depth literature evaluation and computational design. We have validated a method which can identify complementary binders *in vivo* which are unlikely to be obtainable any other way. SYD is a platform technology which allows for rapid detection of novel binding pairs in a matter of days. Domains identified from SYD can allow easy integration with existing biosensor designs as they are complementary to the ligand. Another significant advantage of using SYD is comparison to traditional display techniques is the ability to detect ligand binding domains in a liquid culture, eliminating the need for ligand conjugation or immobilisation. This is problematic as it can alter the binding site of the target ligand or other

associated characteristics and interfere with finding true binders. Diffusing the ligand in media is advantageous as it preserves the native structure and characteristics.

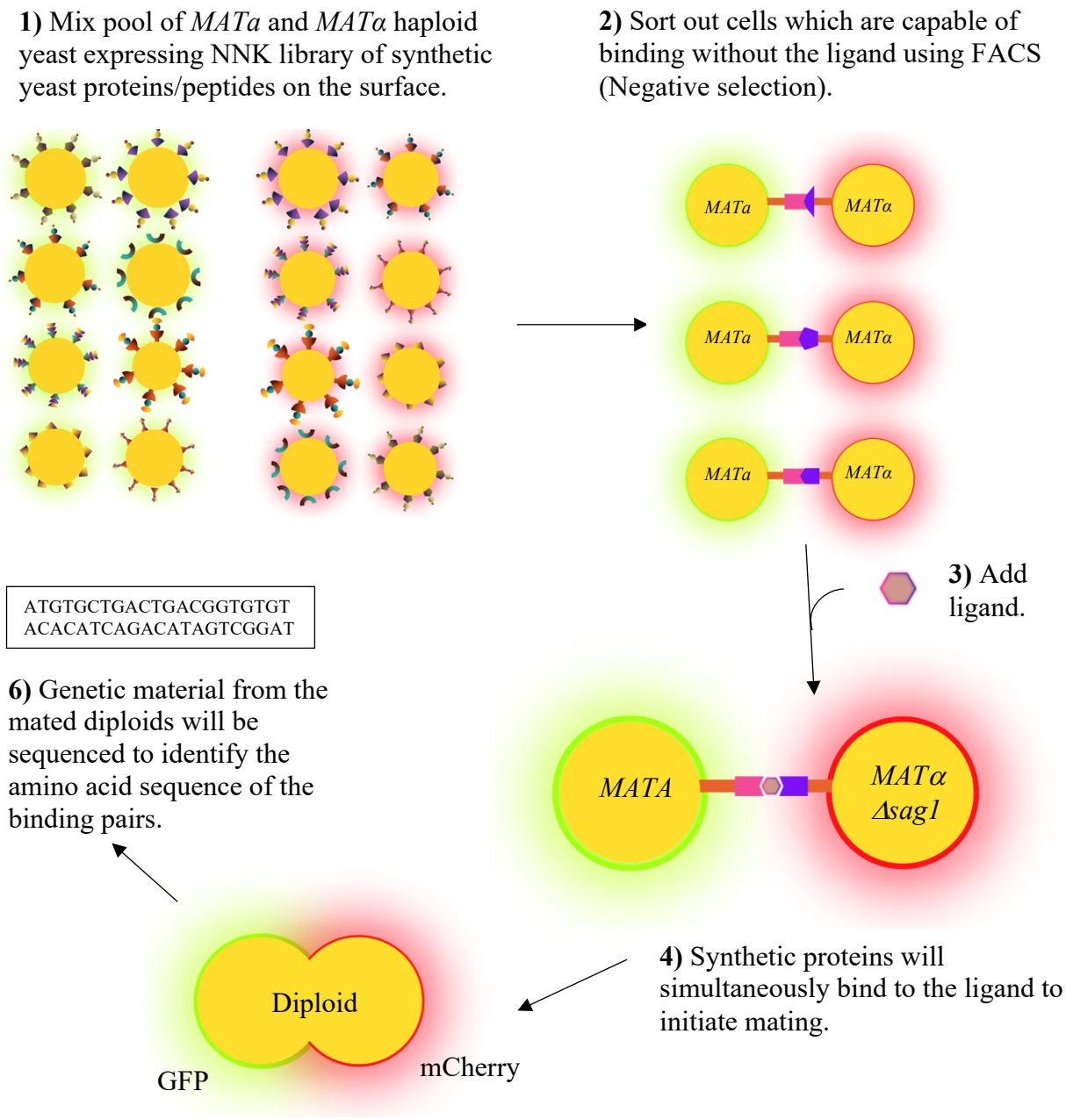

Figure 7: Workflow of Simultaneous Yeast Display (SYD). A synthetic NNK library is expressed on the surface of yeast MATa and MATα cells. The populations undergo negative selection in the first sorting step, where cells which bind to each other without the ligand, are separated out using FACS with gates for cells that are only expressing GFP or mCherry. The yeast cell population is then exposed to the ligand. Synthetic proteins which simultaneously bind to the ligand initiate the mating process to form diploids. The diploids express the GFP and mCherry fluorescent marker and are separated using FACS in the second sorting step. The strains are sequenced to identify the binding domains.

Another advantage of SYD over conventional display methodologies which utilise FACS for identifying fluorescently labelled ligand bound to surface displayed peptide/protein, is the growth-based selection strategy. Yeast display libraries are usually incubated with a fluorescently labelled ligand and analysed using FACS and deep sequencing. While our strains are fluorescent and can be selected via FACS, we utilised growth dependent on selectable markers for identifying diploids. We have chosen yeast strains where diploids can be isolated by growth on selective media, thus eliminating the need to conduct flow cytometry. A growth-based selection strategy is cheaper and easier to scale up and maximise screening efficiency. A major limitation of flow-based cell sorting is the number of assessable diploids per day (tens of millions of cells). Growth-based selection schemes have higher throughput as they allow screening of billions of cells in only a few days. It also requires cheaper equipment as the steps only include incubation in growth media in comparison to FACS which utilises more expensive equipment. Previous studies have also reported a similar amount of time required for plating vs. flow cytometry[66].

2. Methods and Materials

2.1 *E. coli* and *S. cerevisiae* media

E. coli DH5α was used for plasmid construction and storage with growth at 37°C in Luria Broth supplemented with 0.1 mg/mL ampicillin for plasmid selection and maintenance. The *S. cerevisiae* strains used were BY4741, BY4742, and BY4742 Δ*sag1*. The *S. cerevisiae* strains were grown at 30°C in YPD medium (1% yeast extract, 2% peptone, and 2% dextrose in distilled water), unless stated otherwise.

2.2 Construction of Plasmids

2.2.1 DNA Purification

A Monarch Plasmid Miniprep Kit (New England BioLabs) was used to purify plasmid DNA from *E. coli* as per the manufacturer's instructions. DNA fragments used in plasmid assembly were amplified using New England BioLabs Q5 High-Fidelity 2X master mix as per the supplier's instructions. Colony PCR was conducted using GoTaq DNA Polymerase (Promega Corporation) according to suppliers' instructions. PCR products were purified using the QIAquick PCR purification kit (Qiagen) or the Zymoclean Gen DNA Recovery Kit (Zymo Research) or Monarch DNA gel extraction kit (NEB). Primers and gBlocks were designed using the software Geneious Prime® 2022.1.1 and supplied by Integrated DNA Technologies or Genewiz. Table 1 contains a complete list of primers used in this study. Sanger sequencing was carried out by Genewiz and the

sequence data was analysed using the Geneious Prime® 2022.1.1. Plasmids were designed under a schema of modular parts including promoter components such as *TEF1* and the *TDH3* promoter, *ADH1* and *CYC1* terminators, a red fluorescent protein (mCherry) and enhanced green fluorescent protein (eGFP). Gibson Assembly and restriction enzyme/T4 DNA ligase assembly was utilised to construct the plasmids.

2.2.2 Cloning of fluorescent genes

For construction of pTEF1-AGA2-FRB-415 (Table 1), a gBlock with the *FRB* was synthesised and ordered from IDT. Overhang primers RS21 and RS22 (Table 2) were used for PCR amplification of the gBlock to create homology to the destination plasmid (Fig. 8). The destination plasmid pRS416, containing the uracil auxotrophy complementing gene URA3, was linearised using the *Sma1* restriction enzyme. The linearised plasmid and amplified gBlock were assembled using the Gibson Assembly mix "NEBuilder HiFi DNA Assembly Cloning Kit" prior to transformation in *E. coli* DH5α. Colony PCR with primers RS03 and RS04; M13 forward and M13 reverse was conducted to confirm the insertion of the gBlock in the plasmid, with the insert sequence verified via Sanger sequencing. Similarly, for the construction of PRS01, the destination pRS413 plasmid containing a histidine auxotroph complementing gene *HIS3,* was linearised using the *Sma1* restriction enzyme. mCherry was amplified from the PLAC7 plasmid using RS27 and RS28 primers and Gibson assembled in linearised pRS413 prior to transformation in *E. coli* DH5α. mCherry insertion was checked visually under blue light and confirmed with colony PCR using M13 forward and M13 reverse primers.

