Designing cell and protein-based DON biosensors for use in investigating DON-biotransforming microbes

by

Brian Harrison

A thesis submitted to the Faculty of Graduate Studies and Postdoctoral Studies of
The University of Manitoba
in partial fulfillment of the requirements of the degree of

Master of Science

Department of Microbiology
University of Manitoba
Winnipeg

Abstract

Fusarium head blight (FHB) poses a major threat to small-grain cereals, reducing yield, quality, and contaminating grains with the mycotoxin deoxynivalenol (DON). Once contaminated, it is difficult and costly to remove DON from grains. New rapid and low-cost assays for detecting the presence of DON are needed in order to screen for novel DON biotransforming microbial strains. This thesis develops and evaluates two novel biosensors for detecting DON: a cell-based *Escherichia coli* biosensor with a GFP reporter and a protein-based sensor engineered from *Fusarium graminearum* acetyltransferase TRI101.

The Horizon Discovery *E. coli* Promoter Collection was screened using fluorescence-activated cell-sorting, followed by sequencing and fluorescence plate assays for DON-responsive strains. Strain *rmf* (ribosome modulation factor) exhibited a reproducible increase in fluorescence (~1.3-1.4× increase) when incubated in the presence of DON. Re-engineering this strain with the brighter GFP variant mClover3 improved peak ON/OFF fluorescence ratio to ~1.6-1.7. A medium-throughput fluorescence assay was developed and tested on 142 cultured soil microbial strains.

Molecular dynamics (MD) simulations were performed to analyze the behaviour of TRI101 in its *apo* form as well as with a DON ligand. Computational analysis of TRI101 was used to determine the best residue fluorophore conjugation sites to produce a modified TRI101 that could report the presence of DON in a solution. Amino acid residues G421 and A218 were identified as the best candidates and a TRI101-G421C was designed and expressed recombinantly in *E. coli*. When conjugated to TRI101-G421C, 7-diethylamino-3-(4′-maleimidylphenyl)-4-methylcoumarin (CPM) produced a detectable response to DON in fluorescence assay with a decrease of 8-9% in emission intensity at the emission maximum wavelength (480 nm).

Together, these results show that an *E. coli rmf* promoter-based biosensor can inexpensively detect DON in liquid samples with relatively high detection sensitivity. A TRI101-based biosensor demonstrated functionality in detecting DON however assay signal contrast and production of recombinant TRI101 remain as limiting factors. Improvements to TRI101 yield, testing A218C and other TRI101 mutants for improved performance, as well as further enhancing the sensitivity and assay signal contrast for both biosensor designs should be the focus of future research.

Acknowledgements

I would like to thank Fabian Rhoden of the Wieden lab for assistance in the set-up and analysis of molecular dynamics simulations for TRI101 (including the use of his scripts), as well as in the re-parameterization and replacement of DON within the protein-ligand structure and the proper configuring of initial simulation steps and production runs.

I would also like to thank my supervisor Dr. Mathew Bakker as well as my committee members Dr. Sabine Kuss and Dr. HJ Wieden for their constructive feedback and research guidance.

This research was supported by the EvoFunPath program, which provided an award of $18,000 towards my research, as well as awards to my supervisor through the Research Manitoba New Investigator Operating Grant and Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants programs.

Computational resources were provided by the Digital Research Alliance of Canada, with access facilitated through Dr. H. J. Wieden.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

15-ADON — 15-acetyl-deoxynivalenol
3-ADON — 3-acetyl-deoxynivalenol
3-epi-DON — 3-epideoxynivalenol
AMBER / AmberTools — Assisted Model Building with Energy Refinement (MD software suite)
BL21(DE3) — *E. coli* protein-expression strain (λDE3 lysogen)
BME — β-mercaptoethanol
CINC — Computational Identification of Non-disruptive Conjugation Sites (per Smith et al., 2022)
CPPTRAJ — AMBER trajectory analysis program
CPM — 7-diethylamino-3-(4′-maleimidylphenyl)-4-methylcoumarin
ddH$_2$O — double-distilled water
DOM-1 — deepoxy-deoxynivalenol
DON — deoxynivalenol
ELISA — enzyme-linked immunosorbent assay
F5M — fluorescein-5-maleimide
FACS — fluorescence-activated cell sorting
FDA — U.S. Food and Drug Administration
FHB — Fusarium head blight
FRET — Förster (fluorescence) resonance energy transfer
GAFF2 — General AMBER Force Field 2
GFP — green fluorescent protein
GI — gastrointestinal
H-bond — hydrogen bond
HPLC — high-performance liquid chromatography

IDT — Integrated DNA Technologies
IPTG — isopropyl β-D-1-thiogalactopyranoside
LB — Lysogeny Broth
LC-MS/MS — liquid chromatography–tandem mass spectrometry
M9 — M9 minimal medium
MD — molecular dynamics
MDCC — N-(7-dimethylamino-4-methyl-3-coumarinyl)maleimide
MOS — maltooligosaccharides
NCBI — National Center for Biotechnology Information
NetCDF — Network Common Data Form (trajectory format)
NMR — nuclear magnetic resonance
NTA — nitrilotriacetic acid (as in Ni–NTA resin)
OD600 — optical density at 600 nm
ompF — outer membrane porin F
ORF — open reading frame
PBS — phosphate-buffered saline
PDA — potato dextrose agar
PDB — Protein Data Bank
PCR — polymerase chain reaction
PMSF — phenylmethylsulfonyl fluoride
PMEMD — Particle-Mesh Ewald Molecular Dynamics (AMBER engine)
PPII — polyproline II (helix)
(p)ppGpp — guanosine tetra-/pentaphosphate
PTC — peptidyl transferase centre
PQQ — pyrroloquinoline quinone
Q30 — Phred quality score ≥ 30
RFU — relative fluorescence units
RMSD — root mean square deviation
RMSF — root mean square fluctuation
rRNA — ribosomal RNA
SASA — solvent-accessible surface area
SDS-PAGE — sodium dodecyl sulfate–polyacrylamide gel electrophoresis
TIP3P — Three-point transferable intermolecular potential water model
TNTC — too numerous to count
TSA — tryptic soy agar
VMD — Visual Molecular Dynamics

Chapter 1 - General Introduction

1.1 Fusarium diseases in agricultural crops and the burden of DON

Fusarium graminearum is a well-characterized fungal species that causes Fusarium head blight (FHB), a major plant disease affecting cereal grains like wheat, oats, barley, and rye.[1] FHB causes reduced yields and diminished grain quality while also contaminating grains with the mycotoxin deoxynivalenol (DON). In corn, infection by *F. graminearum* causes ear rot, root rot, stalk rot, and Fusarium seedling blight. In addition to *F. graminearum*, DON is also produced by *Fusarium culmorum*, which also infects grains although it is less prevalent in North America. DON is a major virulence factor in FHB and aids in the spread of the disease within the plant.[2] Without mycotoxin production, disease severity is substantially reduced.[3] DON is produced by specific *F. graminearum* chemotypes in two major acetylated forms, 3-ADON and 15-ADON, with 15-ADON being more common in North American strains.[3,4] These chemotypic differences contribute to variation in plant toxin load and the severity of FHB outbreaks across different regions, which shapes the economic and geographical impact of FHB each season.[5]

In a review of the 2022 Canadian growing season, statistics identify FHB as the worst economic disease threat for the prairie provinces.[6] The severity and incidence of FHB in 2022 was unusually high, due to unseasonably heavy precipitation and high heat, particularly during the critical flowering stage of wheat growth as well as during the grain filling period. In particular, Manitoba spring wheat reached incidence rates between 60-100% with durum wheat reaching 68% incidence of FHB in Saskatchewan. On average, Manitoba red spring wheat exhibits the highest DON concentrations of the prairie provinces, averaging between 0.3-1.0 mg DON per kilogram of harvested wheat.[7]

Current strategies for managing FHB and *Fusarium* corn diseases include using resistant cultivars, rotating crops, and applying fungicides with varying efficacy. Despite this, FHB has established itself as a major and worsening economic threat with total losses estimated to be up to 300 million CAD since the early 1990's.[8] In the US, the cumulative loss for US wheat from FHB between the years 1993 to 2014 is estimated at approximately $17 billion USD.[9] As detailed by the Manitoba Crop Alliance, there are three major ways in which FHB causes economic loss: downgrading of grains due to reduced quality, yield loss from kernel damage as well as reduction of harvestable grain, and contamination with mycotoxins such as DON.[10]

When tested levels exceed what is acceptable for human consumption (the FDA sets a limit of 1000µg per kg of raw grain while the Canadian Food Inspection Agency sets a limit of 2000µg/kg)[11] DON-contaminated grain is often fed to livestock as a source of cheap feed.[12] Various impairments result from this chronic exposure to DON, including intestinal damage, impaired immune function, and liver and kidney impairment. In dairy cows, DON exposure reduces milk output and quality. DON contamination has also been detected in milk and eggs, revealing both the high DON concentrations to which livestock are exposed as well as the toxin's resilience to degradation in the intestinal tract, rumen, and animal circulatory systems.[13] Mycotoxin exposure (including DON) leads to slowed growth rates of livestock and can adversely affect the quality of animal products. In humans, DON exposure is typically lower and primarily causes nausea, vomiting and GI disturbances, although its contribution to intestinal damage and disruption to natural gut flora are also of concern.[14,15]

1.2 Chemical and structural characteristics of DON

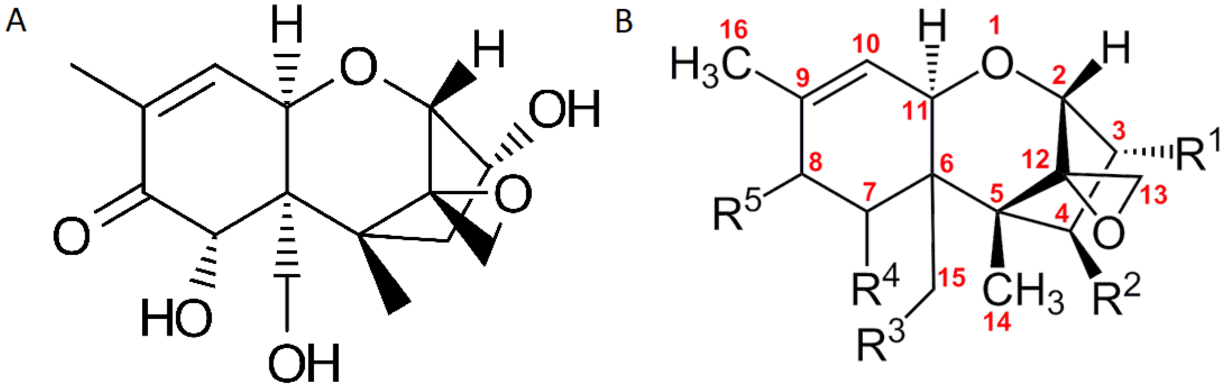


Figure 1. Chemical structure of deoxynivalenol (A) and a generic trichothecene core structure (B) with numbered carbons. 'R' groups are variable among trichothecene variants.

Deoxynivalenol is a trichothecene mycotoxin which belongs to the sesquiterpenoid compound family and is originally derived from three isoprene units.[16] Trichothecenes are further characterized by type from A-D, with DON characterized as type B, sharing this category with nivalenol, as well as 3-ADON and 15-ADON. Type B trichothecenes possess a carbonyl group at C-8 instead of the ester functional group found in type A.[17] T-2 toxin (a type A trichothecene), which has risen in prominence in recent years, likely as a result of rising global temperatures, is more acutely toxic than DON.[18] This is possibly due to the addition of three ester groups (two being alkyl esters) at side groups R^2, R^3, and R^5, increasing the toxin's lipophilicity, allowing it to more easily cross cytoplasmic and mitochondrial membranes.[19,20] However, DON's relative hydrophilicity provides advantages in improved transport through the apoplast and vascular tissues (xylem/phloem) of plants during infection, likely contributing to the increased prevalence of DON and nivalenol in crop FHB over T-2 toxin.[21] DON's moderate polarity and multiple hydroxyl groups are characteristic of small molecules that can participate in non-covalent interactions with biological macromolecules, including proteins and polysaccharide-rich components of the grain endosperm.[22–24] Such interactions may contribute to the association of DON with grain tissues and reduce its susceptibility to simple physical removal methods, such as washing.[25] These chemical features help explain the persistence of DON contamination in agricultural crops consistent with its ability to penetrate plant tissues and translocate systemically.

1.3 DON mechanism of ribotoxicity and cellular impact

The mechanism of toxicity by trichothecene mycotoxins to eukaryotic cells has been well characterized since 1974 by Cundliffe et al.[26] The mode of binding has been described by a structural study capturing DON bound to the 80S eukaryotic ribosome in *S. cerevisiae*.[27] It was found that DON forms multiple interactions with the 25S rRNA (part of the 60S large subunit), including hydrogen bonds and non-polar π-stacking interactions involving the epoxide ring (as formed by C13 in Figure 1) as well as the alkene bond between C9 and C10.[21] For this reason, the epoxide group is regarded as giving DON most of its ribotoxic activity. Through this interaction, DON specifically blocks the peptidyl transferase centre (PTC) of the ribosome, preventing proper positioning of tRNA within the pocket and therefore stalls peptide bond formation. This, in turn, triggers a “ribotoxic stress response”, activating MAP kinase pathways which downstream leads to inflammation, cytokine induction and potentially apoptosis.[16,28] The resulting ribotoxic stress-activated signalling cascade is also responsible for impairing immune function in plant cells and therefore highlights DON’s role in plant pathogenesis.[29] Beyond the immediate stress response, blocking translation halts the replenishment of essential enzymes as well as structural and regulatory proteins required for cell maintenance and growth. Since DON primarily targets the eukaryotic 80S peptidyl-transferase center and prokaryotes have structurally distinct 70S ribosomes, prokaryotic cells are less susceptible. However, exposure in prokaryotes still triggers altered gene expression and stress pathways.[30]

1.4 Challenges in current DON decontamination methods

Once DON has contaminated grains, it becomes especially difficult to remove due to its high chemical stability and resistance to many different types of treatments.[31,32] The C12,13 epoxide and the conjugated double bonds in the C9–C10 region are structural features associated with increased chemical stability and resistance to degradation by hydrolytic and oxidative processes.[22] High temperature processes and treatments up to 350°C have been shown to have a minimal effect on DON levels. Acidic conditions also have limited effect while an alkaline environment can speed up breakdown when combined with high temperatures.[33] While these physical treatments can often be done quickly, they also risk significantly degrading grain quality. DON becomes tightly embedded within the grain endosperm, making washing steps less effective at removing it. A combination of infrared, microwave drying, and pressure washing was found to reduce DON concentrations by as much as 89.0%.[34] Chemical treatments such as sodium carbonate, sodium hydroxide, and chlorine have modest efficacy. Given these limitations, attention has been increasingly shifting towards more biologically tailored detoxification strategies in which DON can be degraded selectively under milder conditions.[35]

1.5 Microbial and enzyme-mediated biotransformation of deoxynivalenol as a promising approach to mycotoxin detoxification of grains

Microbes and microbial enzymes, both natural and modified, have also been explored as a promising way to detoxify DON-contaminated grains.[35] These methods offer the potential for more sustainable decontamination of grains compared to chemical treatments.[36] A bacterial strain of *Eubacterium* sp., a gram-positive obligate anaerobe, was isolated from bovine rumen and found to convert DON into the metabolic derivative DOM-1 or deepoxy-deoxynivalenol by removing the C12,13 epoxide group[37], which as previously stated is critical for toxicity.[38] This bacterium was approved as a feed additive for pigs and ruminants but has encountered difficulty in significantly ameliorating weight loss in pigs given DON contaminated feed in trials.[39] Better efficacy was seen in simulated rumen (anaerobic conditions) against zearalenone, a non-trichothecene *Fusarium* mycotoxin.[40] The enzyme(s) mainly responsible for this detoxification activity were never identified. Similarly, a strain of *Devosia mutans,* a gram-negative obligate aerobe, was found to contain two enzymes, DepA and DepB, which together convert DON to 3-epi-DON, a stereoisomer at the C3 hydroxyl group.[41] This is a somewhat surprising detoxification mechanism that does not involve modification to the epoxide group, and the precise structural basis for why this modified form of DON is less toxic is not fully understood.[42] While both enzymes were found to have functional detoxification activity when expressed and isolated recombinantly, the requirement of both enzymes for full detoxification is a limitation that complicates their commercial use and effectiveness (DepA is also PQQ dependent which further limits effectiveness *in situ*). Almost complete conversion of DON to 3-epi-DON was found in assays (when supplied with PQQ) but its effectiveness against contaminated grains is not established. These studies illustrate both the promise and limitations of applying microbial-based methods towards practical detoxification of DON-contaminated grains. Notably, in these examples the detoxification is demonstrated either in animal gut systems or in vitro, rather than as a direct treatment of contaminated grain.

More recently, a 2024 study modified the glutathione-S-transferase enzyme Fhb7 (from *Thinopyrum elongatum*, a species of wheatgrass) to become even more effective at its established activity opening the DON epoxide ring.[43] While single amino acid substitutions were explored, computational protein design allowed for the creation of a mutant with 10 total amino acid substitutions in which enzyme stability was found to be increased by 266-fold without substantial loss of detoxifying activity. It's important to note that Fhb7 is solely being explored as a means for reducing DON contamination and FHB severity in transgenic plants rather than being applied to grains which are already DON-contaminated. Engineering crops such as wheat with a cloned DON detoxifying enzyme is often viewed as an ideal solution which prevents contamination and severe disease simultaneously. However, not all enzymes would be compatible with plant expression (particularly those with bacterial origin) and even less would be expressed at levels necessary to significantly reduce disease all the while maintaining favourable yield and quality.[44]

1.6 Challenges in the discovery and screening of DON-detoxifying microbes and enzymes

While these discoveries and engineered variants have been important steps towards finding suitable microbes and enzymes for treating contaminated grains (or in preventing contamination and severe disease), the discovery of new DON-detoxifying microbes and enzymes is an extremely expensive, slow and painstaking process. Often the screening of newly isolated microbes for activity against DON requires specialized laboratory methods and equipment such as high-performance liquid chromatography (HPLC), enzyme-linked immunosorbent assays (ELISA), and nuclear magnetic resonance (NMR) spectroscopy, all of which require trained and experienced technicians/researchers.[45,46] HPLC requires extracts to be prepared one at a time (constraining throughput) and filtered, along with precise calibration of the instrument for adequate separation of target analyte peaks. Sample preparation can be costly and certain compounds which escape sample cleanup may co-elute with DON and create interfering peaks. While loss of the DON parent peak can indicate biotransformation has occurred, identifying the resulting product relies on specific reference standards (which may not be readily available). If the transformation product cannot be identified, the fraction can be purified and further analysis using NMR spectroscopy can be pursued to precisely identify the compound, assuming sufficient quantity and purity. Until the modified DON product is identified, small modifications which leave DON's toxicity largely intact can not be readily ruled out. ELISA kits for detecting DON are commercially available, but both sensitivity and specificity can be limited by cross-reactivity and matrix effects. The cost, limits in throughput and technical expertise and in using these tools creates a bottleneck that can greatly slow the screening and discovery of novel DON biotransforming microbes and enzymes. To improve screening scale and save cost and time, simpler DON detection methods and especially those amenable to rapid high-throughput screening of environmentally isolated microbes are greatly needed to improve discovery rates and enhance the overall arsenal of known microbes and enzymes with DON detoxifying activity.

