Structural and Enzymatic Comparison of *Faecalibacterium prausnitzii* GH31 α-Glycosidases

by

Anna Eva Jewczynko

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2022

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.
I understand that my thesis may be made electronically available to the public.

Abstract

The gut microbiome is home to thousands of species of bacteria, that are essential for human digestion, immunity and physiology. *Faecalibacterium prausnitzii* makes up about 5% of a healthy human gut microbiome and a lower abundance of this bacterium has been found in patients with IBD and Crohn's disease. Among an extensive repertoire of carbohydrate active enzymes, *F. prausnitzii* has 2 GH31 enzymes, which are from the same family as Sucrase-Isomaltase and Maltase-Glucoamylase, human digestive enzymes with overlapping and distinguishing substrate specificities. This thesis aims to characterize the substrate specificity, preference and inhibition sensitivity of *F. prausnitzii* GH31 α-glucosidases to better understand the structural features of GH31 enzymes and the biological capabilities of these bacteria. In this thesis, AlphaFoldV2.1.0 was used to create computational models of *F. prausnitzii* α-glucosidases, and the substrate specificity, enzyme kinetics and inhibition parameters are reported. Structurally, these α-glucosidases have the same identified conserved N-terminal and $(\beta/\alpha)_8$ barrel domains, but FpAG1 has an additional conserved domain of unknown function at the C-terminus which is not found in the FpAG2 structure. Both FpAG1 and FpAG2 have α-glucosidase and oligo-1,6-glucosidase activity. The comparative kinetic studies show that FpAG1 has a greater preference for α-1,6 glycosidic linkages, and FpAG2 has a greater preference for α-1,4 glycosidic linkages. The comparative inhibition studies show that the tested α-glucosidase inhibitors, acarbose, miglitol and kotalanol, are more potent in FpAG2 than FpAG1, and acarbose is a weak inhibitor of *F. prausnitzii* α-glucosidases. Distinguishing binding affinities of miglitol and kotalanol in these GH31 enzymes suggest structural differences in the FpAG1 and FpaG2 active sites. Gaining insight on the GH31 α-glucosidases as a component of *F. prausnitzii* metabolism can further our understanding of this community in the human gut microbiome.

Acknowledgements

I would like to start by thanking Dr. David Rose for taking me on as a Master's student. Thank you for giving me the opportunity to explore the field of structural biology, the confidence to take on new challenges, and supporting me throughout my journey as a graduate student. Thank you to my committee members, Dr. Todd Holyoak and Dr. John Honek, for providing insight into my research and helping to guide my work.

Thank you to the Rose Lab and Holyoak Lab members for sharing knowledge of new methods and analyses, helping me troubleshoot the many hurdles I've encountered and celebrating the milestones. In particular, I would like to thank Nardo for always encouraging me to try new techniques, to take that extra step and for providing me with the many needed hallway pep talks. Your infectious excitement for science has influenced my perspective on research and I can't thank you enough for your constant support.

Thank you to my UWaterloo Gaggle for your unwavering friendship and cheerleading throughout these last few years, as we navigated this interesting time in our lives together through *virtual* hang outs and book clubs. Finally, thank you, Mama and Tato, for being my biggest supporters, always providing a listening ear and always being there to remind me of what matters most at the end of the day.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BLAST	Basic Local Alignment Search
CASP14	Critical Assessment of Structure Prediction 14
CAZy	Carbohydrate Active Enzymes database
CDD	Conserved Domain Database
DUF5110	Domain of unknown function 5110
FpAG1	GH31 a-glucosidase found in *F. prausnitzii* A2-165 (QIA42631.1)
FpAG2	GH31 a-glucosidase found in *F. prausnitzii* A2-165 (QIA43029.1)
GAGO	Sigma-Aldrich Glucose (GO) Assay Kit
GDT	Global Distance Test
GH	Glycoside Hydrolase
GOPOD	Megazyme Glucose oxidase/ peroxidase Assay Kit
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β- D-1-thiogalactopyranoside
LB Media	Luria-Bertani Media
LDDT-Cα	Local Distance Differentiation Test on Cα atoms
MalA	*Sulfolobus solfataricus* GH31 α-glucosidase
MES	2-(N-morpholino)ethanesulfonic acid
MGAM	Maltase-Glucoamylase
mRNA	Messenger ribonucleic acid
MSA	Multiple Sequence Alignment
NCBI	National Center for Biotechnology Information
ntMGAM	N-terminal subunit Maltase-Glucoamylase
ntSI	N-terminal subunit of Sucrase-Isomaltase
PAE	Predicted Aligned Error
pI	Isoelectric point
PIPES	Piperazine-N,N′-bis(2-ethanesulfonic acid)
pK_a	Acid dissociation constant
pLDDT	Predicted Score of the LDDT-Cα metric
RMSD	Root-mean-square deviation
Ro-αG1	GH31 α-glucosidase found in *Blautia obeum*
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SI	Sucrase-Isomaltase
TM-score	Template Modelling Score
tRNA	Transfer ribonucleic acid

Chapter 1: Introduction

1.1 Thesis Introduction

The human gut microbiome has been described as an additional human organ, due to its integrated and essential role in human metabolism, immunity and physiology[1]. There are over a thousand species of bacteria and trillions of cells in the gastrointestinal tract[1,2]. The majority of the bacterial phyla composing this microbiome are Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria and Fusobacteria[3]. The relationship between the gut microbiome and human host is bi-directional, meaning that factors such as exercise, diet, and stress[4–6] can contribute to the composition of the gut microbiome, and as a result the gut microbiota can influence factors of human health such as hormones, digestion and the immune system[7–9]. Multiple studies are researching the effects of gut microbial composition on bone density[10], kidney stones[11], fertility[12], type 1 diabetes[13] and mental health[14,15].

Gaining further insight on carbohydrate active enzymes can provide us with information on how the food we eat is digested. Human α-glucosidases, sucrase-isomaltase (SI) and maltase-glucoamylase (MGAM), are maltases at the small intestinal border that completely digest dietary starches in the human digestive tract[16,17]. These heterodimeric enzymes are part of the Glycoside Hydrolase (GH) 31 family and all four subunits hydrolyze the α-1,4 glycosidic bond in maltose[16]. SI is responsible for 60-80% of maltose hydrolysis in the intestinal tract[18]. In addition to the maltase activity, the subunits in SI have distinguishing activities, with the N-terminal subunit capable of hydrolyzing α-1,6 and the C-terminal subunit capable of hydrolyzing α-1,2 glycosidic linkages[16,19,20]. Microbes in the intestinal tract contribute to the digestion of carbohydrates, because as a whole, they contain a greater catalogue of enzymes than the human host[21]. This variety of catalytic activity allows the microbes to utilize carbohydrates that otherwise wouldn't be digested, resulting in about 10% of human's dietary energy being consumed from the products of microbes in the intestine[22].

Faecalibacterium prausnitzii has been identified to compose about 5% of a healthy gastrointestinal bacterial population[23,24]. *F. prausnitzii* is a Firmicute, part of the Ruminococcaceae family and Clostridia genus (Clostridium cluster IV)[24]. Large proportions of this Gram positive anaerobic bacterium[25] have been identified in the colon, with a lower abundance found in the terminal ileum and duodenum[26,27]. *F. prausnitzii* is a significant butyrate producer in humans[28,29]. Butyrate is a short chain fatty acid that has an important role in maintenance of the intestinal barrier[24], signaling and differentiation of immune cells[9,30], and regulation of inflammation reponse[31,32]. Increased amounts of this microbe have been identified in athletes[33] and low carbohydrate high fat diets have resulted in a detrimental effect on *F. prausnitzii* abundance in the gastrointestinal tract[34]. This bacterium has anti-inflammatory properties and a lower abundance has been observed in many gut disorders[28,35], including in patients with Inflammatory Bowel Disease (IBD)[36], Crohn's disease[37] and patients hospitalized with Covid-19[38].

CAZy (Carbohydrate Active Enzymes database; http://www.cazy.org/) is a database for enzymes that are predicted to interact with glycosidic bonds, by either generating, degrading or altering them[39]. Within this database, the classes of enzymes are categorized by structural features determined with the primary amino acid sequence. In the CAZy database, the GH class includes enzymes that are capable of hydrolyzing glycosidic bonds[40,41]. The most common mechanisms for catalysis predicted in this enzyme class include the acid and base catalytic residues, which can hydrolyze the glycosidic bond through either the overall retaining mechanism or the overall inversion of anomeric configuration[42]. In some exceptional cases, the utilization of an acetamido group at the C-2 of the substrate[43] (ex. GH19) or the use of a co-factor[44,45] (ex. GH4 and GH 109) have been identified to play the role of the catalytic nucleophilic residue. Within the class, there are also clans of enzymes categorized by the structural fold. These enzymes are further funneled into families, which are categorized by the structural fold, mechanism and catalytic machinery.

In *Faecalibacterium prausnitzii*, there are 2 glycoside hydrolases found in family 31 on the CAZy database in strain A2-165. For clarity, *F. prausnitzii* GH31 α-glucosidases will be referred to individually as FpAG1 (QIA42631.1; GXM22_05840) and FpAG2 (QIA43029.1; GXM22_08100). The GH31 family includes enzymes in the GH-D clan, which have a $(\beta/\alpha)_8$ barrel, an Asp residue as both catalytic nucleophile and proton donor, and have been predicted to hydrolyze glycosidic bonds through the overall retaining mechanism[46–48]. The CAZy database recognizes the conserved folds in enzymes and their relationship with other glycoside hydrolases. These structural features don't confirm substrate specificity because the determined structural folds are not specific enough to identify local variations to structures, so further analysis will need to be done to determine the role of these α-glucosidases in *F. prausnitzii*.

1.2 Thesis Objectives

The objectives of this project are to characterize the GH31 α-glucosidases found in *F. prausnitzii* through biochemical and structural studies to gain insight into the substrate specificity, substrate preference and inhibition sensitivity of these carbohydrate-active enzymes. These studies will compare the structural features and enzymatic kinetics to identify similarities and differences between the *F. prausnitzii* α-glucosidases. Identifying microbes that are important to the function of the gastrointestinal system and looking at the molecular structures of GH31 α-glucosidases could help gain further insight into the biological capabilities of these microbes and their role in starch digestion.

Chapter 2: Computational Analysis *F. prausnitzii* α-Glucosidases

2.1 Introduction

2.1.1 Sequence Analysis

BLAST (Basic Local Alignment Search) is a tool supported by the National Center for Biotechnology Information (NCBI)[https://blast.ncbi.nlm.nih.gov/Blast.cgi]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2022 February 02], available from: https://www.ncbi.nlm.nih.gov/.[49] This tool compares an extensive library of sequences to gain information about the query sequence, such as identifying homologous sequences or conserved features and domains. The CD-Search in BLAST looks for conserved domains within a protein or nucleotide sequence with RPS-BLAST, which scans a set of position-specific scoring matrices to find associations between the query sequence and conserved domains[50,51]. The Conserved Domain Database (CDD) is a database that includes characterized domains from NCBI and Pfam, SMART, COG, PRK, and TIGRFAM[51]. In the BLAST CD-Search results, a hashtag indicates that the residue site has been annotated in an NCBI-curated domain because it is involved in a conserved feature of the domain. The E value shows the result's significance, as the score given specifies the likelihood that the results were found by chance[50]. The bit score is a statistical determinate of sequence similarity, and it can be used for inferring homology[52]. This score is normalized to the sequence length and database size[52].

EMBOSS Matcher, a tool supported by the European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK, https://www.ebi.ac.uk/Tools/psa/emboss_matcher/, is a local pairwise sequence alignment algorithm, where stretches of the sequence with high similarity are aligned, and the identity and similarity of the sequences are analyzed[53].

2.1.2 Computational Protein Structure Modelling

One approach to the structure of a protein is to predict a model based on the primary amino acid sequence using AlphaFoldV2.1.0[54,55], an artificial intelligence system developed by Deep Mind and the European Molecular Biology Lab – European Bioinformatics Institute EMBL- EBI, (AlphaFold Data Copyright (2021) DeepMind Technologies Limited). With this model, the structural features of the *F. prausnitzii* α-glucosidase sequences can be analyzed. AlphaFoldV2.1.0 is an artificial intelligence program that can predict protein structures with atomic accuracy[54]. AlphaFoldV2.1.0 analyzes the amino acid sequence to predict the orientation and the distance between the amino acids in the 3D structure, and then these parameters are used to create the predicted structure[54]. The Multiple Sequence Alignment is created using publicly available databases uniref90, smallbfd and mgnify, to collect sequences that are thought to be evolutionarily related to the target sequence[54]. The Evoformer module of AlphaFoldV2.1.0 develops a relationship between the protein residues. The MSA alignment and the pair representation of the target protein are updated repeatedly as more information is determined about the sequence[54]. The Structure Prediction module takes this information and fits each residue into a 3D space by predicting a translation and rotation of each distinct residue[54]. The side chains are predicted by determining Chi angles so that there are reasonable angles and bond lengths present in the structure, and a minimization step is used on the final structure to remove clashes[54]. The accuracy of this deep learning algorithm is competitive with experimental structure determination methods, as it has been trained using protein chains found in the PDB. According to the Protein Structure Prediction Center (US National Institute of General Medical Sciences (NIH/NIGMS), © 2007-2020, University of California, Davis), AlphaFoldV2.1.0 was found successful at predicting protein structures within a median distance of 0.96 Å of the experimentally determined protein models, and this is a significant improvement compared to the next-best method, which had a 2.83 Å median distance[54] as stated in the Critical Assessment of Structure Prediction (CASP14) results in 2020

(https://predictioncenter.org/casp14/zscores_final.cgi)

AlphaFoldV2.1.0 outputs confidence estimates alongside its predicted protein structure. The Predicted Score of the LDDT-Cα metric (pLDDT) presents the confidence for intra-domain structure predictions[54]. This parameter is the prediction score of each residue with the Local Distance Differentiation Test on Cα atoms (LDDT-Cα) metric, which analyses the distance of atoms in the model and the stereochemistry of the residues. The pLDDT shows the confidence in the local structure in individual domains of a protein, on a scale of 0 to 100, where any value below 50 predicts structure disorder and no meaningful coordinates, and a score greater than 90 shows high confidence in the local structure[54]. Although the pLDDT can determine confidence in the individual domains, the Predicted Aligned Error (PAE) is calculated to determine the confidence between the different domains in the whole structure. In the PAE 2D plot output, the confidence of the relative position between two residues is calculated, where residue y is aligned to the true structure, and the position error at residue x is measured[54]. This value is used to evaluate the relative position of the domains in the structure. With these metrics, the confidence of the determined AlphaFoldV2.1.0 protein structure can be estimated.

2.1.3 Protein Model Structure Comparison

Tools such as DeepAlign[56,57] from the RaptorX Structure Alignment server developed by the Xu group at University of Chicago (http://raptorx.uchicago.edu/DeepAlign/submit/) analyze protein structures and determine parameters such as the root-mean-square deviation (RMSD), Global Distance Test (GDT), and the template modelling score (TM-score), to compare the similarity of two protein structures. The RMSD represents the average distance between equivalent atoms, such as the Cα in the protein backbone, of two structures[58]. The uGDT/GDT score measures the quality of the structure alignment by analyzing and weighing the RMSD of the residue pairs in the aligned structures[56,59]. The variety of RMSD values between the structures are ranked and the output is a single value (uGDT). The value is then normalized to the shortest protein length to obtain the GDT.

It compares the protein backbone of the aligned structures and the score ranges from 0 to 100, with the higher scores suggesting the structural similarity. A TM-score compares the structural similarity of the two proteins, focusing on the global fold similarity over the local structural variations. The normalized score is between 0 and 1, where 1 indicates a perfect superposition, and a score above 0.6 assumes that there is >90% chance that the protein folds are similar between the two structures[60–62]. These programs will be used to determine key structural similarities and differences of the protein backbone that may account for differentiating enzymatic activities between the *F. prausnitzii* α-glucosidases.

2.2 Chapter Objective

This chapter aims to analyze the primary amino acid sequence and create confident computational protein models of *F. prausnitzii* α-glucosidases. Through programs such as BLAST CD-Search, and RaptorX DeepAlign, FpAG1 and FpAG2 protein models can be investigated to determine structural and functional similarities and differences between these proteins.

2.3 Material & Methods

The sequences used are FpAG1 alpha-glucosidase [Faecalibacterium prausnitzii A2-165] (GenBank: QIA42631.1; CDS: GXM22_05840), 684 residues, and FpAG2 alpha-glucosidase [Faecalibacterium prausnitzii A2-165] (GenBank: QIA43029.1; CDS: GXM22_08100), 671 residues. These sequences were inputted into BLAST CD-Search[50,51], where the sequences were searched against the database CDD v3.19 – 58235 PSSMs, with the "expect value threshold" of 0.01. A composition-based statistics adjustment was applied, the maximum number of hits was 500, and the chosen results mode was concise. To compare these sequences to other GH31 enzymes sequences, a Local Pairwise Sequence Alignment was performed using EMBOSS Matcher, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome

Campus, Hinxton, Cambridge, UK, https://www.ebi.ac.uk/Tools/psa/emboss_matcher/[53]. To predict computational models, these sequences were inputted as sequence_1 into the simplified version of AlphaFold v2.1.0 in the Google Colab Notebook[54]. The single-chain model was created, and the "is_prokaryote" feature relaxation setting "run_relax" was used to consider stereochemical violations. The RaptorX Structure alignment server[56,57]was used to align the whole structures and individual domains, as indicated in Table 3.

2.4 Results and Discussion

2.4.1 FpAG1 and FpAG2 Sequence Analysis

Table 1. Conserved Domains in FpAG1 and FpAG2 Identified with BLAST CD-Search

	Conserved Domain	Residue Range	Hits	Bit Score	E Score	Conserved Residues Identified
FpAG1	GH31_glucosidase_II_MalA domain (Alphα-glucosidase II-like)	149-517	86	525.54	0e+0	W166, D194, I195, I231, W268, W302, D304*, M305, R401, W414, D417*, H419, F450 and H475
	GH31_N (N-terminal domain of glycosyl hydrolase family 31 (GH31))	37-149	587	105.35	8.20e-27	D70
	Domain of unknown function (DUF5110)	632-675	321	39.91	2.60e-04	N/A
FpAG2	GH31_glucosidase_II_MalA domain (Alphα-glucosidase II-like)	150-522	86	523.61	0e+00	W167, D195, I196, I232, W269, W303, D305*, M306, R404, W417, D420*, N422, F453 and H478
	GH31_N (N-terminal domain of glycosyl hydrolase family 31 (GH31))	38-150	587	88.40	6.38e-21	D70

* Catalytic residues identified in sequence

In BLAST CD-Search, the FpAG1 and FpAG2 primary amino acid sequences were analyzed. The conserved domain GH31_glucosidase_II_MalA domain (Alphα-glucosidase II-like) was identified

in both these sequences, corresponding to the residues in the $(\beta/\alpha)_8$ barrel fold. In the N-terminus of these sequences, residues 37-149 in FpAG1 and 37-149 in FpAG2, the conserved domain GH31_N (N-terminal domain of glycosyl hydrolase family 31 (GH31)) was identified. In the FpAG1 sequence, a Domain of unknown function (DUF5110) was identified, which is not present in the FpAG2 sequence. The number of hits, the bit score, the E-score and the conserved residues identified in this domain are also highlighted in this summary. Catalytic residues identified are indicated with an asterisk (*). The alignments can be found in Appendix A.

Table 2. Sequence Identity and Similarity between FpAG1, FpAG2 and GH31 enzymes

	FpAG1		FpAG2	
	Seq ID (%)	Seq Similarity (%)	Seq ID (%)	Seq Similarity (%)
ntMGAM [PDB 2QLY]	28.9	44.2	27.3	45.1
ntSI [PDB 3LPO]	28.1	45.4	26.4	45.7
Ro-αG1 [PDB 3N04]	56.2	71.7	51.0	68.5
MalA [PDB 2G3M]	31.1	49.7	28.3	46.7
FpAG2	50.9	66.9	N/A	N/A

A Local Pairwise Sequence Alignment was performed using EMBOSS Matcher, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK, https://www.ebi.ac.uk/Tools/psa/emboss_matcher/[53], to determine the sequence identity and the sequence similarity.

The primary amino acid sequence can provide insight into the structure and function of enzymes. To determine conserved structural features and key residues, the sequences of the *F. prausnitzii* α-glucosidases were analyzed using BLAST CD-Search. A summary of these results can be found in Table 1, and the alignments can be found in Appendix A. A conserved catalytic domain, GH31_glucosidase_II_MalA domain (Alphα-glucosidase II-like) and the same conserved residue sites were identified by BLAST CD-Search in both *F. prausnitzii* primary amino acid sequences. The low E value and high bit scores determined for FpAG1 and FpAG2 sequences indicate a significant

probability of identifying the catalytic domain. Within the *F. prausnitzii* α-glucosidases that share 66.9% sequence similarity, the same residue sites were identified as part of the active site in this domain. These residues have been predicted to be involved in catalysis, substrate specificity and the formation of the active site[63–65]. The nomenclature for carbohydrate-binding sites in GH enzymes indicates that the -1 subsite is the non-reducing end of the carbohydrate. The + 1 subsite is the reducing end of the carbohydrate occupying the active site, adjacent to the region where hydrolysis occurs[66]. In GH31 enzymes, the overall retaining mechanism is the expected mechanism to hydrolyze glycosidic bonds[42]. Due to significant sequence similarity with homologous proteins (Table 2), the same aspartic acid residue positions were identified in FpAG1 and FpAG2 as the catalytic residues, as indicated in Table 1. The catalytic nucleophile is predicted to be Asp304 in FpAG1 (Asp305 in FpAG2) due to experimental data done in ntMGAM, YicI and Ro-αG1[47,64,67]. Due to experimental data done on Ro-αG1[64], Asp417 in FpAG1 (Asp420 in FpAG2) is predicted to act as the proton donor in the enzymatic reaction. A critical residue identified in substrate affinity in the active site of GH31 enzymes is the characteristic tryptophan in the active site (Trp166 in FpAG1/Trp167 in FpAG2). In GH31 enzymes where a tyrosine occupies this position, such as in ntMGAM and MalA, higher specificity for α-1,4 glycosidic linkages has been observed[20,65]. In GH31 enzymes that have α-1,6 catalytic activity, such as the N-terminal subunit of SI (ntSI) and Ro-αG1[20,64], a tryptophan occupies this residue site. Ro-αG1 [PDB ID 3N04] is an α-glucosidase from *Ruminococcus obeum* (renamed *Blautia obeum*), with high similarity with FpAG1 and FpAG2, 71.7 and 68.5%, respectively. Tan et al, 2010 proposed that the tryptophan residue (Trp169) is important for substrate specificity, as this residue was observed to interact with the nonreducing end at the -1 subsite of substrates with a α-1,6 glycosidic bond, such as isomaltose[64]. Replacing this residue with a tyrosine resulted in a dramatic reduction in α-1,6 glycosidic bond catalytic activity[64]. Therefore, due to the conservation of the aspartic acid catalytic residues, the characteristic tryptophan residue in the active site, and the

high sequence similarity, the *F. prausnitzii* α-glucosidases are predicted to hydrolyze α-1,6 glycosidic bonds, enzymatic activity comparable to Ro-αG1.

There are differences in the conserved residues identified by BLAST CD-Search, in the position His419/Asn422. The residue occupying this position is near the possible catalytic proton donor (Asp417/Asp420). The consensus sequence in BLAST CD-Search identifies this site as a serine. The asparagine is expected to be more flexible than the histidine. As these residues are near the proton donor catalytic residue in the active site, this may contribute to the flexibility of that loop in the active site[68,69]. The difference in physicochemical properties between the residues in this conserved site may contribute to the enzymatic reaction in the active site between the *F. prausnitzii* α-glucosidases. In Ro-αG1, the loops in the active site between the determined structures were observed to be very flexible[64]. The presence of a bound ligand, such as isomaltose, caused significant changes in the position of the active site residues[64]. The significant sequence similarity between the *F. prausnitzii* α-glucosidases and Ro-αG1 indicates that a similar effect may be expected in the active site of *F. prausnitzii* α-glucosidases. However, more experimental analysis would be needed to draw confident conclusions.

Within the catalytic domain of both of the *F. prausnitzii* α-glucosidases, there are two inserts present in the sequence. Although the presence of these inserts is consistent with the Ro-αG1 sequence, they are not highly conserved among homologous models[64]. In the BLAST CD-Search FpAG2 sequence analysis, a gap was identified in the conserved catalytic domain, between residues 319-354. This gap was not indicated in the BLAST CD-Search FpAG1 sequence analysis, suggesting greater homology in that region within the catalytic domain when compared to the homologous sequences in the CDD. In the Ro-αG1 protein structure, these inserts, particularly Insert 2, were predicted to contribute to forming a 4 α-helical bundle in the homodimer[64].

A conserved GH31_N (N-terminal domain of glycosyl hydrolase family 31 (GH31) was identified in both *F. prausnitzii* α-glucosidase sequences (Appendix A). There is a difference in

statistics identifying this region, with the N-terminal domain identified in the FpAG1 sequence having higher confidence compared to FpAG2 sequence, indicating greater sequence similarity and homology with other homologous sequences in the database[52] and a reduced likelihood that the results were found by chance[50]. As there is a highly conserved residue in this domain that is expected to interact with the +1 subsite in the active site of GH31 enzymes, the conservation of this domain is of interest for the enzymatic reaction of the active site. The conserved residue indicated at residue site Asp70 in FpAG1 (Asp70 in FpAG2), which is present on a loop in the conserved N-terminal β-sandwich domain, is also found in Ro-αG1, ntMGAM and MalA[63–65]. This residue has been proposed to initiate the enzymatic reaction by interacting with the carbohydrate in the +1 subsite, and mutation of this residue in Ro-αG1 eliminated enzymatic activity[64].

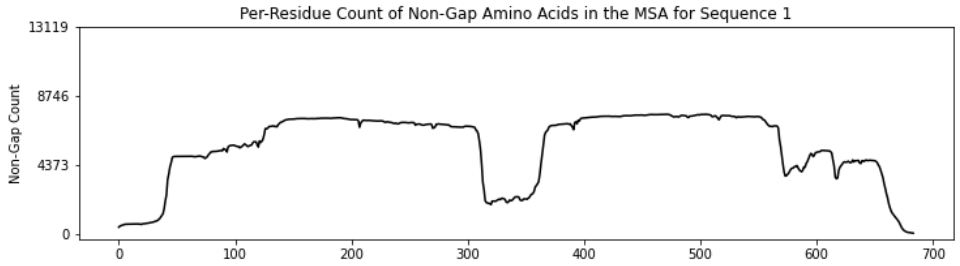
BLAST CD-Search identified an additional conserved domain in FpAG1, which was not identified in FpAG2 (Appendix A). This domain is described as 'Domain of unknown function' (DUF5110), and it is predicted to be involved in carbohydrate binding[51]. The significance of this domain is much lower than the conserved N-terminal and catalytic domains identified in these proteins, as there is an increased likelihood that the results were found by chance[50]. The presence of this domain may be noteworthy when comparing the function of *F. prausnitzii* α-glucosidases. The truncated version of the distal C-terminal domain in FpAG2 compared to FpAG1 may contribute to the lack of recognition as a 'Domain of unknown function 5110' in BLAST CD-Search.

2.4.2 FpAG1 and FpAG2 Computational Protein Structure Modelling

A

```
9608 unique sequences found in uniref90 for sequence 1
3179 unique sequences found in smallbfd for sequence 1
500 unique sequences found in mgnify for sequence 1

13119 unique sequences found in total for sequence 1
```



B

```
9597 unique sequences found in uniref90 for sequence 1
3310 unique sequences found in smallbfd for sequence 1
500 unique sequences found in mgnify for sequence 1

13246 unique sequences found in total for sequence 1
```

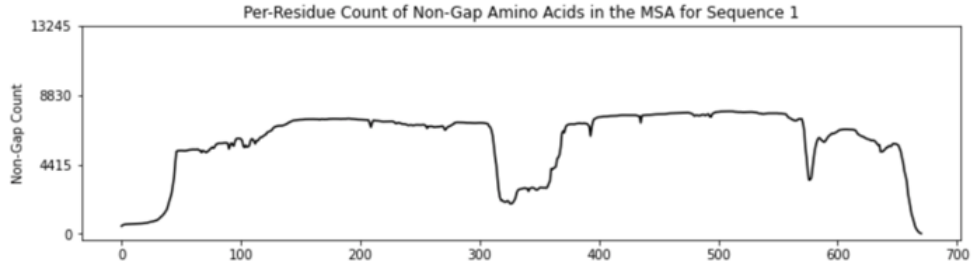


Figure 1. Multiple Sequence Alignment of *F. prausnitzii* GH31 Amino Acid Sequences

This MSA shows the multiple sequence alignments of FpAG1[A] and FpAG2[B] amino acid sequences, generated using AlphaFoldV2.1.0. A visual of the non-gap count of sequences that align with the inputted sequence was presented. For the FpAG1 sequence, there were 13119 unique sequences found and for the FpAG2 sequence, there were 13246 unique sequences found.