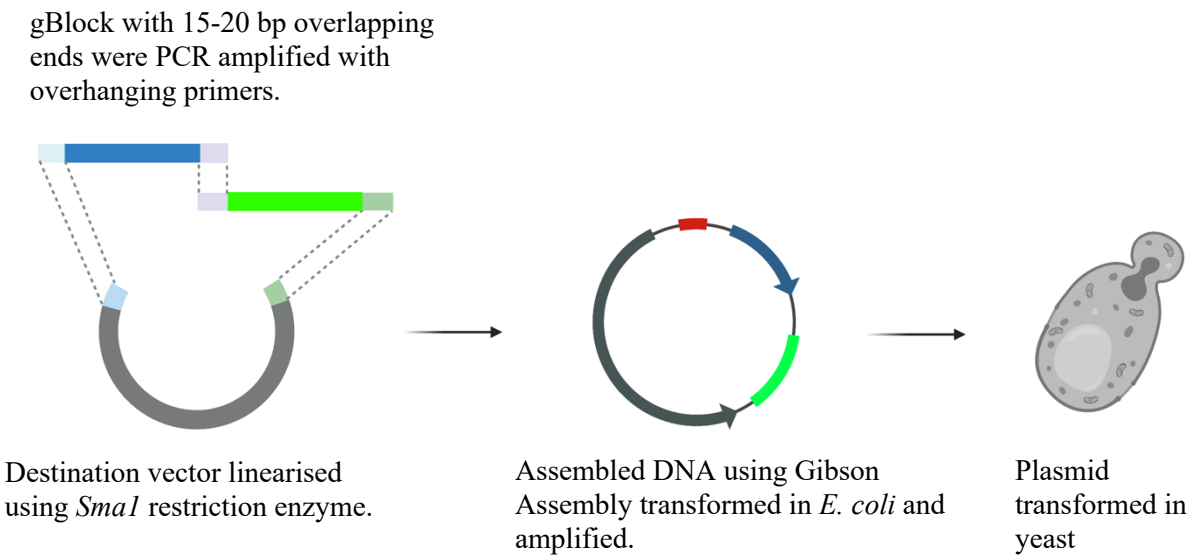

Figure 8: Workflow of the Gibson assembly reaction. Overhang primers with homology to the gBlock and the destination vector were designed with a 15-20 bp overlap. The gBlock was PCR amplified and the destination vector was linearised using *Sma1* restriction enzyme for Gibson assembly reaction. The assembled DNA was transformed and amplified in *E. coli* DH5α and extracted for transformation in yeast. Created with BioRender.com

2.3 Measurement of cellular concentration

Unless otherwise stated, measurement of cellular concentration was performed using a Ultraspec 10 spectrophotometer and optical density (OD) was recorded at 600 nm (OD_{600}). MilliQ water was used as a blank for measuring the cultures and to dilute samples so that readings fell within the linear range of the spectrophotometer (0.05 to 0.5).

2.4 Transformation Protocol

Transformation protocol used was described by Ellis lab[67]. A 5 mL culture of YPD medium was inoculated with a single colony of BY4741 and BY4742 Δ*sag S. cerevisiae*. The inoculated media was incubated at 30°C at the speed of 200 rpm in a shaking incubator overnight. The OD_{600} of the culture was measure and diluted to an OD_{600} of 0.5 in a baffled 250 ml shake flask with 25 mL of YPD medium. The culture was grown for 4 hours at 30°C and 200 rpm. Post incubation, the contents of the shake flask were transferred to a Falcon tube and centrifuged at 2000 rpm for 10 minutes at room temperature. The cells were resuspended in 5 mL of 0.1 M lithium acetate and 7.5 ml of sterile milliQ water and spun again at 2000 rpm for 10 minutes at room temperature. The cells were transferred to a sterile 1.5 mL microcentrifuge tube and resuspended in 100 µL of 0.1 M

lithium acetate, 10 µL freshly boiled herring sperm DNA (10 mg/mL), and 1 µg of transformation DNA. After 30 minutes of incubation at room temperature, 600ul of 50 % PEG-3350, 90 µL of 1M lithium acetate, 100 µL of DMSO was added to the cells. After a brief vortex, the cells were incubated for 30 minutes at room temperature (RT). The cells were heat shocked at 42°C for 7 minutes. After a brief vortex, the cells were heat shocked for 7 minutes again. The cells were centrifuged for 2 minutes at 4200 rpm and resuspended in 250 µL of 5 mM calcium chloride. After a 10-minute incubation at RTP, the cells were washed with 1 mL of 1X phosphate-buffered saline (PBS) containing 8 g/L NaCl, 0.2 g/L KCl, 1.42 g/L Na_2HPO_4, 0.24 g/L KH_2PO_4 at a final pH of 7.4. The cells were centrifuged at 4200 rpm and allowed to recover in YPD medium for 4 hours. The cells were centrifuged and plated with 100 uL of PBS buffer onto selective media plates containing respective selectable marker and Zeocin (200 µg/mL) (Table 1). Plates were incubated for 2 days at 30°C.

2.4 Mating assay

2.5.1 Liquid culture mating assay

The purpose of this experiment was to evaluate the diploid formation rate with yeast strains containing plasmids which express surface proteins which are compatible binders to small molecule progesterone present in liquid culture. Triplicate cultures were grown in 50 mL Falcon tubes in their respective selective media to maintain the introduced plasmids and minimise plasmid loss. All cultures were grown in 1X Yeast nitrogenous base (YNB) (Sigma Aldrich) with 10 g/L glucose and supplemented with amino acids 20 mg/L uracil, 50 mg/L histidine, 100 mg/L leucine as required to maintain plasmid selection. The Falcon tubes were incubated overnight at 30°C with shaking at 200 rpm. 100 µL of each culture was taken for flow cytometry analysis. Each BY4741 culture was diluted to an OD_{600} of 0.01 and BY4742 *Δsag* was diluted to an OD_{600} of 0.03 and mixed for mating in 3 mL yeast peptone dextrose (YPD) rich media in triplicates. Mating assay was conducted with BY4741 strain expressing Aga1p, Aga2p-Pro1a and eGFP and BY4742 *Δsag* strain expressing Aga1p, Aga2p-Pro1b and mCherry to assess protein-ligand-protein interaction (Table 3). Three mating assays were performed to test the mating efficiency in the absence of progesterone, presence of 0.1 µM progesterone and 1 µM progesterone. A mating assay conducted as a positive control comprised wild type BY4741 and BY4742 and for the negative control wild type BY4741 and BY4742 *Δsag* (Table 3). All 3 mL cultures inoculated with opposing mating type strains were incubated at 30°C with shaking at 275 rpm for 4 hours and 17 hours, respectively.

2.5.2 Colony counting

The OD_{600} of the mating assays was measured and normalised to the lowest OD_{600} recorded amongst the experimental cohort. For the 4-hour mating assay the final absorbance was normalised at OD_{600} 1 and for 17-hour mating assay the final absorbance was normalised to OD_{600} 5 in order to ensure equal number of cells are plated from each assay. Samples were spun down and washed with PBS buffer to remove the rich media. All samples were plated in a dilution series of OD_{600} :1, 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4} in PBS buffer, and 100 µL of each dilution was plated on solid SD media supplemented with leucine (100 mg/L) and zeocin (100 µg/mL) and 2% w/v agar. Control samples were plated on YNB agar plates (2% w/v agar) supplemented with 50 mg/L histidine, 100 mg/L leucine, 100 mg/L lysine and 100 µg/mL zeocin to complement auxotrophies and maintain plasmids accordingly. After a 2-day outgrowth, the number of colony forming units per population was estimated by counting colonies at a dilution series of 10^{-1} to assess the number of diploids formed from the mating assay and representative pictures were taken. Dilution series 10^{-1} was taken for the 17-hour mating assay because it had the most number of countable colonies. For the 4-hour mating assay, OD was normalised to 1 and plated on the selective medium.

2.6 Flow cytometry

Samples taken during the liquid culture assay at the stationary phase were assessed using flow cytometry as follows. 100 µl of sample was diluted 100 µl of PBS and transferred to a 96-well flat bottom plate for flow cytometry analysis. Each well was analysed for 10,000 events using BD Accuri C6 flow cytometer with a 488 nm excitation laser and a 533/30 filter for BY4741 strains and 670 LP for BY4742 *Δsag* strains. Samples were normalised by division to the non-fluorescent strain BY4741.

2.7 Statistical Analysis

All data was analysed using Microsoft Excel Version 16.60 and GraphPad Prism 7 software. Mean and standard deviations were calculated and statistical significance was tested using a non-parametric T-test.

Table 1: Plasmids used in this study

Plasmid	Function of plasmid	Source
pTEF1-mCherry-416	Shuttle vector with a constitutive *pTEF1* promoter and *mCherry* inserted at the MCS; Selectable markers are ampicillin and histidine.	Paulsen laboratory
pTEF1-GFP-416	Shuttle vector with a constitutive *pTEF1* promoter and *eGFP* inserted at the MCS; Selectable markers are ampicillin and uracil.	Paulsen laboratory
pTEF1-Pro1b-416	Shuttle vector with a constitutive *pTEF1* promoter and gene coding for progesterone binding protein, Aga2p-Pro1b[63] inserted at the MCS; Selectable markers are ampicillin and uracil.	Paulsen laboratory
pTEF1-Pro1a-pBR322	Shuttle vector with a constitutive *pTEF1* promoter and gene coding for progesterone binding protein, Aga2p-Pro1a [63] inserted at the MCS; Selectable markers are ampicillin and methionine.	Paulsen laboratory
pTEF1-mCherry-413	Shuttle vector with a constitutive *pTEF1* promoter and *mCherry* inserted at the MCS; Selectable markers are ampicillin and histidine.	This thesis
pTEF1-AGA2-FRB-415	Shuttle vector with a constitutive *pTEF1* promoter and *FRB* inserted at the MCS; Selectable markers are ampicillin and uracil.	This thesis
pTEF1-AGA2-FKBP-pBR322	Shuttle vector with a constitutive *pTEF1* promoter and *FKBP* inserted at the MCS; Selectable markers are ampicillin and methionine.	This thesis
pTDH3-AGA1-Zeo-413	Shuttle vector with a constitutive *pTDH3* promoter and *AGA1* inserted at the MCS; Selectable markers are antibiotics ampicillin, zeocin and histidine.	Synthesised by GeneWiz
PRS415	*LEU2* marker cloning/ expression vector	Euroscarf[68]
PRS416	*URA3* marker cloning/ expression vector	Euroscarf[68]
PRS413	*HIS3* marker cloning/ expression vector	Euroscarf[68]

