Rational Engineering of a Loop Region in the *Saccharomyces cerevisiae* β-fructofuranosidase

by

Jerard Gibbon

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in the Faculty of Science at Stellenbosch University

Supervisor: Dr Heinrich Volschenk

Co-supervisor: Dr Kim Trollope

April 2019

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signed: Jerard Gibbon
20 January 2019

SUMMARY

The increased need for chemical processes to become more environmentally friendly and the pressure for these processes to be more efficient have led to the chemical industry looking to biocatalysis. Biocatalysts are organic catalysts that increase the rate of the reaction by lowering the activation energy of the chemical reaction. Although biocatalysts are highly specific and can produce enantiomerically pure products compared to inorganic catalysts, they cannot produce the yields seen with inorganic catalysts and are unable to function under the harsher conditions under which inorganic catalysts operate. With the advent of molecular techniques, these biocatalysts can be tailored to function under harsher conditions associated with the chemical industry. The engineering of enzymes has been a hit-and-miss activity as the interaction between amino acids as well as the folding of a protein is still not well understood. This has led to the search for techniques to understand the interactions of amino acids better and identify the structural features that confer activity, specificity and stability.

In the GH 32 enzyme family, several studies have aimed at understanding the transfructosylating activity or hydrolytic activity present in these enzymes. Fructooligosaccharides (FOS) consist of fructose bound to sucrose as acceptor molecule by either a β-(2-6)-glycosidic bond or a β-(2-1)-glycosidic bond. The fructose can either fructosylate the glucose or the fructose of the sucrose forming neokestose for the former and 1-kestose or 6-kestose for the latter. The FOS can be polymerised with the addition of one fructosyl unit which increases the degree of polymerisation (DP). In this study higher DP FOS is classified as FOS with a DP of 4 and higher. It was previously identified that a TI loop region within the GH 32 enzymes had the potential to limit the formation of higher DP FOS. This study aimed to evaluate the effect the TI loop region had on the DP of FOS within the *Saccharomyces cerevisiae* β-fructofuranosidase (Suc2) making use of rational enzyme engineering.

Employing four rational engineering substitutions, three of the variants were able to produce higher DP FOS. Interestingly, these three variants produced 1^F-fructofuranosyl nystose, an ability which has not been documented in the literature for the *S. cerevisiae* enzyme. The variants were able to produce multiple isomers of higher DP FOS based on their HPLC retention time. Unfortunately, due to a lack of available standards, these could not be identified. The TI loop was able to change the activity of Suc2 as the variants were predominantly transfructosylating enzymes, moving away from the predominantly hydrolytic

activity of Suc2. This was confirmed as the variants increased their total FOS production by more than fourfold.

Additionally, the changes to the TI loop changed the regiospecificity of the variant Suc2 enzyme producing ten times more neokestose than Suc2. The latter indicates that the TI loop contributes to regulating the orientation of the acceptor molecule, as previously described in the literature. Lastly, the changes to the TI loop led to a change in the topology of the catalytic pockets of the variants compared to Suc2, with the variants' catalytic pockets resembling enzymes with high transfructosylating activity.

This functional knowledge of the TI loop in *S. cerevisiae*'s β-fructofuranosidase gained in this study through mutational analysis contributed to a new understanding of how this loop governs the hydrolytic or transfructosylating activity of β-fructofuranosidase enzymes.

OPSOMMING

Die toenemende behoefte aan chemiese prosesse om meer omgewingsvriendelik te word asook die druk vir hierdie prosesse om doeltreffender te wees, het daartoe gelei dat die chemiese industrie belangstelling toon in bio-gekataliseerde prosesse. Biokatalisators is organiese katalisators wat die reaksie tempo verhoog deur die aktiveringsenergie van die chemiese reaksie te verlaag. Biokatalisators is dus aanloklike alternatiewe tot anorganiese katalisators weens hulle hoë ensiemspesifisiteit en vermoë om suiwer enantiomeriese produkte te vorm. Nietemin, produseer biokatalisators laer opbrengste as anorganiese katalisators en is nie in staat om onder die meerderheid van industriële toestande waaraan anorganiese katalisators blootgestel word te funksioneer nie. Met die aankoms van molekulêre tegnieke kan hierdie biokatalisators aangepas word om te funksioneer onder moeiliker toestande wat met die chemiese industrie geassosieer word. Die verandering van ensieme is 'n lukrake aktiwiteit aangesien die interaksie tussen aminosure sowel as die vou van proteïene nog steeds nie goed verstaan word nie. Dit het gelei tot die soeke na tegnieke om die interaksies van aminosure beter te verstaan en die strukturele eienskappe wat aktiwiteit, spesifisiteit en stabiliteit verleen, te identifiseer.

In die GH 32 ensiem familie het verskeie studies daarop gefokus om die transfruktosilerende aktiwiteit of hidrolitiese aktiwiteit wat in hierdie ensieme voorkom, beter te verstaan. Frukto-oligosakkariede (FOS) is fruktose wat gebind word aan 'n akseptor molekule, sukrose, deur óf ’n β-(2-6)-glikosidiese binding of ’n β-(2-1)-glikosidiese binding. Die fruktose kan aan glukose of fruktose in die sukrose bind wat lei tot neokestose in die geval van glucose en 1/6-kestose vir fruktose. Die FOS kan gepolimeriseer word met toevoeging van een fruktosiel-eenheid wat die graad van polimerisasie (DP) verhoog. In hierdie studie word hoër DP van FOS geklassifiseer as FOS met 'n DP van 4 en hoër. Daar is voorheen 'n lus binne die GH 32-ensieme geïdentifiseer met die potensiaal om die vorming van hoë DP in FOS te bepaal. Die doel van hierdie studie was om die effek wat die TI-lus op die bepaling van DP van FOS het binne die *Saccharomyces cerevisiae* β-fruktofuranosidase (Suc2) te bepaal, deur gebruik te maak van ’n rasionele mutasie benadering.

Deur gebruik te maak van vier rasionele benaderings om Suc2 verander, kon drie van die variante hoër DP FOS produseer. Hierdie drie variante het 1^F-fruktofuranosiel nystose geproduseer, wat nie in die literatuur voorheen vir Suc2 gedokumenteer was nie. Die variante kon verskeie isomere van hoër DP FOS op grond van die HPLC retensietyd produseer. Ongelukkig kon dit nie weens gebrek aan beskikbare standaarde geïdentifiseer word nie. Die

TI-lus was in staat om die aktiwiteit van Suc2 te verander, van ’n hoofsaaklik hidrolitiese ensiem na ’n transfruktosilerende ensiem. Dit is bevestig omdat die variante hul totale FOS-produksie met meer as 4 keer verhoog het.

Daarbenewens het die veranderinge aan die TI-lus die regiospesifisiteit van die Suc2-ensiem variante verander, wat daartoe gely het dat 10 keer meer neokestose produseer as Suc2. Dit dui aan dat die TI-lus bydra tot die regulering van die oriëntering van die akseptor molekule, soos voorheen in die literatuur beskryf. Laastens het die veranderinge aan die TI-lus gelei tot 'n verandering in die topologie van die katalitiese holtes van die variante in vergelyking met Suc2, met die variante se katalitiese holtes wat meer ooreenstem met ensieme met ’n hoë transfruktosilerende aktiwiteit.

Die funksionele kennis van die TI-lus in *S. cerevisiae* se β-fruktofuranosidase wat in hierdie studie deur middel van mutasie analise verkry is het bygedra het tot 'n nuwe begrip van hoe hierdie lus die hidrolitiese of transfruktosilerende aktiwiteit in β-fruktofuranosidase-ensieme beheer.

PREFACE

This dissertation is presented as a compilation of 4 chapters. Each chapter of this thesis is introduced separately and has been written according to the instructions of the BIOCHIMICA ET BIOPHYSICA ACTA-PROTEINS AND PROTEOMICS.

Chapter 1 provides a general introduction and delineation of the project aims. The literature review (Chapter 2) provides a broad overview of the enzyme engineering and the techniques involved, the current status and future perspectives of FOS production as well as β-fructofuranosidases with a focus on previous research conducted on these enzymes. The results chapter (Chapter 3) entitled “The rational redesign of a loop region of *Saccharomyces cerevisiae* β-fructofuranosidase” is written as a journal manuscript and has been compiled in that format. Chapter 4 provides an overview of the main findings and conclusions of this study.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude and appreciation to the following persons and institutions for their invaluable contributions to the successful completion of this study:

Dr Kim Trollope, Department of Microbiology, Stellenbosch University, who acted as my co-promoter. I want to thank you, Kim, for being an ear to all my crazy ideas as well as having valued discussion about the project or any academic topic. Furthermore, I want to thank her for helping me through the process and teaching me some valuable research lessons.

Dr Heinrich Volschenk, Department of Microbiology, Stellenbosch University, who acted as my promoter, I want to thank him for all the opportunities that he has given me during my time in his lab as well as sending me on development courses.

Colleagues and the staff of the Department of Microbiology, Stellenbosch University, I want to thank everyone that I crossed paths with for their advice and assistance.