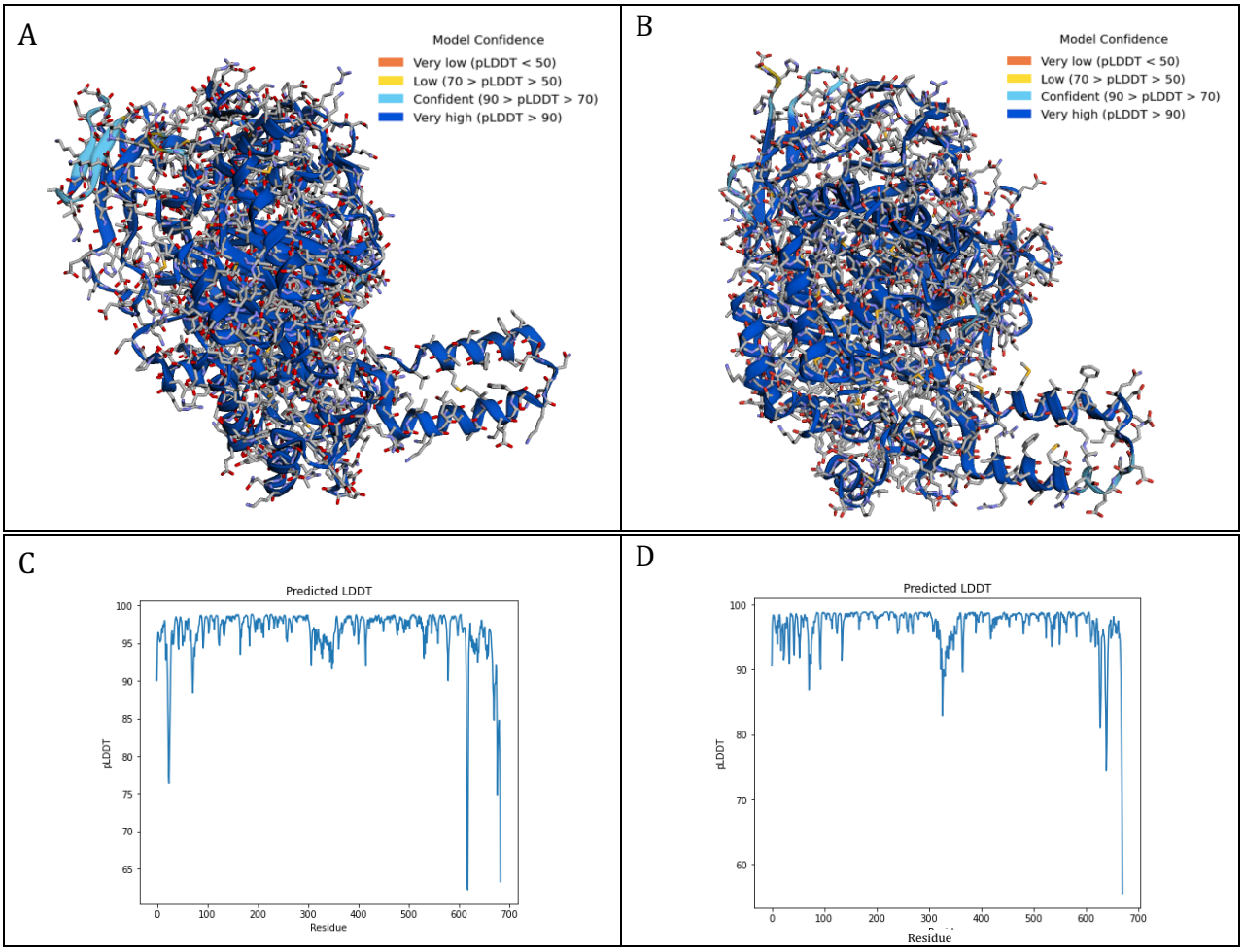


Figure 2. pLDDT of *F. prausnitzii* AlphaFoldV2.1.0 Models

The per residue pLDDT was calculated for both the FpAG1 [A,C] and FpAG2 [B,D] AlphaFoldV2.1.0 structures, where it is presented in A) & B) by colouring the predicted protein structure according to the pLDDT legend, and C) & D) by presenting the data in a visual graph. In FpAG1 [A,C], 96.19% of the residues have a pLDDT score greater than 90, indicating that there is very high confidence in the model. There are residues in the N-terminus (23-28; 72-73) and C-terminus (617, 620, 669-674, 677-683), that fall within the 70 to 90 pLDDT score range, indicating that there is confidence in the structure. At the C-terminus of the FpAG1 model, there are a couple of pLDDT scores between 50 and 70 (618, 619, 684) indicating that there is low confidence in that region. In

FpAG2 [B,D], 96.72% of the residues have a pLDDT score greater than 90, indicating that there is very high confidence. There are residues in the N-terminus (72, 73), alpha helix bundle (327-332; 366), and C-terminus (627-630, 638-642, 668-669), that fall within the 70 to 90 pLDDT score range, indicating that there is confidence in the structure. At the C-terminus of the FpAG2 model, there are 2 pLDDT scores between 50 and 70 (670,671) indicating that there is low confidence in that region.

Computational models of *F. prausnitzii* proteins were created using AlphaFoldV2.1.0 to compare the structures and to identify structural similarities and differences in the features to gain greater insights into these proteins. The computational models were created using AlphaFoldV2.1.0 due to the program's utilization of an in depth MSA and the competitive prediction results compared to experimental results, which was validated by CASP14. The Google Collab version of this algorithm uses a selected portion of the Big Fantastic Database (BFD) and it doesn't use any homologous template structures to create the computational models. As both *F. prausnitzii* proteins have a large number of unique sequences identified in the MSA (13119 for FpAG1 and 13246 for FpAG2), homologous templates would not have been significantly relied upon in the software, and thus not necessary for the structure prediction of these proteins. The Google Collab platform has been validated and for target sequences with larger MSA's, a strong prediction of *F. prausnitzii* α-glucosidase structures is expected. Homologous proteins to the *F. prausnitzii* α-glucosidases, such as Ro-αG1 and MGAM, are oligomers[63,64], and an evolutionary relationship with *F. prausnitzii* α-glucosidases, would suggest the possibility of these proteins forming oligomers. As there is no experimental data to confirm that either FpAG1 or FpAG2 form an oligomer, the monomer of each *F. prausnitzii* α-glucosidase was modelled. Due to this representation, error may be introduced to the models. Future studies are required to confirm the oligomerization of these proteins, but for the purposes of this analysis, monomers will be used to model FpAG1 and FpAG2. Despite the simplified version of AlphaFoldV2.1.0 used to predict these models and the unknown oligomerization of these

F. prausnitzii α-glucosidases, the models predicted are considered to be accurate at representing the structure of the proteins as demonstrated by the accuracy scores obtained.

The pLDDT provides important information regarding the confidence of the predicted structures. In the FpAG1 protein model, there is a pLDDT score greater than 90 in 79.5% of the protein model (between residues 73 and 617), indicating that there is very high confidence in the core of the structure. Conserved residues predicted to be part of the active site have a pLDDT score greater than 90, therefore in an unbound static protein, the position of these residues is predicted to have high confidence. In the FpAG2 protein model, there is a pLDDT score greater than 90 in 37.85% of the protein model (between residues 73 and 327) and 38.90% of the protein model (between residues 366 and 627), which makes up a significant core of the protein model, but there is reduced confidence in the region of Insert 2 in the catalytic domain of the FpAG2 protein model, indicating the difference in confidence between the two protein structures predicted. The placement of these residues in the FpAG2 protein structure and the reduced unique sequences in the MSA may explain the reduced confidence of these residues in the model. Since this Insert 2 α-helical hairpin is predicted to help form an oligomer, as seen in Ro-αG1[64], protein contacts in oligomerization may help stabilize these residues. Residue 366 in FpAG2 also has reduced confidence, but it is present on a loop on the surface of the protein, indicating that the residue may be flexible and hard to place in a static structure.

The significance of the lower confidence of the predicted N-terminal residue 72-73 region in both *F. prausnitzii* predicted structures is that the loop is predicted to be flexible and in very close proximity to the residue expected to be directly involved in the enzymatic reaction (Asp70 in FpAG1 and Asp70 in FpAG2), as demonstrated in the homologous protein, Ro-αG1 (D73)[64]. It is possible that this reduced confidence may be a result of a lack of ligand binding, and the flexibility of the loops results in reduced predicted confidence of this region.

There are residues in the N-terminal and C-terminal regions of the FpAG1 protein model with significantly lower confidence. Reduced and low confidence was also predicted in the C-terminus of the FpAG2 protein model. As these are terminal regions in dynamic protein structures, these regions might just not be stable, resulting in lower structural confidence[70]. The residue region 23-28 is in a loop region on the surface of the protein, which may also contribute to the lower confidence. There is a reduced amount of MSA alignments and pLDDT confidence in the C-terminal of the protein structure, specifically in residues 617-620 and 669-684 of FpAG1 and 668-671 of FpAG2. The C-terminal regions with low confidence are present on the surface of the protein models. The placement of the residues in the structures indicate that the residues may be more flexible and may be harder to place in a static model[70]. Therefore, the majority of residues identified to have low confidence in the pLDDT plot occur in loop regions at the surface of the protein. These regions are expected to be more flexible, which may explain the reduced confidence of the modelled residues.

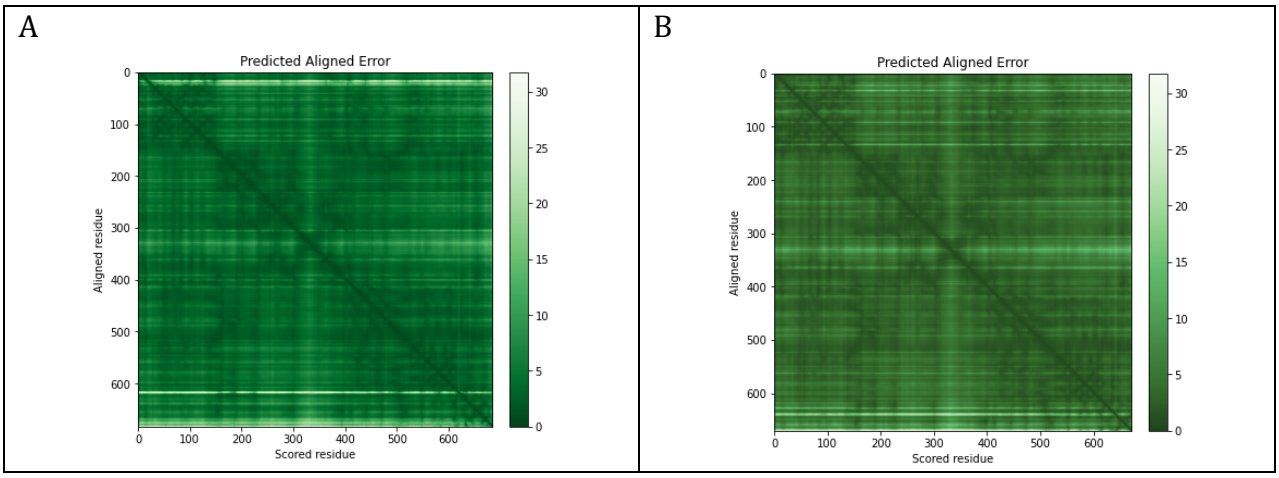


Figure 3. Predicted Aligned Error of *F. prausnitzii* AlphaFoldV2.1.0 Models

The PAE was calculated for both the FpAG1 and FpAG2 AlphaFoldV2.1.0 structures, where it is presented in A & B by colouring based on the expected position error (Å). The darker colour indicates 0 error and lighter colour indicates higher error. In the PAE plot of FpAG1, there is a maximum score of 26.8 Å in the N-terminus prior to aligned residue 50, and in the C-terminus, after

aligned residue 600, there is a maximum score of 29.5 Å. In the PAE plot of FpAG2, there is a maximum score of 29.4 Å in the C-terminus, after aligned residue 600. There is an average PAE of 4.27 Å found in FpAG1 and 4.00 Å found in FpAG2.

The PAE for both FpAG1 and FpAG2 were also analyzed. In FpAG1, there is predicted aligned error present in the N-terminal (maximum 26.8 A) and C-terminal (maximum 29.5 A), regions that correspond to the reduced or low pLDDT scores. In FpAG2, there is predicted aligned error present in the C-terminal (maximum 29.4 A), also regions that correspond to the reduced or low pLDDT score. The proximity to the terminus of the protein structure may increase the flexibility of this region, which could increase the PAE of these regions relative to the rest of the protein. As the average of the PAE 2D plot is low, 4.27 Å in FpAG1 and 4.00 Å in FpAG2, and the colour is generally consistent in both the FpAG1 and FpAG2 models, there is confidence in the relative domain positions in the computational protein structures of *F. prausnitzii* α-glucosidases.

Comparing the MSAs of both FpAG1 and FpAG2, the unique sequences identified have 1% difference in the number of identified alignments. Similarities between the two structures include high amount of unique sequences in the MSA and a reduced number of alignments found in the N-terminal, C-terminal and the predicted Insert 2 α-helical hairpin. These are also the regions that have the lowest confidence in the pLDDT plot, 50 – 70, with lower MSA unique sequences resulting in a lower pLDDT confidence in the protein structures. Differences include different aligned sequences identified in the alpha helix bundle region and the C-terminal of the protein, as the pattern of unique sequences aligned changes. These differences correspond to the differences observed in the pLDDT scores. The high confidence of the pLDDT scores and the low average of PAE indicates that there is high confidence for the computational models predicted by AlphaFold 2.

2.4.3 Protein Model Structure Comparison

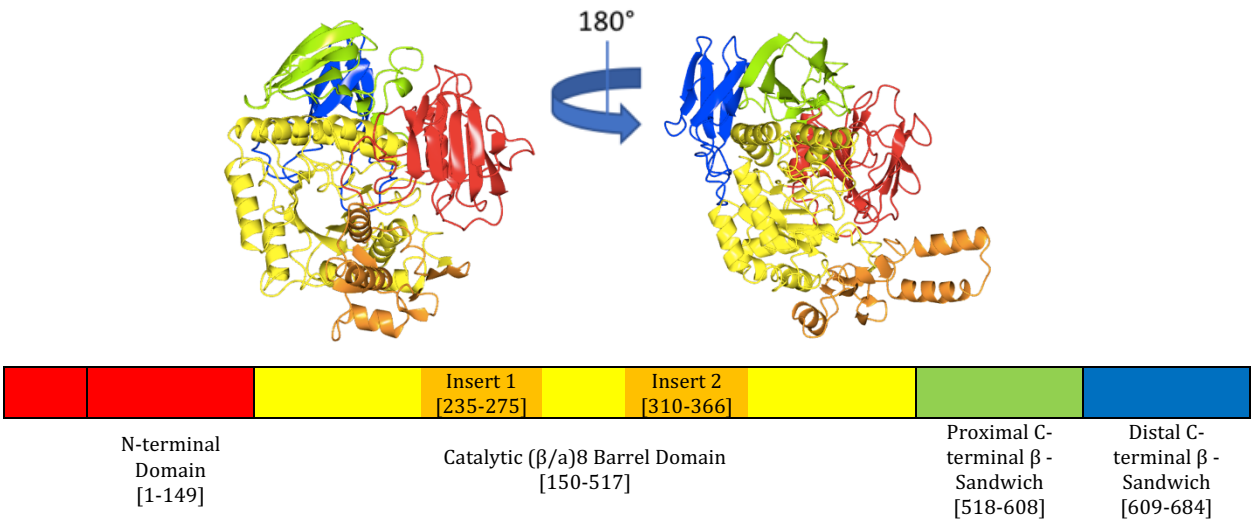


Figure 4. FpAG1 Structural Domains Computational Model

This model shows the location of each domain in FpAG1. N-terminal β sandwich domain (red), $(\beta/\alpha)_8$ barrel (yellow), Insert 1 & Insert 2 (orange), proximal C-terminal β sandwich domain (green) and distal C-terminal β sandwich domain (blue).

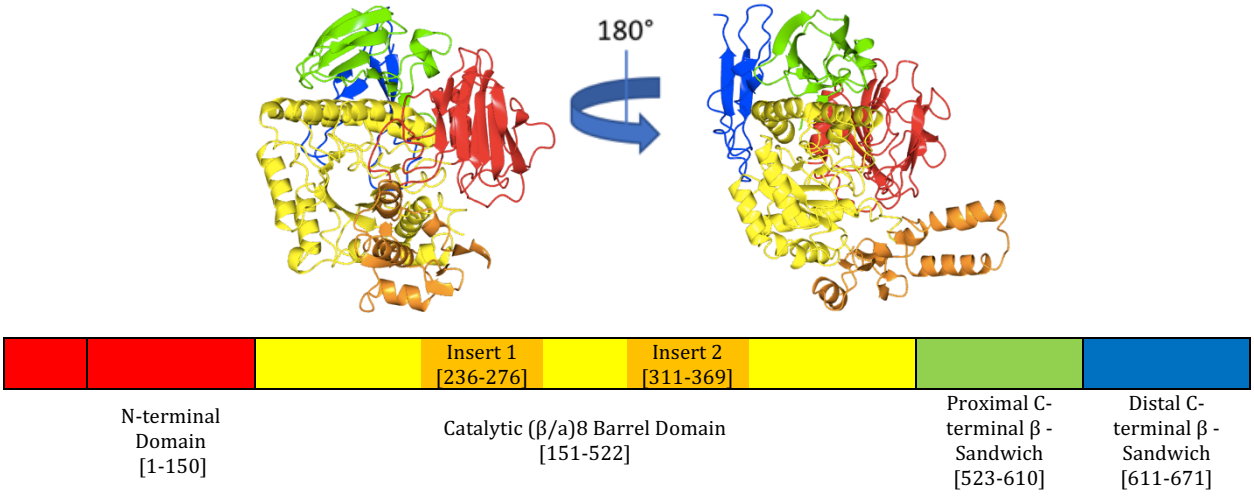


Figure 5. FpAG2 Structural Domains Computational Model

This model shows the location of each domain in FpAG1. N-terminal β sandwich domain (red), $(\beta/\alpha)_8$ barrel (yellow), Insert 1 & Insert 2 (orange), proximal C-terminal β sandwich domain (green) and distal C-terminal β sandwich domain (blue).

The *F. prausnitzii* α-glucosidase sequences and computational protein models predicted by AlphaFoldV2.1.0 have high sequence similarity and structural similarity. Both predicted protein structures contain four domains, including an N-terminal β-sandwich domain, a catalytic $(\beta/\alpha)_8$ barrel domain, a proximal C-terminal β-sandwich domain and a distal C-terminal domain. There are also two inserts within the catalytic domain. In the predicted computational structures, these inserts contain several secondary structure elements and Insert 1 forms an anti-parallel β-sheet, while Insert 2 forms an α-helical hairpin followed by a β-hairpin. There are also two C-terminal β-sandwich domains found in the *F. prausnitzii* α-glucosidase proteins. The proximal domains are very similar between these proteins, but there are structural differences in the distal C-terminal domain, with the FpAG2 distal domain being significantly smaller than the FpAG1 distal C-terminal domain. There are many structural similarities to the GH31 α-glucosidase from *R. obeum* (Ro-αG1, PDB ID 3N04), which was the first bacterial protein solved of subgroup 1 of the GH31 family, as well as the thermophilic archaeon *Sulfolobus solfataricus* α-glucosidase MalA (PDB ID 2G3M) and the N-terminus of human maltase-glucoamylase (ntMGAM, PDB ID 2QLY).

Table 3. RaptorX DeepAlign Comparison of Fragments of *F. prausnitzii* Protein Models

Conserved Structural Domains	Structure Fragments	Length Aligned	Residue Range	Sequence Identity/ Similarity (%)	RMSD (Å)	uGDT/GDT	TM-score
	Whole Structure	665	1-684 1-671	50.9 66.9	1.09	630/94	0.9790
	N-terminal	37	1-37 1-38	37.8 56.8	1.50	33(90)	0.736
N-terminal Domain		112	37-149 38-150	44.6 61.6	0.73	109(97)	0.964
Catalytic Domain		369	149-517 150-522	55.8 70.1	0.74	359(97)	0.990
	N-terminal β Barrel	86	149-234 150-235	61.2 77.6	0.43	86(100)	0.984
	Insert 1	41	235-275 236-276	51.3 69.2	0.36	41(100)	0.968
	Insert 2	57	310-366 311-369	32.7 46.2	1.14	53(93)	0.876
	C-terminal β Barrel	151	367-517 370-522	61.8 75.7	0.53	149(99)	0.988
	Proximal C-terminal β - Sandwich	88	518-608 523-610	51.7 68.5	1.12	84(95)	0.935
	Distal C-terminal β - Sandwich	59	609-684 611-671	41.1 64.3	1.38	53(87)	0.825

The whole structure and fragments of the *F. prausnitzii* protein models were aligned and structural features were compared with the RaptorX DeepAlign program to determine similarity. The conserved N-terminal and catalytic domains both scored a GDT of 97. For FpAG1[FpAG2], the N-terminal sequence (residues 1-37[38]), Insert 2 (residues 310[311]-366[369]) and the distal C-terminal β-Sandwich domains (residues 609[611]-684[671]) have the lowest calculated GDT and TM-scores, indicating the lowest structural similarity between the structures. A Local Pairwise Sequence Alignment was performed using EMBOSS Matcher, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK, https://www.ebi.ac.uk/Tools/psa/emboss_matcher/.[53]

RaptorX DeepAlign aligns structures based on geometric similarity, evolutionary information and hydrogen-bonding similarity[56]. In general, two structures with 50% sequence identity will have an RMSD of ~ 1Å when superimposed[71]. The RMSD of the individual structural domains of these protein models indicates which structural features were introducing bias into the overall structure RMSD. The RMSD is not necessarily the best representative of similarity because it has a lot of introduced error associated with it, specifically it is largely affected by local changes to a domain in a structure, which may not be representative of the superposition of the whole structure[72]. When the structure superimposition is broken down into multiple domains, there are differences in the RMSD calculated, indicating that there are certain folds with greater similarity than others.

The GDT and TM-score were determined to reduce the error introduced by singular deviations in the structure[56]. Both values are normalized to the sequence length, and are therefore more comparable between domains than the RMSD value[56]. The GDT of the whole structure superimposition is 94, indicating that there is high structural similarity of the protein backbone. There was significant structural similarity determined in the N-terminal domain and the $(\beta/\alpha)_8$ barrel catalytic domain, as GDT of 97 were calculated for both domains. The active site is predicted to be present in the $(\beta/\alpha)_8$ barrel domain and the high structural similarity of the protein backbone of this $(\beta/\alpha)_8$ barrel indicates confidence in the similarity of the conserved catalytic domain identified in these *F. prausnitzii* α-glucosidases, even though the sequence identity of this domain is 55.9%. Based on the computational models predicted by AlphaFoldV2.1.0 and the structural comparison done using RaptorX DeepAlign, and there is significant structural similarity between the conserved domains in the *F. prausnitzii* protein structures.

There are structural variations in the N-terminal, distal C-terminal domain and Insert 2 of the $(\beta/\alpha)_8$ barrel domain regions of the *F. prausnitzii* α-glucosidases, with a GDT score less than 95 and a TM-score less than 0.9. The $(\beta/\alpha)_8$ barrel is a common domain found in enzymes, but the inserts present within this domain in *F. prausnitzii* α-glucosidases are not conserved[73]. The structural

similarity of the inserts in the catalytic domain were analyzed in RaptorX DeepAlign. Reduced GDT values of Insert 2 shows the deviation of the protein backbone in that region between these two structures, most significantly in the α-helical hairpin. In this catalytic domain, Insert 2 causes the most significant structural deviation in the $(\beta/\alpha)_8$ barrel domain. Structural deviations were also identified in the N-terminal structure prior to the N-terminal conserved β-sandwich domain and the distal C-terminal β -Sandwich. These structural deviations could contribute to differences in molecular dynamics, protein stability or protein-protein contacts. All of the domains aside from the distal C-terminal domain, are predicted to make protein contacts with the adjacent monomer in the homodimer in Ro-αG1[64]. Structural differences in these regions may contribute to the inter monomeric distance which could affect substrate accessibility to the active sites or molecular dynamics of the protein, which may contribute to the enzymatic activity.

2.5 Chapter Conclusions

AlphaFoldV2.1.0 was used to predict protein structure models with high confidence of *F. prausnitzii* α-glucosidases. The same conserved domains, N-terminal domain (GH31_N (N-terminal domain of glycosyl hydrolase family 31 (GH31))) and $(\beta/\alpha)_8$ barrel domain (GH31_glucosidase_II_MalA domain (Alphα-glucosidase II-like)) were identified in both of the *F. prausnitzii* α-glucosidases, but FpAG1 had an additional domain, DUF5110, identified at the C-terminus. The *F. prausnitzii* α-glucosidases had the same conserved residue sites indicated and the majority of the conserved residues are identical, aside from the H419/N422 residue in close proximity to the predicted proton donor catalytic residue. Despite the structural differences between these GH31 α-glucosidases identified in Insert 2 of the $(\beta/\alpha)_8$ barrel fold catalytic domain and the C-terminal distal domain, these proteins have high structural similarity in the $(\beta/\alpha)_8$ barrel fold catalytic domain and the N-terminal β-sandwich domain (with a GDT of 97). There were no structural differences identified in the protein structures modelled by AlphaFoldV2.1.0, which would suggest functional differences between the FpAG1 and FpAG2 enzymes.

Chapter 3: Construct Design, Protein Expression and Purification

3.1 Introduction

3.1.1 Construct Design

A vector is required to express a protein from a DNA sequence in a recombinant cell system. There are many considerations when designing an expression construct, which include the promotor, terminator and protein machinery for the induction of protein expression. The pET-24b(+) - Novagen expression vector (Sigma Aldrich 69750) is used in bacterial expression systems, and includes a T7 promotor, T7 terminator and LacI gene for isopropyl β- D-1-thiogalactopyranoside (IPTG) protein induction. The DNA sequence chosen to express with this construct is important, as codon bias can affect the expression of recombinant proteins. For messenger ribonucleic acid (mRNA) translation, transfer ribonucleic acids (tRNA) with an anticodon complementary to the mRNA codon, bind a specific amino acid. This amino acid is translated in the position of the corresponding codon. There are a total of 61 tRNAs that could be found in translation, as there are multiple codons possible for most amino acids[74]. Not all organisms express the tRNAs or express the tRNAs at the same levels. This leads to a codon usage bias, as some mRNA sequences from different organisms cannot be properly or efficiently translated[74]. Therefore, optimizing the codons of the sequence to reflect the codon usage of the expression system will increase expression potential of the recombinant protein.

3.1.2 Protein Expression and Purification

Escherichia coli BL21(DE3) (NEB) cells are *E. coli* cells that that do not express proteases Lon or OmpT and, using the T7 promoter, allow controlled expression of target genes. This expression system contains a prophage DE3, which expresses the T7 RNA polymerase under the control of the lacUV5 promoter. Expression of genes controlled by the lacUV5 promoter can be induced by either IPTG or lactose[75]. In IPTG induction, IPTG takes on the role of allolactose, which removes a repressor

from the lac operon, and initiates the expression of the T7 RNA polymerase, thus inducing the expression of the recombinant protein under control of the T7 promoter[75].

Immobilized Metal Affinity Chromatography (IMAC) is a purification method for recombinant proteins with an affinity tag[76]. The affinity tag has affinity to transition metal ions such as nickel or cobalt, because of the electron donor groups of the histidine residues[76]. As the proteins from cell lysate pass through the gravity column, proteins with affinity to the metal coordinated resin with be detained. The addition of an imidazole elution buffer will release the recombinant protein from the metal-coordinated resin because higher concentrations of imidazole outcompete the histidine affinity tag[76]. By increasing the imidazole concentration, proteins with different binding interactions could be eluted with the column. By eluting any contaminating proteins with lower imidazole concentrations, the target protein could be eluted as pure as possible[76].

3.2 Chapter Objectives

The objective of this chapter is to design a construct for both FpAG1 and FpAG2 recombinant proteins for overexpression in BL21(DE3) cells. FpAG1 and FpAG2 will be purified through IMAC and concentrated using centrifugal filtration for further use in downstream applications.

3.3 Material and Methods

3.3.1 Construct Design

The sequences were retrieved from CAZy Database – GH31 enzymes from *Faecalibacterium prausnitzii* A-165 (Genbank Entry: QIA42631.1 [FpAG1] and Genbank Entry: QIA42631.1 [FpAG2]). Each amino acid sequence was inputted into Genscript™ Codon Optimization software, *Escherichia coli* host organism selected, to determine an optimal DNA sequence for protein expression. The amino acid sequence was also inputted into Expasy Peptide Cutter[77] to predict which proteases are least likely to cleave the protein structure. The peptide cleavage sequence for the determined

protease was included at the C-terminal of the sequence. This sequence was inputted between ribosome binding site and the 6-His-tag of the pET-24b vector. This insert was synthesized and cloned into pET-24b by Genscript, and transformed into BL21(DE3) cells for protein expression.

3.3.2 Protein Expression and Purification of FpAG1 and FpAG2

A colony of BL21(DE3) transformed with pET-24b (FpAG1) or pET-24b (FpAG2) were grown overnight in Luria-Bertani (LB) Media with 50 µg/ml kanamycin [78] at 37°C, for 16 hours and 200 rpm. Overnight culture (1 mL) was added to 1 L of LB media and grown at 37 °C and 200 rpm. At an OD_{600} of 1.2, IPTG was added to a final concentration of 0.05 mM and the culture was incubated at 20 °C, for 16 hours and 200 rpm. The cells were pelleted and the wet weight of the BL21(DE3) cells was measured in 50mL conical centrifuge tubes prior to freezing at -80 °C.

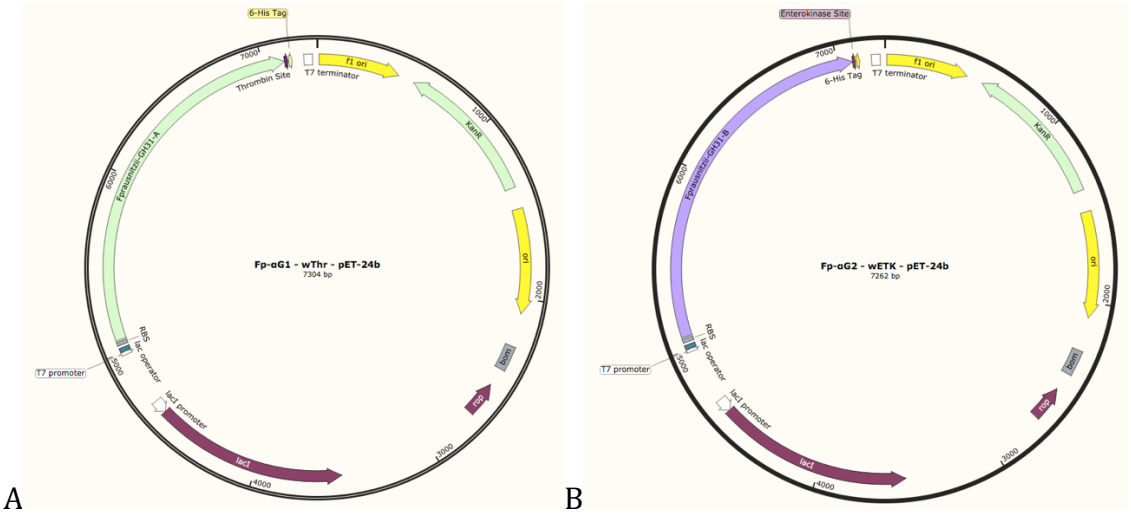
Protein Purification Protocol

1. Approximately 4 g of cells were thawed from -80 °C and resuspended in 40 mL of lysis buffer
2. The cells were passed through the French Press twice at a 1000 psi (metal chamber was refrigerated at 4 °C overnight) in a lysis buffer (50 mM sodium phosphate, 500 mM sodium chloride, 5% Glycerol; pH 8.0)
3. Cell lysate was centrifuged at 7000 x g for 30 minutes at 4 °C
4. Supernatant was centrifuged at 17000 x g at 4°C for 30 minutes
5. Supernatant was diluted with equilibration buffer to a total volume of 200 mL (50 mM sodium phosphate, 300 mM sodium chloride; pH 8.0)
6. The resin was equilibrated as per the manufacturer protocol
7. The lysate was run through the gravity column with 5 mL of HisPur™ Cobalt Resin (Thermo Scientific ™ 0089964)
8. Approximately 500 mL of equilibration buffer was used to wash the buffer, until no protein was detected in the eluate with the Bradford Reagent (Bio-Rad Coomassie Blue G-250).

9. Wash buffer (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole; pH 8.0) was used to wash away any contaminating or weakly bound proteins.
10. A range of elution buffers (50 mM sodium phosphate, 300 mM sodium chloride, [10 – 300] mM imidazole; pH 8.0) were used and 100 mM Imidazole buffer was used to elute the majority of the target protein from the nickel resin (approximate volume 200 mL).
11. Amicon® Ultra-15 Centrifugal Filter Units (30K) (EMD Millipore UFC903008) were used to concentrate the proteins at 3000 x g at 4 °C.
12. The proteins were diluted in the storage buffer: 25mM MES, 100 mM NaCl, pH 6.5.

3.4 Results and Discussion

3.4.1 Construct Design



A B

Figure 6. pET-24b (FpAG1) and pET-24b (FpAG2) DNA Constructs

Designed constructs of FpAG1 (A) and FpAG2(B) target sequences in pET-24b(+) vectors with C-terminal protease cleavage site and 6-Histidine tag.