Table 2: Primers and oligos used in this study

Name	Sequence	Description
RS01	GGTGTCCAAGTTGAAACCATCT	Forward primer to check FKBP gBlock insertion in the pTEF1-AGA2-FKBP-pBR322 plasmid using Colony PCR
RS02	CCAACTTTAATAGTTCGACATCG	Reverse primer to check FKBP gBlock insertion in the pTEF1-AGA2-FKBP-pBR322 using Colony PCR
RS03	TATTTTGTGGCACGAAATGTGGCA	Forward primer to check *FRB* gBlock insertion in the pRS415 plasmid using colony PCR
RS04	GATTCTACGGAAAACGTGGTAATA	Reverse primer to check *FRB* gBlock insertion in the pRS415 plasmid using colony pcr
RS05	CGCAAATATGTAGTAATACGTGGGA	*PRB* integration reverse check primer
RS06	AAAAGGGGGCCAATGTTACG	*PRB* integration forward check primer
RS07	CGACTTGTAACCTCGAGACGCCTAA GGAAAGAAAAAGAAAAAAAAAAGC AGCTGAAATTTTTCTAAATGAAGAA TTATGCAGCTAAGGTAATCAGATCC	Homology based 80 base pairs overlap PCR for knocking out *PRB* and inserting *AGA1*
RS08	AAACTTAAGAGTCCAATTAGCTTCA TCGCCAATAAAAAAACAAACTAAAC CTAATTCTAACAAGCAAAGATGGAA TAAAAAACACGCTTTTTCAGTTCG	Homology based 80 base pairs overlap PCR for knocking out *PRB* and inserting *AGA1*
RS09	AGAACTAGTGGATCCCCCTGCAGCT AAGGTAATCAGATCCA	Zeocin Reverse primer
RS10	TAGCACAGTGATCCTTCAGTAATGT CTTGTTTCT	Zeocin Forward primer
RS11	TGAAGGATCACTGTGCTAATACTCC TCTTCCC	*AGA1* reverse
RS12	AAACAAACAAAATGACATTATCTTT CGCTCA	*AGA1* forward

RS13	GATAATGTCATTTTGTTTGTTTATGT GTGTTTATTCGA	*pTDH3* reverse primer
RS14	ATCGAATTCCTGCAGCCCGAATAAA AAACACGCTTTTTCAGTTCG	*pTHD3* forward primer
RS16	CGA TGT CGA ACT ATT AAA GTT GGA ATA ACG AAT TTC TTA TGA TTT ATG ATT TTT ATT AT	Forward Primer for insertion of *FKBP*
RS17	ATA ATA AAA ATC ATA AAT CAT AAG AAA TTC GTT ATT CCA ACT TTA ATA GTT CGA CAT CG	Reverse Primer for insertion of *FKBP*
RS18	CTG GTT CTG GTG GTG GTG GTG GTG TCC AAG TTG AAA CCA TC	Forward Primer for insertion of *FKBP*
RS19	ACA AAA TAG CGA CTC TGA TCA ATT CAC CAC CAC CAC CAG AAC	Reverse Primer for insertion of *FRB* gblock using Gibson assembly insertion in pRS415
RS21	ATA AAT CAT AAG AAA TTC GTT ATG TTT GGA GAT TCT ACG GAA AA	Reverse Primer for insertion of *FRB* gblock insertion in pRS415
RS22	CTG GTT CTG GTG GTG GTG GTG AAT TGA TCA GAG TCG CTA TTT T	Forward primer for insertion of *FRB* gblock in pRS415
RS23	ATTAGTTAATCCCAACAATATTGTG	To check the insertion of the promoter *ptDH3* in the pRS413 plasmid
RS24	ATTAGGAAAGTTGGTCTTCGCC	To check insertion of both *AGA2* and Zeocin resistance gene in yeast strain; compatible with M13 reverse.
RS25	TCGTGGTTGGGAGGAAGGTG	Diagnostic forward primer for checking *FKBP* insertion.
RS26	CGCCCAATACGCAAACCGC	Diagnostic reverse primer for *FKBP* on the backbone.
RS27	CTAGAACTAGTGGATCCCCCGAGCG ACCTCATGCTATACCT	Reverse primer for *mCherry* integration with pRS413
RS28	GATATCGAATTCCTGCAGCCCGCAC ACACCATAGCTTCAAA	Forward primer for *mCherry* integration with pRS413

RS29	TGTTTTTTATTCTTCTTCATTTAGAAAAATTTCAGCTGCT	Reverse primer for 1st fragment of the 500 base pair overlap for genomic insertion of pTDH3-AGA1-Zeo-413.
RS30	CTAAATGAAGAAGAATAAAAAACACGCTTTTTCAGTTCG	Forward primer for 2nd fragment of plasmid pTDH3-AGA1-Zeo-413 for genomic insertion.
RS31	CAAGCAAAGTGCAGCTAAGGTAATCAGATCCA	Reverse primer for 2nd fragment of pTDH3-AGA1-Zeo-413 for genomic insertion.
RS32	TTAGCTGCACTTTGCTTGTTAGAATTAGGT	Forward primer for integration of 3rd fragment of pTDH3-AGA1-Zeo-413 for genomic insertion.
RS33	AAATTACATATACTCTATATAGCAC	Reverse primer for integration of the 3rd fragment of pTDH3-AGA1-Zeo-413 for genomic insertion.
RS34	GTTGGAGCTTCTATCTTGAC	Diagnostic forward primer to check integration of 500 base pair homology for *AGA1* insertion
RS35	GAATTGCTTGCTTCTGTATCTATAC	Diagnostic forward primer to check integration of first fragment of 500 base pair homology for *AGA1* insertion
RS36	TTGCACCCGACAAATCAGC	Diagnostic reverse primer to check integration of the 500 base pair homology for *AGA1* insertion
RS37	GGCTCTGTCATCGCTCAACGGTTTTAGAGCTAGAAATAGCAAGTTA	Forward primer of CRISPR plasmid 1 for cut site for PRB deletion and *AGA1* insertion
RS38	CGTTGAGCGATGACAGAGCCGATCATTTATCTTTCACTGCG	Reverse primer of CRISPR plasmid 1 for cut site for PRB deletion and *AGA1* insertion
RS39	AAGCGATAGTACCGGCACAGGTTTTAGAGCTAGAAATAGCAAGTTA	Forward primer of CRISPR plasmid 2 for cut site for PRB deletion and *AGA1* insertion

RS40	CTGTGCCGGTACTATCGCTTGATCATTTATCTTTCACTGCG	Reverse complement of CRISPR plasmid 2 for cut site for PRB deletion and *AGA1* insertion
RS41	ACCTCGAGACGCCTAAGGGAATAAAAAACACGCTTTTTCAGTTCG	Forward primer complementary with RS31 (Second fragment of pTDH3-AGA1-Zeo-413 plasmid)
RS42	TTTTTTATTCCCTTAGGCGTCTCGAGGT	Forward primer for the first fragment of the 500 bp 3' region.

Table 3: *S. cerevisiae* strains used in this study

Strain Name	Genotype/Plasmids	Notes	Source
Strain 1: BY4741	*MATa, his3*Δ1, *leu2*Δ0, *met15*Δ0, *ura3*Δ0	Wild type *MATa* strain for positive and negative control of mating assay.	Euroscarf
Strain 2: BY4742	*MATα, his3*Δ1, *leu2*Δ0, *lys2*Δ0, *ura3*Δ0	Wild type *MATα* strain for negative control of mating assay.	Euroscarf
Strain 3: BY4742 Δsag::*KanMX*	*MATα, his3*Δ1, *leu2*Δ0, *lys2*Δ0, *ura3*Δ0	Wild type *MATα* strain with *Sag1* knocked out for positive control of mating assay.	Paulsen laboratory
Strain 4: BY4742 Δsag::*KanMX*	BY4742 Δsag pTEF1-Pro1b-416 + pTEF1-mCherry-416+ pTDH3-AGA1-Zeo-413	*MATα* strain with *Sag1* knocked out with plasmid coding for red fluorescent protein (mCherry), plasmid expressing *Aga1* and plasmid expressing fusion of *Aga2* and *Pro1b*.	This thesis
Strain 5: BY4741	BY4741 + pTEF1-Pro1a-PBR322 + pTEF1-GFP-416+ pTDH3-AGA1-Zeo-413	*MATa* strain for with plasmid coding for green fluorescent protein (eGFP), plasmid expressing *Aga1p* and plasmid expressing fusion of *Aga2* and *Pro1a*.	This thesis

3. Results and Discussion

3.1 Simultaneous Yeast Display (SYD): Platform for creating biosensor binding domain

Simultaneous Yeast Display (SYD) overcomes the limitation of generating complementary pairs of binding domains for target analytes by screening yeast displayed peptide libraries in liquid solution where cell mating is dependent on two peptides simultaneously binding the same target molecule, eliminating the need to immobilise the target molecule. An existing method utilising synthetic yeast agglutination to detect protein-protein interaction was employed to test protein-ligand-protein interaction. The method relies on eliminating native sexual agglutination in opposing yeast mating types to achieve agglutination and expressing synthetic proteins on the surface of *MATa* and *MATα* haploid cells.

3.2 Strain Design

In order to allow haploids of opposite mating types and diploids to be distinguished from one another using flow cytometry, constitutive expression constructs for green and red fluorescent proteins were introduced into *MATa* and *MATα* cells respectively. This was achieved by transforming strain BY4741 *MATa* with GFP expression plasmid pTEF1-GFP-416. To generate a mCherry expressing BY4742 *MATα* strain, we first needed to clone the red fluorescent protein mCherry in pRS416 vector. pTEF1-mCherry-416 was generated using Gibson assembly for producing red fluorescent protein mCherry (Fig 8). A set of forward and reverse primers containing homology arms were designed to amplify mCherry. The primers were designed to have homology between the destination vector and the mCherry. pTEF1-GFP-416 encoding green fluorescent protein and pTEF1-mCherry-416 were introduced in our background strains to allow separation using FACS. Similarly, to test protein-ligand-protein interaction for small molecule rapamycin, plasmids pTEF1-AGA2-FKBP-PBR322 and pTEF1-AGA2-FRB-415 were also built using Gibson assembly. FRB and FKBP heterodimerise in the presence of rapamycin and do not interact in the absence of rapamycin and hence make a good candidate for cell surface fusion protein expression. gBlocks were designed for expression of *FKBP* and *FRB,* and primers used to amplify the gblocks were designed to have homology between the gBlock and destination vector. These plasmids will be used in future experiments to validate SYD.

