CHARACTERIZATION AND RATIONAL DESIGN OF BIOMOLECULAR SENSORS USING MOLECULAR DYNAMICS SIMULATIONS

DUSTIN D. SMITH
Master of Science, University of Lethbridge, 2017

A Thesis
in Partial Fulfillment of the
Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

BIOMOLECULAR SCIENCE

Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

CHARACTERIZATION AND RATIONAL DESIGN OF BIOMOLECULAR SENSORS USING MOLECULAR DYNAMICS SIMULATIONS

DUSTIN D. SMITH

Date of Defence: April 5, 2023

Dr. Hans-Joachim Wieden	Professor	Ph.D.
Dr. Trushar Patel	Associate Professor	Ph.D.
Thesis Co-Supervisors		
Dr. Ute Kothe	Professor	Ph.D.
Thesis Examination Committee Member		
Dr. D. Wade Abbott	Research Scientist	Ph.D.
Thesis Examination Committee Member		
Agriculture and Agri-Food Canada		
Dr. Michael Gerken	Professor	Ph.D.
Internal External Examiner		
Department of Chemistry and Biochemistry.		
University of Lethbridge		
Dr. Alisdair Boraston	Professor	Ph.D.
External Examiner		
Department of Biochemistry and Microbiology		
University of Victoria		
Dr. Paul Hayes	Professor	Ph.D.
Chair, Thesis Examination Committee		

ABSTRACT

Biosensors are analytical devices that use biological components to detect and report the presence of a target molecule. Although useful for a broad range of purposes, biosensors are conventionally designed using laborious methods limiting development to a small number of applications with large commercial value. To overcome this limitation, computational approaches are needed to streamline rational design of protein-fluorophore conjugate-type biosensors. Here, I report and iteratively improve such a biosensor development pipeline based on protein molecular dynamics simulations, exploiting underlying dynamic properties of proteins for biosensor design. As proof-of-concept, I report the construction of several carbohydrate-detecting biosensors which are advantageous compared to previous carbohydrate detection methods, and I use these biomolecular tools to characterize several Carbohydrate Active Enzymes (CAZymes). This research highlights how underlying dynamic features of proteins can be utilized for the design and mechanistic interpretation of biomolecular function in a broad range of applications.

CONTRIBUTIONS OF AUTHORS

The contents of Chapter 2 have previously been published as: Smith, D.D., Girodat, D., Abbott, D.W., and Wieden, H.-J. (2022) Construction of a highly specific and selective carbohydrate-detecting biosensor utilizing Computational Identification of Non-disruptive Conjugation sites (CINC) for flexible and streamlined biosensor design. *Biosensors and Bioelectronics*, 200, 113899. I performed all biochemical characterization in this study, analyzed molecular dynamics data with Dr. Dylan Girodat, and wrote the first draft of the manuscript.

The contents of Chapter 3 have previously been published as: Smith, D.D., King, J.K., Abbott, D.W., and Wieden, H.-J. (2022). Development of a real-time pectic oligosaccharide-detecting biosensor using the rapid and flexible Computational Identification of Non-disruptive Conjugation sites (CINC) biosensor design platform. *Sensors*, 22(3), 948. I wrote the first draft of this manuscript, performed all computational and biochemical characterization therein, and supervised an undergraduate student (Josh King) who performed preliminary assays to characterize the reported biosensors as part of his Independent Study courses (BCHM 2990, BCHM 3990).

Chapter 4 encompasses a currently unpublished format-neutral manuscript where I performed all wet lab and computational experiments reported and wrote the first draft of the manuscript. I wrote Chapters 1 and 5 of this thesis and further

details outlining the contributions of my co-authors to the various manuscripts included in this thesis are highlighted in the respective prefatory sections of these Chapters.

ACKNOWLEDGEMENTS

Thanks to Dr. Hans-Joachim Wieden for taking me in as a student over a decade ago. I've learned a lot working in your lab and appreciate the mentorship you've provided over the years.

I would also like to thank Dr. Trushar Patel for becoming my thesis co-supervisor towards the end of my degree, and for providing me with useful feedback over the course of my program. I would like to acknowledge my committee member and collaborator Dr. D. Wade Abbott for inspiring some foundational aspects of my thesis work related to carbohydrate detection. As well, I would like to thank Dr. Ute Kothe for her useful feedback as my committee member and for the role she and her team had in training me over the years. More recently, I would like to thank my thesis External Examiner Dr. Alisdair Boraston, my Internal External Examiner Dr. Michael Gerken, and my thesis Examination Chair Dr. Paul Hayes for their respective roles in my PhD Defence.

I would like to thank the various students I have worked through my program who provided me with the opportunity to develop my teaching skills. I would also like to thank Davinder, Dylan, Harland, and Luc who have collaborated with me on research projects and/or our patent submission during my time as a PhD student. I would also like to thank the members of the Wieden, Kothe, Patel, and Abbott labs for their constructive feedback and memories over the years. As well, I would

like to thank my friends and family for their support through my graduate program. I also would like to recognize the various funding sources that have supported me throughout my graduate studies including Alberta Innovates, the University of Lethbridge, the Alberta GreenSTEM Program, and Mitacs.

Last, I would like to thank my wonderful wife for her support over the years, and for her ability to still always make me “two dots and a U” :).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA: Auxiliary Activities
ABC: ATP-binding cassette
ARRTI: Alberta RNA Research and Training Institute
CAZome: genome-encoded carbohydrate active enzymes
CAZyme: Carbohydrate Active Enzyme
CINC: Computational Identification of Non-disruptive Conjugation sites
CPM: 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin
DACM: *N*-(7-dimethylamino-4-methylcoumarin-3-yl)maleimide
Dansyl: *N*-[2-(dansylamino)ethyl]maleimide
digalUA: digalacturonic acid
DMF: dimethylformamide
DMSO: dimethyl sulfoxide
E. coli: *Escherichia coli*
EF-Tu: Elongation Factor-Thermo unstable
Fluorescein: fluorescein-5-maleimide
FRET: Förster Resonance Energy Transfer
FW: Formula weight
GDP: Guanosine diphosphate
GH: Glycoside Hydrolase
GTP: Guanosine triphosphate
IMAC: Immobilized Medal Affinity Chromatography
IPTG: isopropyl β-$_D$-1-thiogalactopyranoside
k_{app}: apparent rate
K_D: dissociation rate constant
k_{on}: association rate constant
LB media: Luria-Bertani media
M3: Maltotriose
M7: Maltoheptaose
MD: Molecular Dynamics
MDCC: 7-diethylamino-3-[*N*-(2-maleimidoethyl)carbamoyl]coumarin
MOS: Maltooligosaccharide
OD_{600}: Optical density at 600 nm
PDB: Protein Data Bank
PGA: polygalacturonic acid
PL: Polysaccharide Lyase
PMT: Photomultiplier tube
PN: Product Number
Pyrene: *N*-(1-pyrene)maleimide
Rhodamine red: Rhodamine Red C2 maleimide
RMSD: root-mean squared deviation
RMSF: root-mean squared fluctuation
SBP: Solute-binding protein
SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis
Tx/Tl: Transcription/ Translation

trigalUA: trigalacturonic acid
unsatdigalUA: 4,5-unsaturated digalacturonic acid
Y. enterocolitica: *Yersinia enterocolitica*

CHAPTER 1: LITERATURE REVIEW

1.1. Preface:

My doctoral thesis focuses on the use of computational methods to both understand and utilize the underlying dynamics properties of proteins for bioengineering purposes. Specifically, a substantial portion of my thesis research is dedicated to the use of protein molecular dynamics simulations for the design of protein-fluorophore conjugate-based biosensors. To set the stage for this research, Chapter 1 highlights current methods for biosensor design and their associated applications, as well as a brief overview on protein molecular dynamics simulations and how they may be used to forward the field of biosensor design and biomolecular sensing.

1.2. Biosensors and their applications:

1.2.1 Biosensor Overview. Biosensors are devices that utilize biological components to recognize a molecule of interest and give a detectable output (Mehrotra 2016). The two main components of a biosensor are a recognition element (which displays specificity for the target molecule) and an output element (which produces a detectable signal upon target detection). The types of components used as the aforementioned biosensor elements are vast. Some of the most successful biosensing tools are familiar even to the general public. For example, rapid tests used for the detection of SARS-CoV2 during the COVID19 pandemic (Peeling et al. 2021) or blood glucose monitors used by individuals

affected with diabetes (Heller and Feldman 2008). In the subsequent sections, I will explore three major types of biosensors categorized based on their mechanism of target recognition. These biosensor types are transcription/translation (Tx/Tl) based biosensors, enzyme-based biosensors, and affinity-based biosensors. I will also examine the opportunities for biosensor design that I have identified as part of my thesis work, and my efforts to improve the biosensor development process.

1.2.2. Tx/Tl-based biosensors. Tx/Tl-based biosensors utilize the transcription and/or translational machinery to detect and report the presence of a molecule of interest, often utilizing binding of a target to a regulatory element to alter transcriptional or translational output of a reporter. Output elements used in Tx/Tl-based biosensors have included RNA light-up aptamers, fluorescent proteins, or other colorimetric systems such as LacZ-based reporters (Mehrotra 2016; Ouellet 2016). Tx/Tl-based biosensors have been utilized both *in vivo* and *ex vivo* and have been used for environmental and biomedical applications (Harms et al. 2006; Pardee et al. 2014; Pardee et al. 2016; Shin 2011). For example, a paper-embedded *ex vivo* translation-based biosensor has been used to detect Zika Virus RNA (Pardee et al. 2014; Pardee et al. 2016), and transcription-based biosensors have been used to produce light up RNA aptamers detectable in the presence of their cognate fluorogen (Ying et al. 2018). An advantage of Tx/Tl-based biosensors is their ability to generate high signal-to-noise outputs via signal amplification. In other words, one stimulus can trigger production of multiple rounds of transcription and/or translation to increase an output signal. When a molecule of interest is

capable of influencing transcription or translation levels either directly or indirectly, utilization of Tx/Tl-based biosensors can be a powerful strategy.

1.2.3 Enzyme-based biosensors. Enzyme-based biosensors exploit the catalytic abilities of a biomolecule to generate a detectable signal upon molecule recognition (Ispas et al. 2012). For example, commercially available blood-glucose monitors use glucose oxidase to oxidize glucose, producing hydrogen peroxide that can be detected via an electrode (Heller and Feldman 2008; Mehrotra 2016; Wang 2008). Blood-glucose monitors are widely used by individuals affected by diabetes, and have an estimated market value of $8.8 billion (USD) (Hughes 2009). Prior to the COVID-19 pandemic, blood-glucose monitors represented the largest market value of any biosensor as they serve the need to continually monitor blood-glucose levels and can be used by non-specialists. Recent research has focused on expanding the scope of portable monitoring for physiologically relevant markers based on the principles of the blood-glucose detector. For example, some projects aim to reengineer glucose dehydrogenase to recognize other molecules of interest as allosteric activators of the enzyme (Guo et al. 2015; Guo et al. 2016). In addition to glucose oxidase, other oxidases and dehydrogenases have been used to develop biosensors for clinically relevant molecules, such as cholesterol and ethanol (Rocchitta et al. 2016). Although oxidase and dehydrogenase-utilizing biosensors have their advantages, they are currently limited by the availability of an enzyme able to produce a detectable molecule in response to a target (i.e. availability of an oxidase or dehydrogenase for a certain substrate). The

aforementioned limitation has in recent years pushed research activity towards affinity-based biosensors, particularly during the COVID-19 pandemic where affinity-based biosensors for SARS-CoV-2 dominated the biosensing market.

1.2.4 Affinity-based biosensors. Affinity-based biosensors utilize biological components to couple direct target binding to an output signal. For example, detection of antigens has been achieved via antibody-based methods with colorimetric, fluorescent, or electro-chemiluminescent output (e.g. pregnancy tests (Posthuma-Trumpie et al. 2009), COVID-19 rapid tests (Peeling et al. 2021), and cancer marker-detectors (Juzgado et al. 2017)). DNA/RNA light-up aptamers have also been used in affinity-based biosensors, for example, in the detection of S-adenosylmethionine (Ouellet 2016), theophylline (Endoh et al. 2022), or ionized metals (Wang et al. 2016). Other affinity-based sensors include analyte-binding proteins conjugated with fluorescent dyes (e.g. anion or cation-detecting biosensors (Brune et al. 1994; De Lorimier et al. 2002; Okoh et al. 2006; Smith et al. 2017; Solscheid et al. 2015), ADP-detecting biosensor (Kunzelmann and Webb 2010), amino acid-detecting biosensors (De Lorimier et al. 2002), carbohydrate-detecting biosensors (De Lorimier et al. 2002; Marvin et al. 2011), or a dipeptide-detecting biosensor (De Lorimier et al. 2002)). Biosensors also have been constructed via analyte-binding proteins expressed as fusions with fluorescent protein derivatives (e.g. ATP:ADP-detecting biosensor (Berg et al. 2009), Ca^{2+}-detecting biosensor (Baird et al. 1999), hydrogen peroxide-detecting biosensor (Belousov et al. 2006), and GTP-detecting biosensors (Bianchi-Smiraglia et al.

2017)). Unlike the previously mentioned Tx/Tl or enzyme-based biosensor systems, some affinity-based biosensors can detect molecules of interest in real-time which is advantageous in situations where rapid detection is required. For example, phosphate-detecting biosensors capable of real-time detection have been used *in vitro* to study systems where phosphate release is integral to underlying functional cycles (Brune et al. 1994; Kothe and Rodnina 2006; Savelsbergh et al. 2005; Smith et al. 2017). Some affinity-based biosensors have also been genetically encoded and fully assembled in the cell, allowing for live-cell imaging of analytes (e.g. as in (Bianchi-Smiraglia et al. 2017; Marvin et al. 2011)). Similar to the previously discussed biosensor categories, the development of affinity-based biosensors is largely limited by the availability of an appropriate scaffold with the desired target specificity, as well as our ability to engineer a biosensing system based on said scaffold. These engineering endeavors are however currently laborious and a streamlined approach for the development of affinity-based biosensors is required.

1.2.5 Engineering affinity-based biosensors capable of real-time analyte detection. Affinity-based biosensors are often constructed using a protein found in nature as a recognition element coupled to an output-signal generating element. The protein recognition elements utilized from nature are extensive, and the engineering challenge in coupling analyte-binding by a recognition element to a signal from an output element is not trivial (Newman et al. 2011; Tamura and Hamachi 2014). A commonly used method for engineering affinity-based

biosensors is to conjugate to a protein that harbors the desired target specificity a fluorophore whose fluorescence properties are sensitive to local electronic environments. The goal with these protein-fluorophore conjugate-type biosensors is to produce a concurrent fluorescence change upon target binding – reporting target detection to the user. Recognition elements labeled with environmentally-sensitive fluorophores require selection of a conjugation site that has an altered environment in the *apo* versus ligand-bound state, which often means construction and screening of a large number of candidate biosensors before a suitable biosensor is produced (i.e. trial-and-error-based approaches to select labeling positions). Previous research focused on fluorescent labeling of proteins has attempted to circumvent some of the trial-and-error process by utilizing static protein structural data to place a fluorescent group in close proximity to the ligand-binding site. The rationale in this case is that a bound ligand will directly influence the environment of a fluorophore, resulting in a fluorescence change. However, a drawback to this approach is that a fluorescent group proximal to the ligand binding site can compromise protein function, resulting in altered binding affinity or specificity when compared the original protein (e.g. as in (De Lorimier et al. 2002)). A previously unexplored aspect of biosensor design is the integration of protein dynamics information into biosensor development pipelines, such as identification of small-scale changes in amino acid dynamics in an *apo* vs. ligand-bound state of a protein and subsequent fluorescent labeling of positions with altered dynamics.

1.2.6 Solute-binding proteins (SBPs). SBPs are class of proteins which have diverse ligand-binding specificities ideal for use as recognition elements for the construction of protein-fluorophore conjugate-based biosensors. SBPs and SBP domains are associated with a wide variety of protein complexes, usually for solute transport, signal transduction, or gene regulation (Berntsson et al. 2010). SBPs have been shown to associate with ATP-binding cassette (ABC) transporters (Berger 1973; Berger and Heppel 1974; Quiocho and Ledvina 1996; Tam and Saier 1993), G-protein coupled receptors (Armstrong and Gouaux 2000), ligand gated-ion channels (Armstrong and Gouaux 2000), and other membrane protein complexes (Felder et al. 1999; Gonin et al. 2007; Misono 2002; Mulligan et al. 2009; Neiditch et al. 2006). SBP domains have characterized roles in gene expression, for example, prokaryotic DNA-binding proteins such as the Lac repressor have been shown to act as transcription factors (Friedman et al. 1995; Lewis et al. 1996; Schumacher et al. 1994). The three dimensional structure of SBPs are highly conserved including two domains connected by a hinge region, where ligand binding occurs at the interface between the two domains (Berntsson et al. 2010). In the absence of ligand SBPs largely exist in an open conformation with both domains separated (Quiocho and Ledvina 1996; Tang et al. 2007). Ligand binding stabilizes a closed conformation of the protein via a Venus Fly-trap mechanism (Mao et al. 1982). Due to their analyte specificity and large conformational change upon binding, SBPs have been utilized in the construction of several affinity-based biosensors (Brune et al. 1994; De Lorimier et al. 2002; Okoh et al. 2006; Smith et al. 2017; Solscheid et al. 2015) demonstrating the

potential for novel biosensors construction using SBPs. Of particular relevance to this thesis is the engineering of carbohydrate-detecting biosensors using SBP scaffolds, as such sensing devices would circumvent a number of limitations associated with current carbohydrate-detection methods.

1.2.7 Biosensor properties and experimental considerations. There are a number of important features that must be considered when describing the properties of a biosensor. These features include, but are not limited to, sensitivity (binding affinity), rate of target detection (association rate constant), biosensor stability, biosensor specificity (the molecules a biosensor responds to), and biosensor selectivity (the ability to distinguish between related components). In the case of specificity versus selectivity, imagine a situation where two biosensors are both able to detect GDP and GTP. Biosensor 1 gives the same output for GDP and GTP, and Biosensor 2 gives different outputs for detection of GDP or GTP. In this case, both biosensors have the same specificity (i.e. detect GDP and GTP). However, Biosensor 2 has improved selectivity compared to Biosensor 1 as it can distinguish between GDP and GTP.

With respect to biosensor sensitivity, when utilizing protein scaffolds from nature the properties of a biosensor are often constrained to those of the starting protein (e.g. ligand binding affinity constrained to that of the starting protein scaffold). The aforementioned constraint may limit biosensor dynamic range, as single-site binding biosensors often transition from ~10% to ~90% signal saturation with an

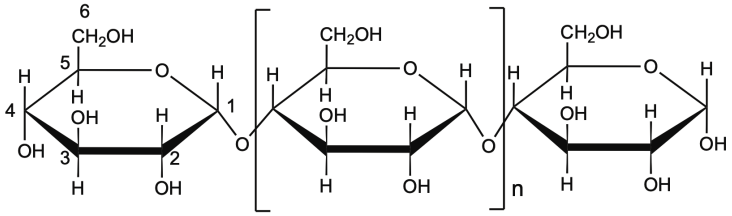
80-fold change in ligand concentration (where the mid-point of this transition occurs at the K_D) (Vallée-Bélisle et al. 2012). A solution to this limitation is to develop a biosensor set capable of responding to a different range of ligand concentrations. Such a biosensor set can therefore be used in experimental conditions where the optimal dynamic range of target detection is outside the range of the starting protein scaffold. In practice, these experimental considerations enable higher throughput enzyme characterization as many different reaction conditions can be tested at once, with many different target detection thresholds to rapidly identify relevant reaction conditions that may otherwise require optimization of input material concentrations. Research of a similar scope has been performed to alter the dynamic range of a phosphate-detecting biosensor, which has since enabled assays where less sensitive phosphate readout was required (Solscheid et al. 2015). Such a biosensor suite with alternative detection thresholds will therefore further enhance the utility of carbohydrate-detecting biosensors by expanding dynamic ranges of ligand detection and further increase application scope.

1.2.8 Carbohydrates and Carbohydrate Active Enzymes. Complex carbohydrates are abundant and have diverse roles in nature. For example, storage polysaccharides, such as starch function as energy storage in plants (Zeeman et al. 2010); and structural polysaccharides such as pectin (e.g. homogalacturonans, substituted galacturonans, rhamnogalacturonans) are components of cell walls (Caffall and Mohnen 2009). Carbohydrate structures are

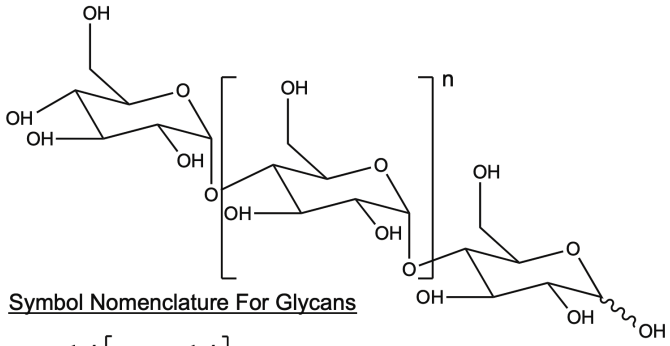
commonly illustrated using all-atom representations, bond-line representations, and the Symbol Nomenclature For Glycans system (Fig. 1.1). Enzymes involved in the breakdown of complex carbohydrates to smaller products are referred to as carbohydrate-active enzymes (CAZymes). CAZyme activity has been utilized in industry for several processes including biofuel production, food processing, and textile finishing (Polaina and MacCabe 2007). There are three main CAZyme classes which depolymerize carbohydrates and are characterized by their substrate activities: glycoside hydrolases (GHs), polysaccharide lyases (PLs), and auxiliary activities (AAs).

(A) Structural representations of amylose

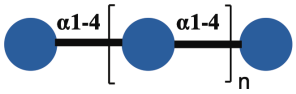
All Atom Representation



Bond-Line Representation

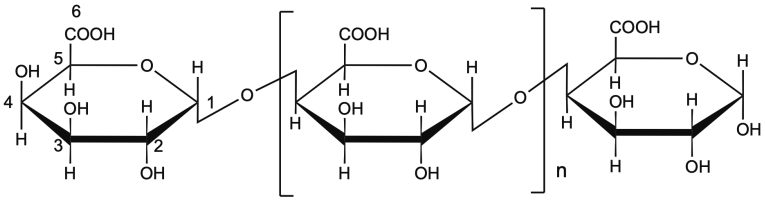


Symbol Nomenclature For Glycans

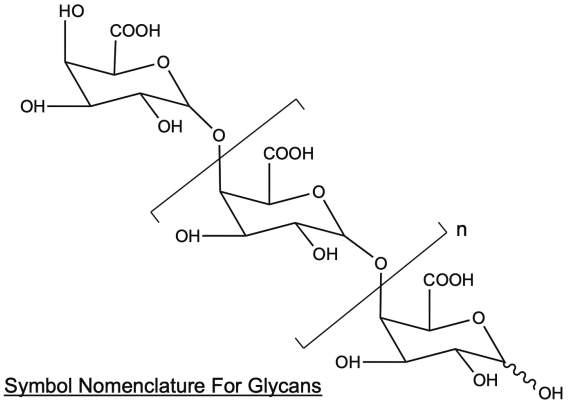


(B) Structural representations of homogalacturonan

All Atom Representation



Bond-Line Representation



Symbol Nomenclature For Glycans

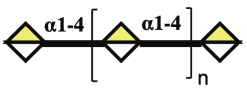


Fig. 1.1 (Figure on previous page). Carbohydrate structural features and representations. (A) Structural features/ representations of amylose (a component of starch). Amylose is a chain of **α** 1-4 linked glucose rings, where the number of linkages (n in the above diagrams) ranges from hundreds to thousands of linkages. All atom, Bond-line, and Symbol Nomenclature For Glycans representations are shown. (B) Structural features/ representations of homogalacturonan (a component of pectin). Homogalacturonan is a chain of **α** 1-4 linked galacturonic acid rings, where the number of linkages (n in the above diagrams) is typically ~100 (Wolf et al. 2009). All atom, Bond-line, and Symbol Nomenclature For Glycans representations are shown.

GHs (EC 3.2.1.-) are a group of CAZymes which hydrolyze the glycosidic bond between carbohydrates, or between a carbohydrate and a non-carbohydrate entity (Lombard et al. 2013). There are two common reaction mechanisms used by GHs: retaining and inverting (Fig. 1.2) - and these mechanisms can be distinguished by comparing the geometry of substituents around the anomeric carbon that is part of the respective reaction. In the case of an inverting mechanism first described by Koshland et al., the reaction generally proceeds in a single step using a general acid and a general base provided by two amino acid side chains (Koshland Jr 1953). For retaining GHs, hydrolysis occurs via a two-step mechanism involving a covalent glycosyl-enzyme, an acid/base and a nucleophile (Koshland Jr 1953). In the case of PLs (EC 4.2.2.-), these enzymes utilize general *syn*- and *anti*- elimination mechanisms to produce a terminal unsaturated hexenuronic acid via β-elimination (Lombard et al. 2013). The PL mechanism has a few key steps and generally begins via proton abstraction from the C5 of the uronic acid sugar ring to destabilize the carbohydrate, followed by charge delocalization onto the C6 carbonyl group and cleavage of the O4-C4 bond (Lombard et al. 2010). Last, AAs are redox enzymes that often act in conjunction

with CAZymes (e.g. lignin breakdown enzymes, which may not act on carbohydrates but are often associated with carbohydrates in the plant cell wall (Lombard et al. 2013)).

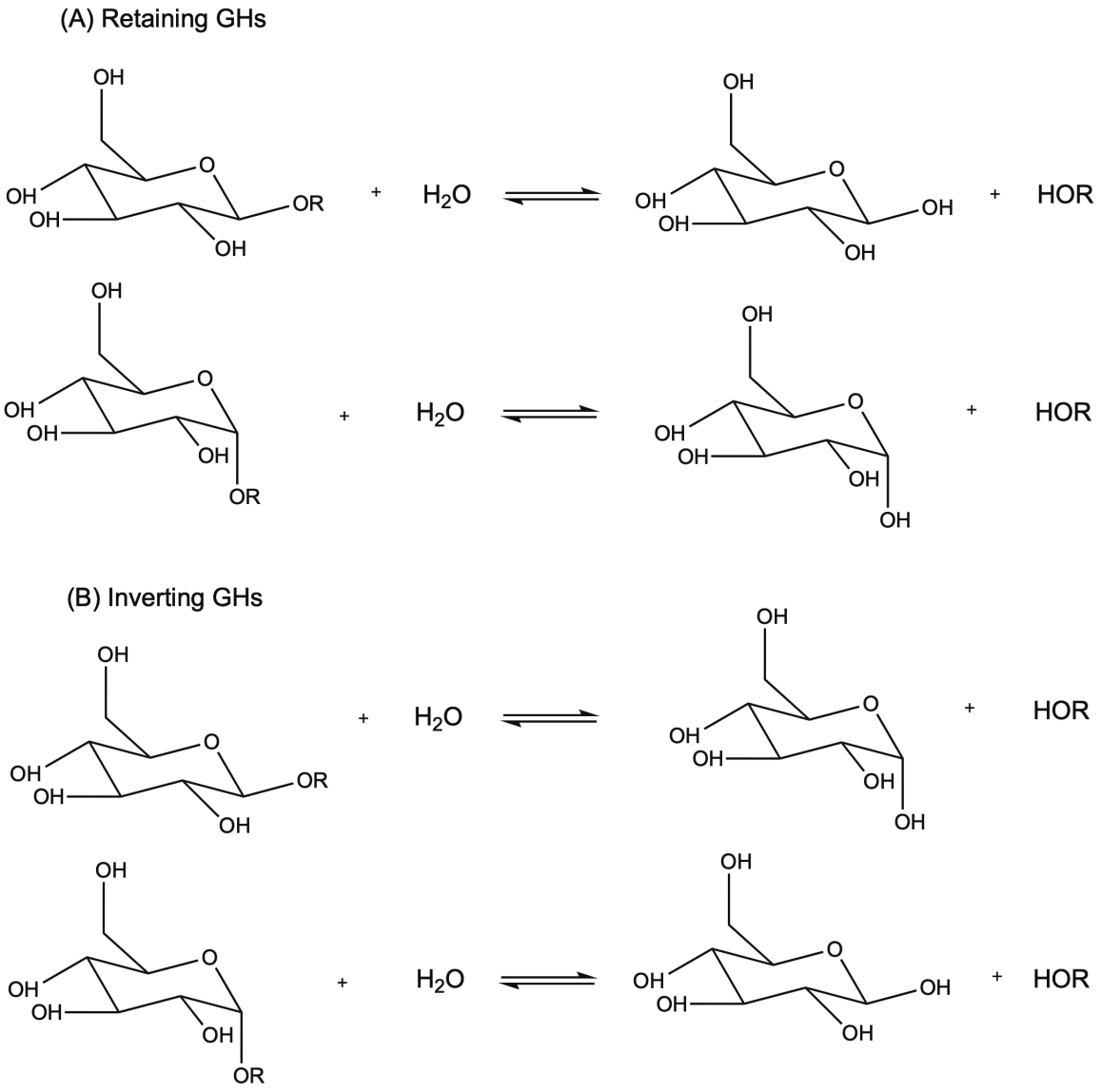


Fig. 1.2. Common reaction mechanisms used by GHs. The most common mechanisms utilized by GHs are either retaining (A) or inverting (B), the names refer to the stereochemistry of the anomeric carbon in the product when compared to the starting carbohydrate. In the above diagram a single glucose ring of a maltooligosaccharide is depicted, however, GHs exist which are active on other carbohydrates as well.

1.2.9 Carbohydrate detection - an opportunity for affinity-based biosensors. Studies of the activity of a CAZyme commonly involve chemical modification of carbohydrate products, followed by analysis via Mass Spectroscopy, Nuclear Magnetic Resonance Spectroscopy, Chromatography, or Electrophoresis (Harvey 2011). Although useful for many applications, these methods often lack selectivity and are not well-suited for research requiring rapid, or high-throughput detection of carbohydrates. An ideal carbohydrate-detection system would be easily scalable for a large number of samples, provide specific detection of a carbohydrate of interest, and report the release of a molecule of interest in real-time enabling detailed CAZyme kinetic assays. To this end, affinity-based biosensors engineered to detect target carbohydrates would prove a transformative solution to the shortcomings of common CAZyme characterization assays. Specifically, the ability to provide real-time monitoring of CAZyme activity, scale well for high-throughput, and have detection specificity defined by that of the solute-binding protein scaffold.