My family, thank you for your patience and support

The **National Research Foundation (NRF)** of South Africa for financial support and funding of the research.

CONTENTS

CHAPTER 1

GENERAL INTRODUCTION AND PROJECT AIMS

Chapter 1: General Introduction and Project Aims

1.1 Introduction

The adoption of biocatalysts has been growing in popularity in the chemical industry as these catalysts are often more efficient, produce less waste and use fewer resources than traditional inorganic catalysts [1]. The biocatalyst industry has experienced significant advances with the advent of molecular techniques which drastically changed the enzyme engineering landscape. The implementation of rational, directed evolution and semi-rational engineering approaches to improve or tailor enzymes have seen increased success rates with a shift in focus to enzymes relevant to industry.

β-fructofuranosidases are a class of enzymes that have been classified by the Carbohydrate-Active enZYmes (CAZy) database in family GH 32 and GH 64 [2]. This enzyme family catalyses hydrolysis and transfructosylation of the β-(2-6)-glycosidic bonds or β-(2-1)-glycosidic bonds between glucose and fructose within sucrose, fructans and fructooligosaccharides (FOS). Many of the enzymes within the GH 32 and 64 family have both activities but favour either the hydrolytic or the fructosyltransferase activity. The interest in these enzymes is due to their application in the sugar industry as they can hydrolyse sucrose into fructose and glucose. The second industrial application of these enzymes has been driven by the nutraceutical industry to produce FOS. The increasing demand for FOS has led to the search for new FOS-producing enzymes as well as engineering existing GH 32 family enzymes to modify catalytic efficiency or specificity. The engineering of the GH 32 enzymes has also led to research in the fundamental understanding of the mechanisms that govern predominantly hydrolytic or transferase activity in these enzymes which share similar sequences and structures. β-fructofuranosidases are promiscuous enzymes and can fructosylate many sugars and sugar-alcohols which make these enzymes attractive for use in the ever-growing nutraceutical market.

FOS is one or more fructose unit(s) bound to a sucrose by either β-(2-6)-glycosidic or β-(2-1)-glycosidic bonds, and the nature of these bonds are linked to the prebiotic properties of FOS [3]. As a prebiotic, FOS in layman's terms, act as food for probiotic bacteria and selectively aids the growth of beneficial bacteria in the gut of the host. There are three types of FOS, i.e., levan-type, inulin-type and neo-series FOS all of which are isomers of each other, with the only difference being the nature of the chemical bonds. The differences between the isomers especially the β-(2-6)-glycosidic or β-(2-1)-glycosidic bonds of FOS led to a higher degree of stimulation in the growth of beneficial bacteria in the gut [3].

While β-fructofuranosidase enzymes are some of the most studied enzymes and have been used to model classical enzyme kinetics [4], this enzyme group has also been extensively mutated to uncover the underlying catalytic mechanisms of these enzymes, specifically to understand the principles between hydrolysis and transfructosylating activity [5–8]. Yeast β-fructofuranosidases are predominantly levan-type and neo-series FOS producers which are regarded to have a higher prebiotic effect than the inulin-type [3]. Yeast β-fructofuranosidases are also predominantly hydrolytic enzymes producing a limited amount of FOS under high sucrose concentrations [6]. There have been multiple attempts to increase the synthesis of FOS by mutating yeast β-fructofuranosidases with varying success. Lafraya *et al.* [6] used rational engineering to mutate the *S. cerevisiae* invertase and increase the production of 6-kestose 10-fold. During this study, Lafraya *et al.* [6] uncovered that the Asn-228 residue contributed to the regiospecificity of the enzyme because a mutation to that position led to an equimolar production of 6-kestose and 1-kestose. Although the engineered *S. cerevisiae* enzyme significantly increased its FOS production, the increase still did not reach the yields of the fructosyltransferase enzyme of *Aspergillus japonicus* (AjFFase, also known as FopA). Trollope *et al.,* [9] investigated the sequence and structural differences between fungal GH 32 enzymes to develop a better understanding of the differences between predominantly hydrolytic and transferase enzymes. During the study, Trollope *et al.,* [9] made use of *in silico* techniques to identify two different motifs to separate fungal GH 32 enzymes with predominantly high hydrolytic and low fructosyltransferase activity from the enzymes with low hydrolytic and high fructosyltransferase activity.

Furthermore, Trollope *et al.,* [9] identified a TI loop region within β-fructofuranosidases that was longer in the high hydrolytic and low fructosyltransferase activity enzymes as opposed to a short loop in the low hydrolytic and high fructosyltransferase activity enzymes. It was hypothesised that this loop potentially plays a role in determining the hydrolytic and transferase activity of the GH 32 family enzymes. However, the hypothesis was made based on purely *in silico* observations with no experimental evidence to support the claims.

1.2 Aims of the Study

This study aimed to investigate the TI loop in the *S. cerevisiae* β-fructofuranosidase and determine its role in the degree of polymerisation and catalytic properties of the enzyme.

The specific objectives of this study were to:

1. Mutate the loop region to create space in the catalytic pocket by the following strategies:
   a) loop deletion (resulting in a minimal loop with β-turn inducing amino acids)
   b) loop substitution (replace Suc2 loop with AjFFase loop)
   c) W48A amino acid substitution (replace bulky amino acid side-chain in the loop)

2. Production and purification of the *S. cerevisiae* β-fructofuranosidase variants in a *SUC2* knockout strain of *S. cerevisiae*.

3. Characterise and compare the catalytic activity of the *S. cerevisiae* β-fructofuranosidase variants using enzymatic assays, differential scanning fluorimetry, structural modelling and molecular docking

1.3 References

[1] R.A. Sheldon, J.M. Woodley, Role of Biocatalysis in Sustainable Chemistry, Chem. Rev. (2017) acs.chemrev.7b00203. doi:10.1021/acs.chemrev.7b00203.

[2] B.I. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics, Nucleic Acids Res. 37 (2009) 233–238. doi:10.1093/nar/gkn663.

[3] S. Kilian, S. Kritzinger, C. Rycroft, G. Gibson, J. Preez, The effects of the novel bifidogenic trisaccharide, neokestose, on the human colonic microbiota, (2002) 637–644.

[4] L. Michaelis, M.L. Menten, Die Kinetik der Invertinwirkung, Biochem Z. 49 (1913) 333–369. doi:10.1021/bi201284u.

[5] M.A. Sainz-Polo, M. Ramírez-Escudero, A. Lafraya, B. Gonza, J. Marín-Navarro, J. Polaina, J. Sanz-Aparicio, Three-dimensional structure of *Saccharomyces* invertase: Role of a non-catalytic domain in oligomerization and substrate specificity, J. Biol. Chem. 288 (2013) 9755–9766. doi:10.1074/jbc.M112.446435.

[6] A. Lafraya, J. Sanz-Aparicio, J. Polaina, J. Marin-Navarro, Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase, Appl. Environ. Microbiol. 77 (2011) 6148–6157. doi:10.1128/AEM.05032-11.

[7] K.M. Trollope, H. Volschenk, J.F. Görgens, Bro R, H.H. Nieuwoudt, Direct, simultaneous quantification of fructooligosaccharides by FT-MIR ATR spectroscopy and chemometrics for rapid identification of superior, engineered β-fructofuranosidases, (2015). doi:10.1007/s00216-014-8417-0.

[8] M. Gimeno-Perez, P. Santos-Moriano, L. Fernandez-Arrojo, A. Poveda, J. Jimenez-Barbero, A.O. Ballesteros, M. Fernandez-Lobato, F.J. Plou, Regioselective synthesis of neo-erlose by the β-fructofuranosidase from *Xanthophyllomyces dendrorhous*, 2014. doi:10.1016/j.procbio.2013.12.018.

[9] K.M. Trollope, N. van Wyk, M.A. Kotjomela, H. Volschenk, Sequence and structure-based prediction of fructosyltransferase activity for functional sub-classification of fungal GH32 enzymes., FEBS J. (2015) 1–15. doi:10.1111/febs.13536.

CHAPTER 2

LITERATURE REVIEW

Exploring the potential of improving biocatalysts to meet future industrial needs

Chapter 2: Exploring the potential of improving biocatalysts to meet future industrial needs

2.1 Introduction – Catalysts are crucial to the chemical industry

Industrial catalysis has progressed from its grassroots in ancient times to today where the majority of commercial chemical reactions utilise some form of catalyst [1]. Catalysts provide several beneficial functions in industrial chemical processes. Catalysts reduce the activation energy of chemical reactions thus increasing the rate of the reactions. By increasing the rate of a reaction, the time to produce the product is substantially less - decreasing operational costs which in turn saves the manufacturer money.

Additionally, a decrease in the activation energy allows a chemical reaction to occur at lower temperatures. Lastly, catalysts can be reused which further increases the cost-effectiveness of industrial chemical conversions. The reusable nature of catalysts is attributed to their function of facilitating chemical reactions without partaking in these reactions - thus preventing their depletion. Although a catalyst cannot be depleted, the disadvantage of reusing a catalyst is a decrease in its effectiveness over time due to contamination of the catalyst's surface. The advantages of using catalysts in industry combined with the demands from governments to enforce legislation to reduce energy consumption and increase efficiency in chemical industries have led to the widely-adopted (90%) use of catalysts in all chemical processes [2].