To eliminate the native sexual agglutination in yeast, the sexual agglutinin gene, *SAG1* was knocked out from wild type BY4742 (*MATα*). Proteins of interest that could potentially enable agglutination were then targeted to the cell surface by fusing them to Aga2p. Previous work has shown that

knocking out native sexual agglutination proteins encoded by *SAG1* and *AGA2* can decrease the efficiency of cell-cell fusion and mating[59,62]. To test the validity of this system, BY4741 and BY4742 Δ*sag1* were built to display complementary binding pairs. To test the capacity of SYD to select for protein-ligand-protein interactions, the heterodimeric binding pair Pro1a and Pro1b which is activated by small molecule progesterone engineered by Tinberg et. al. (shown in Fig. 9) was chosen[63]. BY4741 strain was transformed with plasmid pTEF1-Pro1a-PBR322 for cell surface localisation of the fusion protein Aga2p-Pro1a. BY4742 Δ*sag* was transformed with pTEF1-Pro1b-416 to surface display the fusion protein Aga2p-Pro1b. We overexpressed *AGA1* to achieve a higher efficiency of surface-display by introducing plasmid pTDH3-AGA1-Zeo-413 in all the strains. pTDH3-AGA1-Zeo-413 contains a strong constitutive promoter (pTDH3) driving expression of *AGA1,* and the Zeocin resistance gene to enable plasmid selection. A green fluorescent protein (GFP) was introduced in the BY4741 strain and a red fluorescent protein (mCherry) in BY4742 Δ*sag* strains to allow for detection of diploids when selecting via FACS. Diploids will produce both green and red fluorescence signal. Growth based selection of diploids was conducted using selectable markers leucine, lysine and zeocin.

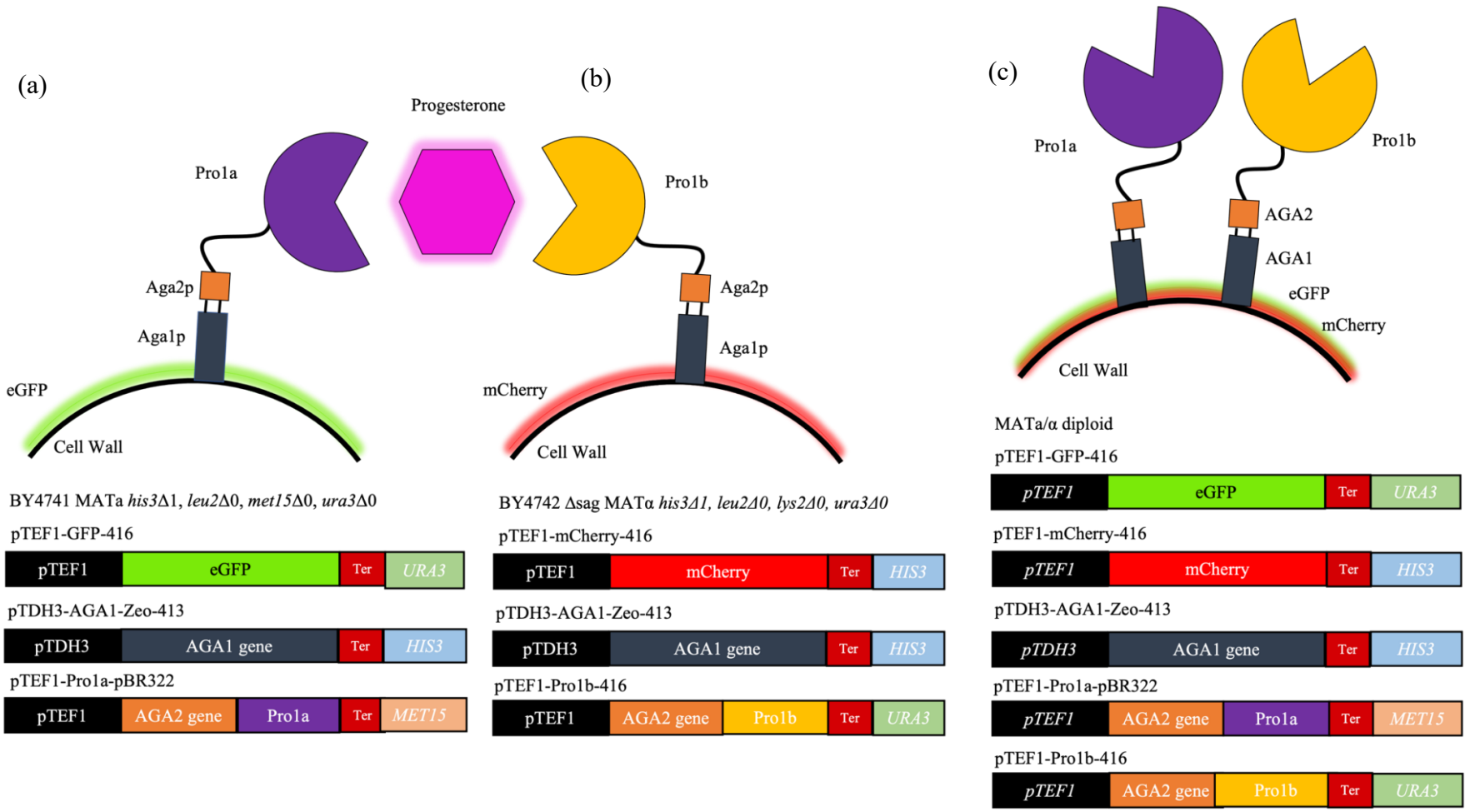

Figure 9: Yeast strains used in the mating assay. Colours of the gene correspond to the colour of the protein in the diagram. (a) BY4741 *MATa* strains contains plasmids pTEF1-GFP-416, pTDH3-AGA1-Zeo-413 and pTEF1-Pro1a-pBR322 for expression of eGFP, Aga1p, and Aga2p-Pro1a which complement uracil, histidine, and methionine. (b) BY4742 Δsag contains plasmid pTEF1-mCherry-416, pTDH3-AGA1-Zeo-413, pTEF1-Pro1b-416 for expression of mCherry, Aga1p, and Aga2p-Pro1b which complement histidine and uracil. c) Pro1a and Pro1b upon binding with small molecule progesterone initiate the sexual agglutination process. The diploids formed from mating of the two strains will contain the plasmids expressing binders Pro1a, Pro1b and both red and green fluorescence and can be selected on media supplemented with leucine and lysine.

3.4 Flow cytometry analysis of fluorescent yeast strains

The primary workflow to test a small molecule with known binders involved selection for diploids based on the presence of different selectable markers present in haploid cells surface-displaying complementary ligand-dependent protein dimers. The reason for performing growth-based assays to test out the initial hypothesis is the ease and speed of diploid detection via plating on selective solid medium. In addition, differentiating fluorescence signal on the flow cytometer can be harder due to the GFP leakage into the mCherrry channel, thus flow cytometry was not chosen as the primary method for selection. However, in future experiments, these strains will be utilised for expressing synthetic peptide/protein libraries so that negative selection can be conducted with flow cytometry prior to introduction of the ligand for characterising protein-ligand-protein interactions. In practice this would involve incubating haploid peptide display library strains of opposite mating type with different fluorescent proteins without the target ligand and using FACS to isolate those cells that do not form diploids. This would screen out random display peptides/proteins that can bind each other or another cell independent of the target ligand.

To test the fluorescence intensity of strains expressing GFP and mCherry, flow cytometry was performed on diluted samples of each strain. Both strains showed a significant difference between the wild type and the strains expressing fluorescent proteins. Fig. 10 (a) shows the mean fluorescence emitted from the BY4742 *Δsag* strain expressing the red fluorescent protein mCherry in comparison to the wild type control strain. There is a significant difference in the mean fluorescence intensity between the control strain (M= 301.7 ± 2 2.9) and strain containing mCherry (M= 7446 ± 314.9) as p value is <0.0001. Fig. 10 (b) shows the mean fluorescence intensity emitted by the opposing mating type carrying the green fluorescent protein, eGFP in comparison to the wild type BY4741 strain. Similarly, there is a significant difference in the mean fluorescence intensity between the control strain (M= 547.2 ± 148.5) and strain containing eGFP (M= 63413.2 ± 39838.9) as the p value is 0.05.