1.7 Principles and design of biosensors for the detection of small molecules

A biosensor is constructed to detect and report the presence of a particular compound within a given sample.[47,48] Rational design separates a biosensor into four main parts. The first and most critical is the recognition element, which binds or "senses" the target compound in some distinguishable way. Recognition elements may be formed from nucleic acids (such as aptamers), antibodies, enzymes, or whole cells (i.e., specific recognition element uncharacterized, or signal amplified through intact cellular signal transduction systems).[47,48] The second component of a biosensor, sometimes referred to as the "transducer", is responsible for transforming the compound's recognition into a measurable signal. This signal may be electrochemical, piezoelectric, optical (such as fluorescence), or a heat signal. Among these, optical signals are perhaps the easiest to measure. An important function of the transducer is that it represents the concentration of the target compound accurately in a measurable way (quantitative or semi-quantitative reporting is more valuable than simple qualitative reporting of presence/absence). The last two components of an effective biosensor, the signal amplification element and the display/readout system, are responsible for amplifying the signal and displaying the output data respectively. Amplification of the reporter's signal is optional and depends on the strength and ease of measurement of the signal being initially generated.

A cellular biosensor has the advantage of being inexpensive to produce, as all required components could reproduce biologically, provided that the proper nutritional requirements and growth environment are met. Typically, in cellular biosensors the recognition element and the way the signal is generated are complex and not always fully characterized, since they rely on the cells' intricate cellular machinery and pathways. Signal reporting is most commonly achieved through the cellular production of a detectable molecule. This detection can commonly be achieved through enzymatic means, in which the cell expresses an enzyme that can convert an added substrate into a coloured, luminescent, or fluorescent product, or through optical detection of a directly expressed fluorescent molecule.[49] Many forms of detection are possible, depending on the specific reporter compound being produced. Variability in the precision and reproducibility of the cell biosensor's function becomes challenging when adapting the sensor to variations in media and substrate, which can alter the production of the reporter as well as potentially mask its signal. Reporting of the target molecule can be indirect and rely on cell replication and growth, introducing further variability.[50]

A model example of a cell-based biosensor with real world applicability is an *Escherichia coli* strain engineered to detect arsenite and arsenate in water, as designed by Stocker et al. (2003).[51] In this biosensor, detection was achieved by fusing the *ars* operon to a reporter "gene" (in this case the *lux*CDABE operon) which produces bioluminescence). Through this design, *E. coli* could function as a quantitative reporter of the presence of arsenic in water samples at a concentration as low as 4μg/L (approximately 50nM).

1.8 Thesis objectives and overview of experimental approaches

The main goal of this thesis was to develop and evaluate two biosensor designs for the rapid and/or low-cost detection of DON, with the long-term aim of facilitating the screening and discovery of DON-biotransforming microbes. Given the current limitations of existing analytical techniques in terms of the cost, throughput, and necessary expertise, this work explored two separate biosensor design strategies that operate at different biological scales and rely on distinct mechanisms of signal generation. This thesis tests the hypothesis that deoxynivalenol (DON) can be detected using low-cost, fluorescence-based biosensors by exploiting either DON-responsive endogenous transcriptional activity in *E. coli* or direct ligand-binding interactions with the trichothecene acetyltransferase TRI101.

The first objective of this thesis was to design and optimize a cell-based biosensor for DON detection using *E. coli* as a whole-cell sensing platform. This approach exploits native *E. coli* promoter activities that respond to DON exposure and further couples these transcriptional responses to the production of a fluorescent reporter protein. By screening a pooled library of promoter-based *E. coli* strains we identified a promoter region which is capable of producing a detectable fluorescent signal in response to DON. The strain was then optimized to improve fluorescence signal contrast and used to screen for DON-biotransforming microbes.

The second objective was to develop a protein-based biosensor for DON using the trichothecene 3-O-acetyltransferase TRI101 from *F. graminearum* as a molecular recognition element. DON detection was explored by utilizing ligand-induced structural changes in TRI101 which alter the local environment of a covalently bound fluorophore, altering fluorescence intensity. Molecular dynamics simulations were used in order to rationally select optimal fluorophore conjugation sites. A TRI101 variant was then expressed recombinantly, labeled with fluorescent probes, and evaluated for its ability to report DON binding in solution.

Chapter 2 – Design of an *E. coli* based biosensor for the detection of DON

2.1 Introduction

Designing a cell-based biosensor for detecting DON is not as straightforward as for a compound like arsenic, in which a defined, well-studied operon exists that can be genetically manipulated to re-route the signal transduction toward production of a measurable output, rather than toward mounting a detoxification/efflux response. For DON biosensing, a suitable genetic counterpart needs to be found in *E. coli* which can be exploited to report the presence of DON in growth media. While the literature on *E. coli*'s interactions and response to DON are limited, screening for DON-responsive genetic elements can be achieved using a collection of *E. coli* strains in which the production of a fluorescent molecule is controlled by varying endogenous genetic elements. The Horizon Discovery *E. coli* Promoter Collection, designed by Zaslaver et al. (2004), offers a commercially available collection of 1,819 reporter strains of *E. coli* in which endogenous promoter "regions" (intergenic sequences located between adjacent open reading frames or ORFs) have been utilized to modify expression levels of GFP in a low-copy plasmid.[52] This effectively allows expression of GFP to mimic expression levels of the gene from which the promoter region was derived from in each unique strain. As such, each strain is named by the gene from which the promoter region was derived. These intergenic sequences were extended 50-100bp into each flanking ORF to ensure all or most transcription factor binding sites for the respective gene were captured within the sequence. By screening this collection, it may be possible to identify fortuitous, innate ability of *E. coli* to detect the presence of DON.

Screening a library of strains at once can be achieved by adapting the protocol of Mahr, R. & Frunzke (2016), in which all strains from the collection were pooled together and screened for promoters responsive to L-phenylanine by using fluorescence activated cell-sorting (FACS) with alternating rounds of positive and negative selection.[53] A similar method can be used to screen for promoters which are instead responsive to DON, and therefore produce higher levels of GFP in the presence of DON, with low levels of GFP in the absence of DON. The final population of filtered cells can then be subjected to amplicon sequencing to determine which strains and promoter regions were enriched through the cell sorting. Identified strains can then be screened individually using fluorescence plate reader assays to determine individual potential as a DON biosensor.

2.2 Materials and Methods

2.2.1 Preparing *E. coli* Promoter Collection and FACS

The entire *E. coli* promoter collection (stored in 96-well plates) was pooled together using a multichannel pipette into 20ml of LB with kanamycin (25µg/ml) in liquid media. The culture was incubated at 37°C for 2 hours on a tube rotator before measuring the OD600 with a spectrophotometer. Cells were transferred into 1ml of fresh LB with kanamycin and DON (500µg/ml) to a final OD600 of 0.01 and incubated for 2.5 hours to ensure adequate cell numbers for sorting, without overgrowing the culture. After incubation, the culture was washed and diluted in 1×PBS to a final OD600 of 0.001. An unsorted control cell culture was set aside for later analysis. Flow cytometry and cell-sorting were performed using a SONY SH800 cell sorter, selecting the 488nm laser with an FL1 filter (FITC channel, 525nm) and a 70µm nozzle. Cells were initially gated based on forward scatter area (FSC-A) and side scatter area (SSC-A) to exclude debris and large aggregates prior to fluorescence-based analysis and sorting. Before cell-sorting, fluorescence readings were taken from 100,000 cells in order to set the appropriate gate for sorting. For the first round of cell-sorting, singlet gating was chosen with a set event rate of 1,000/s and only high fluorescing cells were selected (the top 10% of all fluorescing cells). Sorting was stopped once a total of 500,000 cells had been collected. These cells were resuspended in LB with kanamycin (without DON) and incubated for 2.5 hours. After resuspending cells in 1× PBS, a second round of cell-sorting was performed. This time, selection collected only low fluorescing cells, defined as cells with readings below 1000 relative fluorescence units (RFU). This threshold for low fluorescing cells was based on a previously measured negative control (a promoterless strain from the same collection). Once again, sorting was stopped once 500,000 cells were collected. This process of alternating selection for high fluorescence in the presence of DON and low fluorescence in the absence of DON was repeated until a total of 4 rounds of cell-sorting were performed (see Figure 2). Once completed, the cells were resuspended in 5ml of LB with kanamycin and incubated for 4 hours. Plasmid DNA was extracted from this culture as well as from the unsorted control culture (retained earlier) using the Monarch Plasmid Miniprep Kit, and stored at -20°C.

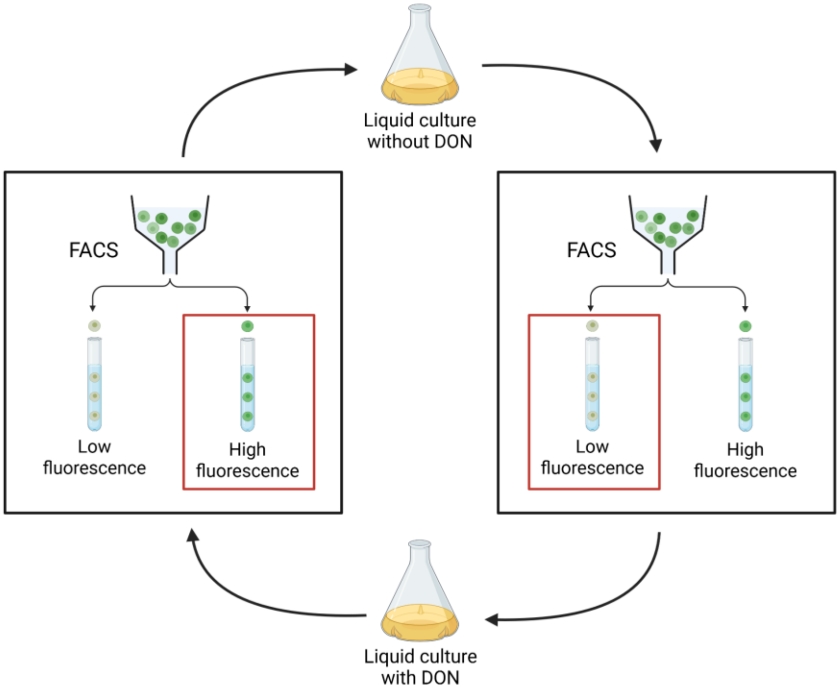


Figure 2. Overview of the FACS screening process for DON-responsive *E. coli* strains from the *E. coli* Promoter Collection. A total of four alternating rounds of positive and negative selection were used to select for DON-responsive strains.

2.2.2 Amplicon sequencing of *E coli* cell-sorted populations

PCR was used to amplify the promoter region from the extracted plasmid DNA using primers flanking this region (see Figure A1). Index sequences were added using the Nextera XT Index V2 kit (set B; Illumina). The library was quantified using the KAPA Library Quantification kit (Roche). Different sets of index sequences were used for the cell-sorted and unsorted control DNA. Sequencing was performed using the iSeq 100 System with a loading DNA volume of 20µl, at a concentration of 50pM, with a PhiX spike-in of 10%. Both sets of reads (i.e., from sorted and unsorted populations) were analyzed by mapping them to the *E. coli* MG1655 genome (NCBI accession: NC_000913.3) using the Geneious Prime "map to reference" tool. A total of 909,305 forward and 420,655 reverse reads were mapped. Gene names that each set of reads belonged to were identified manually and cross-referenced to the strain list provided with the *E. coli* Promoter Collection.

2.2.3 Testing identified *E. coli* strains for DON-responsiveness

The number of *E. coli* strains identified in the final cell-sorted population was quite low, at 44 strains (of 1,819 strains in the initial library). Thus, all 44 strains were individually tested using a fluorescence plate assay for responsiveness to DON. Each strain was individually cultured in 5ml of LB with kanamycin for 4 hours. In a 96-well black microplate with an optical bottom (Costar), 200µl of M9 minimal media with 0.4% succinate and kanamycin (25µg/ml) with vs. without DON (100µg/ml) were inoculated with 5µl of starter culture. Fluorescence intensity of each well was measured over 24 hours at 37°C with shaking, using a Synergy H1 Multimode Reader. Continuous orbital shaking (200rpm) was used with readings taken every 5 minutes. Excitation and emission wavelengths were set to 479 and 520nm, respectively. *E. coli* strain *lacZ* from the same collection was tested using the same assay conditions except that IPTG (1mM) was used to induce expression of GFP, instead of DON.

2.2.4 Design of mClover3 plasmids for candidate single-strain DON biosensors

Sanger sequencing was used to confirm the sequence of the *rmf* promoter region used to control expression of gfp-mut2 (the GFP variant used in the *E. coli p*romoter collection plasmid) in the *rmf* strain from the *E. coli* collection. A new *rmf* plasmid was designed, in which gfp-mut2 was replaced with the brighter GFP variant mClover3 using the pNCS-mClover3 as the cloning vector.[54] A plasmid was also designed in which the upstream intergenic region of the *E. coli* gene outer membrane porin F (*ompF*) was used to control expression of mClover3, due to a prior report in which transcription of *ompF* was shown to be upregulated by a factor of between 2.10 and 4.62 in the presence of DON, with moderate basal expression levels.[30] Oligonucleotides corresponding to each chosen promoter region were synthesized (IDT), with total lengths of 410 nucleotides for the *rmf* promoter region and 804 nucleotides for the *ompF* promoter region. Primers with PciI and NdeI restriction sites were used to amplify each oligo and add these restriction sites to each end using PCR (see Figure A2). The plasmid pNCS-mClover3 was purchased (AddGene), and contained the PciI and NdeI restriction sites required to place each synthesized promoter region directly upstream of the mClover3 ORF, replacing the original T7 promoter (all primers used in the plasmid assemblies are shown in Table B1). A double restriction digest with PciI and NdeI enzymes (NEB) was used to cut each amplified promoter region as well as the pNCS-mClover3 plasmid. T4 DNA ligase (NEB) was used to ligate each promoter region to the plasmid fragments, creating two distinct plasmids, *rmf-mClover3* (assembly shown in Figure A3) and *ompF-mClover3* (assembly as in Figure A3, but not shown).

The ligated plasmids were then transformed into competent DH5α *E. coli* cells using heat shock transformation, plated onto LB agar plates with ampicillin (100µg/ml) and incubated at 37°C until colonies formed (~48 hours). Both new strains (referred to as *rmf-mClover3* and *ompF-mClover3*, respectively) were confirmed with Sanger sequencing (data not shown) before being tested in fluorescence plate assays.

2.2.5 Screening soil microbes for activity using *mClover-rmf*

Strain *rmf-mClover3* was further assessed to determine whether DON detection worked with solid media as it did in broth. Plates were prepared with a thin layer (8 mL in a 9 cm Petri dish) of solid M9 minimal media + 0.4% succinate and ampicillin (100µg/ml). After the media had cooled to around 45°C but before it had solidified, 100µl of *rmf-mClover3* preculture (incubated for 2 hours in 5ml of liquid M9 minimal media) was added to the media and gently mixed before plates were poured. DON was also added to the molten agar at final concentrations of 0, 2, 5, 10, or 30µg/ml. Plates were incubated at 37°C and were analyzed for fluorescence (479nm excitation and 520nm emission) at 4, 8, 16 and 24 hours of incubation (EVOS FL Auto 2 Imaging System, ThermoFisher Scientific). Similar plates were also prepared substituting either gellan gum (0.4%) or agarose (1%) in place of agar to test whether these alternatives improved optical clarity for fluorescence detection.

Soil microbes were cultured on potato dextrose agar (PDA) and tryptic soy agar (TSA) containing DON (30µg/ml), to assess any DON biotransformation activity as revealed by modified fluorescence of the biosensor. The optimal soil dilution factor and incubation time required to obtain plates with numerous (optimally 20-40), well-isolated colonies were determined empirically. Once these conditions were established, new plates were prepared with less media (~8ml per plate to conserve DON usage) containing 10µg/ml of DON. One gram of soil was mixed into ddH_2O to a final volume of 10ml, vortexed for 5-10 minutes until well suspended, and serially diluted. 100µl of 10^{-4} soil dilution was spread onto the TSA+DON plates while the 10^{-2} soil dilution was used for the PDA+DON plates. Plates were then incubated at 28°C until enough well-isolated colonies formed (1-2 days for TSA and 3-5 days for PDA plates).

After colonies formed on the PDA and TSA plates, ~50µl 'plugs' of agar were taken from directly under each colony using P1000 micropipette tips with the ends cut (to make the tip wider). These plugs were then placed into microcentrifuge tubes and incubated at 95°C until the agar had melted (about 10 minutes). Once melted, 30µl of the liquid agar was added to previously prepared 96-well plates (black, with optically clear bottom; Costar) filled with 160µl of M9 minimal media + 0.4% succinate, 0.2% casamino acids and ampicillin (100µg/ml). Wells were then inoculated with 10µl of *rmf-mClover3* preculture (M9 + ampicillin) and analyzed over 24 hours by the Synergy H1 Microplate reader using the same conditions previously specified. Results were compared to a DON control to look for a decrease in produced fluorescence by wells containing samples of the agar plugs.

2.3 Results

2.3.1 Screening *E. coli* Promoter Collection with FACS and identifying potential DON reporting strains

FACS testing with the *lacZ* strain showed a significant increase in mean cellular fluorescence when incubated with IPTG (see Figure 3.) The U66 promoterless strain demonstrated a slightly left skewed distribution in fluorescence levels with >90% of all cells exhibiting between 100 and 1000 RFUs. As seen in Figure 4, positive selection rounds for FACS (rounds 1 and 3) selected roughly the top 10% of fluorescing cells (a threshold chosen to provide moderate selection stringency while avoiding excessive loss of diversity) while negative selection rounds (rounds 2 and 4) selected only cells exhibiting less than 1000 RFUs. The distribution of cells exhibiting less than 1000 RFUs increased from 11.26% in round 2 to 36.25% in round 4.

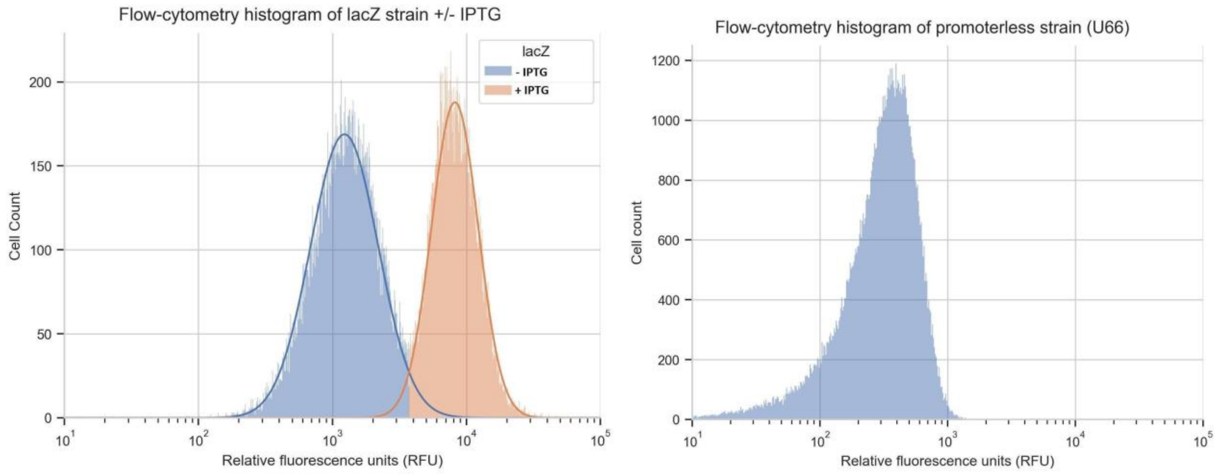


Figure 3. Flow cytometry histograms for *lacZ* strain (left) and U66 promoterless strain (right).