1.2.10 Molecular dynamics simulations and their potential for use in forward-engineering affinity-based biosensors. Molecular dynamics (MD) simulations are a computational method for analyzing molecular systems using Newtonian physics (Badar et al. 2022). The underlying principle of molecular dynamics simulations is that given the positions of all atoms in a biomolecular system, the force exerted on each atom by every other atom can be calculated. The forces applied to each atom are repeatedly re-calculated and used to adjust the position of the constituent atoms as the simulation proceeds. The resulting trajectory

describes, in atomistic detail, the configuration of the biological system as a function of the simulation time. These trajectories can then be analyzed to address relevant biological questions with respect to the underlying dynamics of the system.

Multi-scale approaches have been developed for MD simulations which vary with respect to their resolution and applicability to different sizes of biological systems (Singh and Li 2019). Although MD simulations describe the dynamics of a molecular system at atomic resolution, they cannot describe chemical reactions. To study chemical reaction mechanisms quantum mechanical calculations are often used (Engkvist et al. 2018). But with high-resolution quantum mechanical models, the tradeoff is that current computational capacities limit the use of quantum mechanical calculations to molecular systems with only a few atoms and shorter time scales than alternative methods (typically picometer system sizes, typically picosecond simulation times). At the next stage of molecular resolution, all atom molecular dynamics simulations treat components of a biological system with atomistic detail and are useful for applications where the motion of atoms is required to address research questions (e.g. examination of individual amino acid conformations, or local conformational dynamics (Hollingsworth and Dror 2018)). Reducing resolution even further, coarse-grained molecular dynamics simulations use reduced representations of atoms (i.e. pseudoatoms) to define a group of atoms in a simulation (Singh and Li 2019). This reduction in resolution is accompanied by a reduction in the degrees of freedom of the system, and therefore

reduces the computational demands of a simulation. Coarse-grained simulations are most often used for representing larger biological systems (e.g. ribosomes or biological assemblies) and/or studying events that occur on longer timescales (e.g. protein folding).

Previous work in our group combining MD simulations with rapid-kinetic methods has highlighted how binding of small ligands to their respective proteins can alter the localized dynamics of amino acids distal from the ligand binding site (De Laurentiis et al. 2016; Girodat et al. 2019; Mercier et al. 2015; Rosler et al. 2015). Specifically, De Laurentiis et al. 2016 used a combination of molecular dynamics and stopped-flow data to show that destabilization of a loop between Helix E & F in Elongation Factor-Tu (EF-Tu) compromises ligand binding. Girodat et al. 2019 showed ligand binding to EF-Tu causes dramatic changes in the localized structural dynamics of the protein distal from ligand binding sites. Mercier et al. 2015 described how ligand dissociation kinetics drive local changes in protein structural dynamics in EF-Tu, and Rosler et al. 2015 showed that the translational GTPase YchF undergoes nucleotide-selective changes in the conformation of a flexible loop. This information demonstrates the potential that structural dynamics changes in response to ligand binding have to be exploited as generalization principles for the identification of fluorophore conjugation sites on a broad range of proteins.

1.3 Hypothesis and Objectives:

The overarching hypothesis of this thesis is that biosensors can be designed based on transient and localized small-scale changes in amino acid dynamics. I further propose that based on these features a generalizable, *in silico* workflow for biosensor design using molecular dynamics simulation data and first-principles can be developed. To this end, the following objectives are proposed with the goal of both establishing and iteratively improving the biosensor design platform guided by successes/failures of designs tested in the wet-lab.

Objective 1: Establish initial criteria for biosensor design platform via construction of a maltooligosaccharide (MOS)-detecting biosensor. For this objective, I examine molecular dynamics simulation data of the MalX protein scaffold using a variety of different metrics to predict fluorophore conjugation sites for biosensor development. After screening candidate biosensors *in vitro*, the efficacy of these initial designs can be used to guide improvements to our *in silico* scoring system for Objective 2.

Objective 2: Refine the *in silico* scoring algorithm for biosensor design based on results from Objective 1 for the construction of an oligogalacturonide-detecting biosensor.

Objective 3: Expand the dynamic range of my MOS-detecting biosensor set via construction of variants with altered ligand binding affinity. The construction of

MOS-detecting biosensors with altered ligand binding affinity will have two main purposes. First, expanding the dynamic range assayable by the MOS-detecting biosensor set will enable a broader range of applications where alternative target detection thresholds are required. Second, this objective will allow me to survey the impact of modifying ligand binding affinity on the underlying dynamics/readout of the MOS-detecting biosensor set and to further elucidate underlying biosensing mechanisms. Together, this data will inform future generalizable principles in biosensor rational design.

CHAPTER 2: 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

2.1 Preface:

The contents of this Chapter have previously been published as: Smith, D.D., Girodat, D., Abbott, D.W., and Wieden, H.-J. (2022) Construction of a highly specific and selective carbohydrate-detecting biosensor utilizing Computational Identification of Non-disruptive Conjugation sites (CINC) for flexible and streamlined biosensor design. *Biosensors and Bioelectronics*, 200, 113899. I performed all biochemical characterization in this study, analyzed molecular dynamics data with Dr. Dylan Girodat, and wrote the first draft of the manuscript. The contents of the aforementioned manuscript have been reformatted to adhere to the guidelines of this thesis. The work in this chapter was included in International PCT Patent Application No. CA2021/050110 (Filed January 30, 2021) which has since been filed with the US and EU Patent offices.

Prior to conducting this study, our research group noticed that molecular dynamics simulations could identify small-scale (amino acid level) changes in dynamics upon ligand binding at sites distal to the ligand binding site. We therefore speculated that these small-scale changes in amino acid dynamics could be

exploited as reporter positions for biosensing applications via conjugation with a fluorescent group.

Manuscript authors: Dustin D Smith[1,2], Dylan Girodat[1,2,3], D. Wade Abbott[2,4], and Hans-Joachim Wieden[1,2,5]

Author affiliations:

[1]Alberta RNA Research and Training Institute (ARRTI), University of Lethbridge, Lethbridge, AB, Canada

[2]Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada

[3]Current Address: Department of Chemistry, Fulbright College of Arts and Sciences, University of Arkansas. Fayetteville, AR, USA.

[4]Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, AB, Canada

[5]Current Address: Department of Microbiology, University of Manitoba, Winnipeg, MB, Canada

Author contributions (CRediT format):

Dustin D. Smith: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization.

Dylan Girodat: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing –review & editing, Visualization.

D. Wade Abbott: Conceptualization, Resources, Writing – original draft, Writing – review & editing.

Hans-Joachim Wieden: Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition.

2.2 Abstract:

Fluorescently-labeled solute-binding proteins that alter their fluorescence output in response to ligand binding have been utilized as biosensors for a variety of applications. Coupling protein ligand binding to altered fluorescence output often requires trial and error-based testing of both multiple labeling positions and fluorophores to produce a functional biosensor with the desired properties. This approach is laborious and can lead to reduced ligand binding affinity or altered ligand specificity. Here we report the Computational Identification of Non-disruptive Conjugation sites (CINC) for streamlined identification of fluorophore conjugation sites. By exploiting the structural dynamics properties of proteins, CINC identifies positions where conjugation of a fluorophore results in a fluorescence change upon ligand binding without disrupting protein function. We show that a CINC-developed maltoligosaccharide (MOS)-detecting biosensor is capable of rapid (k_{on} = 20 $\mu M^{-1}s^{-1}$), sensitive (sub-μM K_D) and selective MOS detection. The MOS-detecting biosensor is modular with respect to the spectroscopic properties and

demonstrates portability to detecting MOS released via α-amylase-catalyzed depolymerization of starch using both a stopped-flow and a microplate reader assay. Our MOS-detecting biosensor represents a first-in-class probe whose design was guided by changes in localized dynamics of individual amino acid positions, supporting expansion of the CINC pipeline as an indispensable tool for a wide range of protein engineering applications.

2.3 Introduction:

Biosensors are devices that utilize biological components to detect the presence or amount of molecules of interest and give a quantifiable output (Mehrotra 2016). The market value for biosensors in 2019 was US$ 21.2 billion and is projected to exceed US$ 31.5 billion by 2024 (Markets and Markets 2019). These projections are derived from the impact that biosensors have on, but are not limited to, medicine, agriculture, environmental monitoring, biofuels, and academic research (Harrison and Dunlop 2012; Ispas et al. 2012; Law et al. 2015; Mehrotra 2016; Pardee et al. 2014; Pardee et al. 2016; Rai et al. 2012; Shin 2011; Srilatha 2011; Verma and Bhardwaj 2015). As such, the opportunity to develop novel biosensors capable of specific and sensitive monitoring of compounds at low concentrations and with high dynamic range is growing rapidly (Edwards 2021; Yu et al. 2021; Zeng et al. 2021; Zhang et al. 2018a; Zhang et al. 2018b). A general approach to biosensor development is the conjugation of fluorophores to a particular protein that binds a specific molecule of interest. This approach utilizes the innate specificity and affinity that proteins possess for their ligands and exploits

fluorescence signal changes induced upon protein-ligand interactions. Protein-fluorophore conjugates have been used as anion or cation-detecting biosensors (Belousov et al. 2006; Chen et al. 2013; De Lorimier et al. 2002; Smith et al. 2017; Solscheid et al. 2015), nucleotide-detecting biosensors (Berg et al. 2009; Bianchi-Smiraglia et al. 2017; Kunzelmann and Webb 2010), amino acid detecting biosensors (Ciftci et al. 2020; De Lorimier et al. 2002; Ko et al. 2017; Ko et al. 2016), and carbohydrate-detecting biosensors (De Lorimier et al. 2002; Fonin et al. 2014; Hsieh et al. 2012; Marvin et al. 2011; Tian et al. 2007). The development of biosensors often involves fluorophore labeling near the ligand binding site, exploiting ligand binding events to induce a change in the local environment proximal to the conjugated fluorophore (e.g., as in (Brune et al. 1994; De Lorimier et al. 2002; Gilardi et al. 1994)). However, introducing a fluorescent label in the vicinity of the ligand binding site can perturb the activity of the resulting protein-conjugate, reducing binding affinity or specificity for the respective ligand (Brune et al. 1994; De Lorimier et al. 2002; Gilardi et al. 1994; Toseland 2013). Altogether, stochastic approaches can be expensive, laborious, time consuming and unsuccessful; thus, the development of an *in silico* screening step for selection of non-disruptive fluorophore conjugation sites would accelerate the design and development of new biosensors.

Previous findings by our group have highlighted how binding of small ligands to their respective proteins, such as EF-Tu and YchF, can alter the localized dynamics of amino acids distal from the ligand binding site (De Laurentiis

et al. 2016; Girodat et al. 2019; Mercier et al. 2015; Rosler et al. 2015). The localized structural dynamics responses of amino acids gleaned from these investigations enabled our establishment of a computational pipeline, entitled Computational Identification of Non-disruptive Conjugation sites (CINC), to predict and exploit these observations as generation principles for identification of fluorophore conjugation sites on structurally unrelated proteins. More specifically, we identify how the structural dynamics properties of amino acids are altered upon ligand binding, producing localized environmental changes which can be detected by a conjugated fluorophore distal from the ligand binding site. This approach examines the localized structural dynamics properties of a protein of interest with amino acid resolution through molecular dynamics simulations of the *apo* and ligand-bound states followed by scoring each amino acid position with respect to changes in local dynamics. Our CINC pipeline examines several criteria proposed to influence fluorescence of environmentally-sensitive fluorophores (Fu and Finney 2018; Marvin et al. 2011; Marvin et al. 2019; Rurack and Resch-Genger 2002). Additionally, we have considered the proximity of amino acids with regards to tryptophan residues, enabling access to Förster Resonance Energy Transfer (FRET)-based biosensors via introduction of a fluorophore capable of forming a tryptophan FRET pair. As a proof of concept, we report the development of a biosensor using the CINC pipeline to engineer the MalX protein from *Streptococcus pneumoniae* (Abbott et al. 2010) for specific, continuous, and direct detection of maltooligosaccharides (MOS). Our approach is not limited to ligand binding and in principle is applicable to any protein that undergoes localized

structural dynamics changes as a function of pH, oxidation, temperature, or other factors that influence small scale structural dynamics. In summary, the CINC pipeline is a rapid, unbiased *in silico* screening step for engineering novel protein-fluorophore conjugates. We envision CINC having a transformative role in applications where rapid and automated engineering of diverse biosensor sets is required, for example, in the design of carbohydrate-detecting biosensors which to date have a limited set of tools capable of detection of specific carbohydrates.

2.4 Materials and Methods:

2.4.1 Computational Identification of Non-disruptive Conjugation sites (CINC), molecular dynamics simulations. Structural models of MalX in its *apo* and MOS-bound state were obtained from protein data bank (2XD2 and 2XD3, respectively; (Abbott et al. 2010)) and used for molecular dynamics simulations. Following the approach established in our previous studies (De Laurentiis et al. 2016; Girodat et al. 2019; Mercier et al. 2015; Rosler et al. 2015), each model of the protein was solvated with a TIP3P water box extending 20 Å from any point on the protein and brought to a concentration of 100 mM NaCl. The potential energy of the water and then the entire system was minimized using AMBER FF99SB force fields for 10 000 steps each. The system was equilibrated and subsequently heated to 300 K for 10 000 steps each. Finally, molecular dynamics simulations were performed for 100 ns at a step size of 2 fs using Langevin dynamics to maintain temperature at 300 K. The root-mean squared deviation (RMSD) for both

apo and MOS-bound states remained stable after an initial equilibration of 10 ns at values of ~2.5 Å and ~4 Å respectively (Fig. A1.1).

2.4.2 F_{Score}. Trajectories from each simulation were put through the CINC pipeline where the backbone flexibility, backbone dihedral angles, distance to ligand, distance to tryptophan residues, and solvent accessibility were measured to develop a ranking score (F_{Score}, Fig. 2.1A). These properties were chosen as localized dynamic features of an amino acid that would have the largest impact on the environment of the conjugated fluorophore. Only amino acids with a conservation score of 1 from the ConSurf server (Ashkenazy et al. 2016; Celniker et al. 2013) and that were further than 5Å away from the ligand binding site were assigned an F_{Score} to ensure there would be no impact on ligand binding properties (Fig. 2.1B). Flexibility was determined by measuring the root-mean-square fluctuation (RMSF) of the Cα atoms of each amino acid over the course of both the ligand bound (F_{Lr}) and *apo* (F_{Ar}) simulations (Eq. 2.1). Backbone dihedral angles were measured by determining the ϕ (between atoms C-N-Cα-C) or ψ (between atoms N-Cα-C-N) angles and then using these angles to construct a Ramachandran plot. The Ramachandran plot was then transformed into a 180x180 matrix (Fig. 2.1C, D) and the sum of the difference between the matrices for the ligand bound (B_{Lr}) and *apo* (B_{Ar}) conformation was used to determine the conformational change in the backbone dihedrals at that residue (Eq. 2.1). Distances from each amino acid to the ligand were measured for the ligand bound simulation (Posor et al.) directly, whereas the structure of the ligand-bound

conformation was aligned to the *apo* simulation to measure the distance from each amino acid to the ligand (L_{Ar}) (Eq. 2.1). Tryptophan distances were measured for both ligand-bound (T_{Lr}) and *apo* (T_{Ar}) (Eq. 2.1). In circumstances where there is more than one tryptophan the distances between an amino acid and each tryptophan were averaged to obtain an average amino acid to tryptophan distance. Last, the solvent accessibility of each amino acid was determined and compared between the ligand bound (S_{Lr}) and *apo* (S_{Ar}) conformations (Eq. 2.1). The values measured were averaged ⟨ ⟩ and the absolute value | | of the difference between ligand bound and *apo* conformations was normalized compared to the maximum (∨) value obtained (Eq. 2.1). Using this approach each amino acid that was under consideration could be assigned an F_{score} from 0 to 5, with a higher score indicating a better candidate for fluorophore conjugation.

$$F_{score} = \frac{|\langle S_{Lr}\rangle - \langle S_{Ar}\rangle|}{\vee|\langle S_{Lr}\rangle - \langle S_{Ar}\rangle|} + \frac{|\langle T_{Lr}\rangle - \langle T_{Ar}\rangle|}{\vee|\langle T_{Lr}\rangle - \langle T_{Ar}\rangle|} + \frac{|\langle L_{Lr}\rangle - \langle L_{Ar}\rangle|}{\vee|\langle L_{Lr}\rangle - \langle L_{Ar}\rangle|} + \frac{|\langle F_{Lr}\rangle - \langle F_{Ar}\rangle|}{\vee|\langle F_{Lr}\rangle - \langle F_{Ar}\rangle|} + \frac{\sum|[B_{Lr}] - [B_{Ar}]|}{\vee\sum|[B_{Lr}] - [B_{Ar}]|} \quad \text{(Eq. 2.1)}$$

2.4.3 Biosensor construct design. All MalX variants were engineered to lack the signal secretion peptide and lipoprotein attachment motif found in wild-type *malX* (*Streptococcus pneumoniae*) as previously described (Abbott et al. 2010). *malX* variants were synthesized with flanking 5' *NheI* and 3' *XhoI* restriction sites and subcloned into pET28a yielding pET28a::*malX A134C*, pET28a::*malX A180C,* pET28a::*malX T249C,* and pET28a::*malX E318C* (BioBasic Canada Inc.). Criteria

for selection of these variants was guided by the CINC pipeline and is explained in detail in Section 2.5.2. All constructs encoded an N-terminal poly-Histidine tag fusion.

2.4.4 Expression of MalX variants. *Escherichia coli* BL21(DE3) gold cells (Agilent) transformed with either pET28a::*malX A134C*, pET28a::*malX A180C,* pET28a::*malX T249C,* or pET28a::*malX E318C* were used to inoculate LB media (10 g/L typtone, 5 g/L yeast extract, 10 g/L NaCl) supplemented with 50 mg/L kanamycin to an optical density at 600 nm (OD_{600}) ≈ 0.1. After growth at 37°C with 200 RPM shaking, isopropyl β-$_D$-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM when OD_{600} reached approximately 0.6. Cultures were grown for an additional 3 hours, cells harvested by centrifugation (5000 x g, 10 minutes, 4°C), flash frozen, and stored at -80°C for further use.

2.4.5 Purification of MalX variants. Cell pellets were resuspended in 7 mL/g of Buffer A (20 mM Tris pH 8.0 at 4°C, 0.5 M NaCl, 10 mM imidazole, 7 mM β-mercaptoethanol, 1 mM phenylmethylsulfonylfluoride), and lysozyme added to a final concentration of 1 mg/mL prior to a 30-minute incubation at 4°C. Sodium deoxycholate was added to a final concentration of 12.5 mg per gram of cells, and the suspension was sonicated (Branson Sonifier 450) on ice for 1 min at 50% output and 60% duty cycle (repeated twice, with a 5 min break between cycles). The resulting lysate was centrifuged (3 000 x g, 30 minutes, 4°C) to pellet insoluble debris, followed by additional centrifugation to produce S30 extract (30 000 x g, 45

minutes, 4°C). The obtained S30 extract was applied to Ni^{2+}-Sepharose resin in a batch-chromatography setup (3 mL resin per 1 g of cells opened) and incubated for 30 minutes at 4°C. Ni^{2+}-Sepharose resin was collected by centrifugation (500 x g, 2 minutes, 4°C), and supernatant decanted. The collected Ni^{2+}-Sepharose resin was washed 3 times with 10 resin-volumes of Buffer A, followed by 4 washes with 10 resin-volumes Buffer B (Buffer A with 20 mM imidazole). Bound protein was eluted six times with 1 resin-volume Buffer C (Buffer A with 250 mM imidazole). Samples were analyzed via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) stained with Coomassie Brilliant Blue, and fractions containing each MalX variant were pooled and dialyzed at 4°C into Buffer D for labeling (20 mM Tris pH 7.5 @ 20°C, 0.5 M NaCl, 30 mM imidazole, 20 mL sample into 1 L Buffer D, dialysis tubing molecular weight cut-off 12.4 kDa, 3 changes). Purified protein was flash-frozen with liquid nitrogen and stored at -80°C for further use. Purification yields were typically 100 mg of protein per liter of culture.

2.4.6 Fluorescent labeling of MalX variants. Labeling reactions were performed essentially as described previously (Smith et al. 2017). The respective MalX variants (100 μM concentration in labeling reaction, 15 mL reaction volume) were incubated with 3.5 mL Ni^{2+}-Sepharose resin in Buffer D. Five-fold molar excess of either 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin (CPM, from 20 mM stock in dimethyl sulfoxide (DMSO), Biotium), *N*-(7-dimethylamino-4-methylcoumarin-3-yl)maleimide (DACM, from 20 mM stock in dimethylformamide (DMF), Invitrogen), *N*-[2-(dansylamino)ethyl]maleimide (Dansyl, from 25 mM stock

in DMSO, Sigma Aldrich), fluorescein-5-maleimide (Fluorescein, from 50 mM stock in DMSO, Biotium), 7-diethylamino-3-[*N*-(2-maleimidoethyl)carbamoyl]coumarin (MDCC, from 20 mM stock in DMF, Sigma Aldrich), *N*-(1-pyrene)maleimide (Pyrene, from 20 mM stock in DMF, Sigma Aldrich), or Rhodamine Red C2 maleimide (Rhodamine Red, from 20 mM stock in DMF, Invitrogen) was added dropwise to the labeling mixture corresponding to a final concentration of 500 μM in the labeling reaction. The labeling reaction was subsequently incubated on an end-over-end mixer at room temperature for 2 hours, followed by 12 hours at 4°C. Ni^{2+}-Sepharose resin was collected by centrifugation (500 x g, 2 minutes, 4°C) and the supernatant decanted. The resin was washed six times with three resin-volumes of Buffer D, followed by six elutions with Buffer E (Buffer D with 250 mM imidazole). Labeling procedure samples were analyzed via SDS PAGE, and the respective gels visualized under UV light (312 nm) or stained using Coomassie Brilliant Blue. Fractions containing labeled MalX were pooled and dialyzed into 50 mM Tris pH 7.5 @ 20°C (20 mL sample in 1 L Buffer, 4 changes, molecular weight cut-off 12.4 kDa). The protein recovery from the labeling procedure was ~50% and labeling efficiencies were typically 70% to >90%. Concentration of labeled MalX variants were calculated as described below (Eq. 2.2, 2.3) using the following parameters: 0.164 as a correction factor accounting for MDCC absorption at 280 nm (Brune et al. 1994; Smith et al. 2017), $\varepsilon_{280,\ MalX}$ = 61 310 M^{-1} cm^{-1} is the calculated extinction coefficient based on protein primary sequence (Gasteiger et al. 2005), L is instrument path length in cm, and $\varepsilon_{430,\ MDCC}$ = 46 800 $M^{-1}cm^{-1}$ (Brune et al. 1994; Smith et al. 2017). MalX A134C-MDCC and MalX T249C-MDCC

Biosensor preparations are currently available from Allos Bioscience Ltd. (B1001 and B1002 respectively).

$$[MalX\ variant] = \frac{A_{280} - (A_{430} \times 0.164)}{\varepsilon_{280,MalX} \times L} \quad \text{(Eq. 2.2)}$$

$$[MDCC] = \frac{A_{430}}{\varepsilon_{430,MDCC} \times L} \quad \text{(Eq. 2.3)}$$

Protein concentration for MalX A134C-CPM, MalX A134C-DACM, MalX A134C-Dansyl, MalX A134C-Fluorescein, MalX A134C-Pyrene, and MalX A134C-Rhodamine Red were determined via densitometry analysis (ImageJ (Schneider et al. 2012)) of SDS-PAGE gels using MalX A134C-MDCC samples to form a standard curve. Concentrations of the respective conjugated fluorescent labels were calculated from spectroscopic data by modifying Equation 2.3 using the appropriate wavelength and manufacturer provided extinction coefficients: $\varepsilon_{384,\ CPM}$ = 33 000 $M^{-1}cm^{-1}$, $\varepsilon_{384,\ DACM}$ = 27 000 $M^{-1}cm^{-1}$, $\varepsilon_{340,\ Dansyl}$ = 4 300 $M^{-1}cm^{-1}$, $\varepsilon_{494,\ Fluorescein}$ = 68 000 $M^{-1}cm^{-1}$, $\varepsilon_{339,\ Pyrene}$ = 38 000 $M^{-1}cm^{-1}$, $\varepsilon_{560,\ Rhodamine\ Red}$ = 119 000 $M^{-1}cm^{-1}$. Labeling efficiencies for all protein-fluorophore conjugates were determined using Equation 2.4.

$$Labeling\ efficiency = \frac{[Fluorophore]}{[MalX\ A134C]} \times 100\% \quad \text{(Eq. 2.4)}$$

2.4.7 Carbohydrates and α-amylase. Maltotriose (M3) was purchased from Megazyme (O-MAL3); isomaltotriose was purchased from Carbosynth (OI05352);

maltose (M5885), maltoheptaose (M7753), water soluble starch (S9765), and *Bacillus licheniformis* α-amylase (A3403) were purchased from Sigma Aldrich.

2.4.8 Equilibrium fluorescence experiments. Fluorescence spectrophotometry was performed using a Quanta Master 60 Fluorescence Spectrometer (Photon Technology International, all experiments utilized 1 nm step size, 1 s integration). For experiments in Fig. 2.2, excitation slit widths were 1 nm, emission slit widths were 2 nm, excitation wavelength was 420 nm, and emission wavelength was 425-575 nm. For data presented in Fig. 2.3, excitation and emission wavelengths were the same as in Fig. 2.2, except excitation slit widths were 4 nm, emission slit widths were 8 nm. To obtain K_D, a hyperbolic function (Eq. 2.5) was fit to the data in GraphPad Prism v. 9.0, where Y is the absolute value of the fluorescence change at each [M3], B_{max} is the maximum specific binding in the same units as Y extrapolated to very high ligand concentrations, and K_D is the equilibrium binding constant.

$$Y = \frac{B_{max} \times [M3]}{K_D + [M3]} \qquad \text{(Eq. 2.5)}$$

For data presented in Table A1.1 and Table A1.2, protein concentration was 20 nM (except the MalX A134C-Dansyl conjugate, which was assayed at 1 μM due to lower extinction coefficient). Emission spectra were generally recorded in windows of 150 nm and included the designated Table A1.2 emission maxima wavelengths. ΔI was determined as follows:

$$\Delta I = \frac{max.\ emission\ (MOS\ bound) - max.\ emission\ (apo)}{max.\ emission\ (apo)} \times 100\% \qquad \text{(Eq. 2.6)}$$

2.4.9 Rapid-kinetics measurements. Rapid kinetics experiments were performed using a KinTek SF-2004 stopped-flow apparatus (KinTek Corp.) at 20°C. An excitation wavelength of 420 nm was used for all rapid kinetics experiments, and fluorescence emission was measured after passing a 450 nm long-pass filter. Individual fluorescence time-courses were fit with a one exponential function (Eq. 2.7), where F is fluorescence observed at time t, F_∞ is final fluorescence, A is signal amplitude, and k_{app} is apparent rate (TableCurve, Systat Software).

$$F = F_\infty + A_1 \times \exp(-k_{app}t) \qquad \text{(Eq. 2.7)}$$

2.4.10 Microplate reader experiments. Microplate reader experiments were performed using a SpectraMax i3x plate reader (Molecular Devices, Fluorescence: Kinetic mode, 420 nm excitation (9 nm range), 465 nm emission (15 nm range), PMT gain: high, 6 flashes per read, sampling interval: 1 minute) in a 96 well plate. Individual fluorescence time-courses were fit using a one-exponential function as described in Section 2.4.9 (Eq. 2.7), and data was plotted using GraphPad Prism v. 9.0.

2.5 Results:

2.5.1 Overview. As a proof-of-concept for the CINC pipeline we sought to develop a highly specific and sensitive biosensor for the study of Carbohydrate Active Enzymes (CAZymes). CAZymes are a group of sequence-diverse enzyme families belonging to five different functional classes that modify the linkages or decorations of carbohydrates (Lombard et al. 2013). CAZymes are commonly used in bioindustrial processes, including biofuel production, food processing, and textile finishing (Polaina and MacCabe 2007). In this context, we chose MalX from *Streptococcus pneumoniae* as a candidate for the specific, continuous, and direct detection of MOS (Webb et al. 2008). MOS are oligosaccharides comprised of glucose molecules linked via α-1,4-glycosidic linkages. MalX is a lipid-anchored solute-binding protein (SBP) of an ATP binding cassette (ABC) transporter responsible for the selective transport of MOS produced by extracellular debranching of amylopectin by SpuA (Abbott et al. 2010). MalX binds MOS with a degree of polymerization between 3-8 (K_D ~ 1 μM), and accommodates MOS with degree of polymerization of 9-11 (Abbott et al. 2010). Structural studies of MalX indicate that MOS binding occurs at an interface between the N- and C-terminal domains similar to other SBP-ligand interactions, and MOS binding stabilizes a hinge-bend motion where the two domains move towards each other (Abbott et al. 2010). Together, these properties of MalX made it a promising candidate for the development of a MOS-detecting biosensor. Unlike previous biosensor-design approaches that exploit global protein conformational changes in response to ligand binding, we utilized the CINC pipeline to identify labeling positions based on

localized altered dynamics of individual amino acid positions, potentially limiting the impact on protein function.