(a) (b)

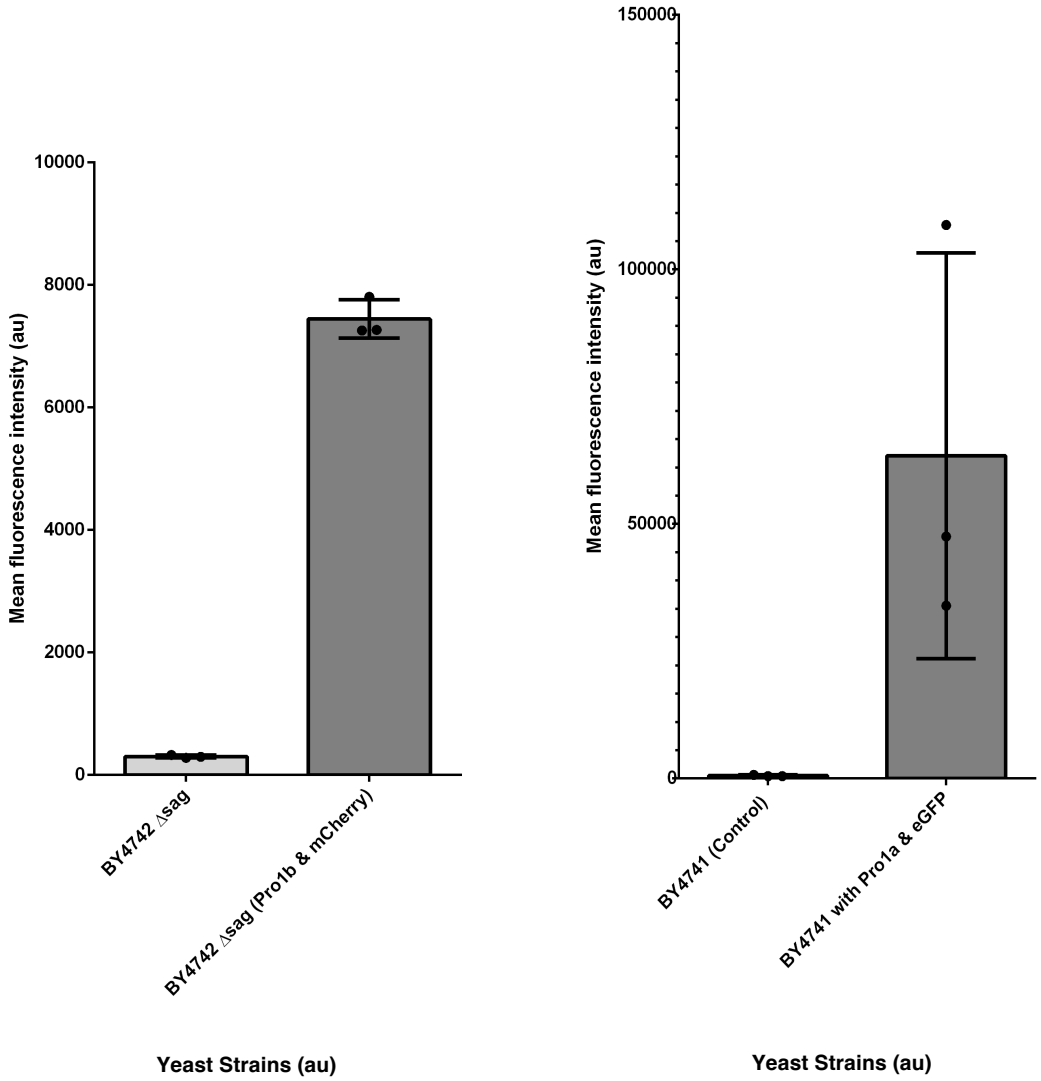

Figure 10: Mean fluorescence intensity of green and red fluorescent strains (a) Mean fluorescence intensity difference between control strain wild type BY4742 Δ*sag* and BY4742 Δ*sag* expressing red fluorescent protein (mCherry) and surface displaying Aga1p and fusion protein Aga2p-Pro1b. (b) Bar chart illustrating the mean fluorescence intensity between control strain wild type BY4741 and BY4741 expressing green fluorescence protein (GFP) and surface displaying Aga1p and fusion protein Aga2p-Pro1b. Fluorescence values are the mean of biological triplicates measured as 10000 events using flow cytometry normalised to the non-fluorescent strain BY4741. Error bars show standard deviation.

3.5 Mating assay for assessing protein-ligand-protein interaction

The protocol published by Younger *et. al* was employed to test its capacity to select for small molecule-protein interactions[59]. Initial experiments showed that BY4741 retained the capacity to interact with BY4742 Δ*sag* to produce diploids. This was inconsistent with previously published work where *SAG1* knockout led to a much more pronounced reduction in mating efficiency. In order to detect differences in mating efficiency, cultures were plated at higher levels of dilution so that subtle differences were more obvious. Hence, several dilutions of the mating assay were plated and a significant difference from the wild type (BY4741 and BY4742) mating was observed when BY4742 *SAG1* knockout strain was used, at a dilution of 10^{-2}. The wild type mating assay at a dilution of 10^{-2} was 1018 cfus and the *SAG1* knockout strain was 9 cfus. To assess protein-ligand-protein interaction, a mating assay was conducted with strains expressing Pro1a and Pro1b on the surface of opposing mating types (Fig.11). Pro1a and Pro1b bind with progesterone and were thus mixed together in equal proportion and shaken in liquid medium which permits the growth of both strains. Mating assays were conducted both in the absence and presence of progesterone to measure difference in mating efficiency upon introduction of the ligand. The same process was repeated with positive control and negative control strains with *SAG1* knocked out. After 17 hours of incubation, serial dilutions of this mixture were plated onto medium selective for the growth of diploids. Colony counting for the 17-hour incubation mating assay was challenging due to high number of diploid colonies, hence the mating assay was optimised using a 4-hour incubation time. This was based on the duration of yeast cell cycle which is ~1.5 hours. 2-3 cell cycles were allowed to isolate true binders and reduce any false positives which could have arisen from cell agglutination resulting from cells settling at the bottom of the Falcon tubes.

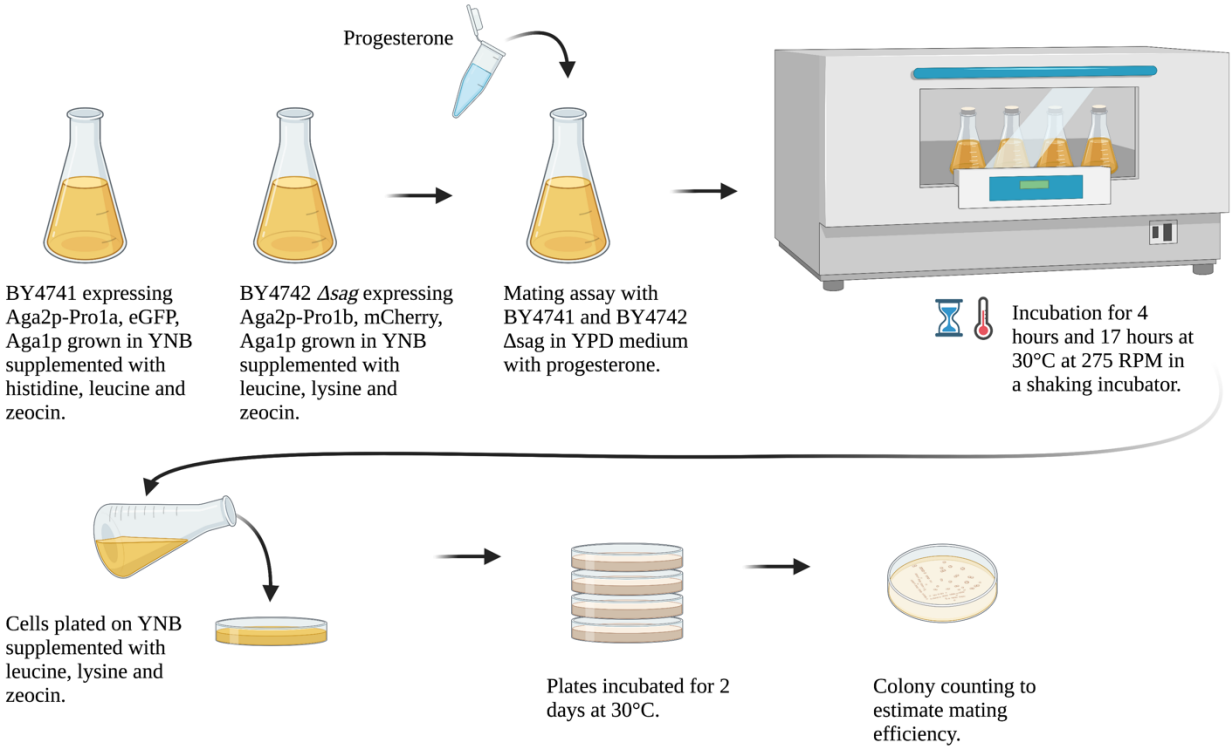

Figure 11: Simplified workflow of the mating assay. Mating assay was conducted with BY4741 strain expressing fusion protein Aga2p-Pro1a on the surface, eGFP and Aga1p and BY4742 Δ*sag* expressing fusion protein Aga2p-Pro1b on the surface, mCherry and Aga1p in the presence of progesterone to test protein-ligand-protein interaction. Dilutions of the mating assay were plated on respective selective media for 2 days prior to colony counting. Created with BioRender.com

As expected, when *SAG1* was deleted from the BY4742 strain, mating efficiency was significantly lower (for 4-hour incubation, *M*= 155.3 ± 95.9 cfus and for 17-hour incubation, *M*= 239.6 ± 254.84 cfus), compared to the wild-type positive controls (*M*= 5333.33 ± 235.70 cfus), indicating that the deletion of *SAG1* reduced the capacity for native sexual agglutination. A trend showing an increase in mating efficiency for the mating assays performed with Strain 4 and Strain 5 (Table 3) for 4 hours and 17 hours was observed when high progesterone was added, in comparison to when no progesterone was added (Fig.12). For 4-hour incubation, the high progesterone treatment group had a mean of 348.8 ± 198.5 cfus and the 17-hour incubation had a mean of 541± 375.2 cfus). In comparison, the mating assay when no progesterone was added had a mean of 0 ± 0 cfus) for 4-hour incubation and 17-hour incubation had a mean of 41.67 ± 52.01 cfus). There was not a significant difference according to statistical analysis between the no progesterone and high-progesterone treatment as the p value =0.06. Interestingly, although there was no significant increase in mating

efficiency between the strains displaying progesterone binding domains and the negative control strains, there was a significant difference in mating efficiency when progesterone was added to the progesterone binding domain displaying strains compared to when no progesterone was added. An important difference between the diploids formed by the control groups and the surface binders was the number of plasmids being carried in the strains. This could be due to the phenomenon of plasmid burden which reduces growth rate in yeast cells, especially diploids in comparison to haploids. Previous work has attributed reduced growth rate in diploid yeast cells due to the energetic costs of maintenance of plasmid DNA and metabolic burden associated with replication machinery[59]. The control strains used for this experiment did not have empty vectors equivalent to those used to express surface proteins, and therefore may have had a higher growth rate during the liquid incubation step, leading to higher cfu counts on agar plates. Given that the mating efficiency of the strains without progesterone is lower than the control mating assay of BY4741 and BY4742 Δ*sag,* it is possible that the test strains had a growth burden from their plasmids or expressed display proteins that led to a lower growth rate and colony count relative to the control assay. Future work replicating this experiment with control strains of BY4741 and BY4742 Δ*sag* containing *AGA1* plasmid, fluorescence plasmids and plasmids containing methionine and uracil marker genes is required to omit this discrepancy, which maybe caused from plasmid burden.