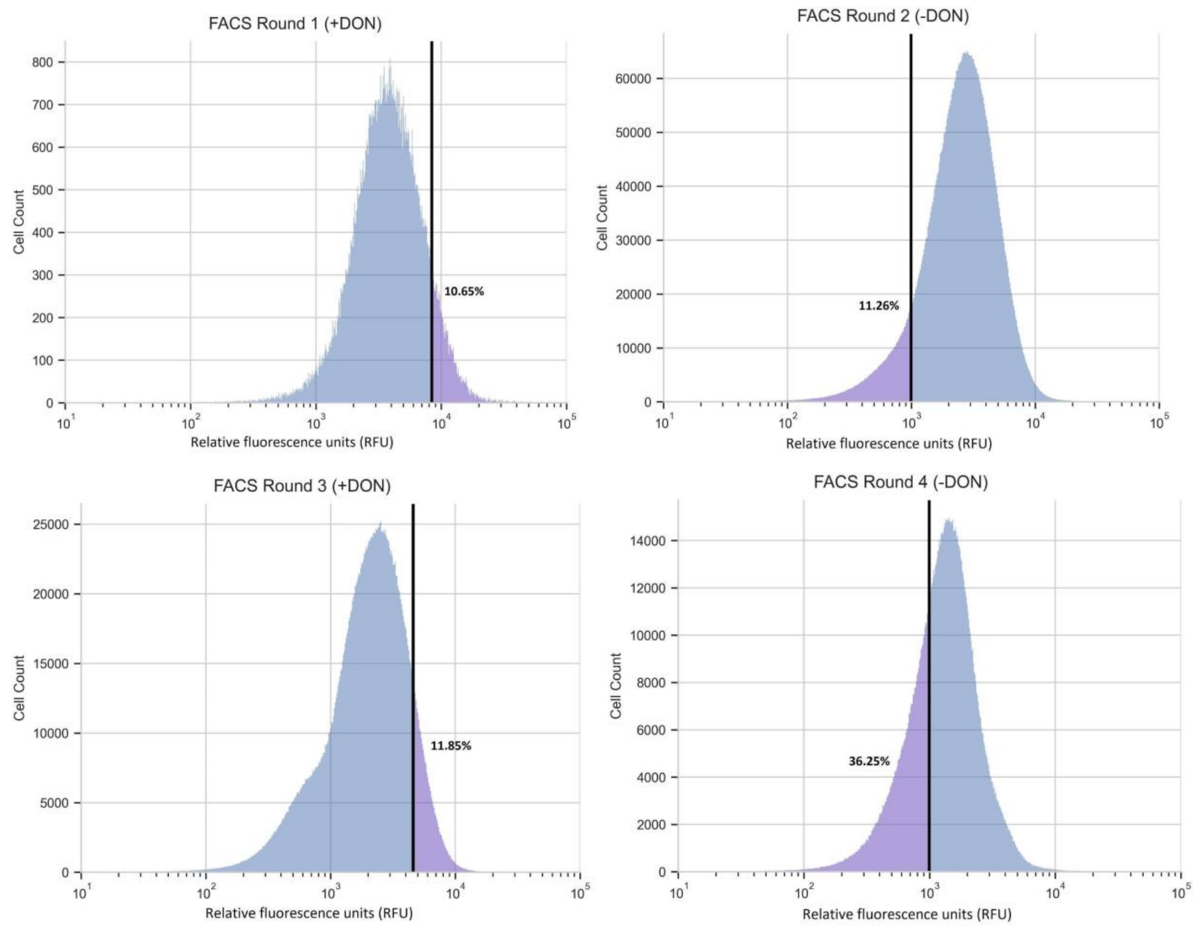


Figure 4. Final FACS histograms and gate thresholds across four sorting rounds for pooled *E. coli* promoter collection (4 rounds).

For the iSeq 100 sequencing run of the sorted and unsorted *E. coli* Promoter Collection populations, the passing filter rate was 74% (indicating that the majority of reads met quality thresholds) and the Q30 was 93% (i.e., 93% of base calls have a Phred score of ≥30, indicating ≤0.1% per-base error rate). In the unsorted population, >500 unique read locations were identified (out of a maximum 1,819 strains). From the sorted population, observed partial promoter region sequences mapped to 50 different locations within the *E. coli* genome. This demonstrates that cell-sorting was effective in reducing the overall pooled strain population by sorting out strains which were unresponsive to DON. Of the 50 mapped locations from the sorted pool, six locations did not match a gene name included within the advertised composition of the Promoter Collection; this may have been a result of an error or lack of update in the collection inventory, or a possible mapping mismatch. The remaining identified 44 strains from the Promoter Collection are listed in Table 1 with their respective read count (i.e., relative abundance within the mixed population after cell sorting). The strains *lacZ, wrbA,* and *serA* were identified as having a large number of reads (>100,000 each) in both the sorted and unsorted populations.

Table 1. List of *E. coli* Promoter Collection strains detected in the cell-sorted pool, by amplicon sequencing.

Reads Range:	**Strains:**
>100,000	*lacZ, wrbA, serA*
40,000–100,000	*argF*
20,000–40,000	*ggt, guaC*
10,000–20,000	*grxA, argI, trxA*
5,000–10,000	*rpiA*
1,000–5,000	*spy, yeeE, pinR, pphA, yehD,* ***rmf***
100–1,000	*yciT, lacI*
10–100	*cbeA, yoaF, cysH, hipB*
1–10	*ycgR, crp, yhiD, lolC, pyrG, tktA, yidI, soxR, ybgS, ymdE, fliC, yqjC, glpD, cueO, pinQ, amiA,cyoA, ymdF, yciZ, mscL, corA, rraB*

Reads Range indicates the relative abundance of each strain within the sorted pool. The *rmf* strain (bolded) was pursued for biosensor development. Reads were 150 bp in length and sequenced by an Illumina iSeq 100. Reads were mapped to *E. coli* K-12 substrain MG1655 genomic sequence (NCBI: NC_000913.3) using the “map to reference” tool in the Geneious Prime software.

2.3.2 Performance and application testing of two potential *E. coli*-based DON biosensors:

LB medium, the standard for culturing *E. coli*, was found to have high autofluorescence that heavily obscured GFP readings from fluorescence plate assays (data not shown). Therefore, M9 minimal medium with 0.4% glucose was initially chosen due to its low baseline autofluorescence. However, growth of *E. coli* in liquid M9 media is also somewhat slow, and it was found that a clear distinction between GFP production in the DON-positive and control wells was not satisfactorily reached within 24 hours, with measurements after 24 hours showing greater variability and reduced reliability. This formulation was adjusted by changing the carbon source to 0.4% succinate with 0.2% casamino acids. This was largely based on expression studies carried out by Zaslaver et al. (2006), who originally designed and tested the *E. coli* Promoter Collection.[52] They found a broad increase in GFP expression levels across many tested strains in the collection with either 0.2% casamino acids or 0.4% succinate compared to supplementing with 0.4% glucose. In our assays, an M9 formulation supplemented with both 0.4% succinate and 0.2% casamino acids produced more reproducible ON/OFF fluorescence ratios and faster *E. coli* growth compared to unsupplemented M9 broth. Increasing the casamino acid concentration beyond 0.2% accelerated growth further but also increased background autofluorescence (data not shown).

From the cell-sorted population, only strain *rmf* (i.e., one out of the 44 tested strains) demonstrated a clear difference in fluorescence when incubated in the presence of DON compared to the control (*rmf* data follows; all other strains, data not shown). Growth between the control and the DON + (50µg/ml) wells remained consistent until approximately 10 hours of incubation when fluorescence levels of the DON-positive wells began to rise (Figure 5). The

ON/OFF fluorescence ratio of *rmf* began to steadily rise from a value of 1 at around 12 hours incubation to reach a peak of ~1.4-1.5 at 24 hours of incubation when measurement ended.

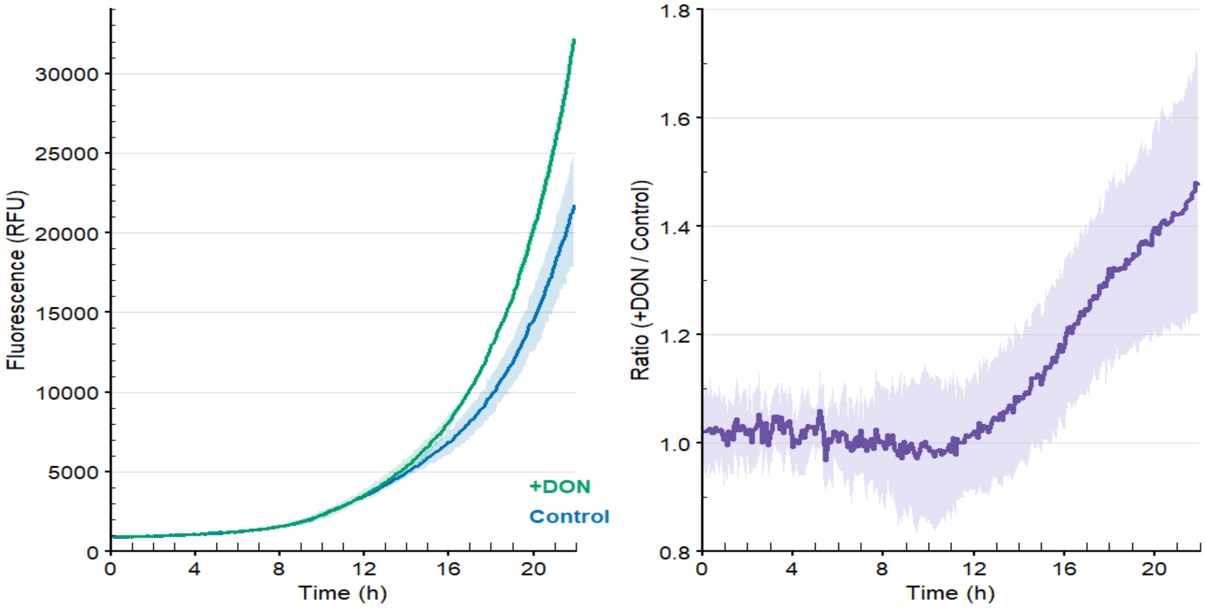


Figure 5. Fluorescence-based assay for *E. coli* strain *rmf* (left) and corresponding ON/OFF fluorescence ratio (right). Excitation and emission wavelengths were set to 479 and 520 nm respectively. A DON concentration of 50µg/ml was used and cells were inoculated in 200µl of M9 minimal media + 0.4% succinate and 0.2% casamino acids. For the fluorescence assay, each plotted line represents the mean of three technical replicate wells measured within the same plate assay, with shading representing ±1 standard deviation from the mean.

The screening of cultured soil microbes using a modified fluorescence plate assay was carried out for a total of 60 isolated fungal colonies and 82 bacterial colonies. Decreases in fluorescence relative to a DON control (10µg/ml) were analyzed but no detectable activity against DON has been identified in these isolates (data not shown).

In an attempt to enhance signal clarity, we placed the *rmf* promoter region upstream of a brighter version of GFP, mClover3. The production of GFP by *rmf-mClover3* + DON (4µg/ml) began to increase over the control at approximately 12 hours post-inoculation, and continued to grow through 22 hours of incubation (Figure 6). Similarly, the ON/OFF fluorescence ratio (Figure 6) steadily improved until reaching a peak of ~1.7 at 22 hours of incubation before decreasing slightly at the end of the assay. RFU readings were much higher for *rmf-mClover3*, peaking around 900,000 RFUs, while fluorescence by the original strain *rmf* from the Promoter Collection peaked at ~100,000 RFUs.

For the fluorescence assay of *ompF-mClover3* (Figure 7), fluorescence intensity of *ompF-mClover3* + DON (4µg/ml) appeared to diverge early from the control. However, the two curves did not diverge substantially until ~19 hours of incubation, slightly earlier than the DON-positive wells, causing the ON/OFF fluorescence ratio to briefly rise until it almost reaches a maximum value of ~ 1.2 (Figure 7).

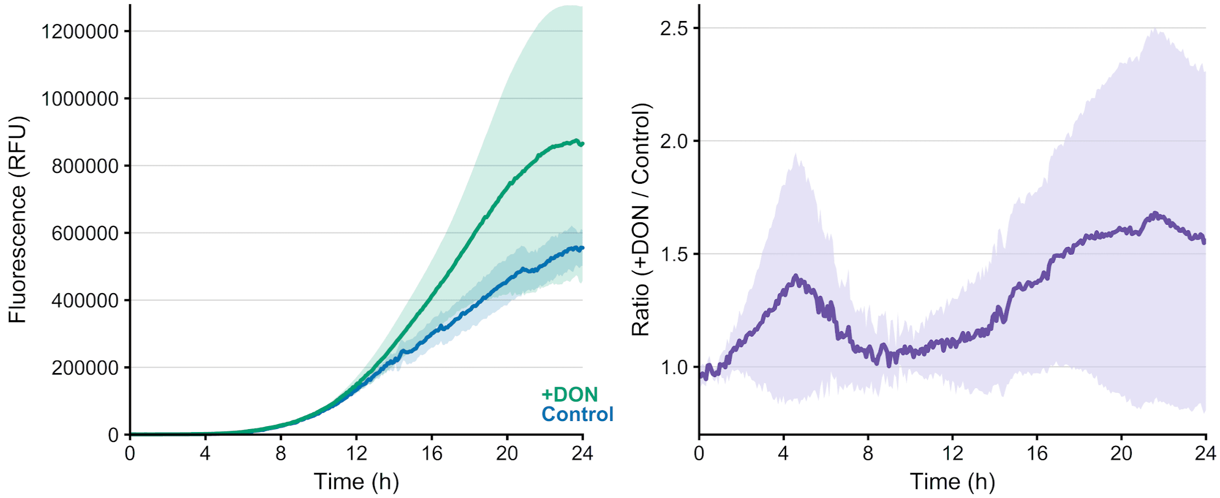


Figure 6. Fluorescence-based assay for *E. coli* strain *rmf-mClover3* (left) and corresponding ON/OFF fluorescence ratio (right). Excitation and emission wavelengths were set to 479nm and 520nm respectively. A DON concentration of 4µg/ml was used and cells were inoculated in 200µl of M9 minimal media + 0.4% succinate and 0.2% casamino acids. For the fluorescence assay, each plotted line represents the mean of three technical replicate wells measured within the same plate assay, with shading representing ±1 standard deviation from the mean.

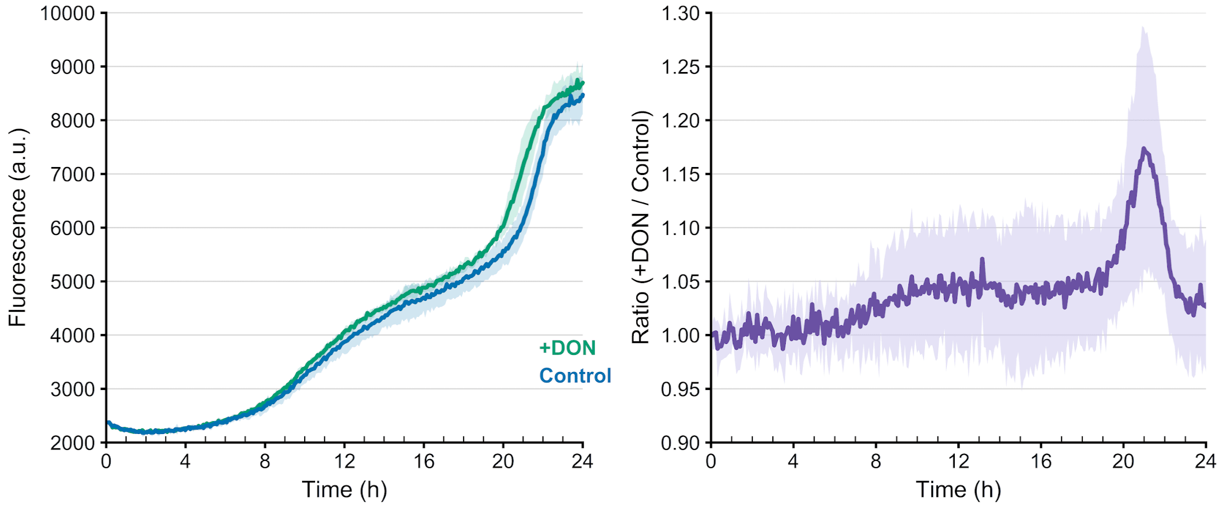


Figure 7. Fluorescence-based assay for *E. coli* strain *ompF-mClover3* (left) and corresponding fluorescence ON/OFF ratio (right). Excitation and emission wavelengths were set to 479nm and 520nm respectively. A DON concentration of 4µg/ml was used and cells were inoculated in 200µl of M9 minimal media + 0.4% succinate and 0.2% casamino acids. For the fluorescence assay, each plotted line represents the mean of three technical replicate wells measured within the same plate assay, with shading representing ±1 standard deviation from the mean.

2.4 Discussion

While FACS eliminated the vast majority of the initially pooled strains from the *E. coli* promoter collection, the final 44 identified cell-sorted strains still needed to be individually evaluated for validation. Attempting to infer biosensor potential of particular strains based on total read count in the pooled amplicon library was obscured by the fact that three strains dominated the library: *lacZ*, *wrbA*, and *serA*. After the experiment, it was discovered these strains occupied multiple wells (each well containing 200µl of glycerol stock) in the twenty-one 96-well plates that make up the entire collection; thus, they were initially present at substantially higher frequencies than the other strains. It is unclear why these three strains specifically were repeated within the library, although *lacZ* would be a common choice as a reference strain in experiments, due to its well-documented induction by IPTG.[55] However, the initial over-representation of these strains in the pool would explain why these strains dominated even in the cell-sorted population. It is likely that sorting constraints were not strict enough to eliminate and prevent overgrowth of particular strains that started out with a slight advantage in the initial growth media. Those three strains aside, the remaining 41 strains that were present after cell sorting were identified by function based on the gene from which the promoter region is derived (according to EcoCyc) as follows: 13 genes contributing to core cellular metabolism and redox, 10 genes involved in gene regulation and stress adaptation (including *rmf*), 7 genes with roles in cell structure and transport, 3 genes involved in DNA and protein post-translational processing, and 8 genes with uncharacterized functions.[56] While these functional annotations helped guide our predictions of which strains would be the most responsive to DON, all 44 strains were experimentally assessed individually in plate fluorescence assays.

The gene *rmf*, named for ribosome modulation factor, is involved in the cellular "hibernation" response in *E. coli* due to stressors like low nutrient availability.[57] Together with Hpf (hibernation promoting factor), Rmf dimerizes the 70S ribosome into 100S complexes, which are translationally inactive, allowing for energy conservation and protecting ribosomes from degradation until favourable growth conditions return. Expression is typically mediated by RelA, a synthetase which binds to the ribosome when a deacylated (or "uncharged") tRNA occupies the ribosomal A site. This produces (p)ppGpp, an "alarmone" which induces *rmf* expression and has been proven to be nearly essential for any *rmf* expression to occur. SpoT is another synthetase (specifically a bifunctional synthetase/hydrolase) which responds to carbon and fatty acid stress to also produce (p)ppGpp and raise *rmf* expression further. When *E. coli* is grown in laboratory culture, *rmf* expression is elevated as cells reach the stationary growth phase and nutrients within the medium have been nearly exhausted.[57] Even at low concentrations, DON may induce an early rise in *rmf* expression due to an imbalance between growth and energy uptake (among other possible mechanisms). DON's primary site of action is the eukaryotic peptidyl transferase center located on the 60S large subunit.[58] DON has reduced affinity and activity against this same peptidyl transferase center pocket in prokaryotes (located on the 50S subunit of the prokaryotic ribosome) due to structural differences.[58]

Taken together, these factors explain why, for *E. coli* strain *rmf*, differences in GFP production become much more pronounced once the cell culture slows down in growth and reaches stationary phase (10-16 hours of incubation time). While *E. coli* cells can withstand very high concentrations of DON (at least as high as 0.5mg/ml, the concentration used in the FACS experiment), it's is likely that as the cell transitions to a metabolically hibernating state, the slight ribotoxic stress imposed by DON compounds the increasingly metabolically starved state of the cell. Consequently, DON-driven stress-response messengers upregulate the *rmf* promoter even as the cells' capacity for protein translation declines. This creates a somewhat counterintuitive biosensor design in which GFP expression is increased under conditions that are somewhat unfavourable for protein synthesis.