2.2 Biomanufacturing and the advancement of biocatalysis

Although inorganic catalysts are widely used in the chemical industry and have helped the chemical industries become more efficient by decreasing their energy utilisation, there is still a need for catalysts to become more efficient in producing fewer by-products and produce less waste [2]. Advances in biocatalysts such as the ability to increase enzyme production through heterologous expression, tailoring of biocatalyst using enzyme engineering and an improved fundamental understanding of biocatalysis have led to increased adoption of biocatalyst usage in the chemical industry.

The progression of biocatalysis has been described by Bornscheuer *et al.,* [3] as the three waves of biocatalysis and by Zhang *et al.,* [4] as biomanufacturing 1.0 to 4.0. According to Zhang *et al.,* [4] the first pre-modern biomanufacturing catalysis involved the fermentation of food sources during the production of products such as bread, beer and wine [5]. Later, Louis Pasteur provided evidence that microorganisms catalyse these fermentations. In the process Pasteur of debunked the 2000-year-old Aristotelian doctrine of spontaneous generation,

leading the way to an explosion of discoveries and developments of biomanufactured products such as the manufacturing of cheese, vinegar and the production of indigo, to name a few [4,5]. Biomanufacturing in this era was often based on solid-state fermentation and usually mixed microorganism cultures [4].

The first biomanufacturing era, biomanufacturing 1.0, was characterised by the use of whole-cell biocatalysts producing primary metabolic products [4]. It is argued that this era started with Chaim Weizmann's acetone, butanol and ethanol (ABE) fermentation in 1915 [6]. The ABE fermentation was regarded as ground-breaking, firstly because it made use of a mono-culture microorganism as opposed to a mixed culture and, secondly, it was a large-scale anaerobic liquid fermentation [7]. The products of the ABE process found application in the making of synthetic rubber, cordite, and were used as fuel [4,7]. The biomanufacturing 2.0 era followed with a shift in focus to the production of secondary metabolites. These secondary metabolites, often produced by microorganisms to defend against other organisms, has found application as human medicine, food flavourings and fragrances [4]. The discovery of different antimicrobials such as sulphonamides, penicillin and streptomycin that coincided with World War II lead to a vast increase in the use of these drugs [4]. The increased demand for antimicrobials led to the development of processes which would increase product yields. These increases were achieved by changing parameters including the optimisation of submerged fermentations [4].

The 1980s was the start of the third biomanufacturing era (biomanufacturing 3.0), characterised by a move away from the conventional production of microbial primary or secondary metabolites towards the use of microorganisms for the production of proteins [4]. These proteins were used either as biocatalysts or in protein-based drugs [3]. This shift was driven by two scientific breakthroughs namely, the introduction of recombinant DNA technology and advances in cell cultures [3]. This era was dominated by the production of pharmaceutical protein-based drugs such as erythropoietin, insulin and human growth hormone [3,4] and the manufacturing of bulk enzymes for biocatalysis included amylases, lipases, glucose isomerases and glucoamylases [8]. The demand for industrial biocatalysts increased the need to produce bulk enzymes at lower prices, which in turn, led to significant developments in more cost-effective manufacturing techniques including high-cell density fermentations, strain selection (super producers) and the development of protein expression hosts [4].

Since the 2000s until the present day, the biomanufacturing 4.0 era has focused on addressing the needs for new products such as renewable energy, artificial food and regenerative medicine [4]. To achieve these goals, numerous tools have been developed such as induced

pluripotent stem cells, metabolic engineering, synthetic biology, systems biology and cascade biocatalysis [9]. Significant progress has been achieved through biocatalysis regarding improved product yields, enzyme titres, scale-up feasibility and sustainability [4]. Currently, research into advanced cell-free techniques such as the Synthetic Pathway Biotransformation (SyPaB) platform has the potential to change the future of biocatalysis. This technique has enormous potential to increase the efficiency of biocatalytic processes as most desired products only require the use of 20 or fewer enzymes, their complexes and coenzymes [10]. Contrary to whole cell biocatalysis, where production of enzymes involved in metabolism, cell division and production of organism-relevant products are essentially regarded as wasted energy, SyPaB makes use of *in vitro* synthetic biosystems which consider only the relevant biosynthetic pathway enzymes involved in producing the final desired end-products. By using this approach, researchers will be able to increase the production of desired products at a higher concentration as well as increasing the efficiency of these processes by converting the energy into the final product rather than biomass [10].

2.3 Biocatalysts – towards a future of greener chemistry

Biocatalysis has gained popularity in the past two decades with the advent of molecular biology techniques [11]. Although enzymes used as biocatalysts provide numerous advantages over inorganic catalysts, including catalytic efficiency, substrate specificity with the added benefit of being environmentally friendly. The knowledge gained in chemistry especially organic chemistry initially resulted in slow adoption of biocatalysis [5,12].

Although yields are important, the effect of these chemical manufacturing processes on the environment has forced industry since the 1980s to shift their focus and holistically assess the manufacturing process from optimal raw material use to decreasing or eliminating waste by avoiding the use or production of toxic and or hazardous chemicals [12,13]. Thus, there was a move away from conventional resource-to-waste processes towards systems where waste is not generated, or all waste products and catalysts are repurposed. With regards to waste repurposing, biocatalysts and catalysts will again play a critical part [12]. By the 1990s it was recognised that eliminating waste not only leads to a decrease in cost because there was no waste handling but also increases profitability as raw materials could be used to its maximum potential [12]. The latter led to the conceptualisation of “green chemistry” which is defined by the U.S. Environmental Protection Agency (EPA) “*as effectively utilising raw material, eliminates wastes and avoids the use of toxic and hazardous reagents and solvent in the manufacturing and application of chemical products*”. Anastas and Warner [14] list the 12 principles for green chemistry as follows: (1) waste prevention instead of remediation, (2) atom

efficiency, (3) less hazardous materials, (4) safer products by design, (5) innocuous solvents and auxiliaries, (6) energy efficient by design, (7) preferably renewable raw materials, (8) shorter synthesis (avoiding derivatisation), (9) catalytic rather than stoichiometric reagents, (10) design products for degradation, (11) analytical methodologies for pollution prevention and (12) inherently safer processes.

These principles were also complemented with increased awareness of sustainability, thus moving away from resource-to-waste to a more circular process [15]. Sustainable technology needs to consume natural resources at a rate that does not deplete the supply over the long term, and the residues generated should not be generated at a rate higher than can be assimilated by the natural environment [12]. The need for greener technologies in chemistry has shifted focus from yields towards producing more atom efficient processes with higher selectivity [12]. The use of both chemical and biocatalysts will be employed to meet these needs for the future; however, biocatalysis is currently better positioned for greener chemical industries as it adheres to all twelve of the “green chemistry” principles [12].

2.4 Changing biocatalysts to meet future needs

The advent of recombinant DNA technology led to the availability of methods to induce changes in genes and the ability to measure the results in real-time [16]. The time needed to effect a change in a biocatalyst and witness its result was significantly decreased. Additionally, recombinant DNA technology also contributed to our understanding of the function of a gene, and the enzyme produced [16]. Recombinant DNA technology challenged two prior views of using enzymes in biocatalysis, i.e., the belief that enzymes are very specific and changing enzymes are a lengthy process [17]. Initially, the specificity of an enzyme’s activity was understood and defined as a “one enzyme one reaction” paradigm, also known as the absolute specificity of an enzyme [17,18]. Similarly, it was widely understood that the changes in organisms to evolve and adapt to new environments were slow and time-consuming [17].

Over the past decades, our understanding of enzyme specificity regarding the lock-and-key and induced fit models has been expanded to the school of thought that enzymes are more flexible than what was initially thought. Enzymes are now considered more “promiscuous” in their ability to catalyse reactions from different substrates or display different activities beyond the lock-and-key and induced fit models of enzyme specificity [18]. Promiscuous enzyme activity has two meanings in the literature. Firstly, it could mean that enzymes are said to be indiscriminate, i.e., different substrates can be accepted and catalysed [19]. The second meaning implies that the secondary reactions catalysed by an enzyme are physiologically

irrelevant to the organism, also known as “moonlighting” [19]. Recently, it was noted that these secondary reactions are more common in enzymes than what was initially assumed. For example, Huang *et al.,* [20] screened 217 enzymes of the haloalkanoic acid dehalogenase family from 86 species against 169 phosphorylated compounds and showed that 204 of these enzymes could catalyse hydrolysis on a median of 15.5 substrates. It was further demonstrated that 101 of these enzymes hydrolysed between 6 and 40 substrates, while 50 enzymes in this group hydrolysed between 41 and 143 substrates [20]. This high level of promiscuity was a critical finding for both evolutionary science as well as enzyme engineering. This finding altered the view of enzymes from rigid entities to more flexible entities capable of catalysing different substrates to some degree and in turn, increased our understanding of enzymes [19].

