2023

Building a High-Throughput Yeast Two-Hybrid Assay to Screen for Novel Chronic Hepatitis B Drugs

Benjamin Glazer

Building a High-Throughput Yeast Two-Hybrid Assay to Screen for Novel Chronic Hepatitis B Drugs

Benjamin Glazer

A Thesis Submitted in Candidacy for Honors at Graduation
Lawrence University
May 1, 2023

I hereby reaffirm the Lawrence University Honor Code.

Table of Contents

Abstract

Hepatitis B is a vaccine-preventable liver disease caused by hepatitis B virus (HBV). Chronic hepatitis B currently affects about 296 million people around the world, killing 820,000 annually. There is currently no cure for chronic hepatitis B, but research over the last decade has identified hepatitis B X protein (HBx) as a promising therapeutic target. Specifically, inhibiting the interaction between HBx and human protein DDB1 (HBx-DDB1) could disrupt HBx function. An essential step in the drug development process is to create an assay that is capable of high-throughput screening of millions of compounds for activity against HBx-DDB1. This research aims to design and optimize a high-throughput capable assay for the HBx-DDB1 interaction using the yeast two-hybrid (Y2H) system in *Saccharomyces cerevisiae*. This assay could eventually be used to screen for novel drug compounds to treat chronic hepatitis B.

Introduction

Hepatitis B

Hepatitis B is a vaccine-preventable disease caused by hepatitis B virus (HBV). HBV infects parenchymal liver cells, causing many forms of liver disease, and is the leading cause of liver cancer. Chronic hepatitis B is usually acquired through vertical transmission of the virus—from mother to child at birth—but HBV can also be transmitted through sexual contact or contact with infected blood. Currently, there are about 296 million people chronically infected with HBV around the world, leading to around 820,000 deaths every year[1,2].

The most widely used hepatitis B vaccine was developed in 1986, and is the first human vaccine to use recombinant DNA[3]. It is a three-dose series given at birth, produced by yeast cells

expressing the HBV surface antigen protein. It is very effective, reducing the number of new cases of chronic hepatitis B to only 5 per 100,000 people in the US in 2020[4]. However, the disease remains prevalent around the world because many countries do not have widespread child vaccination programs for hepatitis B. Child vaccination programs can be difficult to implement for a variety of reasons including cost, infrastructure, and social stigma. The most heavily affected regions of the world include the Western Pacific and Africa[5] (Figure 1).

There are currently seven FDA approved drugs for treating chronic hepatitis B in adults. These treatments can slow HBV reproduction and reduce the risk of serious liver damage, but there is no cure for chronic hepatitis B[6]. Despite dramatic reductions in the price of these drugs and the cost of screening, low-income countries remain disproportionately affected. In low-income countries only 7% of cases are diagnosed and 2% of diagnosed cases are being treated, compared to a 44% diagnosis rate and 30% treatment rate in high-income countries[7,8].

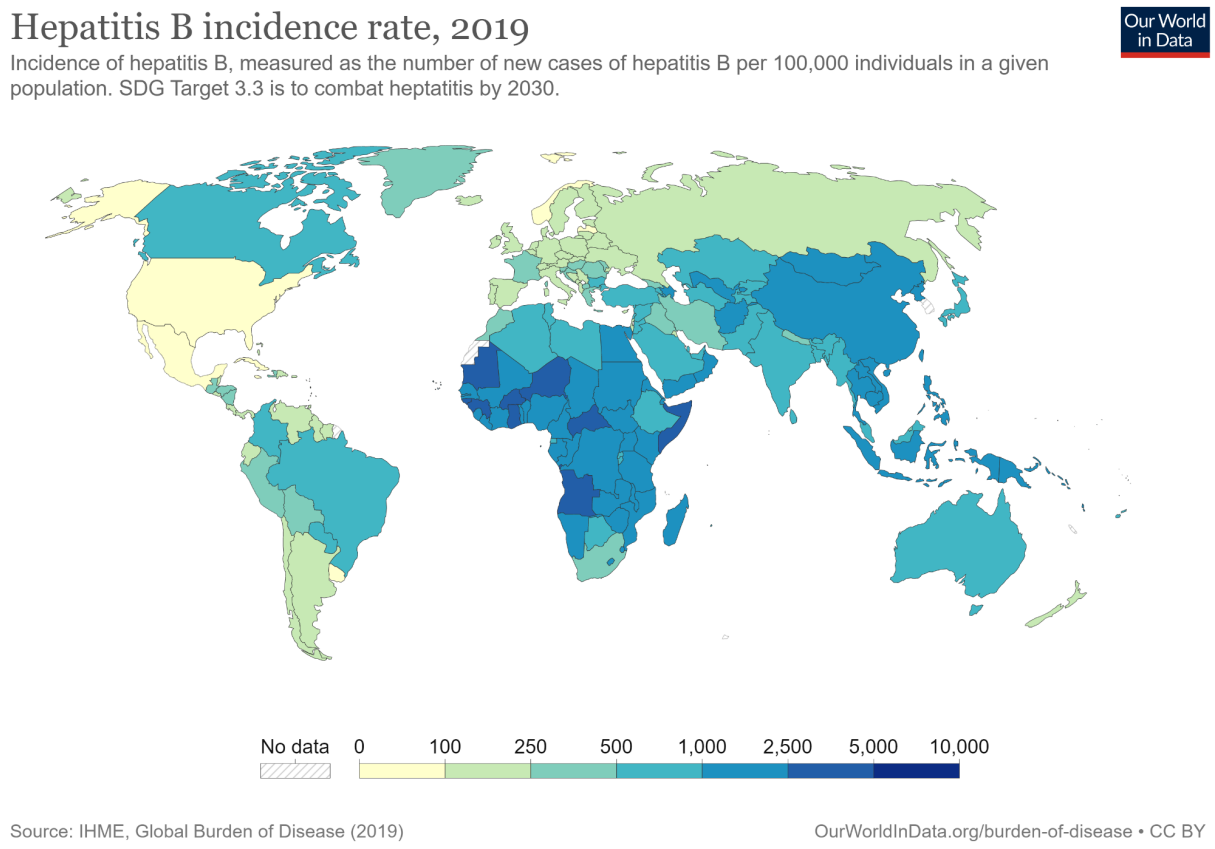

Figure 1. Hepatitis B incidence rate by country in 2019.[9] Darker colors indicate higher incidence rate of hepatitis B.

HBV

Hepatitis B virus is a small, enveloped virus with a gapped dsDNA genome (Figure 2). Enveloped viruses surround their dense protein core (capsid) with a lipid membrane in the process of exiting host cells. This often provides the functional, infectious viral particles (virions) a familiar mask to hide from the host immune system. The HBV capsid is 32 nm in diameter inside a 42 nm diameter envelope[10]. The capsid contains the viral DNA genome, together being referred to as the nucleocapsid. A viral protein embedded in the envelope of HBV virions interacts closely with Na^+-taurocholate cotransporting polypeptide (NTCP), a bile acid

transporter found on the surface of parenchymal liver cells (hepatocytes). This tight interaction supports endocytosis of HBV virions, in which the lipid envelope fuses with the hepatocyte cell membrane, releasing the nucleocapsid into the cytoplasm. The nucleocapsid is transported along microtubules to the nucleus of the cell, where it releases the DNA genome to enter the nucleus through nuclear pore complexes[11,12].

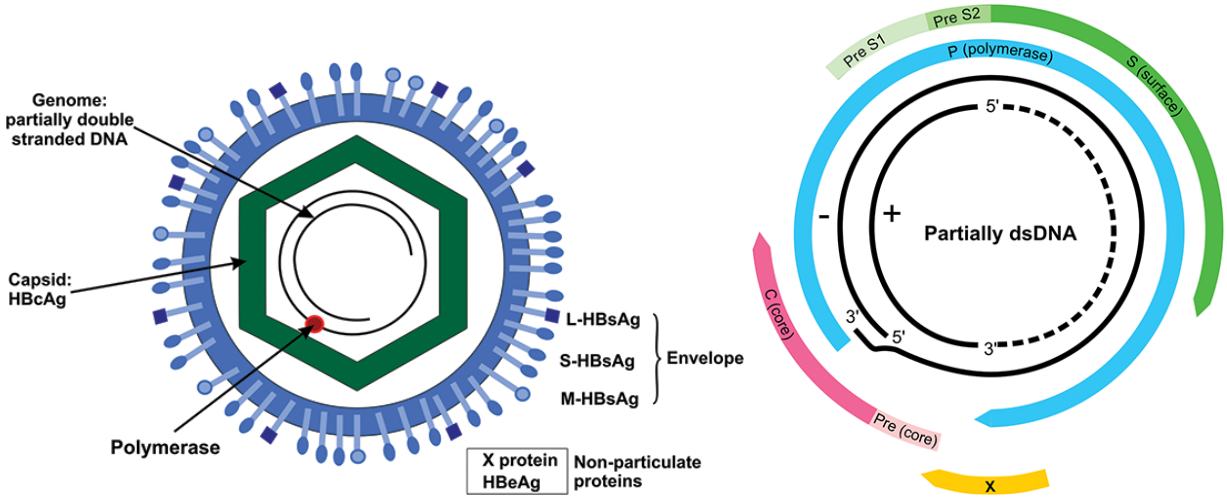

Figure 2. Structure and genome of an HBV virion.[13] In the structure on the left, the envelope is pictured in blue, the capsid in green, and the polymerase in red.

The genome is transported in virions as DNA that is only partially double stranded (gapped dsDNA), but host polymerases fill in the gaps once it enters the nucleus to make a complete, circular dsDNA genome. The complete HBV genome is only about 3.2 kb, containing four distinct genes. These genes are all protein-coding and, as with many viruses, have impressively overlapping open reading frames to conserve space. The four proteins have many names, but can be referred to concisely as C, P, S, and X (Figure 2). C is the core protein, monomers of which form the capsid. P is a polymerase, specifically a reverse transcriptase, used to produce new DNA-based viral genomes from RNA transcripts of the parent genome. S is a surface antigen protein that embeds itself in the lipid envelope upon virion exocytosis from the

host cell, and contributes to hepatocyte tropism and T-cell exhaustion[14]. These three categories of protein are common among viruses, but X protein is not as well understood[15].

Hepatitis B X protein (HBx) is an anti-antiviral protein—it combats a natural human antiviral defense mechanism. HBx protects HBV from the human structural maintenance of chromosome 5/6 complex (SMC5/6), which would normally silence the episomal HBV genome through its chromatin remodeling functions. HBx does this by simultaneously interacting with SMC5/6 and damage specific DNA-binding protein 1 (DDB1). DDB1 is a component of an E3 ubiquitin-ligase complex, which tags defective proteins for degradation. This simultaneous interaction causes the E3 complex to tag SMC5/6 with ubiquitin. This ubiquitin tag tells the host cell that a protein should be degraded, so the tagged SMC5/6 is transported to a proteasome and destroyed (Figure 3). Both direct interference with the HBx-DDB1 interaction and HBx deficiency have been shown to dramatically reduce HBV gene expression in infected human hepatocytes, while HBV gene expression is rescued by SMC5/6 knockdown with RNAi. This mechanism makes the HBx-DDB1 interaction a primary target for drug development[16,17].

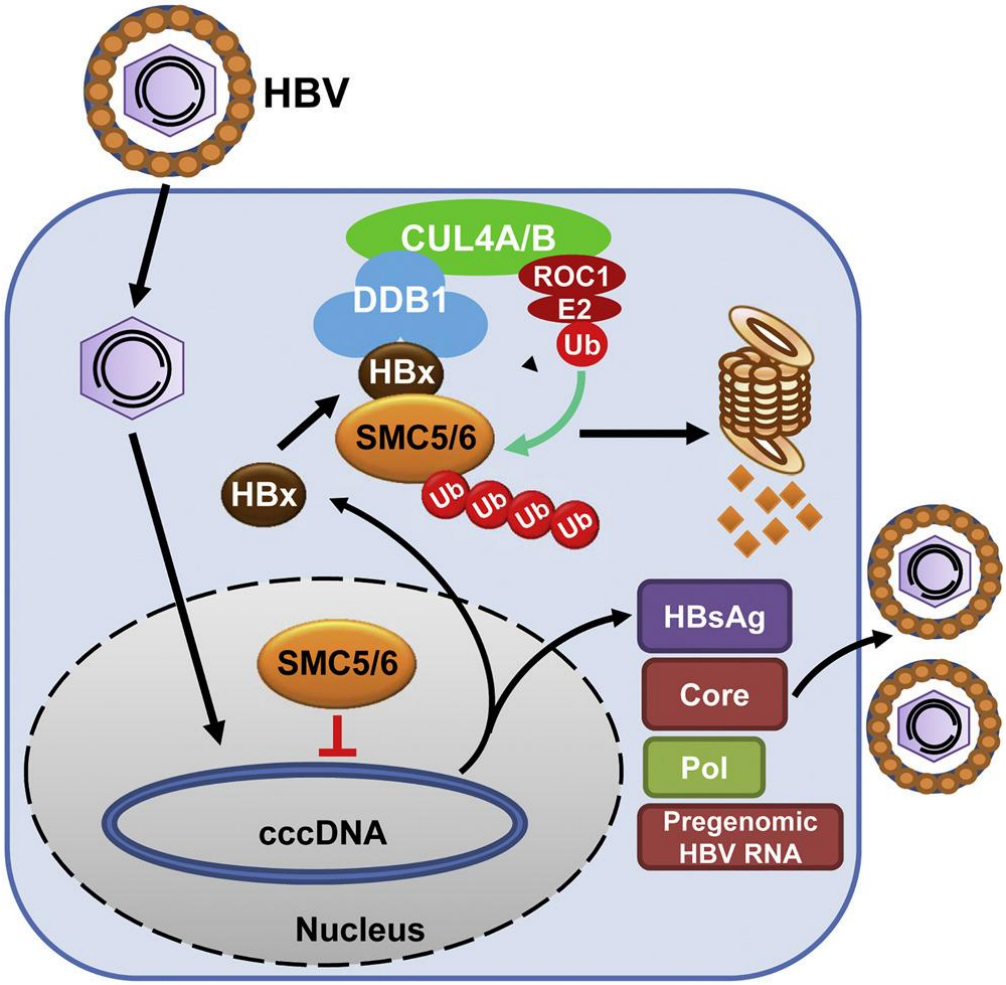

Figure 3. Schematic of HBx mechanism.[16] HBV uses HBx to target SMC5/6 for degradation through the simultaneous interaction of HBx with SMC5/6 and DDB1.