The pET-24b(+) - Novagen expression vector was chosen because it is a high copy vector, a C-terminal His-tag and an option for both IPTG and autoinduction protein expression. The promoter used in the vector affects the method of induction in protein expression. The T7 promoter in the pET vector gives the option of IPTG induction and autoinduction[79,80]. Although these proteins were originally found in *F. prausnitzii,* the recombinant protein was to be expressed in *E. coli*, so the codon bias was adjusted to make sure that the amino acid sequence was optimally expressed in the expression system. Although codon optimization is a way to increase protein yield, changing codons has also been found to affect the translation rate which may be important for protein folding[81]. Codon optimization has also been seen to affect protein folding, stability, and function[74]. The C-terminal His-tag position was chosen because this was not seen to have an effect on α-glucosidase activity in other GH31 proteins expressed in the lab. This His-tag was required for purification by immobilized metal affinity chromatography[76], which could purify the protein to about 95%. In the case that this His-tag would affect either enzymatic activity or protein crystallization in X-ray crystallography, a protease cleavage site was included to ensure that the His-tag could be removed. This was chosen by using a Expasy Peptide Cutter to identify amino acid motifs that would increase the probability of protease cleavage. This resulted in different proteases being chosen for each target to ensure that the protease will remove the affinity tag without affecting the structure of the protein. In FpAG1, a thrombin site was chosen and in FpAG2, an enterokinase site was chosen as thrombin was predicted to cleave the protein at residue 342.

3.4.2 Protein Expression and Purification of FpAG1 and FpAG2

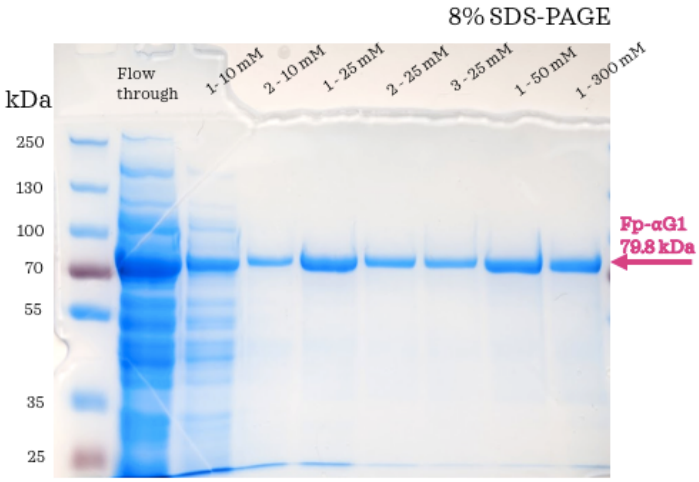


Figure 7. 8% SDS-PAGE gel of FpAG1 Purification

FpAG1 was purified using the purification protocol indicated in section 3.3.2. Proteins were eluted with 10, 25, 50 and 300 mM imidazole concentrations. PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa, was run in lane 1.

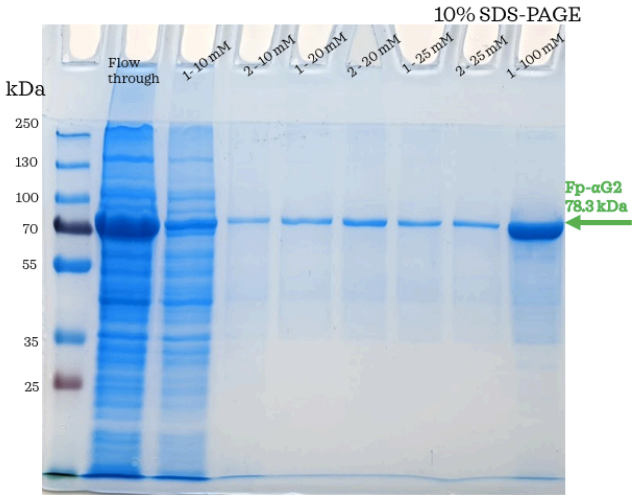


Figure 8. 10% SDS-PAGE gel of FpAG2 Purification

FpAG2 was purified using the purification protocol indicated in section 3.3.2. Proteins were eluted at 10, 25, 50 and 100 mM imidazole concentrations. PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa, was run in lane 1.

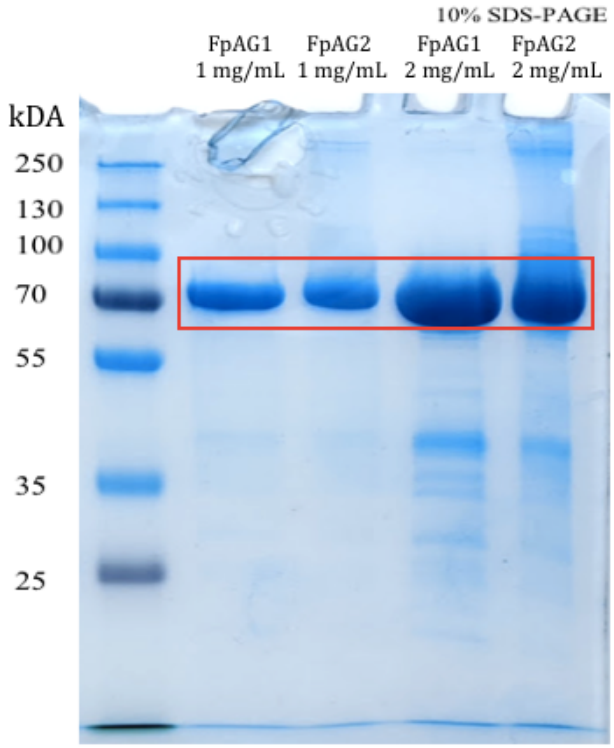


Figure 9. Concentrated Purified FpAG1 and FpAG2 Proteins

FpAG1 and FpAG2 concentrated using the protocol in section 3.2.2, and stored in final buffer 25 mM MES, 100 mM NaCl, pH 6.5. The protein concentration was determined using the NanoDrop 2000E (Extinction coefficient 15.6 for FpAG1 and 15.87 for FpAG2) and 1 mg/mL of protein was run in lanes 2 and 3. A concentration of 2 mg/mL of protein was run in lanes 4 and 5. The target protein is indicated with the red outline. PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa, was run in lane 1. The average FpAG1 concentration purified was 6.4 +/- 0.3 mg FpAG1/gram of BL21(DE3) cells. The average FpAG2 concentration purified was 4.7 +/- 0.5 mg FpAG2/g BL21(DE3) cells.

Cobalt resin was chosen for the IMAC purification because there were fewer contaminating proteins in the purification compared to the Nickel resin. A pH of 8.0 was used in this protein purification, because higher a pH will result in a greater bond with the transition metal ion[76]. The storage buffer 25 mM MES, 100 mM NaCl, pH 6.5, was used because this was the buffer used in a the purification of a similar bacterial α-glucosidase of the GH31 family, from *R. obeum*[64] .

In this purification, additional bands are found at lower molecular weights and at higher molecular weights. The majority of the protein is found at the expected molecular weight of 79.8 kDA and 78.3 kDA for FpAG1 and FpAG2, respectively. Other purification methods were not successful in removing the additional protein bands in the SDS-PAGE gels. Other optimization techniques attempted include changing chelating resin, changing cell lysis method, attempting size exclusion chromatography, changing pH of the storage buffer, including protease inhibitors, changing salt concentration, and attempting ammonium sulfate precipitation. These techniques did not improve the purity of the protein solution beyond the purity obtained through the indicated protocol. These additional bands are predicted to be contaminated by pH and/or protease induced protein aggregates. As the purification protocol was optimized to the point that the proteins maintained constant enzymatic activity for over two weeks, further optimization to the protocol was ceased. Future directions include further purification optimization of these proteins, to obtain a protein that is pure and stable enough for techniques such as X-ray crystallization structure determination.

Throughout the protein purification optimization process, FpAG2 was consistently more difficult to purify. As seen in Figure 9, specifically in lane 4 and 5, where 20 μl of 2 mg/mL of FpAG1 and FpAG2 protein was loaded, there are more bands at greater molecular weights than the expected molecular weight in the FpAG2 purification, compared to the FpAG1 purification. These bands could indicate protein aggregation. When testing for enzymatic activity (more information on the enzymatic assay methods used can be found in Chapter 3), FpAG1 was generally more stable and was able to maintain enzymatic activity for over 2 weeks. FpAG2 required a much shorter duration at a

pH of 8.0 during purification to be able to maintain enzymatic activity for longer than 3 days. By decreasing the duration FpAG2 spent at a pH of 8.0 during purification, the enzymatic activity was maintained for over 2 weeks, when stored at a temperature of 4 °C. When analyzing the computational structures in Chapter 1, FpAG1 had a conserved domain at the C-terminal of the protein, that was not found in FpAG2. This was the DUF5110 that is predicted to be a carbohydrate-binding domain found at the C-terminal of GH31 proteins. This domain is the greatest detected structural difference between FpAG1 and FpAG2 and further studies to investigate whether DUF5110 has a role in protein stability and activity can illuminate the role of this domain.

3.5 Chapter Conclusions

A construct was designed and synthesized for both *F. prausnitzii* proteins. It included promoters for inducible expression by the lac operon, a C-terminal his-tag for purification by IMAC, and a protease cleavage site specific to the target sequence to account for any issues with downstream applications. Significant amounts of protein (FpAG1: 6.4 +/- 0.3 mg FpAG1/gram of BL21(DE3) cell; FpAG2: 4.7 +/- 0.5 mg FpAG2/g BL21(DE3) cells) have been expressed with BL21(DE3) cells and purified through IMAC and centrifugal filtration. Although the purification included a few contaminating bands, the majority of the protein was present in the expected sizes of 79.9 & 78.3 kDa. This protocol was successful in obtaining FpAG1 and FpAG2 protein for further biochemical studies.

Chapter 4: Substrate Specificity and Enzymatic Activity Kinetics

4.1 Introduction

4.1.1 Determining Optimum Conditions for Enzymatic Activity

FpAG1 and FpAG2 are α-glucosidases part of the GH31 family, as predicted by CAZy, and the predicted mechanism in this family, the overall retaining mechanism, is expected to occur with aspartic acid residues, that act as acid/base catalysts. The pH of the protein environment affects the protonation of the aspartic acids, because as the pH decreases, the concentration of H^+ in the buffer increases. The acid dissociation constant (pK_a) of these catalytic residues allows them to donate and accept protons within a specific pH range. A pH profile can represent the likelihood that the catalytic acid is protonated and the catalytic base is deprotonated at a certain pH, resulting in an effective enzymatic reaction. The pH at which enzymes exhibit the greatest amount of enzymatic activity is the optimum pH. Aside from the effects on the catalytic residues in the active site, the pH can also affect the stability of the protein, resulting in a risk of aggregation and irreversibly affecting the catalytic activity[82,83]. The isoelectric point of a protein (pI) is the pH at which the protein doesn't have a net electrical charge. When the pH of the system approaches the pI, the protein solubility may be affected. Reduced protein solubility may lead to reduced enzymatic activity, as protein folding could be influenced[84]. By testing enzymatic activity with a pH range, the glucose released can be measured to determine the best pH for enzymatic activity of the *F. prausnitzii* α-glucosidases.

Temperature can also have a significant effect on the enzymatic rate. When the temperature of a reaction increases, the rate of the enzyme-catalyzed reaction increases as the kinetic energy of molecules in the solution is increased[82]. If the temperature is increased too much, the hydrogen bond and ionic bond network in the tertiary structure will be disrupted, resulting in irreversible denaturation of the protein[85]. This can result in aggregation, improper formation of disulfide bonds or covalent alterations and the protein will be unable to undergo any catalytic reaction[85]. In extreme cases of low temperature, cold denaturation can occur, leading to protein denaturation[86]. It is caused

by the disruption of the interaction between non-polar groups in the protein and water, which is dependent on temperature[86]. The temperature at which cold denaturation of many proteins occurs at is below 0°C[87]. Testing the amount of glucose released at a variety of temperatures can determine the best temperature range for enzymatic activity of the *F. prausnitzii* α-glucosidases.

4.1.2 Substrate Specificity

According to CAZy, the GH31 family includes enzymes that have α-glucosidase, α-galactosidase, α-mannosidase, α-1,3-glucosidase, α-xylosidase, α-glucan lyase, sucrose α-glucosidase, oligo-1,6-glucosidase, isomaltosyltransferase, oligosaccharide α-1,4-glucosyltransferase, α-N-acetylgalactosaminidase, sulfoquinovosidase and α-6-glucosyltransferase activity[88]. Investigating the substrate specificity of GH31 enzymes by measuring whether substrate hydrolysis occurs can illuminate specific activities of these proteins and insight into the substrate specificity of these proteins can provide information on structural aspects of these proteins.

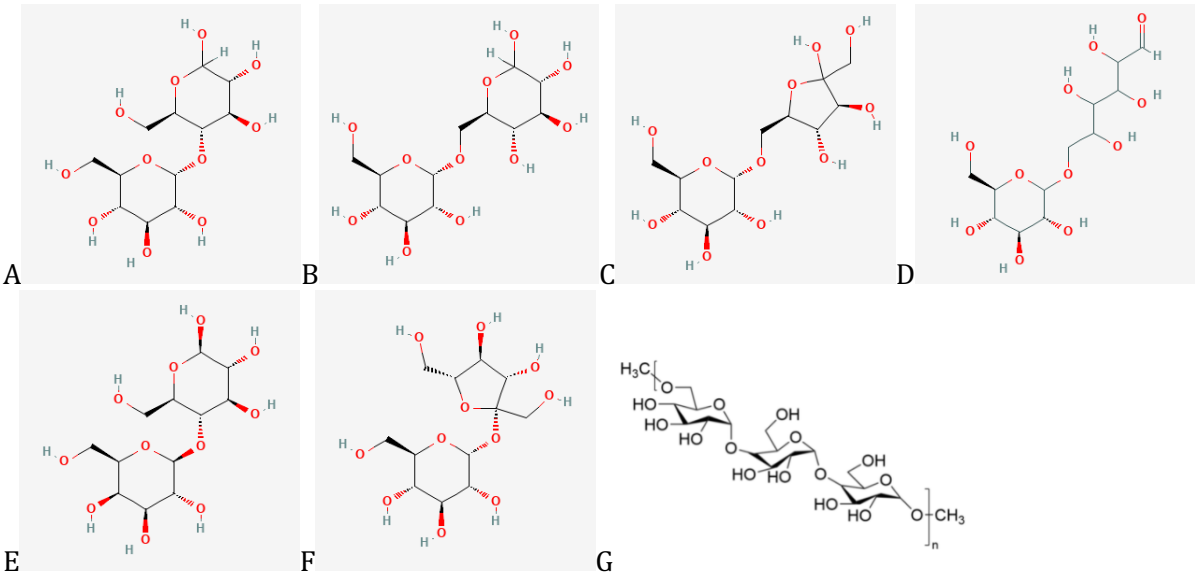


Figure 10. Carbohydrate Structures

These are structures of the discussed in this chapter. A) maltose B) isomaltose C) palatinose D) melibiose E) lactose F) sucrose G) pullulan. These structures were obtained from the National Center for Biotechnology Information, . PubChem Compound Summary for CID 41774. *Precose.*

(2022). Available at: https://pubchem.ncbi.nlm.nih.gov/compound/Precose. (Accessed: 20th February 2022). Pullulan Structure was obtained from Toronto Research Chemicals, https://www.trc-canada.com/product-detail/?P840250.

Substrates with α-1,4, α-1,6 and α-1,2 glycosidic bonds can be hydrolyzed by enzymes that have α-glucosidase activity. Maltose and longer oligosaccharides such as maltotetraose, maltohexaose and maltoheptaose, are different lengths of glucose units linked with α-1,4 glycosidic bonds. These substrates are naturally found in malted grains, peaches and many other common food sources[89]. The malto-oligosaccharides are also the product of starch after treatment with α-amylase[90]. GH31 enzymes with α-glucosidase activity [EC 3.2.1.20] have been observed to hydrolyze these substrates with α-1,4 glycosidic linkages. Isomaltose is a disaccharide containing two glucose units and palatinose is a glucose-fructose disaccharide. Both of these substrates have a α-1,6 glycosidic bond. Isomaltose is produced during the caramelization process of glucose and can be present in a isomaltooligosaccharide mixture, which is added to many processed fruits, canned goods and confections[91,92]. Isomaltose is also generated from starch by amylases[93]. Palatinose is naturally found in honey and sugar cane, and is considered a slow-release glucose-fructose disaccharide because it is broken down more slowly by digestive enzymes[94]. GH31 enzymes with oligo-1,6-glucosidase activity [EC 3.2.1.10] have been observed to hydrolyze these substrates with α-1,6 glycosidic linkages. Sucrose is also a glucose-fructose disaccharide, but it has α-1,2 glycosidic bond[89]. This glucose-fructose disaccharide can be found in many common foods, including fruits, vegetables and table sugar[89]. This sucrose α-glucosidase activity [EC 3.2.1.48] has also been seen in enzymes part of the GH31 family[95]. The ability of the enzyme to hydrolyze specific α-glycosidic bonds can provide insight on substrate specificity and structural aspects of these proteins.

Melibiose, lactose and pullulan are substrates that have structural similarities to the substrates known to be hydrolyzed by enzymes with α-glucosidase activity. Melibiose and lactose

both have galactose at the non-reducing end and a glucose unit at the reducing end. These substrates are broken down by enzymes with α-galactosidase activity [EC 3.2.1.22], which is seen in the GH31 family and β-galactosidase activity [EC 3.2.1.108], which is not seen in this family[88,96]. The difference between these substrates lies in the linkage, where melibiose has a α-1,6 glycosidic linkage and lactose have a β-1,4 glycosidic linkage[89,97]. Lactose is naturally found in mammalian milk[89]. Pullulan is a polysaccharide that is made up of maltotriose units connected by α-1,6 glycosidic bonds[98]. This polysaccharide has many uses in the food industry, including a low-calorie sweetener, replacement for starch in baked goods and a binder in food products[99]. This oligosaccharide has the same glycosidic bonds in other substrates, α-1,4 and α-1,6, but organized differently[98]. It is hydrolyzed by enzymes that have pullulanase (endo-1,6-αglucosidase activity) activity [EC 3.2.1.41], seen in the GH13 family. The ability of the enzyme to hydrolyze β-1,4 glycosidic bonds and α-glycosidic linkages with different carbohydrate units can provide insight on substrate specificity and structural aspects of these proteins.

4.1.3 Enzymatic Activity Kinetics

The substrate preference and reaction efficiency of the enzyme could be analyzed using enzymatic kinetics parameters determined from the Michaelis-Menten equation. The initial reaction rate is the reaction velocity where product generation is linear with time prior to the reaction plateau that occurs due to lack of substrate. The initial enzymatic reaction rate of a range of substrate concentrations is used to create a Michaels-Menten curve to determine kinetic parameters such as V_{max}, K_M and k_{cat}, which are dependent on temperature and pH[82,100,101]. The K_M is a constant which represents the apparent substrate concentration at which half of the maximum velocity is reached and it is used to estimate the substrate's binding affinity. The maximum velocity of the system where the enzyme system is completely saturated with substrate is the V_{max}, and this parameter is dependent on the enzyme concentration[82]. The turnover rate constant is known as k_{cat} and it describes the amount of substrate hydrolyzed per unit of time when each enzymatic active site is

fully saturated[82]. As the substrate concentration increases, the reaction velocity is expected to increase. When the system is saturated with substrate, the maximum reaction velocity of the system is reached above which the change in velocity is expected to be very low[100,101]. Using the k_{cat} and K_M determined from the Michaelis-Menten curves of different substrates, the substrate affinity and reaction efficiency can be used to analyze the enzymatic preference for substrates[100–102].

There are cases of Michaelis-Menten kinetics where the reaction does not reach steady-state. Instead, the reaction rate starts to decrease at high substrate concentrations. This trend fits the substrate inhibition derivation of the Michaelis-Menten model[103]. This derivation introduces a K_i constant to the Michaelis-Menten equation, which is the concentration of inhibitor that will cause the reaction to occur at half of the maximum velocity, after the reaction has reached its peak[103]. In substrate inhibition systems, allosteric inhibition is a mechanism that is predicted to occur, where the binding of a substrate in an additional site will result in a catalytically reduced or inactive enzyme[103]. Due to this regression in activity, the substrate at high concentrations is acting as an inhibitor[103]. Substrate inhibition has been seen in nature to maintain homeostasis, and has been seen as a critical regulatory mechanism in enzymes, such as phosphofructokinase where adenosine 5'-triphosphate inhibits activity at physiological concentrations[104].

A. $v = \frac{Vmax \times [S]}{(Km + [S])}$ B. $v = \frac{Vmax \times [S]}{[Km + ([S] \times \left(1 + \frac{[S]}{\mathbf{Ki}}\right))]}$

Figure 11. Michaelis Menten and Substrate Inhibition Equations

Equation A is the Michaelis-Menten equation, and Equation B is the substrate inhibition derivative of the Michaelis-Menten equation.

4.2 Chapter Objectives

The objectives of this chapter are to illustrate and distinguish the enzymatic activities and substrate specificities of FpAG1 and FpAG2. In this chapter, the pH and temperature will be investigated to determine the best conditions for enzymatic activity of FpAG1 and FpAG2. The substrate specificity of these *F. prausnitzii* GH31 enzymes will be characterized, by testing a variety of substrates, with the suspected possibility of hydrolysis, to determine the hydrolytic ability of these *F. prausnitzii* enzymes. The glycosidic bonds to be tested are α-1,4 (maltose, maltotetraose, maltohexaose, maltoheptaose), α-1,6 (isomaltose, palatinose and melibiose), α-1,2 (sucrose), and β-1,4 (lactose). The kinetic parameters of the hydrolysis of maltose, isomaltose and palatinose will be determined and the enzymatic activity between FpAG1 and FpAG2 will be compared.

4.3 Material and Methods

4.3.1 pH Profile Protocol

The effect of pH on the enzymatic activity of FpAG1 and FpAG2 were tested by diluting substrates and enzymes in a variety of buffers composed of 25mM 2-(N-morpholino)ethanesulfonic acid (MES), Piperazine-N,N′-bis(2-ethanesulfonic acid) (PIPES) or 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 100mM NaCl and a pH range of 5.25 to 8.0. The pH of these buffers was determined at a temperature of 25 °C. Enzymes were diluted ~300x in the final buffer, so any traces of initial buffer were insignificant and not expected to interfere with the assay. The diluted substrate and enzyme in buffers ranging a pH of 5.25 to 8.0, were pre-incubated at the appropriate temperature for 20 minutes prior to combining the volumes in a 1.5 mL microcentrifuge tube. The final concentration of enzymes in the assay was 0.05 mg/mL and the final maltose concentration in assay was 60 mM. The reaction was carried out at 14 °C, 25 °C, or 37 °C for 25 minutes and inactivated at 85 °C for 3 minutes. Each condition for each enzyme was done in

triplicate. The amount of glucose released was determined using the protocol for the modified version of Sigma-Aldrich Glucose (GO) Assay Kit (GAGO) assay as described in Appendix B.

4.3.2 Substrate Specificity Protocol

Maltose (Sigma Aldrich M9171), isomaltose (TRC- I821250), palatinose (Sigma Aldrich P2007), pullulan (Sigma Aldrich P4516), lactose (Fisher Scientific L6), sucrose (Sigma Aldrich S7903), melibiose (TRC- M215503), maltotetraose (TRC-M161520), maltohexaose (TRC-M160100) and maltoheptaose (TRC-M160055) were solubilized resuspended in 25 mM MES, 100 mM NaCl, pH 5.8 buffer. The amount of glucose released by 0.05 mg/mL of FpAG1 and FpAG2 in the presence of substrate (60 mM final concentration) after 30 minutes at 37 °C was measured. The solutions were pre-incubated and the reaction was inactivated at 85 °C for 3 minutes. Each reaction was done in triplicate. The amount of glucose released was determined using the protocol for the modified version of Megazyme Glucose oxidase/ peroxidase Assay Kit (GOPOD) assay as described in Appendix C. The activity assay with malto-oligosacharides was additionally conducted at a pH of 6.4, with a final substrate concentration of 25mM, to compare the amount of glucose released in the presence of maltotetraose to other malto-oligosaccharides (maltose, maltohexaose, and maltoheptaose).

4.3.3 Determination of Enzyme Kinetics Parameters Protocol

The enzyme solution and a range of substrate concentrations were diluted in 25 mM MES, 100mM NaCl, pH 5.8 buffer. The concentrations and volumes used in the activity assay reaction can be found in Table 4. An aliquot of substrate and enzyme was pre-incubated in a 37 °C hot plate for 20 minutes prior to being combined. Multiple time points were measured to determine the rate of the reaction. In this case, the substrate and enzyme volumes were combined, vortexed for 2 seconds on high, and aliquoted evenly among three microcentrifuge tubes. The reaction started when all three tubes were simultaneously placed on the hot plate, where the reaction ran for the indicated amount of time at 37 °C prior to inactivation. In the case where only one time point was obtained,

the distribution of the aliquots was omitted, and the reaction volume was vortexed and immediately placed on the hot plate. The enzyme was inactivated at 85 °C for 3 minutes at the indicated time point. Each reaction was conducted and tested in triplicate. The amount of glucose released was determined using the protocol for the modified version of GOPOD assay as described in Appendix C. A Non-Linear Regression, Michaelis Menten, Substrate Inhibition, Extra sum-of-squares F test with a P value less than 0.05, was performed using GraphPad Prism version 9.0 for Mac, GraphPad Software, La Jolla California USA, www.graphpad.com. This comparison was done to determine the better model fit for the experimental values, to determine the kinetic parameters (Appendix D).

Table 4. Concentrations and Timepoints Used in Kinetic Endpoint Assays

	Maltose	Isomaltose	Palatinose
Range of Substrate Concentration Tested (mM)	9.375 – 800	6.25 – 300	1.95 – 125
Substrate Solution Volume in Assay (uL)	99	44	99
Enzyme Concentration of 1uL enzyme added (mg/mL)	3	2.25	5
Final Enzyme Concentration (mg/mL)	0.03	0.05	0.05
Reaction Time Points (minutes)	3, 6, 9 [FpAG1] & 4, 8, 12 [FpAG2]	10 [FpAG1] & 45 [FpAG2]	3, 6, 9 [FpAG1] & 30, 45, 60 [FpAG2] (for 3.91mM [FpAG2], where 60, 75, 90)

This table outlines all the substrate conditions, assay volumes, and reaction time points used in the endpoint kinetic assays of FpAG1 and FpAG2.

4.4 Results and Discussion

4.4.1 pH and Temperature Profile

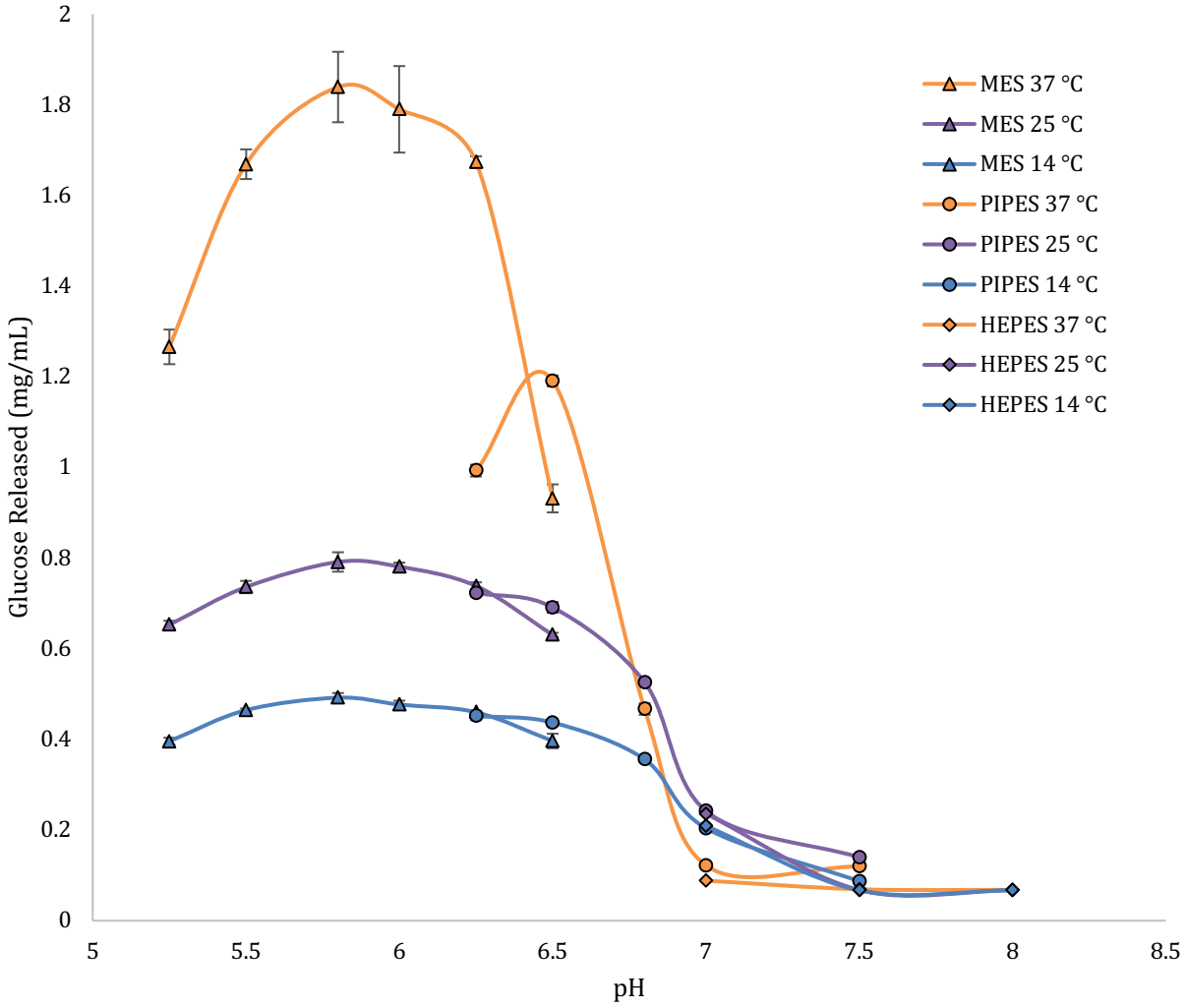


Figure 12. pH and Temperature Profile of FpAG1

This plot presents the amount of glucose produced by enzymatic hydrolysis of 30 mM of maltose with 0.05 mg/mL FpAG1 in a pH range from 5.25 to 8.0 using three different buffers: MES (represented with a triangle symbol), PIPES (represented with a circle symbol) and HEPES (represented with a diamond symbol). Three temperature ranges, 37 °C (orange), 25 °C (purple) and 14 °C (blue), are also displayed in this figure.