Park et al. (2014) studied the transcriptomic effects of DON directly on *E. coli* at concentrations of 2 and 0.2ppm. They incorporated both raw read counts for uninduced and DON-induced samples and a calculated z-score metric that accounts for the magnitude of gene regulation relative to overall read abundance. Higher read counts across both uninduced and induced samples imply higher basal expression of a particular gene. Across both 2 and 0.2ppm trials, *rmf* topped the list for highest z-score, driven by both high basal expression and a modest upregulation factor of 1.61 at 0.2 ppm and 1.78 at 2 ppm. The observed high basal *rmf* expression is not unexpected considering cells were harvested after overnight growth and therefore would be presumed to have reached stationary phase growth. This parallels the experimentally observed maximum ON/OFF fluorescence ratios for both strain *rmf* (~1.3-1.4) and *rmf-mClover3* (~1.6-1.7). Due to the increased brightness of mClover3, fluorescence intensity was also approximately 8-9 times higher than the original *rmf* strain. The fact that no other strains with DON-responsive activity were observed with FACS and plate assays is perhaps unsurprising when analyzing the results of Park et al's transcriptome study; basal expression for the vast majority of these genes is simply not high enough for these differences in fluorescence to be recognized, due to the sensitivity of the plate reader instrument and background/autofluorescence from media and cellular components.

The only other reported gene in the Park et al. (2014) study that consistently reported a high z-score was *ompF,* which has only modest baseline expression but was upregulated by factors of 2.10 and 4.62 for the 0.2ppm and 2ppm trials respectively. Unfortunately, the *E. coli* Promoter Collection does not include an *ompF* promoter strain. Therefore, potential was seen in constructing a plasmid in which the *ompF* intergenic region could be used to control expression of GFP and then be tested for responsiveness to DON. While some strongly downregulated genes were also observed in the transcriptome study, these were not of particular interest. A cell-based biosensor design which relies on decreased fluorescence as a reporting mechanism would be susceptible to false positive results caused by non-DON cell stressors suppressing production of the reporter.

The OmpF protein is a general diffusion porin mainly for transporting small hydrophilic solutes.[59] As such, its increased expression as a response to DON is predicted to be a general stress response mechanism that increases membrane permeability rather than being part of any target removal or efflux of DON from the cell. Therefore, the specificity of such a biosensor would be expected to be low. Unfortunately, the performance of the *ompF-mClover3* strain proved to be sub-optimal when compared to strains *rmf* or *rmf-mClover3*, and as such was not explored further.

There are several reasons why upregulation of gene expression in response to DON exposure, as observed using transcriptomics, may not translate well to expression of GFP in our fluorescence assays. Firstly, even small degrees of background noise in the fluorescence plate assay can obscure differences in GFP expression if basal expression is low. Secondly, transcriptome studies measure mRNA levels, which do not necessarily translate directly to protein levels, depending on post-transcriptional regulation. Lastly, and perhaps most complexly, cloning a gene's upstream intergenic region into a plasmid may not completely capture native transcriptional control factors.[60,61] Key regulatory factors may be missing or distorted, such as distal regulatory sites, chromosomal domains (which aren't fully reproduced in plasmids), as well as differences in gene dosage (even with low-copy plasmids, which can cause titration of gene regulators).

The use of the mClover3 variant of GFP as opposed to the much older GFP-mut2 in theory would increase observed fluorescence by up to 2.5×.[62] The idea behind using a much brighter GFP variant was that it could help overcome sensitivity limitations of the plate reader instrument in detecting fold changes in GFP expression even at generally low expression levels. Higher intensity readings from GFP would also help to reduce the masking effects of autofluorescence created by the media and cellular compounds. The mClover3 also has ~60% increased photostability, meaning brightness intensity levels are maintained longer under continuous (or repeated) excitation.[54]

The *rmf-mClover3* plasmid appeared to function modestly well as a sensor for DON in liquid media, although the relatively small ON/OFF ratio greatly limits further applications where differences in fluorescence cannot be detected as sensitively. We also explored the use of this biosensor in screening cultured soil microbes for DON-detoxifying activity in a high-throughput plate system. When incubated within thin layers of solid media (either containing agar, agarose, or gellan gum), *rmf-mClover3* did not produce visually distinct increases in fluorescence intensity in response to DON when observed using the EVOS plate imaging instrument, even when visualized at multiple incubation timepoints. It's likely that the high degree of light scattering observed within solid media (as well as small levels of autofluorescence) is simply too high for the biosensor's current ON/OFF fluorescence contrast to overcome. Despite this, while a truly high-throughput plate method is not currently achievable with this design, screening of cultured soil microbes was carried out in a way that tries to replicate the currently successful fluorescence plate assay design as much as possible.

While these results demonstrate that our rmf-based *E. coli* DON-biosensor can reproducibly report the presence of DON under defined assay conditions, the data does not confirm the precise molecular or regulatory mechanism that drives this response. In addition, the biosensor's specificity for DON compared to other ribosome-targeting or stress-inducing compounds was not assessed in this study (and therefore cannot be inferred). Therefore, while the biosensor has been empirically validated to report the presence of DON under specific conditions, it cannot be regarded as a mechanistically defined sensor.

Chapter 3 - Rational design of a TRI101-based biosensor for the detection of DON

3.1 Introduction

In contrast to whole-cell biosensors, a protein biosensor has the advantage of being more precise with highly reproducible results, assuming no protein degradation and that the reporter's signal is not interfered with by components of the sample extract. Here, recognition of the target compound is defined by directly binding the designed protein. Reporters can be covalently linked fluorescent molecules (dyes or proteins), redox related (such as with heme groups), FRET systems, may involve enzyme based detection (eg., β-galactosidase), or involve an intrinsic signal by the binding protein itself.[63,64] These biosensors also provide rapid, highly selective, and direct detection of target molecules, greatly limiting the potential for false positives and allowing for real-time detection that does not rely on cellular pathways or metabolism. Sensitivity and dynamic range are also typically substantially improved over cell-based biosensors.

A model protein-based biosensor is described by Smith et al. (2022) in which engineered variants of MalX, a *Streptococcus pneumoniae* protein previously characterized as binding maltooligosaccharides (MOS), were covalently bound to a fluorophore in order to report the presence of MOS through optical based detection.[65] Because MalX undergoes conformational changes when binding MOS, differences in the fluorophore's relative position generates a change in fluorescence intensity. Accurate detection of MOS at concentrations less than 1µM were achievable using this design, marginally better than the cell-based arsenic biosensor. These advances in biosensor design demonstrate the success and versatility of such systems in detecting a wide variety of molecular targets using precise rationale design suited to the specific detection needs.

Building on the principle of rational biosensor design, molecular dynamics (MD) simulations is a way of directly examining the behaviour of proteins and enzymes computationally in simulated solutions that show promising ability to accurately predict real molecular interactions under various conditions and modifications.[66] A single simulation can include many different molecules including cofactors, other proteins (if analyzing protein-protein interactions), as well as ligands. Smith et al. (2022) presents a promising methodology for the rational design of protein biosensors with bound fluorophores, based on computational MD simulations with the target ligand, implemented through the CINC pipeline.[65] Theoretically, such a biosensor could be designed utilizing established crystal structures of the protein/enzyme in interaction with the target ligand. By running parallel MD simulations, both with the ligand present and without, further analysis can establish which residues undergo the largest shift in conformation relative to the rest of the protein due to the presence of the ligand. As such, these residue locations can then be modified to covalently bind a fluorophore, which can then be evaluated in laboratory assays for changes in fluorescence intensity upon ligand binding.

One enzyme with well characterized activity and known binding interactions with DON is TRI101 (the structure and DON binding pocket are visualized in Figure 8), an *F. graminearum* trichothecene 3-O-acetyltransferase.[67] TRI101's main function is to add an acetyl group to the C3 carbon, early in the DON synthesis pathway, converting isotrichodermol to isotrichodermin.[68] At the end of the biosynthetic pathway, TRI8 may remove the C3 acetyl group forming 15-ADON, or remove the C15 acetyl group to form 3-ADON (depending on the Tri8 allele, thus giving rise to the main *F. graminearum* trichothecene chemotypes). Despite functioning earlier in the DON synthesis pathway, TRI101 is also known to bind synthesized DON directly and can acetylate it to 3-ADON *in vitro*.[67,69] In the plant, 3-ADON is often converted back to DON through hydrolysis by plant esterases and deacetylases.[70]

Garvey et al. (2007) were the first study to solve the crystal structure for TRI101 from *F. graminearum*, elucidating the enzyme structure both with its DON and CoA ligands (PDB ID: 3B2S) and in its *apo* state (PBD ID: 2KRM).[67] Therefore, these crystal structures provides the means to analyze TRI101 computationally using MD simulations in order to determine the best fluorophore conjugation sites for designing a DON protein-based biosensor. The 3B2S structure can be utilized for both DON-bound and *apo* simulations. This ensures that variations between the two simulations are strictly limited to either the presence or absence of DON. Simulations including both DON and CoA can further help in validating simulation accuracy and quality, while potentially furthering our understanding of the precise intricacies of TRI101's interactions with DON.[71]

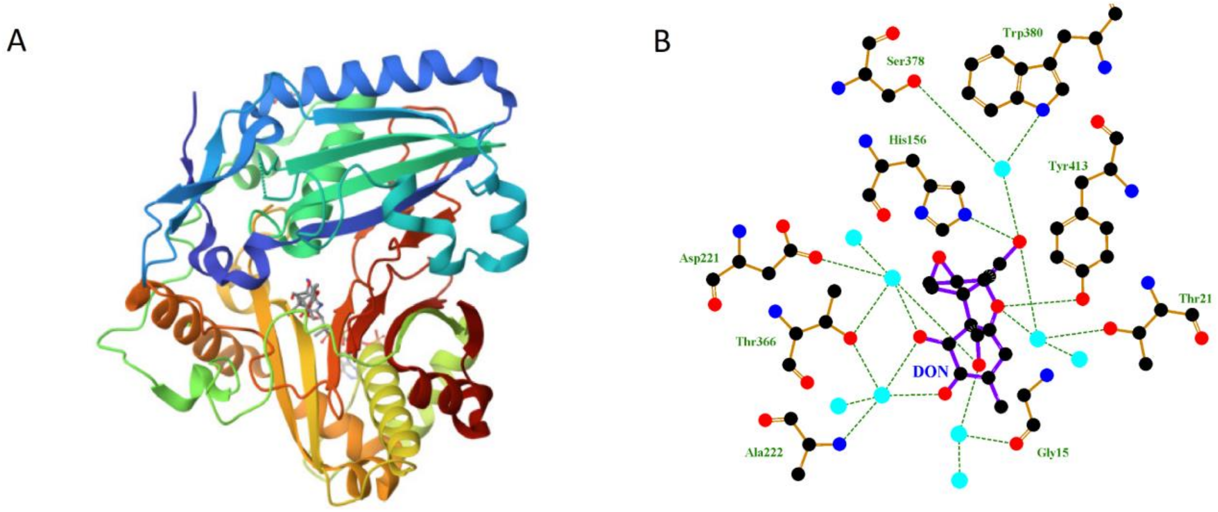


Figure 8. (A) 3D structure of TRI101 from *F. graminearum* binding DON and acetyl-CoA. (PDB entry: 3B2S)[67] (B) LigPlot 2D representation of DON within the TRI101 binding pocket with associated residues and hydrogen bonds.[72] Dotted lines represent hydrogen bonds while blue dots represent water molecules. Both direct and water mediated hydrogen bonds are represented.

3.2 Materials and Methods

3.2.1 Computational system preparation of the TRI101–DON complex

In order to begin preparing molecular dynamics simulations of TRI101, the Protein Data Bank (PDB) crystal structure with accession code 3B2S was used as a starting point. The tleap program (within AmberTools)[73] was used to strip any non-protein molecules (water, ions) from the PDB file so only the protein residues and any ligands remain. The 3B2S protein structure contains both the co-enzyme A and DON ligands. Since our focus is on TRI101-DON interactions, and the fact that two ligands can raise simulation complexity substantially, co-enzyme A was also removed from the PDB file (both ligands were removed to prepare a separate system for the *apo* simulations). There were issues parameterizing the original DON structure, likely due to slightly unusual bond angles around the C11 atom in the crystal structure. Parametrization tools like Antechamber assume that the molecule adopts reasonable covalent geometry.[74] However, if the standard bond length and angle ranges are violated, the program often fails to assign atom typing and bond parameters accurately. A new DON structure was carefully superimposed onto the original DON coordinates in the PDB structure using PyMOL. The structure was then parameterized using Antechamber (AmberTools) using GAFF2 in order to apply the appropriate force fields and compute partial charges. After successfully parameterizing DON, missing hydrogen atoms were then added to the protein and ligand using tleap. The protein-ligand structure was then solvated with tleap by embedding the protein structure in a truncated octahedral TIP3P water box (a commonly used model for water molecules)[75] under periodic boundary conditions (which simulate an infinite, continuous environment). Water molecules were added so any protein or ligand atom was at least 10Å from the box boundary. K^+ and Cl^- counter-ions were added to neutralize the system's overall charge, and then KCl was added to a final concentration of 0.15M (standard for MD simulations). tleap further creates a prmtop file (defining all bonds, angles, charges, molecule types) and inpcrd file (cartesian coordinates for all atoms) which can then be passed on to SANDER for minimization and equilibration. The computational workflow for preparing and running these simulations is summarized in Figure 9. A third simulation was prepared using the 3B2S structure's original double ligand configuration with DON and CoA. Parametrization of CoA was accomplished using its original 3B2S PDB coordinates with topology data generated using Antechamber.

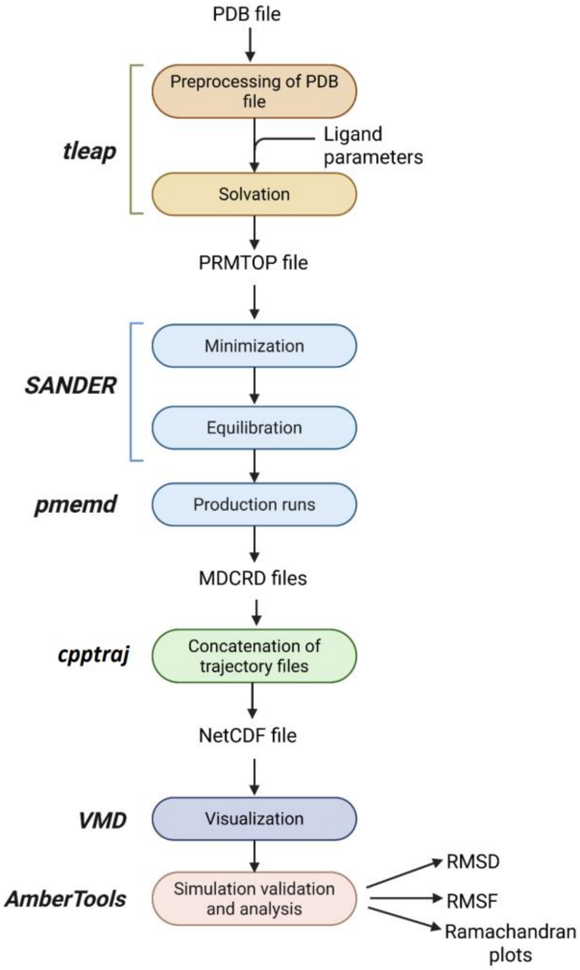


Figure 9. Computational workflow for simulating and analyzing the behaviour of TRI101 in aqueous solution, with and without DON bound as a ligand. The system was prepared with tleap; energy minimization, equilibration, and production molecular dynamics were performed with SANDER/pmemd[76]; and trajectories were analyzed with CPPTRAJ[77]. All programs are part of the AMBER suite. AmberTools was used for downstream analysis of the completed simulations. Antechamber (not shown but included in the AmberTools toolkit) was also needed in order to parameterize a replacement DON molecule for the ligand-bound simulations.

High performance computing servers (Digital Research Alliance of Canada) were used to run the more computationally intensive minimization and production runs using the AMBER suite (version 20.12-20.15). SANDER[76] was used to minimize the system in two steps; solvent and ion relaxation followed by all-atom minimization to remove any bad contacts. SANDER was then further used to heat and then pressurize the system to 300K and 1 atm respectively. The program pmemd was then used to run production simulations, in which five parallel runs were started for TRI101+DON+CoA, TRI101+DON and the TRI101 *apo* systems. Trajectories were saved in NetCDF format every 2 ps, yielding 3000 frames (6 ns total) for analysis. Simulations were visualized with VMD (version 1.9.3) to ensure that anomalies such as loss of ligand did not occur.

3.2.2 Analysis of TRI101 MD trajectories for simulation validation and identification of optimal sites for fluorophore conjugation

Backbone root-mean-square deviation (RMSD) was measured to assess the average displacement of all residue alpha carbons (Cα) in TRI101, from the starting reference structure throughout the simulation.[78] In a well-equilibrated MD simulation, the RMSD rises slightly from a starting value of 1 at the start of the simulation and then plateaus. The relative stability of the plateau (measured by standard deviation) reflects how tightly the trajectory stays in one conformational basin throughout the simulation. Sudden, maintained increases in RMSD are often seen in simulations in which large conformational changes have occurred.[79]

A modified form of the CINC pipeline[65] was developed and used to calculate four individual factors for each TRI101 residue position which affect fluorophore emission energy. Difference in solvent-accessible surface area (SASA) between the apo and ligand-bound simulations measures the degree in which a residue becomes more “buried” or unburied within the protein structure, altering its accessibility to water molecules. Change in psi and phi dihedral angles measures the two backbone dihedral angles of a protein residue: the psi angle which measures rotation about the Cα–C′(carbonyl) bond, and phi angle which measures rotation about the N–Cα bond. Root mean square fluctuation (RMSF) measures how much the backbone atoms of each residue move around their time-averaged position for a simulation. Average distance to the nearest tryptophan residue quantifies how close each position is to aromatic sidechains which can influence fluorescence quenching. After these four factors were calculated with the aid of the AmberTools package (version 23.6), these values were normalized (each value divided by the maximum observed value) before being summed together to generate an Fscore (a composite metric that combines all four normalized factors) for each residue. Calculated distance to ligand for each residue was left out of our Fscore calculations (it was included by Smith et al. 2022) due to the limited predicted effect this factor would have for TRI101, which undergoes quite subtle conformational changes when binding its ligand (i.e., distance to ligand changes between simulation sets is expected to be quite small). Candidate residue locations for fluorophore conjugation were eliminated based on their calculated conservation score using the ConSurf tool.[80] ConSurf generates a multiple sequence alignment (MSA) for a chosen protein with known orthologs and assigns each residue a score (from 1-10) based on the degree of conservation of each amino acid in the chosen protein compared to the orthologous protein sequences at that same position in the alignment. This ensures that the most crucial residues for the structure and function of TRI101 are not considered as candidate sites for mutation to a cysteine followed by fluorophore conjugation. Following the rationale of Smith et al. (2022), all residues with a score higher than two were eliminated as potential fluorophore binding sites, to ensure residues which are the most important for enzyme structure and function are maintained.[65] Residues were also excluded if they were within 5Å of the DON binding pocket, to prevent disruptions to TRI101 binding DON. After this process, a total of 37 candidate residues were identified (shown with their respective Fscores in Figure 12).

CPPTRAJ was also used to measure average minimum distances (as well as standard deviation) and contact frequency (with a set contact boundary of 4Å) between the DON ligand and six binding pocket residues (Thr21, His156, Asp221, Ser378, Trp380, Tyr413) along with a non-binding pocket control (Asp31) for both the TRI101+DON and TRI101+DON+CoA simulations. CPPTRAJ computes minimum distances by identifying the closest pair of atoms between the selected residue and the ligand at each simulation frame. Calculated minimum distances were also analyzed for their distributions throughout each simulated trajectory. Binding free energy calculations were performed using the MMPBSA module in Ambertools to quantify both the total binding free energy and the per-residue energy for the TRI101-DON complex with and without CoA, using the MD trajectory data. Pearson correlation values were calculated based on the minimum distance values for the binding pocket residues throughout the TRI101+DON trajectory in order to quantify correlated residue distance fluctuations relative to DON.