Since Charles Darwin wrote the *Origin of Species* [21] scientists had been fascinated by the prospect of change due to environmental pressures. Although evolution is a slow process, rarely directly observed within macroscopic life [18], with the advent of molecular techniques and the use of microorganisms, it was possible to witness evolution within weeks and months rather than millennia [18]. The shortened timescale has led to an increased understanding of evolution, specifically from an enzymatic view where adaptations of an organism are mainly ascribed to changes in genes being expressed and the enzymes produced [18]. Recently, it was shown that enzyme promiscuity potentially evolved as a survival strategy, as promiscuous enzymes can catalyse a range of substrates at very low levels giving the organism the opportunity to survive [19]. The survival of the organism allows the organism enough time to induce changes in the gene, and these changes increase its activity and specificity for the new substrate [17]. Changes in enzymes which make them more efficient at catalysing a reaction are stably incorporated into the genome [22].

Tyzack *et al.,* [17] described the mechanisms of change within enzymes as either creeping or leaping evolution. Creeping evolution is defined as relatively minor changes over time, i.e., usually point mutations, which leads to small differences in substrate specificity. Leaping evolution, on the other hand, consists of more dramatic changes in the enzyme chemistry over time which influences the activity of the enzyme by either altering the functioning of the enzyme or accepting different substrates [17]. Creeping evolution can be thought of as changes to an enzymes EC number at the fourth digit, for example EC:1.4.3.4 (monoamine oxidase) to EC:1.4.3.2 (L-amino-acid oxidase) whereas leaping evolution results in a change in EC, for instance from EC:1.2.1.10 (acetaldehyde dehydrogenase) to EC:4.1.3.39 (4-hydroxy-2-oxopentanoate pyruvate lyase) [17]. The core structure of the enzyme is rarely changed and provides a robust scaffold for accommodating varying and diverse chemistry [23]. The changes to an enzyme to acquire new functions are based on modifications to

random coils within the protein structure, as these random coils seem to have the highest tendency to produce new functions [24]. The evolution of an enzyme is not a one generation phenomena as seen in nature and the laboratory since tolerance to mutations follows cycles of protein destabilisation and stabilisation [24]. Most mutations are neutral. However, some destabilising mutations will be tolerated since a threshold to the degree of destabilisation exists [24]. If a change is too disruptive, the protein will not be expressed or fail to fold correctly, but other co-occurring or subsequent mutations could compensate for the destabilisation effect and restabilise the enzyme [24]. The process of enzyme evolution is not a linear process and is typically characterised by periods of decreased enzyme fitness as it sacrifices stability for an increase in activity for a new substrate followed by a period of increased stability [25].

Understanding the mechanisms involved in enzyme evolution offer not only new knowledge of enzymes in general but also provide insights into the interactions of the different amino acids in an enzyme and their interactions with one another and how these interactions affect protein structure and function. Furthermore, the effect of mutations can be implemented in enzyme engineering approaches [24,26]. The fundamental knowledge that was presented above had a striking influence on how enzyme engineering is conducted, i.e., using the information gained through these enzyme structure-function experiments meant that this knowledge could be put to the test [27].

2.5 Improving and creating new biocatalysts - broadening the application and understanding of biocatalysis

Enzymes have naturally evolved over millions of years to be most efficient at catalysing biological relevant chemical reactions [19,28]. However, enzyme engineering is necessary due to growing industry demands which require enzymes with new properties such as tolerance to higher substrate loadings, higher activity at lower enzyme dosages, increase stability under high or low pH's and temperatures and the need for non-natural products, all at economically affordable rates [29,30]. Protein engineering aims to address these shortcomings in terms of increasing enzyme stability, substrate specificity and catalysis. An enzyme can be engineered for these new properties following three approaches illustrated in Fig. 2.1. Firstly, a rational design approach in which the changes made to the enzyme has a rationale behind them, i.e., previous experimental knowledge [31]. The second approach is based on directed evolution where evolution on a shortened time-scale is achieved making use of chemical- or error-prone DNA polymerase-derived point mutations as well as gene shuffling to induce modifications within the gene at random [32]. These two methods can also

be combined into a third approach, known as semi-rational design, in order to overcome shortcomings associated with each approach [32].

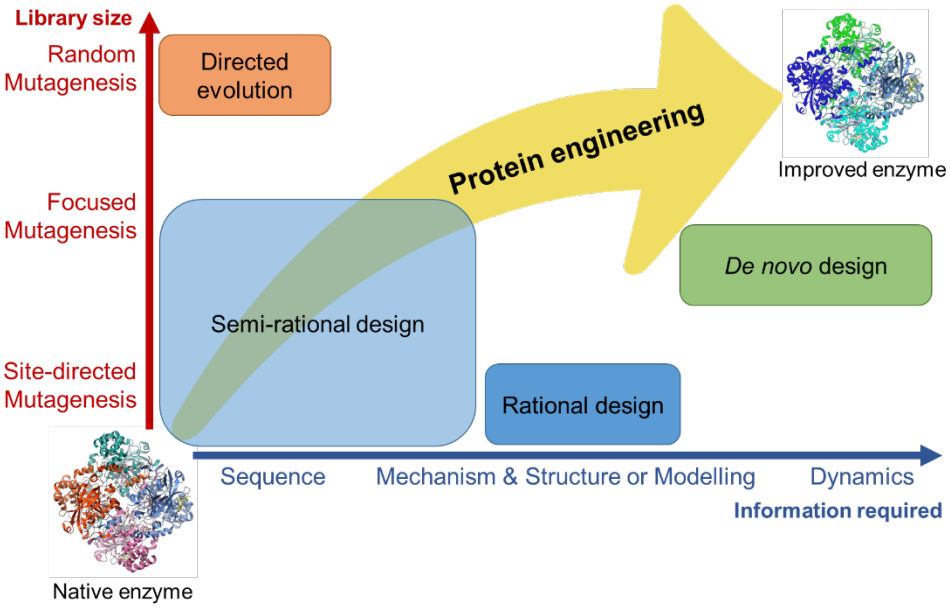

Figure 2.1. Comparison of common protein engineering methods used to improve enzyme properties. The information and library size required for four general methods are compared. Library size is indicated as mutation numbers. Information required was quantified as the complexity of information patterns including sequence, structure, and dynamics (adapted from [33]).

2.5.1 Rational enzyme engineering

One of the first approaches in enzyme engineering was rational enzyme engineering, so called due to the use of structural or sequence data to guide the changes within the enzyme [34]. This approach is also known as a data-driven approach as it relies on the use of functional (biochemical) data, structural data, sequence data or modelling data [32]. The in-depth knowledge of the structural features, i.e., the active site or the catalytic pocket, means that these landscapes can be targeted [35]. The residues in the active site and the surrounding area interacts directly with the substrate through substrate binding or participating in the catalytic process [32]. The residues can also indirectly affect the activity and selectivity as the residues between the direct interacting residues can influence the orientation of the side chains or the folding of the catalytic pocket [35]. Targeting these areas at or near the active site has an advantage as it usually dictates activity and selectivity. Furthermore, only focusing on residues in the active site and the surrounding area results in the creation of smaller variant libraries simplifying the downstream screening/selection procedures.

2.5.1.1 Sequence-based rational enzyme engineering

One of the first methods employed for rational engineering was the use of multiple sequence alignments to gain information regarding which amino acid positions should be targeted for mutagenesis [36]. Alignment of multiple protein sequences, combined with structure homolog modelling, can be used to calculate and identify consensus residues computationally [36]. This method can be used to identify conserved or highly conserved residues within a set of proteins that correlate with protein stability, reactivity or substrate selectivity [36]. Residues conserved between protein homologs are essential to enzyme stability and/or activity thus making them ideal targets for engineering [29]. This method has been successfully used in several industrial enzymes to tailor their thermostability although, it was noted by Ebert and Pelletier [29] that this technique has been undervalued as it can also change the efficiency and activity of the enzyme. Automated programs such as Hot-Spot Wizard 2.0 makes use of various algorithms and modelling programs to target functional and stability hot-spots within proteins, where one of these algorithms usually employs a multiple sequence alignment step [37]. The limitation of this method is that there is a vast amount of proteins sequences in nature which are similar in function, i.e., catalysing the same enzymatic reaction. As a result of their isolation, their sequences homology can be vastly divergent, i.e., low sequence homology/identity [38]. Highly divergent protein sequences can lead to incorrect enzyme classification leading to an oversight of potential information regarding an enzyme.

Xiao *et al.,* [39] improved the thermostability of *Xanthomonas campestris* pectate lyase making use of a multiple sequence alignment. By using ClustaW as well as four thermal stable pectate lyase sequences which they selected based on their reported range of temperatures of 65°C to 69°C, 70°C and 90°C [39]. There was a low sequence identity between the five sequences having an amino acid identity of between 25 to 36%, thus indicating low homology among these proteins [39]. The authors found only nine amino acid positions were unchanging in all four of the thermostable pectate lyases. The researchers then changed these residues and tested their contribution to thermostability or activity. The thermostability data indicated that one of the single mutations increased the T_m with 6°C and extended the enzyme's half-life at 45°C by 23-fold [39]. Three other mutations led to a decrease of between 18 to 53% in the thermostability of the enzyme, thus illustrating a total of four amino acid mutations that affected the thermostability of the enzyme [39]. One of the variants lost all activity while another variant increased the activity of the enzyme by 5-fold. The latter mutation was then combined with the best candidate amino acid change for thermostability resulting in a double mutant with a 5-fold increase in enzyme activity and a 12-fold increase in thermostability [39].