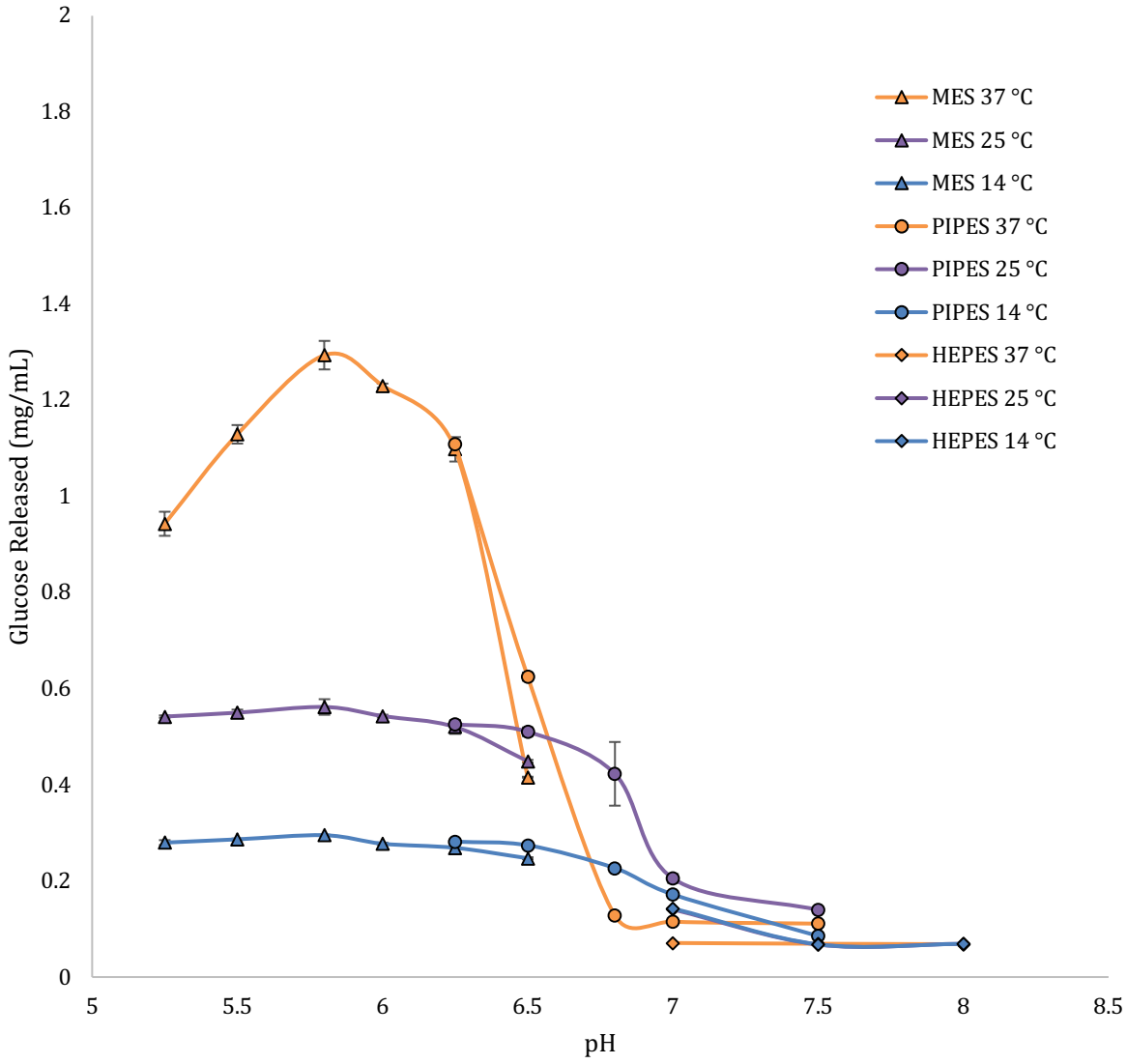


Figure 13. pH and Temperature Profile of FpAG2

This plot presents the amount of glucose produced by enzymatic hydrolysis of 30mM of maltose with 0.05mg/mL FpAG2 in a pH range from 5.25 to 8.0 using three different buffers: MES (represented with a triangle symbol), PIPES (represented with a circle symbol) and HEPES (represented with a diamond symbol). Three temperature ranges, 37 °C (orange), 25 °C (purple) and 14 °C (blue), are also displayed in this figure.

A range of different conditions were tested for FpAG1 and FpAG2 maltose hydrolysis. Three temperatures in approximately 10 °C intervals were tested to determine the effect of temperature on the enzymatic activity of these enzymes. The optimal tested temperature observed for both FpAG1 and FpAG2 was 37 °C, which was the highest temperature tested. Generally, the higher the temperature, the greater the enzymatic rate because of the increased kinetic energy associated with an increase of temperature[82]. *F. prausnitzii* is found in the intestinal tract of the human body, which maintains an average temperature of 36.96 ± 021 °C, P = 0.004[105]. Therefore, due to the native environment of these proteins and thermodynamic principles, the optimal temperature of 37 °C was expected. Higher temperatures were not tested due to the risk of denaturation. When proteins are exposed to temperatures above 40 °C, there is a risk of disrupting the tertiary protein structure, leading to irreversible damage[82].

The maltose hydrolysis of FpAG1 and FpAG2 was tested with pH conditions ranging from 5.25 to 8.0. The greatest amount of glucose was released at a pH of 5.8 for both FpAG1 and FpAG2 (Figure 12 and Figure 13). Both FpAG1 and FpAG2 have a predicted pI of 5.25 [77]. Relative to the peak of glucose release in Figure 12 and Figure 13, enzymatic activity is reduced in the pH range below 5.5. As the protein may be approaching its pI value, it may precipitate out of the solution, thus reducing hydrolytic activity potential[84]. Insignificant amounts of maltose were hydrolyzed at a pH greater than 7.0, suggesting that either the conditions are not optimal for hydrolysis or the change in pH affects the stability of the tertiary structure. This reduction in enzymatic activity could be a result of the pH effect on the tertiary structure of the protein. These pH conditions could also affect the shape of the active site or how the dimer interaction is affected[82]. As aggregation is predicted to occur when these α-glucosidases are exposed to higher pH conditions, this environment may be causing significant structural changes to the protein structure which would prevent maltose hydrolysis[82]. The optimal pH for growth of *F. prausnitzii* was determined to be between 5.7 and 6.7, which is also the pH of the large intestine[24,29]. *F. prausnitzii* has also been found in the duodenum and the terminal ileum, where

the pH is also within this range[24]. Therefore, there is overlap with the pH range where microbial growth is detected and enzymatic activity is measured with these α-glucosidases.

4.4.2 Substrate Specificity of FpAG1 and FpAG2

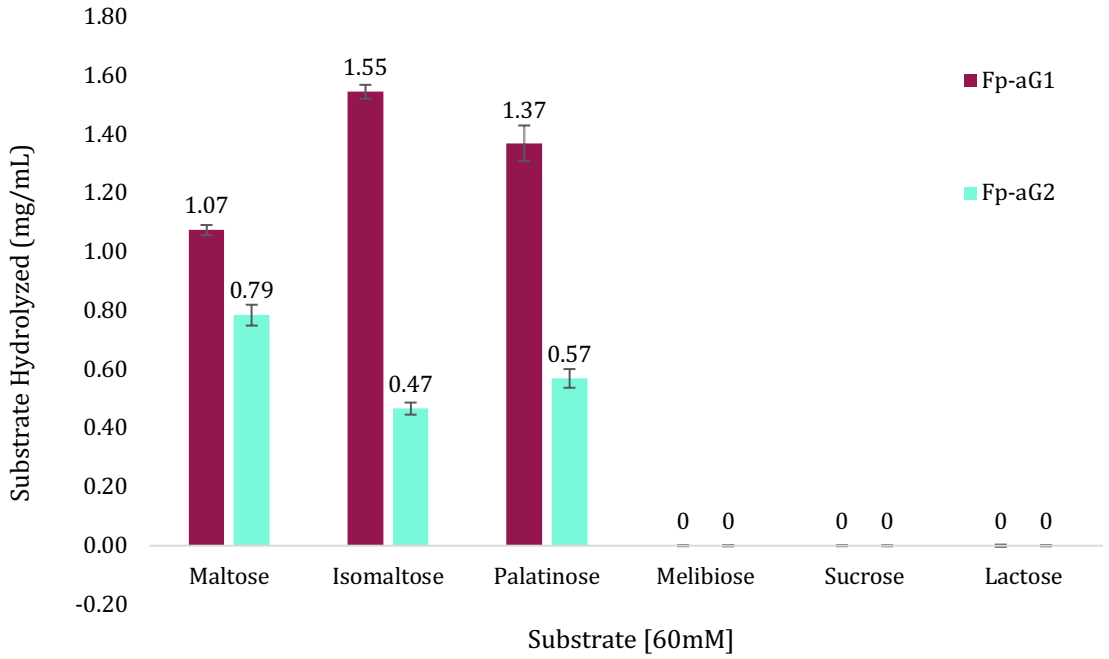


Figure 14. Comparison of Substrate Preference Between FpAG1 & FpAG2

This plot presents the amount of substrate catalyzed by 0.05 mg/mL of FpAG1 & FpAG2, after a 30 minute reaction at 37 °C with 60 mM substrate at a pH of 5.8. The amount of glucose released was detected and the measurements were adjusted to the number of glucose units present in each of the disaccharides. There is substantial hydrolysis of maltose, isomaltose and palatinose observed by FpAG1 & FpAG2, but no significant hydrolysis of melibiose, sucrose and lactose.

As seen in Figure 14, both FpAG1 and FpAG2 hydrolyzed the same disaccharides, maltose, isomaltose and palatinose, as seen by the amount of glucose released in the activity assays. This

indicates that both FpAG1 and FpAG2 have α-glucosidase and oligo-1,6-glucosidase hydrolytic activity. The difference between maltose and isomaltose/palatinose disaccharides is that maltose has α-1,4 glycosidic bond, whereas isomaltose and palatinose have α-1,6 glycosidic bonds. As mentioned in 2.4.1, both *F. prausnitzii* enzymes have a characteristic tryptophan in the active site (Trp166 in FpAG1 and Trp167 in FpAG2) that is predicted to provide substrate specificity for α-1,6 glycosidic bond hydrolysis[64]. Through experimental studies with Ro- αG1 structures by Tan et al, this residue is predicted to stabilize substrates such as isomaltose, as its position allows for hydrogen bonding with the α-1,6 glycosidic bond. As a result, this residue is expected to clash with the α-1,4 glycosidic bond[64]. The hydrolysis of substrates with α-1,6 glycosidic bonds by both FpAG1 and FpAG2 indicates that the tryptophan residue at this site is contributes to α-1,6 substrate specificity. As seen in Figure 14, different amounts of glucose were released from isomaltose and palatinose relative to maltose between the F. prausnitzii enzymes. More glucose was released by FpAG1 when substrates with an α-1,6 glycosidic linkage were hydrolyzed, compared to substrates with an α-1,4 glycosidic linkage. The opposite was observed with FpAG2, where more glucose was released when substrates with an α-1,4 glycosidic linkage were hydrolyzed, compared to substrates with an α-1,6 glycosidic linkage. The difference in isomaltose hydrolysis relative to maltose hydrolysis between FpAG1 and FpAG2 indicates that the presence of this tryptophan residue in the active site is not the only factor contributing to α-1,6 glycosidic linkage substrate preference.

There is significant hydrolytic activity seen with palatinose in Figure 14. This disaccharide has a glucose on the non-reducing end, but it has a fructose at the reducing end. In the Ro-αG1(D307A) crystal structure [PDB 3MKK], only C4'-OH and C3'-OH of isomaltose are involved in the substrate binding on the reducing end of the disaccharide, whereas every hydroxyl group on the non-reducing end of the disaccharide seems to be involved in the hydrogen bonding network[64]. The difference in binding of the reducing end of palatinose could explain the difference in substrate preference between FpAG1 and FpAG2. As the monosaccharide present in the +1 subsite is predicted

to be responsible for interacting with the conserved aspartic residue in the N-terminal domain and initiating the hydrolysis of the substrate, differences in the structure could contribute to the difference in hydrolytic activity of the enzymes.

In enzymatic activity assays with *F. prausnitzii* enzymes, there was no glucose released in the presence of melibiose and sucrose, substrates that have been hydrolyzed by enzymes in the GH31 family. Melibiose was tested to determine whether the *F. prausnitzii* enzymes had α-galactosidase activity. Melibiose has been seen to be broken down by α-galactosidases in the GH31 family such as Pedobacter Gal31 from *Pedobacter heparinus*, PsGal31A[96]. The difference between melibiose and the other disaccharides that showed activity is that there was no glucose at the reducing end of the disaccharide. As seen in Ro-αG1 binding of isomaltose, which contains two glucose units, there are hydrogen bonds between the C4 of the glucose at the non-reducing end and the Asp197 & His478 residues. In galactose, this hydroxyl is in a different orientation, and this may affect the hydrogen bonding network of this substrate binding. The lack of hydrolysis when there is a galactose at the non-reducing end instead of a glucose, suggests that the positioning of C4 hydroxyl is important in substrate binding. Sucrose is hydrolyzed by sucrase α-glucosidases part of the GH31 family, such as the C-terminal subunit of SI[19]. Although sucrose is composed of a glucose at the non-reducing end and fructose at the reducing end, as seen in palatinose, these monosaccharides are linked by an α-1,2 glycosidic bond. Although the glucose is expected to have the same hydrogen bonding network for substrate binding as in palatinose, the fructose would be in a different position, which may interfere with binding at the +1 subsite.

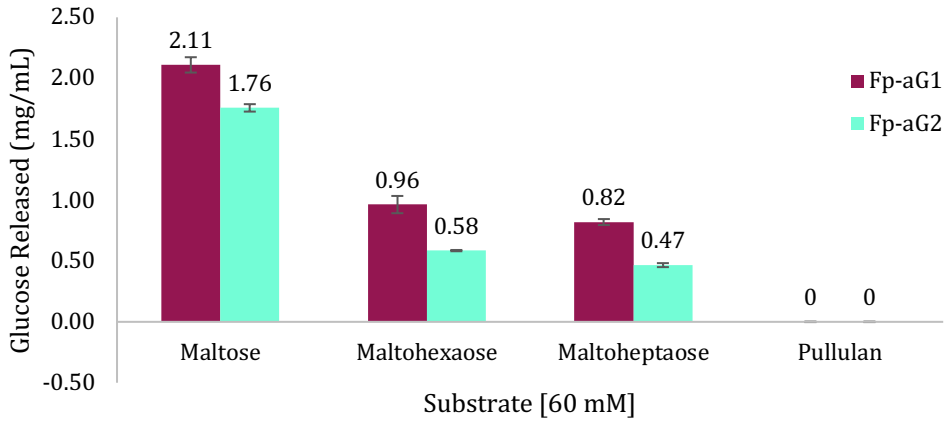


Figure 15. FpAG1 & FpAG2 Substrate Hydrolysis of Oligosaccharides at pH 5.8

This plot presents the amount of substrate catalyzed by 0.05 mg/mL of FpAG1 & FpAG2 after a 30 minute reaction at 37 °C with 60 mM substrate at a pH of 5.8. The amount of glucose released was detected. There is substantial hydrolysis of maltose, maltohexaose and maltoheptaose observed by FpAG1 & FpAG2, but no detected hydrolysis of pullulan.

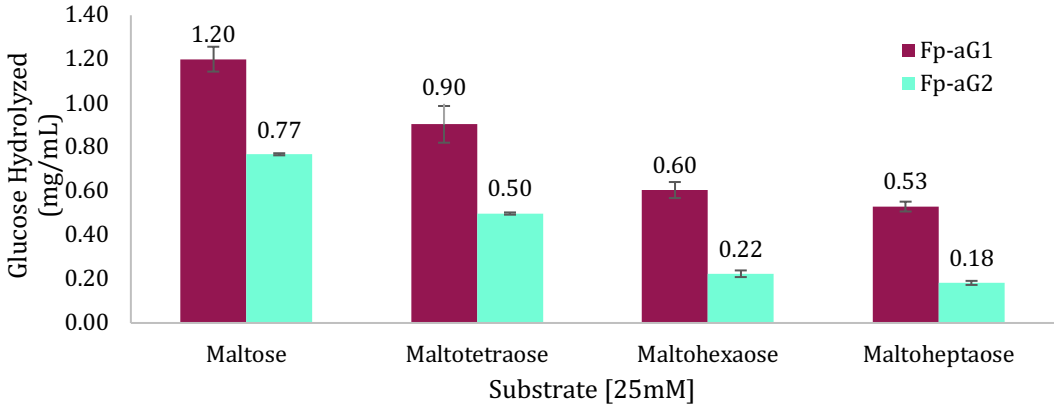


Figure 16. FpAG1 & FpAG2 Substrate Hydrolysis of Malto-oligosaccharides at pH 6.4

This plot presents the amount of substrate catalyzed by 0.05 mg/mL of FpAG1 & FpAG2, after a 30 minute reaction at 37 °C with 25 mM substrate at a pH of 6.5. The amount of glucose released was detected. There is substantial hydrolysis of maltose, maltotetraose, maltohexaose and maltoheptaose observed by FpAG1 & FpAG2.

In enzymatic activity assays with *F. prausnitzii* enzymes, there was no glucose released in the presence of lactose and pullulan, which are substrates that have been hydrolyzed by enzymes in the GH13 and GH1 families. Lactose was tested to determine whether the *F. prausnitzii* enzymes had β-galactosidase activity. Both α-glucosidases in the GH31 family and β-galactosidases in the GH 2 family have an active site in the $(\beta/\alpha)_8$ barrel and hydrolyze carbohydrates through the overall retaining mechanism. Although β-galactosidase activity was not expected due to the lack of structural features, ie. catalytic Glu residues in place of the Asp residues present in the active site of these enzymes, this substrate was tested to confirm the structural hypothesis. Lactose has a galactose at the non-reducing end and a glucose at the reducing end, connected by a β-1,4 glycosidic linkage. The galactose is expected to affect the binding at the -1 subsite as was seen with melibiose. The β-1,4 glycosidic linkage will position the glucose in the reducing end in a different orientation then what is seen in melibiose. The orientation of this glucose at the reducing end avoids binding with the conserved aspartic acid residue in the N-terminal domain (Asp70 in FpAG1, Asp70 in FpAG2), which would prevent the initiation of hydrolysis of lactose. Activity assays with pullulan were conducted with *F. prausnitzii* enzymes to determine whether the enzymes have endo-1,6-α-glucosidase activity. Pullulan is a polysaccharide composed of glucose units connected by both α-1,4 and α-1,6 glycosidic linkages, which are also found in maltose and isomaltose. As seen in Figure 15, neither FpAG1 nor FpAG2 were able to hydrolyze the glycosidic linkages in pullulan. The enzymatic activity test with pullulan suggests that the hydrolysis of both α-1,4 and α-1,6 glycosidic linkages requires a terminal glucose at the non-reducing end of the polysaccharide, but further studies are required to confirm.

Malto-oligosaccharides, including maltose, maltoteraose, maltohexaose and maltoheptaose, were also tested for enzymatic hydrolysis. Multiple enzymatic assays were conducted due to limited resources and all results were included for comparison to other substrates. As seen in Figure 15 and Figure 16, with varying conditions, smaller malto-oligosaccharides released a higher amounts of

glucose. This is seen in both FpAG1 and FpAG2, and the amount of malto-oligosaccharide hydrolyzed relative to maltose generally remained consistent between these two enzymes. This reduction in hydrolysis as the malto-oligosaccharide chain length increased indicates that these enzymes have a preference for shorter chains. The reason for this could be because shorter maltose chains are more accessible to the active site due to their size, are easier to orient into the active site or have a lower thermodynamic cost for binding. In Ro-αG1, the space between the homodimer in the crystal structure, which contained the access to the active sites, was 39 Å[64]. The length of a maltotetraose is about 19 Å[64]. If oligomerization of FpAG1 and FpAG2 were comparable to that observed with Ro-αG1, the accessibility of these longer malto-oligosaccharides to the active site would be restricted, especially since there would be two active sites within this space[64]. The data presented in Figure 15 and Figure 16 is the amount of glucose released by hydrolysis of the terminal glucose. The hydrolysis of malto-oligosaccharides is expected to consist of a mixture of different malto-oligosaccharides. As substrates are hydrolyzed, they break into shorter malto-oligosaccharide chains and the preference of these substrates relative to the original malto-oligosaccharide cannot be confirmed with this assay. Yet, the rate of hydrolysis of longer malto-oligosaccharides, such as maltoheptaose, relative to malto-oligosaccharides such as maltotetraose, in a short time frame can be compared. Further studies need to be done to gain further insight on how longer malto-oligosaccharides are hydrolyzed.

4.4.3 Enzymatic Kinetic Parameters of FpAG1 and FpAG2

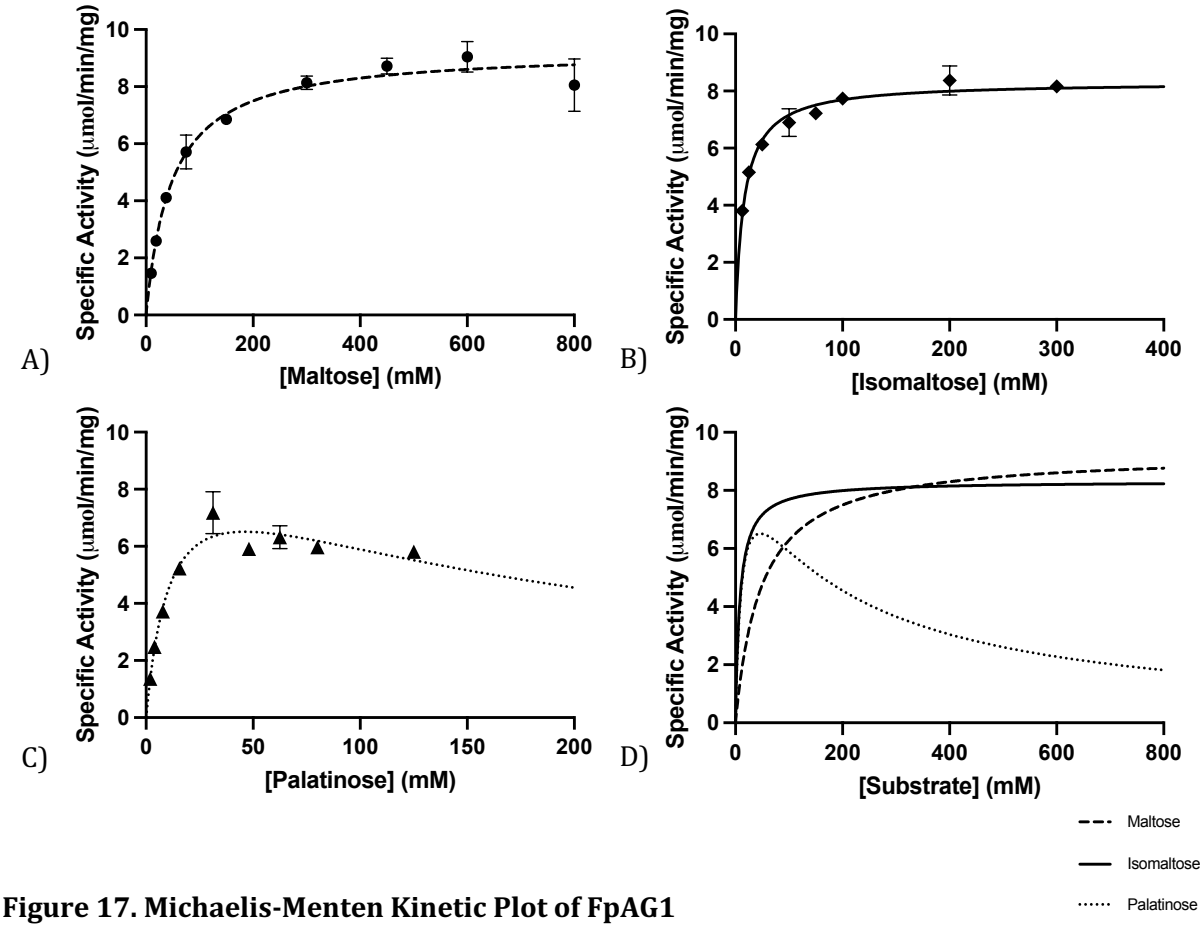


Figure 17. Michaelis-Menten Kinetic Plot of FpAG1

Reaction rates were measured with 25 mM MES, 100 mM NaCl, pH 5.8 buffer, at a temperature of 37 °C. The measured specific activity is the substrate (mM) hydrolyzed per minute, per mg of protein in the reaction. A) Michaelis-Menten kinetic plot of 0.03 mg/mL (0.383 μM) of FpAG1 with a 9.375 mM to 800mM range of maltose concentration. B) Michaelis-Menten kinetic plot of 0.05 mg/mL (0.639 μM) of FpAG1 with a 6.25 mM to 300mM range of isomaltose concentration. C) Substrate inhibition derivative of a Michaelis-Menten kinetic plot of 0.05 mg/mL (0.639 μM) with a 1.95 mM to 125 mM range of palatinose concentration. D) Comparison of the maltose, isomaltose and palatinose enzymatic kinetic plots of FpAG1.

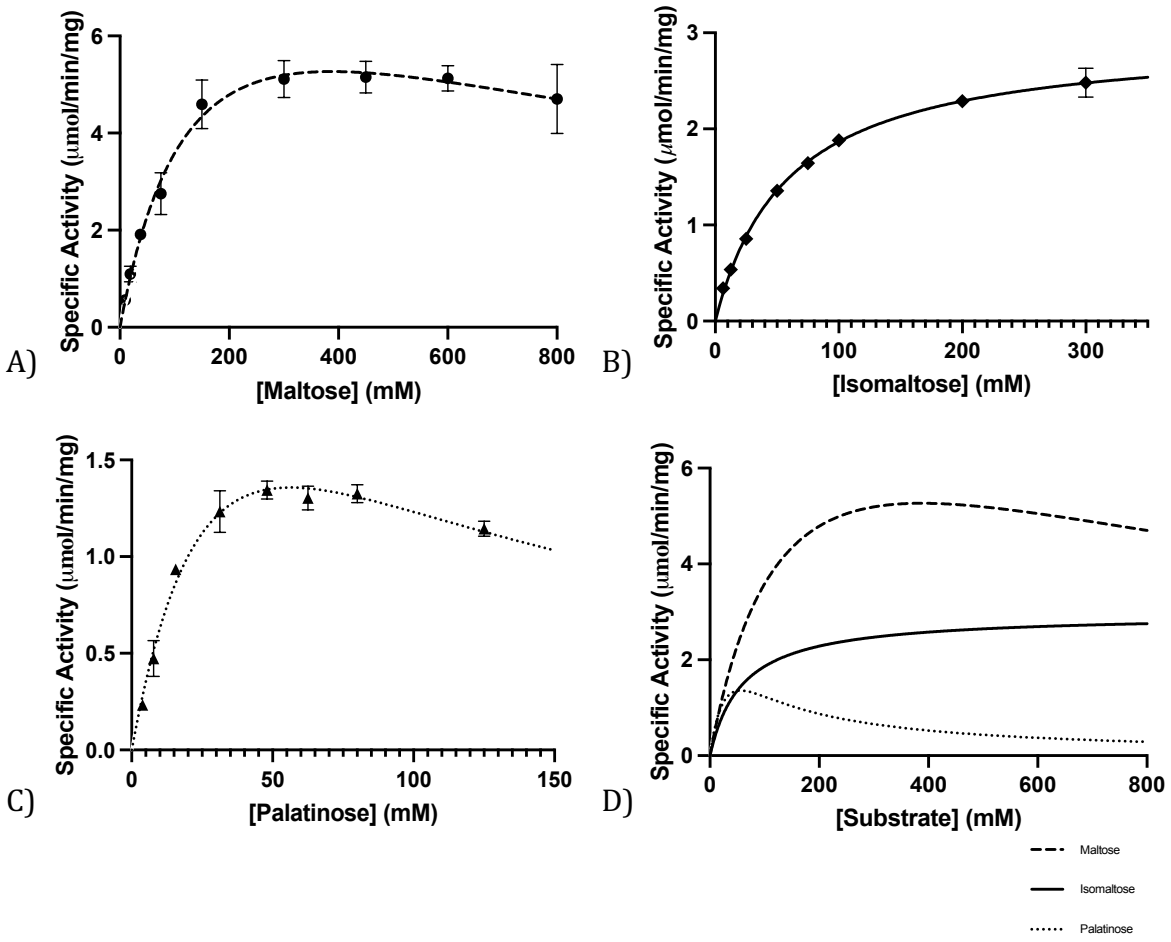


Figure 18. Michaelis-Menten Kinetic Plot of FpAG2

Reaction rates were measured with 25mM MES, 100mM NaCl, pH 5.8 buffer, at a temperature of 37°C. The measured specific activity is the substrate (mM) hydrolyzed per minute, per mg of protein in the reaction. A) Michaelis-Menten kinetic plot of 0.03 mg/mL (0.392 μM) of FpAG2 with a 9.375 mM to 800mM range of maltose concentration. B) Michaelis-Menten kinetic plot of 0.05 mg/mL (0.653 μM) of FpAG2 with a 6.25 mM to 300mM range of isomaltose concentration. C) Substrate inhibition derivative of a Michaelis-Menten kinetic plot of 0.05 mg/mL (0.653 μM) of FpAG2 with a 3.91 mM to 125 mM range of palatinose concentration. D) Comparison of the maltose, isomaltose and palatinose enzymatic kinetic plots of FpAG2.

Table 5. Enzymatic Kinetic Parameters of FpAG1 and FpAG2

Substrate	Maltose		Isomaltose		Palatinose	
	FpAG1	FpAG2	FpAG1	FpAG2	FpAG1	FpAG2
K_M (mM +/- SE; 95% CI)	47.74 +/- 4.32 [39.63, 57.42]	152.4 +/- 35.98; [96.99, 266.3]	8.072 +/- 0.61; [6.86, 9.44]	58.90 +/- 2.69; [53.54, 64.78]	11.72 +/- 2.28; [7.94, 18.07]	43.21+/- 10.50; [28.01, 77.10]
V_{max} (μmol/ min/mg +/- SE; 95% CI)	9.29 +/- 0.19 [8.90, 9.70]	9.41 +/- 1.37; [7.26, 13.66]	8.32 +/- 0.11; [8.08, 8.56]	2.96 +/- 0.05; [2.86, 3.06]	9.81 +/- 0.95; [8.16, 12.45]	3.44 +/- 0.59; [2.57, 5.35]
K_i (mM +/- SE; 95% CI)	N/A	985.90 +/- 364.50; [458.60, 2371]	N/A	N/A	182.40 +/- 55.70; [98.01, 388.60]	73.37 +/- 21.68; [37.47, 130.60]
k_{cat} (s^{-1})	12.12 +/- 0.25	12.01 +/- 1.74	10.86 +/- 0.14	3.78 +/- 0.06	12.80 +/- 1.23	4.39 +/- 0.75
k_{cat}/K_M ($mM^{-1}s^{-1}$)	0.25 +/- 0.02	0.08 +/- 0.02	1.34 +/- 0.10	0.06 +/- 0.00	1.09 +/- 0.24	0.10 +/- 0.03

The kinetic parameters K_M^{app}, V_{max} and K_i^{app}, were estimated with standard error and the 95% confidence interval of these parameters was determined. The Michaelis-Menten kinetic equation was used to determine these parameters, and models that estimate a K_i parameter use the substrate inhibition derivative of the Michaelis-Menten kinetic equation.