Drug Development

The process of developing a drug against a novel target begins with drug discovery. This is the identification of as many compounds as possible that show promising levels of activity against the target. Once a few active compounds are identified, their activity is confirmed and computational chemistry tools are used to assess their feasibility as a safe and deliverable drug in the human body. This process, known as lead identification, is used to select only the most

promising candidates from the library (the "lead" compounds). The lead compounds will then proceed onto lead optimization, where they are given slight chemical modifications with the aim of improving desirable qualities: potency, specificity, nontoxicity, solubility, permeability, etc[18]. Solubility and permeability both contribute to the bioavailability of a drug, which is its ability to enter blood circulation after oral ingestion[19]. Bioavailability is also influenced by chemical changes imparted by hepatic metabolism, often requiring the development of drugs that are only active after such metabolic modifications. Finally, any compounds that still show promise after lead optimization have a chance at entering clinical trials, in which more than 90% of them will fail[20]. So, to develop a clinically relevant drug, a large number of active compounds must be identified during drug discovery. Until we see dramatic improvements across the board in the drug development pipeline, efficient drug discovery is essential.

The standard approach to drug discovery is largely trial and error: testing massive libraries of chemical compounds against a target, one compound at a time. This process is known as high-throughput screening (HTS). Often, millions of compounds must be screened to identify just a handful of promising compounds. To practically enable this approach, HTS requires an assay that is consistent, automatable, and inexpensive. These assays may test for a variety of molecular events, such as ligand binding, enzymatic activity, and protein-protein interactions. They may also use a variety of output signals, such as luminescence, fluorescence, and radioactivity. The most successful HTS assays can reasonably test millions of compounds against a putative therapeutic target[18].

Other screening techniques can achieve even higher throughput than typical HTS assays. Generally, an HTS assay is used to test one compound at a time, and the compounds must come from a previously prepared library of individual compounds. A different system, known as

DNA-Encoded Library Technology (DELT), breaks both of those standards. DELT uses a pooled library of compounds that are linked to DNA barcodes indicating each compound's chemical ancestry. This is achieved by preparing the library through an elegant cycle of batched chemical modifications, barcode extension, pooling, and separation back into batches. This pooled library can then be screened by mixing it with a target protein and washing away the compounds that do not interact with the target. The barcodes of the remaining compounds can then be sequenced to reveal the chemical synthesis pathways of the highest affinity compounds (Figure 4). Through this process, billions and even trillions of compounds can be screened, making DELT three to six orders of magnitude higher-throughput than HTS[21]. However, DELT is currently limited in a few important ways. It only has the capacity to screen compounds by their affinity for a target protein, meaning it is unable to discriminate compounds that functionally interfere with the target from compounds that just passively bind the target. Also, DELT compounds must be synthesized from a limited set of chemical reactions that do not harm the DNA barcode. Finally, the large DNA barcodes on DELT compounds may interfere with the normal interactions of the compounds[22]. Although DELT is an incredibly powerful drug discovery tool, it should be used in conjunction with HTS strategies for these reasons.

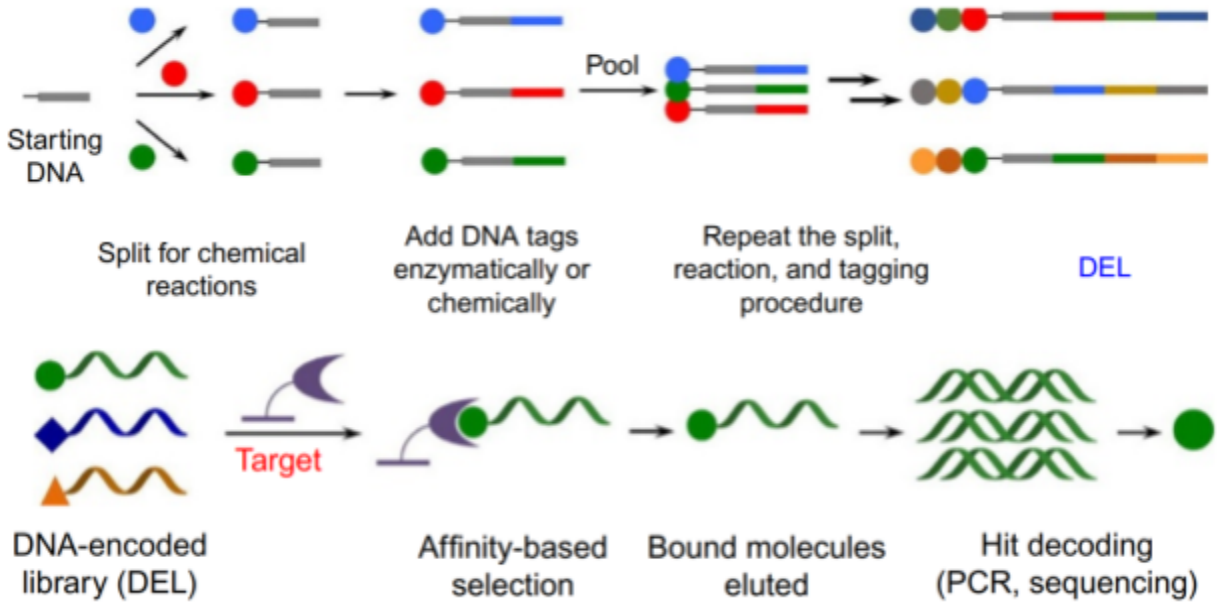

Figure 4. DELT library generation (top) and affinity selection (bottom).[21]

There are also many recent approaches to drug discovery that hope to replace the standard trial-and-error paradigm. These approaches primarily use AI-based predictive modeling enabled by new machine learning tools. For example, active learning models could eventually be used to reliably assess the likelihood that a compound will show activity against the target, thus dramatically narrowing down the list of compounds worth testing *ex silico*. These models are most successful when used in conjunction with HTS, easing the burden on the physical screening process. This can dramatically lower the cost of drug discovery and lessen the stringent requirements for an assay to qualify as HTS-capable. However, AI-based modeling is far from eliminating the utility of HTS[23,24].

These aspects of drug development are relevant to current hepatitis B research. The HBx-DDB1 interaction has already been identified as an exciting new therapeutic target, but there are no clinically available drugs targeting it yet. In 2019, Sekiba et al. created a

medium-throughput assay for the HBx-DDB1 interaction[17]. This assay uses a split luciferase system expressed in a human cell line (Figure 5). They used the assay to screen a library of 817 drugs that are already FDA-approved, and identified nitazoxanide (NTZ) as a compound that weakly interferes with the HBx-DDB1 interaction. NTZ had already been identified as inhibitory to HBV replication in cell culture, but this study specifically demonstrated the mechanism to be inhibition of the HBx-DDB1 interaction. Sekiba et al. also confirmed that NTZ inhibits expression from the HBV genome and restores SMC5 expression in human hepatocytes infected with HBV. However, NTZ has a relatively high IC50 to CC50 ratio in these systems, meaning that the inhibitory dose of NTZ is close to the cytotoxic dose. This has prevented the development of NTZ as a treatment for chronic hepatitis B. Although NTZ may not be clinically viable as a hepatitis B treatment, these preliminary data support the identification of the HBx-DDB1 interaction as a promising therapeutic target. Additionally, the initial success of the split-luciferase assay excellently highlights the potential for a higher-throughput assay to find a better compound.

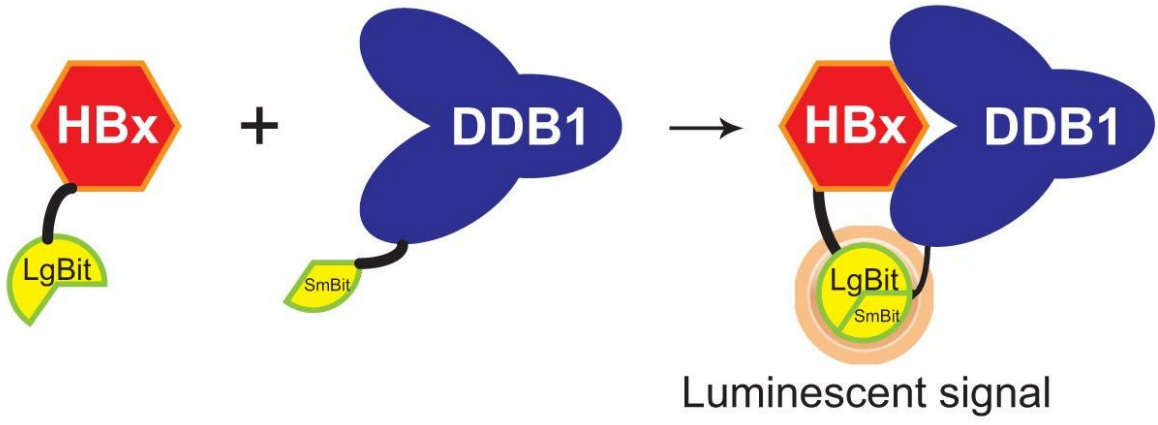

Figure 5. Split luciferase assay for the HBx-DDB1 interaction.[17] When the split domains of the luciferase in yellow are united by the HBx-DDB1 interaction, a luminescent signal is produced.

Yeast Two-Hybrid

The yeast two-hybrid (Y2H) system was pioneered in 1989[25], and it remains an excellent technique for observing robust *in vivo* protein-protein interactions[26,27]. It utilizes the Gal4 transcription factor, an activator that regulates galactose-inducible genes in wild-type yeast. In Y2H, Gal4 is split into its DNA-binding domain (Gal4BD) and its activation domain (Gal4AD), disrupting Gal4 function. Gal4BD is then translationally fused to one protein of interest (the bait) and Gal4AD to the other (the prey). These constructs are transformed into a *Saccharomyces cerevisiae* yeast strain with Gal4 enhancer regions upstream of reporter genes. If and only if the proteins of interest are interacting, Gal4 behaves normally and the reporter genes are expressed (Figure 6). HBx and DDB1 are the bait and prey in the assay I am building, so the addition of a compound that inhibits the HBx-DDB1 interaction should result in loss of expression of the reporter genes (Figure 7).