2.5.2 Selection of fluorophore conjugation sites in MalX using CINC. The CINC pipeline was used to rank-select optimal fluorophore conjugation sites in MalX for development of a novel MOS-detecting biosensor. The CINC pipeline was developed to take into consideration structural dynamics features of amino acids that differ between two distinct states. The properties selected were RMSF, solvent accessible surface area, backbone flexibility, and distance to ligand or tryptophan (Fig. 2.1A). To obtain information regarding changes in the local structural dynamics of each amino acid position in MalX, both the *apo* and MOS-bound MalX states were each subjected to 100 ns molecular dynamics simulations and candidate amino acid positions were assigned an F_{Score} score using the CINC pipeline. Amino acids were considered candidates if they were given a Consurf score of 1, indicating lack of sequence conservation (Ashkenazy et al. 2016; Celniker et al. 2013), to avoid modification of potentially functionally relevant amino acids (Fig. 2.1B). From the F_{Score} data we chose four positions to substitute with cysteine and to subsequently conjugate with a thiol-reactive fluorophore. Positions A134 and A180 were selected as two of the highest overall scorers (F_{Score} 2.7, and 2.0 respectively), and positions T249 and E318 (F_{Score} 1.6, and 1.9 respectively) were chosen as mid-level scorers to test the robustness of F_{Score} (Fig. 2.1B - E). Of these labeling positions, the strongest contributor to F_{score} for each mutant were changes in backbone dihedral angles (A134), changes in RMSF (A180), distance

to Trp (T249), and change in proximity to ligand/solvent accessible surface area (E318) – these variants can therefore be used to examine the predictive power of these parameters in the F_{score} algorithm (Fig. 2.1B - E).

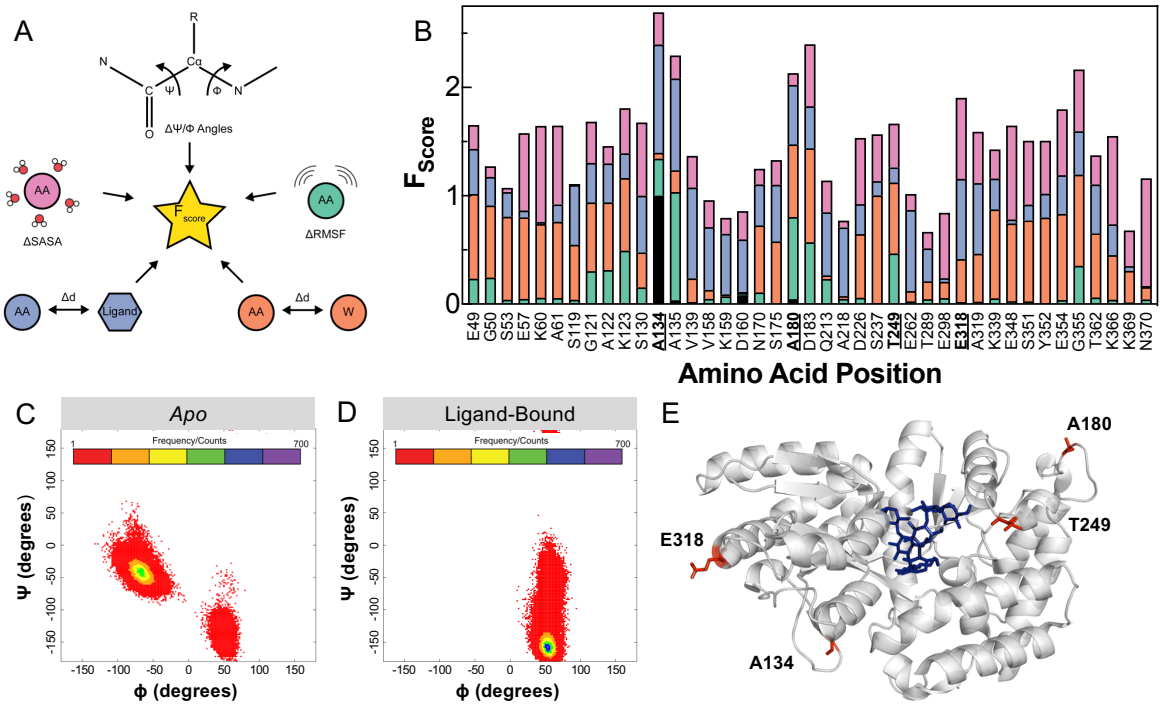


Fig. 2.1. Overview of CINC-based design of MOS-detecting biosensors based on MalX using 100 ns molecular dynamics simulations. (A) Overview of factors that contribute to the F_{Score} of an amino acid position. (B) F_{Score} values and the contributions of individual factors for candidate labeling positions in MalX. Shown are changes in backbone dihedral angles (black), in RMSF (green), in the distance to Trp (orange), in the distance to the ligand (blue), and changes in solvent accessible surface (pink). Underlined Positions indicate sites later selected for fluorophore conjugation. Ramachandran plot for position A134 from 100 ns *apo* (C) or ligand-bound (D) molecular dynamics simulations of MalX. (E) Candidate substitution positions (red) A134, A180, T249, and E318 and their location in MOS-bound MalX (PDB 2XD3, figure generated using PyMOL).

2.5.3 Candidate biosensor response to M3, fluorescence spectrophotometry.

To perform initial screening, each of the four MalX variants were conjugated with MDCC, a diethylaminocoumarin previously used in the construction of other SBP-

based biosensors (Brune et al. 1994; Hanes et al. 2011; Kunzelmann and Webb 2009; Salins et al. 2002; Smith et al. 2017). Each MalX variant labeled with MDCC was examined using fluorescence spectrophotometry, directly measuring the change in fluorescence in response to M3 binding. The highest CINC scorer, A134, and a mid-CINC-scorer T249 exhibited a 20% and 30% fluorescence decrease upon binding of M3, respectively, whereas the MalX variants with MDCC conjugated at A180C and E318C did not exhibit a fluorescence change upon addition of M3 (Fig. 2.2A-D). In the presence of maltose as a ligand we observed little to no fluorescence change for all the variants, suggesting no binding (Fig. 2.2E-H, Fig. A1.2). The aforementioned result is consistent with previous data showing that three pyranose rings at the reducing end of a MOS are required to stabilize a protein-ligand complex with MalX via stacking interactions with aromatic amino acids at subsites 1-3 (Abbott et al. 2010). Therefore, MalX A134C-MDCC and MalX T249C-MDCC selectively detect M3 via a fluorescence change specific to MOS, confirming our initial hypothesis that changes in the local structural dynamics distant from the ligand binding site can be exploited for biosensor design. MalX A134C-MDCC also did not detect isomaltotriose, a compositionally similar carbohydrate to M3 whose α-1,6-glycosidic linkages create a spatially disposed conformation not recognized by MalX (Table A1.1, Fig. A1.2). A larger MOS, maltoheptaose, is detected by MalX A134C-MDCC, further demonstrating biosensor specificity for MOS (Table A1.1, Fig. A1.2). Labeling at positions 134 and 249 does not influence the ligand binding affinity of MalX, as equilibrium titration of each labeled protein with M3 revealed affinities (K_D) of 190 ± 50 nM and

600 ± 200 nM for MalX A134C-MDCC and MalX T249C-MDCC respectively (Fig. 2.3A, B), which are similar to the previously determined affinities of the unlabeled protein measured using UV difference spectroscopy (800 ± 125 nM), and Isothermal Titration Calorimetry (2.0 ± 0.2 μM) (Abbott et al. 2010). Altogether this demonstrates the robustness, sensitivity, and specificity our system for detecting and monitoring sub-μM concentrations of M3 without disruption of the ligand binding properties of wild-type MalX.

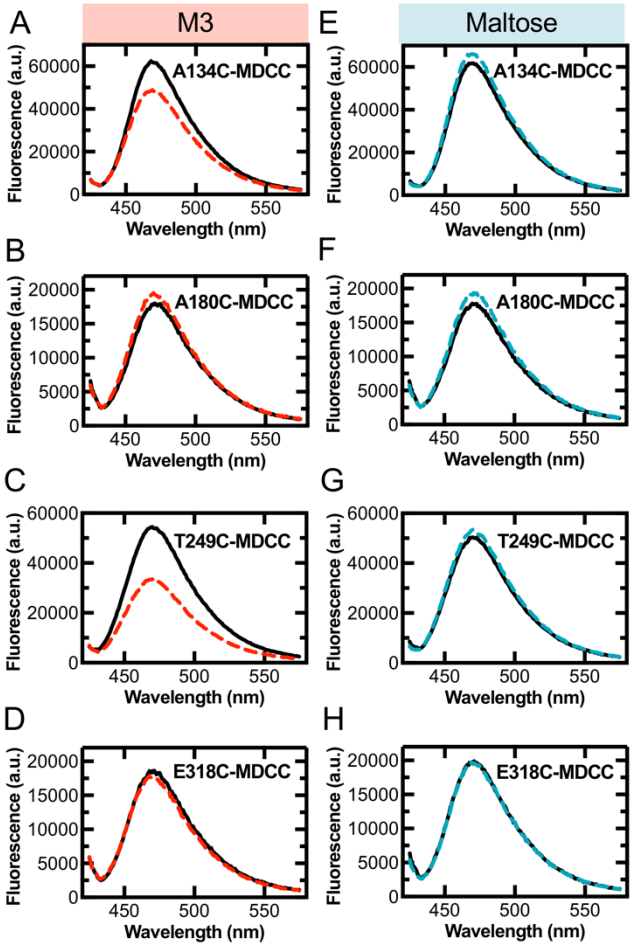


Fig. 2.2. Fluorescence spectra of MOS-detecting biosensors. MOS-detecting biosensor (1 μM) fluorescence in the absence (black solid line) and presence (red dashed line) of 20 μM M3 for (A) A134C-MDCC, (B) A180C-MDCC, (C) T249C-MDCC, and (D) E318C-MDCC. MOS-detecting biosensor in the absence (black solid line) and presence (blue dashed line) of 30 μM maltose for (E) A134C-MDCC, (F) A180C-MDCC, (G) T249C-MDCC, and (H) E318C-MDCC.

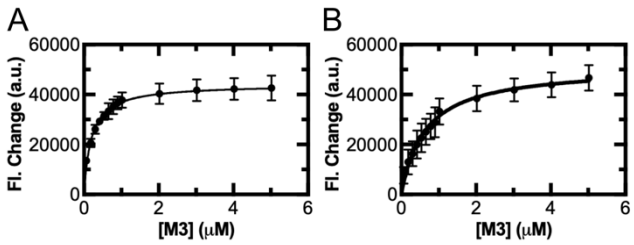


Fig. 2.3. Concentration dependence of M3-binding induced fluorescence change in MOS-detecting biosensors. MOS-detecting biosensor (20 nM) binds M3 with an affinity of (A) A134C-MDCC, 190 ± 50 nM and (B) T249C-MDCC, 600 ± 200 nM (n = 13 data points, best fit ± 95% c.i.) demonstrating the high sensitivity of the biosensor.

2.5.4 Real-time detection of M3 by MalX A134C-MDCC. To assess the ability of the biosensor developed using the CINC pipeline to detect binding in real-time we determined the kinetic parameters of M3 binding to MalX A134C-MDCC using the stopped-flow technique. In essence the stopped-flow apparatus used is a fluorescence spectrophotometer coupled with a rapid mixing device, enabling real-time fluorescence measurements of rapid biomolecular events. Here, MalX A134C-MDCC was rapidly mixed with increasing concentrations of M3, and the resulting fluorescence time-courses representing binding of M3 to the biosensor were recorded. Consistent with first-order binding kinetics the obtained fluorescence time courses were best fit with a one-exponential function, yielding the total fluorescence change (A_1) and the apparent rate of binding (k_{app}) (Eq. 2.7). Similar to the corresponding equilibrium fluorescence spectrophotometry data (see 3.3*)*, rapid mixing of M3 and MalX A134C-MDCC resulted in a fluorescence decrease (Fig. 2.4A, Fig. A1.3). As the observed k_{app} increases with M3 concentration, the association rate constant (k_{on}) for M3 binding to MalX A134C-MDCC was determined from the slope of k_{app} as a function of the M3 concentration

(Fig. 2.4B). The determined k_{on} of 20 ± 2 $\mu M^{-1}s^{-1}$ highlights the rapid M3 detection possible using our engineered biosensor. This further demonstrates that changes in the structural dynamics at sites distant from the ligand binding site can be used as rapid readout for the binding event and do not introduce delays. Rapid M3 detection by MalX A134C-MDCC makes the biosensor amenable to kinetic applications where M3 release is slower than detection by the biosensor (i.e. real-time detection of ligand formation by preceding enzymatic reaction). Interestingly, rapid kinetics analysis of M3 detection by MalX T249C-MDCC resulted in fluorescence time-courses that did not follow pseudo first-order binding kinetics and this biosensor was not utilized in rapid kinetics assays (data not shown) which further emphasizes the role of dihedral angle change as a key selection criterion for the labeling site. Although the complex kinetics observed with MalX T249C-MDCC make it less useful for real-time applications, this biosensor has utility in endpoint or equilibrium applications where sensitive MOS detection is required.

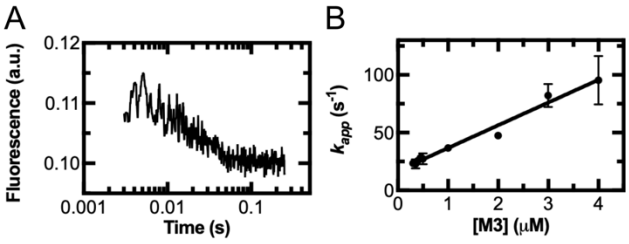


Fig. 2.4. Concentration dependence of the M3-association rate to MOS-detecting biosensor. (A) Representative fluorescence time-course of 2 μM M3 binding 100 nM MalX A134C-MDCC. (B) MalX A134C-MDCC (100 nM) fluorescence signal change across a variety of M3 concentrations (n = 4 for each data point, mean ± s.d.). When fit with a linear function the resulting slope yields an association rate constant of 20 ± 2 $\mu M^{-1}s^{-1}$ (n = 9, best fit ± s.d.).

2.5.5 CINC-developed MalX A134C can be conjugated to a variety of fluorophores yielding biosensor with select spectroscopic properties. To examine the influence that the fluorophore species have on our CINC-developed biosensors, we conjugated MalX A134C to a variety of fluorophores with various structural and spectroscopic properties. Ultimately, we examined the impact of different fluorophore families (four) and linker compositions (three) (Fig. A1.4). To address linker composition, the coumarin fluorophores (CPM and DACM) were conjugated to MalX A134C and compared to the previously characterized MDCC conjugate. MDCC contains an N-ethylcarbamoyl linker between the coumarin and maleimide group, CPM contains a phenyl linker between the coumarin and maleimide group, and DACM contains no linker (Fig. A1.4). To examine the impact of fluorophore family on the MalX A134C-based biosensor Dansyl (naphthalene family), Fluorescein (xanthene family), Pyrene (pyrene family), and Rhodamine Red (xanthene family) were conjugated to MalX A134C and their response to a MOS examined (Fig. 2.5, Fig. A1.4, Table A1.2). Direct excitation of all tested fluorophores gave a detectable response to M3 with changes in peak fluorescence intensity ranging from 5 – 29%, where differences in the magnitude of the reported fluorescence changes are likely due to the different shapes/ sizes of the utilized fluorophores probing slightly different environments. MalX A134C labeled with dansyl (which is capable of forming a FRET pair with tryptophan) also reported a fluorescence change upon M3 addition after excitation of tryptophan at 280 nm (Table A1.2), thereby expanding the spectral range of our MOS-biosensor set to report emission maxima of 380 nm to 580 nm. Together, these results demonstrate

that a wide variety of fluorophores can probe the altered environment detected via our CINC pipeline, enabling rapid development of biosensor libraries with variable spectroscopic properties facilitating multiplexing of such assays.

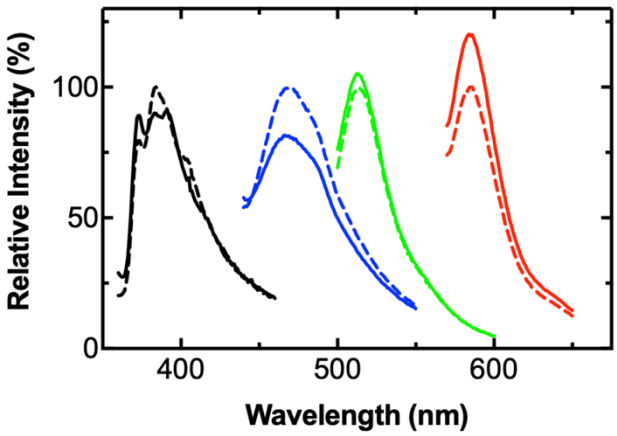


Fig. 2.5. MOS-detecting biosensor can be conjugated to a variety of fluorophores to alter fluorescence properties. Fluorescence emission spectra of MalX A134C conjugated to Pyrene (black), MDCC (blue), Fluorescein (green) or Rhodamine Red (red). Dashed spectra indicate MOS-detecting biosensor alone (20 nM), and solid spectra indicate MOS-detecting biosensor in the presence of MOS (10 μM).

2.5.6 Detection of MOS released from α-amylase-catalyzed starch degradation using MalX A134C-MDCC. To examine the utility of MalX A134C-MDCC in characterizing enzyme activity, we examined formation of MOS in real-time via *Bacillus licheniformis* α-amylase-catalyzed degradation of starch. Using the stopped-flow technique, all experimental conditions with α-amylase resulted in a time-dependence of the observed fluorescence signal decrease slower than the rate of M3 binding to MalX A134C-MDCC (Fig. 2.6A). The observed apparent rate of MOS generation is dependent on α-amylase concentration and fitting this dependence with a linear equation revealed a rate constant of 0.7 ± 0.1 $\mu M^{-1} min^{-1}$ (Fig. 2.6B), while the amplitude of the signal change was independent of α-

amylase concentration (Fig. 2.6C). To demonstrate portability of our MalX A134C-MDCC biosensor to different detection platforms, we demonstrate that the kinetics of the α-amylase concentration-dependent rate can also be determined using a commonly used microplate reader assay (Fig. 2.6D). In agreement with our stopped-flow assay, the apparent rate of MOS generation is dependent on α-amylase concentration and fitting with a linear function yielded a rate constant of 1.4 ± 0.5 $\mu M^{-1} min^{-1}$ (Fig. 2.6E), while the amplitude of the signal change was independent of α-amylase concentration (Fig. 2.6F). Together, these findings demonstrate the utility of the CINC-designed MalX A134C-MDCC biosensor in detecting MOS release as part of coupled enzymatic assays enabling the characterization of the functional cycle of α-amylase.

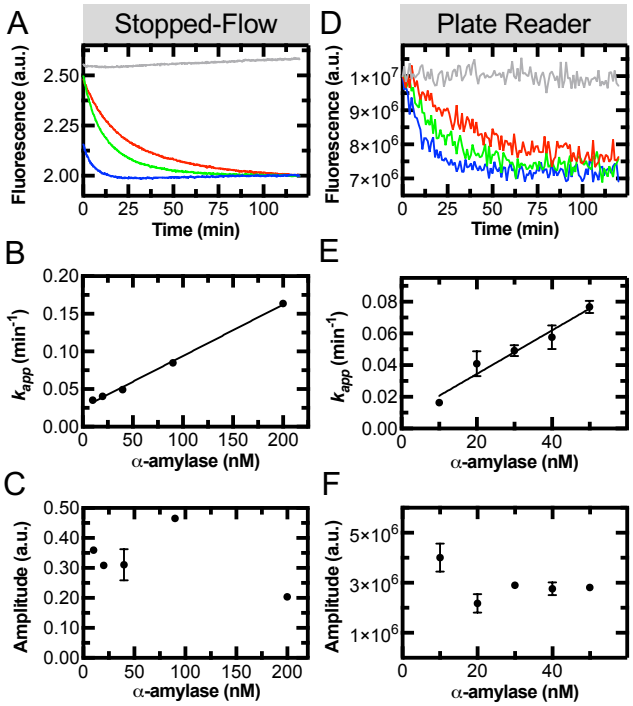


Fig. 2.6. MOS-release from α-amylase-catalyzed starch de-polymerisation reported by MOS-detecting biosensor. (A) Representative fluorescence-time courses of MalX A134C-MDCC (1.5 μM) in the presence of 100 mg/L starch and α-amylase (0 nM, grey; 40 nM, red; 90 nM, green; 200 nM, blue) using the stopped-flow method. (B) Concentration dependence of α-amylase on the apparent rate (k_{app}) using a stopped-flow assay, MOS release into solution occurs with a rate constant of 0.7 ± 0.1 $\mu M^{-1}min^{-1}$ (error reflects 95% c.i. in fit). (C) Amplitudes of MalX signal change remain constant as a function of α-amylase concentration using the stopped-flow method. (D) Representative fluorescence-time courses of MalX A134C-MDCC (1.5 μM) in the presence of 100 mg/L starch and α-amylase (0 nM, grey; 10 nM, red; 30 nM, green; 50 nM, blue) using a microplate reader. (E) Concentration dependence of α-amylase on apparent rate (k_{app}) using a microplate reader assay, MOS release into solution occurs with a rate constant of 1.4 ± 0.5 $\mu M^{-1}min^{-1}$ (error reflects 95% c.i. in fit). (F) Amplitudes of MalX signal change remain constant as a function of α-amylase concentration using microplate reader assay (n = 3 for each data point, error bars on individual data points represent s.d.).

2.6 Discussion:

Due to the vast structural and functional diversity of carbohydrates found in nature, the up-front costs of biosensor development using conventional approaches has limited the access to such advanced sensors for the majority of

targets. However, to determine substrate specificity and kinetics of a CAZyme, suitable assays must be available. Commonly, carbohydrate activity assays are often laborious, involve non-continuous steps (i.e. manual sampling and/or downstream chemical derivatization), non-specific (i.e. detect many carbohydrates at once) or are linked (i.e. depend on other enzymes or cofactors) and don't lend themselves well to high-throughput or multiplexing. Biosensor-based assays for real-time, specific detection of target carbohydrates will be advantageous and pivotal for unlocking the detailed enzymatic characterization of genome-encoded CAZymes (the CAZome). The CINC pipeline employed here allows researchers to circumvent these limitations by enabling a rapid computer-aided design and screening step to identify promising target positions for the subsequent introduction of fluorescence reporter groups, overcoming a critical hurdle for the streamlined (accelerated) and routine construction of novel biosensors for detection of user-defined reaction products.

As a proof of concept, we have demonstrated the utility of the CINC pipeline for the rational design of a novel biosensor capable of detecting M3. Of the four positions initially selected in MalX for substitution and subsequent fluorescent labeling based on their F_{Score}, two resulted in biosensors that gave a response to M3. Therefore, our current algorithm has a 50% success rate in identifying labeling positions that do not influence ligand binding properties. The described approach greatly minimizes time and cost of developing efficient, specific, and sensitive biosensors. As such the CINC-based design platform likely outperforms

conventional approaches for developing protein-fluorophore conjugates, which produce biosensors that respond to the desired ligand in 20-30% of designs – often with altered binding affinity when compared to the unmodified protein (e.g. as in (Brune et al. 1994; De Lorimier et al. 2002)). As well, the fluorescent labeling sites selected by CINC are unique compared to conventional approaches. For example, in static crystal structures of both *apo* and MOS-bound MalX (PDB 2XD2 and 2XD3), position A134 adopts a loop secondary structure. However, our short molecular dynamics simulations suggest that position 134 in *apo* MalX adopts a secondary structure that exists in equilibrium between an α-helix and a loop conformation (Fig. 2.1C). It is therefore beneficial to utilize CINC to highlight changes in local dynamics that otherwise would go unnoticed using conventional approaches. The CINC pipeline will have broad applicability for biosensor design, as it is based on first-principles, which in turn can be examined in a large number of biomolecules. However, additional testing with different protein scaffolds (data not shown) will further demonstrate the flexibility (i.e. general applicability to different protein scaffolds/ adaptability to different input signals) of the scoring algorithm. Results for the four tested MalX variants will enable improved selection criteria for the development of future biosensors utilizing the CINC pipeline, whereby parameters used to calculate F_{Score} can be weighed differently guided by the success/failure of candidate biosensors response to ligand. For example, the largest contributor to F_{Score} for position 134 in MalX are changes in dihedral angles – suggesting future iterations of CINC could emphasize changes in dihedral in the F_{Score} criteria. A future iteration of CINC prioritizing changes in backbone dihedral

angles may also improve the success rate of the algorithm, as large-scale changes in backbone dihedral angles were only seen in the successful candidate labeled at position 134 (Fig. 2.1 B-D). As well, further optimization of F_{Score} cut-off values may improve selection criteria in future iterations of CINC. For example, this report used F_{score} values > 1.5 which led to successful biosensor designs utilizing positions T249C and A134C for fluorophore conjugation. However, it may be possible to enhance selection criteria with a more stringent F_{Score} cut-off (e.g. $F_{Score} > 2.5$).

In addition to streamlining the identification and selection of non-disruptive labeling positions in a protein of interest remote from the ligand binding site, the successfully identified positions are also robust with respect to fluorophore selection. MalX A134C conjugated with various environmentally sensitive fluorophores was able to report M3 binding at different wavelengths across the visible spectrum. This highlights an additional advantage of the CINC pipeline over previously developed labeling approaches, as many non-CINC developed biosensors heavily favor a specific protein-fluorophore combination to provide detection of a molecule of interest (e.g. as in (Brune et al. 1994; De Lorimier et al. 2002)), whereas CINC has identified an altered environment that can be reported by a wide variety of fluorescent groups with different spectral properties. Therefore, our system lends itself to rapid construction of biosensors with user-defined spectroscopic properties and may be easily amenable to development of multiplexed biosensor assays due to the lack of a strong fluorophore preference in CINC-developed biosensors. Such multiplexed biosensor assays would be

desirable in a number of situations. For example, the detection of multiple carbohydrate ligands in complex mixtures of enzymes or microbes would be possible. In principle, this will enable evaluation of how multiple substrates are being processed simultaneously in environments such as the gut microbiome or in biofuel feedstocks.

With respect to the developed MOS-detecting biosensor, the fluorophore conjugation site was selected via small-scale changes in local dynamics distal from the MOS binding site. The kinetic properties of ligand binding for MalX A134C-MDCC are similar to previously developed phosphate-detecting biosensors based on *E. coli* phosphate binding protein (Table A1.3) ((Brune et al. 1994; Okoh et al. 2006; Smith et al. 2017) which has enabled detailed kinetic analysis of the timing and role of phosphate release in various systems (Kothe and Rodnina 2006; Muretta et al. 2015; Savelsbergh et al. 2005; Smith et al. 2017). In the current study, we examine MOS-release from *B. licheniformis* α-amylase via our MOS-detecting biosensor. Because our MOS-detecting biosensor binds MOS rapidly and with high affinity, we propose that the observed fluorescence change by the MOS-detecting biosensor is rate-limited by the availability of free MOS in solution produced via starch de-polymerization and subsequent dissociation of the formed MOS from the starch degrading enzyme. Because of the non-destructive measurement of MOS in real time, when combined with information about the chemical step of glyosidic bond cleavage (e.g. obtained via quench flow), these biosensors can be used to characterize the contributions of the cleavage step and

the timing of product release to the overall reaction cycle. Using prior approaches critical information about the catalysis and product release step cannot be deconvoluted, whereas our biosensor-based approach enables characterization of de-polymerization enzymes and their rational design, optimizing catalysis of de-polymerization relative to product release. The protein scaffold MalX is capable of binding MOS with degree of polymerization of 3-9 with similar affinity (Abbott et al. 2010), so it is likely that MalX A134C-MDCC similarly detects MOS with degrees of polymerization 4-6 and 8-9 in addition to the detection of M3 and maltoheptaose reported here. Together, the development of biomolecular tools like the MOS-detecting biosensor will provide an alternative solution to detect and discern between the formation of nascent carbohydrate in real-time and enable rapid characterization of CAZymes, including their enzymatic properties.

2.7 Conclusion:

We describe the CINC platform for development of novel protein-fluorophore conjugates capable of responding to an input signal, such as ligand binding, by changing their fluorescence intensity. CINC exploits small-scale changes in amino acid dynamics for biosensor design, which is advantageous for the design of biomolecular tools that do not undergo large-scale global conformational changes upon ligand binding. An additional advantage of CINC over conventional biosensor design approaches is that it does not rely on locating the fluorescent reporter group in close proximity to the ligand binding site, which often leads to proteins with altered binding properties relative to the wild-type

protein (ultimately leading to a smaller number of failed sensor designs). The CINC pipeline is rapid, and in most cases the time from conceptualization over final biosensor design and validation is less than one week. At this time, CINC uses high-resolution structural data for the protein scaffold in its *apo* and ligand-bound state. However, expansion of biosensor libraries using CINC can be further accelerated by coupling with next-generation protein structure modeling (Kuhlman and Bradley 2019) thereby circumventing the requirement for high-resolution structural data and increasing the number of protein scaffolds accessible to CINC by orders of magnitude.

2.8 Declaration of Competing Interest:

DDS, DG, DWA, and HJW are Joint Inventors on International PCT Patent Application No. CA2021/050110 (Filed January 30, 2021)

2.9 Acknowledgements:

This research was in part enabled by support provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca). We thank the University of Lethbridge for the use of SynBridge (Lethbridge, AB, Canada), which was used for fluorescence plate reader experiments. We would like to thank Dr. Ute Kothe for the use of the fluorescence spectrophotometer, Dr. Ata Ghavidel for critical feedback on MOS-release assays, and Emily Wilton for critical reading of the manuscript.

2.10 Chapter 2 Supplementary Information:

Supplementary information associated with Chapter 2 is available in Appendix 1.

CHAPTER 3: DEVELOPMENT OF A REAL-TIME PECTIC OLIGOSACCHARIDE-DETECTING BIOSENSOR USING THE RAPID AND FLEXIBLE COMPUTATIONAL IDENTIFICATION OF NON-DISRUPTIVE CONJUGATION SITES (CINC) BIOSENSOR DESIGN PLATFORM

3.1 Preface:

The contents of this Chapter have previously been published as: Smith, D.D., King, J.K., Abbott, D.W., and Wieden, H.-J. (2022). Development of a real-time pectic oligosaccharide-detecting biosensor using the rapid and flexible Computational Identification of Non-disruptive Conjugation sites (CINC) biosensor design platform. Sensors, 22(3), 948. My contribution to this work was I performed all computational and biochemical characterization reported in the manuscript and supervised an undergraduate student (Josh King) who performed preliminary assays to characterize the reported biosensors as part of his Independent Study courses (BCHM 2990, BCHM 3990). I also wrote the first draft of the published paper, which has been reformatted to meet the guidelines of this thesis.