3.2.3 Expression and purification of the G412C TRI101 variant

A novel variant of TRI101, named TRI101-G421C, was codon-optimized for expression in *E. coli* and synthesized as a double-stranded DNA gBlock (Integrated DNA Technologies). The gene was cloned into the kanamycin-selectable bacterial expression vector pET-28a(+) (see Figure A5 for the plasmid map) using the multiple-cloning site restriction sites BamHI and XhoI. In TRI101-G421C, glycine residue 421 was substituted for a cysteine residue to create a site that will facilitate binding to a fluorophore containing a maleimide group. In addition, the original three cysteine residues present in TRI101 had to be modified to non-cysteine residues while trying to maintain normal enzyme structure and function, in order to prevent fluorophore conjugation with these residues (through maleimide thiol chemistry). ConSurf was once again used to determine the most conserved non-cysteine residues at these residue locations in TRI101 orthologs. This resulted in the cysteines at locations 175, 290, and 424 being substituted for leucine, alanine, and glutamic acid, respectively. The fluorophore-cysteine bond is formed when the cysteine's thiol group adds across the maleimide double bond, resulting in a succinimide thioether linkage.[81] A C-terminal His-tag was added to TRI101-G421C for purification purposes.

The TRI101-G421C expressing plasmid was transformed into One Shot™ BL21(DE3) Chemically Competent *E. coli* cells (ThermoFisher). For expression of TRI101, cells were pre-cultured in LB + kanamycin (50 µg/ml) in 15ml loose cap culture tubes at 37°C with 150rpm of shaking until an OD of 1.2 was reached (2-3 hours). 1ml of pre-culture was transferred to 500ml of fresh LB + kanamycin within 2L sterile flasks (3 flasks in total). Flasks were incubated at 37°C with 150rpm of shaking until an OD of 1.2 was reached (3-4 hours). Cultures were then induced with IPTG (from a sterile 1M stock) at a final concentration of 0.5mM and cultures were incubated at 16°C overnight (12-16 hours). During the first extraction of TRI101, an uninduced flask was also prepared for reference on a protein gel. Once IPTG induction completed, cell pellets were collected by spinning cultures down in 50ml centrifuge tubes at 4700rpm and 4°C for 20 minutes. Pellets were combined, washed twice in 50ml ddH_2O and weighed.

The cell pellet (typically between 3-4g) was resuspended in 7 ml/g of lysis buffer (20mM Tris pH 8.0, 0.5M NaCl, 10mM imidazole, 7mM β-mercaptoethanol, 1mM phenylmethylsulfonyl fluoride (PMSF), 1mg/ml lysozyme), and incubated for 30 minutes at 4°C (closely following the protein purification procedure described by Smith et al., 2022).[65] Tween-20 was added to a final concentration of 0.5%. The sample was sonicated (50% output 40% duty cycle; Branson Sonifier

450) while on ice for 3×1 minute intervals, separated by 5 minutes to cool. The lysate was centrifuged at 4°C at 3,000xg for 30 minutes followed by 30,000xg for 45 minutes. A sample of the supernatant (before purification) was taken for analysis by western blot (SDS-PAGE 12% resolving gel) with anti-His antibody (6×-His Tag Monoclonal Antibody; Invitrogen) in combination with the SuperSignal Pierce West Pico PLUS chemiluminescent substrate (ThermoFisher). The membrane was imaged by a GelDoc supporting chemiluminescent imaging (see Figure A6).

For purification, the supernatant was passed through a resin- packed column (5ml of Cytiva Ni Sepharose 6 Fast Flow resin in a 15ml gravity flow column) already equilibrated with binding buffer (20mM Tris pH 8.0, 0.5 M NaCl, 20mM imidazole). The supernatant was passed through the column three times at 4°C. The resin column was then washed five times with 30ml of fresh binding buffer. Bound protein was eluted with 10ml of elution buffer (20mM Tris pH 8.0, 0.5 M NaCl, 300mM imidazole), which was passed through the column 3 times.

3.2.4 Fluorophore labeling and testing of TRI101-G421C

Labeling of TRI101-G421C was attempted with five different fluorophores: fluorescein-5-maleimide (F5M), dansyl, pyrene, 7-diethylamino-3-(4-maleimidylphenyl)-4-methylcoumarin (CPM), and 7-diethylamino-3-(4′-maleimidylphenyl)-4-methylcoumarin-3-carboxamide (MDCC). 10ml of protein eluate was buffer-exchanged into PhoS buffer (50mM TRIS-HCl pH 7.5, 300mM NaCl, 1mM $MgCl_2$, 10mM imidazole, 15% glycerol, 1mM PMSF) using 22mm, 10 kDa SnakeSkin™ dialysis tubing (Thermo Scientific). The sample was dialyzed overnight at 4°C with gentle stirring against 1 L of buffer, followed by a second dialysis step in fresh buffer.

Importantly, although beta-mercaptoethanol (BME) helps keep the cysteine residue reduced by preventing disulfide formation, BME must not be present in the buffer when attempting to conjugate the fluorophore to the protein, since BME's thiol group will react with and quench the maleimide group of the fluorophore before it can bind the cysteine. For each fluorophore, 0.5ml of Ni Sepharose 6 Fast Flow resin previously equilibrated with PhoS buffer was added to a 1.5ml microcentrifuge tube. The concentration of purified protein was determined spectrophotometrically at 280nm (NanoDrop, Thermo Fisher Scientific). 500µl of purified protein was added to each tubes, with the fluorophores (1 per tube) added to a final concentration 5× the protein concentration. Tubes were wrapped in foil and incubated at 4°C with gentle rotation on an orbital shaker, before being centrifuged at 500rpm for 2 minutes to spin down the resin. The supernatant was removed and the resin was washed in 1ml of PhoS buffer three times before being eluted with 500µl of elution buffer (PhoS buffer with 500mM imidazole). Elution was repeated 3 times (or more) with fresh elution buffer and elution fractions were measured for protein concentration, followed by pooling of fractions with high protein concentration readings. The pooled protein solution was then dialyzed in PhoS buffer (excluding imidazole and glycerol) twice as previously described.

Fluorescence of TRI101 conjugated with each of the fluorophores (1µM) was measured (Horiba Scientific fluorometer) with and without 10µM of DON, at the appropriate excitation wavelength and emission range for each fluorophore. Excitation wavelengths of 340nm, 387nm, 493nm, 420nm, and 340nm were used for dansyl, CPM, fluorescein, MDCC, and pyrene respectively.[82–85]

3.3 Results

3.3.1 Analysis of TRI101 molecular dynamics simulation data

I successfully modeled the dynamics of the TRI101 protein during ligand binding by DON or by DON + co-enzyme A, with 5 in silico technical replicates. For the molecular dynamics simulations of TRI101 with a DON ligand but without CoA, four of the six binding pocket residues (Thr21, His156, Asp221, Ser378, Trp380, Tyr413) maintained at least 30% contact frequency (within 4Å of DON throughout 30% of the trajectories). His156 reported the highest degree of contact with DON at 96.69% while maintaining the closest average distance (3.113Å) of all six residues. Thr21 had a contact frequency of only 0.23%, despite its relatively low mean minimum distance of 5.531Å. This was similar to Trp380, which also maintained close distance to DON but a low contact frequency of 6.54%. Per-residue free energies (the calculated contribution of each residue to the overall free energy of TRI101 binding DON) were relatively strong for each binding pocket residue except His156, which maintained a slightly positive ΔG. Tyr413 maintained the strongest apparent association with DON at -0.8838 kcal/mol.

In the simulations including both DON and co-enzyme A, five of the six binding pocket residues (not Thr21) reduced their average distance to DON by an average of 0.527Å, while contact frequencies increased for all except Thr20 (see Table 2). All calculated free energies increase significantly in the CoA bound trajectories, with His156 exhibiting the fourth-strongest per-residue binding contribution in the CoA-bound state, highlighting its distinct role in enhancing DON stabilization within the binding pocket exclusively when CoA is bound.

Four of the six binding site residues (not Trp380 or Thr21) adopted a slightly bimodal distribution (Figure 10) in distance to DON over the modeled trajectories with DON as the only ligand. Pearson *r* correlation coefficients were calculated from these minimum distance trajectories in order to quantify the coordinated movement between binding pocket residues compared to the non-binding pocket residue control Asp31 (see Figure A4). When CoA is additionally bound, however, these bimodal features are lost and distributions become significantly more consolidated and uniform. When CoA is bound, Trp380 demonstrates a statistically significant decrease in distance to DON (see Figure 10).

Table 2. Binding pocket residue-ligand contact analysis during molecular dynamics simulations for TRI101 with DON (blue cells) and for TRI101 with both DON and CoA (green cells).

Residue	Mean Minimum Distance to Ligand (Å)	Maximum Observed Minimum Distance (Å)	Contact Frequency (%)	First Frame Minimum Distance (Å)	DON binding free energy (kcal/mol)
Thr21	5.531 ± 0.701	7.875	0.23	6.201	-0.0447 (26)
	5.665 ± 0.285	6.905	0.00	5.593	-0.1391 (25)
His156	3.113 ± 0.391	4.698	96.69	3.098	+0.2032 (450)
	2.871 ± 0.144	3.832	100.00	2.779	-1.4093 (4)
Asp221	4.572 ± 0.814	7.216	29.73	3.730	-0.1006 (21)
	4.020 ± 0.666	7.124	59.62	4.061	-0.7788 (7)
Ser378	3.660 ± 0.379	4.973	84.25	3.269	-0.2567 (12)
	3.464 ± 0.255	4.666	99.07	3.496	-0.5619 (14)
Trp380	5.242 ± 0.760	7.670	6.54	3.995	-0.1203 (19)
	3.559 ± 0.249*	4.714	94.92	3.290	-0.5876 (12)
Tyr413	3.263 ± 0.582	5.341	86.02	3.539	-0.8838 (3)
	2.907 ± 0.173	3.689	100.00	2.864	-2.5737 (1)
Asp31	35.139 ± 0.562	37.018	0.00	35.053	0.0000 (292)
	33.970 ± 0.489*	35.407	0.00	34.048	-0.0001 (210)
				Total DON binding energy for TRI101 (kcal/mol)	-18.38 ± 0.91
					-33.74 ± 0.80

Residue to ligand distance calculations were made using the CPPTRAJ (V6.18.1) within AmberTools (Version 23.6) using one full production trajectory of 3000 frames (saved every 2 ps) for TRI101 bound to DON (highlighted in blue) and TRI101 bound to both DON and CoA (highlighted in green). Binding free energies (ΔG) were calculated using MM/GBSA[86] decomposition of the TRI101 MD trajectory data using AmberTools. Bracketed numbers represent how each energy value ranks out of all 451 residues in the simulation. Asp31 is used as the non-binding pocket control residue. *The difference of means for Trp380 is statistically significant, $t(5998) = 99.9$, p < 10^(-300). The difference of means for Asp31 is also significant, $t(5998) = 78.4$, p < 10^(-300).

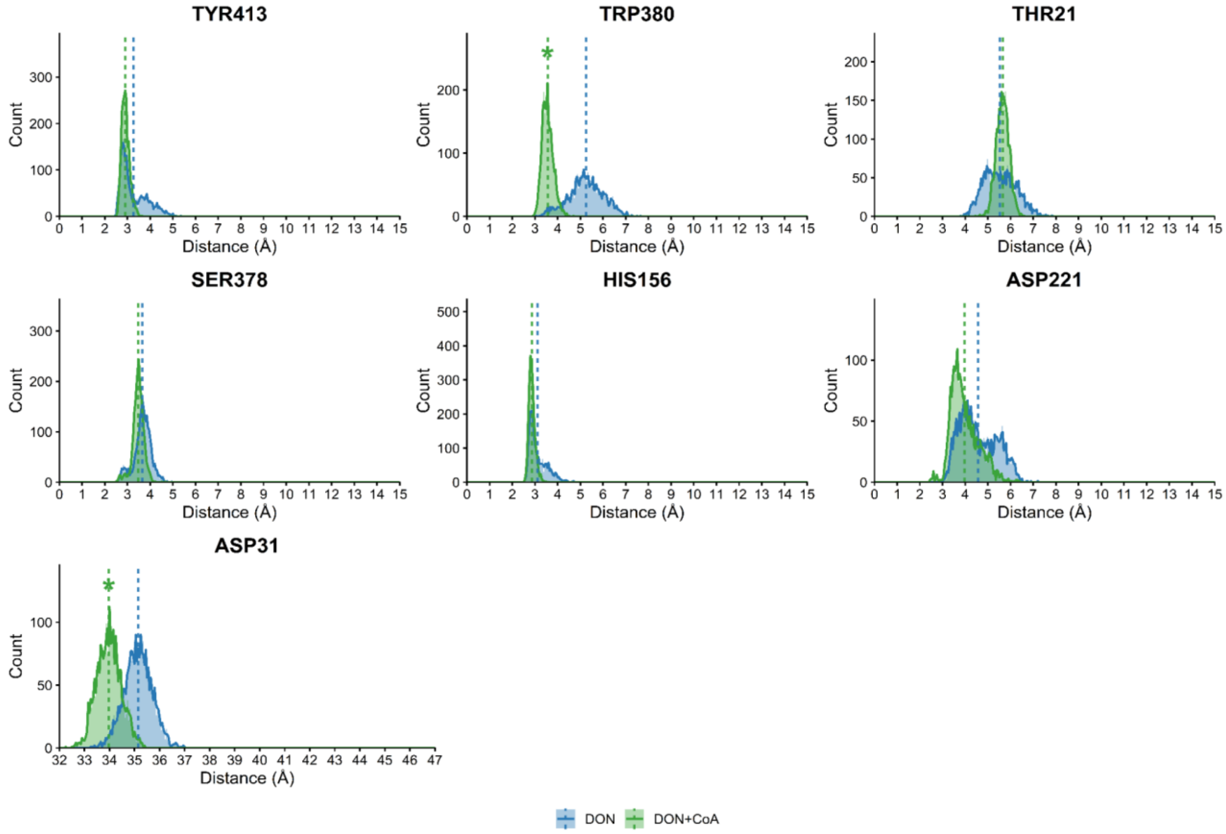


Figure 10. Histograms showing the minimum distance between DON and each of the six analyzed TRI101 binding-pocket residues during MD simulations containing the DON ligand alone (blue colour) or both DON and CoA ligands (green colour). Asp31 is included as a non-binding pocket residue control.

Root mean square deviation (RMSD) values were analyzed and plotted for the five *apo* and five DON-bound TRI101 simulation trajectories. On average, the standard deviation in RMSD values was 0.320Å for the *apo* simulations and 0.286Å for the DON-bound simulations (Figure 11) indicating that each set of simulations are relatively stable and well equilibrated. No large jumps in RMSD values were observed throughout any of the simulations, indicating that there are no alternate protein conformations being adopted throughout each trajectory.

Based on molecular dynamics simulations of TRI101 in its *apo* state and in interaction with DON, integrated into a single Fscore per residue, 37 amino acid residues were explored for their potential as fluorophore conjugation sites (Figure 12). Both Trp distance and SASA were relatively low and consistent across the residues, while RMSF contributed substantially to the Fscore only for E279. Differences in psi-phi angles provided a greater than 50% contribution to the total Fscore for five residues: E63, A218, G219, P419, and G421. These residues were also the top five scoring candidate residue locations in TRI101 for fluorophore conjugation. For residues A218 and G421, the shifts in psi-phi angles became especially apparent when visualizing their respective Ramachandran plots for each set of simulations (Figure 13). For A218, the residue backbone initially adopted what is known as a polyproline II (PPII)-like conformation, which indicates the backbone is in a more flexible and solvent exposed state.[87] However, with DON in the binding pocket, this conformation shifted and the backbone adopted some more constrained, helix-like torsions.[88] G421, on the other hand, shifted from a left-handed helix conformation (*apo* state) to a right-handed helix when bound with DON, while also adopting some extended, beta-sheet-like torsions (which glycines can easily adopt).[89] Due to glycine's lack of a beta-carbon, this movement appears to be a hinge-like pivot that may allow for nearby residues to move closer to DON upon binding.

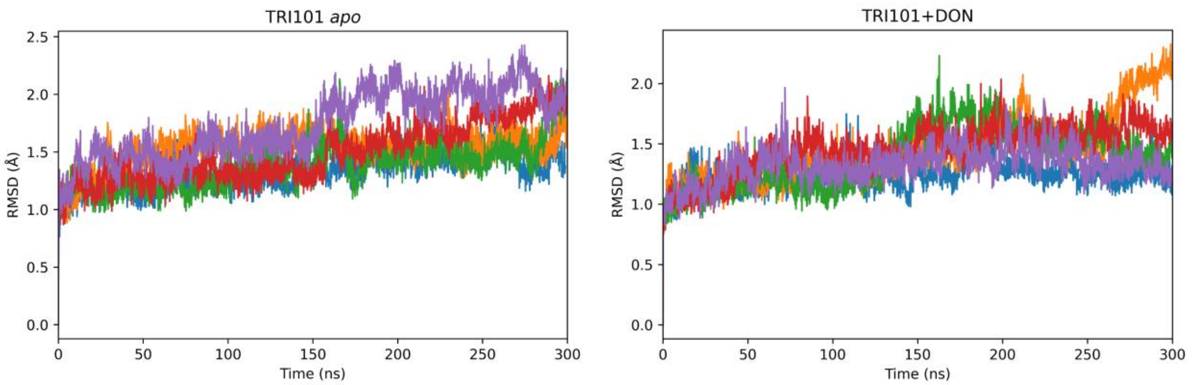


Figure 11. Root mean square deviation (RMSD) plots for both the *apo* and ligand-bound TRI101 trajectories. Each of the five replicate runs are represented by a different colour on each plot.

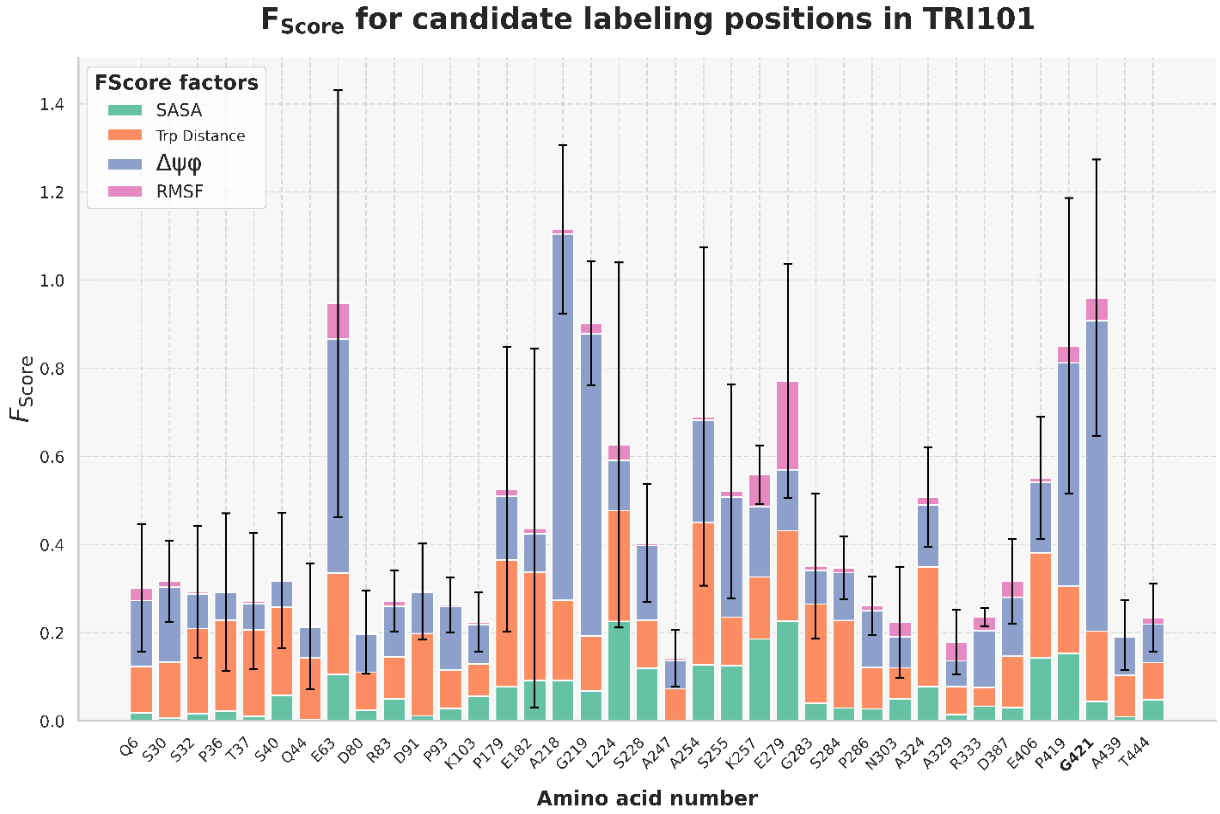


Figure 12. Total calculated Fscore values for candidate fluorophore conjugation sites in TRI101. G421, the residue selected as a fluorophore conjugation site for biosensor testing, is indicated in bold. Fscore components include SASA (change in solvent exposure), Trp distance (average distance to the nearest tryptophan), Φ/Ψ angles (backbone dihedral angle shifts), and RMSF (backbone flexibility). Error bars indicate ±1 standard deviation.