The example of Xiao *et al.,* [39] illustrates the potential of sequence-based approaches, i.e., combining sequence alignments with rational amino acid selection. It needs to be mentioned

that Xiao *et al.,* [39] only selected four target protein sequences out of thousands in the NCBI protein database, thus highlighting both the potential and the weakness of this technique. The weakness being that Xiao *et al.* [39] made the selection using their experience in the field to select potential candidates for mutagenesis, which will be difficult to reproduce in a purely objective manner or by an automated system. However, the multiple sequence alignment approach can be an invaluable tool in selecting potential targets for mutagenesis and is included in numerous pipelines of automated programs to identify potential targets for mutagenesis [37,39,40].

2.5.1.2 Structure-based rational enzyme engineering

Using the structural data of a protein is often more reliable to identify the functional enzyme homologs because the secondary and tertiary structure of a protein is more conserved than the primary structure. The reason for this is the limited number of natural protein folds, i.e., there can be multiple sequences that will produce the same fold which combines to form the protein structure [41]. This conservation in protein structures can be used to identify other structures that are similar in folding but is much more diverse in terms of the amino acid sequence [42]. As a result, it is possible to identify an unknown protein based on structural similarities to known proteins, as well as potentially identify the function and potential ligands of the unknown protein [42]. Enzyme function and ligand predictions are possible due to the structural folds being more conserved and by default increases the likelihood that known residues interacting with the substrate in the homolog are also interacting with the substrate of the unknown protein [42]. Important residues, i.e. residues that confer activity, specificity and ligand interaction which have been researched on the homologs can inform the researcher about the potential function, specificity and ligands of the unknown protein due to the similarity to the known homolog [43]. This functional information can aid in enzyme engineering to substitute or delete amino acids in an unknown homolog due to its similarity to a known and characterised enzyme.

The complex multidimensional features of a protein structure also have its drawbacks as it is more computationally demanding to analyse, i.e., it requires more computational power and more complex algorithms compared to one-dimensional multiple sequence alignments [29]. However, the increase in computational power of the general desktop computer and optimisations of software, allows these techniques to be used by non-expert users [29]. Structural data can be used in combination with sequence data to identify targets for engineering, i.e., a researcher can identify an area which is similar between two comparatively and functionally similar enzymes. This identified area can then be modified by using the amino

acids found in a sequence alignment from a similar family of enzymes [32]. Amino acids can be changed based on their reactivity to assess the potential role of a residue in the protein. For example, by replacing a tryptophan residue, which is known to be reactive and contribute to substrate interaction, with an alanine that is chemically non-reactive. Furthermore, changes to hydrophobic amino acid residues can influence the shape of active site pockets thus changing the structure of the protein which can be either deleterious or advantageous.

2.5.1.3 Modelling-based rational enzyme engineering

The use of sequence and structural-based techniques have helped evolutionary scientists and protein engineers to identify potential enzyme ancestors along with the changes which have taken place to produce diverse enzymes [44,45]. The wealth of information gained from these approaches is invaluable, as these have facilitated the identification of potentially critical areas or hotspots in proteins as well as narrowing down the possible amino acids to consider for substitution, deletion or insertion. However, this can be a time-consuming process [31]. With the increase in computational power of the general personal computer and the optimisation of the programs needed to do *in silico* simulations, it has become more and more viable for the non-expert user [29]. Simulation techniques such as molecular docking and molecular dynamics will be discussed below to highlight the potential and shortcomings of these techniques.

Molecular docking. Molecular docking was developed as a tool for structural molecular biology and computer-assisted drug design [46]. The purpose of ligand-protein molecular docking is to predict the main binding mode or modes of a ligand in a known or unknown protein three-dimensional structure [46]. The basic principle of docking is the search of the three-dimensional space to isolate an area of the protein's active site in which the ligand or desired ligand will fit (induced fit). Using statistical scoring of the ligand-protein complex energies, complexes can be ranked accordingly to their binding affinity [46]. This method has numerous applications. It can be used as a way to evaluate a protein's active site to identify potential ligands for applications such as *de novo* protein design or to identify inhibitors [47,48].

Furthermore, this method can be used in enzyme engineering to identify clashes of a ligand and amino acids, key amino acids that are involved in substrate binding and the orientation of the ligand [29]. This method has a few weaknesses, one being reliability. To increase reliability and to ensure accurate results, this method requires a ligand that has been crystallised with the protein of interest or a structural homolog [49]. Enzyme-ligand crystallography data is needed to superimpose theoretical pose after docking for validation purposes. Another factor that contributes to the failure of docking experiments is the significant protein changes during

ligand binding [29]. The latter has been addressed by the algorithms used in molecular docking protocols by defining the flexibility of the bonds in the ligand [50]. Molecular docking often fails to identify the correct binding mode and has a high rate of false-positives in identifying mutation hotspots [29]. Despite these shortcomings, the use of molecular docking remains an invaluable tool to identify potential drug targets and screening of potential ligands that an active site can potentially accommodate [51].

***Molecular dynamics*.** Molecular dynamics is the science of simulating the movement of a system of particles [52]. Molecular dynamics can be applied to small systems such as atoms to larger systems such as galaxies. The core principles of molecular dynamics stay the same; one needs knowledge of the interaction potential for the particles from which the forces can be calculated, and the equations of motion governing the dynamics between the particles [52]. The advantage that molecular dynamics possess above static methods such as molecular docking is the ability to simulate movement. Proteins undergo conformational change at a variety of timescales. On the shortest of these timescales, bonds vibrate and side chains rotate. On longer timescales, macroscale motions can take place; loops ‘open’ and ‘close’, and domains can twist relative to each other or move on hinge-like regions. This movement is crucial as it gives more information about the protein’s interaction with the ligand [53]. These insights can contribute to informing the researcher about protein stability or the interactions with the ligand, which would not have been apparent in the static environment [54]. Molecular dynamics makes use of algorithms to simulate movements of the protein and ligand. The simulated movement can be used to evaluate subtle changes to protein structure which can potentially affect the enzyme’s function. These subtle changes can be changes to the active site due to ligand binding (conformational changes) or due to changes of an enzyme in a harsher environment (e.g., lower or higher temperatures) leading to the protein structure changes which affects the enzyme’s function [55]. Molecular dynamics can thus highlight the potentially hidden information that is not observable using techniques such as sequence-based, structural-based and molecular docking methods and give valuable insight to observed experimental phenomena [54].

As with most of the techniques listed above the process is not without its inaccuracies and three primary sources of inaccuracies can plague molecular dynamics simulations. The first of which is the error in the accuracy of the force field calculations. The second source of error involves the extent of the conformational sampling and is known as the sampling error which occurs due to the constraint on computing time and the methods which limit the flexibility and sampling of the protein [54]. The third and final limitation is the determination of the correct simulation time by selecting between economy (shorter simulations) and sampling (longer simulations). The constraints on time pose a challenge; thus there needs to be a balance

between economy and sampling. Although these problems are still present today, advances in software development, force field development, high-performance computing and parallelisation schemes, have allowed simulations to increase in time and accuracy. The latter combined with the advances in the validation programs within molecular dynamic software has increased the accuracy and reliability of molecular dynamics simulations [54].

The application of molecular dynamics in protein design/engineering can inform the user either directly or indirectly [54]. The direct application is the use of the information gained from the simulation to guide the rational design of a mutagenesis strategy or using the simulation results to generate a testable hypothesis [55]. Another use of molecular dynamics is using the technique in an indirect manner which makes use of the simulations to screen, rank and evaluate designs based on the results that were obtained during the simulations [56]. Using molecular dynamics to gain general insights into the protein dynamics can help to explain the relationship between protein-ligand interactions. It offers an understanding of the dynamic nature of the protein and how it relates to the protein's molecular mechanisms as well as its stability [54].

2.5.2 *De novo* protein design

Most of the enzyme engineering approaches discussed thus far can be described as protein redesign, i.e., taking pre-existing proteins found in nature and then improving them by using molecular methods such as rational engineering, directed evolution (section 2.5.3) or semi-rational design (section 2.5.4) [54]. *De novo* design is a new direction that is starting to become more and more prevalent as our understanding of proteins and their structure increase [31]. *De novo* protein design differs from rational engineering, directed evolution and semi-rational design in that it starts with the existing chemical and biochemical knowledge of proteins to design proteins from scratch [57]. Novel or new enzymes can be created by grafting new folds with known function into existing scaffolds as well as creating new activities by grafting active sites onto another protein class scaffold to produce new activities [32].