In this manuscript, we refine our scoring algorithm in CINC, based on the results reported in Chapter 2. This research further illustrates the strength of our CINC approach that can be modified based on the successes (or failures) of previous designs. In this report, we demonstrate that our $F_{Score2.0}$ rank-scoring system outperforms the original F_{Score} algorithm used in Chapter 2 with respect to success rates in producing a second set of carbohydrate-detecting biosensors.

Manuscript authors: Dustin D. Smith[1,2], Joshua P. King[1,2], D. Wade Abbott[2,3], and Hans-Joachim Wieden[1,2,4]

Author affiliations:

[1]Alberta RNA Research and Training Institute (ARRTI), University of Lethbridge, Lethbridge, AB, Canada

[2]Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada

[3]Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, AB, Canada

[4]Department of Microbiology, University of Manitoba, Winnipeg, MB, Canada

Author Contributions (CRediT format):

Dustin D. Smith: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision, Project Administration

Josh P. King: Methodology, Investigation, Writing – Original Draft, Writing – Review & Editing

D. Wade Abbott: Conceptualization, Resources, Writing – Original Draft, Writing – Review & Editing

Hans-Joachim Wieden: Conceptualization, Methodology, Resources, Writing – Original Draft, Writing – Review & Editing, Supervision, Project Administration, Funding Acquisition

3.2 Abstract:

Fluorescently-labeled solute-binding proteins that change their fluorescent output in response to ligand binding are frequently used as biosensors for a wide range of applications. We have previously developed “Computational Identification of Non-disruptive Conjugation sites” (CINC), an *in silico* pipeline utilizing molecular dynamics simulations for the rapid design and construction of novel protein-fluorophore conjugate-type biosensors. Here we report an improved *in silico* scoring algorithm for use in CINC and its use in the construction of an oligogalacturonide-detecting biosensor set. Using both 4,5-unsaturated and saturated oligogalacturonides, we demonstrate that signal transmission from the ligand binding pocket of the starting protein scaffold to the CINC-selected reporter positions is effective for multiple different ligands. The utility of an oligogalacturonide-detecting biosensor is shown in Carbohydrate Active Enzyme (CAZyme) activity assays, where the biosensor is used to follow product release upon polygalacturonic acid (PGA) depolymerization in real time. Our oligogalacturonide-detecting biosensor set represents a novel enabling tool integral to the rapidly expanding platform for biosensor-based carbohydrate detection, whereas the CINC pipeline will continue to enable the rational design of

biomolecular tools to detect additional chemically distinct oligosaccharides and other solutes.

3.3 Introduction:

Biosensors are analytical devices that use biological components to detect target molecules and produce a measurable output. A popular approach to biosensor design is the conjugation of a fluorescent group to a solute-binding protein that harbors specificity for the target molecule, exploiting ligand-induced fluorescence changes in the protein-fluorophore conjugate to detect a target (Edwards 2021; Toseland 2013). These protein-based biosensors have been utilized for amino acid detection (Ciftci et al. 2020; De Lorimier et al. 2002; Ko et al. 2017; Ko et al. 2016), anion or cation detection (Belousov et al. 2006; Chen et al. 2013; De Lorimier et al. 2002; Smith et al. 2017; Solscheid et al. 2015), nucleotide detection (Berg et al. 2009; Bianchi-Smiraglia et al. 2017; Kunzelmann and Webb 2010), and carbohydrate detection (De Lorimier et al. 2002; Fonin et al. 2014; Höfig et al. 2018; Hsieh et al. 2012; Marvin et al. 2011; Smith et al. 2022a; Tian et al. 2007). Protein-based biosensors are often constructed by placing a fluorescent group in close proximity to the ligand binding site, exploiting ligand binding events to change the local environment probed by the fluorescent group. However, placement of the fluorescent group in the vicinity of the ligand-binding site can interfere with ligand binding, thereby reducing binding affinity or altering ligand specificity (e.g. as in (Brune et al. 1994; De Lorimier et al. 2002; Gilardi et al. 1994; Toseland 2013)). Further innovation is therefore required to move beyond

these trial-and-error-based approaches which are laborious and often disrupt the underlying binding properties of the protein scaffold.

To this end, we have previously developed CINC, an *in silico* screening step for rational selection of fluorophore conjugation sites distal from the ligand binding site (Smith et al. 2022a). CINC was originally designed based on observations in our previous work that binding of ligands to their respective binding sites altered localized dynamics at distal sites (De Laurentiis et al. 2016; Girodat et al. 2019; Mercier et al. 2015; Rosler et al. 2015). CINC utilizes explicit solvent molecular dynamics simulations of proteins in their *apo* and ligand-bound state to identify changes in localized dynamics as a consequence of ligand binding and maps these changes to the corresponding amino acids. Changes in amino acid dynamics in response to an input (i.e. ligand binding) are calculated and ranked using our in-house developed "F_{Score}" algorithm (Smith et al. 2022a). The aforementioned changes in amino acid dynamics upon ligand binding are then exploited for biosensor design via the placement of a fluorescent group at the selected site. To generate candidate biosensors, selected amino acid sites are substituted with cysteine and subsequently conjugated with a thiol reactive fluorophore. CINC was previously used to develop protein-based biosensors for detection of maltooligosaccharides (Smith et al. 2022a), and evaluation of the efficacy of prior design considerations has since been used to improve our scoring algorithm. Here we present a streamlined and improved version of the CINC pipeline scoring algorithm ($F_{Score2.0}$) informed by computational and experimental

data in our previous report (Smith et al. 2022a). Our prior study with CINC examined ligand-induced changes in Root Mean Squared Fluctuation (RMSF), changes in distance to tryptophan, changes in solvent accessibility, changes in proximity to ligand, and changes in backbone dihedral angles when scoring candidate fluorophore conjugation sites (Smith et al. 2022a). However, experimental data suggested that examination of changes in backbone dihedral angle dynamics were critical to successful biosensor design and could alone be used to improve the efficacy of CINC designs (Smith et al. 2022a), but this hypothesis had not been tested. The $F_{Score2.0}$ ranking algorithm emphasizes changes in backbone dihedral angle dynamics when scoring potential candidate biosensors to improve the efficacy of our design pipeline. We then use CINC to engineer TogB from *Yersinia enterocolitica* (Abbott and Boraston 2007b) for direct, specific, and continuous monitoring of both saturated and 4,5-unsaturated oligogalacturonides. In addition to testing the efficacy of CINC, use of TogB as a protein scaffold in the CINC pipeline presents a unique biosensor engineering challenge with respect to the ligand binding properties of TogB. Specifically, we investigate if CINC-detected changes in amino acid dynamics are consistent when the protein scaffold binds chemically distinct ligands within the same binding pocket. Such information will be useful for applications where the desired biosensors not only maintain the ligand-binding properties of the protein scaffold, but also must be capable of faithfully reporting the binding of the whole ligand range via signal transmission from the ligand-binding site to the reporter, producing a detectable output.

3.4 Materials and Methods:

3.4.1 Molecular dynamics simulations. Structural models of TogB in its *apo*, digalacturonic acid (digalUA)-bound state, 4,5-unsaturated digalacturonic acid (unsatdigalUA)-bound state, and trigalacturonic acid (trigalUA)-bound states were obtained from protein databank (2UVG, 2UVH, 2UVI, 2UVJ, respectively (Abbott and Boraston 2007b)). Each protein structure was solvated using a TIP3P water box extending 10 Å from all atoms in the protein and the system was neutralized with Na^+ ions. Potential energy of water was minimized using AMBERFF99S for 10 000 steps, followed by potential energy minimization of the entire system using AMBERFF99S and GLYCAM force fields (Case 2016; Kirschner et al. 2008; Singh et al. 2016). Subsequently the system was heated to 300 K for 10 000 steps, and molecular dynamics simulations were performed using the AMBER suite (Case 2016) for 100 ns (3 replicates each for *apo*, unsatdigalUA-bound, digalUA-bound, and trigalUA-bound states) at 2 fs step size where temperature was maintained using Langevin dynamics. Trajectories were combined using cpptraj (Roe and Cheatham III 2013), and the root-mean squared deviation (RMSD) for both *apo* and ligand-bound states of TogB remained stable after an initial equilibration (Figure A2.1).

3.4.2 Adaptation of CINC and development of $F_{Score2.0}$. The CINC pipeline examines small-scale changes in dynamic features of amino acid positions that impact a conjugated fluorophore. By design, the F_{Score} ranking system is flexible and can weigh changes in the aforementioned criteria differently guided by wet-

lab data. Because CINC is based on first-principles it has broad utility for applications (e.g. custom biosensor design) where high-resolution structural data of an input protein scaffold is available. Results of our previous study suggested that changes in dynamics of backbone dihedral angles between the *apo* and ligand bound state may alone be used as a metric for producing effective biosensors (Smith et al. 2022a). As well, our prior study using CINC had not examined algorithm predictive ability and consistency in cases where the protein scaffold bound multiple different types of ligands. To this end, the scoring algorithm used for the current study focuses entirely on changes in backbone dihedral angle dynamics, and data was examined for the *apo* vs. various different ligand bound states of TogB. Backbone dihedral angles were calculated using VMD (Humphrey et al. 1996) by determining ϕ (between atoms C-N-Cα-C) and ψ (between atoms N-Cα-C-N) angles throughout each simulation constructing a Ramachandran plot for each amino acid position. Each Ramachandran plot was transformed into a 180x180 matrix, and the sum of the absolute differences between the ligand-bound (B_{Lr}) and *apo* (B_{Ar}) matrices were determined for each amino acid position, resulting in a single value for each position representing the difference between the *apo* and ligand-bound state plot (performed using in-house developed R scripts). Values were normalized by dividing by the largest difference value in the data set, resulting in $F_{Score2.0}$ values ranging from 0 - 1 (Eq. 3.1). For TogB, amino acids directly involved in binding-site interactions with ligand (Abbott and Boraston 2007b) were not considered as candidates for fluorophore conjugation. $F_{Score2.0}$

values from each replicate were averaged, and averages as well as individual replicates were plotted using GraphPad Prism v. 9.0 (GraphPad Software).

$$F_{score2.0} = \frac{\sum|[B_{Lr}]-[B_{Ar}]|}{\vee\sum|[B_{Lr}]-[B_{Ar}]|} \quad \text{(Eq. 3.1)}$$

3.4.3 Construct design. All *togB* genes were engineered to lack the signal secretion peptide found in wild-type *togB* as previously described (Abbott and Boraston 2007b). Variants of *togB* were synthesized with flanking 5' *NdeI* and 3' *XhoI* restriction sites then subcloned into pET28a yielding pET28a::*togB K99C*, pET28a::*togB F247C*, pET28a::*togB A284C*, pET28a::*togB K362C*, and pET28a::*togB D363C* (BioBasic Canada Inc.). Genes for *yePL2b and yeGH28* were subcloned into the 5' *NheI* and 3' *XhoI* sites of pET28a, yielding pET28a::*yePL2b* and pET28a::*yeGH28*.

3.4.4 Overexpression and purification of TogB variants. LB + kanamycin media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 50 µg/mL kanamycin) was inoculated to an optical density at 600 nm (OD_{600}) ≈ 0.1 with *Escherichia coli* BL21(DE3) gold cells transformed with pET28a::*togb* variants. Cultures were incubated at 37°C with 200 RPM shaking until OD_{600} ≈ 0.6, then cooled to 16°C at 200 RPM shaking for one hour prior to induction with 300 µM isopropyl-β-D-thiogalactopyranoside (IPTG). Cultures were grown at 16°C with 200 RPM shaking for 16 hours prior to harvest by centrifugation (5000 × *g*, 15 minutes, 4°C).

Cells were resuspended in 7 mL of Buffer A per gram of cells (Buffer A: 20 mM Tris-Cl (pH 8.0 at 4°C), 500 mM NaCl, 10 mM imidazole, 10% glycerol, 7 mM β-mercaptoethanol, 1 mM phenylmethylsulfonylfluoride). Lysozyme was added to a final concentration of 1 mg/mL and the cell suspension was incubated on ice for 30 minutes with periodic inversion. The cell suspension was then mixed on ice for 5 minutes with 12.5 mg sodium deoxycholate per gram of cells. The mixture was sonicated (Branson Sonifier 450, Danbury, CT, USA) for 30 seconds at 50% output and 50% duty cycle (repeated once, with a 5-minute break between cycles). Insoluble cellular debris was pelleted by centrifugation (3 000 × *g*, 30 minutes, 4 °C), followed by further centrifugation of the supernatant to collect the S30 fraction (30 000 × *g*, 45 minutes, 4°C). S30 supernatant was loaded onto a 5 mL gravity flow column with Ni^{2+} Sepharose IMAC resin (GE Lifesciences) equilibrated with Buffer A. The column was subsequently washed three times with 3 column-volumes of Buffer A, and four times with 3 column-volumes of Buffer B (Buffer A with 20 mM imidazole) to remove weakly bound proteins. TogB protein variants were eluted in 5 mL fractions using Buffer C (Buffer A with 250 mM imidazole) and analyzed by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) with Coomassie Brilliant Blue staining. Fractions containing TogB variants were pooled and buffer exchanged to Buffer D (20 mM Tris-Cl pH 8.0 at 4°C, 30 mM imidazole, 500 mM NaCl, 10% glycerol) using VivaSpin 20 concentrator columns with 10 kDa molecular weight cut off (GE Lifesciences, 15 mL Buffer D added to 5 mL sample, concentrated to 5 mL, repeated 3 times). Purification yields were typically 20 - 100 mg of protein per liter of culture, and purity was typically

>95% based on ImageJ (Schneider et al. 2012) densitometry analysis of Coomassie Brilliant Blue Stained SDS-PAGE.

3.4.5 Overexpression and purification of YePL2b and YeGH28. Overexpression of YePL2b and YeGH28 was done using the same procedure as overexpression of TogB variants, except final concentration of IPTG used for induction was 200 μM (McLean et al. 2015). Purification procedure for YePL2b and YeGH28 was the same as the purification of TogB variants except fractions in Buffer C were buffer exchanged into Buffer E (20 mM Tris-Cl (pH 8.0 @ 20°C)) using dialysis (30 mL sample in 500 mL Buffer E, 4 changes, molecular weight cut off 12.4 kDa Sigma PN: D0530). Purification yields were typically 60 – 100 mg per liter of culture, and purity was typically >95% based on ImageJ (Schneider et al. 2012) densitometry analysis of Coomassie Brilliant Blue Stained SDS-PAGE. Protein concentrations were determined using extinction coefficients $\varepsilon_{280,\ YePL2b}$ = 114 835 $M^{-1}cm^{-1}$, and $\varepsilon_{280,\ YeGH28}$ = 68 425 $M^{-1}cm^{-1}$ calculated from primary sequence data using ExPASy ProtParam (Gasteiger et al. 2005). Purified proteins were flash frozen in liquid nitrogen and stored at -80°C for future use.

3.4.6 Fluorescent labeling of TogB variants. TogB variants (100 μM concentration in labeling reaction, 5 mL labeling reaction volume) were each incubated with 2 mL Ni^{2+} Sepharose IMAC resin in Buffer D. 7-Diethylamino-3-[N-(2-maleimidoethyl)carbamoyl]coumarin (MDCC; Sigma PN: 05019; 25 mM stock in dimethylformamide) was added at five-fold molar excess to each TogB variant,

corresponding to a 500 μM concentration in the labeling reaction. Labeling reactions were subsequently incubated at 4°C for 16 hours in an end-over-end mixer. Mixtures were centrifuged (500 × *g*, 2 minutes, 4°C) to collect Ni^{2+} Sepharose IMAC resin and supernatant removed. Ni^{2+} Sepharose IMAC resin was washed six times with three resin-volumes of Buffer D (500 × *g*, 2 minutes per wash, 4°C). Bound protein was eluted six times (500 × *g*, 2 minutes, 4°C) using 1 resin-volume of Buffer F (Buffer D with 250 mM imidazole) per fraction. Samples from the labeling procedure were examined using SDS-PAGE and the resulting gels imaged (460 nm light, Cy2 Filter, Amersham Imager 600, GE Lifesciences) to confirm the presence of the MDCC label prior to staining with Coomassie Brilliant Blue. Fractions containing the desired protein-fluorophore conjugate were pooled and dialyzed into Buffer G at 4 °C (50 mM Tris-Cl (pH 8.0 @ 20°C), 500 mM NaCl, 10% glycerol; 4 mL sample into 500 mL Buffer G, 3 changes, molecular weight cut-off 12.4 kDa Sigma PN: D0530). Recovery of protein from the labeling procedure was commonly ~50% and labeling efficiency ranged from 60 – 80%. Concentrations of MDCC-conjugated TogB variants were determined using spectrophotometry and using Equations 3.2 – 3.4. Parameters used were as follows: A_{280} is the absorbance at 280 nm, A_{430} is the absorbance at 430 nm, $\varepsilon_{280,\ TogB}$ = 90 300 $M^{-1}cm^{-1}$ is the extinction coefficient of TogB at 280 nm calculated using ExPASy ProtParam based on protein primary sequence (Gasteiger et al. 2005), 0.164 is used to correct for MDCC absorbance at 280 nm (Brune et al. 1994), *L* is instrument pathlength in cm, $\varepsilon_{430,\ MDCC}$ = 46 800 $M^{-1}cm^{-1}$ (Brune et al. 1994). Purified protein-fluorophore conjugates were flash frozen in liquid nitrogen

and stored at -80°C for future use. Preparations of the TogB D363C-MDCC Biosensor (PN: B1003) and custom biosensor design services using CINC are currently available from Allos Bioscience Ltd.

$$[TogB] = \frac{A_{280} - (A_{430} \times 0.164)}{\varepsilon_{280,TogB} \times L} \quad \text{(Eq. 3.2)}$$

$$[MDCC] = \frac{A_{430}}{\varepsilon_{430,\mathrm{MDCC}} \times L} \quad \text{(Eq. 3.3)}$$

$$Labeling\ Efficiency = \frac{Equation\ 3}{Equation\ 2} \times 100\% \quad \text{(Eq. 3.4)}$$

3.4.7 Carbohydrates. UnsatdigalUA was produced using methods similar to those described previously (Abbott and Boraston 2007a, b). Polygalacturonic acid (PGA, PN:P-PGACT, Megazyme) was dissolved in water at 20 mg/ml and dialyzed into water to remove small carbohydrate impurities (3.5 kDa molecular weight cut-off (Spectra/Por3), 50 mL sample into 2 L water, 2 changes). A 50 mL solution of 10 mg/mL PGA was digested overnight at 20°C with 1 μM YePL2b exopolygalacturonate lyase in 1 mM Tris-Cl (pH 8.0). The sample solution was evaporated to dryness using a SpeedVac, and re-dissolved in 2 mL water followed by addition of 8 mL of ethanol and 0.5 mL of acetic acid. The tube was then stored at 4°C for 24 hours, followed by centrifugation (14 000 × *g*, 10 minutes, 20°C). The pellets were washed with the same acidified ethanol solution and centrifuged (14 000 × *g*, 10 minutes, 20°C). Supernatants from the two centrifugations were pooled and evaporated to dryness using SpeedVac. Reaction products were examined via Thin Layer Chromatography using Silica 60 plates (Millipore) to confirm

production of unsatdigalUA (mobile phase 2:1:1 butanol: acetic acid: water, plates stained in 1% orcinol (PN: O1875, Sigma) in 70:3 ethanol: sulfuric acid, removed from stain and heated using Bunsen burner). Concentrations of unsatdigalUA were determined by mass (FW = 352.3 g/mol (Abbott and Boraston 2007b)), and confirmed in solution using spectrophotometry and $\varepsilon_{230,\ unsatdigalUA}$ = 5 200 $M^{-1}cm^{-1}$ (Abbott and Boraston 2007a; Shevchik et al. 1999). Preparations of unsatdigalUA are currently available from Allos Bioscience Ltd. (PN: C1002). TrigalUA (PN: T7407) and galacturonic acid (PN: 48280) were purchased from Sigma, and digalUA (PN: O-GALA2) was purchased from Megazyme.

3.4.8 Equilibrium fluorescence measurements. A Quanta Master 60 Fluorescence Spectrometer was utilized for fluorescence spectrophotometry measurements (Photon Technology International; excitation wavelength 420 nm, emission wavelength 440 – 520 nm, excitation slit widths: 3 nm, emission slit widths: 6 nm, step size: 1 nm, integration: 1 s). All equilibrium binding measurements were performed in Buffer H (50 mM Tris-Cl (pH 8.0 @ 20 °C), 500 mM NaCl) at 20 °C with MDCC-conjugated TogB variants at a concentration of 100 nM. Carbohydrate concentrations were at least three-fold higher than the previously reported affinity (K_D) values describing binding to TogB (Abbott and Boraston 2007b) and were as follows: unsatdigalUA: 16 μM, digalUA: 48 μM, trigalUA: 570 μM. Fluorescence emission spectra were plotted using GraphPad Prism v. 9.0 (GraphPad Software).

3.4.9 Rapid kinetics measurements. A KinTek SF-2004 (Kintek Corp.) rapid mixing device (stopped-flow apparatus) was used to collect rapid kinetics data. Experiments were performed at 20 °C, fluorescence was excited at 420 nm, and fluorescence emission was detected after a 450 nm long-pass filter (NewPort Corp.). All experiments in the stopped-flow apparatus were performed in Buffer H. Individual fluorescence time-courses were fit with a one-exponential function (Eq. 3.5), or a two-exponential function (Eq. 3.6), where F is the fluorescence observed at time t, F_∞ is the final fluorescence, A the signal amplitude and, k_{app} the apparent rate (TableCurve, Systat Software). To obtain K_D values, a hyperbolic function was fit to the data using GraphPad Prism v. 9.0 (Eq. 3.7), where Y indicates ligand binding at a given [carbohydrate] and B_{max} indicates maximum plateau value.

$$F = F_\infty + A_1 \times \exp\left(-k_{app1}t\right) \quad \text{(Eq. 3.5)}$$

$$F = F_\infty + A_1 \times \exp\left(-k_{app1}t\right) + A_2 \times \exp\left(-k_{app2}t\right) \quad \text{(Eq. 3.6)}$$

$$Y = \frac{B_{max} \times [carbohydrate]}{K_D + [carbohydrate]} \quad \text{(Eq. 3.7)}$$

3.5 Results:

3.5.1 Use of CINC to select fluorophore conjugation positions in TogB. TogB *apo*, TogB-unsatdigalUA, TogB-digalUA, and TogB-trigalUA were each subjected to 100 ns molecular dynamics simulations in triplicate and analyzed using the CINC pipeline to determine amino acid $F_{Score2.0}$ values. Small-scale changes in amino acid dynamics in the *apo* vs. ligand-bound states of TogB were evident when using

$F_{Score2.0}$, with several mid to high-scoring positions distal from the ligand binding site (Fig. 3.1). Conjugation of a fluorescent group distal to the ligand-binding site is preferred to modification of the binding site, as fluorophore conjugation near the binding pocket can lead to proteins with altered or reduced ligand affinity and/or specificity (e.g. as in (De Lorimier et al. 2002)). To validate the robustness of $F_{Score2.0}$ (i.e. for use in cases that may have subtle or drastic changes in amino acid dynamics), the calculated $F_{Score2.0}$ values were ranked (Fig. 3.1, data not shown) and five mid and high scoring positions were selected to construct candidate biosensors labeled with a thiol-reactive fluorophore (Fig. 3.2, Fig. A2.2, Fig. A2.3). This included the highest scoring position in TogB (363, $F_{Score2.0}$ = 1), which is located approximately 20 Å from the bound ligand in the unsatdigalUA-bound TogB structure (PDB 2UVI).

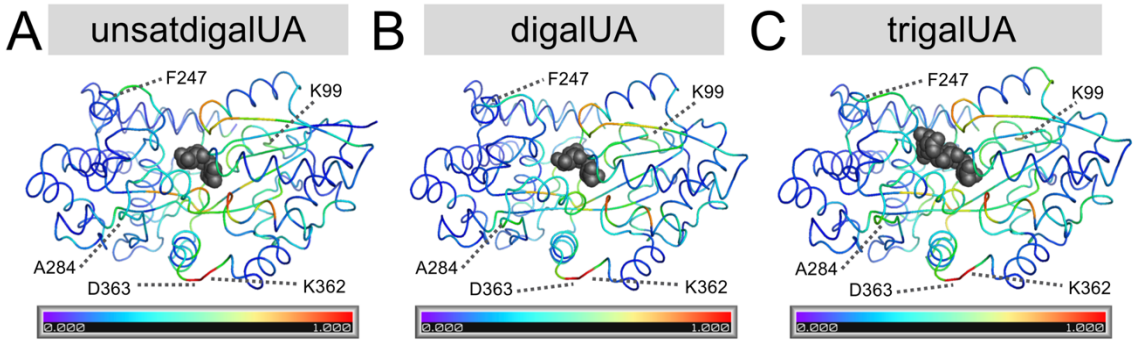


Fig. 3.1. Small-scale changes in amino acid dynamics upon ligand binding identified using $F_{Score2.0}$. Average $F_{Score2.0}$ values for each *apo* vs. ligand-bound state are projected onto their corresponding PDB structures: TogB bound to unsatdigalUA (A, PDB 2UVI), TogB bound to digalUA (B, PDB 2UVH), and TogB bound to trigalUA (C, PDB 2UVJ). Ligand is shown using grey spheres, and protein backbone is shown as ribbon coloured according to $F_{Score2.0}$ values. Candidate labeling positions selected based on $F_{Score2.0}$ values are also indicated.

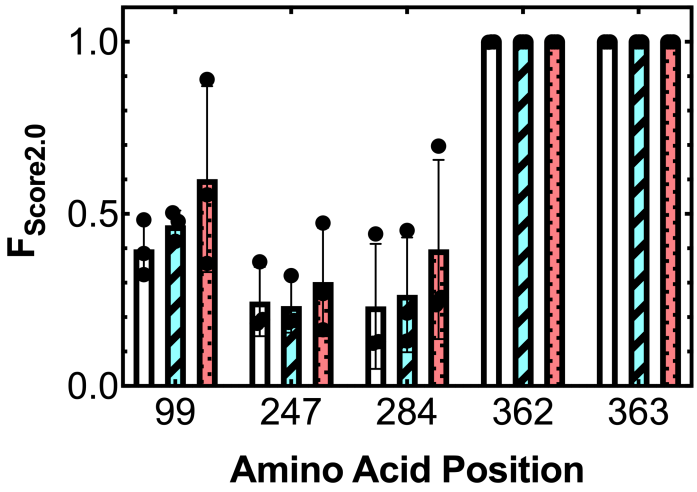


Fig. 3.2. $F_{Score2.0}$ values reflecting changes in dynamics of backbone dihedral angles at candidate labeling positions. Changes in dynamics of candidate labeling positions relative to dynamics in the TogB *apo* state are shown for TogB•unsatdigalUA (white bars), TogB•digalUA (cyan striped bars), and TogB•trigalUA (red dotted bars). Each bar reflects the results of 3 replicates ± s.d. with results from individual molecular dynamics simulations superimposed on the plot (black dots).

3.5.2 Oligogalacturonide detection by candidate biosensors. Each TogB variant was conjugated to MDCC, a diethylaminocoumarin that is commonly used in the construction of solute-binding protein-based biosensors due to its size, cost-effectiveness and low probability to affect solubility of the conjugate (Brune et al. 1994; Hanes et al. 2011; Kunzelmann and Webb 2009; Smith et al. 2022a; Smith et al. 2017). The MDCC conjugated TogB variants were then examined for their ability to detect unsatdigalUA, digalUA, and trigalUA using fluorescence spectrophotometry. Four of the five TogB-MDCC conjugates were able to selectively detect binding of the target carbohydrates and did not alter their fluorescence in the presence of the non-specific ligand, galacturonic acid (Table 3.1, Fig. A2.4, Fig. A2.5). To better understand the behaviour of the Biosensor as well as its scaffold protein, and dependent on the downstream application, its

performance under a wide range of physicochemical conditions (e.g. altered pH, ionic strength and temperature) or in the presence of additional analytes in complex mixtures needs to be characterized. Together, these results demonstrate that examining changes in dihedral angle amino acid dynamics in the CINC pipeline alone is effective at rapidly informing biosensor rational design and streamlining biosensor development.

Table 3.1. Response of fluorescently labeled TogB variants to ligand. Fluorescently labeled TogB variants were incubated in the absence and presence of saturating concentrations of unsatdigalUA, digalUA, and trigalUA (saturating concentrations defined as ligand concentration at least three-fold above previously reported dissociation constants (Abbott and Boraston 2007b)). Labeled TogB variants were also incubated in the absence and presence of a non-specific carbohydrate galacturonic acid. Values reported indicate percentage change in peak fluorescence intensity after addition of ligand (n = 1).

	w/ 16 μM unsatdigalUA	w/ 48 μM digalUA	w/ 570 μM trigalUA	w/ 1710 μM galacturonic acid
TogB K99C-MDCC	+2%	-4%	-4%	-2%
TogB F247C-MDCC	-14%	-10%	-20%	-1%
TogB A284C-MDCC	-31%	-31%	-30%	-1%
TogB K362C-MDCC	-32%	-25%	-29%	-2%
TogB D363C-MDCC	-60%	-39%	-44%	-1%

3.5.3 Rapid kinetics of unsatdigalUA and digalUA detection by TogB D363C-MDCC. TogB D363C-MDCC, the highest CINC scorer, displayed the largest fluorescence change in response to the target carbohydrates (Table 3.1, Fig. A2.4). However, for use in CAZyme characterization assays a detailed knowledge of the underlying kinetics of ligand binding to TogB D363C-MDCC is required. Kinetic parameters for ligand binding to TogB D363C-MDCC were determined using a the stopped-flow method, a rapid mixing device coupled with a fluorescence spectrophotometer which enables real-time monitoring of

biomolecular events. In agreement with equilibrium-state fluorescence data (*vide supra*), mixing of TogB D363C-MDCC with unsatdigalUA (Figure 3.3A) or digalUA (Figure 3.3B) resulted in a fluorescence decrease. The resulting fluorescence time-courses were best fit with a one-exponential function to obtain A_1 and k_{app} (Eq. 3.5). A_1 values of the fluorescence time courses obtained at increasing concentrations of unsatdigalUA (Fig. 3.3C) or digalUA (Figure 3.3D) were plotted against ligand concentration and a hyperbolic function (Eq. 3.7) was fit to the data to determine K_D = 1.3 ± 0.5 μM for unsatdigalUA and K_D = 6 ± 1 μM for digalUA. The K_D value obtained for unsatdigalUA binding to TogB D363C-MDCC is similar to affinity values for the unmodified protein determined previously via isothermal titration calorimetry (K_D = 5 ± 1 μM) or UV-difference spectroscopy (K_D = 3.2 ± 0.1 μM) (Abbott and Boraston 2007b). As well, the K_D determined for digalUA binding to TogB D363C-MDCC is comparable to previously reported binding affinity data determined for the unmodified protein using isothermal titration calorimetry (K_D = 16 ± 2 μM) and UV-difference spectroscopy (K_D = 11.8 ± 0.5 μM) (Abbott and Boraston 2007b). Apparent rate values obtained from the fluorescence time-courses were plotted against ligand concentration and fit with a linear function to determine k_{on} = 18.6 ± 0.7 $\mu M^{-1}s^{-1}$ for unsatdigalUA (Fig. 3.3E) and k_{on} = 6 ± 1 $\mu M^{-1}s^{-1}$ for digalUA (Fig. 3.3F). TrigalUA binding to TogB D363C-MDCC was also examined using the stopped-flow method, but under the conditions tested binding was likely too fast and occurred in the dead-time of the stopped-flow apparatus (data not shown). Together, these results demonstrate that the CINC pipeline is robust in selecting fluorophore conjugation sites that do not disrupt ligand binding,

and the resulting TogB D363C-MDCC biosensor is capable of rapid, sensitive ligand detection.