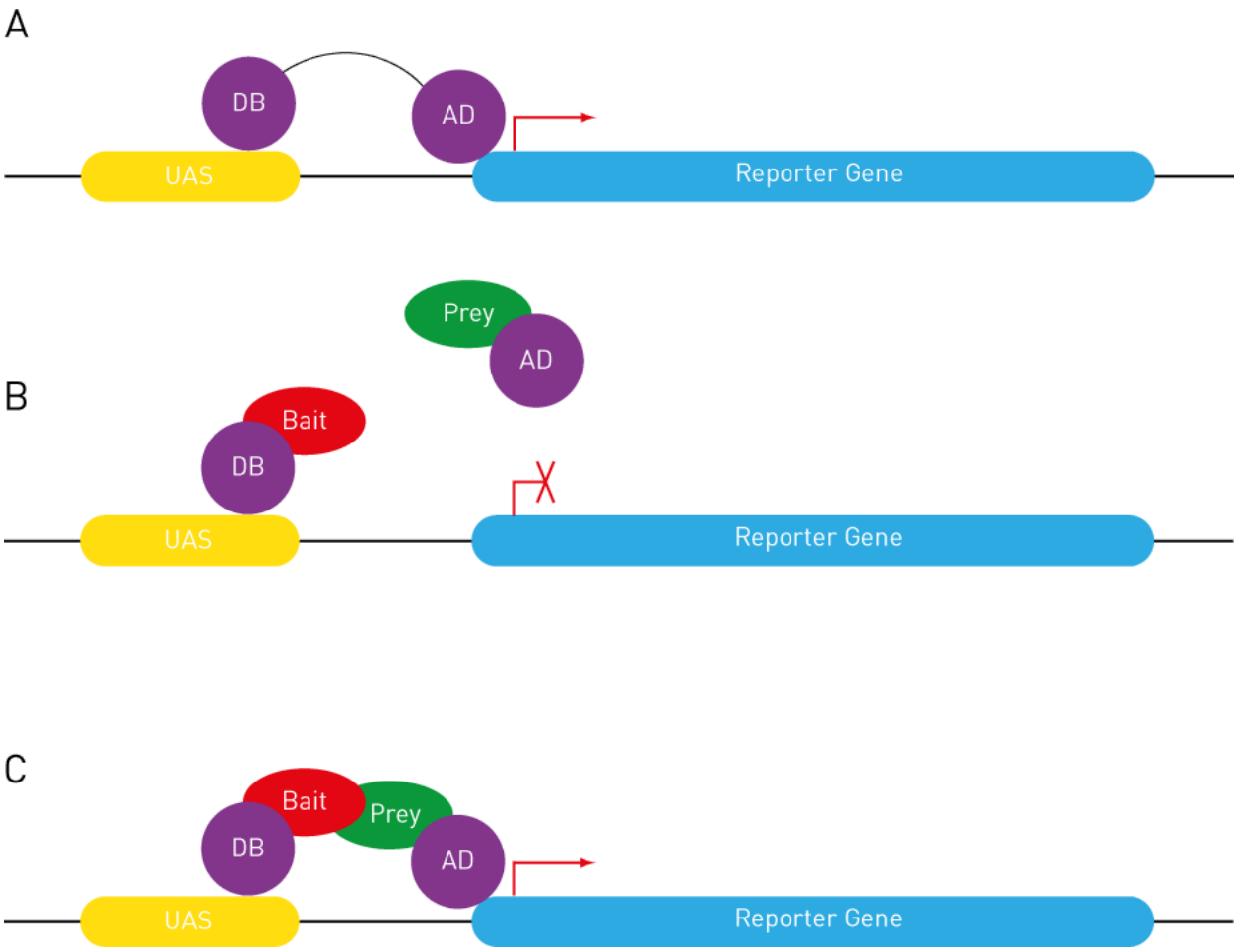

Figure 6. Basics of Y2H.[28] (A) Whole Gal4 activates reporter genes. (B) Split Gal4 is non-functional when the bait and prey are not interacting. (C) Split Gal4 regains function when bait and prey interact, activating the reporter genes.

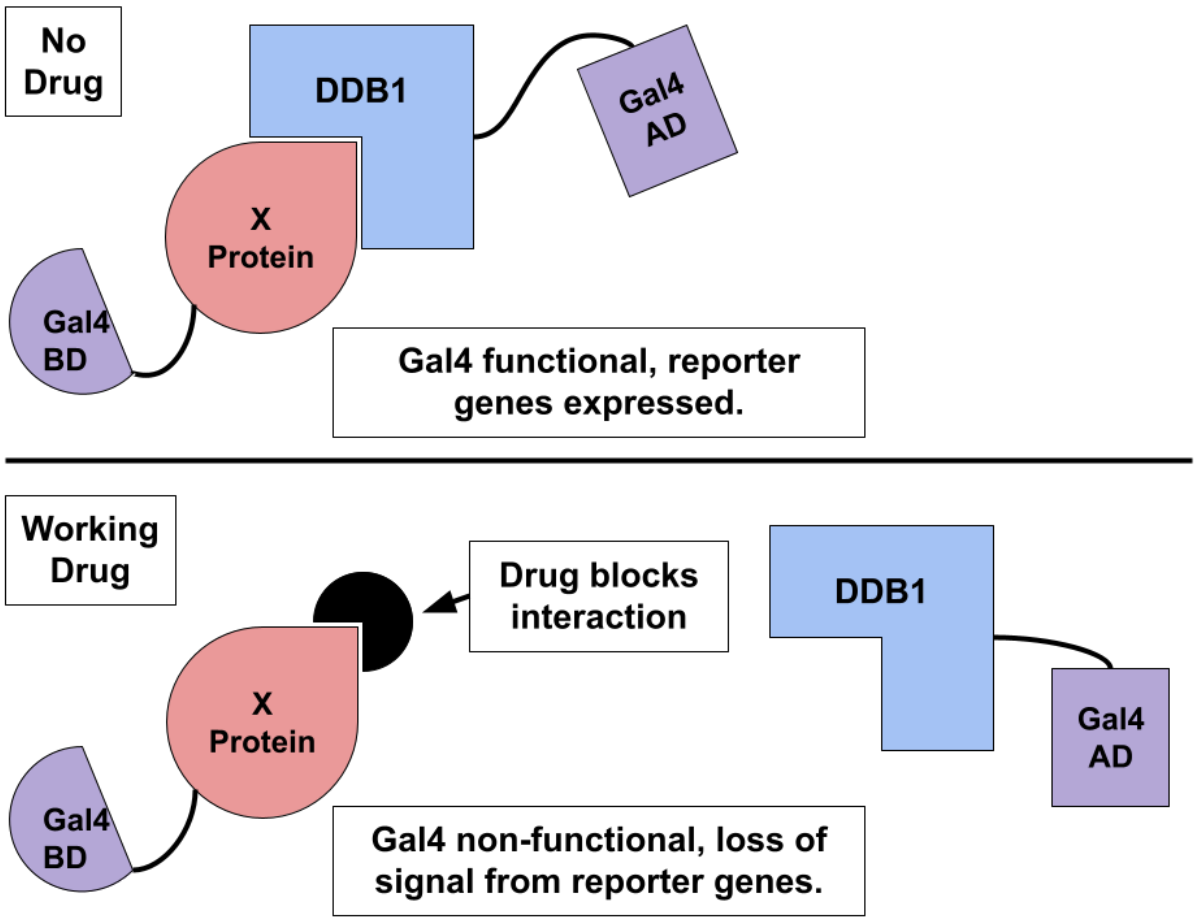

Figure 7. HBx-DDB1 Y2H assay. Split Gal4 domains are pictured in purple, HBx in red, and DDB1 in blue.

Y2H began as an assay for detecting known protein-protein interactions one at a time, but is now primarily used for detection of many unknown protein-protein interactions[26,29,30]. To achieve this, a pooled library of yeast containing different prey constructs is mated with yeast containing a single bait construct. This application is not directly relevant to the HBx-DDB1 Y2H assay as the interacting proteins are already known and constant, but some of the techniques from the newer Y2H systems should be considered in optimizing the assay for HTS. In the library mating Y2H systems there are typically multiple reporter genes that confer survival traits under regulation from different Gal4 recognition sites. By selecting for multiple survival

genes simultaneously, background growth can be mostly eliminated. Following this standard could increase the signal to noise ratio in the HBx-DDB1 assay, which is crucial to HTS. In addition, the use of different Gal4 recognition sites for each reporter reduces the false positive rate by preventing sequence-specific bypassing of the bait[26].

Y2H is a versatile system, but it has a few limitations. While false positives can be easily managed, false negatives can be harder to prevent. Niche viral proteins like HBx have been heavily selected for optimal expression in humans, and may not be expressed properly in heterologous systems like *S. cerevisiae*. This can be caused by issues with folding, post-translational modifications, or nuclear localization. Folding and localization issues can be diagnosed by the addition of epitope tags into the bait and prey constructs for use in western blots. Improperly folded proteins are usually recognized and degraded, which would be visible on a blot. Folding and localization issues might be alleviated by truncating the proteins of interest to express only their interacting domains, rearranging the fusion constructs, or even by using an alternative Y2H strain with different available chaperones. The HBx-DDB1 interaction is known to occur between a well-characterized α-helical motif in HBx and a hydrophobic pocket in DDB1[31], so truncating these proteins is feasible. Post-translational modification issues can be harder to identify if they are not already known, but can be solved by co-expression of the modifying enzyme.

Once a Y2H assay is working, there are many ways to improve it for HTS applications. The most significant improvement will come from replacing the reporter genes. The preferred detection methods for HTS are luminescence and fluorescence[18]. Luminescence releases light as the result of a luciferase enzyme cleaving a luciferin substrate. Luminescent assays have almost no background noise, but the luciferin can be expensive[32]. Fluorescence releases light when

stimulated with a different wavelength of light, but can have moderate background noise. The best reporters use a luminescent system that synthesizes its own luciferin (autoluminescence), but this can be hard to engineer. Naturally bioluminescent organisms contain metabolic pathways to create their own luciferin. If the enzymes in a luciferin synthesis pathway are known, they can be harnessed to create new autoluminescent organisms. In 2018, Kotlobay et al. elucidated a eukaryotic luciferin synthesis pathway in the fungus *Neonothopanus nambi* and demonstrated its function in other eukaryotic species[33]. Building this pathway in *S. cerevisiae* requires the addition of five enzymes: one luciferase (Luz), two for luciferin synthesis from caffeic acid (HispS and H3H), one for recycling used luciferin (CPH), and one to provide a necessary post-translational modification to HispS (Figure 8). Although caffeic acid is a common metabolite, *S. cerevisiae* do not naturally produce it. Three additional genes can be introduced to enable caffeic acid synthesis from tyrosine, or caffeic acid can simply be added to the growth media as it is relatively inexpensive.

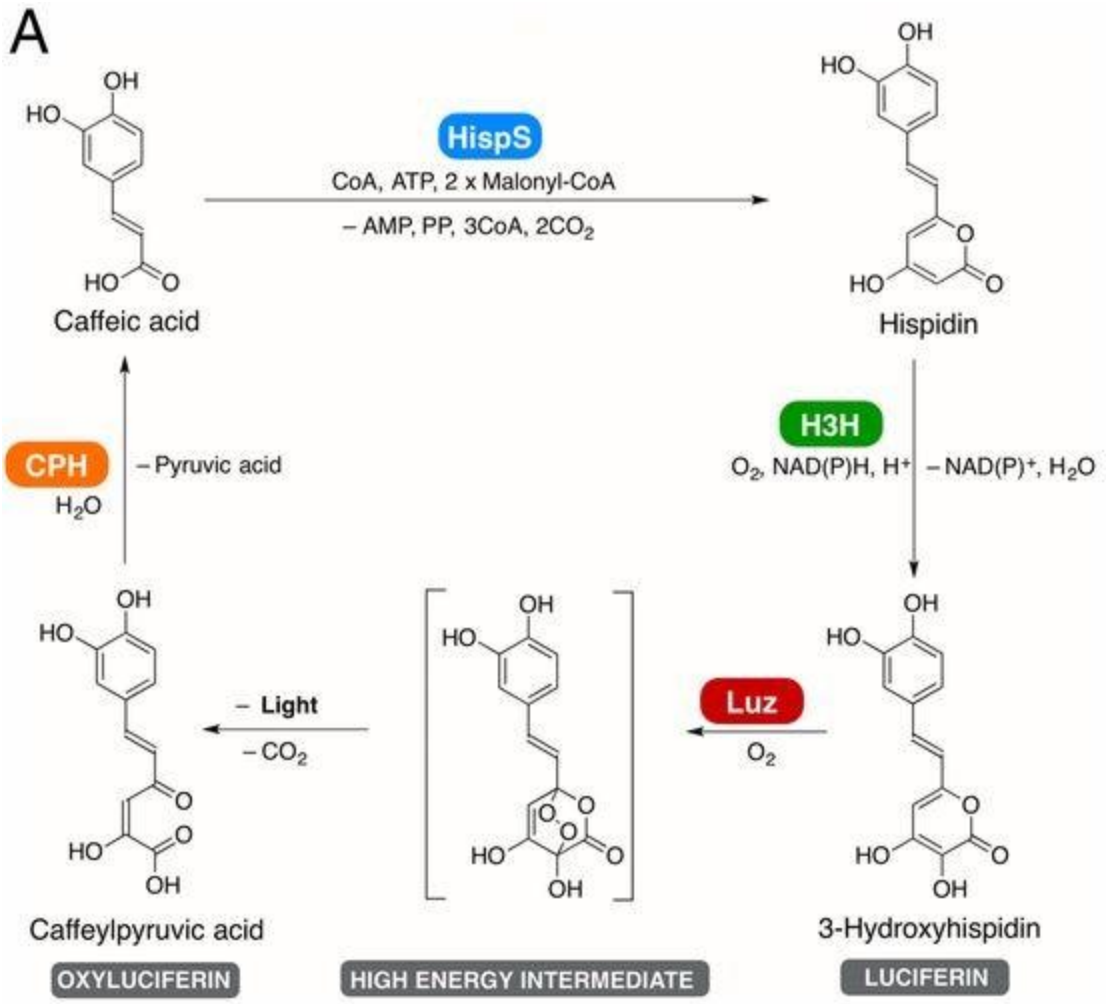

Figure 8. *Neonothopanus nambi* luciferin synthesis.[33]

Two other optimizations could improve the reliability of the assay even further: genome integration and codon optimization. Integrating the bait and prey constructs into the genome of the yeast, rather than expressing them from plasmids, will stabilize their expression. This will allow the yeast to be grown in rich non-selective media, boosting assay signal and speed. Codon optimization is the process of replacing rare codons in coding sequences with more common, synonymous codons. The prevalence of synonymous codons varies dramatically across different species. The low tRNA levels associated with rare codons can appreciably decrease the

translation rate of mRNAs containing rare codons, limiting a protein's expression[34]. Codon optimization of the Y2H constructs and autoluminescent reporter genes could provide another small boost to the assay signal and speed.