De novo protein design is the complete theoretical design of the protein scaffold, the active site or the entire protein [32]. **"**True" *de novo* design starts with the generation of a natural or unnatural chemical ligand [58]. After the design of the ligand, the active site is designed around this ligand, and the active residues are then selected. After selection using simulations, i.e., molecular dynamics, the active residues are fine-tuned. After the active site is finalised, a scaffold that will be able to house the active site is then built. The scaffold design focuses on providing the optimal stability for the specific process to be catalysed. The residues that will help with substrate stability is selected in a similar way - amino acids are selected such that

they will stabilise the substrates and increase specificity [59]. After the theoretical protein is designed, it then undergoes a series of optimising simulations until it is favourable for synthesis and testing in the laboratory [60]. *De novo* protein design follows an iterative process of design, build, test, and learn where the predicted designed protein's physical and theoretical properties can be analysed. These can be used to change the algorithms to suit the designed protein better [31].

Limited knowledge of the protein's structure-function relationship is one of the biggest hurdles to the *de novo* protein design process as there is currently no accurate method to predict the potential effects of an amino acid mutation on an enzyme. Highlighting that current protein structure-function knowledge is not yet sufficient to design a protein from scratch [59]. The pure *de novo* design is in its beginning phase but has the potential to revolutionise both the protein engineering field and the biocatalysis industry. This method will provide scientists with the ability to design a protein from scratch with tailored characteristics.

2.5.3 Enzyme engineering by directed evolution.

Rational enzyme design has immense potential, and with the solving of the protein folding problem and correlating structures with the function of enzymes, true *de novo* design will be the future of designer enzymes [9]. However, the protein folding problem still exists, and the number of gene sequences far exceeds the available good quality protein crystal structures [61]. Combined with the relative lack of functional data means that, for the moment, the data-driven rational design approach is only feasible for a few well-studied enzymes [62]. As an alternative approach, directed evolution, a technique pioneered by Frances Arnold and Willem Stemmer, makes use of short time-scale evolutionary-based techniques to induce changes within a protein of interest [63,64]. Directed evolution approaches do not need any experimental enzyme data or structure-function knowledge and use techniques that are well understood, albeit with some limitations [59]. Directed evolution is a very effective method to increase an enzyme's stability as well as its activity and can scale the fitness landscape of an enzyme relatively quickly [65]. This ability has made directed evolution the go-to approach and has been combined with rational approaches where it illustrated exceptional performance [66]. There are multiple techniques to incorporate random changes within the gene sequence of choice with the most common being error-prone PCR (ep-PCR) and DNA-shuffling techniques [66].

The ep-PCR method makes use of a non-proof reading DNA polymerase enzyme combined with suboptimal reaction conditions such as an increase or decrease in $MgCl_2$ concentration or changes in the concentration of specific deoxyribonucleotide triphosphates [67]. The

increased error rate of the PCR reaction leads to the generation of gene variant libraries with random changes in their sequence, that translates into roughly three to seven amino acid changes in the enzyme [68]. DNA-shuffling is a method that produces mutations based on the random recombination of gene fragments. As described in the original work by Stemmer [69], deoxyribonuclease (DNase I) is used to cleave and fragment genes which are then recombined randomly using a PCR reaction to produce a full-length sequence. The production of full-length sequences is followed by screening or a selection process to select for the fittest recombinant enzyme which will then undergo further reiterative cycles of DNA-shuffling [69]. There is a vast array of different techniques within the DNA-shuffling approaches to produce more diverse and interesting enzymes [70]. The advantages of ep-PCR and DNA-shuffling techniques is that there is no need for prior knowledge of the protein's sequence or structure, and due to the random nature of the mutations it covers more sequence space and produces unique mutations that would be difficult or nearly impossible to rationalise [34,70].

One of the biggest challenges with directed evolution to cover the amount of sequence space is the need to screen large numbers of variants [57]. One of the reasons for these large libraries is the potential combinations within a protein. There are 20 natural amino acids which mean the possible combination of sequences is 20^N where N is the length of the protein [71]. This correlates, for example, for a relatively short protein of N = 100 amino acids having 20^{100} possible sequences [71]. Furthermore, large libraries of variants need to be screened as certain amino acid changes do not translate into favourable functional or structural changes within the enzyme [59]. ep-PCR has other limitations such as the relatively low error rate due to the redundancy of the genetic code which means that changes in codons do not necessarily translate into amino acid changes at the protein level [72]. Most of these problems can be solved by using a gene shuffling approach, which has the advantage of introducing genes from different sources and building an enzyme with greater stability or activity [73]. Despite these bottlenecks and limitations, directed evolution is still a highly successful approach to consider when producing enzymes that are of industrial value [59]. Furthermore, the knowledge gained from random changes and recombination also contributes to our understanding of the fundamental aspects of protein folding and structure-function relationships.

2.5.4 Enzyme engineering by semi-rational design

Both rational design and directed evolution have led to significant successes in enzyme engineering, but both techniques have advantages and disadvantages [74]. Combining these two approaches, called semi-rational design, provides a method with more benefits than using

only rational design or directed evolution, while also reducing the major disadvantages [64]. One of the current semi-rational strategies, site-saturation mutagenesis, makes use of a rational approach such as CASTing, sequence analysis or using structural methods (superimposition) to identify key areas or ”hot spots” followed by substitution of a site with all 19 natural amino acids alternatives [75]. Using “hot spots” effectively reduces the library size several folds for screening as the sequence space being explored has been confined to a smaller area. Similarly “hot spot” areas can also be replaced by sequences from similar enzyme families or different enzyme families using routine cloning techniques [76]. Another powerful method in semi-rational design starts with *de novo* designing of enzymes followed by ep-PCR, gene shuffling or site saturation to increase enzymes fitness [28].

Combining the two methods using rational approaches to identify “hotspots” with random mutagenesis methods, such as site-saturation mutagenesis, effectively reduces the library sizes required for screening, especially if the desired trait such as selectivity or specificity is targeted, since most of the changes that affect these activities (selectivity or specificity) are in a 10 Å radius around the active site [59]. These libraries are known as “smart libraries” as the mutagenesis effort is focused on a specific area, thus reducing the sequence space to a smaller portion, therefore requiring smaller screening libraries [77]. The semi-rational design has some negatives in that the majority of the sequence space is not explored. The exploration of the sequence space can be enhanced by not only targeting “hotspot” areas but also combining it with ep-PCR. By combining ep-PCR after site-saturation of a “hot spot”, means the sequence space is explored to a larger extent and could lead to improvements outside what would have been reached using purely rational design methods while at the same time decreases the library size to manageable levels [64]. Another enzyme engineering strategy that has been explored with success is the use of a randomised approach during the first round of engineering to identify “hot spots” followed by the use of these residues as the target for the second round mutagenesis [64]. Semi-rational design methods have experienced numerous successes while also contributing to our critical understanding of the fundamentals of enzymes in the absence of structural information [64].

Computational methods have had a significant impact on protein engineering using semi-rational design by increasing productivity thereby, decreasing experimental time [28]. Even though the semi-rational design process reduces the number of mutants in a library, it still constitutes a vast amount of variants that often requires specialised, costly equipment and expertise to be screened [64]. Computational methods can be used as a virtual screening tool and consequently has the potential to increase the effectiveness of semi-rational design [28]. Using tools such as bioinformatics, modelling and *de novo* designs, random mutations can be generated within a sequence, and the library can be screened virtually by eliminating variants

with inconsistent folds [28,64]. Virtual screening effectively reduces the number of clones to be screened by several orders of magnitude. This decrease is achieved by selecting the optimal sequences based on the lowest conformational energies as well as staying with the mutations of the highest frequency to reduce the experimental screening to a manageable size [64].

The future of protein engineering of biocatalysts or the understanding of the specific enzymes will probably make use of a combination of both directed evolution and rational engineering approaches, assisted by computational methods, as it leads to the faster acquisition of fundamental knowledge that is critical for the modification of the enzymes [28,29,31,78].

2.6 Introduction to β-fructofuranosidases

β-fructofuranosidases is a class of enzymes belonging to the Glycoside Hydrolase (GH), family GH 32 according to the Carbohydrate-Active enZymes Database (CAZy) classification system [79]. β-fructofuranosidases are known to hydrolyse sucrose, but some can catalyse the transfructosylation reaction resulting in the formation of fructooligosaccharides (FOS) [80]. β-fructofuranosidases are used in industry to hydrolyse sucrose as well as to produce FOS. The latter has gained more interest as there is a growing need from consumer looking for alternatives to conventional medicines [81]. FOS being a nutraceutical has gained interest for its' use as a prebiotic and a sugar substitute [82]. β-fructofuranosidases is a promiscuous enzyme and has been identified to fructosylate different acceptors which makes it an ideal candidate for the use in the biotechnological landscape [83]. Piedrabuena et al. [83] illustrated the range of promiscuity of β fructofuranosidase with the *Schwanniomyces occidentalis* enzyme, which was able to fructosylate various mono-, di- and oligosaccharides. The *S. occidentalis* β-fructofuranosidase was tested using forty compounds to determine their ability to act as acceptors for the enzyme. Out of the forty compounds, seventeen compounds ranging from mono-, di- and oligosaccharides as well as alditols and glycosides were fructosylated. The two products produced at the highest levels were fructosylated mannitol classified as 1-O-β-D-fructofuranosyl-D-mannitol at a concentration of 44 g/l and fructosylated erythritol classified as 1- and 4-O-β-D-fructofuranosyl-D-erythritol produced at a concentration of 35 g/l [87]. The promiscuity of β-fructofuranosidases can be used in the nutraceutical industry to develop new products.