Between FpAG1 and FpAG2, the same disaccharides were hydrolyzed, although differences in substrate preference between the enzymes were observed. Kinetic enzymatic assays were conducted to investigate these differences. Through the analysis of the initial rates, the fit of the traditional Michaelis-Menten and the substrate inhibition derivative models was compared and a statistical analysis was completed to determine the kinetic model for the dataset (Appendix D). For FpAG1, only the model utilizing palatinose was fit with the substrate inhibition derivative, but in FpAG2, both maltose and palatinose utilization models were fit with the substrate inhibition model. Substrate inhibition has been observed in α-D-glucosidases and α-D-galactosidases, where inhibition

was observed at concentrations exceeding 5 mM and 50mM, respectively, so there is evidence that substrate inhibition affects similar protein structures[103,106,107]. Due to statistical analysis comparison of the data sets (Appendix D), the kinetic models with the better fit were chosen for the kinetic analysis and are used to compare the enzymatic activity between the substrates tested and between the FpAG1 and FpAG2 enzymes.

Kinetic parameters K_M^{app} and k_{cat}^{app} were estimated for FpAG1 and FpAG2 hydrolysis of maltose to compare the difference in hydrolytic efficiency of these enzymes. Maltose was hydrolyzed with both FpAG1 and FpAG2, but FpAG2 was observed to have lower hydrolytic efficiency. Although the K_M^{app} estimated for FpAG2 maltose hydrolysis is approximately 3 times greater, the k_{cat}^{app} estimated is very similar. In FpAG2, the substrate inhibition derivative of the Michaelis Menten kinetic model is also used to analyze the kinetics, and substrate inhibition may be weakly influencing the efficiency of maltose hydrolysis (K_i^{app} of 985.9 +/- 364.5 mM). The substrate affinity and substrate inhibition are factors predicted to contribute to the difference in maltose hydrolysis identified between FpAG1 and FpAG2.

The K_M^{app} for maltose, isomaltose and palatinose follow the same pattern between FpAG1 and FpAG2, where the K_M^{app} for maltose is significantly greater than the K_M^{app} of isomaltose and palatinose hydrolysis. This suggests that for both FpAG1 and FpAG2, maltose, which has a α-1,4 glycosidic linkage, has lower affinity compared to substrates with α-1,6 glycosidic linkages, isomaltose and palatinose. The presence of the characteristic tryptophan residue may explain why the substrates with an α-1,6 glycosidic linkage have a lower K_M^{app} compared to the substrates with an α-1,4 glycosidic linkage in both *F. prausnitzii* enzymes. This residue is predicted to stabilize the α-1,6 glycosidic bond in disaccharides as its position allows for hydrogen bonding with the α-1,6 glycosidic bond, but as a result, it is expected to clash with the α-1,4 glycosidic bond[64]. Therefore, as the K_M^{app} of maltose relative to the K_M^{app} 's of isomaltose and palatinose is consistent between the *F. prausnitzii*

enzymes, substrate affinity is not expected to contribute to the differences in substrate preference (k_{cat}/K_M) observed between FpAG1 and FpAG2.

The turnover rates, k_{cat}^{app}, were estimated for FpAG1 and FpAG2 hydrolysis of maltose, isomaltose and palatinose to compare the hydrolytic efficiency of these enzymes. In FpAG1, the k_{cat}^{app} parameters estimated between these three substrates are very similar. With similar turnover rates but significantly different K_M^{app}'s, the substrate affinity has a more significant effect on the catalytic efficiency of FpAG1. The k_{cat}/K_M ratios indicate that FpAG1 has greater substrate preference for isomaltose, followed by palatinose and maltose. In FpAG2, the k_{cat}^{app} parameters estimated between these three substrates are significantly different, with the k_{cat}^{app} 's estimated for isomaltose and palatinose being 31.1% and 36%, respectively, of the k_{cat}^{app} estimated for maltose. The reduction of turnover rate can be a result of a difference in the molecular dynamics of the protein, which could be affected by oligomerization of the proteins, flexibility of loops in the active site and suboptimal environmental conditions[69,108–110]. As both the K_M^{app} 's and k_{cat}^{app}'s estimated between these three substrates are different, both the substrate affinity and turnover rate of the hydrolytic reaction are expected to contribute to the difference in catalytic efficiency observed in FpAG2 between the substrates. The k_{cat}/K_M ratios indicate that FpAG2 has a greater substrate preference for palatinose, followed by maltose and isomaltose. Substrate inhibition is predicted to occur in FpAG2 hydrolysis of maltose and palatinose. The k_{cat}/K_M ratio does not take into account the effect of substrate inhibition on the catalytic efficiency, which may put the palatinose predicted hydrolytic efficiency parameter into context when comparing the values to the substrate hydrolysis measured in Table 5 as it has a very low estimated K_i of 73.37 +/- 21.68. As indicated in Chapter 2.3, there is a difference in the conserved residues in close proximity to the possible electrophilic catalytic residue (His419/Asn422). FpAG1 and FpAG2 have residues with different physicochemical properties in this position, which may contribute to the flexibility of the loop, and different substrate preferences. Further research should be conducted to determine whether this residue is important for substrate

affinity or turnover rate in the enzymatic reaction in the active site between the *F. prausnitzii* α-glucosidases. Therefore, despite the similarity of the *F. prausnitzii* protein structures and the conserved residues predicted to be involved in the active site, there are differences observed in the substrate hydrolysis of FpAG1 and FpAG2.

4.5 Chapter Conclusions

Within the parameters tested, FpAG1 and FpAG2 have a broad pH spectrum at a temperature of 37 °C. These enzymatic activity studies reveal that significant α-glucosidase and oligo-1,6-glucosidase activity was seen with both FpAG1 and FpAG2, as maltose, malto-oligosaccharides up to 7 glucose units and disaccharides with an α-1,6 glycosidic linkage and a glucose on the non-reducing end (isomaltose and palatinose) were hydrolyzed. Both FpAG1 and FpAG2 were not able to hydrolyze melibiose, sucrose, lactose and pullulan, indicating that α-galactosidase, β-galactosidase, sucrase or pullulanase activity was not detected. The comparative kinetic studies show that FpAG1 has a greater preference for α-1,6 glycosidic linkages and FpAG2 has a greater preference for α-1,4 glycosidic linkages. In FpAG1, the difference in substrate preference is expected to be a result of the difference in substrate affinity as seen by the K_M^{app}. In FpAG2, the difference in substrate preference is expected to be a result of a combination of substrate affinity, turnover rate and substrate inhibition, as seen by the K_M^{app}, k_{cat}^{app} and substrate inhibition K_i^{app}. Between FpAG1 and FpAG2 maltose utilization, substrate affinity is predicted to contribute the higher maltose hydrolysis observed in FpAG1 compared to FpAG2.

Chapter 5: Inhibition of *F. prausnitzii* GH31 α-glucosidases

5.1 Introduction

Over the last couple of decades, inhibitors have been tested for clinical use for diabetes treatment[111]. With this treatment, human α-glucosidases are inhibited to prevent hydrolysis of ingested carbohydrates so that the amount of free glucose is limited. This is done to reduce and control the glucose levels in diabetic patients[111]. Inhibitors have also been used to determine key residues involved in substrate binding and conformational changes in the active site. The two main α-glucosidase inhibitors that have been approved for clinical use are pseudo-saccharides, acarbose and miglitol[111]. Acarbose is a pseudo-tetrasaccharide and it consists of a valientol ring attached to a maltotriose chain through an imino bridge (Figure 19). The pK_a of this inhibitor is 5.1. Acarbose is synthesized from *Actinoplanes* species through commercial microbial fermentation[112]. This inhibitor is expected to reversibly and competitively bind in the active site of GH31 and GH13 enzymes[113,114]. Clinically, acarbose has been observed to poorly absorb into the intestines and therefore is primarily acting within the gut[115]. Miglitol is a pseudo-monosaccharide that is structurally similar to glucose , but it includes an ethanolamine (Figure 19). Both acarbose and miglitol contain an amine in their structure, but they differ in size and the location of the amine relative to the -1 subsite of the active site. Miglitol is semi-synthetic, as it is derived from 1-deoxynojirimycin[116]. 1-deoxynojirimycin is produced naturally in plants like *Commelina communis,* or in organisms such as *Bacillus* and *Streptomyces*[117]. Miglitol is expected to reversibly and competitively bind to the active site of α-glucosidases such as sucrase, isomaltase and glucoamylase to prevent the hydrolysis of other oligosaccharides[116]. This inhibitor is seen to be mostly absorbed in the small intestine and the pK_a is 5.9[118]. Acarbose and miglitol are well studied inhibitors and mimic maltotetraose and glucose, respectively.

Another area of α-glucosidase inhibition research is testing sulphur-based inhibitors, such as kotalanol, to determine its effectiveness for clinical use and to determine protein structural features

that would contribute to this binding affinity[119–122]. There are a variety of inhibitors that were either isolated or derived from compounds discovered in *Salacia reticulata*, which is a traditional Ayurvedic medicine[122,123]. One of these inhibitors is kotalanol, which has a 1-deoxy-4-thioarabinofuranosyl sulfonium cation and a 1-deoxy-D-heptosyl-3-sulfate anion[123] (Figure 19). This inhibitor has a permanent positive charge[124,125]. Small amounts of kotalanol can be extracted from the plant, but it can also be manufactured synthetically[119,123,126]. Kotalanol is expected to competitively bind to the active site of the α-glucosidases such as sucrase and isomaltase[123]. Research with these sulphur-based inhibitors and their derivatives has provided a lot of structural insight on the active site and the residues involved in binding.

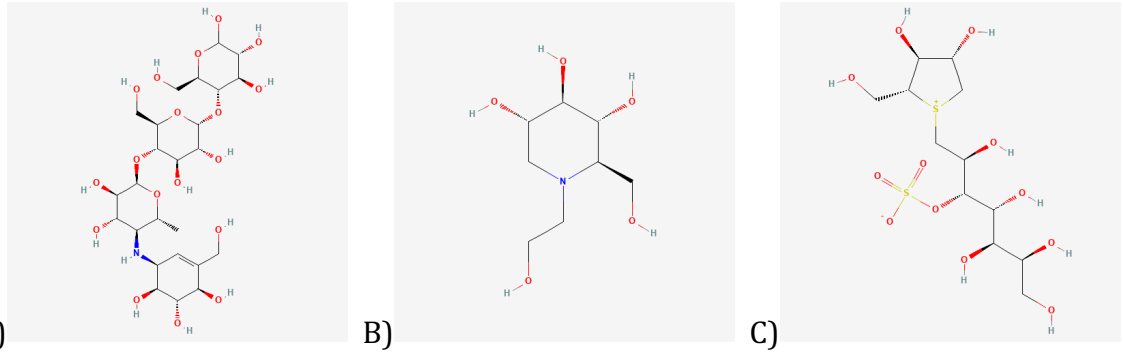


Figure 19. α-glucosidase Inhibitor Structures

These are structures the α-glucosidase inhibitors discussed in this chapter. A) Structure of acarbose B) Structure of miglitol C) Structure of active compound from *Salacia reticulata*, Kotalanol. These structures are obtained from the National Center for Biotechnology Information, . PubChem Compound Summary for CID 41774. *Precose.* (2022). Available at: https://pubchem.ncbi.nlm.nih.gov/compound/Precose. (Accessed: 20th February 2022) [127].

The effect of an inhibitor on an enzyme can be determined through an IC_{50} dose-response sigmoidal curve[128,129]. The equation used to model this relationship is seen below. The IC_{50} dose-response relationship is dependent on the enzyme concentration, substrate concentration and the mode of inhibition[130,131]. To quantify the reduction of enzymatic activity as a response to the inhibitor, the rate of reaction in the presence of inhibitor is normalized to the rate of reaction when no inhibitor is present. This results in the enzymatic activity of the enzymes when no inhibitor is present is 100% and the response at maximum inhibition of the enzymes being 0%. This ensures that the values inputted into the dose-response curves represent only the response of the different inhibitor concentrations in a defined system[132].

$$R = \frac{1}{1 + \left(\frac{\mathrm{IC}_{50}}{[I]}\right)^{n}}$$

Figure 20. IC_{50} equation

R is the response in the presence of different inhibitor concentrations. [I] is the inhibitor concentration. n is the Hill Slope coefficient. IC_{50} is the inhibitor concentration that causes a 50% reduction in enzymatic activity.

The IC_{50} measures the potency of the inhibitor. It is the concentration required to inhibit a system, a specific enzyme concentration and substrate concentration by 50%[130]. The Hill coefficient describes the steepness of the dose-response curve and it indicates how an inhibitor interacts with enzyme as the concentration of the inhibitor increases[133]. It also describes active site cooperativity in inhibitor binding. A standard Hill slope coefficient is 1, which means that the inhibitor response is increasing at a steady rate, independent from any cooperativity effects[134]. Cooperative binding means that when a ligand is bound to a macromolecule, the binding of additional ligands is affected, either positively or negatively[134].

K_i is the dissociation constant describing the binding affinity relationship between the enzyme and the inhibitor[135]. K_i is dependent on buffer conditions and temperature, but is independent of enzyme concentration[136]. The Cheng-Prusoff equations (Table 6) define the theoretical relationship between the K_i and the IC_{50} of an inhibitor using the mode of inhibition, IC_{50}, substrate concentration, enzyme concentration and K_M of the substrate[131]. These equations assume that there is no cooperativity between the inhibitor and enzyme, binding does not occur through any complex inhibition mechanisms, and that the inhibition mechanism is reversible[136]. Within these equations, classical inhibition and tight-binding inhibition are taken into consideration. In classical inhibition models, both the substrate and inhibitor concentrations are in excess of the enzyme concentration and therefore the assumption is that the total concentration and free concentration of substrate and inhibitor is the same[137]. With tight-binding inhibitors, the K_i value is closer to the concentration of the enzyme in the solution, and the assumption of classical inhibition models is no longer held[136,137]. If the free inhibitor concentration and total inhibitor concentration cannot be assumed to be the same in an inhibition assay, tight-binding inhibition equations are required[136,137]. Therefore, the determination of the K_i is dependent on the mode of inhibition and the type of binding that the inhibitor is presenting[136]. The K_i is an important parameter to determine when comparing inhibitors because it provides a constant and comparable value between other inhibitor/enzyme relationships[130].

Table 6. Summary of Calculations for K_i Calculation from IC_{50} value

Modes of Inhibition	Classical Inhibition	Tight-binding Inhibition
Competitive	$K_i = \frac{IC_{50}}{(S/K_m + 1)} \begin{cases} if\ S = K_m, & K_i = IC_{50}/2 \\ if\ S >> K_m, & K_i << IC_{50} \\ if\ S << K_m & K_i \cong IC_{50} \end{cases}$	$K_i = \frac{(IC_{50} - E/2)}{(S/K_m + 1)}$
Uncompetitive	$K_i = \frac{IC_{50}}{(K_m/S + 1)} \begin{cases} if\ S = K_m, & K_i = IC_{50}/2 \\ if\ S >> K_m, & K_i \cong IC_{50} \\ if\ S << K_m & K_i << IC_{50} \end{cases}$	$K_i = \frac{(IC_{50} - E/2)}{(K_m/S + 1)}$
Noncompetitive	$K_i = IC_{50}$ when $S = K_m$ *or* $S >> K_m$ *or* $S << K_m$	$K_i = IC_{50} - E/2$

These are Cheng-Prussof equations that define the relationship between the IC_{50} and the K_i of an inhibitor, in conditions of competitive, uncompetitive and non-competitive inhibition for both Classical and Tight-binding inhibition [131,136,137].

5.2 Chapter Objectives

In this chapter, the inhibition studies of FpAG1 and FpAG2 in complex with acarbose, miglitol and kotalanol are described. Through IC_{50} curves, calculation of the K_i^{app}, and the comparison of inhibitor binding networks observed in ntMGAM, Ro-αG1 and ntSI co-crystallized protein structures, the binding mechanism of these inhibitors is predicted and compared between FpAG1 and FpAG2. The objective of these inhibition studies is to gain insight on the active sites and to identify differences between the active sites of *F. prausnitzii* α-glucosidases.

5.3 Material and Methods

Acarbose (TRC-A123500) and miglitol (TRC-M344200) were purchased from Toronto Research Chemicals, and kotalanol was synthesized by the Pinto laboratory at Simon Fraser University[119]. Concentrated enzyme and the maltose/inhibitor solution were preincubated for

20 minutes in 25mM MES, 100mM NaCl, pH 5.8 buffer at 37°C, prior to combining for a final enzyme concentration of 0.05 mg/mL enzyme (0.626μM FpAG1 and 0.636μM FpAG2), 60mM maltose and inhibitor (final inhibitor concentrations found in Table 7). The reaction was stopped after 25 minutes (or 30 minutes in IC_{50} curve data set) at 37°C by inactivating the enzyme at 85C for 3 minutes. Each reaction was done in triplicate. The amount of glucose released was determined using the protocol for the modified version of GOPOD assay as described in Appendix C. The data was normalized to the maltose control and presented as a percentage of maltose hydrolysis when no inhibitor was present. A Non-Linear Regression, Dose Response – Inhibition, was performed using GraphPad Prism version 9.0 for Mac, GraphPad Software, La Jolla California USA, www.graphpad.com.

Table 7. Inhibitor Concentration Ranges Used in Inhibition Assays

	Acarbose (μM)	Miglitol (μM)	Kotalanol (μM)
FpAG1 IC_{50} curve	500 – 25000	1 - 60	1 – 60
FpAG2 IC_{50} curve	125 - 7500	0.125 - 10	0.25 - 15

The web-based tool "IC50-to-Ki: a web-based tool for converting IC_{50} to K_i values for inhibitors of enzyme activity and ligand binding"[136] was used to convert the experimentally determined IC_{50} values to K_i values. The enzyme concentration, substrate concentration, apparent K_M and apparent IC_{50} were used to determine the K_i of the inhibitor. Enzyme concentration was determined by converting the monomer of the homodimer into molarity and this was done under the assumption that the protein was pure and homogenous. A Non-Linear Regression, Dose Response – Inhibition, was performed using GraphPad Prism version 9.0 for Mac, GraphPad Software, La Jolla California USA, www.graphpad.com.

5.4 Results and Discussion

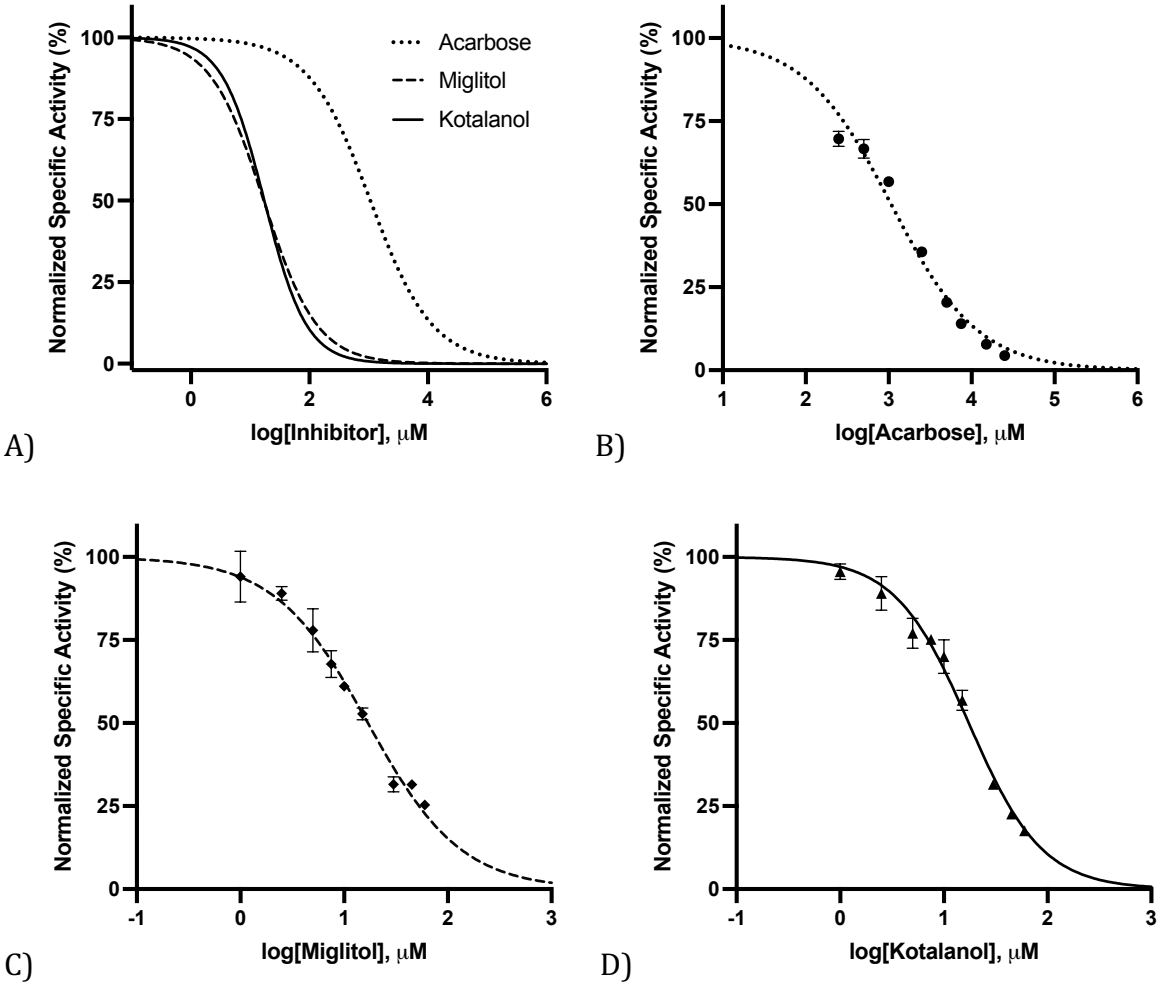


Figure 21. FpAG1 IC_{50} curves with Acarbose, Miglitol and Kotalanol

These plots show the IC_{50} curves of inhibitors acarbose, miglitol and kotalanol, where the amount of 60 mM maltose hydrolyzed by 0.05 mg/mL (0.639 μM) FpAG1 in 25 mM MES, 100 mM NaCl, pH 5.8 buffer, after 30 minutes at 37 °C, was detected. A) This plot is a comparison of the acarbose, miglitol and kotalanol IC_{50} curves; B) Acarbose IC_{50} curve; C) Miglitol IC_{50} curve; D) Kotalanol IC_{50} curve. These measurements were normalized to a maltose control.

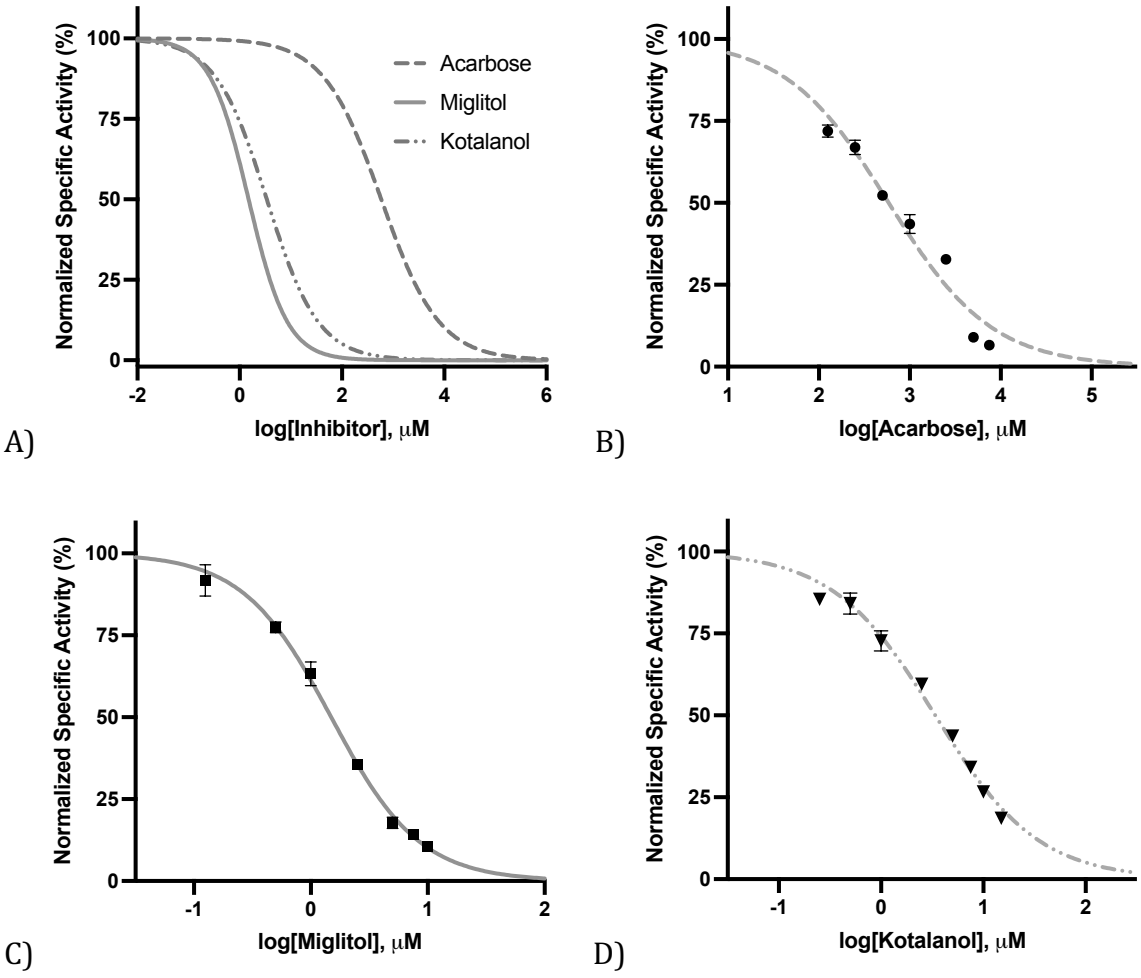


Figure 22. FpAG2 IC_{50} curves with Acarbose, Miglitol and Kotalanol

These plots show the IC_{50} curves of inhibitors acarbose, miglitol and kotalanol, where the amount of 60 mM maltose hydrolyzed by 0.05 mg/mL (0.653 μM) FpAG2 in 25 mM MES, 100 mM NaCl, pH 5.8 buffer, after 30 minutes at 37 °C, was detected. A) This plot is a comparison of the acarbose, miglitol and kotalanol IC_{50} curves; B) Acarbose IC_{50} curve; C) Miglitol IC_{50} curve; D) Kotalanol IC_{50} curve. These measurements were normalized to a maltose control.

Table 8. Summary of All IC_{50} values of FpAG1 and FpAG2 Inhibition Assays

	FpAG1				FpAG2			
	IC_{50} (μM)	R^2	Hillslope	K_i^{app} (μM)	IC_{50} (μM)	R^2	Hillslope	K_i^{app} (μM)
Acarbose	1069 +/- 57.68	0.99	-0.83 +/- 0.04	473.54 +/- 25.42	586.14 +/- 45.08	0.97	-0.77 +/- 0.05	420.33 +/- 32.11
Miglitol	16.94 +/- 0.69	0.98	-0.97 +/- 0.04	7.37 +/- 0.16	1.49 +/- 0.04	0.99	-1.15 +/- 0.03	1.07 +/- 0.03
Kotalanol	17.34 +/- 0.56	0.99	-1.22 +/- 0.05	7.54 +/- 0.11	3.39 +/- 0.12	0.99	-0.86 +/- 0.03	2.43 +/- 0.09

These values are the determined IC_{50} values with standard error, the R^2 value of the fitted curve and the Hill Slope calculated through the non-linear regression analysis. The K_i with standard error was determined under the assumption of competitive inhibition. The K_i for miglitol and kotalanol were calculated using the tight-binding inhibition equation, and the remaining K_i's were calculated using the classical inhibition model (Appendix E).

The K_i of each of these inhibitors was determined to compare the binding affinity of these inhibitors with each of the *F. prausnitzii* α-glucosidases and to estimate an approximate value that can be compared with the Ki's determined in other α-glucosidases. In intestinal α-glucosidases, acarbose, miglitol and kotalanol show competitive inhibition[113,121,123,138,139]. These inhibitors have also been co-crystalized and determined to bind in the active sites of α-glucosidases in the GH31 family[20,120,140]. Without any additional data regarding how these inhibitors interact with FpAG1 and FpAG2, the assumption that will be used in this analysis is that the mode of inhibition is competitive and that these inhibitors can completely reduce enzymatic activity at high enough inhibitor concentrations. Future studies will have to be

completed to confirm this assumption. The computational models were superimposed with the co-crystallized protein structures to compare the residues in the active site in attempt to identify differences in FpAG1 and FpAG2 protein structures that may explain the difference in inhibition observed experimentally. Although positional differences in the residues are observed when active sites of GH31 enzymes are superimposed, the residues in the active site are expected to be flexible as seen in the crystal structures of Ro-αG1[64] and a static computational model may not adequately show the binding of the inhibitors in the active site. By investigating the inhibition effect of these pseudosaccharides on FpAG1 and FpAG2, insight into the key residues in the active site can provide information on the key differences between these *F. prausnitzii* α-glucosiadses.