Another limitation which needs to be addressed in future work is validation of our current findings with other known protein-ligand-protein interactions such as FKBP-Rapamycin-FRB. Increasing the concentration of rapamycin should ideally increase the affinity of FRB and FKBP domains to simultaneously bind with rapamycin and show a linear increase in mating efficiency.

(a) (b)

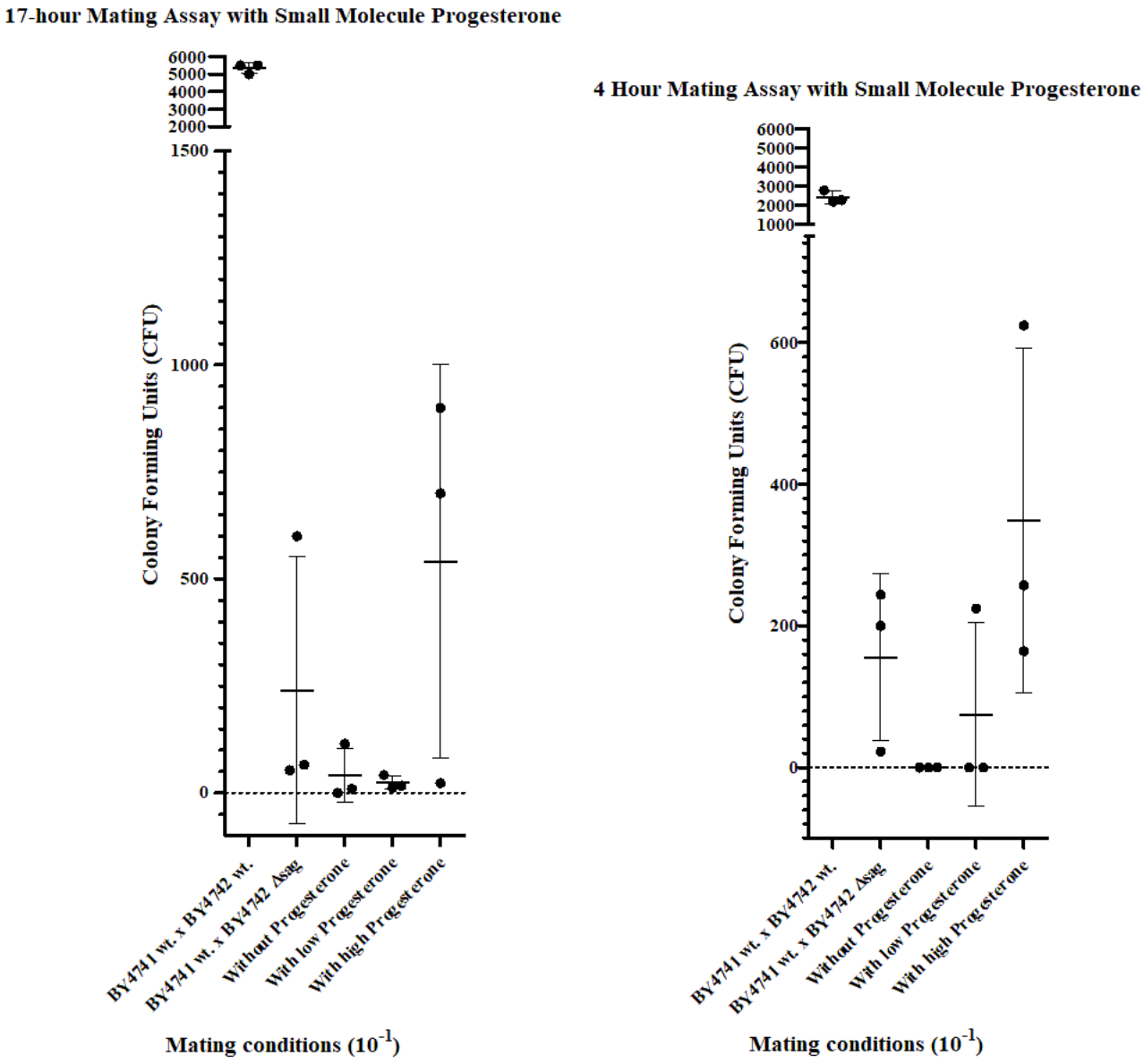

Figure 12: Mating efficiency of protein-ligand-protein interaction mating assay. Mating efficiency comparison between positive control (wild type BY4741 and BY4742), negative control (BY4741 and BY4742 Δ*sag*) and strains with progesterone binders Pro1a and Pro1b in the absence and presence of low (0.1µm) and high (1µm) concentration of progesterone incubated for **(a)** 17 hours and **(b)** 4 hours. Mating efficiency was determined by calculating the colony forming units at a dilution of 10^{-1} for 17 hour mating and no dilution for 4 hour assay. Columns and are the mean of biological triplicates. Error bars show standard deviation.

4. Future Directions

Previous work conducted by Younger et al. reported the capacity of yeast cells to recover mating efficiency when complementary binding domains are expressed on the surface[59]. We have demonstrated initial results of recovering mating efficiency upon introduction of a ligand with complementary binding domains expressed on the surface of opposing mating types of yeast in liquid culture. Due to the short timeframe of this thesis and limitations imposed by COVID-19, we were only able to test a known protein-ligand-protein interaction for the small molecule progesterone to test the feasibility of this approach. We aim to replicate the mating assay experiment with more biological replicates, higher progesterone concentration and more genetically comparable control strains. Our work is promising for building a universal platform for generating novel pairs of binding domains for incorporation into biosensors. The future aim of SYD is to screen diverse libraries of displayed peptides or proteins against the target ligand to generate novel binding pairs which can be integrated in biosensors. We aim to transform our background strains to display peptide libraries and test it against a multitude of medically relevant ligands that span a wide range of size and complexity. Our goal is to transform our background strains with an NNK library (a 6-codon library which can code all 20 amino acids) fused with Aga2p and characterize whole library-on-library protein interactions against medically relevant ligands. The peptides with affinity to the ligand will be enriched or isolated and identified using SYD. Next generation sequencing will be utilised for identifying the binders from the plasmid extracts. If successful, SYD could dramatically improve the ease of biosensor creation and enable the use of biosensors across multiple industries.

4.1 Future target biomarkers for generating medical biosensors using SYD binding domains

In the medical industry, diagnostic testing is currently heavily based on biomarker testing conducted in laboratories which requires expensive equipment, long waiting time and trained personnel. Biosensors offer easy, cheap and rapid solution for point-of-care diagnostics. Our area of focus is cancer diagnostics as cancer is one of the leading causes of death worldwide, estimated at 10 million deaths and 19.3 million new cases per year[69]. The national expenditure of the US alone on cancer care in 2020 is $208.9 billion[70]. This poses a huge burden on society by impacting the quality of life of citizens and increasing mortality rates. Early diagnosis and novel drugs are vital for promoting effective treatment, reducing patient mortality, and clinical monitoring in the patients.

Current commercially used cancer diagnostic techniques utilise Fluorodeoxyglucose (18F) [FDG] which is a radioactive analog of glucose[71]. This radiotracer is administered to the patient and a PET (Positron Emission Topography) scanner identifies the distribution of the FDG (due to the positron-emitting radionucleotide fluorine-18) due to the higher uptake of glucose to identify tumours[71]. This is an invasive method of diagnosis compared to molecular biosensors which can detect cancer biomarkers in patient serum. However, the biggest roadblock in commercialising cancer-specific biosensors is the specificity towards the biomarker.

A platform for generating novel matching binding domains which can be easily integrated in existing biosensor designs can improve identification and quantification of biomarkers. Since the size limitations of targetable molecules in the SYD system are unclear, we aim to screen a range of clinically relevant biomarkers from extracellular vesicles, proteins, and metabolites to find the optimal size range. It was not practical to conduct library-on-library work during this thesis, however we aim to utilise SYD (simultaneous yeast display) to build medical biosensors against cancer biomarkers in the future. Our chosen cancer biomarkers in order of decreasing complexity and size are: prostasomes, AMF/PGI, AMFR/gp78, Tn antigen and sarcosine. Testing the SYD system against these biomarkers will simultaneously test the range of molecular targets that SYD is effective with and generate pairs of cancer biomarker specific binding peptides that could potentially be developed into cancer biosensors.

4.1.1 Prostasomes

An emerging target molecule for molecular diagnostics are extracellular vesicles (EVs) such as exosomes, as they carry multiple bioactive molecules (DNA, RNA, proteins, lipids) secreted by a parent cell, including tumor cells. As EVs can be detected non-invasively in every bodily fluid, they make an excellent window into understanding the state of the cells and progression of cancer.

Prostate-originating EVs called prostasomes are a validated diagnostic biomarker for prostate cancer. Prostasomes are multivesicular bodies (mean diameter, 150 nm) that are released as a form of cell-cell communication by malignant prostate acinar cells. Invasive growth of malignant prostate cells may cause these prostasomes which are normally released into seminal fluid, to appear in peripheral circulation[72]. They are blood biomarkers as the blood of healthy men is normally devoid of prostasomes, thus their secretion in blood is indicative of pathology of prostate cancer[73]. Furthermore, tumour aggressiveness has been correlated by increased levels of prostasomes via proximity ligation assay (PLA) making them promising biomarkers for prostate cancer detection[72].

Utilising prostasomes for SYD could enable generation of multiple peptides which can be utilised in assays such as PLA with a combinatorial signal from both a protein and a nucleic acid, or to generate binding pairs for novel protein-based biosensors.