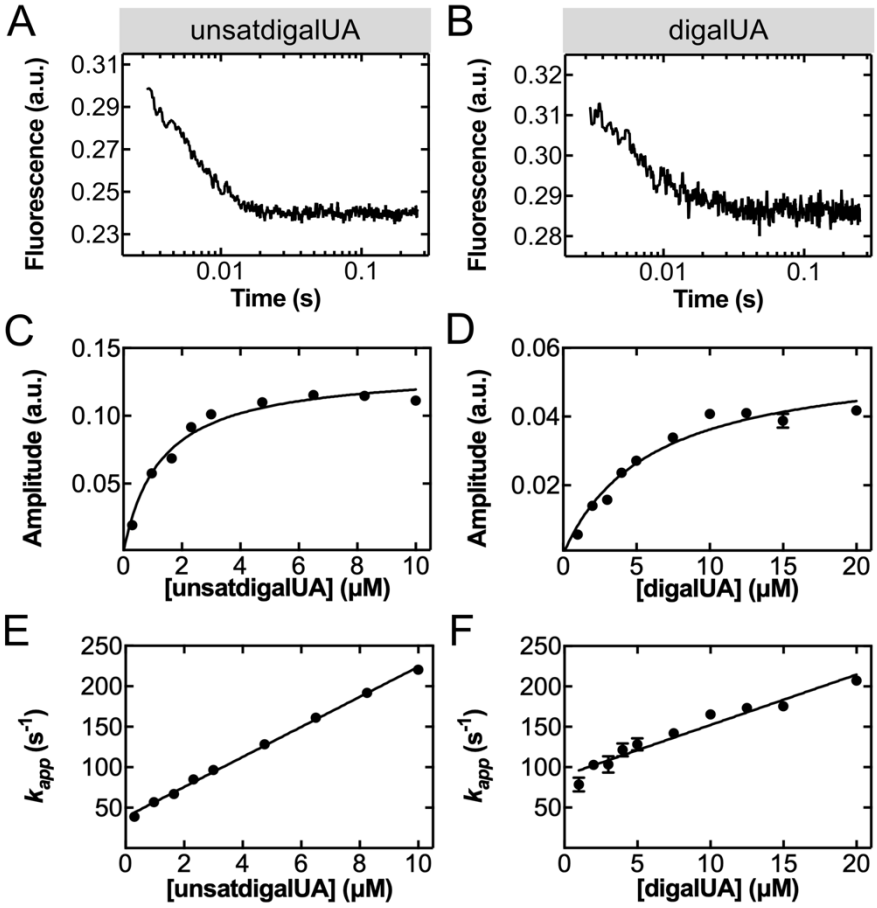


Fig. 3.3. Concentration dependence of unsatdigalUA-association rate and digalUA-association rate. Representative fluorescence time-course of 100 nM TogB D363C-MDCC binding 10 μM unsatdigalUA (A) or 10 μM digalUA (B). Fluorescence time-courses were obtained for 100 nM TogB D363C-MDCC binding to ligand across a range of carbohydrate concentrations (0.3 – 10 μM for unsatdigalUA, and 1 – 20 μM for digalUA). Fluorescent-time courses were fit with a one-exponential function (Eq. 3.5) to determine Amplitude and k_{app}. Amplitudes of signal change were plotted against concentrations of unsatdigalUA (C) and digalUA (D) and fit with a hyperbolic function (Eq. 3.7) to determine dissociation constant (K_D = 1.3 ± 0.5 μM for unsatdigalUA, K_D = 6 ± 1 μM for digalUA). k_{app} was plotted against concentrations of unsatdigalUA (E) and digalUA (F) and fit with a linear function to determine association constants (k_{on} = 18.6 ± 0.7 $\mu M^{-1}s^{-1}$ for unsatdigalUA, and 6 ± 1 $\mu M^{-1}s^{-1}$ for digalUA). Each data point in panels C – F reflects mean ± s.d. at the indicated carbohydrate concentration (n = 3).

3.5.4 Detection of oligogalacturonide release from a polysaccharide lyase and a glycoside hydrolase. To demonstrate the utility of the TogB D363C-MDCC biosensor for characterizing enzyme activity, we examined oligogalacturonide release from CAZyme-catalyzed degradation of PGA in real time. YePL2b is an exo-acting polysaccharide lyase from *Y. enterocolitica* that cleaves PGA using β–elimination producing unsatdigalUA as the major product (Abbott and Boraston 2007a; McLean et al. 2015). Using the stopped-flow method, a solution containing TogB D363C-MDCC and YePL2b was rapidly mixed with PGA, and the resulting fluorescent time-courses were recorded and best fit with a two-exponential function (Fig. 3.4, Eq. 3.6). The observed fluorescence decrease represents the real-time detection of the unsatdigalUA as in the absence of either PGA, or YePL2b, the fluorescence of TogB D363C-MDCC remains unchanged over the time-course of the measurement (Fig. 3.4A). The values determined from the aforementioned data reveal a rapid initial burst phase (k_{app1} = 39 $\pm$ 8 s^{-1}), consistent with the first round of product release from the CAZyme. Whereas the slower second phase (k_{app2} = 0.033 $\pm$ 0.005 s^{-1}) likely represents a multiple turnover phase. The multiple turnover phase is slower than the burst phase as it requires the cleaved (now shorter) PGA to reoccupy the +1 and +2 subsites of YePL2b (Abbott and Boraston 2007a) after cleavage of the (longer) PGA and the dissociation of the unsatdigalUA product (Table 3.2).

To demonstrate that the TogB D363C-MDCC biosensor can also detect saturated oligogalacturonide products generated by PGA hydrolysis (i.e. digalUA),

we investigated product release from YeGH28. YeGH28 is an exo-acting glycoside hydrolase from *Y. entercolitica* that hydrolyzes PGA producing digalUA as the major product (Abbott and Boraston 2007c; Liao et al. 1999). Using the stopped-flow method, we were able to observe, similar to YePL2b, a bi-phasic fluorescence decrease in the presence of TogB D363C-MDCC, YeGH28, and PGA that is not observed in the negative control conditions (Fig. 3.4B). Fitting the obtained fluorescence time-courses with a two-exponential function also showed an initial burst phase (k_{app1} = 33 ± 1 s^{-1}), likely representing the first round of hydrolysis and product release, followed by a slower multiple turnover phase (k_{app2} = 0.0210 ± 0.0005 s^{-1}; Table 3.2). Together, these findings demonstrate that TogB D363C-MDCC is a robust detection platform for both unsatdigalUA and digalUA and can be used in complex solutions to examine enzyme activity in real-time.

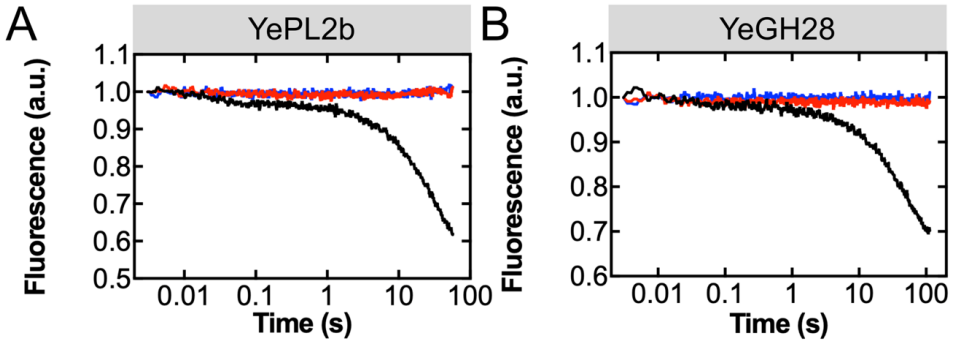


Fig. 3.4. Oligogalacturonide-release from CAZyme-catalyzed degradation of polygalacturonic acid detected by TogB D363C-MDCC. Representative fluorescence time-courses for product released by 250 nM YePL2b (A, Black fluorescence time-course) or 250 nM YeGH28 (B, Black fluorescence time-course) in the presence of 0.5 mg/L PGA detected by 250 nM TogB D363C-MDCC. Negative controls in the absence of CAZyme (Red fluorescence time-courses) and in the absence of PGA (Blue fluorescence time-courses) are shown.

Table 3.2. Oligogalacturonide release fit parameters obtained via CAZyme catalyzed PGA degradation. Fitting of a two-exponential function (Eq. 3.6) to biphasic fluorescence time-courses of oligogalacturonide release shown in Figure 3.4 detail underlying enzyme kinetic parameters. Fit parameters for a polysaccharide lyase (YePL2b) and a glycoside hydrolase (YeGH28) are reported (mean ± s.d., n = 6 replicates for each enzyme).

	A_{inf} (a.u.)	A_1 (a.u.)	k_1 (s^{-1})	A_2 (a.u.)	k_2 (s^{-1})
YePL2b	0.545 ± 0.02	0.05 ± 0.03	39 ± 8	0.37 ± 0.03	0.033 ± 0.005
YeGH28	0.565 ± 0.002	0.03 ± 0.01	33 ± 1	0.30 ± 0.01	0.021 ± 0.001

3.6 Discussion:

In this report, we expand on our carbohydrate-detecting biosensor toolkit using CINC and demonstrate the utility of their use in CAZyme activity assays. CAZymes are a group of structurally and functionally diverse enzymes that modify linkages or decorations in carbohydrates (Lombard et al. 2013), and are involved in a large number of industrial processes (Polaina and MacCabe 2007). Improved methods for carbohydrate detection will enable detailed kinetic analysis of CAZymes, thereby improving their implementation in industrial processes. We utilized CINC to engineer a biosensor based on TogB from *Y. enterocolitica* for the detection of breakdown products of PGA. TogB is the solute-binding protein of an ATP-binding cassette transporter (ABC transporter), where solute binding occurs at the interface of the N- and C-terminal domains of TogB (Abbott and Boraston 2007b). Structural studies indicate that upon solute binding TogB undergoes a hinge-bend motion where the N- and C-terminal domains move towards each other (Abbott and Boraston 2007b). TogB binds both unsatdigalUA and digalUA with micromolar affinity, and has weaker affinity for trigalUA (Abbott and Boraston 2007b). The binding specificity of TogB was ideal for construction of a biosensor to detect breakdown products of PGA, as TogB can bind major breakdown

products produced by both polysaccharide lyases (unsaturated products, e.g. as in (Abbott and Boraston 2007a; McLean et al. 2015)) and glycoside hydrolases (saturated products, e.g. as in (Abbott and Boraston 2007c; Liao et al. 1999)). An engineered biosensor set based on TogB can therefore enable the detailed kinetic analysis of two different CAZyme classes capable of depolymerizing PGA using different catalytic mechanisms. In general, prior approaches to detect carbohydrates are limited because they are not selective (detect many different carbohydrates in a single reaction), do not detect in real-time, requiring endpoint derivatization, and do not scale well for high-throughput (e.g. require linked enzyme assays or utilize low-throughput techniques such as mass spectroscopy) (Harvey 2011). Saturated oligogalacturonide detection methods suffer from many of the aforementioned limitations (Ström and Williams 2004), whereas detection methods for unsaturated oligogalacturonides can mitigate some of these drawbacks via absorbance-based readout of 4,5-unsaturation in the target molecule(s) (Abbott and Boraston 2007a; Shevchik et al. 1999). The TogB-based biosensor circumvents all of these limitations and represents a first-in-class method for the real-time detection of oligogalacturonides released from CAZymes.

An underlying challenge in developing biosensor-based assays for CAZyme product detection is inefficiencies in initial biosensor development. However, the CINC platform can enable the on-demand design of biosensors for specific targets based on simple protein scaffolds. The widespread availability of biosensor-based assays where detection of specific carbohydrates occurs in real time will be

transformative in elucidating and harnessing the enzymatic properties of genome-encoded CAZymes, and the CINC pipeline reported here will allow researchers to circumvent prior bottlenecks in the biosensor design process by enabling rapid computer-aided design and screening. Of the five candidate biosensors selected using CINC, four resulted in biosensors that were capable of detecting the target carbohydrates. With a success rate of 80%, CINC using $F_{Score2.0}$ outperforms prior approaches of selecting fluorescent labeling positions based on structural data whose designs typically respond to the desired ligand 20-30% of the time and often result in altered binding affinity when compared to the unmodified protein (e.g. as in (Brune et al. 1994; De Lorimier et al. 2002)). The $F_{Score2.0}$ algorithm also outperforms the prior version of F_{Score}, which had a success rate of 50% (Smith et al. 2022a). Users of CINC could further improve the success rate by increasing stringency and selecting only high-scoring candidate labeling positions based on changes in dihedral angle dynamics. The two highest scoring positions in this study (Position 362, $F_{Score2.0} = 0.9997 \pm 0.0002$; and Position 363 $F_{Score2.0} = 1$) resulted in functional biosensors with large fluorescence changes in response to ligand binding. In the case of the lower-scoring variants tested (Positions 99, 247, 284), the $F_{Score2.0}$ values of individual simulations showed more variability than the highest scorers (Fig. 3.2). This increased spread of $F_{Score2.0}$ values for the lower-scoring positions was due to partial shifts in the equilibrium between different microstates in the *apo* vs. ligand-bound simulations. For example, Position 284 largely exists in two different microstates for both the *apo* and ligand-bound simulations (Fig. A2.3). Upon ligand binding, occupancy of these states is partially

shifted – altering the amino acid dynamics of this individual position in a subset of the single molecules. Whereas the high $F_{Score2.0}$ value for Position 363 is due to a complete shift in equilibrium of amino acid dynamics, stabilized by ligand binding (Fig. A2.2). We therefore speculate that complete (or near-complete) shifts in microstate occupancy in *apo* vs. ligand-bound states are generally preferred to partial shifts. Additionally, it is important to note that it is difficult to speculate on the occupancy of specific microstates at ambient temperatures without molecular dynamics simulations. For example, in the case of Positions 362 and 363 in TogB, the starting X-ray crystal structures demonstrate different dihedral angles when comparing the apo vs. ligand-bound states. However, these static X-ray crystal structures represent local energy minima and can be constrained by contacts within the crystal lattice. It was therefore critical to examine local structural dynamics and dihedral conformations in an environment that more accurately reflected physiological conditions (e.g. at room temperature and solvated) where the biomolecule was provided with additional energy and the opportunity to explore alternative conformations. Consistent with this, we did indeed observe in several positions additional populations of dihedral angles not present in the starting structures (e.g. see Fig. A2.3). For some protein scaffolds, complete microstate shifts in the *apo* vs. ligand-bound state may not be present for any amino acid positions (for example, previous CINC-designed maltooligosaccharide-detecting biosensors were based on a partial shift in microstates (Smith et al. 2022a)). It is therefore important to note that our work demonstrating the design of biosensors based on partial shifts in microstate populations of *apo* vs. ligand-bound

conformations will enable construction of biosensors for a broad range of protein scaffolds.

One of the candidate TogB-based biosensors, TogB K99C-MDCC, did not detect any of the target carbohydrates *in vitro* (Table 3.1). We hypothesise that although Position 99 was a mid-range scorer, the change in backbone dihedral angles may not have positioned a conjugated fluorophore into an altered environment long enough (Fig. 3.2, Fig. A2.6). It is important to note that the current version of CINC does not quantify residence-times in the respective sub-states (including ones with different dihedral angles), but rather reports a probability that the dihedral angles differ upon ligand binding. Because the experimentally determined fluorescence changes are ensemble measurements this might result in averaging of the fluorescence of the different states preventing their detection *in vitro*. Of the four engineered TogB-based biosensors that reported a fluorescence change upon ligand binding, all four detected unsatdigalUA, digalUA, and trigalUA (Table 3.1). With prior structure-based approaches for biosensor design, the fluorescent group is often placed in close proximity to the ligand-binding site relying on direct interaction between the ligand and fluorescent group in order to elicit a fluorescence change. Using CINC, no such direct fluorophore-ligand interactions exist as the fluorophore is positioned distal from the ligand binding site. Of particular importance to the design of a TogB-based biosensor is that TogB completely internalizes its target ligands upon binding (Abbott and Boraston 2007b), complicating conventional biosensor design

approaches utilizing fluorophore-ligand interaction. CINC-designed biosensors are able to overcome this challenge as they do not rely on fluorophore-ligand interactions in order to produce a fluorescence change. Instead, CINC-designed biosensors focus on changes in amino acid dynamics at locations distal from the ligand binding site. On this background it was important to determine if signal transmission from the binding site to reporter positions occurs similarly for each ligand-bound state of TogB. Specifically, the ligand-bound states of TogB differ in that they include a discriminatory interaction between the unsaturated ligand and S271 of TogB when compared to binding of the saturated ligands. As well as an additional hydrogen bond between disaccharides and Y276 of TogB when compared to the trigalUA-bound complex (Abbott and Boraston 2007b). Indeed, the changes in amino acid dynamics identified via CINC in the TogB *apo* vs. ligand bound states were capable of reporting ligand binding despite the different interactions present in the TogB•unsatdigalUA, TogB•digalUA and TogB•trigalUA complexes. Our aforementioned *in silico* predictions were validated with *in vitro* characterization of the TogB-based biosensors and demonstrated that signal transmission from the ligand binding site to the CINC-designed reporter sites is likely as reliable as directly sensing the bound ligand.

In addition to streamlining the *in silico* CINC pipeline and improving its success rate via the implementation of $F_{Score2.0}$, this work further demonstrates that biosensors produced by this pipeline do not perturb the ligand binding properties of the starting protein scaffold. The TogB D363C-MDCC biosensor was

engineered based on localized amino acid dynamics changes distal from the ligand binding site, and the engineered fluorescent protein conjugate has ligand binding affinity comparable to the unmodified protein (*vide supra*). The ligand binding properties of TogB D363C-MDCC are consistent with previously developed solute binding protein-based biosensors for the detection of phosphate and maltooligosaccharides (Brune et al. 1994; Okoh et al. 2006; Smith et al. 2022a; Smith et al. 2017) (Table A2.1). The previously developed phosphate-detecting biosensor has facilitated detailed mechanistic investigations into the timing and function of phosphate release in various systems (Kothe and Rodnina 2006; Muretta et al. 2015; Savelsbergh et al. 2005; Smith et al. 2017), whereas the maltooligosaccharide-detecting biosensor has enabled detailed kinetic analysis of CAZymes that release maltooligoaccharides as part of their functional cycle (Smith et al. 2022a). In the current report, we examine unsatdigalUA release by a polysaccharide lyase (YePL2b) and digalUA release by a glycoside hydrolase (YeGH28) during PGA digestion using TogB D363C-MDCC. TogB D363C-MDCC binds unsatdigalUA and digalUA with high affinity and rapid rates. We therefore speculate that observed fluorescence decreases by TogB D363C-MDCC are rate-limited by the availability of free oligogalacturonides in solution produced via CAZyme-catalyzed PGA degradation. Together with the previously developed MOS-detecting biosensors and others, the oligogalacturonide-detecting biosensors reported here are novel enabling tools integral to our emerging biosensor platform for rapid characterization of CAZymes. Moving forward, biosensor-based assays for the detection and discrimination between

carbohydrates that may differ with minor chemical variances (e.g. degree of polymerization, chemical modification, branch points) are amenable to multiplexing and high-throughput screening assays (e.g. for simultaneous detection of multiple carbohydrates using biosensors with different fluorescence properties). Alternatively, the design of microfluidic devices with multiple carbohydrate-detecting biosensors in succession would circumvent the need for biosensors to have distinct fluorescence properties for detailed analysis of carbohydrate content. Such assays would have clear utility in rapid testing experiments and enable characterization of complex carbohydrate-containing solutions and/or metabolic processes, such as biofuel feedstock metabolism or the gut microbiome. Furthermore, this class of biosensors also provides the additional capability of distinguishing between the formation of a nascent carbohydrate and its release into bulk solution. For example, when combined with information about the chemical step of glyosidic bond cleavage (e.g. using quench flow), one can distinguish between the chemical cleavage step and the release of product from a CAZyme site. Which in turn would provide critical information for detailed kinetic analysis of de-polymerization enzymes paramount to their rational design. An additional future consideration with the CINC pipeline is the incorporation of fluorescent protein(s) at one or more positions with high $F_{Score2.0}$ values rather than small fluorescent dyes. For example, a recent report has utilized molecular dynamics simulations to propose that the sensing ability of a FRET-based glucose-detecting biosensor is due to careful positioning of the donor fluorescent protein (Reinartz et al. 2021). It would therefore be beneficial to develop fluorescent

protein-based biosensors guided by CINC for either FRET-based readouts or single-labeled biosensors using circular permuted fluorescent proteins (Kostyuk et al. 2019).

3.7 Conclusions:

The CINC pipeline is capable of detecting and exploiting small-scale changes in amino acid dynamics between the *apo* and ligand-bound states of an input protein scaffold for biosensor design. The CINC pipeline is rapid, simple, and has general applicability in the design of novel protein-fluorophore conjugates. In terms of success rates of designs, CINC using $F_{Score2.0}$ outperforms conventional biosensor design methods and the prior F_{Score} algorithm used in CINC. A current limitation of CINC is its dependence on high resolution structural data of a protein in its *apo* and ligand-bound state. However, moving forward expansion of biosensor libraries can be accelerated by coupling CINC with next-generation protein modelling, such as AlphaFold (Jumper et al. 2021), circumventing the current requirement for high-resolution structural data for each protein examined.

3.8 Declaration of Competing Interest:

DDS, DWA, and HJW are Joint Inventors on International PCT Patent Application No. CA2021/05/01/10 (Filed January 30, 2021).

3.9 Acknowledgements:

This research was in part enabled by support provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca). We thank the University of Lethbridge for the use of SynBridge (Lethbridge, AB, Canada), which was used for imaging of SDS-PAGE gels. We would like to thank Dr. Ute Kothe for use of the fluorescence spectrophotometer, Dr. Dylan Girodat for feedback on molecular dynamics data and his role in the development of the CINC biosensor design pipeline/ F_{Score} algorithm, Dr. Xing Xiaohui for preparation of unsatdigalUA, Dr. Darryl Jones for his input on oligogalacturonide-release assays, Dr. Davinder Kaur Dhalla for feedback on molecular dynamics data, and Dr. Harland Brandon/ Emily Wilton for critical reading of the manuscript.

3.10 Chapter 3 Supplementary Information:

Supplementary Information for Chapter 3 can be found in Appendix 2

CHAPTER 4: RE-ENGINEERING OF A COMPUTATIONALLY-DESIGNED BIOSENSOR SWITCHES SIGNALLING MODE AND IMPROVES DYNAMIC RANGE

4.1 Preface:

Chapter 4 has been written as a format-neutral manuscript for submission as a brief communication letter (or technical note) to, for example, a journal such as *Computational and Structural Biotechnology Journal* or *PLoS Computational Biology*. This chapter encompasses our design, and initial characterization of biosensors based on MalX A134C-MDCC which have been modified to create a broad range of sensor binding affinities. As well, some of these modifications were shown to alter the underlying mechanism by which the MalX-based biosensors respond to ligand. Taken together with the data in Chapters 2 and 3, our results suggest additional improvements to the CINC workflow to maximize its utility in rational biosensor design. I performed all wet lab and computational experiments reported in this chapter and wrote the first draft of the manuscript.

Authors: Dustin D. Smith[1,2], D. Wade Abbott[2,3], and Hans-Joachim Wieden[1,2,4]

Affiliations: [1]Alberta RNA Research and Training Institute (ARRTI), University of Lethbridge, Lethbridge, AB, Canada

[2]Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada

[3]Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, AB, Canada

[4]Current Address: Department of Microbiology, University of Manitoba, Winnipeg, MB, Canada

Projected Author Contributions upon Manuscript publication (CRediT format):

Dustin D. Smith: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision, Project Administration

D. Wade Abbott: Conceptualization, Resources, Writing – Original Draft, Writing – Review & Editing

Hans-Joachim Wieden: Conceptualization, Methodology, Resources, Writing – Original Draft, Writing – Review & Editing, Supervision, Project Administration, Funding Acquisition

4.2 Abstract:

A fluorescent, computationally-designed maltooligosaccharide (MOS)-detecting biosensor was re-engineered to both alter its ligand-binding affinity and reverse the underlying sensing mechanism. The dynamic range of the biosensor was expanded through the introduction of a series of amino acid substitutions in the starting protein scaffold (MalX from *Streptococcus pneumoniae*), which resulted in a biosensor set with binding affinities spanning over five orders of magnitude. The impact of the introduced substitutions on the underlying mode

signal generation was screened *in silico* using our previously reported Computational Identification of Non-disruptive Conjugation sites (CINC) pipeline. CINC utilizes structural molecular dynamics simulations to examine protein molecular dynamics with amino acid-level resolution. Using CINC, we demonstrate that re-engineering of our MOS-detecting biosensor set resulted in biosensors with two distinct output modes which differed based on local conformational changes at the fluorescently modified reporter position (readout based on removal of a unique *apo* conformation versus readout based on addition of a unique ligand-bound conformation). Together, these results suggest additional improvements to the CINC workflow which maximize its utility in rational biosensor design.

4.3 Introduction:

Biosensors are analytical devices used for the detection of target ligands. A common mode of biosensor development is the conjugation of a fluorescent group to a solute-binding protein which, upon ligand binding, alters its fluorescent output (Dwyer and Hellinga 2004; Edwards 2021). This approach exploits the ligand binding properties and dynamics of a starting protein scaffold to transmit a signal to a fluorescent reporter upon ligand recognition. Current approaches for biosensor development are largely focused on trial-and-error based approaches, where a large number of candidate biosensors with different combinations of fluorophores and conjugation sites are tested *in vitro* to produce a functional sensor (e.g. this concept is well illustrated in (Brune et al. 1994; De Lorimier et al. 2002). These approaches are laborious. To overcome this limitation we have previously

developed the Computational Identification of Non-disruptive Conjugation sites (CINC) pipeline for rational biosensor design *in silico* (Smith et al. 2022a). CINC utilizes molecular dynamics simulations in combination with an in-house developed scoring algorithm (F_{Score}, and later $F_{Score2.0}$) to select fluorophore conjugation sites in a protein scaffold (Smith et al. 2022b).

We have previously used CINC to develop carbohydrate-detecting biosensors to aid in the characterization of Carbohydrate Active Enzymes (CAzymes), which are a series of functionally diverse enzymes that modify linkages or decorations on carbohydrates (Lombard et al. 2013). CAzymes have a variety of bioindustrial applications (Polaina and MacCabe 2007), however, industrial implementation of CAzymes require detailed characterization of enzymatic features (e.g. substrate preference, product(s) produced, reaction rates, etc.). To this end, biosensor-based approaches are advantageous over other commonly used methods for carbohydrate detection in that product detection is specific, can be achieved in real time, and is amenable to high-throughput assays (e.g. screening in 384 well plate-based fluorescence assays). However, biosensor-based approaches require the upfront development of these biomolecular tools for detection of the desired target molecules which can be cost-prohibitive. CINC streamlines the biosensor design process and has been used to construct biosensors for the sensitive detection of MOS (Smith et al. 2022a) and oligogalacturonides (Smith et al. 2022b). However, with our prior work the binding properties of the biosensor as well as the mechanism of biosensor response to

ligand were limited to that of the initial protein scaffold. CINC is well-positioned to maintain the ligand binding properties of the starting protein scaffold; however, a logical next step in its progression is to expand its scope to generate novel biosensors with user-defined properties (e.g. altered binding affinity, or altered readout mechanism). Our previously developed MOS-detecting biosensor displayed rapid (association rate constant (k_{on}) = 20 ± 2 $\mu M^{-1}s^{-1}$) and sensitive (dissociation constant (K_D) = 190 ± 50 nM) binding of its target ligand, which is beneficial in the study of reactions that produce MOS rapidly. However, these properties also necessitate reaction conditions where the sensor is present in excess over MOS - as essentially all MOS binds the sensor under most reaction conditions. As well, tight ligand binding by a biosensor can render these tools to be sensitive to contaminations present in solutions (e.g. MOS present in starch solutions). To this end, we sought to develop a biosensor library based on our MOS-detecting biosensor with an increased dynamic range of ligand detection to enable applications where sub-stoichiometric biosensor usage relative to target carbohydrate is required. As well, we demonstrate an additional application for CINC, in that it can be used to describe underlying mechanisms for biosensor response – enabling forward engineering of biosensors with well-defined responses to an input signal.