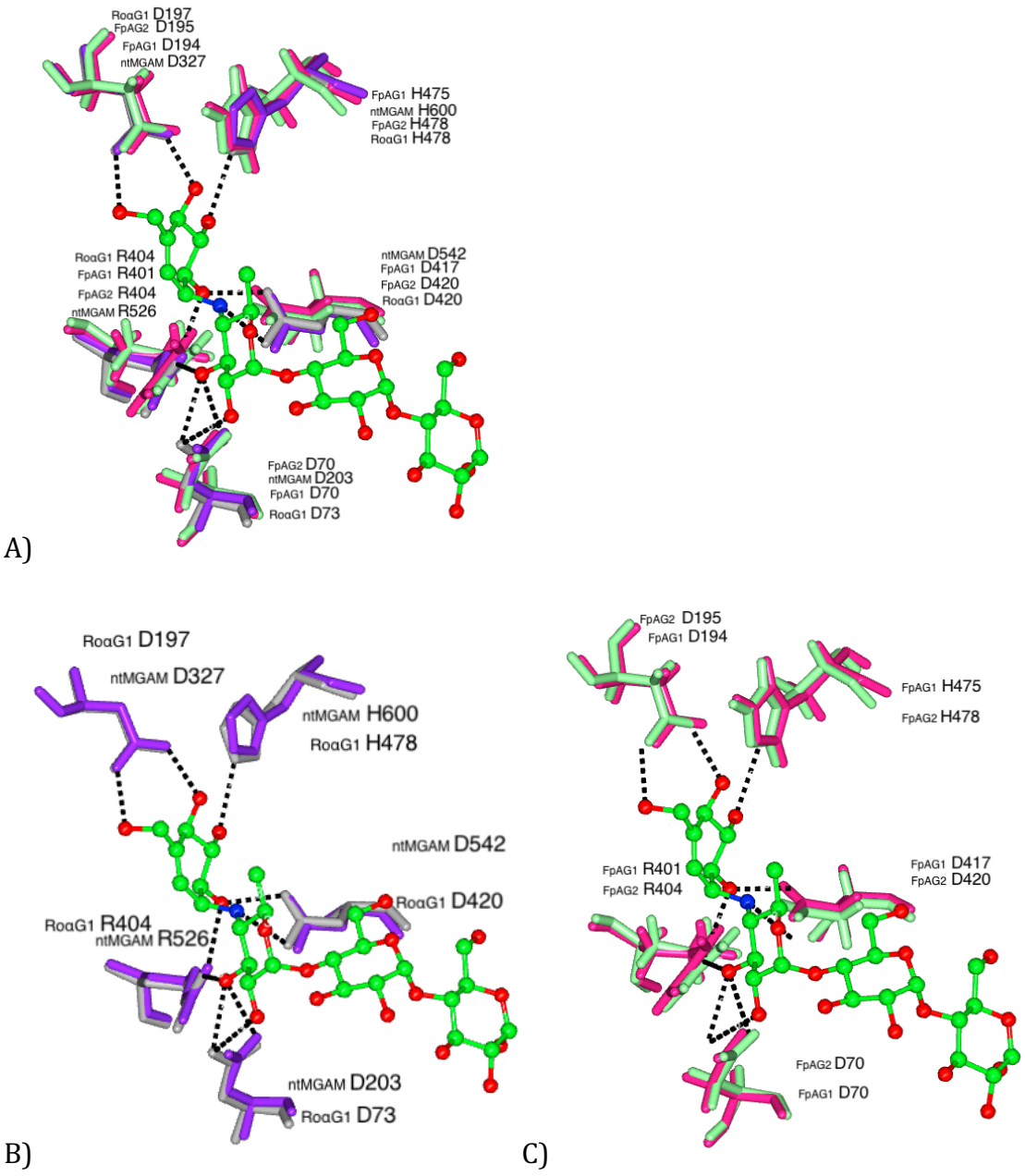


Figure 23. Superimposed Protein Models with ntMGAM/Acarbose Structure and Ro-αG1 (W167Y)/Acarbose Structure

In A) superimposed ntMGAM (grey), Ro-αG1 (W167Y) (purple), FpAG1(pink), and FpAG2 (green) structure residues identified to bind with the inhibitor in the -1 and +1 subsites. In B),

ntMGAM/acarbose structure [PDB 2QMJ] residues (grey) with superimposed Ro-αG1 (W167Y)/ acarbose structure [PDB 3PHA] residues (purple) residues identified to bind with the inhibitor in the -1 and +1 subsite. In C), FpAG1 structure residues (pink) with superimposed FpAG2 structure residues (green) predicted to bind with the inhibitor in the -1 and +1 subsite. These residues are conserved and superimpose well with those seen in ntMGAM [PDB 2QMJ] and Ro-αG1 (W167Y) [PDB 3PHA]. Structures were superimposed with CCP4MG version 2.10.11[141].

Due to the similarities in the active site between FpAG1 and FpAG2 with ntMGAM and Ro-αG1, the valientol ring of acarbose is predicted to have the same $^2H_3$ half-chair conformation as seen in co-crystallized protein structures[3]. In these crystallized structures, the imino bridge of acarbose forms an electrostatic interaction with the proton donor aspartic acid in the active site of ntMGAM and Ro-αG1(W169Y)[120,140]. In the ntMGAM/acarbose structure, Asp 327, His600, Asp 542 and Arg526 interact with the valientol ring in the -1 subsite. Asp542, Arg526 and Asp203 interact with the adjacent ring in the +1 subsite through hydrogen bonds[63]. The binding of acarbose to Ro-αG1(W169Y) [PDB 3PHA] was observed to be nearly identical[140], as the acarbose binding residues superimpose well between these two crystal structures. In FpAG1 and FpAG2 computational models, the conserved residues in the active site superimpose well with those in ntMGAM and Ro-αG1(W167Y) [PDB 3PHA] (Figure 23) and there is no obvious difference in the residues predicted to be involved in the binding of acarbose. Acarbose was found to inhibit ntMGAM and ntSI with a K_i of 62 +/- 13 μMand 14 +/- 1 μM, respectively[20,121]. These values are significantly lower than the K_i^{app}'s determined for FpAG1 (473.54 +/- 25.42 μM) and FpAG2 (420.33 +/- 32.11 μM). The K_i^{app} of acarbose for FpAG1 and FpAG2 is within the same magnitude

and therefore there is not a significant difference in binding affinity of acarbose between these two enzymes.

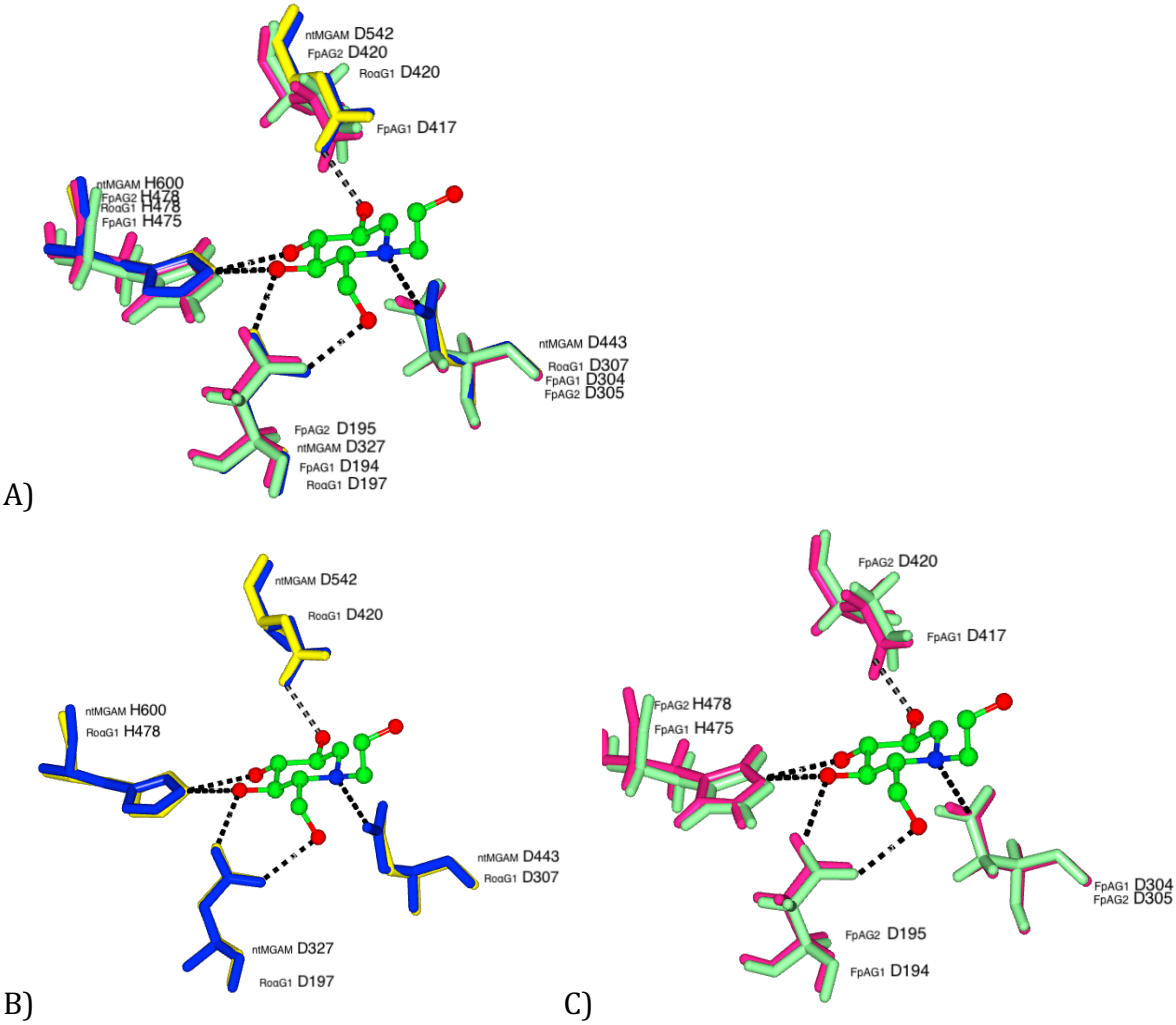


Figure 24. Superimposed Protein Models with ntMGAM/Miglitol Structure and Ro-αG1 (W167Y)/Miglitol Structure

In A) superimposed ntMGAM (yellow), Ro-αG1 (W167Y) (blue), FpAG1(pink), and FpAG2 (green) structure residues identified to bind with the inhibitor in the -1 subsite. In B), ntMGAM/miglitol structure [PDB 3L4W] residues (yellow) with superimposed Ro-αG1 (W167Y)/miglitol structure [PDB 6CA3] residues (blue) residues identified to bind with the

inhibitor in the -1 subsite. In C), FpAG1 structure residues (pink) with superimposed FpAG2 structure residues (green) predicted to bind with the inhibitor in the -1 subsite. These residues are conserved and superimpose well with those seen in ntMGAM [PDB 3L4W] and Ro-αG1 (W167Y) [PDB 6CA3]. Structures were superimposed with CCP4MG version 2.10.11[141].

Similarly to acarbose, miglitol is predicted to have the same binding observed in ntMGAM and Ro-αG1, due to the structural similarities of these co-crystallized protein structures with FpAG1 and FpAG2. In this experimental data, the ring amine of miglitol has a $^4C_1$ structure and forms an electrostatic interaction with the nucleophilic catalyst in the active site of ntMGAM and Ro-αG1(W169Y) [PDB 6CA3][120,140]. In these structures, miglitol took on a chair structure, which is seen during substrate binding, instead of an oxacarbenium transition state[120,140]. In ntMGAM/miglitol structure, the small inhibitor was observed to make hydrogen bonds with Asp327, His600 and Asp542, and the nitrogen was in close proximity to the catalytic nucleophile of ntMGAM, Asp443[120]. In the Ro-αG1/miglitol structure, the same residues were involved in binding this inhibitor[140]. A difference in the active site between the ntMGAM/miglitol and Ro-αG1/miglitol crystal structures was that the ethanolamine in the miglitol caused conformational change in the ntMGAM/miglitol structure, where the Trp406 shifted and the Met444 residue shifted to fill this void[120]. In the Ro-αG1/miglitol structure, no structural conformation occurred[140]. In FpAG1 and FpAG2, these residues are conserved and superimpose well with those in ntMGAM and Ro-αG1(W167Y) (Figure 23) and there is no obvious difference in the residues predicted to be involved in the binding of miglitol. With these static computational protein models, it is unknown whether the ethanolamine of this inhibitor causes structural conformations in these *F. prausnitzii* enzymes. Although there is no obvious difference in the

residues predicted to bind to miglitol, there is an 8-fold difference in binding affinity between FpAG1 and FpAG2, indicating that there are differences in the -1 subsite that would contribute to these experimental results. These results are also not consistent with the K_M^{app} estimated in 4.4.3, where FpAG2 had higher K_M^{app} compared to FpAG1, for maltose, isomaltose and palatinose. This is of interest particularly since miglitol is predicted to take on a chair structure as seen in substrate binding. Miglitol was found to inhibit ntMGAM with a K_i of 1.0 +/- 0.1 μM, which is very similar to the K_i calculated with FpAG2 (1.07 +/- 0.03uM), and lower than the K_i calculated with FpAG1 (7.365 +/- 0.16 μM)[120]. This similarity in miglitol K_i^{app} between ntMGAM and FpAG2 may indicate structural similarities within these active sites, and further research should be done to identify these structural features and the difference in binding of miglitol between FpAG1 and FpAG2.

In inhibition studies with both FpAG1 and FpAG2, miglitol has a lower K_i^{app} than acarbose by two orders of magnitude. The difference between these inhibitors, the ring shape in the -1 subsite (valientol vs. six-membered ring with an amine) and the structural conformation of the compound ($^2H_3$ half-chair vs $^4C_1$ structure), could influence the hydrogen bonds interactions with the active site residues and thus the binding affinity of these inhibitors. Miglitol also has a pK_a of 5.9, compared to pK_a 5.1 of acarbose, and at an assay pH of 5.8, the miglitol amine is more likely to be protonated and carry a positive charge than acarbose, thus improving the electrostatic interaction. The difference in size between acarbose and the smaller inhibitor miglitol, may also be a reason why acarbose is a less effective inhibitor than miglitol in both FpAG1 and FpAG2. Thermodynamics of binding and accessibility of the active site may contribute to this 100-fold difference in inhibition between miglitol and acarbose. The reduced binding affinity of acarbose in the inhibition assays is consistent with the reduced rate of hydrolysis of longer malto-

oligosaccharides, such as maltotetraose, relative to smaller substrates such as maltose, seen in 4.4.2. Therefore, acarbose is a weak inhibitor of both FpAG1 and FpAG2, relative to the smaller inhibitor tested, miglitol.

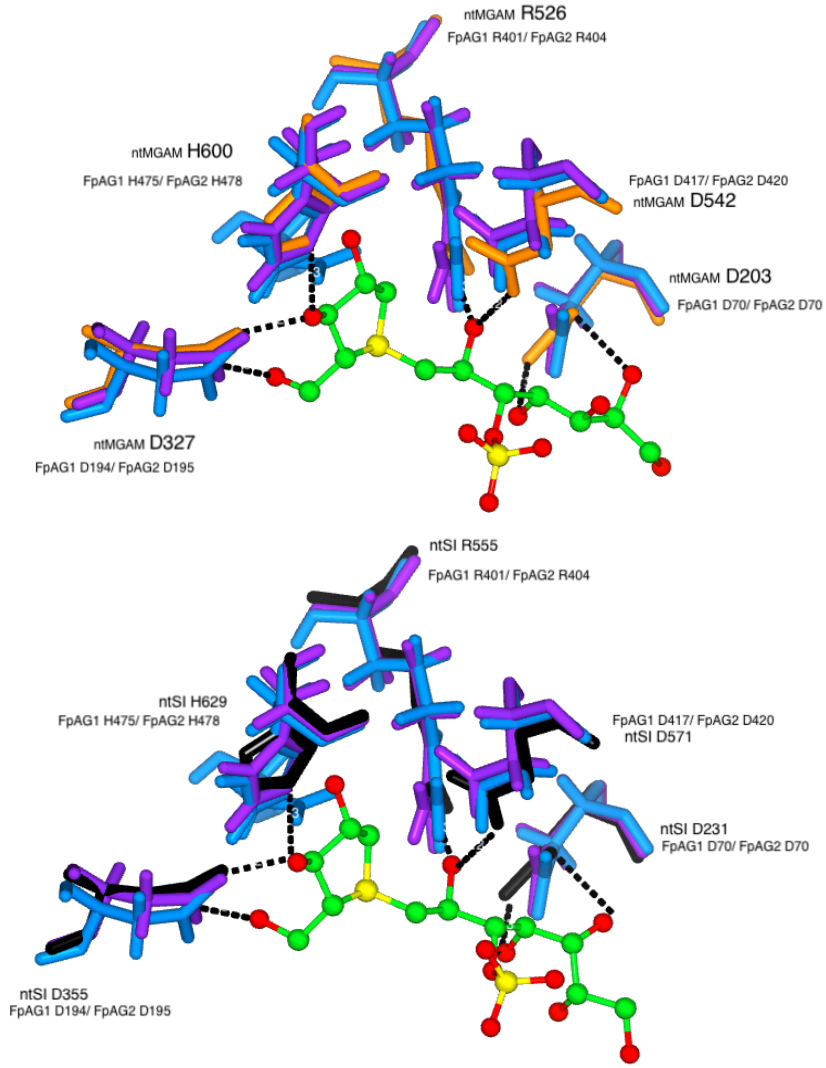


Figure 25. Superimposed Protein Models with ntMGAM/Kotalanol Structure and ntSI/Kotalanol Structure

In A), ntMGAM/kotalanol structure [PDB 3L4V] residues (orange) identified to bind with the +1 subsite with FpAG1 (purple) and FpAG2 (blue) residues superimposed. In B),

ntSI/kotalanol structure [PDB 3LPP] residues (black) identified to bind with the +1 subsite with FpAG1 (purple) and FpAG2 (blue) residues superimposed. Aside from the slight positional differences, these residues are conserved and superimpose well with those seen in ntSI and ntMGAM. Structures were superimposed with CCP4MG version 2.10.11[141] . The residues superimposed in this figure are found in Appendix F.

The binding of kotalanol in ntMGAM and ntSI was compared to FpAG1 and FpAG2, due to the structural similarity of the active site, as seen in Figure 25. In the ntMGAM/kotalanol and ntSI/kotalanol co-crystallized protein structures, there are 8 residues involved in hydrogen bond network of the ring in the -1 subsite of the active site (Appendix F)[20]. These residues superimpose well in the ntMGAM and ntSI co-crystallized protein structures and also with the residues in the FpAG1 and FpAG2 computational protein structures, as seen in Figure 25. Therefore, due to the conservation of the residues and their positions in the active site, the same hydrogen bonding network is expected in the binding of kotalanol in the -1 subsite of FpAG1 and FpAG2 active sites, as seen in ntMAGM and ntSI.

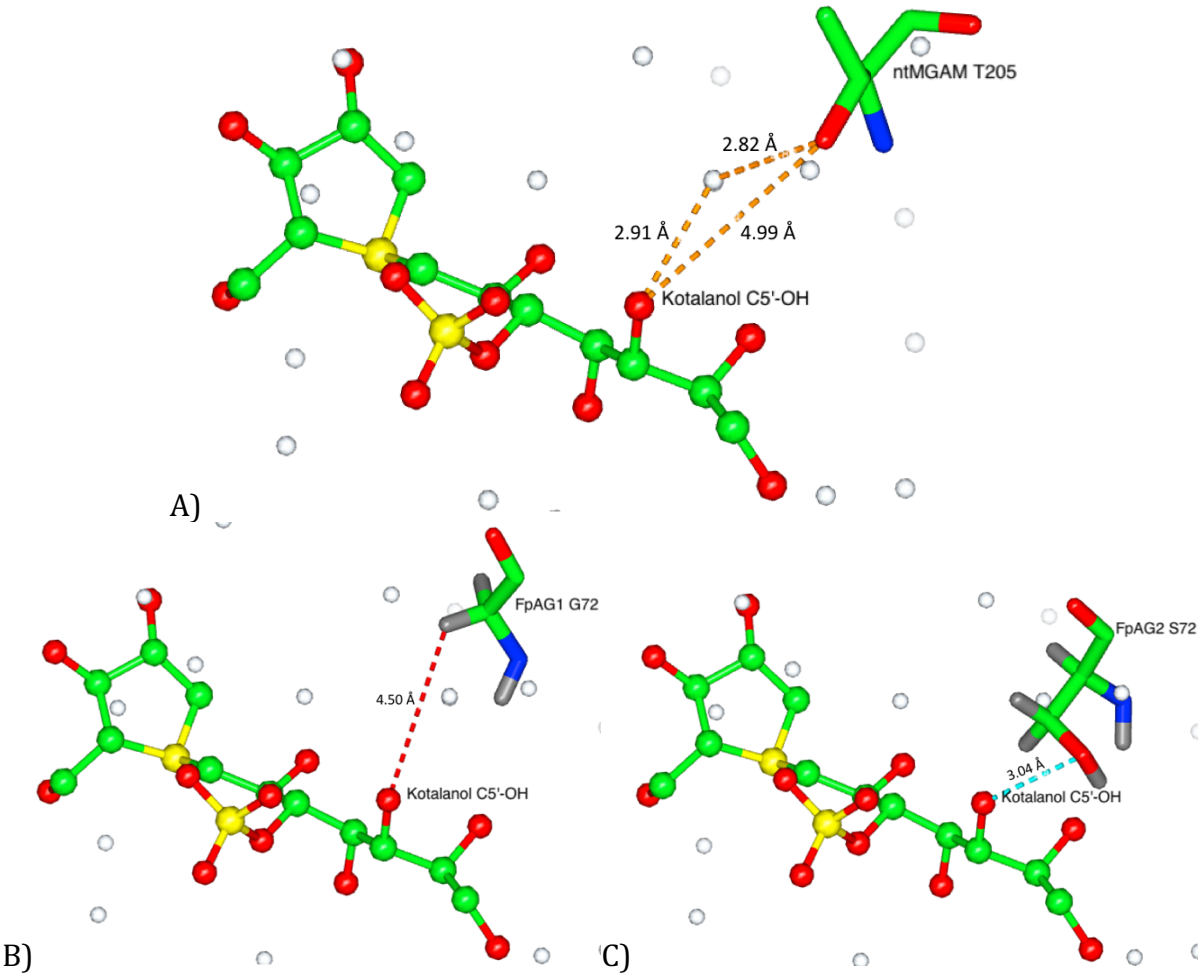


Figure 26. Superimposed Protein Models with ntMGAM/Kotalanol Structure and Predicted Residue Binding to C5'-OH of Kotalanol

In A), ntMGAM/kotalanol structure [PDB 3L4V], Thr205 from ntMGAM is forming a hydrogen bond with C5'-OH via a water molecule. In B), ntMGAM/kotalanol ligand [PDB 3L4V] is superimposed with the +1 subsite with FpAG1 and Gly72 is shown in the same position as Thr205 from ntMGAM. This residue is 4.50 Å away from the C5'-OH of kotalanol. In C), ntMGAM/kotalanol ligand [PDB 3L4V] is superimposed with the +1 subsite with FpAG1 and Ser72 is shown in the same position as Thr205 from ntMGAM. This residue is 3.04 Å away from

the C5'-OH of kotalanol. Although there are slight positional differences, the hydroxyl from serine may form a hydrogen bond with C5'-OH of kotalanol as seen in ntMGAM. Structures were superimposed with CCP4MG version 2.10.11[141].

Between ntSI and ntMGAM, there are differences observed in the binding of the polyhydroxylated tail of kotalanol interacting with the +1 subsite residues. In the ntMGAM +1 subsite, there is a hydrogen bond between Thr205 and the C5'-OH of kotalanol through a water molecule[20]. In the FpAG1 protein model, there is a Gly72 in this residue position, and in the FpAG2 protein model, there is a Ser72 in this residue position. Threonine is a polar residue and the hydroxyl group can interact with the C5'-OH of kotalanol through a water molecule. The glycine in this position in FpAG1 cannot form a hydrogen bond, and therefore the binding of kotalanol in the +1 subsite in FpAG1, is not expected to occur in the manner seen in ntMGAM. The Ser72 in FpAG2 is 3.04 Å away from the C5'-OH of kotalanol, which is within the distance to form a hydrogen bond[142] and therefore this is a possible binding mechanism of kotalanol in the FpAG2 +1 subsite. This difference in the physiochemistry of residues could explain the difference in kotalanol inhibition observed experimentally between FpAG1 and FpAG2, as a loss of a hydrogen bond can affect the binding affinity of a ligand, and there is an 3-fold difference in K_i of kotalanol observed between FpAG1 and FpAG2.

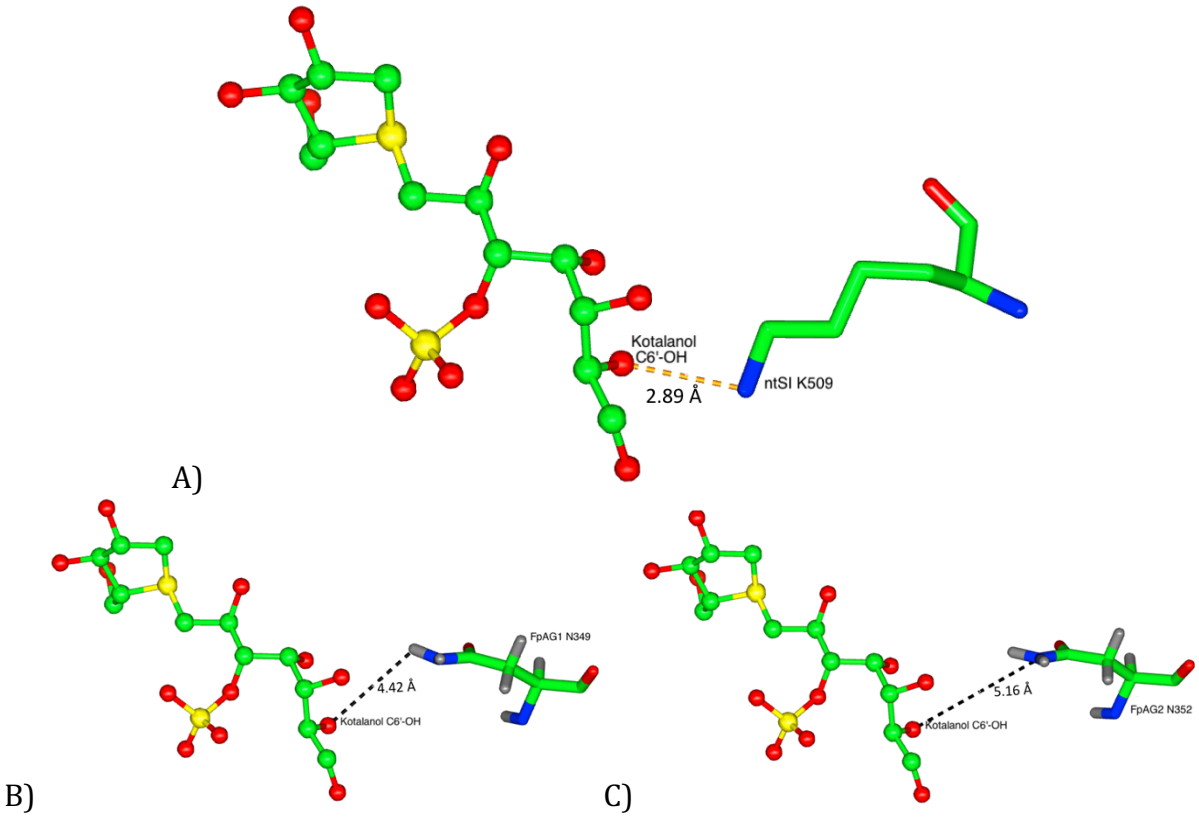


Figure 27. Superimposed Protein Models with ntSI/Kotalanol Structure and Predicted Residue Binding to C6'-OH of Kotalanol

In A), ntSI/kotalanol structure [PDB 3LPP] residues (Lys 509) identified to bind with kotalanol in the +1 subsite. Lys509 from ntSI is forming a hydrogen bond with C6'-OH. In B) ntSI/kotalanol structure ligand [PDB 3LPP] is superimposed with FpAG1, and Asn349 from FpAG1 is shown in the same position as Lys509 from ntSI. Asn349 is 4.42 Å away from C6'-OH of kotalanol. In C), ntSI/kotalanol ligand [PDB 3LPP] is superimposed with FpAG2, and Asn352 from FpAG2 is shown in the same residue position as Lys509 from ntSI. Asn352 is 5.16 Å away from C6'-OH of kotalanol. Structures were superimposed with CCP4MG version 2.10.11[141].

The binding mechanism of kotalanol in ntSI is different than that observed in ntMGAM because in the Thr205 residue position of ntMGAM, there is a Leu233 in ntSI. Leucine cannot

form hydrogen bonds, and therefore in this hydrogen binding network, there is no hydrogen bond with the C5'-OH of kotalanol via a water molecule, as seen in the ntMAGM/kotalanol co-crystallized protein structure. The tryptophan present in the ntSI [PDB 3LPP] (in the same position as the Trp166 in FpAG1 and Trp167 in FpAG2), causes a displacement of kotalanol sulfate group by more than 1 Å away from its position in ntMGAM [PDB 3L4V][20]. Instead of forming a hydrogen bond with the C5'-OH of kotalanol, in the ntSI +1 subsite, there is a 2.89 Å hydrogen bond between Lys509 and the C6'-OH of kotalanol [20]. In the superimposed FpAG1 and FpAG2 structures with ntSI, there is an Asn349 in the FpAG1 protein model and a Asn352 in the FpAG2 protein model in this residue position. These residues in the FpAG1 and FpAG2 protein structures can participate in hydrogen bonding, but based on the static computational models, Asn349 from FpAG1 is 4.423A and Asn352 in FpAG2 is 5.160A away from the C6'-OH of kotalanol (Figure 27). These residues are further away from kotalanol than Lys 509, and unless they form hydrogen bonds via water molecules, it is unlikely that these residues are involved in binding the polyhydroxylated chain of kotalanol.

Kotalanol was found to inhibit ntMGAM and ntSI with a K_i of 0.19 +/- 0.03 μM and 0.6 +/- 0.06 μM, respectively[20,121], values which are significantly lower than the K_i^{app} calculated with FpAG1 (7.54 +/- 0.11 μM) and FpAG2 (2.43 +/- 0.09 μM). The difference in physiochemical properties and position of the residues in the +1 subsite involved in the hydrogen bonding network with polyhydroxylated tail of kotalanol, may illuminate the difference in inhibitor binding between these proteins. If the binding of kotalanol in the +1 subsite of the *F. prausnitzii* α-glucosidases follows the same binding as seen in ntMGAM, the presence of residues capable of forming hydrogen bonds would increase the binding affinity of this sulfonium ion inhibitor and would offer an explanation why kotalanol is a greater inhibitor in FpAG2 compared to FpAG1. It

is possible that a different binding mechanism is occurring in the *F. prausnitzii* α-glucosidases than those mechanisms identified in ntSI and ntMGAM, and further studies have to be done to identify the difference in binding of kotalanol in FpAG1 and FpAG2.

The kotalanol and miglitol Ki's in FpAG1 are not significantly different, whereas the kotalanol K_i^{app} in FpAG2 is approximately double the miglitol K_i^{app} in FpAG2. The difference in ring shape would result in differences in hydrogen bond interactions in the -1 subsite. The presence of the long polyhydroxylated tail interacting with the +1 subsite would also affect the inhibitor binding affinity, as it could either improve or interfere with the binding of the inhibitor. As the residues predicted to interact with kotalanol in the -1 subsite superimpose well in ntMAGM, ntSI, FpAG1 and FpAG2, the structural features that affect the binding of kotalanol between FpAG1 and FpAG2 is predicted to occur in the +1 subsite. The greater binding affinity of miglitol and kotalanol in FpAG2 compared to FpAG1 in inhibition assays is not consistent to the enzymatic assays done in Chapter 3, as FpAG2 generally had higher K_M^{app}'s and lower rates of hydrolysis compared to FpAG1. This difference in the active site may suggest the possibility of untested substrate specificities. Differences in experimental inhibition activity between these enzymes suggest the influence of key structural differences in the +1 subsite between these *F. prausnitzii* α-glucosidases.

It is interesting to note that ntMGAM has a greater preference for short oligosaccharides with α-1,4 glycosidic linkages over α-1,6 glycosidic linkages[20], which was seen with the FpAG2 enzymatic activity (4.4.2). In ntSI, there is preference for both α-1,6 glycosidic linkages and α-1,4 glycosidic linkages[20], which is comparable to the substrate specificity seen in FpAG1 (4.4.2). The residues indicated in the +1 subsite in kotalanol binding are adjacent to the conserved aspartic acid residue in the N-terminal domain that has a role is initiating hydrolysis of carbohydrates[63].

This observation suggests that there may be similarities between the architecture of the +1 subsite of the FpAG2 and the ntMGAM active sites, which are not present in FpAG1 and ntSI. Further research should investigate the residues in the +1 subsites to determine whether these residues have a role in the preference between α-1,4 and α-1,6 glycosidic linkages.