In this research, my aim is to build an HTS-capable assay for the HBx-DDB1 protein-protein interaction. This interaction is essential to the survival of HBV in chronically infected individuals, and an HTS-capable assay could enable the discovery of a drug that interferes with this interaction. I aim to build this assay using a Y2H system with an autoluminescent reporter and other optimizations to increase assay efficacy and decrease assay cost. I hope to eventually see this work used in a high-throughput screen for novel hepatitis B drugs.

Materials and Methods

Cloning of Y2H fusion constructs

Four Y2H fusion proteins were constructed: Gal4BD-HBx, Gal4AD-DDB1, Gal4BD-DDB1, and Gal4AD-HBx (henceforth abbreviated as GBD-X, GAD-DDB1, GBD-DDB1, and GAD-X, respectively). Although only one bait and prey are necessary for a Y2H assay, assembling all four possible Y2H fusion proteins enables testing the proteins of interest in both roles: bait or prey. Plasmids containing these constructs were assembled with a traditional cloning procedure using vectors pGBDU-C1 and pGAD-C1, which were gifts from Elizabeth Craig. Inserts were amplified via polymerase chain reaction (PCR) using the proofreading Q5 polymerase from New England Biolabs (NEB; Ipswich, MA, USA). The PCR primers were designed to add flanking BamHI and SalI restriction sites to the amplicons, as well as a flexible linker sequence between the Gal4 domain and the bait or prey protein. This linker encodes the flexible peptide sequence GSGGSG, decreasing the chance of interference between the two proteins in the fusion construct. All primers were ordered from Integrated DNA Technologies (Coralville, IA, USA). The HBx insert was PCR-amplified from a plasmid reconstruction of the HBV genome, which was a gift from Dan Loeb. The DDB1 insert was PCR-amplified from pDDB1, a gift from Cheryl Arrowsmith. The vectors and inserts were column purified (Bio Basic; Markham, Ontario, Canada), digested with restriction enzymes BamHI and SalI (NEB) and purified with gel extraction (Qiagen; Hilden, Germany). Digested vectors and inserts were mixed in a 1:3 molar ratio, respectively, and ligated with T4 ligase (NEB) to form plasmids containing the complete fusion constructs.

Ligated plasmids were transformed into competent DH5α *Escherichia coli* (Molecular Cloning Laboratories; South San Francisco, CA, US). Transformed *E. coli* were streaked onto

LB Miller agar media containing 100 ug/mL ampicillin. Isolated transformant colonies were grown in LB Miller ampicillin broth overnight and miniprepped (Bio Basic). Proper assembly of miniprepped plasmids was verified with Sanger sequencing by ELIM Biopharmaceuticals (Hayward, CA, USA).

Cloning of Myc-tagged Y2H constructs

3xMyc epitope tags (3M) were cloned into each of the four Y2H constructs to create GBD-3M-X, GAD-3M-DDB1, GBD-3M-DDB1, and GAD-3M-X. The 3xMyc tag is a short, repeated peptide sequence that can be specifically targeted with high-affinity commercially produced antibodies. This eliminates the need for costly custom antibodies that usually generate less reproducible results.

The four Y2H plasmids were PCR-amplified to produce linear vectors flanked with BamHI and XmaI restriction sites. The 3xMyc insert was PCR-amplified to produce an amplicon with the same restriction sites. The vectors and inserts were digested with BamHI-HF and XmaI (NEB). Quick CIP (NEB) was added to the vector digestion reactions to decrease background after transformation resulting from vector self-ligation. Digested vectors and inserts were column purified. Ligation, transformation, and preparation of the 3xMyc-tagged Y2H plasmids were identical to the prior fusion construct cloning. Proper assembly of Myc-tagged plasmids was verified with an informative restriction digest using BbsI (NEB).

Cloning of truncated HBx Y2H constructs

Truncated HBx (tX) versions of Y2H constructs containing HBx were generated using Golden Gate Assembly, creating GBD-3M-tX and GAD-3M-tX. These constructs contain a

version of HBx with residues 2-44 and 141-154 removed. This has been shown to approach the largest possible truncation while maintaining HBx function[35].

Primers were designed to amplify around the pGBD-3M-X and pGAD-3M-X backbones, and within the HBx coding sequence, leaving out the undesired residues. In addition, the primers added PaqC1 restriction sites. PaqC1 is a type IIS restriction enzyme, meaning it cuts outside of its recognition sequence. This is useful because it enables seamless cloning, so modifications in the middle of coding sequences can be made easily. In addition, these unique restriction enzymes can be used for a simplified plasmid assembly procedure known as Golden Gate Assembly, which combines the digestion and ligation steps of traditional cloning into one reaction.

The PCR-amplified vectors and inserts were mixed in a 1:2 molar ratio, respectively, with PaqCI (NEB) and T4 ligase for Golden Gate Assembly. Ligation, transformation, and preparation of the truncated HBx plasmids were identical to the prior fusion construct cloning. Proper assembly of pGBD-3M-tX and pGAD-3M-tX was verified with an informative restriction digest using FspI (NEB).

Transformation and culturing of S. cerevisiae Y2H strains

Verified Y2H construct plasmids were transformed into competent PJ69-4A *S. cerevisiae* to produce Y2H strains. PJ69-4A was a gift from Elizabeth Craig. PJ69-4A is a Y2H-compatible strain of *S. cerevisiae* that has deleted *gal4* and *gal80*, allowing galactose-independent, Y2H-specific activation of Gal4 regulated genes[36]. Competent PJ69-4A cells were generated using the Frozen-EZ Yeast Transformation II Kit from Zymo Research (Irvine, CA, USA). Functional Y2H strains require both the bait and the prey constructs, so plasmid transformations were performed sequentially, with single-plasmid transformants being used to generate more competent cells for the second plasmid transformation. Transformed yeast were cultured on

dropout media lacking the appropriate amino acids for the selection genes in their plasmids, and single-colony isolates were selected for proliferation and storage. Cells were grown in selective liquid culture to an OD_{600} of about 1.0, then stored at -80°C in 25% glycerol.

In total, ten *S. cerevisiae* strains were generated for use in this research, detailed in Table 1. These strains include two complete Y2H testing strains, a positive control strain (generated using a plasmid containing reconstructed Gal4 provided by Eric Lewellyn), a negative control strain containing empty Gal4 domain vectors, one strain containing each Myc-tagged Y2H construct, and two strains containing the truncated HBx constructs.

Strain	Description	Plasmids
XBDPa	HBx Bait, DDB1 Prey, MATa	pGBD-X, pGAD-DDB1
DBXPa	DDB1 Bait, HBx Prey, MATa	pGBD-DDB1, pGAD-X
rGAL4a	Reconstructed Gal4, MATa	pGal4, pGBDU-C1
sGAL4a	Split Gal4, MATa	pGBDU-C1, pGAD-C1
XBMa	HBx Bait Myc tag, MATa	pGBD-3M-X
DPMa	DDB1 Prey Myc tag, MATa	pGAD-3M-DDB1
DBMa	DDB1 Bait Myc tag, MATa	pGBD-3M-DDB1
XPMa	HBx Prey Myc tag, MATa	pGAD-3M-X
tXBMa	Truncated HBx Bait Myc tag, MATa	pGBD-3M-tX
tXPMa	Truncated HBx Prey Myc tag, MATa	pGAD-3M-tX

Table 1. *S. cerevisiae* strains produced in this research and their descriptions.

Auxotrophic rescue testing

The PJ69-4A background strain for Y2H systems used in this research contains three reporter genes under Gal4 regulation: *HIS3*, *ADE2*, and *lacZ*. *HIS3* and *ADE2* encode enzymes

required for synthesis of essential amino acids histidine and adenine, respectively, and would typically be constitutively expressed in *S. cerevisiae*. However, the naturally occurring versions of these genes have been broken in the PJ69-4A parent strain, so that only cells with a functioning Y2H system will be able to synthesize these amino acids. Therefore, when PJ69-4A-derived Y2H strains are inoculated onto dropout media lacking histidine or adenine, growth is an indicator of bait-prey interaction.

Single-colony isolates from strains XBDPa, DBXPa, rGAL4a, and sGAL4a were streaked onto dropout media lacking histidine or adenine (US Biological; Swampscott, MA, USA) and incubated at 30°C for 5-7 days before imaging.

Protein extraction from S. cerevisiae

Proteins from Y2H strains containing Myc-tagged constructs and controls were extracted using trichloroacetic acid (TCA) precipitation in preparation for western blot. Yeast were grown in selective media to an OD_{600} of about 1.0. The equivalent of 5 mL of culture at OD_{600} 1.0 was taken from each culture and centrifuged briefly at 10,000 RCF. Pellets were resuspended in 1 mL water, pelleted again, then resuspended in 20% TCA. Suspensions were flash frozen in liquid nitrogen, then thawed on ice. About 200 uL of 0.2 mm glass grinding beads were added to each sample and vortexed vigorously for 3 minutes with breaks on ice. 750 uL 5% TCA was added to each sample, vortexed, then placed in a fresh microtube for 10 minutes centrifugation at 14,000 RCF 4°C. Resulting pellets were washed with 500 uL cold acetone, then resuspended in 50 uL 1M TRIS pH 8.0 and about ten glass grinding beads. 80 uL 2X Laemmli Sample Buffer (Bio-Rad Laboratories; Hercules, CA, USA) with 5% β-mercaptoethanol was added to each sample, vortexed, and boiled for 5 minutes at 95°C. Samples were centrifuged at 10,000 RCF for

5 minutes and 100 uL of supernatant was transferred to a new microtube for immediate loading onto a polyacrylamide gel.

Western blot of Myc-tagged Y2H constructs

15 uL of freshly prepared protein extracts were loaded onto a precast 20% polyacrylamide gel (NuSep; Germantown, MD, USA) with Precision Plus Protein Kaleidoscope Standard (Bio-Rad). Gel was run in a Mini-PROTEAN Tetra Cell (Bio-Rad) in Running Buffer (25 mM Tris, 190 mM glycine, 0.1% SDS) at 100 V for 1.5 hours. The gel contents were then transferred to a nitrocellulose membrane (Schleicher & Schuell BioScience; Keene, NH, USA) in Transfer Buffer (25 mM Tris, 190 mM, 20% methanol) at 25 V for 8 hours at 4°C. Subsequent steps were all performed at room temperature with agitation.

The membrane was blocked with 5% nonfat instant milk (Nestle; Vevey, Switzerland) in TBST (20 mM Tris pH 7.5, 150 mM NaCl, 0.1% Tween 20) for 1 hour, then incubated with mouse anti-c-Myc (Santa Cruz Biotechnology; Dallas, TX, USA) primary antibody diluted to 1 ug/mL in blocking solution for 2 hours. The membrane was washed four times with TBST for 5 minutes each, then incubated with anti-mouse HRP-linked secondary antibody (Invitrogen; Thermo Fisher Scientific; Waltham, MA, USA) diluted to 0.16 ug/mL in blocking solution for 1 hour. The membrane was washed four times with TBST for 5 minutes each, then two times with TBS for 5 minutes each. 1 mL of SuperSignal West Femto chemiluminescent HRP substrate (Thermo Fisher Scientific) was added directly to the membrane for a 5 minute incubation prior to imaging.

Results

Validation of Y2H plasmids

Every Y2H construct plasmid was assessed for proper assembly after cloning. Plasmids pGBD-X, pGAD-DDB1, pGBD-DDB1, and pGAD-X were validated with Sanger sequencing across the ligation junctions, where errors are most likely. Figure 9 shows an alignment from one of these reactions verifying the expected sequence across the 3' junction of HBx in pGBD-X. Trace data across all eight junctions from the four Sanger-validated plasmids matched the expected sequences exactly (data not shown).