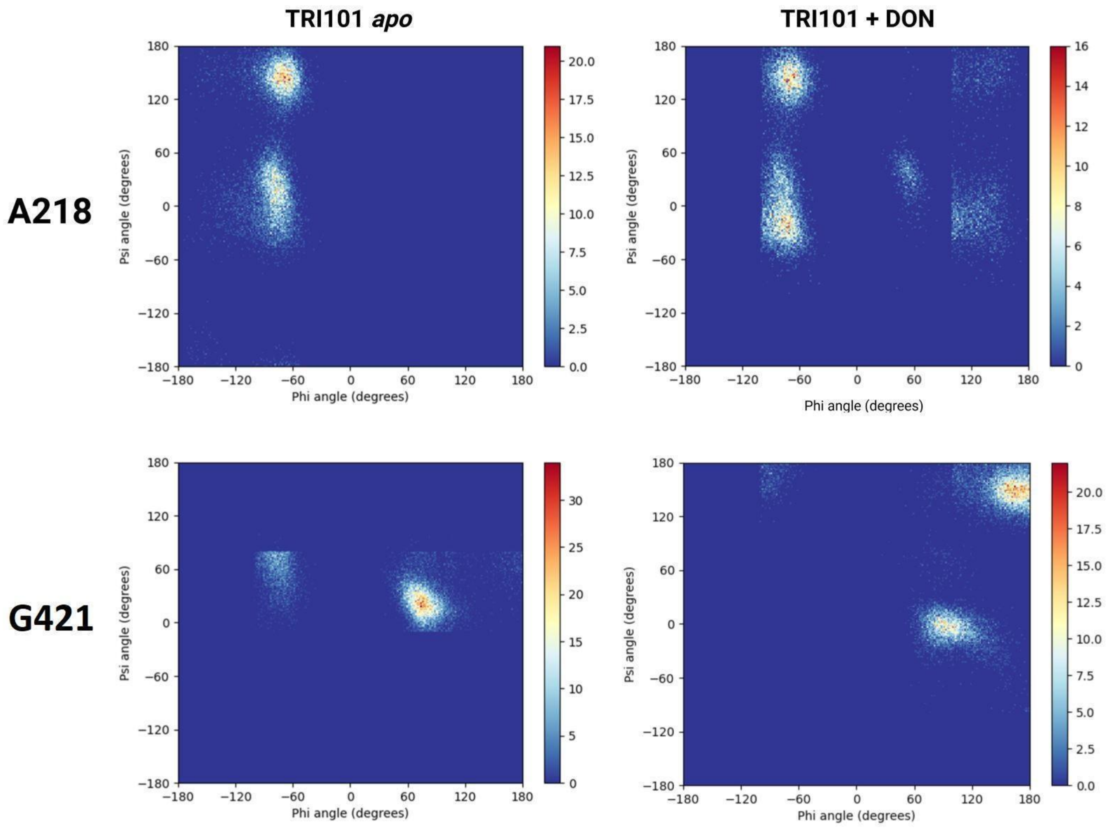


Figure 13. Ramachandran plots for TRI101 residues alanine 218 and glycine 421. Psi-phi angles are averaged across the 5 parallel simulation runs for both ligand-bound and *apo* MD simulations. Colour scale ranges from blue (low occupancy of φ/ψ angles) to red (high occupancy).

3.3.2 Expression optimization and fluorescence analysis of fluorophore-labeled TRI101-G421C

Initial SDS-PAGE analysis did not immediately show a large expression band around the expected 55kDa for TRI101-G421C. Therefore, different induction conditions were attempted, including varying IPTG and L-rhamnose concentrations and induction times. Western blot analysis was used to see if low levels of expression were being obscured in SDS-PAGE by other *E. coli* proteins. Anti-6×-His monoclonal antibody binding with chemiluminescent staining revealed that TRI101-G421C expression was occurring but at lower than expected levels. Lower concentrations of IPTG (0.2mM, adjusted down from 1mM) gave a very slight boost to expression as confirmed with Western blot (Figure A6).

Analysis of the fluorophore-labeled TRI101-G421C samples revealed no significant changes in fluorescence intensity when DON was added to the solution for MDCC, fluorescein, pyrene, and dansyl (Figure 14). However, in the case of the CPM fluorophore there was a depression near the emission maximum (c. 480nm) when DON was present (Figure 14). This depression is calculated to be around a 8-9% reduction in fluorescence intensity.

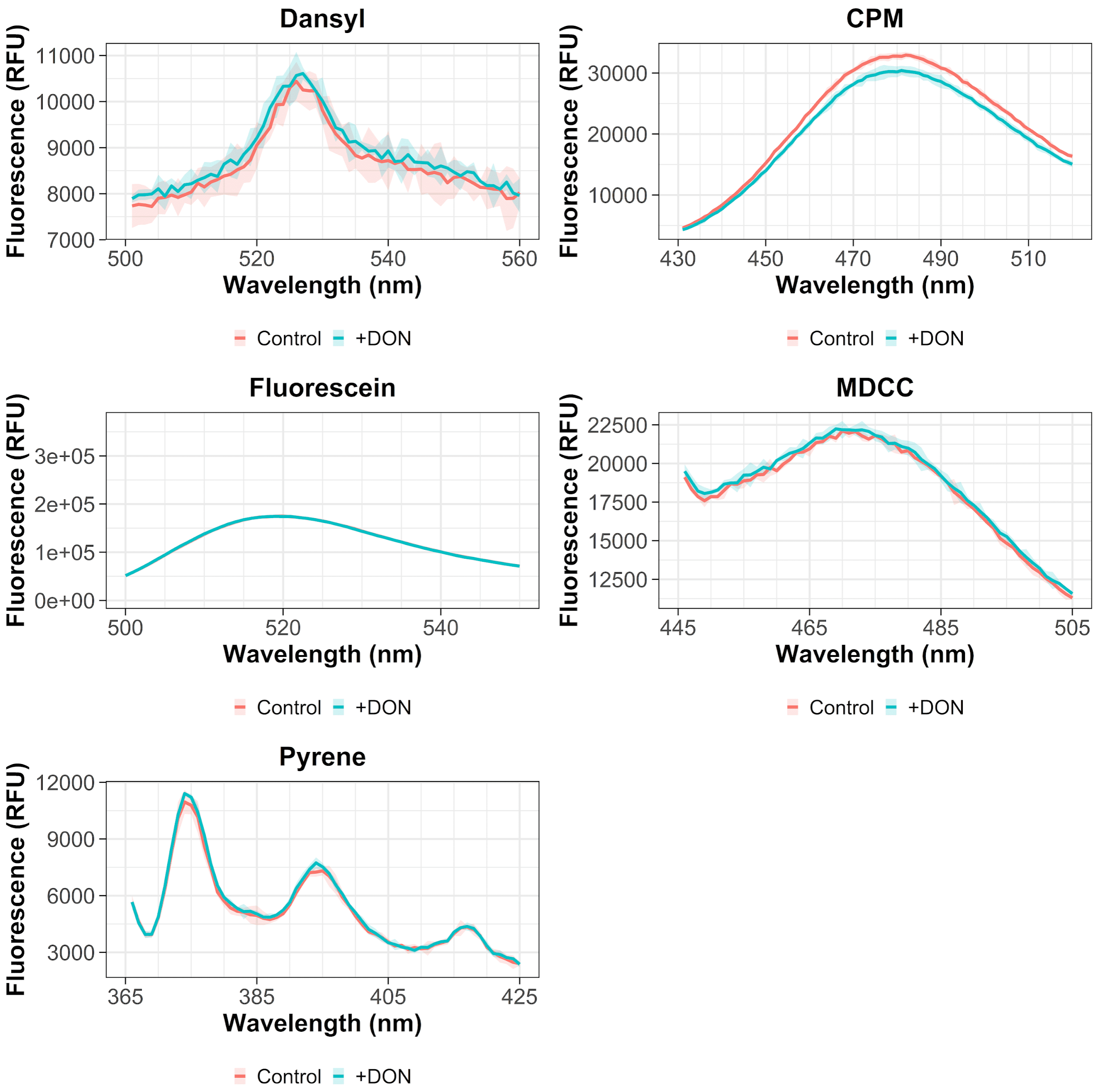


Figure 14. Fluorescence emission of the TRI101-G421C protein conjugated to five different fluorophores. The TRI101 variant G421C was conjugated to the fluorophores dansyl, CPM, fluorescein, MDCC, and pyrene. Labeled proteins were purified and diluted to a concentration of 1µM. Fluorescence emission intensity was measured with a Horiba Scientific fluorometer in the absence of DON or with 10µM of DON. Excitation wavelengths were selected according to each fluorophore's excitation maximum: 340 nm (dansyl), 387 nm (CPM), 493 nm (fluorescein), 420 nm (MDCC), and 340 nm (pyrene).[82–85] Each trace represents the mean of three technical replicates, while shaded regions indicating ±2 standard deviations. For the fluoroscein measurements, the emission scan range overlapped with the excitation wavelength, introducing

potential excitation light contamination. Therefore, the absolute emission intensities and accuracy of these data should be interpreted with caution.

3.4 Discussion

There are a few major factors which make designing a TRI101-based DON biosensor potentially trickier than the example set by Smith et al. (2022) with their MalX-based biosensor.[65] Firstly, TRI101 does not undergo the kind of large, ligand-induced motions when binding DON that MalX does with its large domain closure when binding maltose oligosaccharides). Therefore, even with perfect fluorophore placement, it is possible that observed changes in amplitude signals will be more subtle than those observed for MalX, consistent with the smaller ligand-induced changes in residue positioning and flexibility predicted by the CINC pipeline for TRI101. MalX also does not associate with any cofactors, while TRI101 uses acetyl-CoA as an acetyl source. Acetyl-CoA was not used in our functional tests of TRI101-G421C as a biosensor because of the expectation that it would result in acetylation of DON, which would be counterproductive to the biosensor's function, as enzymatic conversion would chemically alter the target analyte during detection. However, the absence of acetyl-CoA is predicted to lower DON's binding affinity for TRI101 somewhat (based on more general studies on the behaviour of acetyl-transferases).[90] With the aid of our MD simulations, we were able to confirm that DON maintained strong associations with certain binding pocket residues throughout the simulation even in the absence of acetyl-CoA, as demonstrated by the average minimum distances between DON and binding pocket residues (Table 2). Thus, we anticipated that this biosynthetic enzyme involved in the formation of DON could be repurposed into a reporter for the presence of DON in a solution.

Minimum distance calculations for DON and six binding pocket residues including data from TRI101 trajectories with both DON and CoA ligands revealed interesting insights about DON-TRI101 binding that were not explicitly revealed by the TRI101-DON and TRI101 *apo* simulations alone. While TRI101 appears to bind DON strongly in the absence of CoA, the minimum distance distributions shown in Figure 10 (and more specifically their slightly bimodal characteristics) show that DON still has space to fluctuate (or "jiggle") within the binding pocket. The restriction of binding pocket residue movement (and therefore DON movement) once CoA is bound implies that DON is being bound even tighter within the pocket and precisely positioned for acetylation, as further evidenced by stronger system (TRI101) and per-residue binding free energies for DON. Consistent with this pattern, Garvey et al. (2007) describe His156 as the catalytic base that deprotonates DON at the C3-hydroxyl group.[67] This specific activity primes DON for the 3-O-acetyl transfer from acetyl-CoA. Our per-residue free-energy analysis shows that His156 contributes very weakly and unfavorably to DON binding in the absence of CoA (+0.2032 kcal/mol), but becomes strongly stabilizing once CoA is bound (–1.4093 kcal/mol), suggesting His156 plays a large role in sensing CoA binding and thus priming DON for acetylation. Not only does elucidating these interactions provide deeper insight into TRI101's complex interactions with DON, but it also strongly supports the overall rational and validity of our MD simulation approach to biosensor design.

Backbone RMSD plots for the TRI101 with DON and TRI101 *apo* simulations showed early plateaus that remained relatively stable throughout each production run, indicating these were well-equilibrated and stable simulations. A rapidly reached and maintained RMSD plateau is typically seen as evidence of a well set-up simulation with a stable protein conformation.[66,91] The standard deviation values across each set of simulations were minimal, further indicating that

TRI101 maintained its native-like conformation throughout each simulation and did not fluctuate significantly.

When initially visualized with VMD, the simulations did show signs of what are known as periodic boundary condition artifacts, where DON would seemingly disappear and reappear randomly (data not shown). In reality, the ligand was crossing the water box boundary, causing it's coordinates to be periodically 'wrapped' to the opposite side of the simulation box. This can negatively affect protein-ligand calculations performed using these simulation files. To correct this issue, the simulations were processed with CPPTRAJ to 're-image' the system so that DON remains within the unit cell which is nearest to the protein.

CPPTRAJ was used to calculate four of the five CINC factors as described by Smith et al. (2022), namely solvent-accessible surface area, change in psi-phi angles, root-mean square fluctuation, and average distance to tryptophan.[65] We omitted distance to the ligand as a contributing Fscore factor, due to its likely unsubstantial contribution to choosing a candidate fluorophore conjugation site for TRI101. Since TRI101 does not undergo a large conformational change (unlike MalX's domain closure when binding MOS), we expected differences in distance to DON between each set of simulations to be extremely small and to contribute little to prediction of optimal conjugation sites for fluorophores.

Distance to tryptophan scores were predictably the most uniform across the top-scoring residues, with an average score of 0.21 and a range of 0.10 to 0.35 across all simulation replicate trajectories. This makes sense considering there are a total of six tryptophan residues in TRI101, all somewhat equally distributed across the protein structure. Therefore any movement a residue may make during a simulation away from one tryptophan is likely towards one or more of the others. RMSF provided very little contribution to the total Fscore for the 37 candidate residues, with E279 receiving the largest RMSF score (meaning it fluctuated to a relatively high degree around its average position during the ligand-bound simulations compared to the *apo* simulations). SASA made higher contributions to total Fscore than did RMSF, but remained comparatively minor overall. While Smith et al. (2022) designed and utilized their CINC pipeline and formulas in order to identify top residue location candidates for fluorophore conjugation, they retrospectively found change in psi-phi angles to be the factor that served as by far the largest predictor to find the best performing sites. Of the top scoring TRI101 residues, G421 and A218 demonstrated significant visual differences in psi and phi angles in their respective Ramachandran plots. Given that alterations in these angles are expected to be the primary driver of variations in fluorophore fluorescence emission intensity between protein states, these two residues became our top two candidates (with G421 as #1) for recombinant TRI101 expression, fluorophore conjugation, and fluorescence assays in order to develop the most effective DON biosensor. When comparing psi-phi angle changes between these two most promising residues, G421 appears to adopt a larger psi-phi reorientation which is permitted due to glycine's lack of a beta carbon and increased torsional space. These properties may give this residue location much more potential for shifts in emission intensity of a bound fluorophore, compared to A218.

TRI101 protein production remained limited, despite exploring various induction conditions. Therefore, a 16°C overnight induction was settled on as the safest protocol due to its established use for TRI101 recombinant expression in *E. coli* by Garvey et al. (2007).[67] Given TRI101's eukaryotic origin, expression in yeast was a tempting alternative that may have produced higher protein yields, but a bacterial expression system was pursued following the precedent of the TRI101 crystallography study and to take advantage of the faster and lower relative cost of expression in *E. coli*. Different formulations of lysis buffers and protocols (including testing various sonication times, detergents, and Ni-NTA purification protocols) were tried with little improvement. Garvey et al. (2007) mention that TRI101 solubility was improved by the addition of acetyl-CoA in the cell lysate, however that was not tried due to the relative amounts of acetyl-CoA that would be required when taking into account its high cost.

It would be worth exploring why CPM produced subtle changes in amplitude fluorescence intensity at the emission maximum when testing its performance in TRI101-G421C, compared to the other four tested fluorophores. Firstly, CPM is a UV excited maleimide dye and has excitation and emission maxima around 384nm and 470, respectively.[92] This makes CPM slightly more sensitive to changes in local protein microenvironments such as nearby aromatic residues and polarity.[92,93] CPM is also weakly fluorescent until it reacts with a thiol group.[94] Therefore, background from unreacted CPM still present in the solution is less likely to obscure shifts in fluorescence intensity.

We cannot determine from the obtained data whether the observed reduction in fluorescence intensity in the TRI101-G421C-CPM sensor results from quenching effects arising from proximity to aromatic residues, changes in fluorophore solvent exposure, RMSF or psi-phi angles, or a combination of these factors. Therefore, although the data demonstrates that a TRI101-based fluorescent biosensor for DON is feasible, they do not establish G421 as the optimal fluorophore conjugation site. Additionally, our molecular dynamics simulations only demonstrate the behaviour of native TRI101 residues and therefore they may not fully capture the structural effects of cysteine substitution and subsequent fluorophore binding.

Chapter 4 - General Discussion and Conclusions

Across the 44 identified and tested strains from the *E. coli* Promoter Collection, only one strain, *rmf,* demonstrated a reproducible increase in the production of GFP in the presence of DON. Experimental ON/OFF fluorescence ratios were in the range of ~1.3-1.4. This ratio was enhanced to ~1.6-1.7 when paired with a brighter variant of GFP, mClover3. This change in reporter variant boosted absolute fluorescence which may in turn have improved the assay's signal-to-noise ratio. DON concentrations as low as 2µg/ml were found to be detectable in liquid M9 media using this strain. An *ompF*-based promoter strain we generated also using mClover3 showed suboptimal performance in fluorescence plate assays, only reaching maximum ON/OFF fluorescence ratios of ~1.1-1.2. When attempting to adapt strain *rmf-mClover3* to a high-throughput screening design using solid media, the biosensor's performance was impaired by light scattering and autofluorescence uniquely inherent to the solid medium. A non-high-throughput screening method was designed in which individually cultured soil microbes could be screened using the same fluorescence plate assay. A total of 60 fungal colonies and 82 bacterial colonies were screened using this modified assay design, however no DON-modifying activity was identified.

Further improvements could be made to *rmf-mClover3*'s performance through genetic changes to the *rmf* promoter region such as promoter element duplication, whole site duplication (in case of any currently unidentified regulatory elements in the intergenic region), or by FACS-based selection of favourable mutants in which basal production and dynamic range are improved. Very low-copy plasmids could also be tried to attempt to reduce any titration artifacts.