4.1.2 Tn Antigen

Tn antigen (GalNAcα1-O-Ser/Thr) is a carbohydrate antigen with a molecular weight of 308.29 Da which has been associated with tumours and not observed in peripheral normal tissue or blood cells[74]. The expression of this antigen has been correlated with all types of human carcinoma. Membrane proteins express Tn antigen and/or truncated O-glycans due to blockage of the normal O-glycosylation pathway indicating lack of homeostasis[75]. Normally, the Tn antigen is efficiently modified by the enzyme T-synthase in the Golgi apparatus but due to misfolding, the glycosylation is altered and produces truncated glycan structures[75]. Tn antigen is responsible for creating an immune suppressive microenvironment driving tumour growth and metastasis[76]. Developing binding peptides or glycoproteins against O-glycans can enable in determining the size range and complexity SYD can operate in.

4.1.3 PGI/AMF and AMFR/gp78

Phosphoglucose isomerase/autocrine motility factor (PGI/AMF) is an essential enzyme of catabolic glycolysis and anabolic gluconeogenesis. PGI/AMF is a moonlighting protein, which means it has different functions in different environments of the cell. Intracellularly, PGI acts as a glycolytic enzyme, however once secreted in the extracellular environment it plays the role of autocrine motility factor (AMF), a cytokine which stimulates tumor cell motility in an autocrine manner[77]. PGI/AMF is 55kDa in size binds to the receptor AMFR/gp78 which is 78 kDa in size. AMF stimulates the migration and proliferation of endothelial cells upon binding with its receptor AMFR/gp 78 which is a seven-transmembrane protein. AMFR/gp78 is highly regulated in a variety of human tumours and their microenvironment, however not in adjacent normal tissue. gp78 overexpression correlates with invasiveness and metastasis progression of the tumour[78]. It causes the tumour cell to detach from the primary site and promote the cell motility[79].

Elevated levels of AMF in serum have been observed in colorectal, lung, gastrointestinal, kidney, breast, and prostate cancer patients. AMF is found to be involved in the development of metastases in vivo[80]. Ahmad et. al, have developed an electrochemical biosensor for AMF detection using

the inhibitor N-(5-phosphate-D-arabinoyl)-2-amino ethanamine[81]. Designing a cheaper protein-based biosensor for the detection and monitoring of AMF protein in human fluids could be a significant scientific breakthrough in oncology.

4.1.4 Sarcosine

Sarcosine is an N-methyl derivative of the amino acid glycine which can be detected non-invasively in urine[82]. Sarcosine has been suggested as an oncometabolite which is 89.9 Da in size. Urine sarcosine concentrations are elevated during prostate cancer progression to metastasis[82]. Sarcosine is associated with angiogenesis through the PI3K/Akt/mTOR pathway inducing invasion in prostate cancer cells. Currently, there is demand for accurate and precise detection of low concentration of sarcosine in a cheap, rapid, and reliable manner. Sarcosine is extremely small in comparison to the abovementioned targets and will enable in understanding the limits of SYD.

5. Conclusion

Biosensors have enabled detection of broad range of metabolites and have a variety of applications across fields like point-of-care diagnostics, environmental monitoring, drug discovery and treatment. They rely on the biorecognition elements such as DNA, RNA, transcription factor or proteins for specificity and selectivity to the target ligand. While there are a range of transduction and signal mechanisms and a variety of biosensor designs which can be applied for construction of successful biosensors, the limiting factor is finding specific biorecognition elements which bind to the target analyte. Most nucleic based and protein-based biosensor designs require two binding domains which simultaneously bind to the target ligand to produce a signal. When a native pair doesn't exist, time consuming and difficult protein engineering projects need to be undertaken for developing new domains. However, there is no such technique for finding novel binding protein pairs for incorporation into protein-based biosensors. Conventional techniques such as microarrays, yeast display and phage display can produce one binding peptide but are not optimized for detecting co-binders in a free solution. Simultaneous Yeast Display (SYD) is a methodology proposed to overcome this limitation and aims to generate binding peptide pair to a target ligand without fluorescently labelling or immobilizing the ligand. The peptides generated can be easily integrated in existing generic biosensor architectures. This thesis has demonstrated that SYD can be used to detect protein-ligand-protein interactions in addition to the protein-protein interactions that it was originally designed for. Future testing and exploitation of SYD will involve using cancer biomarkers such as prostasomes, PGI, AMFR, Tn antigen and sarcosine to test the limitations of the

system and provide binding domains for eventual use in cancer biosensors. The field of synthetic biology can greatly benefit from development of this technique to develop robust industrial and medical biosensors.

6. References:

1. Pickel, D., Manucy, G. P., Walker, D. B., Hall, S. B. & Walker, J. C. Evidence for canine olfactory detection of melanoma. *Applied Animal Behaviour Science* **89**, 107–116 (2004).
2. Fields, R. D. The shark's electric sense. *Sci Am* **297**, 74–80 (2007).
3. Banerjee, A., Maity, S. & Mastrangelo, C. H. Nanotechnology for biosensors: A Review. *arXiv:2101.02430 [physics]* (2021).
4. Liu, H., Ge, J., Ma, E. & Yang, L. 10 - Advanced biomaterials for biosensor and theranostics. in *Biomaterials in Translational Medicine* (eds. Yang, L., Bhaduri, S. B. & Webster, T. J.) 213–255 (Academic Press, 2019). doi:10.1016/B978-0-12-813477-1.00010-4.
5. Naresh, V. & Lee, N. A Review on Biosensors and Recent Development of Nanostructured Materials-Enabled Biosensors. *Sensors* **21**, 1109 (2021).
6. Laupland, K. B. & Valiquette, L. The changing culture of the microbiology laboratory. *Can J Infect Dis Med Microbiol* **24**, 125–128 (2013).
7. Pitt, J. J. Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *Clin Biochem Rev* **30**, 19–34 (2009).
8. Campuzano, S., Yáñez-Sedeño, P. & Pingarrón, J. M. Molecular Biosensors for Electrochemical Detection of Infectious Pathogens in Liquid Biopsies: Current Trends and Challenges. *Sensors (Basel)* **17**, 2533 (2017).
9. Haleem, A., Javaid, M., Singh, R. P., Suman, R. & Rab, S. Biosensors applications in medical field: A brief review. *Sensors International* **2**, 100100 (2021).
10. Bhalla, N., Jolly, P., Formisano, N. & Estrela, P. Introduction to biosensors. *Essays Biochem* **60**, 1–8 (2016).
11. Insights, M. Glucose Biosensor Market Size to Reach $31 Billion by 2022 | Million Insights. https://www.prnewswire.com/news-releases/glucose-biosensor-market-size-to-reach-31-billion-by-2022--million-insights-681600061.html.
12. Newman, J. D. & Turner, A. P. F. Home blood glucose biosensors: a commercial perspective. *Biosens Bioelectron* **20**, 2435–2453 (2005).
13. Nguyen, H. H., Lee, S. H., Lee, U. J., Fermin, C. D. & Kim, M. Immobilized Enzymes in Biosensor Applications. *Materials (Basel)* **12**, 121 (2019).
14. Campbell, K. *et al.* Comparison of ELISA and SPR biosensor technology for the detection of paralytic shellfish poisoning toxins. *J Chromatogr B Analyt Technol Biomed Life Sci* **877**, 4079–4089 (2009).
15. Jl, F. Update on pregnancy testing. *Primary care* **13**, (1986).
16. Ahn, C. H. *et al.* Disposable smart lab on a chip for point-of-care clinical diagnostics. *Proceedings of the IEEE* **92**, 154–173 (2004).
17. Lisowski, P. & Zarzycki, P. K. Microfluidic Paper-Based Analytical Devices (μPADs) and Micro Total Analysis Systems (μTAS): Development, Applications and Future Trends. *Chromatographia* **76**, 1201–1214 (2013).
18. Tsagkaris, A. S., Pulkrabova, J., Hajslova, J. & Filippini, D. A Hybrid Lab-on-a-Chip Injector System for Autonomous Carbofuran Screening. *Sensors* **19**, 5579 (2019).
19. Park, M. Surface Display Technology for Biosensor Applications: A Review. *Sensors (Basel)* **20**, 2775 (2020).
20. Liu, L. S., Wang, F., Ge, Y. & Lo, P. K. Recent Developments in Aptasensors for Diagnostic Applications. *ACS Appl. Mater. Interfaces* **13**, 9329–9358 (2021).
21. Prante, M., Segal, E., Scheper, T., Bahnemann, J. & Walter, J. Aptasensors for Point-of-Care Detection of Small Molecules. *Biosensors* **10**, 108 (2020).
22. Jayasena, S. D. Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin Chem* **45**, 1628–1650 (1999).
23. Ruscito, A. & DeRosa, M. C. Small-Molecule Binding Aptamers: Selection Strategies, Characterization, and Applications. *Frontiers in Chemistry* **4**, 14 (2016).