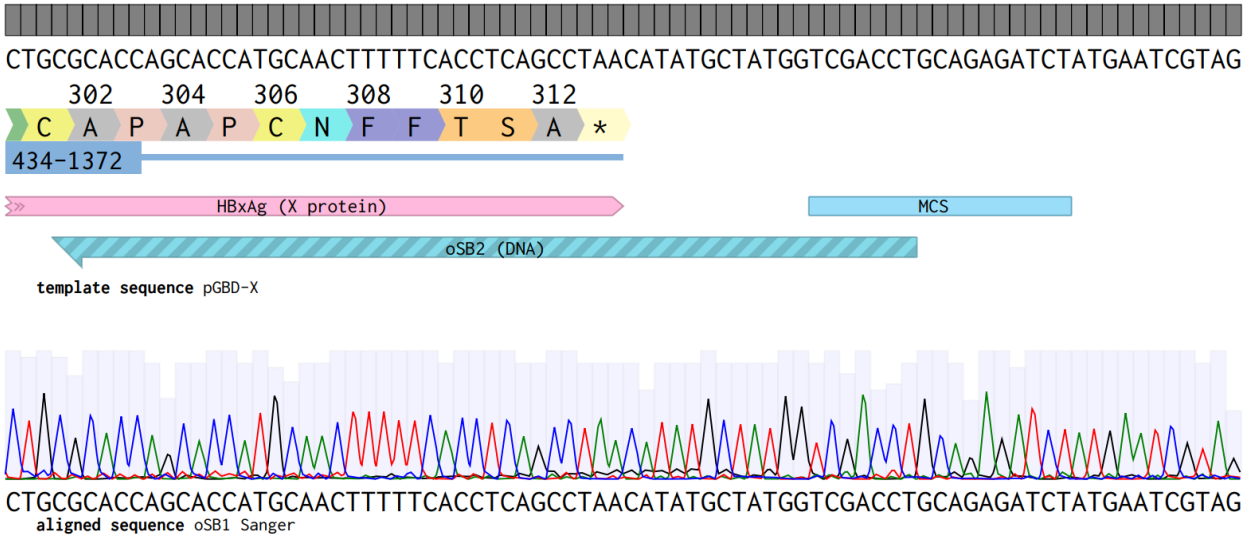

Figure 9. Representative alignment of Sanger sequencing data confirming successful ligation of pGBD-X. The displayed chromatogram was generated from a primer 400 bp upstream. 40 bp on either side of the 3' junction of HBx (pink) are displayed. Also displayed are the reverse sequencing primer (oSB2) used to verify the 5' junction, the remnants of the multiple cloning site (MCS), and the C-terminal translation of HBx. Alignment and annotations performed with Benchling.

Myc-tagged and truncated HBx plasmids were verified with informative restriction digests. These digests are designed to cut the target plasmid into distinctly different fragments when the insert is present than when the vector self-ligates. This is usually achieved by using a restriction enzyme with a cut site in the desired insert. The DNA fragments are run on an agarose gel and the resulting banding patterns are compared to a digestion of the vector plasmid as a negative control. Figure 10 shows the results of this validation for pGAD-3M-X. Informative digests were used to validate the remaining plasmids as well (Appendix).

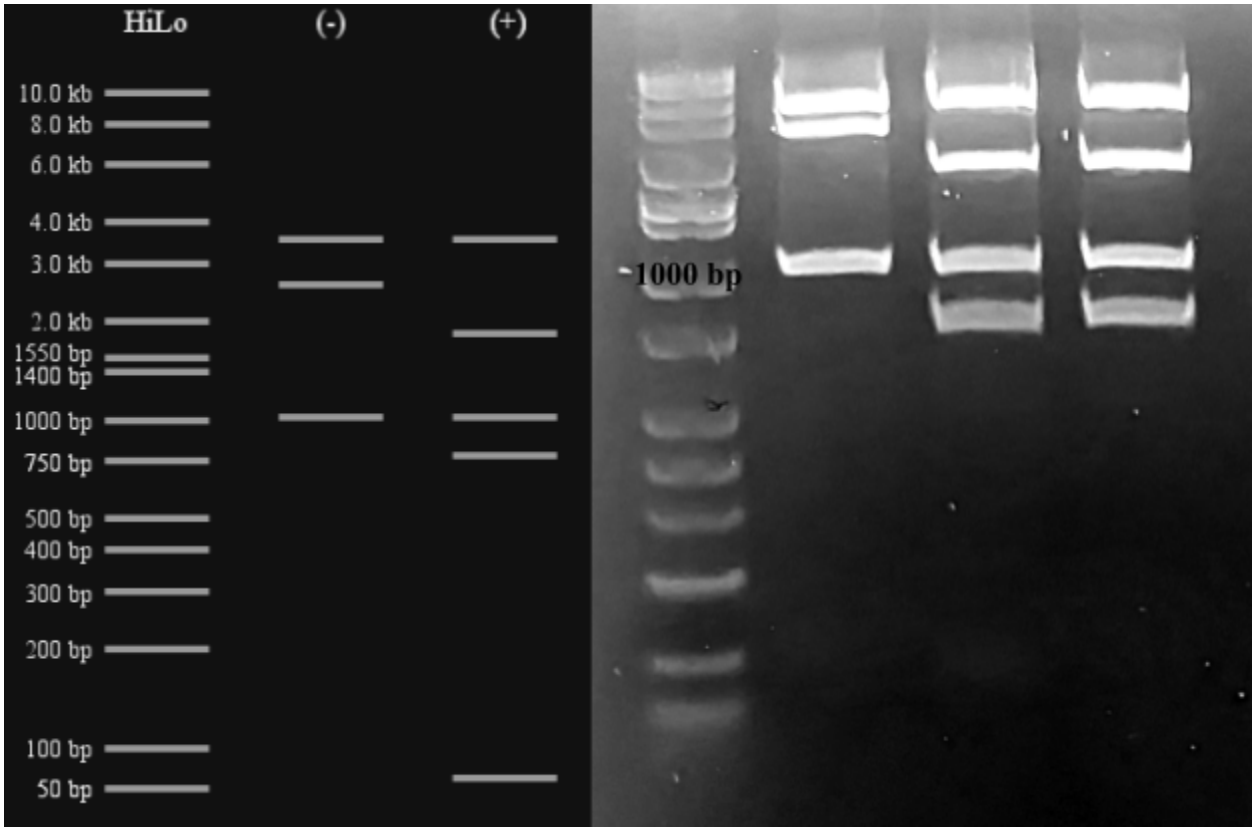

Figure 10. Informative restriction digest with BbsI on a 2% agarose gel confirming successful ligation of pGAD-3M-X. Anticipated banding patterns are displayed in a virtual digest on the left, with (-) and (+) representing unsuccessful or successful cloning, respectively. Left to right on the gel: Hi-Lo DNA marker (Bionexus; Kansas City, MO, USA), pGAD-X negative control, transformation isolate 1, and isolate 2.

Qualitative evaluation of Y2H strains

Y2H strains XBDPa and DBXPa were tested for expression of all three reporter genes and compared to control strains rGAL4a and sGAL4a. Y2H-regulated expression of the *HIS3* and *ADE2* reporter genes can rescue the respective auxotrophries of the PJ69-4A parent strain. Therefore, they can easily be assayed on dropout histidine or dropout adenine media, where growth indicates a functional Y2H system.

Figure 11 shows the two experimental and two control *S. cerevisiae* strains after 7 days of growth at 30°C on dropout histidine media. Substantial growth is clearly visible in the positive control and DBXPa strains, with both showing many colonies >1 mm in diameter, compared to almost no healthy growth from the negative control or XBDPa strains. This is an exciting indicator for the assay in DBXPa, but it is surprising that there is such a striking difference between the two experimental strains. In addition, while the growth may look similar between rGAL4a and DBXPa, healthy colonies were only obvious in DBXPa after about 5 days of growth, compared to about 3 days for rGAL4a.

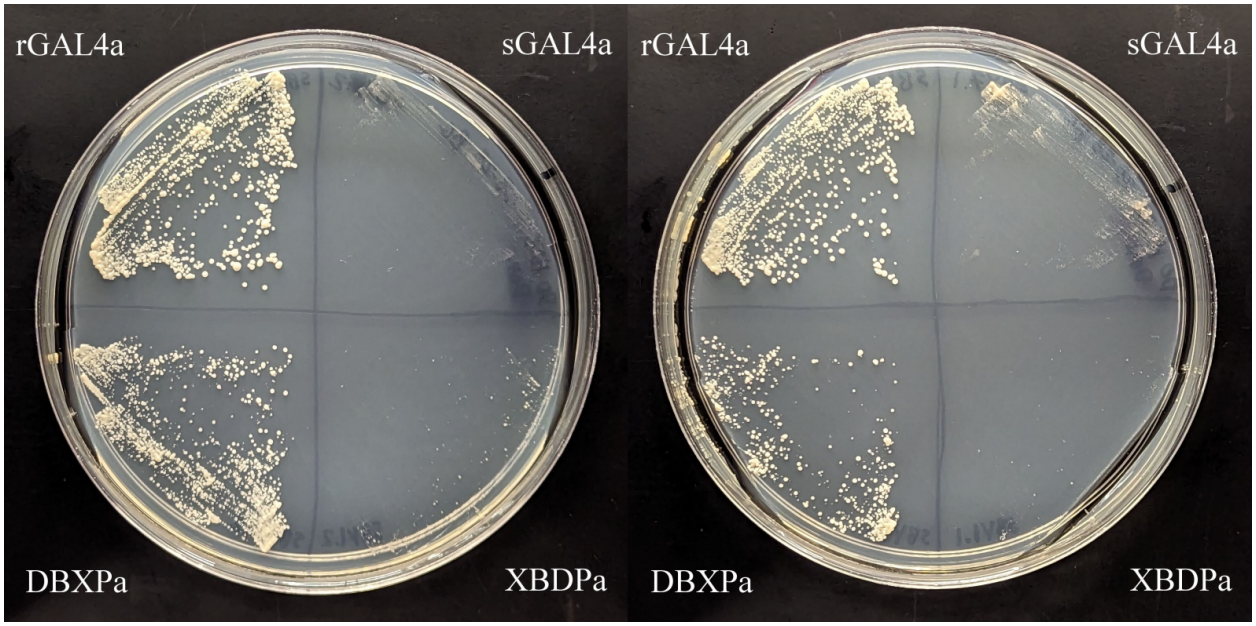

Figure 11. Y2H strain growth on dropout histidine media. Four strains were streaked in replicate, with rGAL4a as a positive control, sGAL4a as a negative control, and experimental strains DBXPa and XBDPa.

Figure 12 shows the same strains on dropout adenine media after 5 days of growth. A similar pattern is seen on the dropout adenine media as on the dropout histidine media: rGAL4a and DBXPa are clearly growing and sGAL4a and XBDPa are clearly not growing. However, it is also much more obvious on these plates that DBXPa is significantly less healthy than rGAL4a. DBXPa growth is much thinner than rGAL4a and has a deep pink coloration.

The pink coloration of the growth is a unique feature of adenine auxotrophy. The substrate for ADE2 is P-ribosylaminoimidazole, an intermediate in the adenine synthesis pathway that can be oxidized into a red pigment[37]. If yeast are ADE2 deficient, then this intermediate accumulates and eventually turns the colonies red. As such, pink coloration can be used as a very rough approximation of ADE2 deficiency.

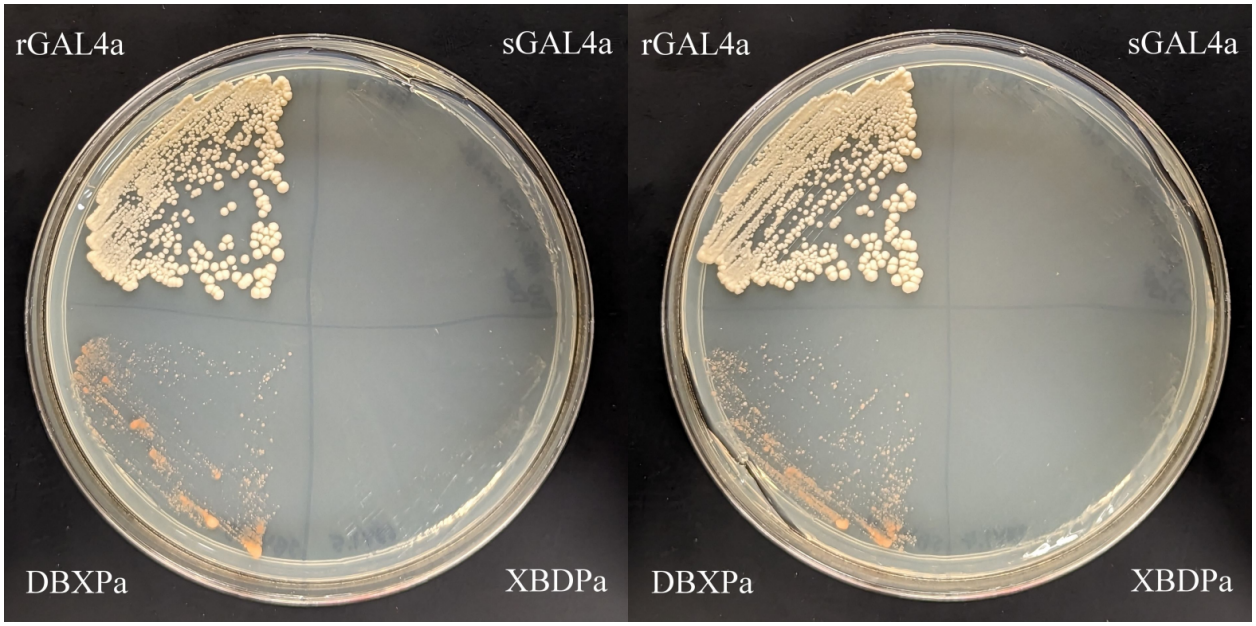

Figure 12. Y2H strain growth on dropout adenine media. Four strains were streaked in replicate, with rGAL4a as a positive control, sGAL4a as a negative control, and experimental strains DBXPa and XBDPa. Pink coloration on dropout adenine media is a rough indicator of ADE2 deficiency.