Using molecular dynamics simulations and protein crystal structures, the *F. graminearum* protein TRI101 was analyzed computationally to determine where targeted amino acid substitutions could improve its function as a protein-based DON biosensor. TRI101 was simulated in both its ligand-bound and *apo* states, which revealed some previously undescribed binding pocket residue dynamics, and allowed for the computational calculation of four protein-residue dynamics factors, which were used to determine optimal fluorophore conjugation sites for engineering a TRI101-based DON biosensor. DON's stability within the binding pocket and optimal RMSD values demonstrated the relevance and accuracy of the built simulations. DON persistently associated with the pocket without the acetyl-CoA cofactor, aided by non-H-bond pocket dynamics like Ser378's steric packing near the epoxide group and intermittent, rotamer-mediated contact of His156 near the DON C3 hydroxyl group. This behaviour is consistent with His156's previous characterization as a catalytic de-protonator of DON at this site. Our analysis of candidate fluorophore locations supported the idea that backbone psi-phi angle changes are the strongest predictor of optimal sites for fluorophore conjugation. Among the high-scoring residues, G421 and A218 demonstrated the clearest shifts in the generated Ramachandran plots between the two simulation sets. This supports their selection as the best sites for fluorophore labeling.

The low experimentally observed expression levels of TRI101 when recombinantly expressed in *E. coli* greatly limited the potential for further, more rigorous testing. Conjugation of fluorescein-5-maleimide to the G421C site produced the most detectable fluorescence intensity changes among the five tested dyes. This result is perhaps consistent with fluorescein's particular qualities of demonstrating higher brightness, lower background/autofluorescence effects, as well as high, reproducible labeling yields under standard conditions. Together, these findings indicate that the mechanisms and theory behind a TRI101 DON biosensor are sound, however protein yield and assay sensitivity have become the main challenges. While the difference in peak amplitude appears to be too slight for TRI101-G421C-CPM to function as a reliable DON biosensor, the results do confirm that the rational design of a TRI101-based DON biosensor is on the right track, although challenges with TRI101 expression and assay sensitivity remain.

In future work, additional testing should be performed on TRI101-A218C, and sites E63, G219, and P419 also merit further testing based on their high calculated scores. Tools developed in this work contribute to the effort of screening at a throughput that will be adequate to discover novel enzymes for detoxifying DON, which could find diverse commercial applications.

References

1. Amarasinghe, C. C., Simsek, S., Brûlé-Babel, A. & Fernando, W. G. D. Analysis of deoxynivalenol and deoxynivalenol-3-glucosides content in Canadian spring wheat cultivars inoculated with *Fusarium graminearum*. *Food Additives & Contaminants: Part A – Chemistry, Analysis, Control, Exposure & Risk Assessment* **33**, 1254–1264 (2016).
2. Bönnighausen, J., Schauer, N., Schäfer, W. & Bormann, J. Metabolic profiling of wheat rachis node infection by *Fusarium graminearum* – decoding deoxynivalenol-dependent susceptibility. *New Phytologist* **221**, 459–469 (2019).
3. Scherm, B. et al. Altered trichothecene biosynthesis in TRI6-silenced transformants of *Fusarium culmorum* influences the severity of crown and foot rot on durum wheat seedlings. *Molecular Plant Pathology* **12**, 759–771 (2011).
4. Bamforth, J. et al. A survey of species and ADON genotype on Canadian wheat grain. *Frontiers in Fungal Biology* **3**, 1062444 (2022).
5. Ward, T. J. et al. An adaptive evolutionary shift in Fusarium head blight pathogen populations is driving the rapid spread of more toxigenic *Fusarium graminearum* in North America. *Fungal Genetics and Biology* **45**, 473–484 (2008).
6. Canada. Botany and Plant Pathology Division, Canada. Division of Botany & Plant Research Institute (Canada). *Annual Report of the Canadian Plant Disease Survey* (2022).
7. Bianchini, A., Horsley, R., Jack, M. M., Kobielush, B., Ryu, D., Tittlemier, S., Wilson, W. W., Abbas, H. K., Abel, S., Harrison, G., Miller, J. D., Shier, W. T., & Weaver, G. Don occurrence in grains: A North American perspective. *Cereal Foods World* **60**, 32–56 (2015).
8. Alberta Fusarium Action Committee. *Alberta Fusarium Graminearum Management Plan.*

Edmonton: Alberta Agriculture and Rural Development. https://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/agdex5210/$file/110_632-3.pdf?OpenElement (2012).

9. Xu, M. et al. Combatting Fusarium head blight: advances in molecular interactions between Fusarium graminearum and wheat. *Phytopathology Research* **4**, 37 (2022).
10. Newly Developed Prairie-Wide Fusarium Head Blight (FHB) Risk Maps. *Manitoba Crop Alliance* https://mbcropalliance.ca/directory/production-resources/prairie-wide-fusarium-head-blight-risk-maps/ (2024).
11. Trucksess, M. W., Bao, L., Weaver, C. M. & White, K. D. Determination of deoxynivalenol in processed foods. *Journal of AOAC International* **93**, 1236–1242 (2010).
12. Recharla, N., Park, S., Kim, M., Kim, B. & Jeong, J. Y. Protective effects of biological feed additives on gut microbiota and the health of pigs exposed to deoxynivalenol: a review. *Journal of Animal Science and Technology* **64**, 640–653 (2022).
13. Muñoz-Solano, B., Lizarraga Pérez, E. & González-Peñas, E. Monitoring mycotoxin exposure in food-producing animals (cattle, pig, poultry, and sheep). *Toxins* **16**, 218 (2024).
14. Pinton, P. & Oswald, I. P. Effect of deoxynivalenol and other Type B trichothecenes on the intestine: a review. *Toxins* **6**, 1615–1643 (2014).
15. EFSA Panel on Contaminants in the Food Chain (CONTAM) *et al.* Risks to human and animal health related to the presence of deoxynivalenol and its acetylated and modified forms in food and feed. *EFSA J* **15**, e04718 (2017).
16. Pestka, J. J. Mechanisms of deoxynivalenol-induced gene expression and apoptosis. *Food Additives & Contaminants: Part A – Chemistry, Analysis, Control, Exposure & Risk*

Assessment **25**, 1128–1140 (2008).

17. Shank, R. A., Foroud, N. A., Hazendonk, P., Eudes, F. & Blackwell, B. A. Current and future experimental strategies for structural analysis of trichothecene mycotoxins--a prospectus. *Toxins* **3**, 1518–1553 (2011).
18. Verheecke-Vaessen, C. et al. Interacting climate change environmental factors effects on *Fusarium langsethiae* growth, expression of Tri genes and T-2/HT-2 mycotoxin production on oat-based media and in stored oats. *Fungal Biology* **123**, 618–624 (2019).
19. Janik, E. et al. T-2 toxin-the most toxic trichothecene mycotoxin: metabolism, toxicity, and decontamination strategies. *Molecules* **26**, 6868 (2021).
20. Kaukoranta, T., Hietaniemi, V., Rämö, S., Koivisto, T. & Parikka, P. Contrasting responses of T-2, HT-2 and DON mycotoxins and *Fusarium* species in oat to climate, weather, tillage and cereal intensity. *European Journal of Plant Pathology* **155**, 93–110 (2019).
21. Wegulo, S. N. Factors influencing deoxynivalenol accumulation in small grain cereals. *Toxins* **4**, 1157–1180 (2012).
22. Gomes, A. R., Varela, C. L., Tavares-da-Silva, E. J. & Roleira, F. M. F. Epoxide containing molecules: A good or a bad drug design approach. *European Journal of Medicinal Chemistry* **201**, 112327 (2020).
23. Li, Y. et al. Study of the interaction of deoxynivalenol with human serum albumin by spectroscopic technique and molecular modelling. *Food Additives & Contaminants: Part A – Chemistry, Analysis, Control, Exposure & Risk Assessment* **30**, 356–364 (2013).
24. Philippe, S., Saulnier, L. & Guillon, F. Arabinoxylan and (1→3),(1→4)-beta-glucan deposition in cell walls during wheat endosperm development. *Planta* **224**, 449–461 (2006).

25. Lancova, K. et al. Fate of trichothecene mycotoxins during the processing: milling and baking. *Food Additives & Contaminants: Part A – Chemistry, Analysis, Control, Exposure & Risk Assessment* **25**, 650–659 (2008).

26. Cundliffe, E., Cannon, M. & Davies, J. Mechanism of inhibition of eukaryotic protein synthesis by trichothecene fungal toxins. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 30–34 (1974).

27. Wang, W. et al. The ribosome-binding mode of trichothecene mycotoxins rationalizes their structure-activity relationships. *International Journal of Molecular Sciences* **22**, 1604 (2021).

28. Zhou, H.-R., Jia, Q. & Pestka, J. J. Ribotoxic stress response to the trichothecene deoxynivalenol in the macrophage involves the SRC family kinase Hck. *Toxicological Sci*ences **85**, 916–926 (2005).

29. Armer, V. J., Urban, M., Ashfield, T., Deeks, M. J. & Hammond-Kosack, K. E. The trichothecene mycotoxin deoxynivalenol facilitates cell-to-cell invasion during wheat-tissue colonization by *Fusarium graminearum*. *Molecular Plant Pathology* **25**, e13485 (2024).

30. Park, J. et al. Transcriptome analyses to understand effects of the *Fusarium* deoxynivalenol and nivalenol mycotoxins on *Escherichia coli*. *Journal of Biotechnology* **192**, 231–239 (2014).

31. Xie, C. et al. Detoxification of deoxynivalenol and its derivatives in *Fusarium* contaminated wheat through soaking and germination. *Food Control* **155**, 110084 (2024).

32. Kushiro, M. Effects of milling and cooking processes on the deoxynivalenol content in wheat. *International Journal of Molecular Science* **9**, 2127–2145 (2008).

33. Wolf, C. E. & Bullerman, L. B. Heat and pH alter the concentration of deoxynivalenol in an aqueous environment. *Journal of Food Protection* **61**, 365–367 (1998).

34. Yener, S. & Köksel, H. Effects of washing and drying applications on deoxynivalenol and zearalenone levels in wheat. *World Mycotoxin Journal* **6**, 335–341 (2013).

35. Yao, Y. & Long, M. The biological detoxification of deoxynivalenol: A review. *Food amd Chemical Toxicology* **145**, 111649 (2020).

36. Zhu, Y., Hassan, Y. I., Lepp, D., Shao, S. & Zhou, T. Strategies and methodologies for developing microbial detoxification systems to mitigate mycotoxins. *Toxins* **9**, 130 (2017).

37. EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP) *et al.* Safety and efficacy of microorganism DSM 11798 as a technological additive for all avian species. *EFSA J* **15**, e04676 (2017).

38. Maresca, M. From the gut to the brain: journey and pathophysiological effects of the food-associated trichothecene mycotoxin deoxynivalenol. *Toxins* **5**, 784–820 (2013).

39. Sayyari, A. et al. Effects and biotransformation of the mycotoxin deoxynivalenol in growing pigs fed with naturally contaminated pelleted grains with and without the addition of *Coriobacteriaceum* DSM 11798. *Food Additives & Contaminants: Part A – Chemistry, Analysis, Control, Exposure & Risk Assessment* **35**, 1394–1409 (2018).

40. Debevere, S. et al. Evaluation of the efficacy of mycotoxin modifiers and mycotoxin binders by using an in vitro rumen model as a first screening tool. *Toxins* **12**, 405 (2020).

41. Carere, J., Hassan, Y. I., Lepp, D. & Zhou, T. The identification of DepB: An enzyme responsible for the final detoxification step in the deoxynivalenol epimerization pathway in 17-2-E-8. *Frontiers in Microbiology* **9**, 1573 (2018).

42. Xu, H. et al. Microbial detoxification of mycotoxins in food and feed. *Critical Reviews in Food Science and Nutrition* **62**, 4951–4969 (2022).

43. Yang, J. et al. Enzymatic degradation of deoxynivalenol with the engineered detoxification enzyme Fhb7. *JACS Au* **4**, 619–634 (2024).

44. Buyel, J. F. Plant molecular farming - integration and exploitation of side streams to achieve sustainable biomanufacturing. *Frontiers in Plant Science* **9**, 1893 (2018).

45. Singh, J. & Mehta, A. Rapid and sensitive detection of mycotoxins by advanced and emerging analytical methods: A review. *Food Science & Nutrition* **8**, 2183–2204 (2020).

46. Agriopoulou, S., Stamatelopoulou, E. & Varzakas, T. Advances in analysis and detection of major mycotoxins in foods. *Foods* **9**, 518 (2020).

47. Naresh, V. & Lee, N. A review on biosensors and recent development of nanostructured materials-enabled biosensors. *Sensors* **21**, 1109 (2021).

48. Colomer-Farrarons, J., Ll., P., Ivon, A. & Samitier, J. Portable bio-devices: design of electrochemical instruments from miniaturized to implantable devices. in *New Perspectives in Biosensors Technology and Applications* (InTech, 2011). doi:10.5772/17212.

49. Close, D. M., Ripp, S. & Sayler, G. S. Reporter proteins in whole-cell optical bioreporter detection systems, biosensor integrations, and biosensing applications. *Sensors* **9**, 9147–9174 (2009).

50. Moraskie, M. et al. Microbial whole-cell biosensors: Current applications, challenges, and future perspectives. *Biosensors and Bioelectronics* **191**, 113359 (2021).

51. Stocker, J. et al. Development of a set of simple bacterial biosensors for quantitative and rapid measurements of arsenite and arsenate in potable water. *Environmental Science &*

Technology **37**, 4743–4750 (2003).

52. Zaslaver, A. et al. Just-in-time transcription program in metabolic pathways. *Nature Genetics* **36**, 486–491 (2004).
53. Mahr, R., von Boeselager, R. F., Wiechert, J. & Frunzke, J. Screening of an *Escherichia coli* promoter library for a phenylalanine biosensor. *Applied Microbiology and Biotechnology* **100**, 6739–6753 (2016).
54. Bajar, B. T. et al. Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Scientific Reports* **6**, 20889 (2016).
55. Sweet, C. R. Expression of recombinant proteins from lac promoters. *Methods in Molecular Biology* **235**, 277–288 (2003).
56. Keseler, I. M. et al. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Research* **45**, D543–D550 (2017).
57. Izutsu, K., Wada, A. & Wada, C. Expression of ribosome modulation factor (RMF) in *Escherichia coli* requires ppGpp. *Genes to Cells* **6**, 665–676 (2001).
58. Pierron, A. et al. Microbial biotransformation of DON: molecular basis for reduced toxicity. *Scientific Reports* **6**, 29105 (2016).
59. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and Molecular Biology Reviews* **67**, 593–656 (2003).
60. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology* **14**, 638–650 (2016).
61. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* **13**, 227–232 (2012).
62. Rodriguez, E. A. et al. The growing and glowing toolbox of fluorescent and photoactive

proteins. *Trends in Biochemical Sciences* **42**, 111–129 (2017).

63. Joshi, K. M. et al. Advancing protein biosensors: redefining detection through innovations in materials, mechanisms, and applications for precision medicine and global diagnostics. *RSC Advances* **15**, 11523–11536 (2025).

64. Rai, P. et al. Protease detection in the biosensor era: A review. *Biosensors and Bioelectronics* **244**, 115788 (2024).

65. Smith, D. D., Girodat, D., Abbott, D. W. & Wieden, H.-J. Construction of a highly selective and sensitive carbohydrate-detecting biosensor utilizing Computational Identification of Non-disruptive Conjugation sites (CINC) for flexible and streamlined biosensor design. *Biosensors Bioelectron*ics **200**, 113899 (2022).

66. Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. *Neuron* **99**, 1129–1143 (2018).

67. Garvey, G. S. & Rayment, I. *Crystal structure of Fusarium graminearum TRI101 complexed with coenzyme A and deoxynivalenol*. Protein Data Bank (2007).
PDB ID: 3B2S. https://doi.org/10.2210/pdb3b2s/pdb

68. McCormick, S. P., Alexander, N. J., Trapp, S. E. & Hohn, T. M. Disruption of TRI101, the gene encoding trichothecene 3-O-acetyltransferase, from *Fusarium sporotrichioides*. *Applied Environmental Microbiology* **65**, 5252–5256 (1999).

69. Khatibi, P. A. et al. Bioprospecting for trichothecene 3-O-acetyltransferases in the fungal genus *Fusarium* yields functional enzymes with different abilities to modify the mycotoxin deoxynivalenol. *Applied and Environmental Microbiology* **77**, 1162–1170 (2011).

70. Schmeitzl, C. et al. Identification and characterization of carboxylesterases from

Brachypodium distachyon deacetylating trichothecene mycotoxins. Toxins **8**, 6 (2015).

71. Hospital, A., Goñi, J. R., Orozco, M. & Gelpí, J. L. Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry.* **8**, 37–47 (2015).
72. Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling* **51**, 2778–2786 (2011).
73. Case, D. A. et al. AmberTools. *Journal of Chemical Information and Modeling* **63**, 6183–6191 (2023).
74. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **25**, 1157–1174 (2004).
75. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926–935 (1983).
76. Case, D. A. et al. The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **26**, 1668–1688 (2005).
77. Roe, D. R. & Cheatham, T. E., 3rd. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation* **9**, 3084–3095 (2013).
78. Ormeño, F. & General, I. J. Convergence and equilibrium in molecular dynamics simulations. *Communications Chemistry* **7**, 26 (2024).
79. Ping, J., Hao, P., Li, Y.-X. & Wang, J.-F. Molecular dynamics studies on the conformational

transitions of adenylate kinase: a computational evidence for the conformational selection mechanism. *BioMed Research International* **2013**, 628536 (2013).

80. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research* **44**, W344–50 (2016).
81. Baldwin, A. D. & Kiick, K. L. Tunable degradation of maleimide-thiol adducts in reducing environments. *Bioconjugate Chemistry* **22**, 1946–1953 (2011).
82. Amadei, M. et al. Genetic incorporation of dansylalanine in human ferroportin to probe the alternating access mechanism of iron transport. *International Journal of Molecular Sciences* **24**, 11919 (2023).
83. Abdel-Salam, E. T., Anwar, Z. M., Ibrahim, A. Z. & Musatfa, H. M. A rapid fluorescence quenching assay for total levothyroxine quantification in pharmaceutical and supratherapeutic serum samples. *Journal of Fluorescence* **35**, 2025 (2025).
84. Alexandrov, A. I., Mileni, M., Chien, E. Y. T., Hanson, M. A. & Stevens, R. C. Microscale fluorescent thermal stability assay for membrane proteins. *Structure* **16**, 351–359 (2008).
85. Brune, M., Hunter, J. L., Corrie, J. E. & Webb, M. R. Direct, real-time measurement of rapid inorganic phosphate release using a novel fluorescent probe and its application to actomyosin subfragment 1 ATPase. *Biochemistry* **33**, 8262–8271 (1994).
86. Kollman, P. A. et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of Chemical Research* **33**, 889–897 (2000).
87. Stapley, B. J. & Creamer, T. P. A survey of left-handed polyproline II helices. *Protein Sci*ence **8**, 587–595 (1999).

88. Richardson, J. S. & Richardson, D. C. Amino acid preferences for specific locations at the ends of alpha helices. *Science* **240**, 1648–1652 (1988).

89. Richardson, J. S. & Richardson, D. C. The de novo design of protein structures. *Trends in Biochemical Sciences* **14**, 304–309 (1989).

90. Salah Ud-Din, A. I. M., Tikhomirova, A. & Roujeinikova, A. structure and functional diversity of GCN5-related N-acetyltransferases (GNAT). *International Journal of Molecular Sciences* **17**, 1018 (2016).

91. Bornot, A., Etchebest, C. & de Brevern, A. G. Predicting protein flexibility through the prediction of local structures. *Proteins* **79**, 839–852 (2011).

92. Miljus, T., Sykes, D. A., Harwood, C. R., Vuckovic, Z. & Veprintsev, D. B. GPCR solubilization and quality control. *Methods in Molecular Biology* **2127**, 105–127 (2020).

93. Klymchenko, A. S. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Accounts of Chemical Researchs* **50**, 366–375 (2017).

94. Lemmon, C. A., Ohashi, T. & Erickson, H. P. Probing the folded state of fibronectin type III domains in stretched fibrils by measuring buried cysteine accessibility. *Journal of Biological Chemistry* **286**, 26375–26382 (2011).