2.6.1 The classification of GH 32 family enzymes

Sugar is an all-engulfing term that can mean many things to different people but, to a glycobiologist, it is a monosaccharide. These monosaccharides can be rearranged and interlinked in different ways and can be present in sugars themselves or linked to amongst others proteins, lipids, nucleic acids and antibiotics [62]. As a result of the pervasive role of sugars in biology, a large variety and diversity of enzymes acting on these glycoconjugates, oligo-, and polysaccharides exist [62,79]. The CAZy database is a knowledge-based electronic resource specialising in enzymes which is responsible for the build and breakdown of complex carbohydrates and glycoconjugates [62]. The CAZy database started with efforts to classify cellulases into distinct families based on their sequence similarities. This family classification, a system based on protein sequence and structural similarities, were extended to all known glycoside hydrolases [62]. One of the defining features of the CAZy classification system is that it defines families based on significant amino acid sequence similarity and at least one biochemically characterised enzyme [79]. Using this conservative approach of classification makes the database successful in classifying enzymes into distinctive families with a particular activity based on significant sequence similarity. However, simultaneously this classification method excludes sequences that are too dissimilar [79]. Another feature of the CAZy database is that it only allows the analysis of protein sequences released by GenBank (https://www.ncbi.nlm.nih.gov/genbank/), to avoid accession number changes/conflicts which can be associated with incomplete sequences [79].

As mentioned, β-fructofuranosidases are a classed under the CAZy GH 32 family (http://www.cazy.org/). The GH 32 family is a family that consists of 14 different enzyme types [79]. These include: invertase (EC 3.2.1.26), endo-inulinase (EC 3.2.1.7), β-2,6-fructan 6-levanbiohydrolase (EC3.2.1.64), endo-levanase (EC 3.2.1.65), exo-inulinase (EC 3.2.1.80), fructan β-(2,1)-fructosidase/1-exohydrolase (EC 3.2.1.153), fructan β-(2,6)-fructosidase/6-exohydrolase (EC n3.2.1.154), sucrose:sucrose 1-fructosyltransferase (EC 2.4.1.99), fructan:fructan 1-fructosyltransferase (EC 2.4.1.100), sucrose:fructan 6-fructosyltransferase (EC 2.4.1.10), fructan:fructan 6G-fructosyltransferase (EC 2.4.1.243), levan fructosyltransferase (EC 2.4.1.-), [retaining] sucrose:sucrose 6-fructosyltransferase (6-SST) (EC 2.4.1.-), and cycloinulooligosaccharide fructanotransferase (EC 2.4.1.-).

The GH 32 and GH 64 families harbour normal hydrolases but also non-Leloir type transferases in the form of fructosyltransferases involved in fructan biosynthesis [80]. The GH 32 and GH 64 families are structurally similar and form the GH-J clan. This clan shares structural features such as the N-terminal β-propeller domain. The N-terminal domain is a

5-bladed β-propeller (I-V) and each of these blades form a central cleft or a tube structure [88, 89]. The β-propeller domain tube structure harbours the catalytic residues which are D23, E204, and D152, collectively known as the catalytic triad [86]. These three residues were identified in three conserved motives by Pons *et al.* [86] as "WMN**D**PNG" which is the nucleophile, "**E**C" the acid/base catalyst and finally the "R**D**P" which is believed to be a transition state stabiliser. The GH 32 family enzymes also contain a C-terminal domain that is absent in the GH 64 family. This domain consists of β-sheets which are formed of four antiparallel β-strands labelled A-D which is connected by variable loops and displays a characteristic "W" topology [84].

2.6.2 Catalytic mechanism of the GH 32 β-fructofuranosidases

The β-fructofuranosidase enzyme was central to the development of the enzyme kinetics field. Michaelis and Menten [87] used a β-fructofuranosidase (*Saccharomyces cerevisiae* invertase) to model enzyme kinetics with its substrate, sucrose. Koshland and multiple of his colleagues pioneered studies of enzymatic mechanisms using the *S. cerevisiae* invertase leading to our current understanding of the inner workings of enzymes [88]. Reddy and Maley, [89] identified the catalytic D23 residue in invertase and later the E204 residue. The positions of the D23 and E204, have been used to identify these amino acid positions in other β-fructofuranosidases for classification purposes [86].

The CAZy database classifies enzymes that produce or hydrolyse FOS in the GH 32 and GH 68 families, both these families retain the anomeric configuration of the anomeric carbon, while their catalytic mechanism follows a double displacement reaction using an enzyme-covalent intermediate illustrated in Fig.2.2 [90]. Given that these enzymes' catalysis follows a double displacement mechanism, the first phase is initiated by the anomeric carbon of the sugar which is carboxylated by the nucleophile and produces a fructose-enzyme intermediate [90]. The acid/base catalyst acts as the acid and donates a proton to glycosyl leaving the active site [90]. The second phase which is the deglycosylation step follows, where the acid/base catalyst acts as the base and removes a proton from the fructosyl acceptor which can be either a water or a sugar molecule [90]. A water molecule then protonates the acid/base catalyst to perform a nucleophilic attack on the anomeric carbon of fructose which leads to the hydrolysis of the fructose-enzyme intermediate [91].

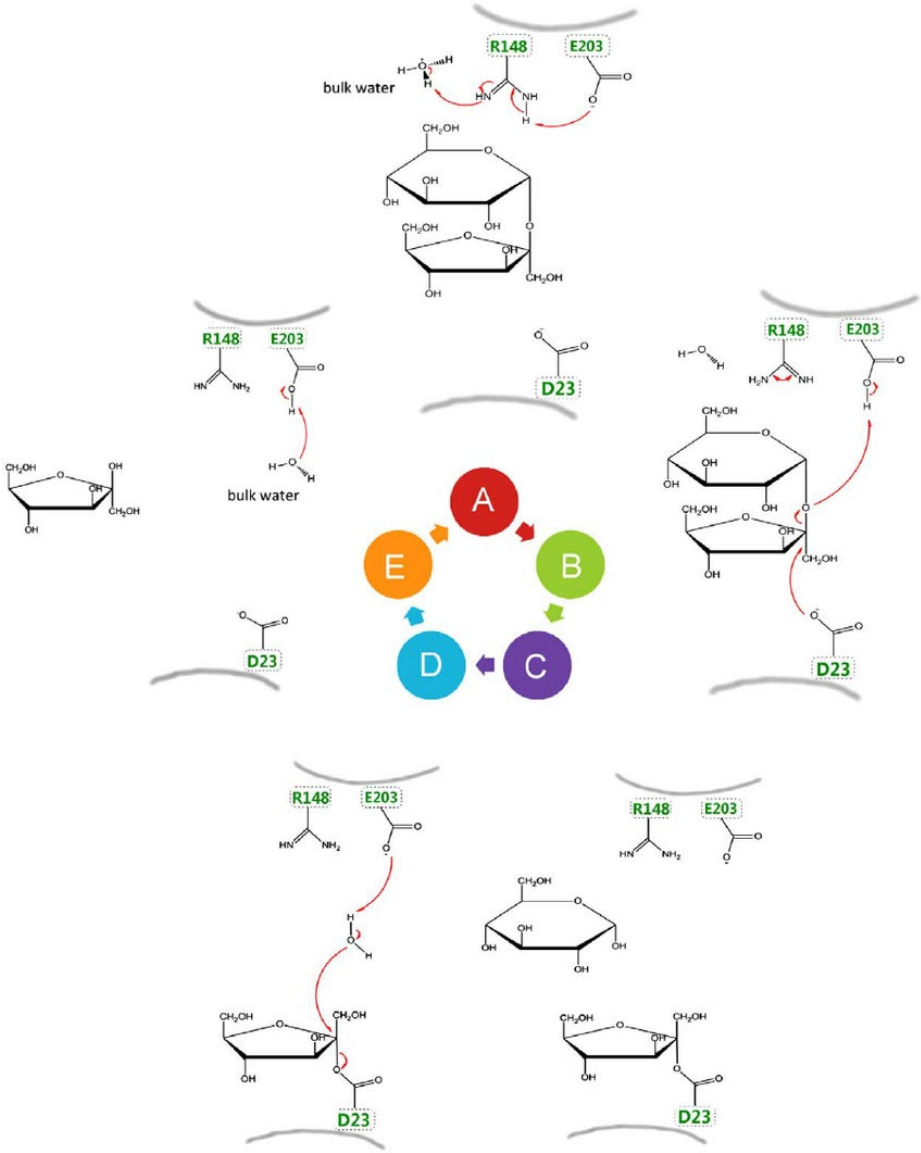

Figure 2.2. The proposed reaction mechanism for the catalytic process in the enzyme clan GH-J adapted from [90]. **(A)** When sucrose binds in the active site, the pK_a of E203 increases and a neighbour (R148) protonates it which can extract a proton from a nearby water molecule. **(B&C)** The glycosidic oxygen in sucrose is protonated by acid/base E203; the nucleophile D23 attacks the anomeric carbon, forming a covalent enzyme-fructosyl complex and releasing glucose from the active pocket. **(D&E)** A water molecule is split, reprotonating the acid/base and leaving the hydroxyl group on fructose. The enzyme is ready to accept another sucrose molecule after fructose has left the active pocket