The final reporter gene, *lacZ*, can be used in various chromogenic and chemiluminescent assays. The Y2H strains were tested with three such assays, but a signal was only ever observed from rGAL4a. It is likely that only a Galacton-Star (Thermo Fisher Scientific) assay would be sensitive enough to observe the *lacZ* signal from the current Y2H strains, which may be worth testing in the future.

Western blot assessment of Y2H construct expression and stability

Protein was extracted from single-transformant yeast strains containing Myc-tagged Y2H constructs for western blotting. Figure 13 shows the resulting banding. The first five lanes contain protein from cells with experimental Y2H constructs. There are extremely faint bands at the expected sizes in four of these lanes, with no band for DPMa (lane 2). Lanes 6 and 7 contain

negative control strains and show no banding as expected. Lanes 8 and 9 contain positive control strains and show significant banding around the expected size, as well as at many other sizes. These positive control strains contain a genome integrated construct with a 13xMyc tag.

As approximately the same amount of cells were used for each protein extraction, comparison of expression between the experimental constructs should be possible. Among the experimental constructs, GBD-3M-X appears to have the highest expression by a significant margin. The DDB1 constructs appear to have generally lower expression than the full length HBx constructs. The truncated HBx construct GAD-3M-tX band is so faint that it might be an artifact, and is clearly not an improvement on the full length HBx constructs.

Despite using approximately the same amount of cells for each protein extraction, quantitative comparison between the positive controls and experimental strains is not possible. This is due to many factors: 1) The 13xMyc tag will generate a higher signal than the 3xMyc tags on the Y2H constructs. 2) There is only one copy of the genome integrated positive control proteins, while the Y2H constructs are on plasmids with high-copy 2μ origins of replication (40-60 copies per cell). 3) The proteins are expressed under different strength promoters. Despite this, it seems reasonable to qualitatively evaluate the massive difference in signal as an indicator that something is limiting the expression or stability of the Y2H constructs, particularly GAD-3M-DDB1.

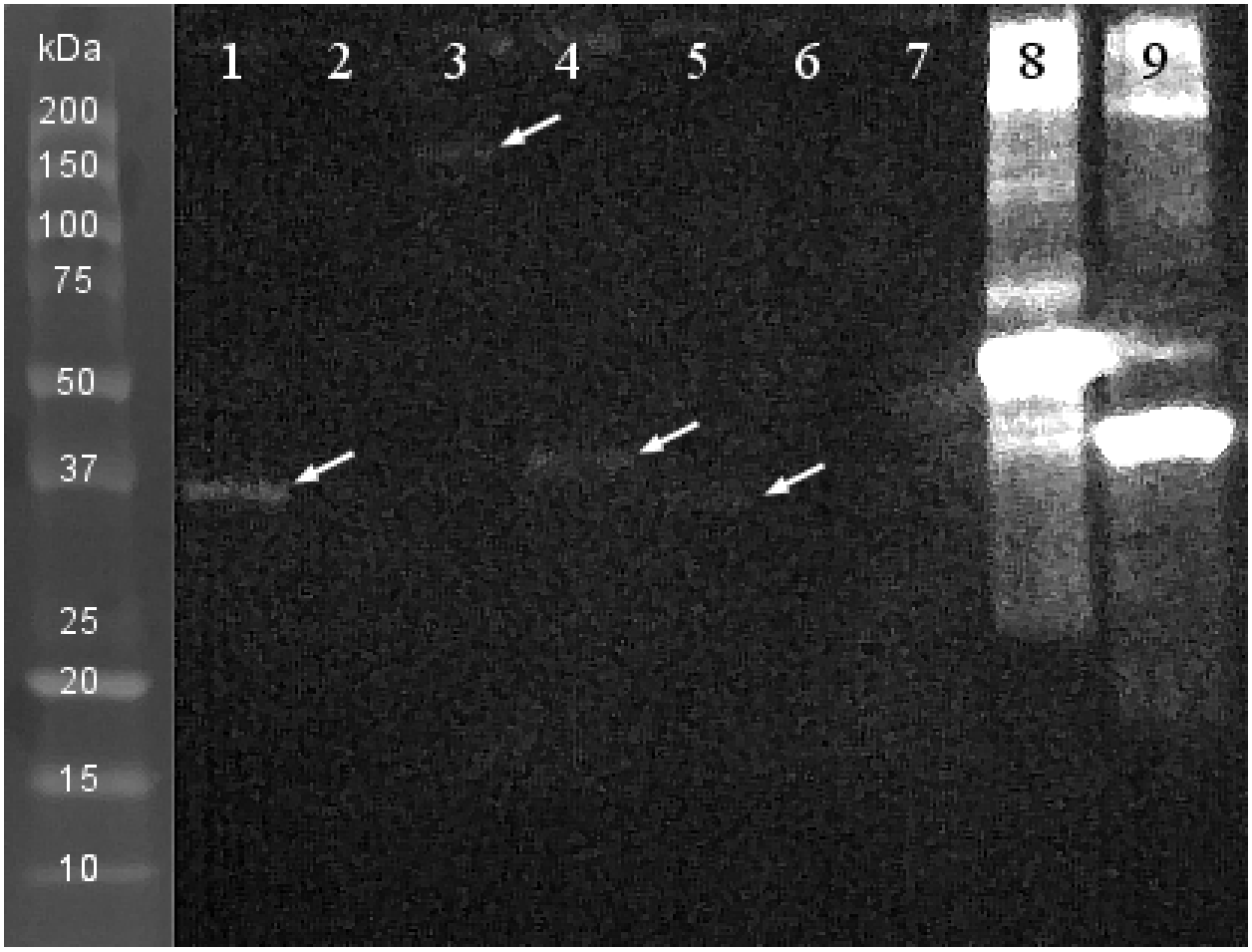

Figure 13. Western blot of Myc-tagged Y2H constructs from *S. cerevisiae* protein extractions. Protein was extracted with TCA precipitation from single-transformant yeast strains and control strains. Left to right: Kaleidoscope Protein Standard, 1) XBMa, 2) DPMa, 3) DBMa, 4) XPMa, 5) tXPMa, 6) PJ69-4A (negative control for Y2H strains), 7) ELY359 (negative control for ELY strains), 8) ELY269.1 (positive control with genome-integrated Rbd2-13xMyc), 9) ELY270.2 (theoretically redundant with ELY269.1). Expected sizes of Myc-tagged proteins: 1) 38.53 kDa, 2) 146.96 kDa, 3) 148.95 kDa, 4) 36.54 kDa, 5) 30.79 kDa, 6) N/A, 7) N/A, 8) 45.14 kDa, 9) 45.14 kDa. Arrows are pointing at extremely faint banding that are close to the expected sizes for their lane. Image annotated, despeckled, and sharpened in ImageJ.

Discussion

Current assay capabilities and limitations

In its current state, the DBXPa Y2H strain demonstrates a weak ability to assay the HBx-DDB1 interaction, while the XBDPa strain shows no signal. From the auxotrophic rescue testing, it appears that DBXPa has an excellent *HIS3* signal but poor *ADE2* signal, but this may not be the case. *HIS3* is known to exhibit leaky expression in many yeast strains, expressing at low levels without Gal4 activation. This phenomenon could be interacting with the low levels of Y2H activity in DBXPa to inconsistently modulate the observed signal. For this reason, many large library screens utilizing the Y2H system recommend adding around 5 mM 3-amino-triazole, an inhibitor of the *HIS3* gene product, to the growth media[38]. Adenine auxotrophic rescue is generally considered a more robust reporter system.

Consistent between the histidine and adenine auxotrophy testing was the lack of signal from XBDPa, despite a clear signal from DBXPa. This result was unexpected, and could provide insight into the current obstacles of the assay with further investigation.

The presence of a clear signal from DBXPa is very promising, as it shows that a Y2H assay is possible with these protein pairs. Importantly, it demonstrates that HBx and DDB1 are likely able to fold correctly in *S. cerevisiae* to some extent. However, the current assay signal is far too low to be HTS-capable.

Next steps for improving the assay

A promising first step to improving the assay is to attempt truncating DDB1, rather than HBx. The results of the western blot suggest that XBDPa might not show any Y2H activity because GAD-3M-DDB1 is being degraded completely. Performing another western blot with

3xMyc-tagged split Gal4 domains that aren't fused to other proteins and 3xMyc-tagged full-length Gal4 would be excellent controls to confirm that the issue isn't elsewhere in the assay. Previous research shows that the BPB domain (residues 400-704) of DDB1 is not necessary for HBx binding, which would be a 305 residue 33.6 kDa deletion from the 1324 residue 127 kDa protein[31,39]. Such a large deletion could significantly lower the chances for folding errors during expression in a heterologous system.

If expression remains a problem, techniques to increase gene transcription or to improve stability after folding might help. The current promoter for the Y2H constructs is ADH1, which is a medium strength promoter. Simply replacing this with a strong promoter could improve the assay signal. Changing the activation and DNA-binding domains of the fusion constructs to protein domains with higher activity than Gal4 could also improve signal, as well as folding. Seeing as only one of the DDB1 constructs appeared fully degraded, it is possible that the Gal4 activation domain does not interact well with DDB1. Alternatively, integrating stabilizing peptides into the fusion constructs or trans-acting chaperones into the genome could improve the stability of these expressed constructs.

The assay could also benefit from integration of the Y2H constructs into the genome. While this will likely lower the number of Y2H genes per cell, they will be more stably expressed. In addition, yeast strains with genome integrated constructs do not need to be grown in selective media, which will substantially improve their growth rate and health.

Finally, the assay signal could be dramatically improved with an alternative reporter gene. By using the *Neonothopanus nambi* luciferin synthesis pathway, it could be possible to create an autoluminescent reporter for Y2H assays with a very high signal to noise ratio.

Research significance

This preliminary Y2H assay demonstrates that *S. cerevisiae* are a plausible host for the study of HBx and DDB1. If the assay can be sufficiently improved, it may be used to aid in drug discovery for the treatment of chronic hepatitis B. Chronic hepatitis B affects hundreds of millions of people around the world, killing 820,000 every year, and yet there remains no cure.

One of the biggest bottlenecks to drug development is drug discovery: the identification of small compounds that have an effect against the desired therapeutic target. This typically requires the screening of millions of drugs, which can be a time-consuming and expensive process. However, yeast are cheap to maintain, their handling is easily automatable, and they grow relatively quickly. This makes them an excellent candidate for the high-throughput screening needed in drug discovery. As the Y2H assay improves, we will continue to learn more about the intricacies of the Y2H system. Although the yeast two-hybrid assay is old, it remains a powerful tool for the study of many protein interactions. Optimizing the Y2H assay for HTS opens the door to many more Y2H-based assays in the future.

Acknowledgments

I thank Marissa Zintel for starting this project with me and her significant contributions during its development. I thank Judith Humphries and Eric Lewellyn for their wonderful guidance through this research. I thank Brian Piasecki for his excellent feedback. I thank Synthetic Biology Club for joining me in parts of this adventure.

References

1. CDC. About Global hepatitis B. *Centers for Disease Control and Prevention* https://www.cdc.gov/globalhealth/immunization/diseases/hepatitis-b/about/index.html (2022).
2. Hepatitis B. https://www.who.int/news-room/fact-sheets/detail/hepatitis-b.
3. First recombinant DNA vaccine for HBV. https://www.nature.com/articles/d42859-020-00016-5.
4. 2020 Hepatitis B | Viral Hepatitis Surveillance Report | CDC. https://www.cdc.gov/hepatitis/statistics/2020surveillance/hepatitis-b.htm (2022).
5. Hyun Kim, B. & Ray Kim, W. Epidemiology of Hepatitis B Virus Infection in the United States. *Clin. Liver Dis.* **12**, 1 (2018).
6. Hepatitis B Foundation: Treatment Options. https://www.hepb.org/treatment-and-management/treatment/.
7. Sohn, E. The global fight against hepatitis B is benefitting some parts of the world more than others. *Nature* **603**, S58–S59 (2022).
8. Regions Dashboard – CDA Foundation. https://cdafound.org/polaris-regions-dashboard/.
9. Hepatitis B incidence rate. *Our World in Data* https://ourworldindata.org/grapher/hepatitis-b-incidence-sdgs.
10. Assembly and Release of Hepatitis B Virus. http://perspectivesinmedicine.cshlp.org/content/5/12/a021394.short.
11. Diogo Dias, J., Sarica, N. & Neuveut, C. Early Steps of Hepatitis B Life Cycle: From Capsid Nuclear Import to cccDNA Formation. *Viruses* **13**, 757 (2021).
12. Zhang, Z., Zehnder, B., Damrau, C. & Urban, S. Visualization of hepatitis B virus entry – novel tools and approaches to directly follow virus entry into hepatocytes. *FEBS Lett.* **590**,

1915–1926 (2016).