4.4 Materials and Methods:

4.4.1 Molecular dynamics simulations and calculation of $F_{Score2.0}$ values. Structural models of *apo* (PDB 2XD2) and MOS-bound MalX (PDB 2XD3) were

modified *in silico* to create variants hypothesized to impact ligand binding. Wild-type MalX, MalX D102A, MalX N195A, MalX K216A, MalX W273A, and MalX Q372A were examined using molecular dynamics-based workflows previously established by our group (De Laurentiis et al. 2016; Girodat et al. 2019; Mercier et al. 2015; Rosler et al. 2015). Briefly, each model was solvated using a TIP3P water box that extended 20 Å from any point on the protein and the system neutralized with Na^+ ions. Potential energy of the solvent, and then the entire system was minimized for 10 000 steps each using AMBER FF99SB and GLYCAM forcefields (Case 2016; Kirschner et al. 2008; Singh et al. 2016). The system was then heated for 10 000 steps to 300 K, and molecular dynamics simulations performed using the AMBER suite for 1 μs (3 replicates for each protein for both *apo* and MOS-bound states). Trajectories were combined using cpptraj (Roe and Cheatham III 2013), and $F_{Score2.0}$ values comparing *apo* to MOS-bound states were determined as previously described (Smith et al. 2022b). Briefly, $F_{Score2.0}$ values describe changes in backbone dihedral angle populations of each individual amino acid position in the *apo* versus MOS-bound molecular dynamics simulations. In each data set, $F_{Score2.0}$ values can range from 0 – 1, where a value of 1 would indicate that there is no overlap between dihedral angle populations in the *apo* versus ligand-bound state, and 0 indicates complete overlap. For a more in-depth discussion of the $F_{Score2.0}$ calculation, please refer to our prior study (Smith et al. 2022b).

4.4.2 *In vitro* production of MOS-detecting biosensors. MalX variants were produced without the signal secretion peptide and lipoprotein attachment motif found in wild-type *Streptococcus pneumoniae malX* as previously described (Abbott et al. 2010). Each *malX* variant incorporated our previously published A134C substitution for conjugation of a thiol reactive probe (Smith et al. 2022a), as well as substitutions proposed to impact ligand binding affinity (MalX A134C, MalX D102A/A134C, MalX N195A/A134C, MalX K216A/A134C, MalX W273A/A134C, and MalX Q372A/A134C). The previously listed *malX* variants were synthesized and subcloned into the *NheI* and *XhoI* sites of pET28a (meaning each construct contained an N-terminal poly-Histidine tag, constructs produced by BioBasic Canada Inc.). MalX variants were overexpressed, and purified using Ni^{2+}-affinity chromatography as previously described (Smith et al. 2022a) and purification yields were typically 100 mg of protein per liter of culture. MalX variants were then fluorescently labelled with the fluorophore 7-diethylamino-3-[*N*-(2-maleimidoethyl)carbamoyl]coumarin (MDCC) as previously described (Smith et al. 2022a; Smith et al. 2017). Labeling efficiencies were typically 80% to >95%, and protein recovery from the labeling procedure was typically ~50%. Concentration of each labeled MalX variant was determined as described below (Eq. 4.1, 4.2) with the following parameters: 0.164 is a correction factor which accounts for MDCC absorption at 280 nm (Brune et al. 1994; Smith et al. 2017), $\varepsilon_{280,\ MalX}$ = 61 310 M^{-1} cm^{-1} (except for MalX W273A/A134C, which had an extinction coefficient of 55 810 M^{-1} cm^{-1}). L is instrument path length in cm, and $\varepsilon_{430,\ MDCC}$ = 46 800 $M^{-1}cm^{-1}$ (Brune

et al. 1994; Smith et al. 2017). Labeling efficiency for all protein-fluorophore conjugates were calculated using Equation 4.3.

$$[MalX\ variant] = \frac{A_{280} - (A_{430} \times 0.164)}{\varepsilon_{280,MalX} \times L} \quad \text{(Eq. 4.1)}$$

$$[MDCC] = \frac{A_{430}}{\varepsilon_{430,MDCC} \times L} \quad \text{(Eq. 4.2)}$$

$$Labeling\ efficiency = \frac{[Fluorophore]}{[MalX\ variant]} \times 100\% \quad \text{(Eq. 4.3)}$$

4.4.3 Equilibrium fluorescence measurements. A Quanta Master 60 (Photon Technology International) fluorescence spectrophotometer was used to determine the dissociation constant of MalX A134C/W273A-MDCC at 20°C (1 nm step size, 1 s integration, excitation slit widths: 4 nm, emission slit widths were 8 nm, excitation wavelength was 420 nm, and emission wavelength was 440-480 nm). To obtain K_D for MalX A134C/W273A-MDCC, a titration of maltotriose (M3, purchased from Megazyme, P.N. O-MAL3) was performed and a hyperbolic function fit to the data using GraphPad Prism v. 9.0 (Eq. 4.4). Y was the absolute fluorescence change at each [M3], B_{max} was maximum binding in the same units as Y, and K_D was the equilibrium binding constant.

$$Y = \frac{B_{max} \times [M3]}{K_D + [M3]} \quad \text{(Eq. 4.4)}$$

4.4.4 Rapid-kinetics measurements. A KinTek SF-2004 stopped-flow apparatus (KinTek Corp.) was used for all rapid kinetics measurements (20°C, excitation wavelength 420 nm, emission measured after passing through 450 nm long-pass filter). Fluorescence time-courses were fit with a one-exponential function (Eq. 4.5), where *F* represented fluorescence observed at time *t*, F_∞ represented final fluorescence, A was amplitude of signal change, and k_{app} was the apparent rate (TableCurve, Systat Software). To obtain K_D for MalX variants listed in Table 4.1 (except MalX A134C/W273A-MDCC), a titration of M3 was performed and a hyperbolic function fit to amplitude changes for each variant (Eq. 4.4). Values for k_{on} were determined from the slope of a linear function fit to a plot of k_{app} vs. [M3] as previously described (Smith et al. 2022a).

$$F = F_\infty + A_1 \times \exp(-k_{app}t) \quad \text{(Eq. 4.5)}$$

4.5 Results:

4.5.1 *In silico* analysis and re-engineering of MalX A134C-MDCC. To generate a library of MOS-detecting biosensors with different MOS affinities, we used MalX A134C-MDCC as a template. MalX binds MOS through a network of hydrogen-bonding and stacking interactions at an interface between its N and C terminal domains (Abbott et al. 2010). In order to lower the inherent high affinity of MalX for MOS, alanine substitutions were introduced to either disrupt hydrogen bonding interactions with ligand, disrupt stacking interactions with ligand, or hinder the hinge-bend motion stabilized by ligand binding (Fig. 4.1A). For example, the

D102A variant of MalX was constructed to disrupt hydrogen-bonding interactions between D102 and the O2/O3 positions of the third glucose residue in the bound MOS. Whereas an N195A substitution was introduced to disrupt the hydrogen bonding network involving N195 and several oxygens in the MOS structure (Fig. 4.1A). The W273A variant was designed to disrupt a stacking interaction between W273 and the first glucose ring in a MOS at MalX Subsite 1 (Abbott et al. 2010), and the K216A/ Q372A variants were hypothesized to destabilize the “closed” state of MalX (Fig. 4.1A).

With the aforementioned amino acid substitutions of our MOS-detecting biosensor in mind, we hypothesized that some amino acid substitutions may impact the dynamics of the reporter position in our previously developed MalX A134C-MDCC biosensor. Dynamics at the reporter position are important to examine in this context as understanding molecular determinants behind biosensor signal transmission from the ligand binding site to the reporter position is critical in forward engineering of these biomolecular tools. To this end, we utilized CINC to assess the dynamics properties of the reporter position (A134) of all proposed variants using molecular dynamics simulations and an in-house developed algorithm ($F_{Score2.0}$, (Smith et al. 2022b)). $F_{Score2.0}$ is a fraction describing differences in backbone dihedral angle populations in the *apo* vs. ligand bound state of a protein scaffold. $F_{Score2.0}$ values can range from 0 – 1, where 0 indicates complete overlap of Ramachandran Plots when comparing the *apo* versus ligand-bound states, and a value of 1 indicates no overlap. In the case of each MalX

variant, mean values for $F_{Score2.0}$ at Position A134 ranged from 0.32 to 0.6 (Fig. 4.1B). Interestingly, all MalX variants except MalX D102A displayed at least one "high scoring" replicate where $F_{Score2.0}$ was > 0.5. In the case of D102A, all molecular dynamics simulation replicates clustered near $F_{Score2.0}$ values of ~0.3. As well, several MalX variants displayed altered modes by which their high $F_{Score2.0}$ values were achieved (Figs. A3.1 – A3.6). For example, with MalX wt the highest $F_{Score2.0}$ value arose from the appearance of a unique dihedral conformation at Position 134 in the ligand-bound state (Fig. A3.1). Whereas in the case of MalX N195A, its highest $F_{Score2.0}$ value arose from the removal of a unique *apo*-state dihedral conformation at Position 134 upon ligand binding (Fig. A3.3). Following this *in silico* screening step, all proposed variants of the MalX A134C-MDCC biosensor were biochemically characterized to determine if they indeed were able to detect MOS with an altered affinity relative to the starting MalX A134C-MDCC biosensor.

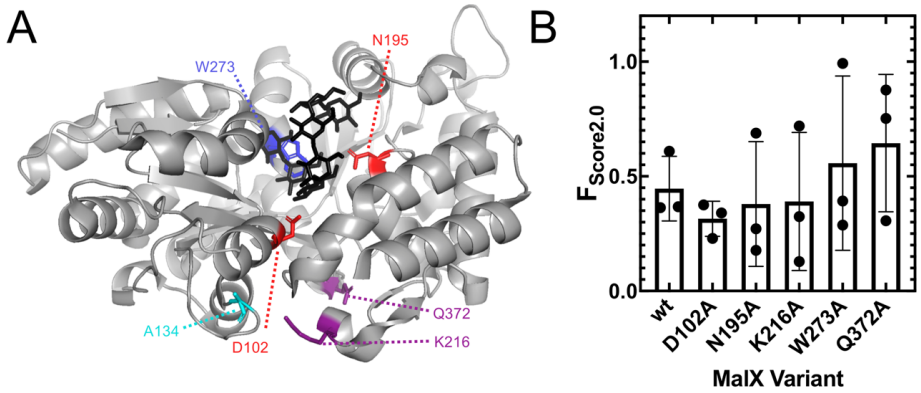


Fig. 4.1. Proposed substitutions of MalX to alter MOS-binding affinity and their impact on dynamics of the A134 reporter position. (A) The MalX protein scaffold is shown as a grey cartoon, with the bound MOS shown in stick representation (black). Candidate amino acid positions for substitution to alanine are shown in blue (disrupt ligand stacking), red (disrupt ligand hydrogen bonding), and purple (hypothesized to disrupt "closed" state of MalX). The reporter position, A134, is shown in cyan. (B) $F_{Score2.0}$ values for wild-type MalX, and each candidate MalX variant examined using CINC. $F_{Score2.0}$ values reflect changes in backbone dihedral angle dynamics in the *apo* vs. MOS-bound states of MalX. Each bar represents mean $F_{Score2.0}$ values (n = 3, ± s.d.), and results from individual molecular dynamics simulations are shown as black dots.

4.5.2 Biochemical characterization of MalX-based biosensors. MalX variants were fluorescently labelled with MDCC, a diethylaminocoumarin previously used in the construction of many solute-binding protein-based biosensors including our previous MalX-based biosensor (Smith et al. 2022a; Smith et al. 2017; Smith et al. 2022b). MalX-based biosensors with their respective amino acid substitution were examined for their ability to detect M3 using both equilibrium fluorescence spectroscopy and the stopped-flow method. Of the designed biosensor variants, 4 out of 5 were able to detect ligand over the MOS concentration range tested (Table 4.1, Fig. A3.7). K_D values for ligand binding by our MOS-detecting biosensor library spanned over 5 orders of magnitude, and in general were accompanied by

decreases in k_{on}. Included in this assessment were biosensors carrying the substitutions K216A and Q372A, which displayed a decrease in binding affinity of approximately one order of magnitude compared to our original MalX-based biosensor. The affinities for our engineered biosensor library for M3 range from hundreds of nanomolar to the millimolar range, thereby realizing our goal of developing an engineered toolset for detection of MOS over a broad concentration range.

Table 4.1. Computational characterization and ligand binding properties of MalX-based biosensor library. Three categories of MalX variants were constructed in order to generate a broad range of ligand binding affinities. $F_{Score2.0}$ values examining the impact of each substitution on the reporter *in silico*, as well as wet-lab kinetic data for M3 binding by MDCC labelled conjugates is shown.

MalX Variant	Category	$F_{Score2.0}$[1] (mean, range)	K_D (μM)	k_{on} ($\mu M^{-1}s^{-1}$)
A134C	N/A	0.4 (0.36 – 0.61)	0.190 ± 0.050[2]	20 ± 2[2]
D102A/A134C	H-bonding	0.3 (0.23 – 0.38)	N.C.	N.C.
A134C/N195A	H-bonding	0.4 (0.18 – 0.69)	230 ± 90	0.113 ± 0.008
A134C/K216A	Inter-domain interface	0.4 (0.13 – 0.72)	3.2 ± 0.8	9 ± 1
A134C/W273A	Ligand Stacking	0.6 (0.39 – 0.99)	5900 ± 200	N.D.
A134C/Q372A	Inter-domain interface	0.6 (0.31 – 0.88)	2.7 ± 0.3	13 ± 1

[1] $F_{Score2.0}$ values were measured in molecular dynamics simulations without A134C substitution as listed in Fig. 4.1.
[2] (Smith et al. 2022a)
N.C.: No fluorescent signal change detected upon ligand addition – variant cannot detect M3 over the concentration range tested.
N.D.: Not determined

4.6 Discussion:

Of the MalX A134C-MDCC biosensor variants, only the D102A substitution completely abolished the biosensor's ability to respond to M3. There are two main explanations for this. First, it is possible that the D102A substitution may have disrupted the binding site of MalX to such an extent that ligand binding is no longer

detectable. Second, MalX D102A/A134C-MDCC is still capable of binding MOS, but the binding event no longer triggers a MOS-dependent signal change of the reporter system. The different behaviour of the D102A variant is consistent with the observation that the results from the *in silico* analysis of MalX D102A are unique in that none of the replicates displayed an $F_{Score2.0}$ value above 0.5 – whereas all other variants had at least one high-scoring replicate (Fig. 4.1B). Upon further investigation of the Ramachandran plots obtained via molecular dynamics simulations for each MalX variant, the importance of these "high-scoring" populations becomes even more apparent (Figs. A3.1-3.6). Specifically, all MalX variants except D102A contained sub-populations in the reported Ramachandran plots which were unique to either the *apo* or MOS-bound states (i.e. an alternative conformation in either the *apo* or MOS-bound state, Figs A3.1-3.6). It is therefore likely that the observed replicates with high $F_{Score2.0}$ values represent a conformation that, when conjugated to a fluorescent reporter, are responsible for the observed fluorescence changes within the biosensor ensemble upon ligand binding (i.e. an "active" subset of the pool of molecules). In the case of D102A, this variant does not access a high-scoring conformation (Fig. A3.2) which explains why the biosensor no longer responds to M3. In our prior study, several other candidate biosensors which responded to their target ligands also had at least one "higher-scoring" replicate when using the $F_{Score2.0}$ criteria (Smith et al. 2022b). Therefore, taken together with this work we recommend that future users of CINC and the $F_{Score2.0}$ algorithm prioritize selection of reporter positions that have at least

one high scoring replicate in single molecule molecular dynamics simulations rather than focusing on averaged mean $F_{Score2.0}$ values.

With respect to the observed Ramachandran plots for the MalX variants utilized in this study, we further propose that the mode by which biosensing is achieved was altered for a number of our MalX variants (Fig 4.2). Specifically, in the case of wt MalX upon ligand binding a population of dihedral angles is present with a unique conformation not observed in the *apo* state (Fig. 4.2B, A3.1). We therefore speculate that the mechanism of readout for MalX A134C-MDCC is due to the addition of a sub-population of molecules in the ligand-bound state that have an altered fluorescence state compared to the other molecules in the ensemble (i.e. readout based on a "active" biosensing state in the ligand-bound state). Similarly, our MalX variants which utilized amino acid substitutions distal from the ligand binding site of MalX also displayed an "addition of a ligand-bound state" mode (Fig. A3.4, A3.6). However, MalX N195A and MalX W273A both exhibit an alternative mode of biosensing where their high $F_{Score2.0}$ values were due to the removal of a unique population from their *apo* state upon ligand binding (Fig. 4.2C, A3.3, A3.5). These observations further detail the power of the CINC algorithm in that it details how point mutations within the ligand binding site (i.e. N195A, W273A) can impact the dynamics of the distal reporter position and alter the mode of output generation. Whereas point mutations distal from the ligand binding site (i.e. K216A, Q372A) maintained the original "addition of population in ligand-bound state" mode of biosensing observed with our original MalX A134C-MDCC

biosensor. Taken together, it may therefore be a promising strategy to screen amino acid substitutions in CINC as a means to forward-engineer biosensing devices where “active” sub-populations are enriched or where specific response modes are desired.

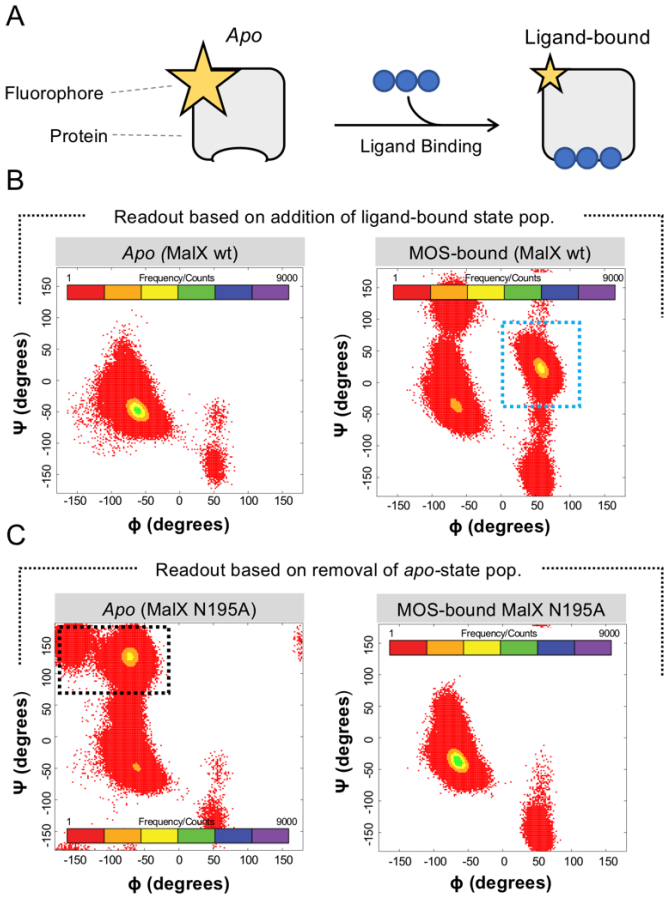


Fig. 4.2. Modes of readout for ensemble biosensor fluorescence changes. (A) A general schematic describing ensemble fluorescence in a biosensor upon ligand binding. However, in the case of our MalX-based biosensors fluorescence change of the ensemble is due to differing underlying sensing mechanisms. (B) In MalX wt, upon ligand binding the addition of a unique dihedral state at Position 134 which is not observed in the *apo* simulations (cyan hashed box) is responsible for the resulting fluorescence change of the ensemble. These sensors are therefore described as their readout being based on the addition of a population in the ligand-bound state. (C) Alternatively, in *apo* MalX N195A the molecule adopts unique dihedral angles (black hashed box), and ligand binding removes this dihedral conformation from the observed equilibrium. Therefore, this biosensor would be described by an alternative sensing mechanism whereby readout is based on removal of a population present in the *apo* state.

With the results reported in this study our MOS-detecting biosensor library now encompasses tools with ligand binding affinities spanning over five orders of magnitude which will be useful where a broad dynamic range of MOS detection is required. This improvement is similar to previous work that increased the dynamic range of a phosphate-detecting biosensor, thereby improving applicability of biosensor-based phosphate detection assays to additional systems (Solscheid et al. 2015). Moving forward, these tools in combination with other CINC-designed biosensors (which utilize variety of fluorophores with alternative spectroscopic properties (Smith et al. 2022a)), will enable multiplexing of CINC-designed biosensors further streamlining CAZyme characterization workflows.

4.7 Declaration of Competing Interest:

DDS, DWA, and HJW are Joint Inventors on International PCT Patent Application No. CA2021/05/01/10 (Filed January 30, 2021).

4.8 Acknowledgements:

This research was enabled by support provided by the Digital Research Alliance of Canada (www.alliancecan.ca). We thank Dr. Dylan Girodat for feedback on molecular dynamics data, and Landon McCabe for his preliminary work on this project during his Independent Study coursework.

4.9 Chapter 4 Supplementary Information:

Supplementary Information for Chapter 4 can be found in Appendix 3

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

The research presented in this thesis further demonstrates ways how protein molecular dynamics can shed light on our understanding of biological systems and how these systems can be engineered for useful purposes. One of my main research questions in the outset of this project was: can protein-based biomolecular tools be designed using our knowledge of their underlying dynamic features? It is clear from the work presented in Chapters 2-4, that protein-based biosensor design can be reliably achieved using the CINC pipeline (Smith et al. 2022a; Smith et al. 2022b). For clarity, I have summarized key considerations related to the use of molecular dynamics simulations for the design of protein-fluorophore conjugates in point form below:

In selecting candidate labeling positions:

- Examining changes in dihedral angles in *apo* vs. ligand bound state was the most reliable means tested for selecting a fluorophore conjugation site (Chapter 2)
- Complete (or near complete) shifts in dihedral angle populations are preferred to partial shifts (Chapter 3)
- $F_{Score2.0}$ currently does not take into consideration the magnitude of dihedral angle changes upon ligand binding. It just reports the extent of population overlap. One should additionally consider the magnitude of dihedral angle changes as minor alterations in backbone conformation may not be

sufficient to induce a reorientation of the conjugated fluorophore into an altered environment upon ligand binding (Chapter 3)

- In the case of selecting a fluorophore conjugation site based on partial shifts in dihedral angle populations, one should give preference to high-scoring single replicates over mean $F_{Score2.0}$ values (i.e. a high-scoring conformation in an ensemble, Chapter 4).

Moving forward, CINC could be improved by amending the $F_{Score2.0}$ algorithm to take into consideration the third and fourth bullet points above as manual intervention is currently required to examine these features. As well, it is possible to apply CINC beyond SBPs. For example, CINC has been used to construct a novel biosensor using the translational GTPase EF-Tu as a protein scaffold and resulted in a biosensor that alters its fluorescent output in response to GDP/GTP levels (Unpublished data from Luc Roberts; data was however included in International PCT Patent Application No. CA2021/050110). EF-Tu has a drastically different tertiary structure than the two-domain solute-binding proteins reported in this thesis, demonstrating that CINC is flexible with respect to a variety of protein scaffolds that can be used as input in the biosensor design pipeline. As a future direction, one could increase the power of the CINC *in silico* screening step via the inclusion of fluorophores in simulations to provide increased resolution with respect to the environment the fluorophore actually is exposed to. However, the inclusion of fluorophores in simulations would increase the computational demands of the pipeline exponentially dependent on the number of conjugation

sites/fluorophores to be examined. Still, as computational power increases the inclusion of fluorophores in simulations for biosensor design will become more feasible and aid in understanding underlying biosensing mechanisms.

In general, protein molecular dynamics simulations are an emerging field critical to bridge the gap between structure and function. As an analogy, for decades researchers have manually solved protein structures using x-ray crystallography and related methods. These research projects to date have resulted in in approximately 190,000 deposited structures into Protein Data Bank (PDB). These structural research projects were laborious and have since enabled Artificial Intelligence protein structural prediction algorithms such as AlphaFold (Jumper et al. 2021). In 2021, an open-source version of AlphaFold was released, and to date over 200 million structures have been deposited in AlphaFoldDB. Of course, there are caveats when comparing structures in AlphaFoldDB versus structures in PDB, but the impact of these computational tools on structural biology is undeniable. Beyond these tools for the prediction of static structures, it is clear that the next level of biological complexity is mass adaptation of computational tools to describe the dynamic properties of biomolecules. Moving forward, one could envision the impact of coupling next-generation protein folding algorithms with our CINC pipeline to generate candidate biosensors in high-throughput. Such an endeavour is currently impractical given the computational infrastructure that would be required to use CINC in high-throughput. However, the research presented in this thesis shows that the CINC framework is feasible for

computational biosensor design and utilizing a similar workflow for high-throughput biosensor design is possible. This would, for example, revolutionize the biosensor industry and enable detection of a broad range of biomolecules. Coupling these devices with portable detection tools (e.g. a handheld spectrophotometer that attaches to a mobile phone), one could envision on-site/ point-of-care detection of industrial or physiologically-relevant biomolecules. Alternatively, quantification of analytes in a complex mixture using a microfluidic device with a series of biosensors in succession could be developed to further apply portable detection methods. Last, the general use of the molecular dynamics-based tools outlined in this thesis have utility in understanding underlying properties of biomolecules, which has further implications in health and disease. In summary, this is a rapidly evolving field with many exciting long-term innovation opportunities which can build on the work presented in this thesis.

REFERENCES:

Abbott, D.W., Boraston, A.B., 2007a. J. Biol. Chem. 282(48), 35328-35336.

Abbott, D.W., Boraston, A.B., 2007b. J. Mol. Biol. 369(3), 759-770.

Abbott, D.W., Boraston, A.B., 2007c. J. Mol. Biol. 368(5), 1215-1222.

Abbott, D.W., Higgins, M.A., Hyrnuik, S., Pluvinage, B., Lammerts van Bueren, A., Boraston, A.B., 2010. Mol. Microbiol. 77(1), 183-199.

Armstrong, N., Gouaux, E., 2000. Neuron 28(1), 165-181.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. Nucleic Acids Res. 44(W1), W344-W350.

Badar, M.S., Shamsi, S., Ahmed, J., Alam, M., 2022. Molecular dynamics simulations: concept, methods, and applications. Transdisciplinarity, pp. 131-151. Springer.

Baird, G.S., Zacharias, D.A., Tsien, R.Y., 1999. Proc. Natl. Acad. Sci. U.S.A. 96(20), 11241-11246.

Belousov, V.V., Fradkov, A.F., Lukyanov, K.A., Staroverov, D.B., Shakhbazov, K.S., Terskikh, A.V., Lukyanov, S., 2006. Nat. Methods 3(4), 281-286.

Berg, J., Hung, Y.P., Yellen, G., 2009. Nat. Methods 6(2), 161-166.

Berger, E.A., 1973. Proc. Natl. Acad. Sci. U.S.A. 70(5), 1514-1518.

Berger, E.A., Heppel, L.A., 1974. J. Biol. Chem. 249(24), 7747-7755.

Berntsson, R.P.-A., Smits, S.H., Schmitt, L., Slotboom, D.-J., Poolman, B., 2010. FEBS Lett. 584(12), 2606-2617.

Bianchi-Smiraglia, A., Rana, M.S., Foley, C.E., Paul, L.M., Lipchick, B.C., Moparthy, S., Moparthy, K., Fink, E.E., Bagati, A., Hurley, E., 2017. Nat. Methods 14(10), 1003.

Brune, M., Hunter, J.L., Corrie, J.E., Webb, M.R., 1994. Biochem. 33(27), 8262-8271.

Caffall, K.H., Mohnen, D., 2009. Carbohydr. Res. 344(14), 1879-1900.

Case, D.A., Betz, R.M. Cerutti, D.S., Cheatham III, T.E., Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T.S., LeGrand, S., Li, P., Lin, C., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, K.M., Kollman, P.A. , 2016. AMBER 2016. University of California, San Francisco., https://ambermd.org/doc12/Amber16.pdf.

Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., Ben-Tal, N., 2013. Isr. J. Chem. 53(3-4), 199-206.

Chen, T.-W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., Looger, L.L., Svoboda, K., Kim, D.S., 2013. Nature 499(7458), 295.

Ciftci, D., Huysmans, G.H., Wang, X., He, C., Terry, D., Zhou, Z., Fitzgerald, G., Blanchard, S.C., Boudker, O., 2020. Sci. Adv. 6(22), eaaz1949.

De Laurentiis, E.I., Mercier, E., Wieden, H.-J., 2016. J. Biol. Chem. 291(44), 23136-23148.

De Lorimier, R.M., Smith, J.J., Dwyer, M.A., Looger, L.L., Sali, K.M., Paavola, C.D., Rizk, S.S., Sadigov, S., Conrad, D.W., Loew, L., 2002. Protein Sci. 11(11), 2655-2675.

Dwyer, M.A., Hellinga, H.W., 2004. Curr. Opin. Struct. Biol. 14(4), 495-504.

Edwards, K.A., 2021. Talanta Open, 100038.

Endoh, T., Tan, J.-H., Chen, S.-B., Sugimoto, N., 2022. Anal. Chem.

Engkvist, O., Norrby, P.-O., Selmi, N., Lam, Y.-h., Peng, Z., Sherer, E.C., Amberg, W., Erhard, T., Smyth, L.A., 2018. Drug Discov. Today 23(6), 1203-1218.

Felder, C.B., Graul, R.C., Lee, A.Y., Merkle, H.-P., Sadee, W., 1999. AAPS Pharmsci. 1(2), 7-26.

Fonin, A., Povarova, O., Staiano, M., D'Auria, S., Turoverov, K., Kuznetsova, I., 2014. Opt. Mater. 36(10), 1676-1679.

Friedman, A.M., Fischmann, T.O., Steitz, T.A., 1995. Science 268(5218), 1721-1727.

Fu, Y., Finney, N.S., 2018. RSC Adv. 8(51), 29051-29061.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.e., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. The proteomics protocols handbook, 571-607.

Gilardi, G., Zhou, L.Q., Hibbert, L., Cass, A.E., 1994. Anal. Chem. 66(21), 3840-3847.

Girodat, D., Mercier, E., Gzyl, K.E., Wieden, H.-J., 2019. J. Am. Chem. Soc. 141(26), 10236-10246.