5.5 Chapter Conclusions

All the tested inhibitors, acarbose, miglitol and kotalanol, inhibited maltose hydrolysis of FpAG1 and FpAG2, and all these α-glucosidase inhibitors were found to be more potent in FpAG2 than FpAG1. These inhibition studies revealed that the smaller inhibitors, miglitol and kotalanol, have binding affinities that are two orders of magnitude lower than the K_i^{app} of acarbose in both FpAG1 and FpAG2, indicating that acarbose is a relatively weak inhibitor of *F. prausnitzii* α-glucosidases. Although miglitol and kotalanol had comparable binding affinities in FpAG1, the K_i^{app} of kotalanol was two-fold greater than the K_i^{app} of miglitol in FpAG2. The differences in the inhibition of miglitol and kotalanol observed between FpAG1 and FpAG2, suggest distinguishing features in the -1 and +1 subsite of the active sites of *F. prausnitzii* α-glucosidases.

Chapter 6: Conclusions and Future Directions

6.1 Conclusions

In this thesis, the substrate specificities, enzymatic kinetics and inhibition of FpAG1 and FpAG2 by α-glucosidase inhibitors were investigated. Confident computational protein models of these GH31 enzymes were created using AlphaFoldV2.1.0. High structural similarity of the protein backbone was observed, but there was a difference in the conserved domains identified in the structures. Kinetic studies showed that these *F. prausnitzii* GH31 enzymes have the same substrate specificity but different substrate preferences, as FpAG1 had greater preference of α-1,6 and FpAG2 had greater preference of α-1,4 glycosidic linkages. Lastly, inhibition studies elucidated distinguishing binding affinities of inhibitors, suggesting different structural architecture in the active sites of FpAG1 and FpAG2.

The microbes in the gastrointestinal tract play an important role in human health that has not yet been fully deciphered. As *F. prausnitzii* is expected to compose about 5% of a healthy gut microbiota, its contribution to various aspects of human health and digestion is noteworthy. It is important to understand the starch digestion process in the gastrointestinal tract, not only by human digestive enzymes, but also by the microbes that we are reliant on. Gaining insight on the key residues and structural features of the GH31 enzymes can elucidate the hydrolytic mechanism of these enzymes and explain the distinguishing substrate specificities and preferences seen among this family. Learning more about the architecture in the active site of GH31 enzymes can help identify the key roles of these residues to further the research of α-glucosidase inhibitors for development of future treatments for diabetes. Despite this characterization of FpAG1 and FpAG2, more work is required to fully understand the

distinguishing substrate preference of these GH31 α-glucosidases and their role in *F. prausnitzii* metabolism.

6.2 Future Directions

1. Co-crystallizing FpAG1 & FpAG2 with ligands can provide more information on the key residues involved in substrate specificity of GH31 enzymes. Investigating the binding of ligands with α-1,6 glycosidic linkages, such as isomaltose and palatinose, can provide more insight on the key residues involved in α-1,6 hydrolyzing activity and *F. prausnitzii* α-glucosidase function. By mutating the catalytic aspartic acids to prevent a catalytic reaction, the binding mechanism of substrates such as maltose, isomaltose and palatinose can be analyzed in these GH31 enzymes, an approach used to co-crystallized a Ro- αG1 mutant with isomaltose[64].
2. There is a difference in conserved domains between the *F. prausnitzii* α-glucosidases, in particular the presence of the Domain of unknown function, DUF5110, in FpAG1 that is not present in FpAG2. Due to the high similarity of the protein backbone and key residues involved in hydrolytic activity, these proteins may be a good model to investigate the purpose of this domain. This domain has been predicted to be involved in carbohydrate binding, so studies on the enzymatic activity of FpAG1 & FpAG2 proteins that incorporate or eliminate this domain, may elucidate its role on enzymatic function. As there was a difference observed in the stability of these recombinant proteins, and the presence of this conserved domain was the greatest global structural difference between these proteins, the role of DUF5110 on protein stability can also be investigated.

References

1. Guinane, C. M. & Cotter, P. D. Role of the gut microbiota in health and chronic gastrointestinal disease: Understanding a hidden metabolic organ. *Therap. Adv. Gastroenterol.* **6**, 295–308 (2013).

2. Sekirov, I., Russell, S. L., Caetano M Antunes, L. & Finlay, B. B. Gut microbiota in health and disease. *Physiol. Rev.* **90**, 859–904 (2010).

3. Chase, D., Goulder, A., Zenhausern, F., Monk, B. & Herbst-Kralovetz, M. The vaginal and gastrointestinal microbiomes in gynecologic cancers: A review of applications in etiology, symptoms and treatment. *Gynecol. Oncol.* **138**, 190–200 (2015).

4. Mailing, L. J., Allen, J. M., Buford, T. W., Fields, C. J. & Woods, J. A. Exercise and the Gut Microbiome: A Review of the Evidence, Potential Mechanisms, and Implications for Human Health. *Exerc. Sport Sci. Rev.* **47**, 75–85 (2019).

5. Singh, R. K. *et al.* Influence of diet on the gut microbiome and implications for human health. *J. Transl. Med.* **15**, 1–17 (2017).

6. Madison, A. & Kiecolt-Glaser, J. K. Stress, depression, diet, and the gut microbiota: human– bacteria interactions at the core of psychoneuroimmunology and nutrition. *Curr. Opin. Behav. Sci.* **28**, 105–110 (2019).

7. Yoon, K. & Kim, N. Roles of sex hormones and gender in the Gut Microbiota. *J. Neurogastroenterol. Motil.* **27**, 314–325 (2021).

8. Oliphant, K. & Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome: Major fermentation by-products and their impact on host health. *Microbiome* **7**, 1–15 (2019).

9. Belkaid, Y. & Hand, T. Role of the Microbiota in Immunity and inflammation. *Cell* **157**, 121–141 (2014).

10. Li, S. *et al.* Gut microbiome and osteoporosis. *Bone Jt. Res.* **9**, 524–530 (2020).

11. Stanford, J., Charlton, K., Stefoska-Needham, A., Ibrahim, R. & Lambert, K. The gut

microbiota profile of adults with kidney disease and kidney stones: A systematic review of the literature. *BMC Nephrol.* **21**, (2020).

12. Sasaki, H. *et al.* Distinctive subpopulations of the intestinal microbiota are present in women with unexplained chronic anovulation. *Reprod. Biomed. Online* **38**, 570–578 (2019).

13. Zheng, P., Li, Z. & Zhou, Z. Gut microbiome in type 1 diabetes: A comprehensive review. *Diabetes. Metab. Res. Rev.* **34**, 1–9 (2018).

14. Clapp, M. *et al.* Gut Microbiota's Effect on Mental Health: The Gut-Brain Axis. *Clin. Pract.* **7**, 131–136 (2017).

15. Butler, M. I., Mörkl, S., Sandhu, K. V., Cryan, J. F. & Dinan, T. G. The Gut Microbiome and Mental Health: What Should We Tell Our Patients?: Le microbiote Intestinal et la Santé Mentale : que Devrions-Nous dire à nos Patients? *Can. J. Psychiatry* **64**, 747–760 (2019).

16. Nichols, B. L. *et al.* Mucosal maltase-glucoamylase plays a crucial role in starch digestion and prandial glucose homeostasis of mice 1-3. *J. Nutr.* **139**, 684–690 (2009).

17. Van Beers, E. H., Büller, H. A., Grand, R. J., Einerhand, A. W. C. & Dekker, J. Intestinal brush border glycohydrolases: Structure, function, and development. *Crit. Rev. Biochem. Mol. Biol.* **30**, 197–262 (1995).

18. Hamaker, B. R., Lee, B. H. & Quezada-Calvillo, R. Starch digestion and patients with congenital sucrase-isomaltase deficiency. *J. Pediatr. Gastroenterol. Nutr.* **55**, 24–28 (2012).

19. Gray, G. M., Lally, B. C. & Conklin, K. A. Action of intestinal sucrase-isomaltase and its free monomers on an alpha-limit dextrin. *J. Biol. Chem.* **254**, 6038–6043 (1979).

20. Sim, L. *et al.* Structural Basis for Substrate Selectivity in Human Maltase-Glucoamylase and Sucrase-Isomaltase. *J. Biol. Chem.* **285**, 17763–17770 (2010).

21. Zhang, Y. *et al.* Impacts of gut bacteria on human health and diseases. *Int. J. Mol. Sci.* **16**, 7493–7519 (2015).

22. McNeil, N. I. The contribution of the large intestine to energy supplies in man. *Am. J. Clin.*

Nutr. **39**, 338–342 (1984).

23. Leylabadlo, H. E. *et al.* The critical role of Faecalibacterium prausnitzii in human health: An overview. *Microb. Pathog.* **149**, (2020).

24. Lopez-Siles, M., Duncan, S. H., Garcia-Gil, L. J. & Martinez-Medina, M. Faecalibacterium prausnitzii: From microbiology to diagnostics and prognostics. *ISME J.* **11**, 841–852 (2017).

25. Zou, Y. *et al.* Characterization and description of Faecalibacterium butyricigenerans sp. nov. and F. longum sp. nov., isolated from human faeces. *Sci. Rep.* **11**, 1–13 (2021).

26. Ahmed, S. *et al.* Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl. Environ. Microbiol.* **73**, 7435–7442 (2007).

27. Nadal, I., Donant, E., Ribes-Koninckx, C., Calabuig, M. & Sanz, Y. Imbalance in the composition of the duodenal microbiota of children with coeliac disease. *J. Med. Microbiol.* **56**, 1669–1674 (2007).

28. Duncan, S. H., Hold, G. L., Harmsen, H. J. M., Stewart, C. S. & Flint, H. J. Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as Faecalibacterium prausnitzii gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* **52**, 2141–2146 (2002).

29. Lopez-Siles, M. *et al.* Cultured representatives of two major phylogroups of human colonic Faecalibacterium prausnitzii can utilize pectin, uronic acids, and host-derived substrates for growth. *Appl. Environ. Microbiol.* **78**, 420–428 (2012).

30. Zheng, D., Liwinski, T. & Elinav, E. Interaction between microbiota and immunity in health and disease. *Cell Res.* **30**, 492–506 (2020).

31. Maslowski, K. M. *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**, 1282–1286 (2009).

32. Segain, J. P. *et al.* Butyrate inhibits inflammatory responses through NFκB inhibition: Implications for Crohn's disease. *Gut* **47**, 397–403 (2000).

33. Oliveira, C. B. *et al.* Gut microbiota of elite female football players is not altered during an

official international tournament. *Scand. J. Med. Sci. Sport.* 1–11 (2021). doi:10.1111/sms.14096

34. Murtaza, N. *et al.* The effects of dietary pattern during intensified training on stool microbiota of elite race walkers. *Nutrients* **11**, 1–14 (2019).

35. Miquel, S. *et al.* Faecalibacterium prausnitzii and human intestinal health. *Curr. Opin. Microbiol.* **16**, 255–261 (2013).

36. Cao, Y., Shen, J. & Ran, Z. H. Association between faecalibacterium prausnitzii reduction and inflammatory bowel disease: A meta-analysis and systematic review of the literature. *Gastroenterol. Res. Pract.* **2014**, (2014).

37. Sokol, H. *et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 16731–16736 (2008).

38. Yeoh, Y. K. *et al.* Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut* **70**, 698–706 (2021).

39. Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**, D571–D577 (2022).

40. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316 (1991).

41. Henrissat, B. & Bairoch, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781–788 (1993).

42. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 (1995).

43. Terwisscha van Scheltinga, A. C. *et al.* Stereochemistry of Chitin Hydrolysis by a Plant Chitinase/Lysozyme and X-ray Structure of a Complex with Allosamidin: Evidence for Substrate Assisted Catalysis. *Biochemistry* **34**, 15619–15623 (1995).

44. Rajan, S. S. *et al.* Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD+/Mn2+-dependent phospho-α-glucosidase from Bacillus subtilis.

Structure **12**, 1619–1629 (2004).

45. Liu, Q. P. *et al.* Bacterial glycosidases for the production of universal red blood cells. *Nat. Biotechnol.* **25**, 454–464 (2007).

46. Quaroni, A. & Semenza, G. Partial amino acid sequences around the essential carboxylate in the active sites of the intestinal sucrase.isomaltase complex. *J. Biol. Chem.* **251**, 3250–3253 (1976).

47. Lovering, A. L., Seung, S. L., Kim, Y. W., Withers, S. G. & Strynadka, N. C. J. Mechanistic and structural analysis of a family 31 α-glycosidase and its glycosyl-enzyme intermediate. *J. Biol. Chem.* **280**, 2105–2115 (2005).

48. Okuyama, M. *et al.* Carboxyl group of residue Asp647 as possible proton donor in catalytic reaction of α-glucosidase from Schizosaccharomyces pombe. *Eur. J. Biochem.* **268**, 2270–2280 (2001).

49. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **50**, D20–D26 (2022).

50. Marchler-Bauer, A. & Bryant, S. H. CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res.* **32**, 327–331 (2004).

51. Lu, S. *et al.* CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* **48**, D265–D268 (2020).

52. Pearson, W. R. An Introduction to Sequence Similarity ('Homology') Searching. *Curr. Protoc. Bioinforma.* 1–9 (2013). doi:10.1002/0471250953.bi0301s42.An

53. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).

54. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

55. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).

56. Wang, S., Ma, J., Peng, J. & Xu, J. Protein structure alignment beyond spatial proximity. *Sci. Rep.* **3**, (2013).

57. Wang, S., Peng, J. & Xu, J. Alignment of distantly related protein structures: Algorithm, bound and implications to homology modeling. *Bioinformatics* **27**, 2537–2545 (2011).

58. Maiorov, V. N. & Crippen, G. M. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology* **235**, 625–634 (1994).

59. Zemla, A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res.* **31**, 3370–3374 (2003).

60. Zhang, Y. & Skolnick, J. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).

61. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins Struct. Funct. Genet.* **57**, 702–710 (2004).

62. Xu, J. & Zhang, Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895 (2010).

63. Sim, L., Quezada-calvillo, R., Sterchi, E. E., Nichols, B. L. & Rose, D. R. Human Intestinal Maltase – Glucoamylase : Crystal Structure of the N-Terminal Catalytic Subunit and Basis of Inhibition and Substrate Specificity. *J. Microbiol.* **375**, 782–792 (2008).

64. Tan, K. *et al.* Novel α-glucosidase from human gut microbiome: Substrate specificities and their switch. *FASEB J.* **24**, 3939–3949 (2010).

65. Ernst, H. A. *et al.* Structure of the Sulfolobus solfataricus α-Glucosidase: Implications for Domain Conservation and Substrate Recognition in GH31. *J. Mol. Biol.* **358**, 1106–1124 (2006).

66. Davies, G. J., Wilson, K. S. & Henrissat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases [1]. *Biochem. J.* **321**, 557–559 (1997).

67. Nichols, B. L. *et al.* The maltase-glucoamylase gene: Common ancestry to sucrase-isomaltase with complementary starch digestion activities. *Proc. Natl. Acad. Sci. U. S. A.*

100, 1432–1437 (2003).

68. Huang, F. & Nau, W. M. A conformational flexibility scale for amino acids in peptides. *Angew. Chemie - Int. Ed.* **42**, 2269–2272 (2003).

69. Mhashal, A. R. *et al.* Modeling the Role of a Flexible Loop and Active Site Side Chains in Hydride Transfer Catalyzed by Glycerol-3-phosphate Dehydrogenase. *ACS Catal.* **10**, 11253–11267 (2020).

70. Subramani, A. & Floudas, C. A. Structure Prediction of Loops with Fixed and Flexible Stems. *J. Phys. Chem.* **116**, 6670–6682 (2012).

71. Chothia, C. & Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823–826 (1986).

72. Kufareva, I. & Abagyan, R. Methods of protein structure comparison. *Methods Mol. Biol.* **857**, 231–257 (2012).

73. Nagano, N., Orengo, C. A. & Thornton, J. M. One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.* **321**, 741–765 (2002).

74. Mauro, V. P. & Chappell, S. A. A critical analysis of codon optimization in human therapeutics Optimizing codon usage for increased protein expression. *Trends Mol. Med.* **20**, 604–613 (2014).

75. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986).

76. Bornhorst, B. J. A. & Falke, J. J. Purification of Proteins Using Polyhistidine Affinity Tags. *Protein Expr. Purif.* **326**, 245–254 (2000).

77. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* 571–607 doi:10.1385/1592598900

78. Cold Spring Harbor Protocols, . LB (Luria-Bertani) liquid medium. *Cold Spring 597 Harb Protoc* **1**, pdb.rec8141. (2006).

79. Komura, R., Aoki, W., Motone, K., Satomura, A. & Ueda, M. High-throughput evaluation of T7 promoter variants using biased randomization and DNA barcoding. *PLoS One* **13**, 1–16 (2018).

80. Rong, M., Biao, H., Mcallister, W. T. & Durbin, R. K. Promoter specificity determinants of T7 RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 515–519 (1998).

81. Stadler, M. & Fire, A. Wobble base-pairing slows in vivo translation elongation in metazoans. *Rna* **17**, 2063–2073 (2011).

82. Robinson, P. K. Enzymes: principles and biotechnological applications. *Essays Biochem.* **59**, 1–41 (2015).

83. Lan, H., Liu, H., Ye, Y. & Yin, Z. The Role of Surface Properties on Protein Aggregation Behavior in Aqueous Solution of Different pH Values. *AAPS PharmSciTech* **21**, 1–13 (2020).

84. Zhou, K., Zou, R., Stephanopoulos, G. & Too, H. P. Enhancing solubility of deoxyxylulose phosphate pathway enzymes for microbial isoprenoid production. *Microb. Cell Fact.* **11**, 1–8 (2012).

85. Lepock, J. R. *et al.* Influence of Transition Rates and Scan Rate on Kinetic Simulations of Differential Scanning Calorimetry Profiles of Reversible and Irreversible Protein Denaturation. *Biochemistry* **31**, 12706–12712 (1992).

86. Privalov, P. L. Cold denaturation of protein. *Crit. Rev. Biochem. Mol. Biol.* **25**, 281–306 (1990).

87. Sanfelice, D. & Temussi, P. A. Cold denaturation as a tool to measure protein stability. *Biophys. Chem.* **208**, 4–8 (2016).

88. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495 (2014).

89. Qi, X. & Tester, R. F. Lactose, Maltose, and Sucrose in Health and Disease. *Mol. Nutr. Food Res.* **64**, 1–9 (2020).

90. Pan, S. *et al.* Maltooligosaccharide-forming amylase: Characteristics, preparation, and application. *Biotechnol. Adv.* **35**, 619–632 (2017).

91. Hyeon, C. L., Jin, H. K., Sang, Y. K. & Jung, K. L. Isomaltose production by modification of the fructose-binding site on the basis of the predicted structure of sucrose isomerase from 'Protaminobacter rubrum'. *Appl. Environ. Microbiol.* **74**, 5183–5194 (2008).

92. Sugisawa, A. H. & Edo, H. The Thermal Degradation of Sugars I. Thermal Polymerization of Glucose. *J. Food Sci.* **31**, 1365–2621 (1966).

93. Peyrot des Gachons, C. & Breslin, P. A. S. Salivary Amylase: Digestion and Metabolic Syndrome. *Curr. Diab. Rep.* **16**, (2016).

94. Holub, I. *et al.* Novel findings on the metabolic effects of the low glycaemic carbohydrate isomaltulose (Palatinose™). *Br. J. Nutr.* **103**, 1730–1737 (2010).

95. Conklin, K. A., Yamashiro, K. M. & Gray, G. M. Human intestinal sucrase isomaltase. Identification of free sucrase and isomaltase and cleavage of the hybrid into active distinct subunits. *J. Biol. Chem.* **250**, 5735–5741 (1975).

96. Miyazaki, T., Ishizaki, Y., Ichikawa, M., Nishikawa, A. & Tonozuka, T. Structural and biochemical characterization of novel bacterial α-galactosidases belonging to glycoside hydrolase family 31. *Biochem. J.* **469**, 145–158 (2015).

97. Tanaka, S., Shinoki, A. & Hara, H. Melibiose, a Nondigestible Disaccharide, Promotes Absorption of Quercetin Glycosides in Rat Small Intestine. *J. Agric. Food Chem.* **64**, 9335–9341 (2016).

98. Cheng, K. C., Demirci, A. & Catchmark, J. M. Pullulan: Biosynthesis, production, and applications. *Appl. Microbiol. Biotechnol.* **92**, 29–44 (2011).

99. Singh, R. S., Kaur, N. & Kennedy, J. F. Pullulan production from agro-industrial waste and its applications in food industry: A review. *Carbohydr. Polym.* **217**, 46–57 (2019).

100. Johnson, K. & Goody, R. The Original Michaelis Constant. *Biochemistry* **50**, 8264–8269 (2012).

101. Michaelis, L. & Menten, M. L. Die Kinetik der Invertinwirkung. *Biochem. Z.* **49**, 333–369

(1913).

102. Johnson, K. A. New standards for collecting and fitting steady state kinetic data. *Beilstein J. Org. Chem.* **15**, 16–29 (2019).

103. Kaiser, P. M. Substrate inhibition as a problem of non-linear steady state kinetics with monomeric enzymes. *J. Mol. Catal.* **8**, 431–442 (1980).

104. Reed, M. C., Lieb, A. & Nijhout, H. F. The biological significance of substrate inhibition: A mechanism with diverse functions. *BioEssays* **32**, 422–429 (2010).

105. Pearson, J. *et al.* Pulmonary artery and intestinal temperatures during heat stress and cooling. *Med. Sci. Sports Exerc.* **44**, 857–862 (2012).

106. de Burlet, G. & Sudaka, P. Propriétés catalytiques de l'α-glucosidase neutre du rein humain. *Biochimie* **59**, 7–14 (1977).

107. Carchon, H. & De Bruyne, C. K. Purification and properties of Coffee-Bean a-D-Galactosidase. *Carbohydr. Res.* **41**, 175–189 (1975).

108. Marcos, E., Crehuet, R. & Bahar, I. Changes in dynamics upon oligomerization regulate substrate binding and allostery in amino acid kinase family members. *PLoS Comput. Biol.* **7**, (2011).

109. McDonald, A. G. & Tipton, K. F. Parameter Reliability and Understanding Enzyme Function. *Molecules* **27**, 1–18 (2022).

110. Malabanan, M. M., Amyes, T. L. & Richard, J. P. A Role for Flexible Loops in Enzyme Catalysis. *Curr. Opin. Struct. Biol.* **20**, 702–710 (2010).

111. Dirir, A. M., Daou, M., Yousef, A. F. & Yousef, L. F. *A review of alpha-glucosidase inhibitors from plants as potential candidates for the treatment of type-2 diabetes*. *Phytochemistry Reviews* **1**, (Springer Netherlands, 2021).

112. Xue, Y. P., Qin, J. W., Wang, Y. J., Wang, Y. S. & Zheng, Y. G. Enhanced production of acarbose and concurrently reduced formation of impurity C by addition of validamine in fermentation of Actinoplanes utahensis ZJB-08196. *Biomed Res. Int.* **2013**, (2013).

113. Calder, P. C. & Geddes, R. Acarbose is a competitive inhibitor of mammalian lysosomal acid α-d-glucosidases. *Carbohydr. Res.* **191**, 71–78 (1989).

114. Brayer, G. D. *et al.* Subsite mapping of the human pancreatic α-amylase active site through structural, kinetic, and mutagenesis techniques. *Biochemistry* **39**, 4778–4791 (2000).

115. Standl, E. & Schnell, O. Alpha-glucosidase inhibitors 2012-cardiovascular considerations and trial evaluation. *Diabetes Vasc. Dis. Res.* **9**, 163–169 (2012).

116. Samulitis BK, Goda T, Lee SM, K. O. Inhibitory mechanism of acarbose and 1-deoxynojirimycin derivatives on carbohydrases in rat small intestine. *Drugs Exp. Clin. Res.* **13**, 517–24 (1987).

117. Onose, S. *et al.* Production of the α-glycosidase inhibitor 1-deoxynojirimycin from Bacillus species. *Food Chem.* **138**, 516–523 (2013).

118. Goda, T. *et al.* Effects of miglitol, an α-glucosidase inhibitor, on glycaemic status and histopathological changes in islets in non-obese, non-insulin-dependent diabetic Goto-Kakizaki rats. *Br. J. Nutr.* **98**, 702–710 (2007).

119. Jayakanthan, K., Mohan, S. & Pinto, B. M. Structure Proof and Synthesis of Kotalanol and De- O -sulfonated Kotalanol , Glycosidase Inhibitors Isolated from an Herbal Remedy for the Treatment of Type-2 Diabetes. 5621–5626 (2009).

120. Sim, L. *et al.* New Glucosidase Inhibitors from an Ayurvedic Herbal Treatment for Type 2 Diabetes : Structures and Inhibition of Human Intestinal Maltase-Glucoamylase with Compounds from Salacia reticulata †. 443–451 (2010). doi:10.1021/bi9016457

121. Rossi, E. J. *et al.* Inhibition of recombinant human maltase glucoamylase by salacinol and derivatives. **273**, 2673–2683 (2006).

122. Nasi, R., Sim, L., Rose, R. & Pinto, B. M. Synthesis and glycosidase inhibitory activities of chain-modified analogues of the glycosidase inhibitors salacinol and blintol. *Carbohydr. Res.* **342**, 1888–1894 (2007).

123. Yoshikawa, M., Murakami, T., Tashiro, K. & Matsuda, H. Kotalanol, a Potent α-Glucosidase

Inhibitor with Thiosugar Sulfonium Sulfate Structure, from Antidiabetic Ayurvedic Medicine Salacia reticulata. *Chem. Pharm. Bull.* **46**, 1339–1340 (1998).

124. Muraoka, O. *et al.* Synthesis of a nitrogen analogue of salacinol and its α-glucosidase inhibitory activity. *Chem. Pharm. Bull.* **49**, 1503–1505 (2001).

125. Ghavami, A., Johnston, B. D. & Pinto, B. M. A new class of glycosidase inhibitor: Synthesis of salacinol and its stereoisomerst. *J. Org. Chem.* **66**, 2312–2317 (2001).

126. Yoshikawa, M. *et al.* Salacinol, potent antidiabetic principle with unique thiosugar sulfonium sulfate structure from the Ayurvedic traditional medicine Salacia reticulata in Sri Lanka and India. *Tetrahedron Lett.* **38**, 8367–8370 (1997).

127. National Center for Biotechnology Information, . PubChem Compound Summary for CID 41774. *Precose.* (2022). Available at: https://pubchem.ncbi.nlm.nih.gov/compound/Precose. (Accessed: 20th February 2022)

128. De Lean, A., Munson, P. J. & Rodbard, D. Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. *Am. J. Physiol. Endocrinol. Metab. Gastrointest. Physiol.* **4**, (1978).

129. Fomenko, I., Durst, M. & Balaban, D. Robust regression for high throughput drug screening. *Comput. Methods Programs Biomed.* **82**, 31–37 (2006).

130. Burlingham, B. T. & Widlanski, T. S. An intuitive look at the relationship of Ki and IC50: A more general use for the dixon plot. *J. Chem. Educ.* **80**, 214–218 (2003).

131. Yung-Chi, C. & Prusoff, W. H. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108 (1973).

132. Weimer, M. *et al.* The impact of data transformations on concentration-response modeling. *Toxicol. Lett.* **213**, 292–298 (2012).

133. Shoichet, B. K. Interpreting steep dose-response curves in early inhibitor discovery. *J. Med. Chem.* **49**, 7274–7277 (2006).

134. Prinz, H. Hill coefficients, dose-response curves and allosteric mechanisms. *J. Chem. Biol.*

3, 37–44 (2010).

135. Cleland, W. W. The kinetics of enzyme-catalyzed reactions with two or more substrates or products: I. Nomenclature and rate equations. *Biochim. Biophys. Acta* **67**, 104–137 (1963).

136. Cer, R. Z., Mudunuri, U., Stephens, R. & Lebeda, F. J. IC 50 -to- K i : a web-based tool for converting IC 50 to K i values for inhibitors of enzyme activity and ligand binding. **37**, 441–445 (2009).

137. Copeland, R. A., Lombardo, D., Giannaras, J. & Decicco, C. P. Estimating Ki Values for Tight-Binding Inhibitors from Dose-Response Plots. *Bioorg. Med. Chem.* **5**, 1947–1952 (1995).

138. He, H. & Lu, Y. H. Comparison of inhibitory activities and mechanisms of five mulberry plant bioactive components against α-glucosidase. *J. Agric. Food Chem.* **61**, 8110–8119 (2013).

139. Bischoff H. Pharmacology of a-glucosidase inhibition. *Eur. J. Clin. Invest.* **24**, 3–10 (1994).

140. Tan, K., Tesar, C., Wilton, R., Jedrzejczak, R. P. & Joachimiak, A. Interaction of antidiabetic a -glucosidase inhibitors and gut bacteria a -glucosidase. **27**, 1498–1508 (2018).

141. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 386–394 (2011).

142. Jiang, L. & Lai, L. CH···O hydrogen bonds at protein-protein interfaces. *J. Biol. Chem.* **277**, 37732–37740 (2002).

143. Sigma-Aldrich Co. *Product Information Glucose (GO) Assay Kit*. **1**, (2019).

144. Woolridge, E., Turchi, S. L. & Edwards, J. R. Experimental Section The Peroxidase-Glucose Oxidase Enzyme System in the Undergraduate Laboratory. *Biochem. Educ.* **14**, 82–83 (1986).

145. Kapustka, L. A., Annala, A. E. & Swanson, W. C. The peroxidase-glucose oxidase system: A new method to determine glucose liberated by carbohydrate degrading soil enzymes.

Plant Soil **63**, 487–490 (1981).