13. Bell, T. G., Kramvis, A., Bell, T. G. & Kramvis, A. *The Study of Hepatitis B Virus Using Bioinformatics*. *Bioinformatics - Updated Features and Applications* (IntechOpen, 2016). doi:10.5772/63076.

14. Kim, J. H. *et al.* Circulating serum HBsAg level is a biomarker for HBV-specific T and B cell responses in chronic hepatitis B patients. *Sci. Rep.* **10**, 1835 (2020).

15. Liang, T. J. Hepatitis B: The virus and disease. *Hepatology* **49**, S13–S21 (2009).

16. Murphy, C. M. *et al.* Hepatitis B Virus X Protein Promotes Degradation of SMC5/6 to Enhance HBV Replication. *Cell Rep.* **16**, 2846–2854 (2016).

17. Sekiba, K. *et al.* Inhibition of HBV Transcription From cccDNA With Nitazoxanide by Targeting the HBx–DDB1 Interaction. *Cell. Mol. Gastroenterol. Hepatol.* **7**, 297–312 (2019).

18. Wildey, M. J., Haunso, A., Tudor, M., Webb, M. & Connick, J. H. Chapter Five - High-Throughput Screening. in *Annual Reports in Medicinal Chemistry* (ed. Goodnow, R. A.) vol. 50 149–195 (Academic Press, 2017).

19. Savjani, K. T., Gajjar, A. K. & Savjani, J. K. Drug Solubility: Importance and Enhancement Techniques. *Int. Sch. Res. Not.* **2012**, e195727 (2012).

20. Dowden, H. & Munro, J. Trends in clinical success rates and therapeutic focus. *Nat. Rev. Drug Discov.* **18**, 495–496 (2019).

21. Goodnow, R. A. & Davie, C. P. Chapter One - DNA-Encoded Library Technology: A Brief Guide to Its Evolution and Impact on Drug Discovery. in *Annual Reports in Medicinal Chemistry* (ed. Goodnow, R. A.) vol. 50 1–15 (Academic Press, 2017).

22. Halford, B. How DNA-encoded libraries are revolutionizing drug discovery. *Chemical & Engineering News*

https://cen.acs.org/articles/95/i25/DNA-encoded-libraries-revolutionizing-drug.html.

23. Luo, Y., Peng, J. & Ma, J. Next Decade's AI-Based Drug Development Features Tight Integration of Data and Computation. *Health Data Sci.* **2022**, (2022).

24. Ghislat, G., Rahman, T. & Ballester, P. J. Recent progress on the prospective application of machine learning to structure-based virtual screening. *Curr. Opin. Chem. Biol.* **65**, 28–34 (2021).

25. Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).

26. Lin, J.-S. & Lai, E.-M. Protein–Protein Interactions: Yeast Two-Hybrid System. in *Bacterial Protein Secretion Systems: Methods and Protocols* (eds. Journet, L. & Cascales, E.) 177–187 (Springer, 2017). doi:10.1007/978-1-4939-7033-9_14.

27. 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).

28. Yeast 2-Hybrid. *Singer Instruments* https://www.singerinstruments.com/resource/yeast-2-hybrid/.

29. Hamdi, A. & Colas, P. Yeast two-hybrid methods and their applications in drug discovery. *Trends Pharmacol. Sci.* **33**, 109–118 (2012).

30. Fashena, S. J., Serebriiskii, I. & Golemis, E. A. The continued evolution of two-hybrid screening approaches in yeast: how to outwit different preys with different baits. *Gene* **250**, 1–14 (2000).

31. Li, T., Robert, E. I., van Breugel, P. C., Strubin, M. & Zheng, N. A promiscuous α-helical motif anchors viral hijackers and substrate receptors to the CUL4–DDB1 ubiquitin ligase machinery. *Nat. Struct. Mol. Biol.* **17**, 105–111 (2010).

32. Heintz, V. J., Wang, L. & LaCount, D. J. NanoLuc luciferase as a quantitative yeast two-hybrid reporter. *FEMS Yeast Res.* **21**, foab069 (2021).

33. Kotlobay, A. A. *et al.* Genetically encodable bioluminescent system from fungi. *Proc. Natl. Acad. Sci.* **115**, 12728–12732 (2018).

34. Hu, W., Sweet, T., Chamnongpol, S., Baker, K. & Coller, J. Co-translational mRNA decay in *Saccharomyces cerevisiae*. (2009).

35. Ramakrishnan, D. *et al.* Hepatitis B Virus X Protein Function Requires Zinc Binding. *J. Virol.* **93**, e00250-19 (2019).

36. James, P., Halladay, J. & Craig, E. A. Genomic Libraries and a Host Strain Designed for Highly Efficient Two-Hybrid Selection in Yeast. *Genetics* **144**, 1425–1436 (1996).

37. Smirnov, M. N., Smirnov, V. N., Budowsky, E. I., Inge-Vechtomov, S. G. & Serebrjakov, N. G. Red pigment of adenine-deficient yeast *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* **27**, 299–304 (1967).

38. Van Criekinge, W. & Beyaert, R. Yeast two-hybrid: State of the art. *Biol. Proced. Online* **2**, 1–38 (1999).

39. Lv, L. *et al.* Discovery of a molecular glue promoting CDK12-DDB1 interaction to trigger cyclin K degradation. *eLife* **9**, e59994 (2020).

Appendix

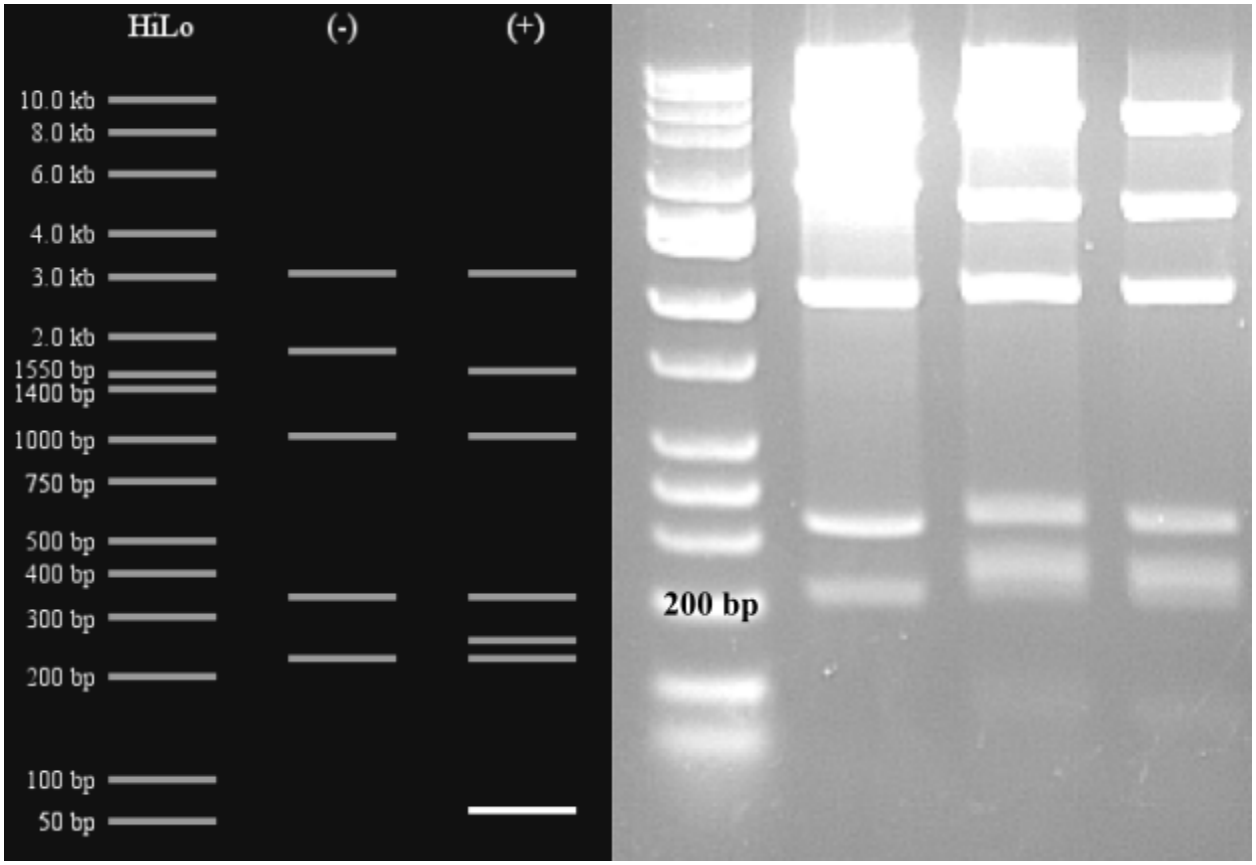

Figure 14. Informative restriction digest with BbsI on a 2% agarose gel confirming successful ligation of pGBD-3M-X. Anticipated banding patterns are displayed in a virtual digest on the left, with (-) and (+) representing unsuccessful or successful cloning, respectively. Left to right on the gel: Hi-Lo DNA marker, pGBD-X negative control, transformation isolate 1, and isolate 2. Both isolates were successful ligations, indicated by the appearance of a doublet at just over 200 bp.

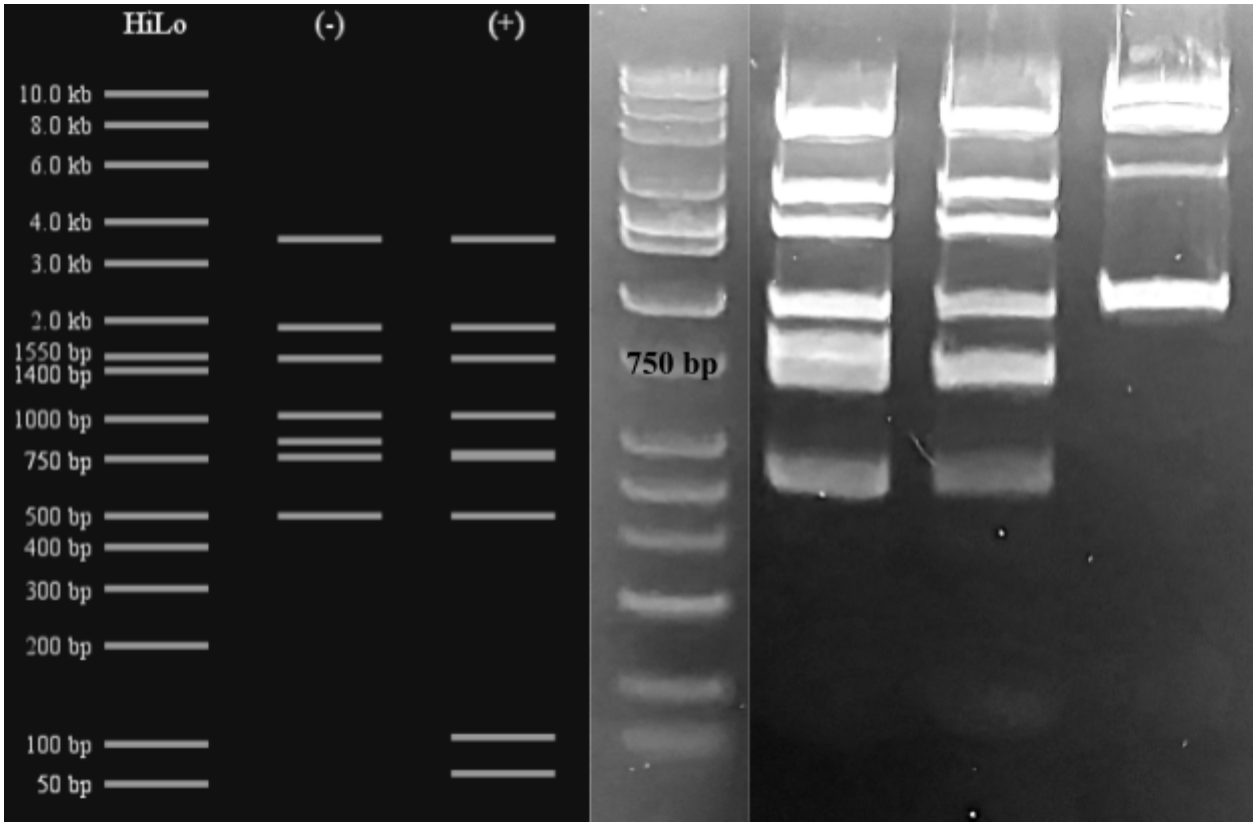

Figure 15. Informative restriction digest with BbsI on a 2% agarose gel confirming successful ligation of pGAD-3M-DDB1. Anticipated banding patterns are displayed in a virtual digest on the left, with (-) and (+) representing unsuccessful or successful cloning, respectively. Left to right on the gel: Hi-Lo DNA marker (digitally spliced from the other side of the gel), pGAD-DDB1 negative control, transformation isolate 1, and isolate 2. The banding from isolate 2 is hard to interpret, but could result from incomplete BbsI digestion. Isolate 1 was a successful ligation, indicated by the loss of the doublet at 750 bp and the gain of a very faint band at 100 bp.