Gonin, S., Arnoux, P., Pierru, B., Lavergne, J., Alonso, B., Sabaty, M., Pignol, D., 2007. BMC Struct. Biol. 7(1), 11.

Guo, Z., Johnston, W.A., Stein, V., Kalimuthu, P., Perez-Alcala, S., Bernhardt, P.V., Alexandrov, K., 2015. ChemComm 52(3), 485-488.

Guo, Z., Murphy, L., Stein, V., Johnston, W.A., Alcala-Perez, S., Alexandrov, K., 2016. J. Am. Chem. Soc. 138(32), 10108-10111.

Hanes, J.W., Chatterjee, D., Soriano, E.V., Ealick, S.E., Begley, T.P., 2011. ChemComm 47(8), 2273-2275.

Harms, H., Wells, M.C., van der Meer, J.R., 2006. Appl. Microbiol. Biotechnol. 70(3), 273-280.

Harrison, M., Dunlop, M., 2012. Front. Microbiol. 3, 360.

Harvey, D.J., 2011. J. Chromatogr. B 879(17), 1196-1225.

Heller, A., Feldman, B., 2008. Chem. Rev. 108(7), 2482-2505.

Höfig, H., Otten, J., Steffen, V., Pohl, M., Boersma, A.J., Fitter, J.r., 2018. ACS Sens. 3(8), 1462-1470.

Hollingsworth, S.A., Dror, R.O., 2018. Neuron 99(6), 1129-1143.

Hsieh, H.V., Sherman, D.B., Andaluz, S.A., Amiss, T.J., Pitner, J.B., 2012. J. Diabetes Sci. Technol. 6(6), 1286-1295.

Hughes, M.D., 2009. J. Diabetes Sci. Technol. 3(5), 1219-1223.

Humphrey, W., Dalke, A., Schulten, K., 1996. J. Mol. Graphics 14(1), 33-38.

Ispas, C.R., Crivat, G., Andreescu, S., 2012. Anal. Lett. 45(2-3), 168-186.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., 2021. Nature 596(7873), 583-589.

Juzgado, A., Soldà, A., Ostric, A., Criado, A., Valenti, G., Rapino, S., Conti, G., Fracasso, G., Paolucci, F., Prato, M., 2017. J. Mater. Chem. B 5(32), 6681-6687.

Kirschner, K.N., Yongye, A.B., Tschampel, S.M., González-Outeiriño, J., Daniels, C.R., Foley, B.L., Woods, R.J., 2008. J. Comput. Chem. 29(4), 622-655.

Ko, W., Kim, S., Lee, H.S., 2017. Org. Biomol. Chem. 15(41), 8761-8769.

Ko, W., Kim, S., Lee, S., Jo, K., Lee, H.S., 2016. RSC Adv. 6(82), 78661-78668.

Koshland Jr, D., 1953. Biological reviews 28(4), 416-436.

Kostyuk, A.I., Demidovich, A.D., Kotova, D.A., Belousov, V.V., Bilan, D.S., 2019. Int. J. Mol. Sci. 20(17), 4200.

Kothe, U., Rodnina, M.V., 2006. Biochem. 45(42), 12767-12774.

Kuhlman, B., Bradley, P., 2019. Nat. Rev. Mol. Cell Biol. 20(11), 681-697.

Kunzelmann, S., Webb, M.R., 2009. J. Biol. Chem. 284(48), 33130-33138.

Kunzelmann, S., Webb, M.R., 2010. ACS Chem. Biol. 5(4), 415-425.

Law, J.W.-F., Ab Mutalib, N.-S., Chan, K.-G., Lee, L.-H., 2015. Front. Microbiol. 5, 770.

Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., Lu, P., 1996. Science 271(5253), 1247-1254.

Liao, C.-H., Revear, L., Hotchkiss, A., Savary, B., 1999. Can. J. Microbiol. 45(5), 396-403.

Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P.M., Henrissat, B., 2010. Biochem. J. 432(3), 437-444.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2013. Nucleic Acids Res. 42(D1), D490-D495.

Mao, B., Pear, M., McCammon, J., Quiocho, F., 1982. J. Biol. Chem. 257(3), 1131-1133.

Markets, and Markets 2019. Biosensors Market by Type (Sensor patch and embedded device), Product (Wearable and nonwearable), Technology (Electrochemical and optical), Application (POC, Home Diagnostics, Research Lab, Food & Beverages), and Geography - Global Forecast to 2024. pp. 1-90.

Marvin, J.S., Schreiter, E.R., Echevarría, I.M., Looger, L.L., 2011. Proteins: Struct. Funct. Genet. 79(11), 3025-3036.

Marvin, J.S., Shimoda, Y., Magloire, V., Leite, M., Kawashima, T., Jensen, T.P., Kolb, I., Knott, E.L., Novak, O., Podgorski, K., 2019. Nat. Methods 16(8), 763-770.

McLean, R., Hobbs, J.K., Suits, M.D., Tuomivaara, S.T., Jones, D.R., Boraston, A.B., Abbott, D.W., 2015. J. Biol. Chem. 290(35), 21231-21243.

Mehrotra, P., 2016. J. Oral Biol. Craniofac. Res. 6(2), 153-159.

Mercier, E., Girodat, D., Wieden, H.-J., 2015. Sci. Rep. 5, 7677.

Misono, K.S., 2002. Mol. Cell. Biochem. 230(1-2), 49-60.

Mulligan, C., Geertsma, E.R., Severi, E., Kelly, D.J., Poolman, B., Thomas, G.H., 2009. Proc. Natl. Acad. Sci. U.S.A. 106(6), 1778-1783.

Muretta, J.M., Rohde, J.A., Johnsrud, D.O., Cornea, S., Thomas, D.D., 2015. Proc. Natl. Acad. Sci. U.S.A. 112(46), 14272-14277.

Neiditch, M.B., Federle, M.J., Pompeani, A.J., Kelly, R.C., Swem, D.L., Jeffrey, P.D., Bassler, B.L., Hughson, F.M., 2006. Cell 126(6), 1095-1108.

Newman, R.H., Fosbrink, M.D., Zhang, J., 2011. Chem. Rev. 111(5), 3614-3666.

Okoh, M.P., Hunter, J.L., Corrie, J.E., Webb, M.R., 2006. Biochem. 45(49), 14764-14771.

Ouellet, J., 2016. Front. Chem. 4, 29.

Pardee, K., Green, Alexander A., Ferrante, T., Cameron, D.E., DaleyKeyser, A., Yin, P., Collins, James J., 2014. Cell 159(4), 940-954.

Pardee, K., Green, A.A., Takahashi, M.K., Braff, D., Lambert, G., Lee, J.W., Ferrante, T., Ma, D., Donghia, N., Fan, M., 2016. Cell 165(5), 1255-1266.

Peeling, R.W., Olliaro, P.L., Boeras, D.I., Fongwen, N., 2021. Lancet Infect. Dis. 21(9), e290-e295.

Polaina, J., MacCabe, A.P., 2007. Industrial enzymes. Springer

Posor, Y., Eichhorn-Gruenig, M., Puchkov, D., Schöneberg, J., Ullrich, A., Lampe, A., Müller, R., Zarbakhsh, S., Gulluni, F., Hirsch, E., 2013. Nature 499(7457), 233-237.

Posthuma-Trumpie, G.A., Korf, J., van Amerongen, A., 2009. Anal. Bioanal. Chem. 393(2), 569-582.

Quiocho, F.A., Ledvina, P.S., 1996. Mol. Microbiol. 20(1), 17-25.

Rai, V., Acharya, S., Dey, N., 2012. J. Biomater. Nanobiotechnol. 3(2A), 315.

Reinartz, I., Sarter, M., Otten, J., Höfig, H., Pohl, M., Schug, A., Stadler, A.M., Fitter, J., 2021. Sensors 21(12), 4144.

Rocchitta, G., Spanu, A., Babudieri, S., Latte, G., Madeddu, G., Galleri, G., Nuvoli, S., Bagella, P., Demartis, M.I., Fiore, V., 2016. Sensors 16(6), 780.

Roe, D.R., Cheatham III, T.E., 2013. J. Chem. Theory Comput. 9(7), 3084-3095.

Rosler, K.S., Mercier, E., Andrews, I.C., Wieden, H.-J., 2015. J. Biol. Chem. 290(30), 18650-18661.

Rurack, K., Resch-Genger, U., 2002. Chem. Soc. Rev. 31(2), 116-127.

Salins, L.L., Goldsmith, E.S., Ensor, M.C., Daunert, S., 2002. Anal. Bioanal. Chem. 372(1), 174-180.

Savelsbergh, A., Mohr, D., Kothe, U., Wintermeyer, W., Rodnina, M.V., 2005. EMBO J. 24(24), 4316-4323.

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. Nat. Methods 9(7), 671.

Schumacher, M.A., Choi, K.Y., Zalkin, H., Brennan, R.G., 1994. Science 266(5186), 763-770.

Shevchik, V.E., Condemine, G., Robert-Baudouy, J., Hugouvieux-Cotte-Pattat, N., 1999. J. Bacteriol. 181(13), 3912-3919.

Shin, H.J., 2011. Appl. Microbiol. Biotechnol. 89(4), 867-877.

Singh, A., Tessier, M.B., Pederson, K., Wang, X., Venot, A.P., Boons, G.-J., Prestegard, J.H., Woods, R.J., 2016. Can. J. Chem. 94(11), 927-935.

Singh, N., Li, W., 2019. Int. J. Mol. Sci. 20(15), 3774.

Smith, D.D., Girodat, D., Abbott, D.W., Wieden, H.-J., 2022a. Biosensors Bioelectron. 200, 113899.

Smith, D.D., Girodat, D., Wieden, H.-J., Selinger, L.B., 2017. Anal. Biochem. 537, 106-113.

Smith, D.D., King, J.P., Abbott, D.W., Wieden, H.-J., 2022b. Sensors 22(3), 948.

Solscheid, C., Kunzelmann, S., Davis, C.T., Hunter, J.L., Nofer, A., Webb, M.R., 2015. Biochem. 54(32), 5054-5062.

Srilatha, B., 2011. J. Nanomed. Nanotechnol. 2(7), 1000123.

Ström, A., Williams, M.A., 2004. Carbohydr. Res. 339(10), 1711-1716.

Tam, R., Saier, M., 1993. Microbiol. Rev. 57(2), 320-346.

Tamura, T., Hamachi, I., 2014. ACS Chem. Biol. 9(12), 2708-2717.

Tang, C., Schwieters, C.D., Clore, G.M., 2007. Nature 449(7165), 1078.

Tian, Y., Cuneo, M.J., Changela, A., Höcker, B., Beese, L.S., Hellinga, H.W., 2007. Protein Sci. 16(10), 2240-2250.

Toseland, C.P., 2013. J. Chem. Biol. 6(3), 85-95.

Vallée-Bélisle, A., Ricci, F., Plaxco, K.W., 2012. J. Am. Chem. Soc. 134(6), 2876-2879.

Verma, N., Bhardwaj, A., 2015. Appl. Biochem. Biotechnol. 175(6), 3093-3119.

Wang, H., Cheng, H., Wang, J., Xu, L., Chen, H., Pei, R., 2016. Talanta 154, 498-503.

Wang, J., 2008. Chem. Rev. 108(2), 814-825.

Webb, A.J., Homer, K.A., Hosie, A.H., 2008. J. Bacteriol. 190(1), 168-178.

Wolf, S., Mouille, G., Pelloux, J., 2009. Molecular plant 2(5), 851-860.

Ying, Z.-M., Xiao, H.-Y., Tang, H., Yu, R.-Q., Jiang, J.-H., 2018. ChemComm 54(64), 8877-8880.

Yu, Z., Cai, G., Liu, X., Tang, D., 2021. Anal. Chem. 93(5), 2916-2925.

Zeeman, S.C., Kossmann, J., Smith, A.M., 2010. Annu. Rev. Plant Biol. 61, 209-234.

Zeng, R., Wang, W., Chen, M., Wan, Q., Wang, C., Knopp, D., Tang, D., 2021. Nano Energy 82, 105711.

Zhang, K., Lv, S., Lin, Z., Li, M., Tang, D., 2018a. Biosensors Bioelectron. 101, 159-166.

Zhang, K., Lv, S., Lu, M., Tang, D., 2018b. Biosensors Bioelectron. 117, 590-596.

APPENDIX 1: SUPPLEMENTARY INFORMATION FOR CHAPTER 2

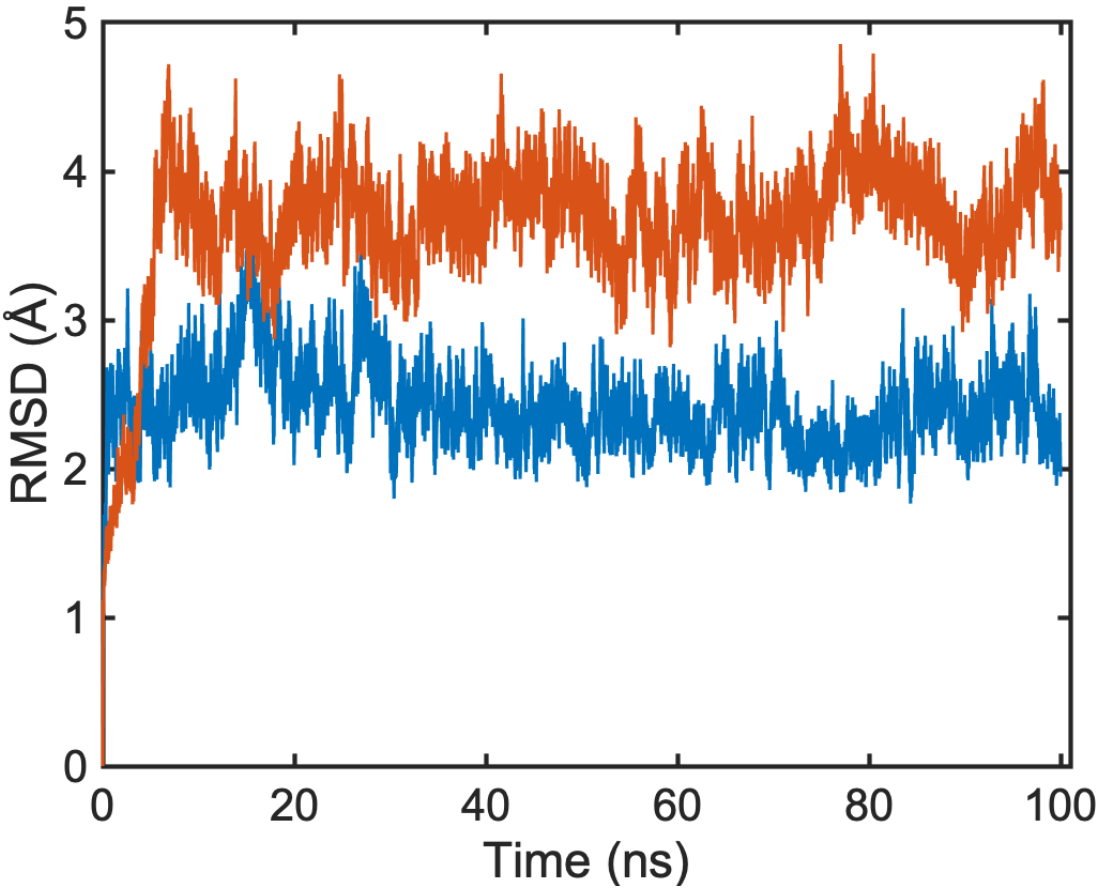


Fig. A1.1. RMSD plots of MalX apo (blue) and ligand-bound (red) for the duration of the molecular dynamics simulations. Both plots show an initial increase in RMSD during initial equilibration relative to starting structures (PDB 2XD2 and PDB 2XD3 respectively) and remain stable for the duration of simulations.

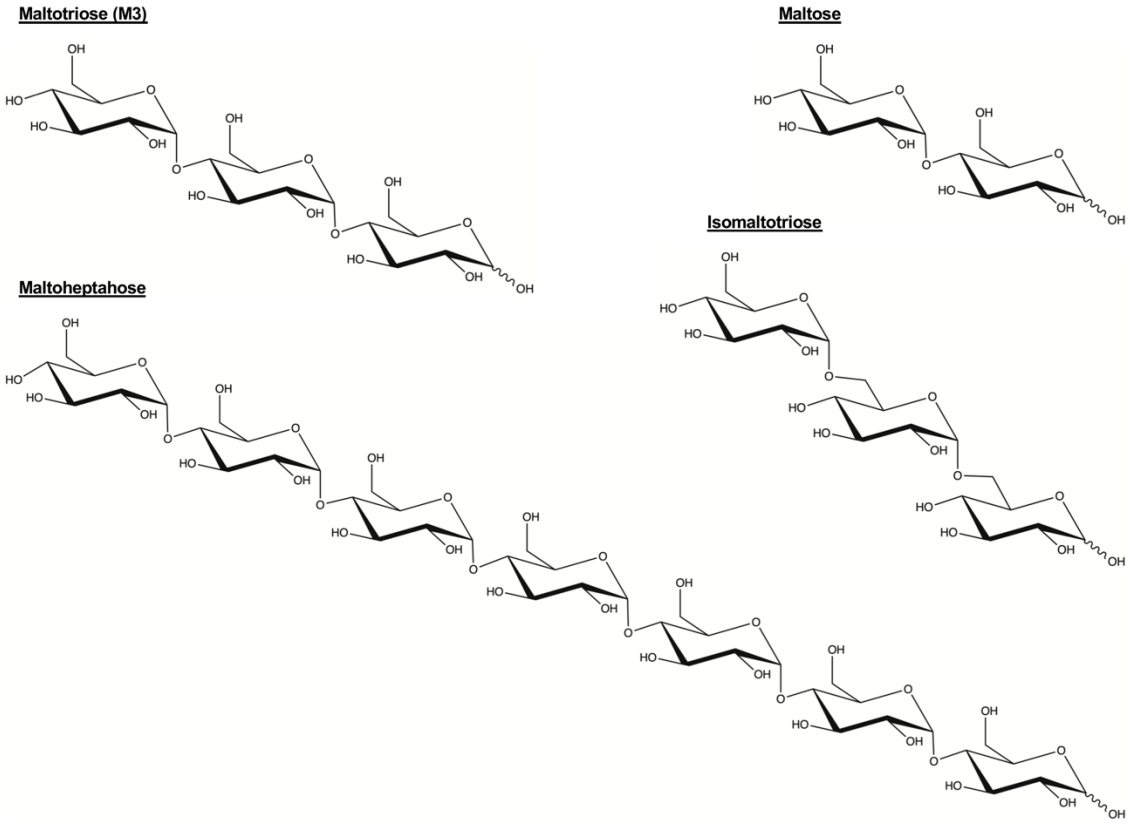


Fig. A1.2. Structural diagrams of carbohydrates used in this study (generated using ChemDraw Prime 20.0). All carbohydrates examined are compositionally similar (comprised only of glucose) and contain similar anomeric linkage configurations (all α linkages). Carbohydrates possess varying degrees of polymerization and distinct positional linkages (in the case of maltose, M3, and maltoheptaose: α-1,4-glycosidic bonds are present; whereas in isomaltotriose α-1,6-glycosidic bonds are present).

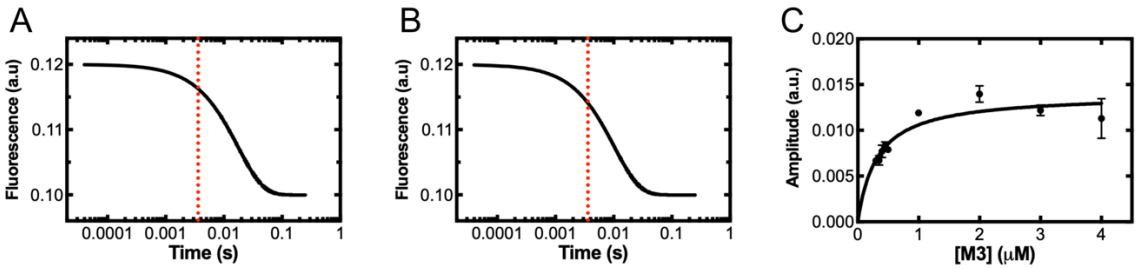


Fig. A1.3. Impact of stopped-flow dead-time on signal change amplitude of MalX A134C-MDCC upon mixing with M3. The stopped-flow apparatus has an ~3ms mixing delay before data is collected, which can result in uncertainty in signal change amplitude when compared to equilibrium state fluorescence data (e.g. collected by fluorescence spectrophotometry). Using the kinetics data produced in this study, we have modeled M3 binding to MalX A134C-MDCC to illustrate the impact of stopped-flow dead-time on the fluorescence time course reported in Fig. 4A. We plot a model one exponential function (Eq. 2.7) for MalX A134C-MDCC binding to M3 at ligand concentrations of 2 μM (A) and 4 μM (B) using equation $F = F_{\infty} + A_1 \times \exp(-k_{app}t)$, where F_{∞}= 0.1, A_1 = 0.02 (expected amplitude of signal change), and k_{app} = 57 (A) or 97 (B) (k_{app} determined from equation of line of best fit in Fig. 4b (y = 20x + 17). In each plot, a hashed red vertical line indicates the stopped-flow dead-time, where data points before the ~3ms mixing delay will not be collected. In the examples examined, approximately 20% (A) and 30% (B) of the expected fluorescence signal changes occur in the dead-time of the instrument and would not be visible in their respective fluorescence time courses. Amplitudes of signal changes from the stopped-flow data shown in Figure 4 are also reported (C, n = 9 data points, 4 replicates each).

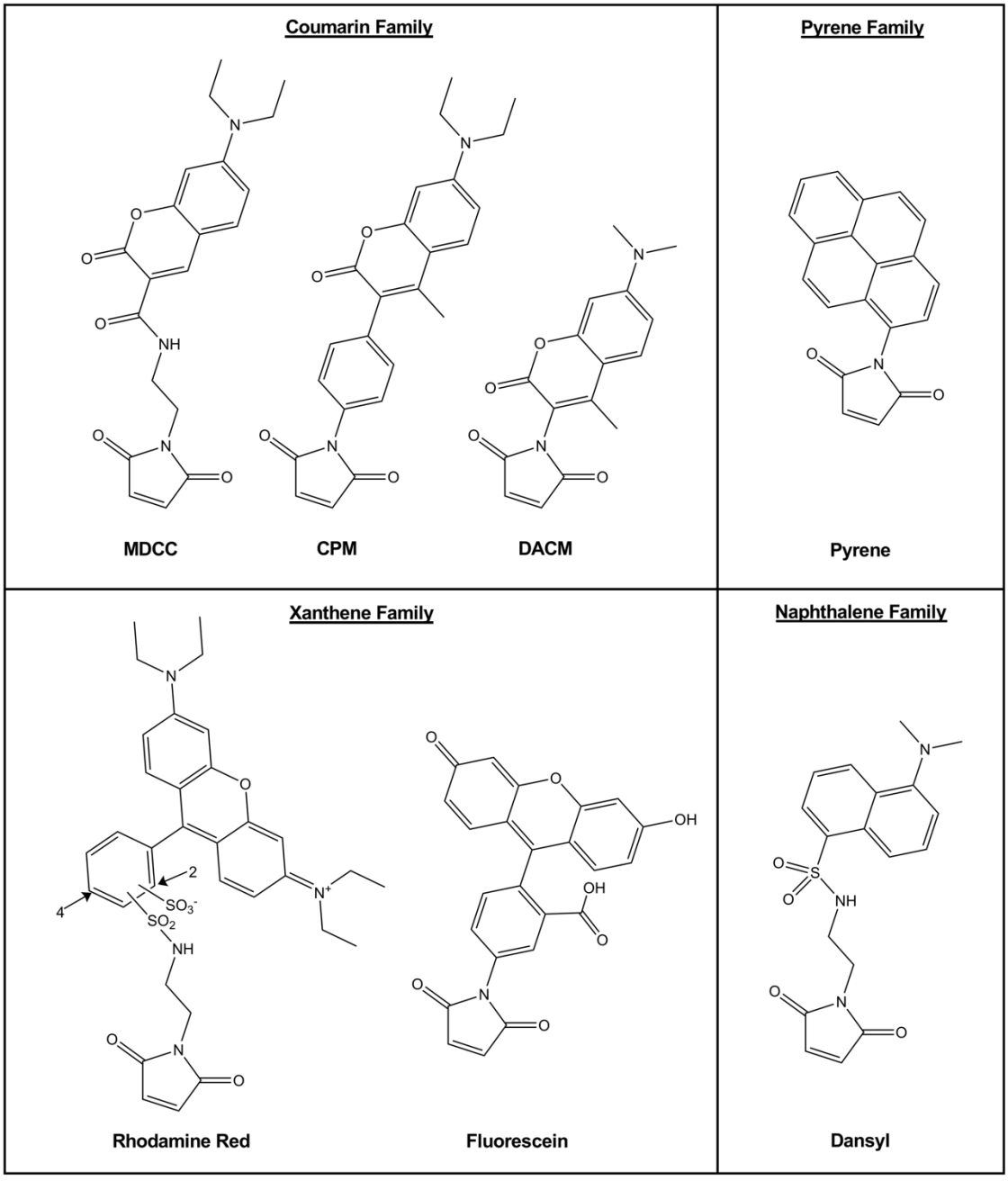


Fig. A1.4. Structural diagrams of thiol-reactive fluorophores used in this study (generated via ChemDraw Prime 16.0).

Table A1.1. MalX A134C-MDCC responds specifically to MOS. Changes in peak fluorescence intensity (ΔI) of MalX A134C-MDCC were examined upon addition of MOS (M3 or maltoheptaose) or upon addition of either a compositionally similar carbohydrate (isomaltotriose), or a structurally similar carbohydrate that contains an insufficient degree of polymerization to stabilize a ligand-bound state of MalX (maltose).

	M3 (10 μM)	maltoheptaose (10 μM)	isomaltotriose (10 μM)	maltose (35 μM)
ΔI	-29 ± 4 %	-28 ± 1 %	+0.1 ± 2 %	+1 ± 2 %

Table A1.2. MalX A134C conjugated to a variety of fluorophores selectively detects MOS target molecule M3. ΔI indicates change in peak fluorescence intensity upon addition of M3 (10 μM) or maltose (15 μM).

Fluorophore	Ex. max (nm)	Em. max (nm)	ΔI upon M3 addition	ΔI upon maltose addition
Pyrene	340	380	-13.4 ± 0.5 %	-2 ± 1 %
CPM	385	465	-19 ± 2%	-0.7 ± 0.4%
DACM	420	465	-17 ± 2%	+0.1 ± 0.9%
MDCC	420	465	-29 ± 4 %	+0.6 ± 1%
Fluorescein	495	515	+9 ± 5%	-1 ± 2%
Dansyl	280	350/520	-10 ± 2%/ +6.1 ± 0.6%	-2.8 ± 0.3%/ -2 ± 1%
	340	520	+4.6 ± 0.74%	-1 ± 1%
Rhodamine Red	560	580	+19 ± 6%	-0.2 ± 0.7%

Table A1.3. Comparison of kinetic properties between MalX A134C-MDCC and HC PhoS A197C-MDCC (Smith et al. 2017).

	HC PhoS A197C-MDCC	MalX A134C-MDCC
Species detected	Phosphate	MOS
ΔI upon ligand addition	+70% (Smith et al. 2017)	-29 ± 4 % (for M3 detection)
k_{on} ($\mu M^{-1}s^{-1}$)	33 ± 2 (Smith et al. 2017)	20 ± 2 (for M3 detection)
K_D (nM)	~100 (Brune et al. 1994)	190 ± 50 (for M3 detection)

Appendix 1 References:

Brune, M., Hunter, J.L., Corrie, J.E., Webb, M.R., 1994. Biochem. 33(27), 8262-8271.

Smith, D.D., Girodat, D., Wieden, H.-J., Selinger, L.B., 2017. Anal. Biochem. 537, 106-113.

APPENDIX 2: SUPPLEMENTARY INFORMATION FOR CHAPTER 3

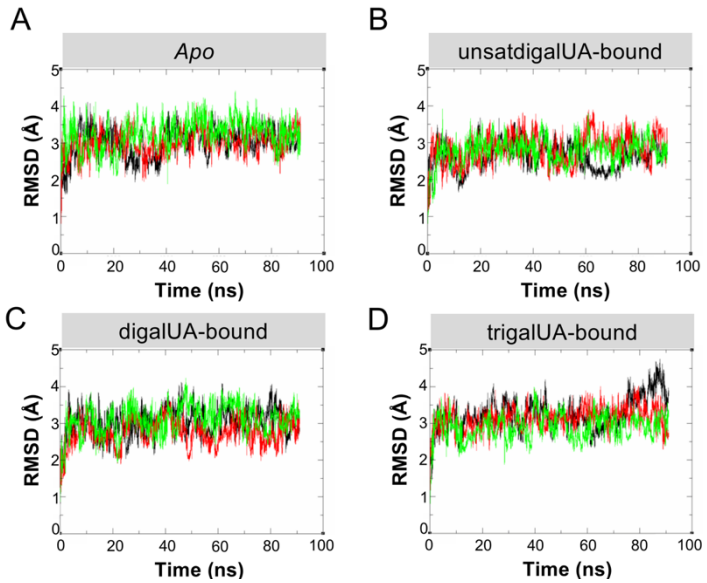


Fig. A2.1. RMSD plots of *apo* (A), unsatdigalUA-bound (B), digalUA-bound (C), and trigalUA-bound (D) TogB for the duration of molecular dynamics simulations. Mass-weighted RMSD fits were calculated using cpptraj (Roe and Cheatham III 2013) for each successive frame in the trajectory when compared to the starting structures. The initial 10 ns of the 100 ns simulations prior to system equilibration were removed before scoring (first 10 ns not shown). Replicates for each condition (R1: Black, R2: Red, R3: Green) remain stable for the duration of the simulations at an average RMSD of approximately 3Å from the starting structures.