95. BioRender. BioRender.com. https://biorender.com. 2025.

96. Benchling. Benchling.com. https://benchling.com. 2025

Appendix A: Supplementary Figures

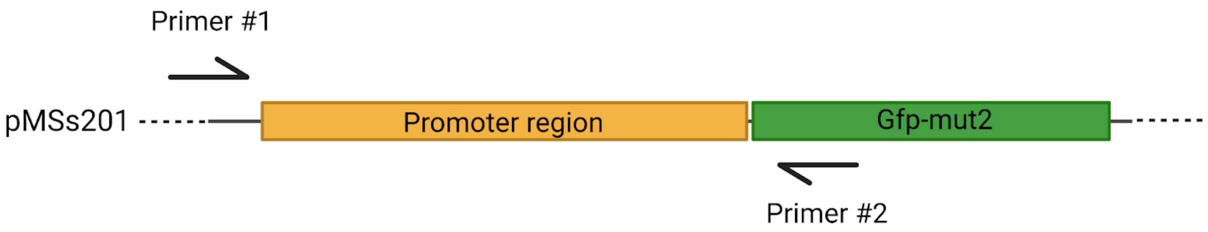


Figure A1. Primer placement for amplicon generation using plasmid DNA extract from final population obtained from cell-sorting. Created with Biorender.com.[95]

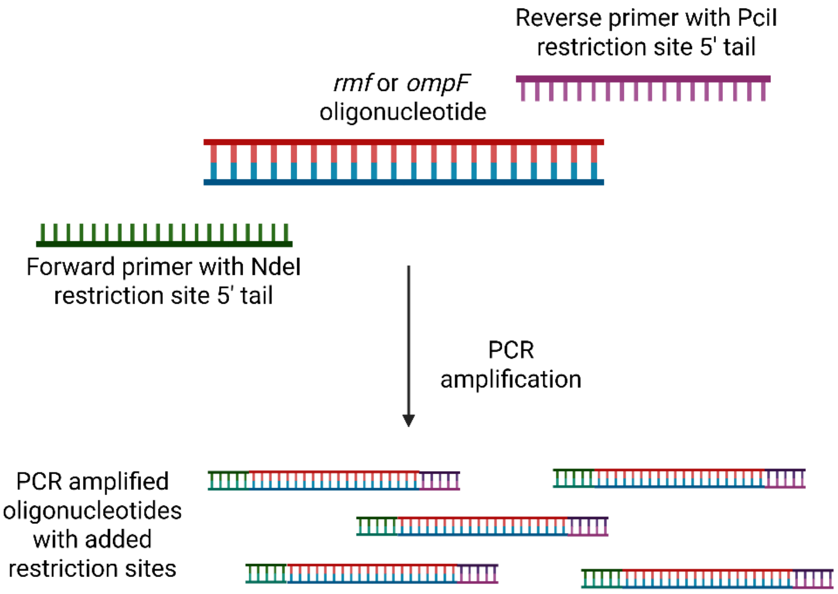


Figure A2. Conceptual diagram demonstrating the addition of restriction sites to both *rmf* and *ompF* oligonucleotides prior to cloning within the pNCS-mClover3 plasmid. Created with Biorender.com.[95]

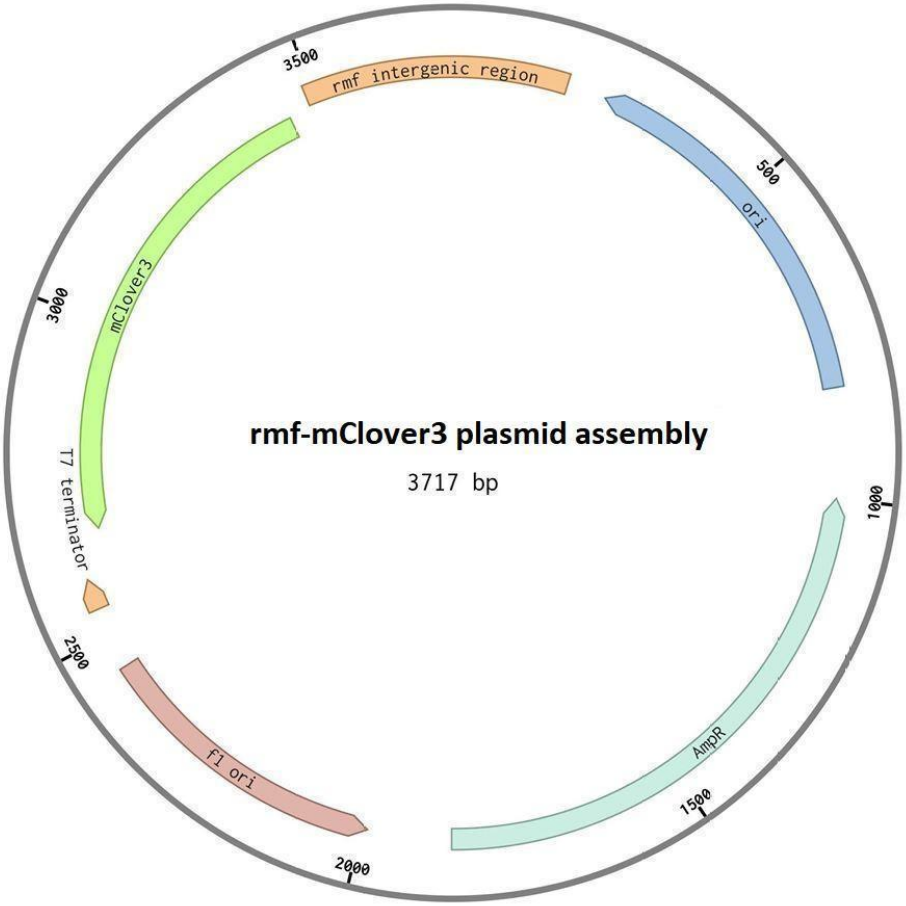
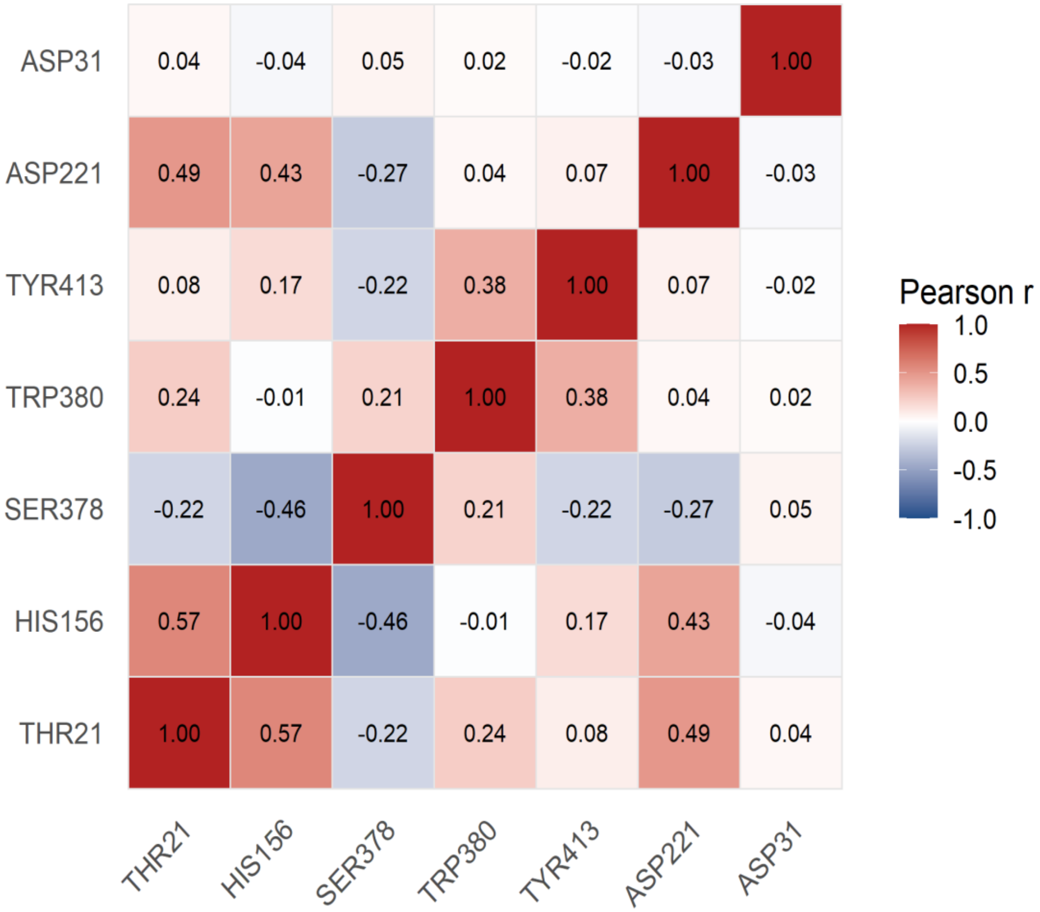


Figure A3. Plasmid map for the assembled *rmf*-mClover3 plasmid. The pNCS-mClover plasmid (Addgene) was used as a backbone with the rmf intergenic region placed upstream of the mClover3 gene and replacing the T7 promoter. Created with Benchling.[96]



	THR21	HIS156	SER378	TRP380	TYR413	ASP221	ASP31
ASP31	0.04	-0.04	0.05	0.02	-0.02	-0.03	1.00
ASP221	0.49	0.43	-0.27	0.04	0.07	1.00	-0.03
TYR413	0.08	0.17	-0.22	0.38	1.00	0.07	-0.02
TRP380	0.24	-0.01	0.21	1.00	0.38	0.04	0.02
SER378	-0.22	-0.46	1.00	0.21	-0.22	-0.27	0.05
HIS156	0.57	1.00	-0.46	-0.01	0.17	0.43	-0.04
THR21	1.00	0.57	-0.22	0.24	0.08	0.49	0.04

Figure A4. Pearson correlation coefficients were computed from the minimum distances to DON across the TRI101–DON trajectory (as summarized in Table 2 and Figure 10). For each residue, the minimum-distance time series (extracted from the TRI101-DON trajectory) was correlated pairwise with that of the other residues on a frame-by-frame basis. Asp31 is included in the calculations as a non-binding pocket control.

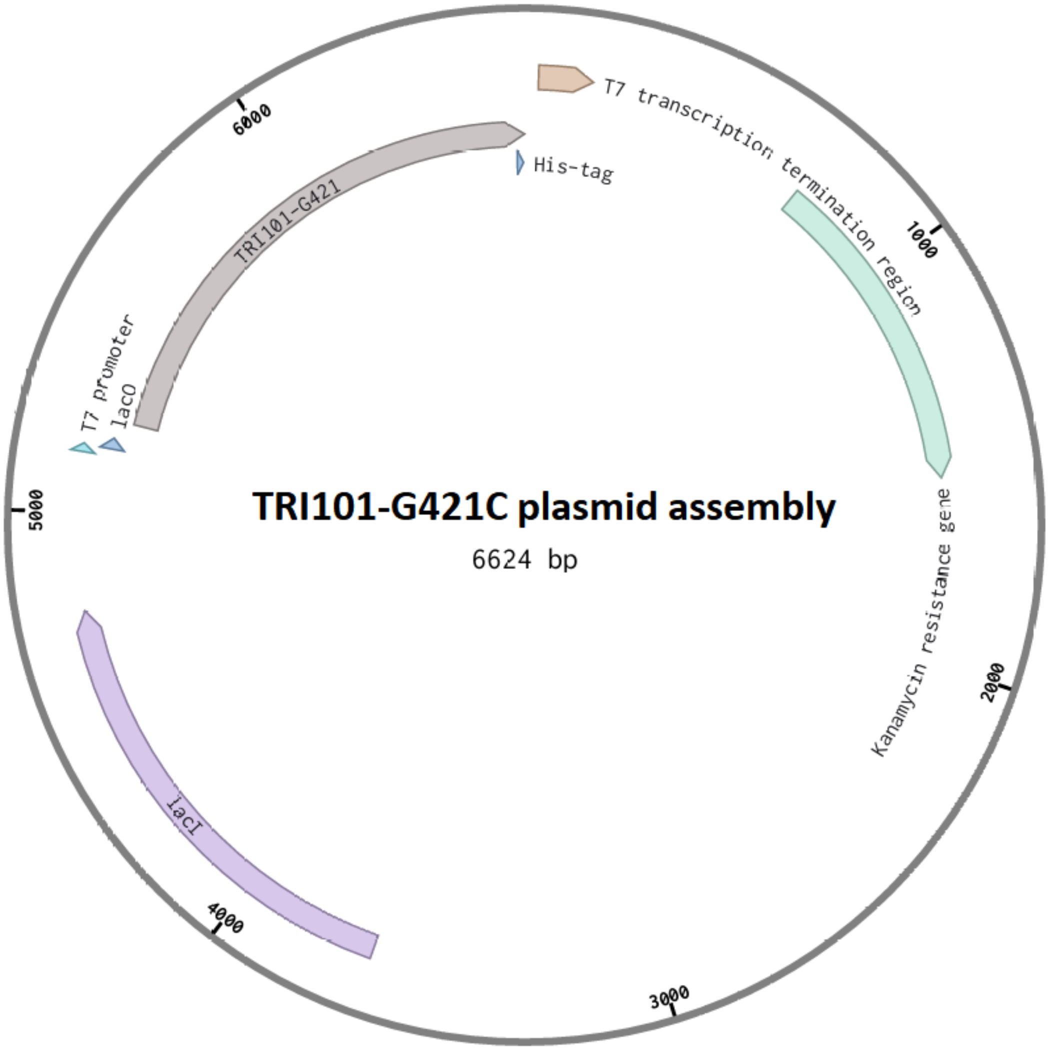


Figure A5. Plasmid map of the TRI101-G421C expression plasmid pET-28a(+)-TRI101-G421C. The TRI101 ORF was cloned into the pET-28a(+) vector using the multiple cloning site restriction sites BamHI and XhoI. A C-terminal His-tag was added for purification purposes. Created with Benchling.[96]

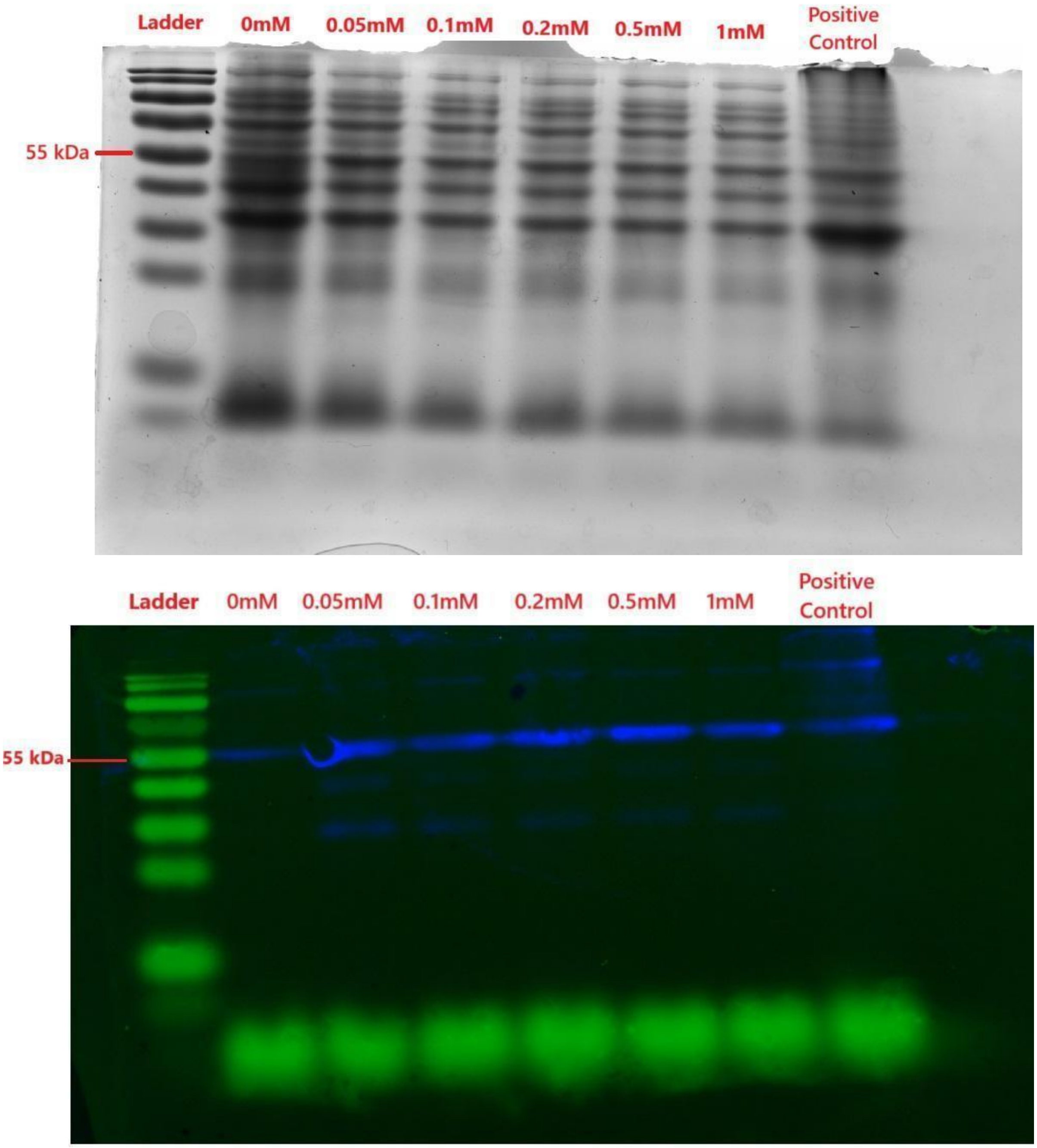


Figure A6. Expression of TRI101-G421C, as visualized by a standard SDS-PAGE gel (12%) with coomassie blue stain (top) and by Western blot with an anti-His antibody and chemiluminescent staining (bottom). Lanes are labeled by the concentration of IPTG used during culture induction, with no IPTG added for the positive control. The ladder is Thermo Scientific PageRuler Plus Prestained Protein Ladder and the 55 kDA ladder band (approximate size of recombinant TRI101-G421C) is indicated.

Appendix B: Supplementary Tables

Table B1. List of primers used in rmf-mClover3 and ompF-mClover3 plasmid assemblies and for confirmation of proper assembly.

Primer	**Target**	**Assay type**	**Sequence (5'-3')**	**Function**
rmf-NsiI.F	*rmf* oligo	endpoint PCR	TTTTTTATGCATTTAAAGTGCTGGCCGGTG	Amplify and add restriction sites to the *rmf* oligonucleotide
rmf-PciI.F	*rmf* oligo	endpoint PCR	TTTTTTACATGTTTAAAGTGCTGGCCGGTG	
rmf-NdeI.R	*rmf* oligo	endpoint PCR	TTTTTTCATATGCATTTCTTTTGAGCGTCCGG	
ompF-NsiI.F	*ompF* oligo	endpoint PCR	TTTTTTATGCATTCGGTCTTGGTTTTGAACGTC	Amplify and add restriction sites to the *ompF* oligonucleotide
ompF-PciI.F	*ompF* oligo	endpoint PCR	TTTTTTACATGTTCGGTCTTGGTTTTGAACGTC	
ompF-NdeI.R	*ompF* oligo	endpoint PCR	TTTTTTCATATGTCTACTTTGTTGCCATCTTTGT	
rmf.F	*rmf-mClover3* plasmid	endpoint PCR	ATTGACGGCAGTTATGATTCG	Check for proper insertion of *rmf* oligo upstream of mClover3 gene
ompF.F	*ompF-mClover3* plasmid	endpoint PCR	GGCATTCTGGATGTCTGAAAG	Check for proper insertion of *ompF* oligo upstream of mClover3 gene
EGFP-mClover3.R	*rmf-mClover3* and *ompF-mClover3* plasmids	endpoint PCR	GGTGCAGATGAACTTCAGG	Check for proper insertion of *rmf* and *ompF* oligos upstream of mClover3 gene