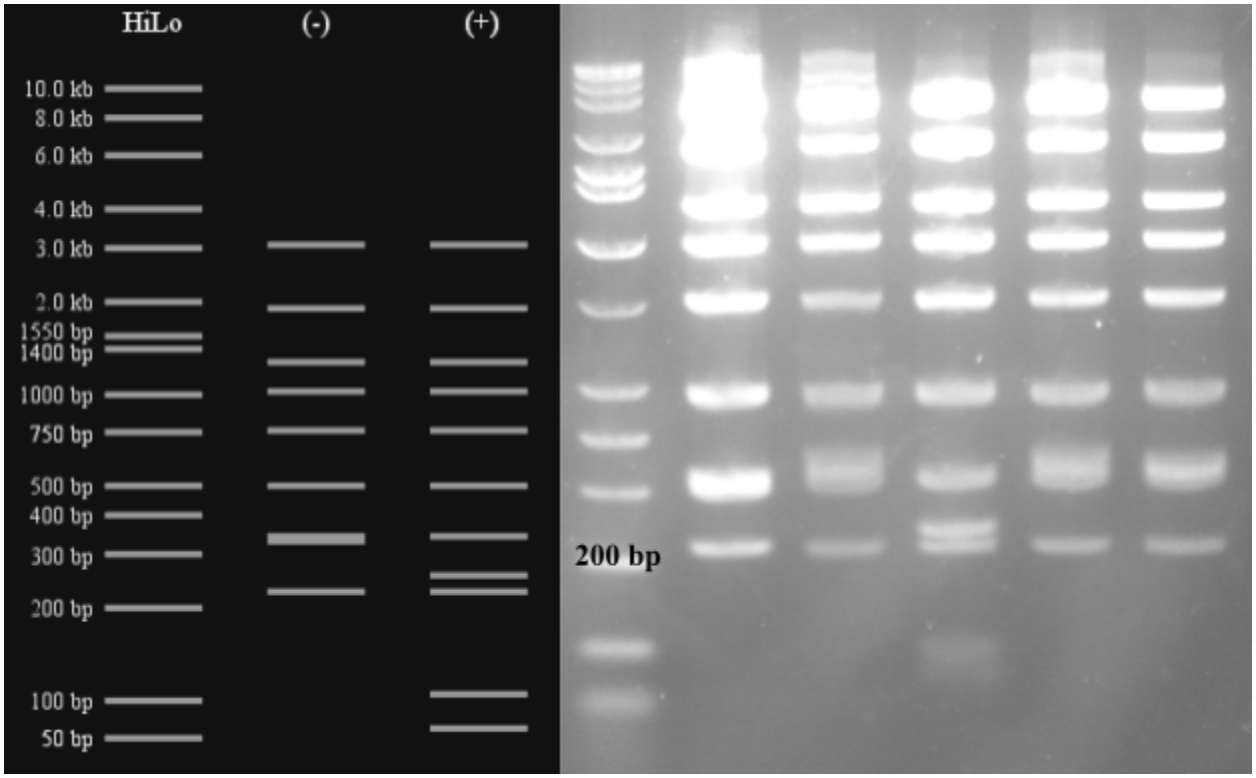

Figure 16. Informative restriction digest with BbsI on a 2% agarose gel confirming successful ligation of pGBD-3M-DDB1. Anticipated banding patterns are displayed in a virtual digest on the left, with (-) and (+) representing unsuccessful or successful cloning, respectively. Left to right on the gel: Hi-Lo DNA marker, pGBD-DDB1 negative control, transformation isolate 1, isolate 2, isolate 3, and isolate 4. Only isolate 2 was a successful ligation, indicated by the appearance of a doublet around 200 bp and a faint band at 100 bp.

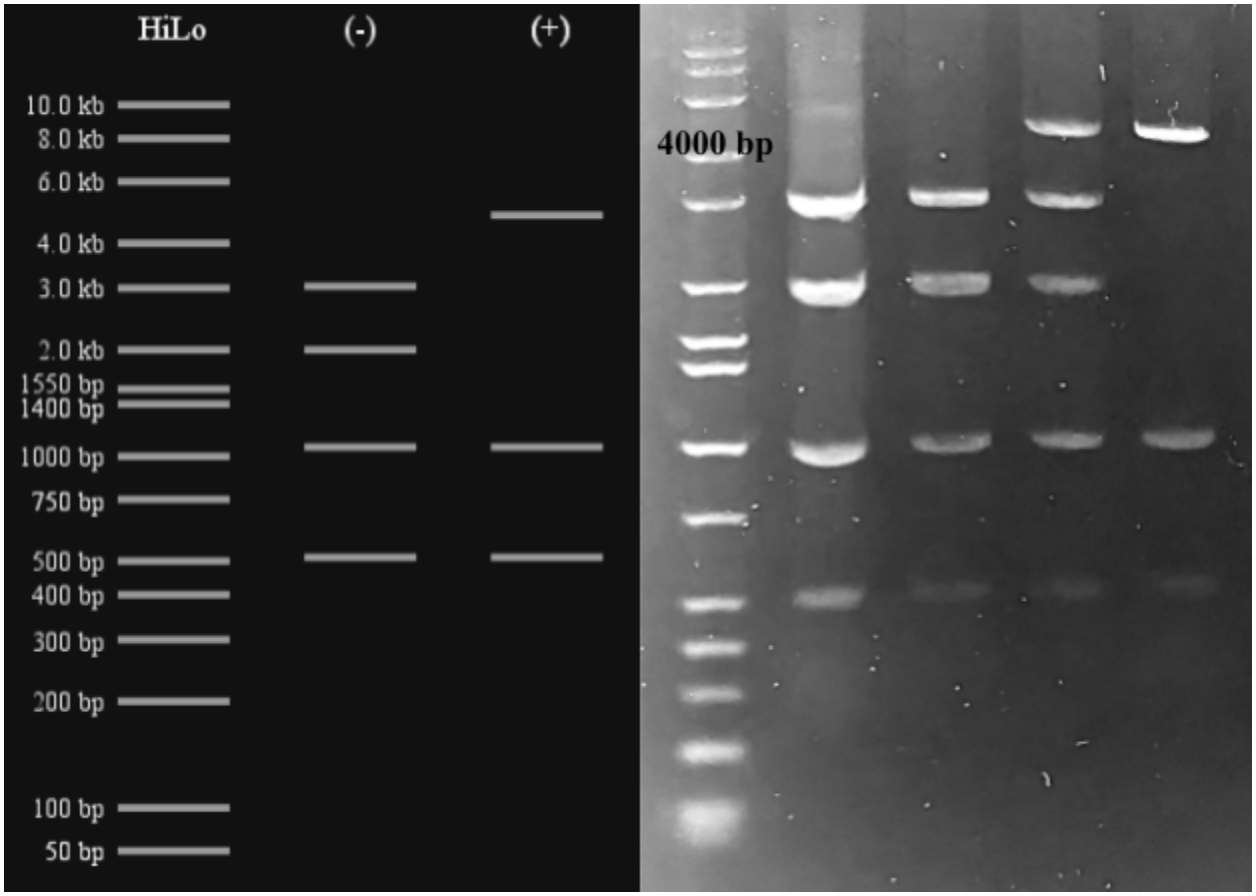

Figure 17. Informative restriction digest with BbsI on a 2% agarose gel confirming successful ligation of pGBD-3M-tX. Anticipated banding patterns are displayed in a virtual digest on the left, with (-) and (+) representing unsuccessful or successful cloning, respectively. Left to right on the gel: Hi-Lo DNA marker, pGBD-3M-X negative control, transformation isolate 1, isolate 2, and isolate 3. The banding from isolate 2 is hard to interpret, but could be explained by overlapping transformation colonies that only appeared as one colony. Only isolate 3 was a successful ligation, indicated by the appearance of a band at 5,000 bp and the disappearances of the 2,000 and 3,000 bp bands.

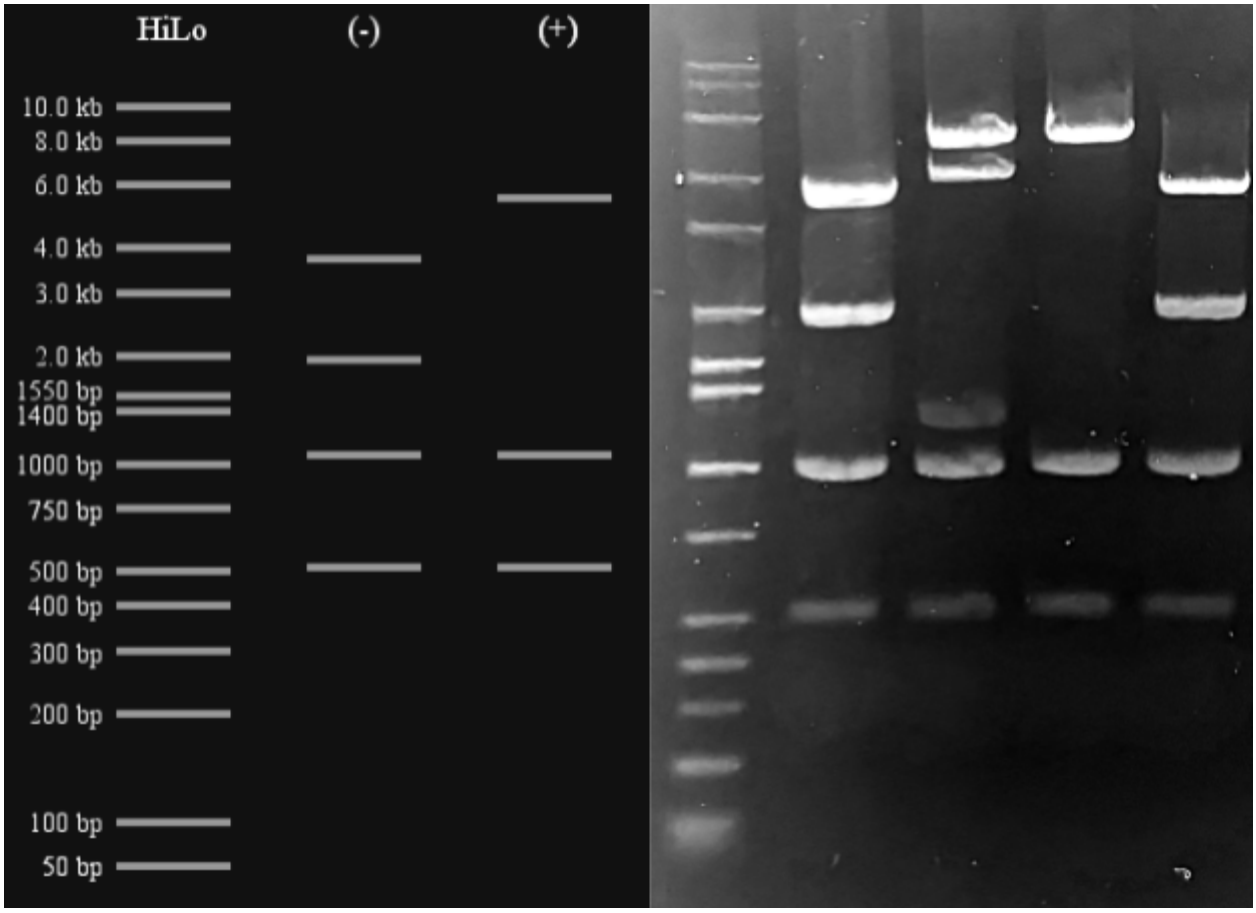

Figure 18. Informative restriction digest with BbsI on a 2% agarose gel confirming successful ligation of pGAD-3M-tX. Anticipated banding patterns are displayed in a virtual digest on the left, with (-) and (+) representing unsuccessful or successful cloning, respectively. Left to right on the gel: Hi-Lo DNA marker, pGAD-3M-X negative control, transformation isolate 1, isolate 2, and isolate 3. The banding from isolate 1 is hard to interpret, but could be explained by BbsI star activity. Only isolate 2 was a successful ligation, indicated by the appearance of a band around 6,000 bp and the disappearances of the 2,000 and 4,000 bp bands.