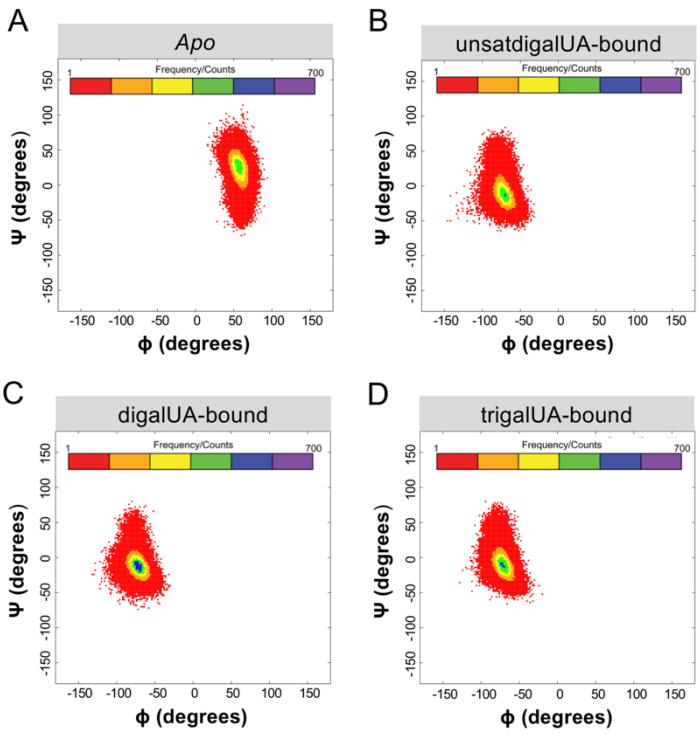


Fig. A2.2. Amino acid position D363 ($F_{Score2.0}$ = 1) in TogB adopts unique dihedral angles in *apo* vs. ligand-bound states. Representative Ramachandran Plots of D363 during molecular dynamics simulations of TogB. Amino acid dihedral angles are shown for TogB in its *apo* state (A), unsatdigalUA-bound state (B), digalUA-bound state (C), and trigalUA-bound state (D).

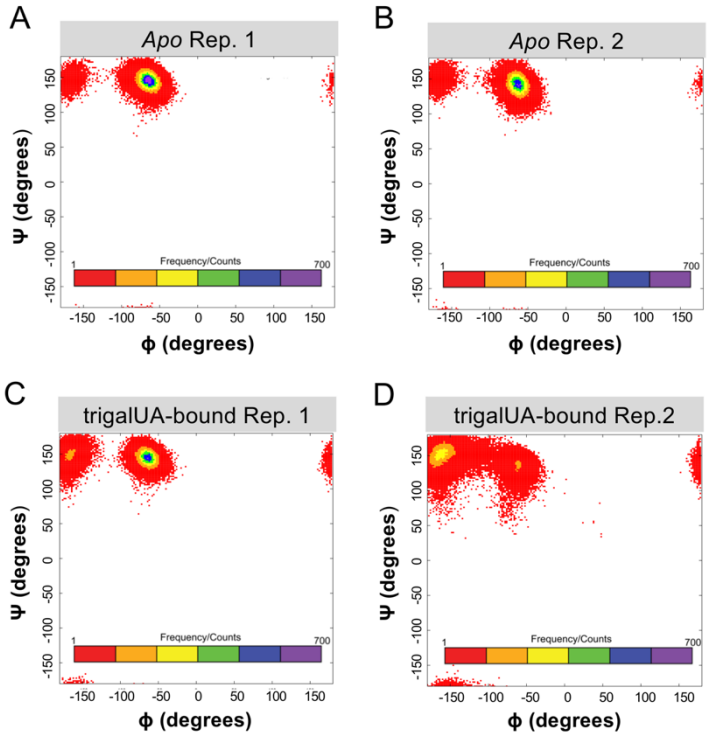


Fig. A2.3. Amino acid position A284 ($F_{Score2.0}$ = 0.4 ± 0.3 for trigalUA) in TogB exhibits a partial shift in the equilibrium between conformations upon ligand binding. Dihedral angles in *apo* (A, B) and trigalUA-bound (C, D) states are shown for two molecular dynamics simulation replicates, and two populations of dihedral angles are present for both states. The propensity for Position 284 to exist in one population of dihedral angles vs. another is altered upon ligand binding, leading to a change in fluorescent output by the ensemble *in vitro* upon ligand binding (Table 1).

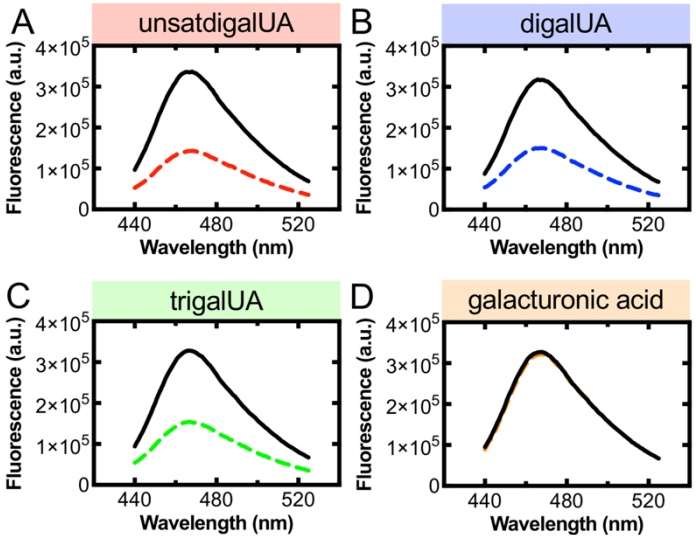


Fig. A2.4. Representative fluorescence emission spectra of TogB D363C-MDCC response to ligand. TogB D363C-MDCC (100 nM) in the absence (black solid lines)

and presence of 16 μM unsatdigalUA (A, dashed red line), or 48 μM digalUA (B, dashed blue line), or 570 μM trigalUA (C, dashed green line), or 1710 μM galacturonic acid (D, dashed orange line).

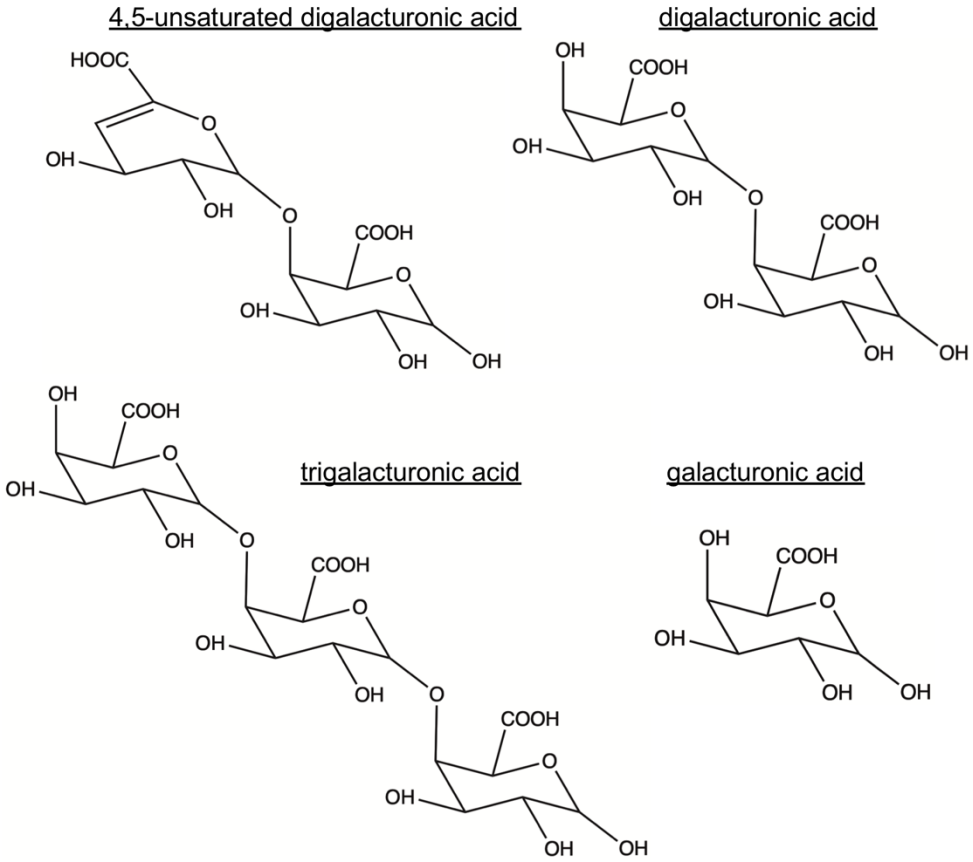


Fig. A2.5. Structural diagrams of carbohydrates utilized in this study (generated using ChemDraw 20.1). Galacturonic acid is a monomer, whereas unsatdigalUA, digalUA, and trigalUA are oligos of galacturonic acid joined via α-1,4- glycosidic bonds.

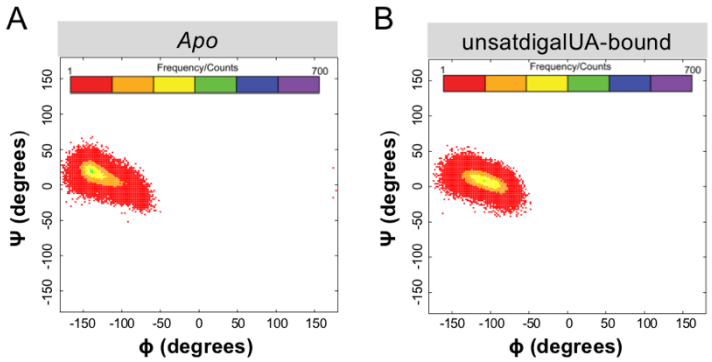


Fig. A2.6. Amino acid Position 99 in TogB ($F_{Score2.0}$ = 0.40 ± 0.08 for unsatdigalUA) exhibits a moderate shift in dihedral angles upon ligand binding. Dihedral angles from representative apo (A) and unsatdigalUA-bound (B) 100 ns molecular dynamics simulations are shown. There is a clear population shift upon ligand binding (resulting in a mid-range $F_{Score2.0}$ value), however the residence times in the respective microstates is not reflected and occupancy may not be long enough to result in an environment change of a conjugated fluorophore that can contribute to the *in vitro* (ensemble) measurement.

Table A2.1. Comparison of kinetic properties between TogB D363C-MDCC, MalX A134C-MDCC(Smith et al. 2022a), and HC PhoS A197C-MDCC(Smith et al. 2017).

	HC PhoS A197C-MDCC	MalX A134C-MDCC	TogB D363C-MDCC
Species detected	Phosphate	Maltooligosaccharides	Oligogalacturonides
ΔI upon ligand addition	+70% (Smith et al. 2017)	-29%[a] (Smith et al. 2022a)	-60% [b]
k_{on} ($\mu M^{-1}s^{-1}$)	33 ± 2 (Smith et al. 2017)	20 ± 2 [a] (Smith et al. 2022a)	18.6 ± 0.7 [b]
K_D (μM)	~0.100 (Brune et al. 1994)	0.190 ± 0.050 [a] (Smith et al. 2022a)	1.3 ± 0.5 [b]

[a] *For Maltotriose detection*
[b] *For unsatdigalUA detection*

Appendix 2 References:

Brune, M., Hunter, J.L., Corrie, J.E., Webb, M.R., 1994. Biochem. 33(27), 8262-8271.
Roe, D.R., Cheatham, T.E., 3rd, 2013. J Chem Theory Comput 9(7), 3084-3095
Smith, D.D., Girodat, D., Wieden, H.-J., Selinger, L.B., 2017. Anal. Biochem. 537, 106-113.
Smith, D.D., Girodat, D., Abbott, D.W., Wieden, H.-J., 2021. Biosensors Bioelectron., 113899.

APPENDIX 3: SUPPLEMENTARY INFORMATION FOR CHAPTER 4

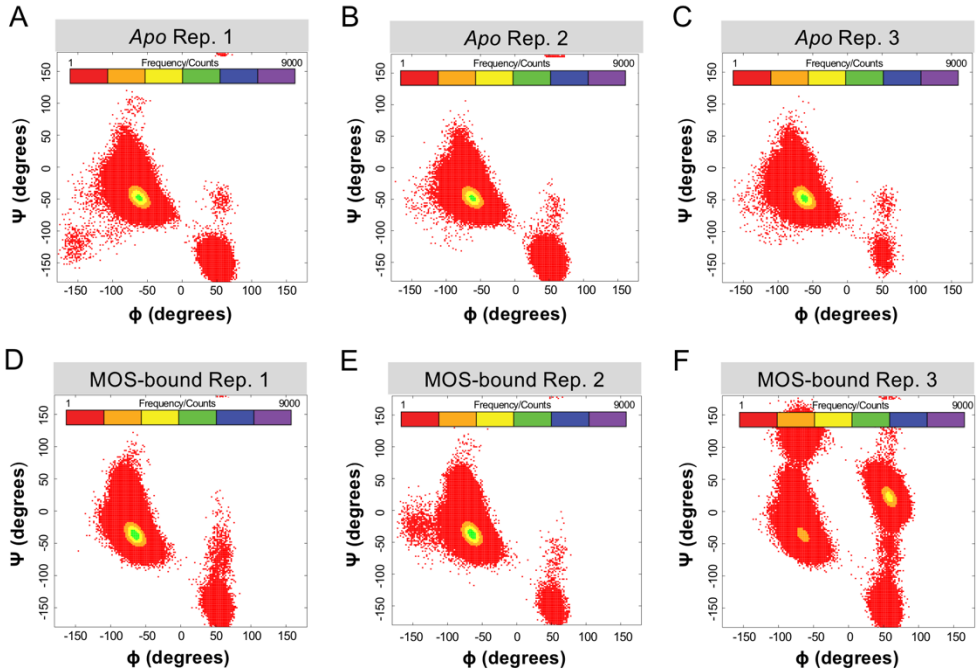


Fig. A3.1. Ramachandran plots of amino acid Position 134 (reporter position) of MalX (wt) during 1μs molecular dynamics simulation. Replicates from *apo* simulations are (A-C) and from MOS-bound (D-F) are shown. In the MOS-bound state, an additional and unique population of phi/psi angles is observed in Replicate 3 (F) which is not present in the other simulations (i.e. a readout based on addition of a population upon ligand binding).

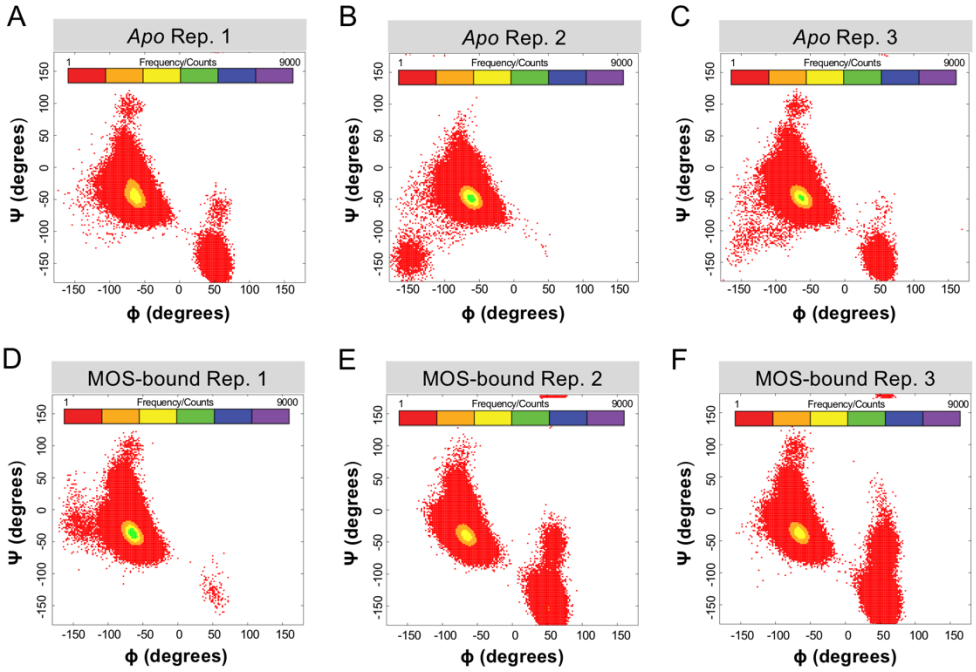


Fig. A3.2. Ramachandran plots of amino acid Position 134 (reporter position) of MalX D102A during 1μs molecular dynamics simulation. Replicates from *apo* simulations are (A-C) and from MOS-bound (D-F) are shown. In contrast to the other variants analyzed in this study, none of the above listed replicates had $F_{Score2.0}$ values above 0.5.

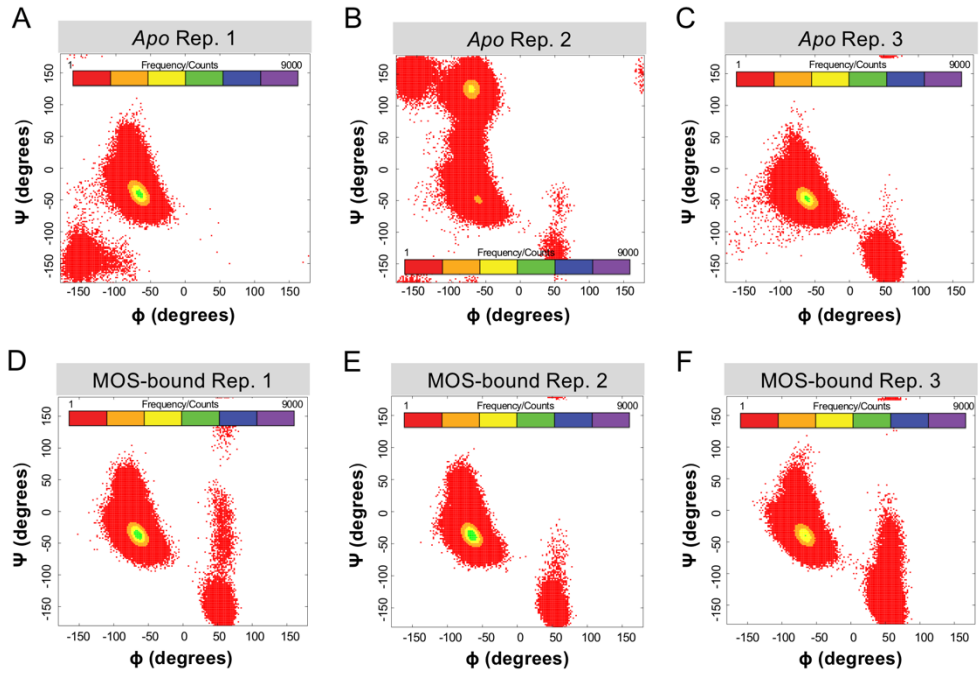


Fig. A3.3. Ramachandran plots of amino acid Position 134 (reporter position) of MalX N195A during 1μs molecular dynamics simulation. Replicates from *apo*

simulations are (A-C) and from MOS-bound (D-F) are shown. When comparing the Ramachandran plots in the *apo* and MOS-bound states, Replicate 2 in the *apo* simulations (B) displays a sub-population not observed in any of the MOS-bound simulations. (i.e. readout based on removal of an *apo* state population upon ligand binding).

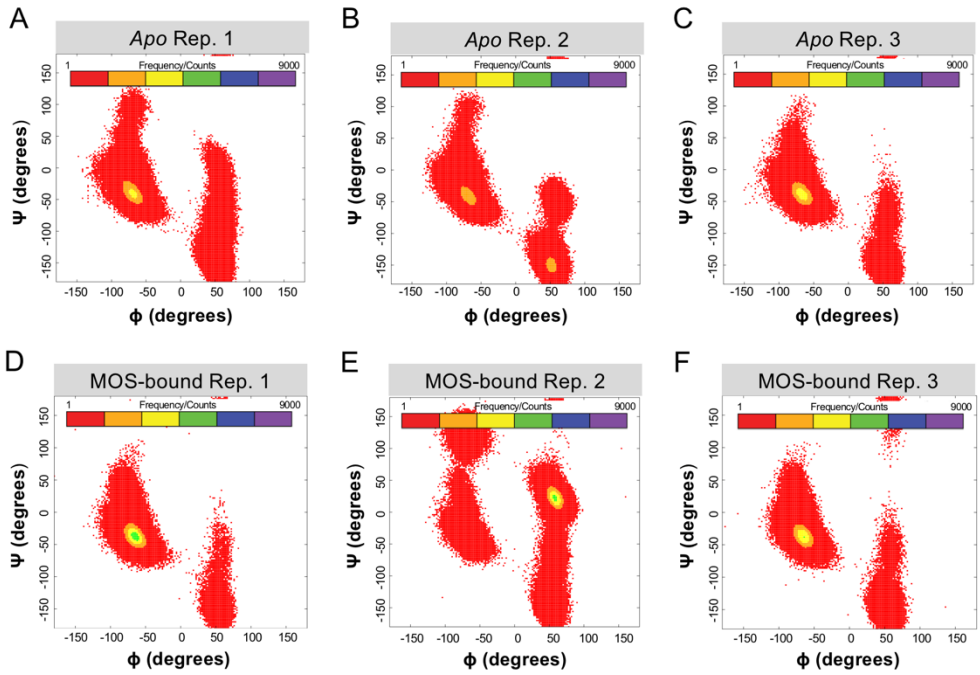


Fig. A3.4. Ramachandran plots of amino acid Position 134 (reporter position) of MalX K216A during 1μs molecular dynamics simulation. Replicates from *apo* simulations are (A-C) and from MOS-bound (D-F) are shown. In the MOS-bound state, an additional population of phi/psi angles is observed in Replicate 2 (E) which is not present in the other simulations (i.e. a readout based on addition of a population upon ligand binding).

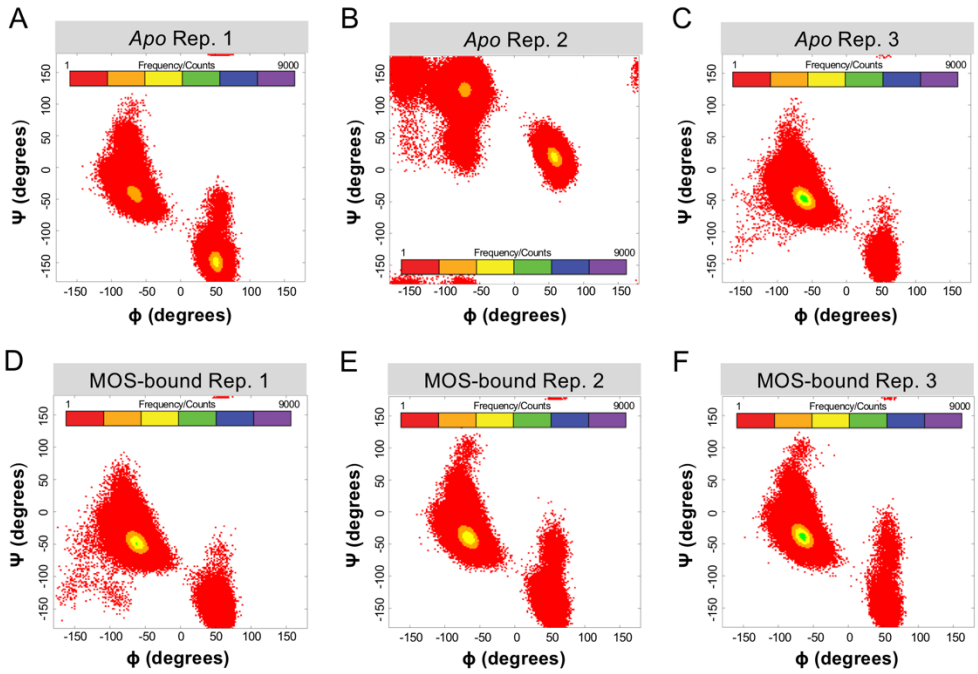


Fig. A3.5. Ramachandran plots of amino acid Position 134 (reporter position) of MalX W273A during 1μs molecular dynamics simulation. Replicates from *apo* simulations are (A-C) and from MOS-bound (D-F) are shown. When comparing the Ramachandran plots in the *apo* and MOS-bound states, Replicate 2 in the *apo* simulations (B) displays a sub-population not observed in any of the MOS-bound simulations (i.e. readout based on removal of an *apo* state population upon ligand binding).

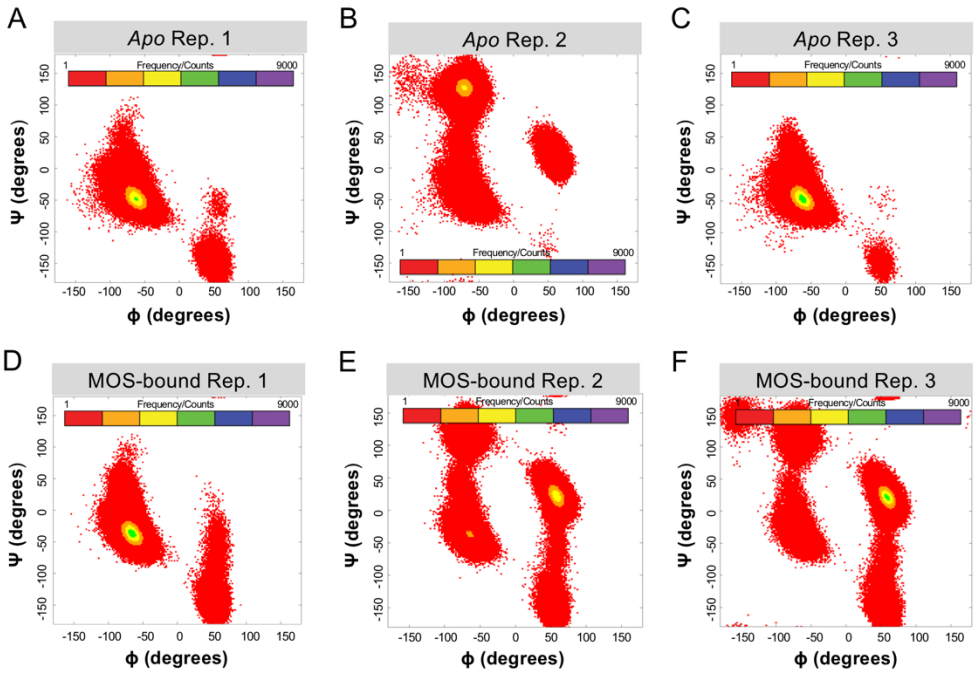


Fig. A3.6. Ramachandran plots of amino acid Position 134 (reporter position) of MalX Q372A during 1μs molecular dynamics simulation. Replicates from *apo* simulations are (A-C) and from MOS-bound (D-F) are shown. In the MOS-bound state, an additional population of phi/psi angles is observed in Replicates 2 & 3 (E, F) which is not present in the other simulations (i.e. a readout based on addition of a population upon ligand binding).

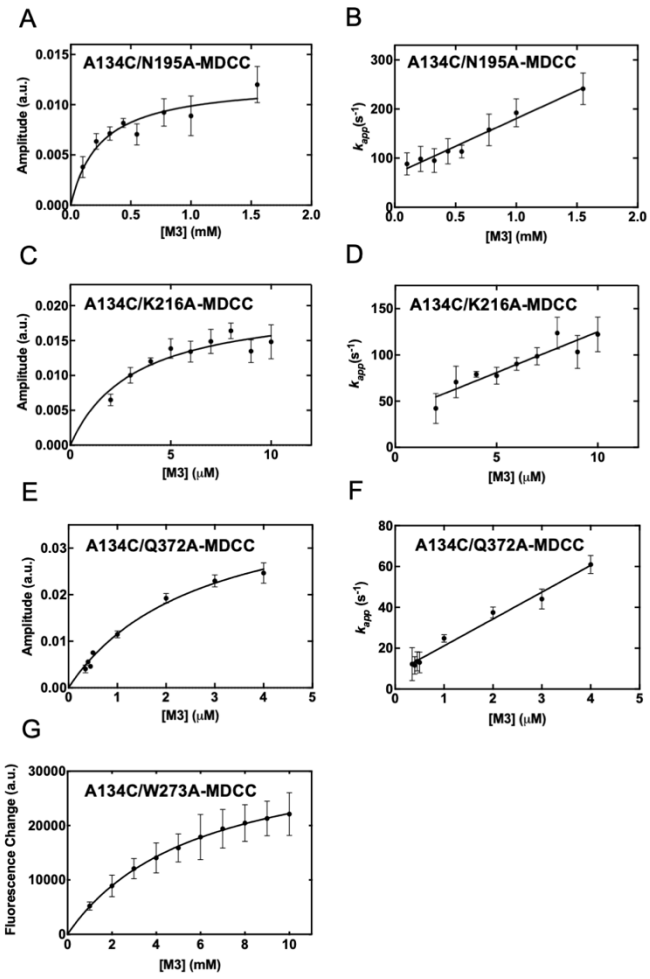


Fig. A3.7. Biochemical characterization of MalX variants using equilibrium fluorescence spectroscopy and the stopped-flow method. The stopped-flow method was used to determine binding affinity (K_D), and rate constant (k_{on}) of M3 binding to MalX variants by titrating M3 and fitting the resulting fluorescence time-courses with a single exponential function. Amplitudes (A) and apparent rates (k_{app}) (B) of these signal changes for MalX A134C/N195A-MDCC titrated with M3 are shown. Fitting of Amplitude values with a hyperbolic function yields K_D = 230 ± 90 **μ**M, and fitting of k_{app} values with a linear function yields k_{on} = 0.113 ± 0.008 **μ**$M^{-1}s^{-1}$. Signal change amplitude (C) and apparent rate (k_{app}) values (D) for MalX A134C/K216A-MDCC titrated with M3 are shown. Fitting of Amplitude values with a hyperbolic function yields K_D = 3.2 ± 0.8 **μ**M, and fitting of k_{app} values with a linear function yields k_{on} = 9 ± 1 **μ**$M^{-1}s^{-1}$. Signal change amplitude (E) and apparent rate (k_{app}) values (F) for MalX A134C/Q372A-MDCC titrated with M3 are shown. Fitting of Amplitude values with a hyperbolic function yields K_D = 2.7 ± 0.3 **μ**M, and fitting of k_{app} values with a linear function yields k_{on} = 13 ± 1 **μ**$M^{-1}s^{-1}$. Last, for MalX A134C/W273A-MDCC, a titration of M3 was analyzed via equilibrium fluorescence spectroscopy assay was performed. Fitting a hyperbolic function to the resulting fluorescence changes resulted in K_D = 5.9 ± 0.2 mM (G).