24. Zhang, Y., Lai, B. S. & Juhas, M. Recent Advances in Aptamer Discovery and Applications. *Molecules* **24**, 941 (2019).
25. Findeiß, S., Etzel, M., Will, S., Mörl, M. & Stadler, P. F. Design of Artificial Riboswitches as Biosensors. *Sensors (Basel)* **17**, 1990 (2017).
26. Mehdizadeh Aghdam, E., Hejazi, M. S. & Barzegar, A. Riboswitches: From living biosensors to novel targets of antibiotics. *Gene* **592**, 244–259 (2016).
27. Hoang Trung Chau, T., Hoang Anh Mai, D., Ngoc Pham, D., Thi Quynh Le, H. & Yeol Lee, E. Developments of Riboswitches and Toehold Switches for Molecular Detection—Biosensing and Molecular Diagnostics. *International Journal of Molecular Sciences* **21**, 3192 (2020).
28. McKeague, M. & Derosa, M. C. Challenges and opportunities for small molecule aptamer development. *J Nucleic Acids* **2012**, 748913 (2012).
29. Aydindogan, E., Balaban, S., Evran, S., Coskunol, H. & Timur, S. A Bottom-Up Approach for Developing Aptasensors for Abused Drugs: Biosensors in Forensics. *Biosensors (Basel)* **9**, 118 (2019).
30. Carpenter, A. C., Paulsen, I. T. & Williams, T. C. Blueprints for Biosensors: Design, Limitations, and Applications. *Genes (Basel)* **9**, 375 (2018).
31. Gallegos M T, Schleif R, Bairoch A, Hofmann K, & Ramos J L. Arac/XylS family of transcriptional regulators. *Microbiology and Molecular Biology Reviews* **61**, 393–410 (1997).
32. Ramos Juan L. *et al.* The TetR Family of Transcriptional Repressors. *Microbiology and Molecular Biology Reviews* **69**, 326–356 (2005).
33. Shi, S., Choi, Y. W., Zhao, H., Tan, M. H. & Ang, E. L. Discovery and engineering of a 1-butanol biosensor in Saccharomyces cerevisiae. *Bioresource Technology* **245**, 1343–1351 (2017).
34. Pandian, G. N. & Sugiyama, H. Designing Biomimicking Synthetic Transcription Factors for Therapeutic Gene Modulation. in *Cell-Inspired Materials and Engineering* (eds. Wang, D. O. & Packwood, D.) 135–161 (Springer International Publishing, 2021). doi:10.1007/978-3-030-55924-3_6.
35. Ansari, A. Z. & Mapp, A. K. Modular design of artificial transcription factors. *Current Opinion in Chemical Biology* **6**, 765–772 (2002).
36. Stein, V. & Alexandrov, K. Synthetic protein switches: design principles and applications. *Trends Biotechnol* **33**, 101–110 (2015).
37. Merkx, M., Golynskiy, M. V., Lindenburg, L. H. & Vinkenborg, J. L. Rational design of FRET sensor proteins based on mutually exclusive domain interactions. *Biochemical Society Transactions* **41**, 1201–1205 (2013).
38. Constantinou, A. & Polizzi, K. M. Opportunities for bioprocess monitoring using FRET biosensors. *Biochemical Society Transactions* **41**, 1146–1151 (2013).
39. Tao, R. *et al.* Genetically encoded fluorescent sensors reveal dynamic regulation of NADPH metabolism. *Nat Methods* **14**, 720–728 (2017).
40. Okuda, J. & Sode, K. PQQ glucose dehydrogenase with novel electron transfer ability. *Biochemical and Biophysical Research Communications* **314**, 793–797 (2004).
41. Guo, Z. *et al.* Engineered PQQ-Glucose Dehydrogenase as a Universal Biosensor Platform. *J. Am. Chem. Soc.* **138**, 10108–10111 (2016).
42. F. Turner, A. P. Biosensors: sense and sensibility. *Chemical Society Reviews* **42**, 3184–3196 (2013).
43. Morales, M. A. & Mark Halpern, J. Guide to Selecting a Biorecognition Element for Biosensors. *Bioconjug Chem* **29**, 3231–3239 (2018).
44. Martin-Yken, H. Yeast-Based Biosensors: Current Applications and New Developments. *Biosensors* **10**, 51 (2020).
45. Alhajj, M. & Farhana, A. Enzyme Linked Immunosorbent Assay. in *StatPearls* (StatPearls Publishing, 2021).
46. Vegas, A. J., Fuller, J. H. & Koehler, A. N. Small-molecule microarrays as tools in ligand discovery. *Chem Soc Rev* **37**, 1385–1394 (2008).

47. Haab, B. B., Dunham, M. J. & Brown, P. O. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biology* **2**, research0004.1 (2001).
48. Yao, C. *et al.* Screening of the Binding of Small Molecules to Proteins by Desorption Electrospray Ionization Mass Spectrometry Combined with Protein Microarray. *J. Am. Soc. Mass Spectrom.* **26**, 1950–1958 (2015).
49. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246 (1989).
50. Wong, J. H. *et al.* A yeast two-hybrid system for the screening and characterization of small-molecule inhibitors of protein–protein interactions identifies a novel putative Mdm2-binding site in p53. *BMC Biology* **15**, 108 (2017).
51. Chen, Y.-C., Rajagopala, S. V., Stellberger, T. & Uetz, P. Exhaustive benchmarking of the yeast two-hybrid system. *Nat Methods* **7**, 667–668; author reply 668 (2010).
52. Brückner, A., Polge, C., Lentze, N., Auerbach, D. & Schlattner, U. Yeast Two-Hybrid, a Powerful Tool for Systems Biology. *Int J Mol Sci* **10**, 2763–2788 (2009).
53. Cherf, G. M. & Cochran, J. R. Applications of yeast surface display for protein engineering. *Methods Mol Biol* **1319**, 155–175 (2015).
54. Wu, C.-H., Liu, I.-J., Lu, R.-M. & Wu, H.-C. Advancement and applications of peptide phage display technology in biomedical science. *Journal of Biomedical Science* **23**, 8 (2016).
55. Yang, X. *et al.* Development of novel surface display platforms for anchoring heterologous proteins in Saccharomyces cerevisiae. *Microb Cell Fact* **18**, 85 (2019).
56. Peltomaa, R., Benito-Peña, E., Barderas, R. & Moreno-Bondi, M. C. Phage Display in the Quest for New Selective Recognition Elements for Biosensors. *ACS Omega* **4**, 11569–11580 (2019).
57. Sidhu, S. S., Lowman, H. B., Cunningham, B. C. & Wells, J. A. Phage display for selection of novel binding peptides. *Methods Enzymol* **328**, 333–363 (2000).
58. Bowen, J. *et al.* Screening of Yeast Display Libraries of Enzymatically Treated Peptides to Discover Macrocyclic Peptide Ligands. *Int J Mol Sci* **22**, 1634 (2021).
59. Younger, D., Baker, D., Berger, S. & Klavins, E. High-throughput characterization of protein–protein interactions by reprogramming yeast mating | *Proceedings of the National Academy of Sciences* **114**, no. 46 12166-12171 (2017)
60. Herskowitz, I. Cellular differentiation, cell lineages, and transposable genetic cassettes in yeast. *Curr Top Dev Biol* **18**, 1–14 (1983).
61. Usher, J. The Mechanisms of Mating in Pathogenic Fungi—A Plastic Trait. *Genes (Basel)* **10**, 831 (2019).
62. Merlini, L., Dudin, O. & Martin, S. G. Mate and fuse: how yeast cells do it. *Open Biol* **3**, 130008 (2013).
63. Jester, B. W., Tinberg, C. E., Rich, M. S., Baker, D. & Fields, S. Engineered Biosensors from Dimeric Ligand-Binding Domains. *ACS Synth. Biol.* **7**, 2457–2467 (2018).
64. Banaszynski, L. A., Liu, C. W. & Wandless, T. J. Characterization of the FKBP.rapamycin.FRB ternary complex. *J Am Chem Soc* **127**, 4715–4721 (2005).
65. Galán, A. *et al.* Library-based display technologies: where do we stand? *Mol. BioSyst.* **12**, 2342–2358 (2016).
66. Malmberg, C. Evaluation of flow Cytometry as Replacement for Plating in in vitro Measurements of Competitive Growth under Antibiotic Stress. (2013).
67. Yeast Transformation - High Efficiency · Benchling. https://benchling.com/will_shaw/f/lib_QVmbt0aU-crispr-guide-public/prt_YuiWM2fv-yeast-transformation-high-efficiency/edit.
68. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122**, 19–27 (1989).

69. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* **71**, 209–249 (2021).
70. Financial Burden of Cancer Care | Cancer Trends Progress Report. https://progressreport.cancer.gov/after/economic_burden.
71. Almuhaideb, A., Papathanasiou, N. & Bomanji, J. 18F-FDG PET/CT Imaging In Oncology. *Ann Saudi Med* **31**, 3–13 (2011).
72. Tavoosidana, G. *et al.* Multiple recognition assay reveals prostasomes as promising plasma biomarkers for prostate cancer. *Proc Natl Acad Sci U S A* **108**, 8809–8814 (2011).
73. Aalberts, M., Stout, T. A. E. & Stoorvogel, W. Prostasomes: extracellular vesicles from the prostate. *Reproduction* **147**, R1–R14 (2014).
74. Fuster, M. M. & Esko, J. D. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat Rev Cancer* **5**, 526–542 (2005).
75. Matsumoto, Y. *et al.* Identification of Tn antigen O-GalNAc-expressing glycoproteins in human carcinomas using novel anti-Tn recombinant antibodies. *Glycobiology* **30**, 282–300 (2019).
76. Hakomori, S. Tumor-associated carbohydrate antigens defining tumor malignancy: basis for development of anti-cancer vaccines. *Adv Exp Med Biol* **491**, 369–402 (2001).
77. Jeffery, C. Multifunctional proteins: examples of gene sharing. *Annals of medicine* **35**, 28–35 (2003).
78. Yanagawa, T. *et al.* Overexpression of autocrine motility factor in metastatic tumor cells: possible association with augmented expression of KIF3A and GDI-β. *Lab Invest* **84**, 513–522 (2004).
79. Watanabe, H., Shinozaki, T., Raz, A. & Chigira, M. Expression of autocrine motility factor receptor in serum- and protein-independent fibrosarcoma cells: implications for autonomy in tumor-cell motility and metastasis. *Int J Cancer* **53**, 689–695 (1993).
80. Nabi, I. R., Watanabe, H., Silletti, S. & Raz, A. Tumor cell autocrine motility factor receptor. *EXS* **59**, 163–177 (1991).
81. Ahmad, L. *et al.* Novel N-substituted 5-phosphate-d-arabinonamide derivatives as strong inhibitors of phosphoglucose isomerases: Synthesis, structure-activity relationship and crystallographic studies. *Bioorganic Chemistry* **102**, 104048 (2020).
82. Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910–914 (2009).