146. Megazyme. *Validation Report : D-Glucose Assay Kit (GOPOD Format) (cat.no. K-GLUC)*. (2021).

147. Megazyme. *Gopod Reagent Enzymes Assay Procedure (Gopod-Format)*. (2019).

Appendix A– BLAST CD-Search Results

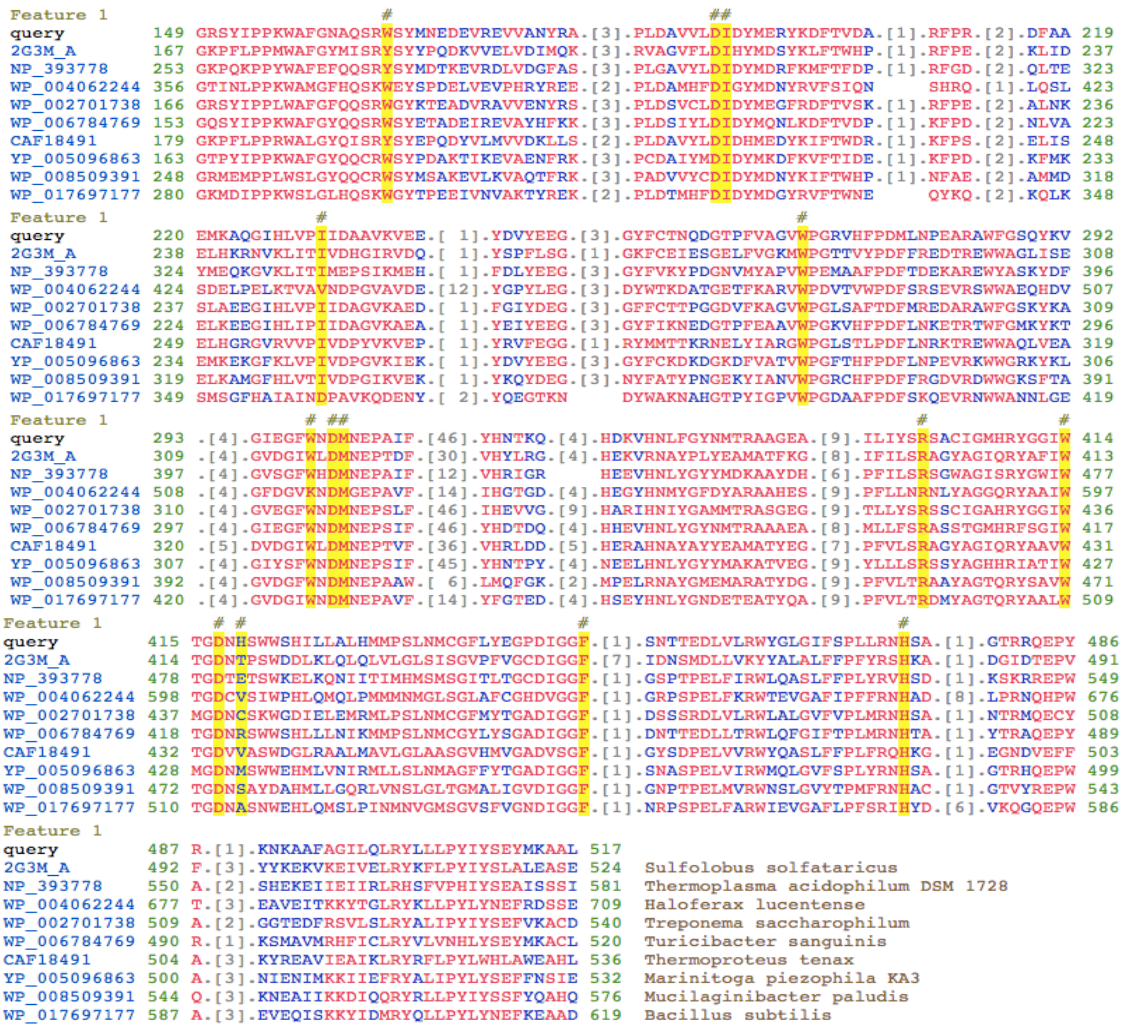


Figure 28. BLAST CD-Search Results on FpAG1 Sequence Catalytic Domain

The *F. prausnitzii* sequences were inputted into BLAST CD-Search and 86 hits were found in the CDD for the FpAG1 catalytic domain (residues 149-517), corresponding with the residues in the $(\beta/\alpha)_8$ barrel fold. The bit score is 525.54 and an E value of 0e+0 was determined for the GH31_glucosidase_II_MalA domain (Alphα-glucosidase II-like). The conserved residues identified in the active site are W166, D194, I195, I231, W268, W302, M305, R401, W414, H419, F450 and H475. The catalytic residues were identified to be D304 and D417. The top 10 listed sequences are included with a 2.0 colour bit threshold (with red colour indicating high conservation at that threshold).

Figure 29. BLAST-CD Results on FpAG1 Sequence N-terminal Domain

In BLAST CD-Search, 587 hits were found in the CDD for the FpAG1 sequence N-terminal domain (residues 37-149). The bit score is 105.35 and an E value of 8.20e-27 was determined for the GH31_N (N-terminal domain of glycosyl hydrolase family 31 (GH31)). D70 is the conserved residue identified to interact with the active site. The top 10 listed sequences are included with a 2.0 colour bit threshold (with red colour indicating high conservation at that threshold).

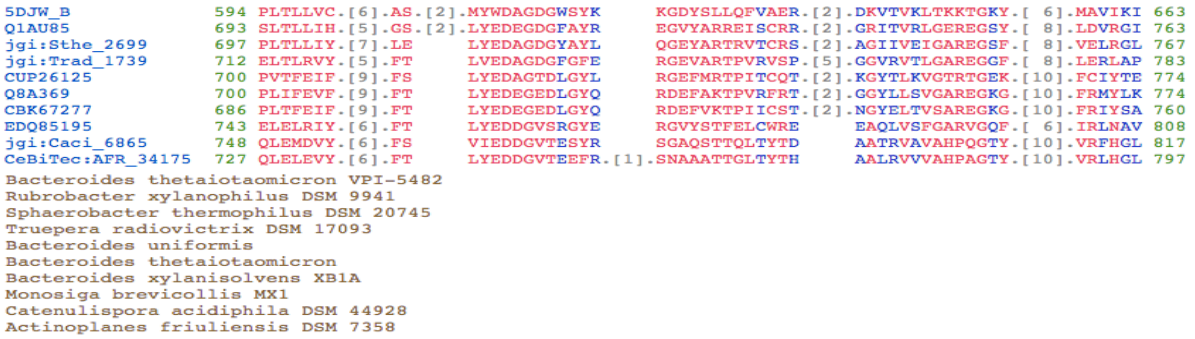


Figure 30. BLAST-CD Results on FpAG1 C-terminal domain

In BLAST CD-Search, 321 hits were found in the CDD for the FpAG1 sequence C-terminal domain (residues 632-675). The bit score is 39.91 and an E value of 2.60e-04 was determined for the DUF5110 (Domain of unknown function (DUF5110)). There were no identified conserved residues. The top 10 listed sequences are included with a 2.0 colour bit threshold (with red colour indicating high conservation at that threshold).

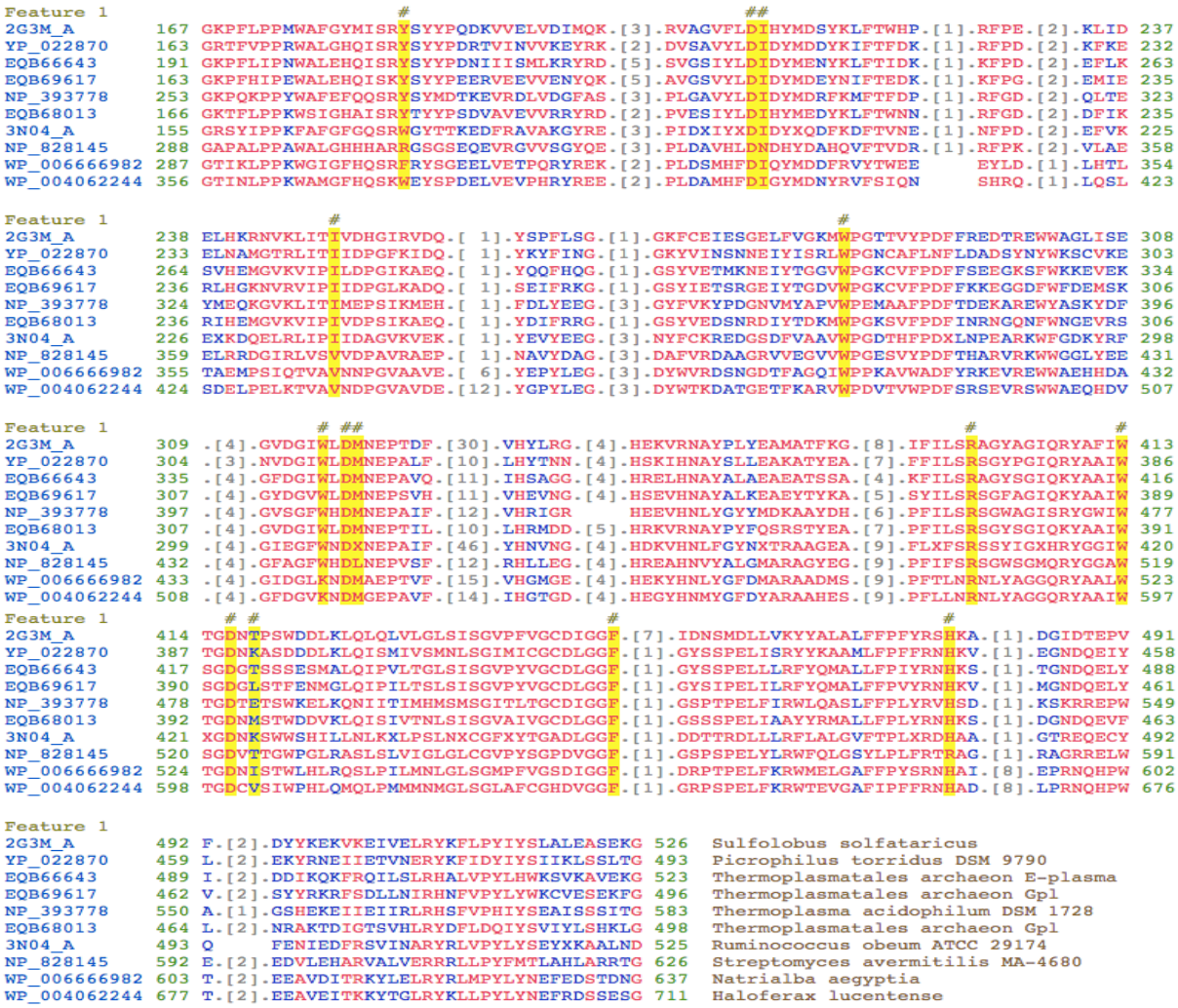


Figure 31. BLAST-CD Results on FpAG2 Sequence Domain

In BLAST CD-Search and 86 hits were found in the CDD for the FpAG2 sequence catalytic domain (residues 150-522), corresponding with the residues in the $(\beta/\alpha)_8$ barrel fold. The bit score is 523.61 and an E value of 0e+00 was determined for the GH31_glucosidase_II_MalA domain (Alphα-glucosidase II-like). There is also a gap in this domain in FpAG2 sequence, between residues 319-354. The conserved sites identified in the active site are W167, D195, I196, I232, W269, W303, D305, M306, R404, W417, D420, N422, F453 and H478. The catalytic residues were identified to be D305 and D420. The top 10 listed sequences are included with a 2.0 colour bit threshold (with red colour indicating high conservation at that threshold).

Figure 32. BLAST-CD Results on FpAG2 Sequence N-terminal Domain

In BLAST CD-Search, 587 hits were found in the CDD for the FpAG2 sequence N-terminal domain (residues 38-150). The bit score is 88.40 and an E value of 6.38e-21 was determined for the GH31_N (N-terminal domain of glycosyl hydrolase family 31 (GH31)). D70 is the conserved residue identified to interact with the active site. The top 10 listed sequences are included with a 2.0 colour bit threshold (with red colour indicating high conservation at that threshold).

Appendix B– GAGO Assay Modified Protocol

Glucose (GO) Assay Kit (GAGO) (Sigma-Aldrich GAGO20) can be used to measure the enzymatic activity of the expressed proteins, by measuring the amount of free glucose produced in the assay[143]. Glucose oxidase, peroxidase and *o*-Dianisidine are added to a solution containing the inactivated reaction solution. The glucose oxidase will react with any free glucose and O_2, and produce hydrogen peroxide[144]. In the presence of peroxidase, the hydrogen peroxide reacts with o-dianisidine through a reduction reaction. This causes the o-Dianisidine to change colour to brown, which could be measured at an optimal wavelength of 450nm[144]. The amount of oxidized o-dianisidine is a direct measurement of the amount of glucose present in the assay volume[144]. The absorbance of reduced o-Dianisidine will be converted to a concentration of glucose, based on a standard curve[143,145,146].

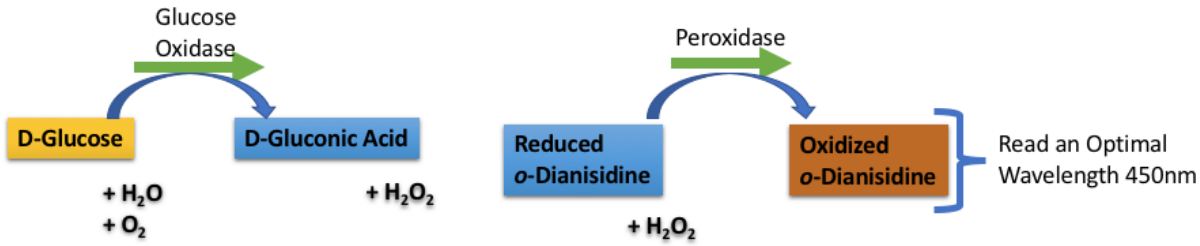


Figure 33. Visual representation of GAGO Assay reaction

Ratios of reagent solutions relative to glucose standard and sample volume were taken from the protocol provided by the manufacturer, and adjusted to be read in a 96-well plate. To create a standard curve, 5 μl of glucose solution was added to 96-well non-binding plate, with concentrations ranging from 0.1mg/mL to 1.0 mg/mL diluted in 25mM MES buffer, 100 mM NaCl, pH 5.8, and 20 μl volume of Milli-Q water, as well as 100 μl of GAGO reagent was added to the well with a repeater pipette. When testing samples, the 5 μl glucose solution was replaced with sample volume. The plate was immediately incubated at 37°C in the spectrometer. Endpoint was read at 30 minutes at 450nm, as this is the optimal wavelength of reduced o-Dianisidine. Samples that measured an absorbance outside of the standard curve range were diluted and re-tested to ensure that the absorbance of the reaction was within a detectable range.

As per the manufacturer's protocol, an equivalent volume (125 μl) of 6M H_2SO_4 was added to terminate the assay reagent reaction and stabilize the developed colour of the reagent reaction after the 30 minute incubation at 37°C. The absorbance was read at a wavelength of 540nm.

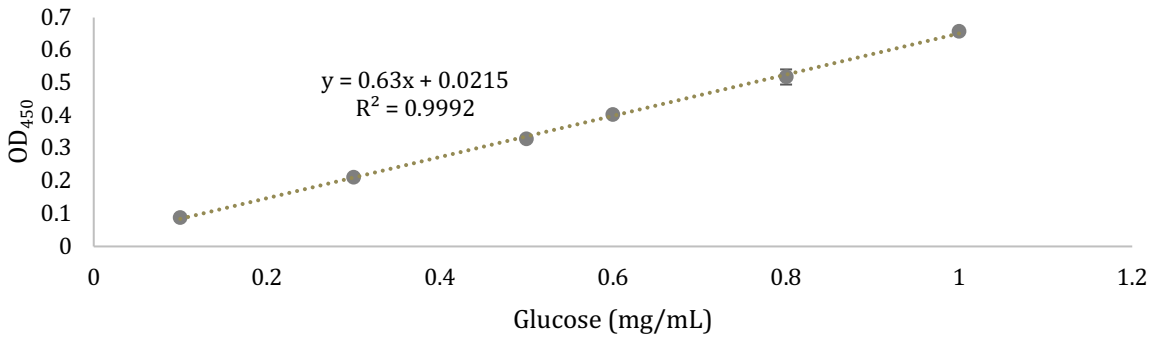


Figure 34. GAGO Assay Linear Relationship @ 450nm

This figure shows the relationship between the amount of glucose reacting with the GAGO assay reagent and the absorbance measured at a wavelength of 450 nm. Five replicates of the standard curve were done to validate the linear relationship.

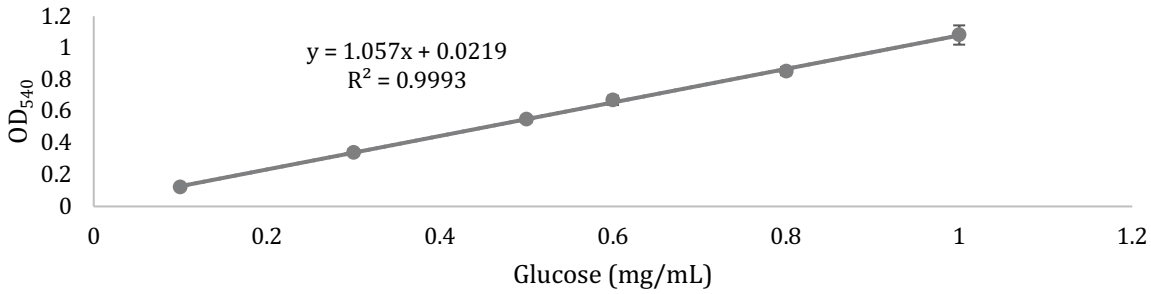


Figure 35. GAGO Assay Linear Relationship @ 540nm

This figure shows the relationship between the amount of glucose reacting with the GAGO assay reagent and the absorbance measured at a wavelength of 540 nm. The linear regression analysis of this relationship can be found in Table 9.

The GAGO assay was adapted from validated protocols to the 96-well format to test the enzymatic activity of the *Faecalibacterium prausnitzii* proteins. The adaptations were implemented to increase the number of samples that could be measured with the purchased assay reagents and to save time executing the experiment, as smaller volumes are easier to work with and measure in a 96-well plate. This setup allowed the plate to be shaken every 2 minutes so that the sample and assay reagent are distributed uniformly. Conducting the measurements in a 96-well plate also allowed the development of the colour as a response to the presence of glucose to be monitored, which ensured that the assay reaction was developing consistently.

Two versions of the GAGO assay standard curve were tested to confirm a linear relationship. At the end of the 30 minute incubation in the spectrophotometer, an absorbance

reading was immediately taken at 450nm. The absorbance of the standard curve was also measured at a wavelength of 540nm after the addition of 6M H_2S0_4. For the glucose standard curve created using the GAGO assay reagent at a wavelength of 540nm, a statistical analysis was done to confirm whether the relationship between the amount of glucose present and the detected absorbance was linear. The adapted assay reagent protocols showed that there was a linear relationship between the absorbance and glucose concentration.

Without the addition of 6M H_2S0_4, there are fewer factors that could introduce error to the measurement. Due to the greater precision of data points observed in the measurements prior to the addition of 6M H_2S0_4, the values measured at 450nm were used in this analysis. With the quick dispersion of the assay reagent through the use of the repeater pipette, the measurements being taken at the same time and the reaction progression being monitored to ensure that the assay reagents are developing consistently, the error associated with taking the measurement prior to the addition of H_2S0_4 is not expecting to be significant.

Table 9. Linear Regression Analysis of GAGO Standard Curve Adapted to 96-Well Plate

Linear Regression					
Data source: Data 1 in Notebook1					
Col 7 = 0.0219 + (1.057 * Col 1)					
N = 6					
R = 1.000	Rsqr = 0.999	Adj Rsqr = 0.999			
Standard Error of Estimate = 0.010					
	Coefficient	Std. Error	t	P	
Constant	0.0219	0.0085	2.575	0.062	
Col 1	1.057	0.0136	77.834	<0.001	
Analysis of Variance:					
	DF	SS	MS	F	P
Regression	1	0.598	0.598	6058.177	<0.001
Residual	4	0.000395	9.87E-05		
Total	5	0.598	0.12		
Normality Test (Shapiro-Wilk)	Passed	(P = 0.862)			
Constant Variance Test:	Passed	(P = 0.060)			
Power of performed test with alpha = 0.050: 1.000					

This table shows the linear regression of the glucose curve measured with the GAGO assay reagent. The data was fit with a linear regression using SigmaPlot version 7.0, from Systat Software, Inc., San Jose California USA, www.systatsoftware.com. The R square of the GAGO assay standard curve was 0.999, and this linear relationship passed the Normality Test (Shapiro-Wilk) with a p value of 0.862 and the constant variance test with a p value of 0.060.

Appendix C– GOPOD Assay Modified Protocol

The GOPOD reagent assay is a colourimetric assay that measures the amount of glucose present in the assay well[146]. In this assay, the glucose-oxidase oxidizes glucose and this results in the production of hydrogen peroxide. The hydrogen peroxide is then used to combine the p-hydroxybenzoic acid and 4-aminoantipyrine, to create quinoneimine dye, through a redox reaction facilitated by peroxidase[147]. The reagent enzymes are optimally active at a temperature between 40-50°C and the reaction requires 20 minutes for completion[147]. The colour formed by the creation of the quinoneinine dye is stable for 2 hours after colour development[147]. The absorbance of reduced quinoneimine dye will be converted to a concentration of glucose, based on a standard curve[143,145,146].

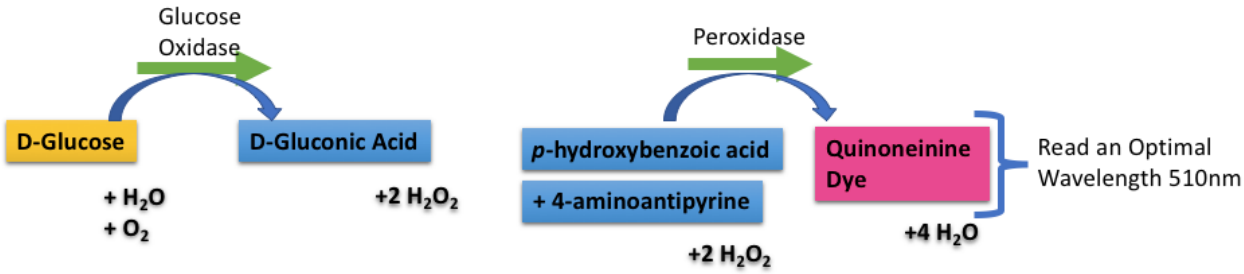


Figure 36. Visual representation of Megazyme Assay reaction

GOPOD assay was adjusted to 96-well plate. Ratios of assay reagent relative to glucose standard and sample volume were taken from protocol provided by manufacturer, and adjusted to a final volume of 250 μL, to be read in a 96-well plate[147]. To create a standard curve, 10 μl of glucose solution were added to 96-well non-binding plate, with concentrations ranging from 0.1mg/mL to 1.0 mg/mL diluted in 25mM MES buffer, 100 mM NaCl, pH 5.8 to be consistent with the sample buffer. To test samples, a 10 μL volume was added to the 96-well plate. A volume of 240ul of GOPOD reagent was added to the well with a repeater pipette to keep volumes dispensed uniform and to ensure that all reactions started simultaneously. The plate was immediately incubated at 45C in the spectrometer. Readings were taken every 2 minutes and endpoint readings were taken after the absorbance had plateaued. This incubation time consistently fell in line with the manufacturer provided incubation times of 20 minutes at 510nm[147]. Samples that measured an absorbance outside of the standard curve range were diluted and re-tested to ensure that the absorbance of the reaction was within a detectable range.

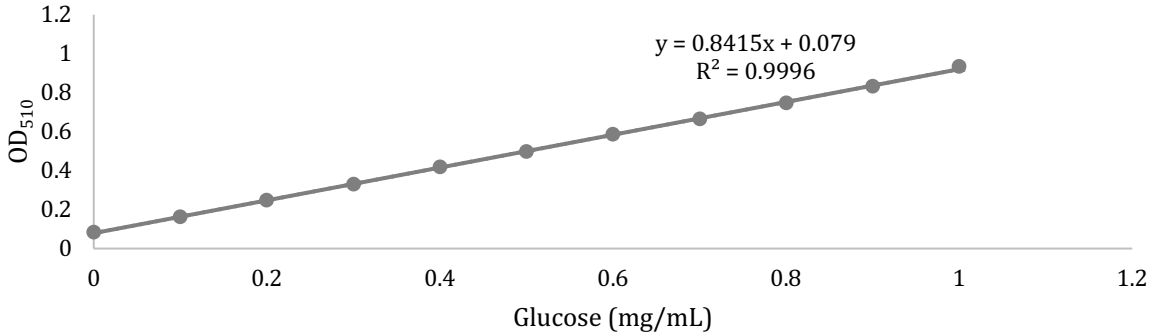


Figure 37. Megazyme Assay Linear Relationship

This figure shows the relationship between the amount of glucose present in the Megazyme assay well and the absorbance measured at a wavelength of 510 nm. The linear regression analysis of this relationship can be found in Table 10. Glucose standard curve was done in duplicate for the linear relationship validation.

Table 10. Linear Regression Analysis of GOPOD Standard Curve

Linear Regression					
Col 2 = 0.0790 + (0.841 * Col 1)					
N = 11					
R = 1.000	Rsqr = 1.000	Adj Rsqr = 1.000			
Standard Error of Estimate = 0.006					
	Coefficient	Std. Error	t	P	
Constant	0.079	0.00322	24.519	<0.001	
Col 1	0.841	0.00544	154.591	<0.001	
Analysis of Variance:					
	DF	SS	MS	F	P
Regression	1	0.779	0.779	23898.52	<0.001
Residual	9	0.000293	3.26E-05		
Total	10	0.779	0.0779		
Normality Test (Shapiro-Wilk)	Passed	(P = 0.103)			
Constant Variance Test:	Passed	(P = 0.082)			
Power of performed test with alpha = 0.050: 1.000					

This table shows the linear regression of the glucose curve measured with the GOPOD assay reagent. The data was fit with a linear regression using SigmaPlot version 7.0, from Systat Software, Inc., San Jose California USA, www.systatsoftware.com. The R square of the GAGO assay standard curve was 1.000, and this linear relationship passed the Normality Test (Shapiro-Wilk) with a p value of 0.103 and the constant variance test with a p value of 0.082.

Appendix D– Statistical Analysis of Enzymatic Kinetic Models

Table 11. Non-Linear Regression Fit Comparison of Substrate Inhibition and Michaelis-Menten Kinetic Models

	Michaelis-Menten				Substrate Inhibition			
Enzyme/ Substrate	**Degrees of Freedom**	**R square**	**Sum of Squares**	**Sy.x**	**Degrees of Freedom**	**R square**	**Sum of Squares**	**Sy.x**
FpAG1/ Maltose	25	0.9689	5.885	0.485	24	0.9720	5.292	0.470
FpAG1/ Isomaltose	22	0.9569	2.308	0.324	21	0.9569	2.308	0.331
FpAG1/ Palatinose	25	0.8841	10.66	0.653	24	0.9422	5.312	0.470
FpAG2/ Maltose	24	0.9418	4.921	0.453	23	0.9696	2.575	0.335
FpAG2/ Isomaltose	22	0.9944	0.07437	0.058	21	0.9944	0.07437	0.059
FpAG2/ Palatinose	22	0.8791	0.4653	0.145	21	0.9746	0.09787	0.068

A Non-Linear Regression, Michaelis Menten, Substrate Inhibition, was performed using GraphPad Prism version 9.0 for Mac, GraphPad Software, La Jolla California USA, www.graphpad.com.

Table 12. Comparison of Non-Linear Regression Fit of Kinetic Models

Enzyme/ Substrate	**Model with Higher Probability**	**P Value**	**F (DFn, DFd)**	**R^2**	**95% CI V_{max}**	**95% CI K_M**
FpAG1/ Maltose	Michaelis-Menten	0.1138	2.693 (1, 24)	0.9689	[8.903, 9.696]	[39.63, 57.42]
FpAG1/ Isomaltose	Michaelis-Menten	>0.9999	$1.83e^{-14}$ (1.21)	0.9569	[8.080, 8.558]	[6.855, 9.441]
FpAG1/ Palatinose	Substrate Inhibtion	<0.0001	24.17 (1, 24)	0.9422	[8.162, 12.45]	[7.935, 18.07]
FpAG2/ Maltose	Substrate Inhibtion	0.0001	20.96 (1, 23)	0.9696	[7.258, 13.66]	[96.99, 266.3]
FpAG2/ Isomaltose	Michaelis-Menten	N/A	N/A	0.9944	[2.860, 3.061]	[53.54, 64.78]
FpAG2/ Palatinose	Substrate Inhibtion	<0.0001	78.83 (1, 21)	0.9746	[2.574, 5.345]	[28.01, 77.10]

A Non-Linear Regression, Michaelis Menten, Substrate Inhibition, Extra sum-of-squares F test with a P value less than 0.05, was performed using GraphPad Prism version 9.0 for Mac, GraphPad Software, La Jolla California USA, www.graphpad.com. The Null Hypothesis is Michaelis Menten and Alternative Hypothesis is Substrate Inhibition.

Appendix E- Calculated K_i values of FpAG1 and FpAG2 Inhibition Assays

Table 13. Calculated K_i values of FpAG1 Inhibition Assays

	Classical Inhibition – K_i (μM)			Tight-Binding Inhibition – K_i (μM)		
	Comp	Uncomp	Noncomp	Comp	Uncomp	Noncomp
Acarbose	473.54 +/- 25.42	595.14 +/- 31.94	1068.69 +/- 57.36	473.67 +/- 25.56	595.32 +/- 32.12	1069.00 +/- 57.68
Miglitol	7.365 +/- 0.16	9.26 +/- 0.21	16.62 +/- 0.37	7.51 +/- 0.31	9.44 +/- 0.38	16.94 +/- 0.69
Kotalanol	7.54 +/- 0.11	9.48 +/- 0.13	17.02 +/- 0.24	7.68 +/- 0.25	9.65 +/- 0.31	17.34 +/- 0.56

The enzyme concentration (0.638 μM), substrate concentration (60 mM), apparent K_M (47.4 mM) and apparent IC_{50} were used to calculate the Ki of the inhibitor. Both the classical inhibition and tight-binding inhibition equations were calculated and the Ki used in the analysis is highlighted.

Table 14. Calculated K_i values of FpAG2 Inhibition Assays

	Classical Inhibition – K_i (μM)			Tight-Binding Inhibition – K_i (μM)		
	Comp	Uncomp	Noncomp	Comp	Uncomp	Noncomp
Acarbose	420.33 +/- 32.11	165.48 +/- 12.64	585.81 +/- 44.75	420.56 +/- 32.34	165.58 +/- 12.73	586.14 +/- 45.08
Miglitol	0.84 +/- (-0.20)	0.33 +/- (-0.08)	1.16 +/- (-0.28)	1.07 +/- 0.03	0.42 +/- 0.01	1.49 +/- 0.04
Kotalanol	2.20 +/- (-0.15)	0.87 +/- (-0.06)	3.06 +/- (-0.21)	2.43 +/- 0.09	0.96 +/- 0.03	3.39 +/- 0.12

The enzyme concentration (0.653 μM), substrate concentration (60 mM), apparent K_M (152.4 mM) and apparent IC_{50} were used to calculate the K_i of the inhibitor. Both the classical inhibition and tight-binding inhibition equations were calculated and the K_i used in the analysis is highlighted.

The K_i values for classical competitive inhibition and tight-binding competitive inhibition were calculated. With this web tool, when the values for classical and tight-binding inhibitors are similar, you can assume that the binding kinetics can be interpreted as classical kinetics[136]. If the K_i value determined with the tight-binding equation was similar to the K_i determined with the classical inhibition equation, it means that the amount of the enzyme used was not significant enough and it cancelled out of the equation. The difference between the K_i determined with the classical equation and tight-binding equation was less than 2% for all interactions except miglitol/FpAG2 which had a 27% difference and kotalanol/FpAG2 which had a 10% difference. In these enzyme/inhibitor interactions, the tight-binding inhibition equation will be used because it cannot be assumed that the free inhibitor and total inhibitor concentrations are the same in this system. The K_i of kotalanol with FpAG2 using the tight-binding inhibitor equation is 1.07 +/- 0.03 μM. The tight-binding equation was also used to calculate the K_i of kotalanol with FpAG2, which had a resulting K_i of 2.43 +/- 0.09 μM.

Appendix F– Predicted Residues Involved in the Binding of Kotalanol in the -1 subsite

Table 15. Residues Involved in the Binding of Kotalanol in the -1 subsite

	ntSI	ntMGAM	FpAG1	FpAG2
1	D472	D443	D304	D305
2	D571	D542	D417	D420
3	H629	H600	H475	H478
4	D355	D327	D194	D195
5	R555	R526	R401	R404
6	D231	D203	D70	D70
7	W435	W406	W268	W269
8	F479	F450	F311	F312

These are the residues identified in the binding of kotalanol in the -1 subsite of ntMGAM and ntSI[20]. The corresponding residues in FpAG1 and FpAG2 are listed in Table 15.