Reverse hydrolysis is an equilibrium process which takes place in two steps. In the initial step, a fructosyl group from sucrose (the donor) interacts with the β-fructofuranosidase enzyme, in turn forming the enzyme-donor complex. This complex then reacts with the hydroxyl group of the acceptor (sucrose or FOS), resulting in a product, i.e. FOS and the release of both glucose and free enzyme. In the hydrolysis reaction, the complex reacts with the hydroxyl group of the acceptor (water) resulting in free fructose, glucose and enzyme. This hydrolysis/reverse

hydrolysis is dependent on the concentration of sucrose as an increase in sucrose concentration leads to a decrease in water. The concentration of oligosaccharides produced is dependent on the conversion reactions by the enzyme-donor complex as well as the substrate concentrations. This is simplified by the diagram below, where Glu-Fru represents sucrose, Fru is the fructosyl group and E the enzyme.

$$\begin{array}{c} \text{Suc/Fos-OH} \\ \text{Glu-Fru + E} \rightleftharpoons \text{E-Fru-Glu} \rightleftharpoons \text{FOS + E + } H_2O + \text{Glu} \\ \Downarrow + H_2O \\ \text{Fru-H + E} \end{array}$$

During transfructosylation, the enzyme-fructose complex will transfer its fructosyl group to an acceptor molecule (sucrose or FOS). This process is also a two-step process. The first step of the reaction is characterised by the formation of an activated donor-enzyme complex. Once formed, water and sucrose will compete for the fructosyl group within the complex. The products (kestose and nystose) can be substrates for β-fructofuranosidases. Therefore, this reaction is usually a kinetically controlled process. Thus, the FOS concentration is dependent on both the rate of hydrolysis and synthesis of oligosaccharides [92]. In these reactions it is routine to observe: an increase in the total FOS concentration, then reaching a period during which FOS production stabilises followed by a decrease. Transfructosylation is schematically represented below: Glu-Fru is sucrose responsible for the donation of the fructosyl group, Suc/FOS are acceptor molecules.

$$\begin{array}{c} \text{Suc/FOS} \\ \text{Glu-Fru + E} \rightleftharpoons \text{ Glu + E-Fruc} \rightleftharpoons \text{FOS + E + } H_2O \\ \Downarrow + H_2O \\ \text{Fru + E} \end{array}$$

The enzymes responsible for the transfer of the fructosyl group to a molecule of sucrose or FOS are known as fructosyltransferases. These enzymes increase FOS by one fructose molecule. Fructosyltransferases has a low affinity for water as an acceptor which in turn dictates lower hydrolysis. Fructosyltransferase produces high concentrations of FOS due to the higher affinity for sucrose or FOS thereby achieving a higher FOS yield even in low substrate concentrations [92]. This differentiation between hydrolytic and transfructosylating activities has been demonstrated in plants but is still controversial in FOS production by microbial enzymes [92].

2.6.3 β-fructofuranosidases and their importance in functional food

The β-fructofuranosidase enzyme is a major player in the nutraceutical industry. These enzymes are valuable for the production of the prebiotic - fructooligosaccharides (FOS). There is an increased demand for functional foods supplemented with nutraceuticals as current day consumers are taking control of their health and well-being by focusing on their diets to achieve the desired outcomes [93]. Apart from β-fructofuranosidases products being used as a functional food, these enzymes are also promiscuous and can fructosylate other substrates (maltose, trehalose, raffinose, maltotriose). This property makes it an ideal enzyme for producing other functional foods [94]. These enzymes are thus important in the development of novel nutraceuticals [83].

Nutraceuticals are defined as any product that is derived from a food source, that confers not only basic nutritional value but also provides additional health benefits [95]. Nutraceuticals are organised into three segments: functional foods, functional beverages and dietary supplements [82]. The field of functional foods is a combination of food science, nutrition and medicine which started in Japan in the 1980s. Functional food is defined as a food ingredient that provides health benefits beyond what was already known [82]. The Food for Specific Health Uses (FOSHU) has added to the basic definition. According to the FOSHU the classification of functional foods must have specific nutritional requirements as follow:

(1) Food must show effectiveness in clinical studies
(2) Food should be safe for use in clinical and nonclinical studies
(3) The active component of the food must be determined

These requirements do differ from country to country although there is a consensus that these foods should convey some form of health benefit. Functional foods are categorised into three groups:

(1) Regular food with naturally occurring bioactive substances, e.g., dietary fibre
(2) Food that has been supplemented with bioactive substances, e.g., probiotics and antioxidants
(3) Derived food components added to traditional foods, e.g., prebiotics

2.7 FOS classification

Fructans are polymers of fructose with a terminal glucose residue, produced by 15% of flowering plants as well as microorganisms [96]. Fructans in plants serve as reserve carbohydrates and have been linked to the plant stress response as they are implicated as

protectants in plants under adverse conditions (e.g., drought and cold) [96]. Plant fructans range in the degrees of polymerisation (DP) and contain between 10 to 200 fructosyl units. Bacteria also makes use of the fructans to store energy. Pathogenic bacteria use fructans to create a protective layer surrounding the cell to evade detection and protect against bacteriostatic compounds [82]. Fructans are classified based on their structure as they differ in length, DP, branching, the bonds between the fructose units and the position of the glucose molecule [97].

The main types of fructans are levan-type, inulin-type, and neoseries fructans. In levan-type fructans the fructose unit is linked by a β-(2→6) glycosidic bond, the levan-type fructan series starts with 6-kestose [96]. The inulin-type fructans contain fructose units linked by β-(2→1) glycosidic bonds and this series starts with 1-kestose. The final type fructans are the neoseries which contain mixed bonds - the bonds can be β-(2→1) or β-(2→6) glycosidic bonds, with an elongation at both C1 and C6 positions of the glucose residue [96].

FOS is a sucrose that is bound to successive fructose units using the linkage β-(2→1) (inulin-type) or β-(2→6) (levan type) glycosidic bonds, with the general structure of GF_n with the G being glucose, the F being fructose and the "n" is the number of fructose units [82]. Commercial FOS is sold purely as the inulin-type of FOS, and this coincides with most studies performed on this type of FOS [98]. The DP of all types of FOS is one of the determining factors for its classification and was set at lower DP's ranging from 2 to 10 fructose units. The latter has further been refined with the classification of short-chain fructooligosaccharides (scFOS) which only have 1 to 5 fructose units bound to a terminal glucose [98]. The two ways of producing FOS are the partial hydrolysis of inulin with inulinases, which produce fructan consisting of only β-(2→1) glycosidic bonds, but with a DP of greater than 4, or via enzymatic fermentation with sucrose to produce FOS with a DP lower than 10 [99].

2.7.1 The potential of FOS as a prebiotic

Prebiotics was defined by Gibson and Roberfroid, [100] as "*a nondigestible food ingredient that beneficially affects the host by selectively stimulating the growth and or activity of one or a limited number of bacteria in the colon, and thus improve host health*." A prebiotic should reach the colon without being degraded or hydrolysed in the process and should serve as a substrate for beneficial bacteria [82]. According to Gibson and Roberfroid, [100] for a food ingredient to be classified as a prebiotic, it must have the following criteria:

(1) It should not be hydrolysed nor absorbed in the upper gastrointestinal tract

(2) It must be a selective substrate for one or a limited number of beneficial bacteria commensals to the colon and should either stimulate grow or metabolically active
(3) Should be able to alter the colonic flora in favour of a healthier population
(4) Induce a luminal or systematic effect which is beneficial to the host health

FOS is a prebiotic as neither the pancreas nor the small intestine produces enzymes that can catalyse the hydrolysis of the β-(1→2) glycosidic bond between two fructose units thus reducing the energy value of FOS as only the β-(2→1) glycosidic bond between the glucose and fructose can be hydrolysed [100]. The advantage of this is conferred in the human gut - bacteria which populate the gut produce the required enzymes and make use of FOS as their carbon source [101]. Strains of *Bifidobacterium* and *Lactobacilli* in the colon produce enzymes such as β-fructofuranosidases which can catalyse the hydrolysis reaction [102,103]. As a result, FOS is fermented in the gut into short-chain fatty acids (SCFA) including acetate, butyrate, and propionate which can be absorbed by the body and used as energy [82]

2.8 Overview of mutagenesis approaches in β-fructofuranosidases

The β-fructofuranosidases have been the subject of extensive mutagenesis studies because of their ability to produce FOS [104]. Mutagenesis of the GH 32 family enzymes has had a strong research focus to understand the mechanisms behind the catalytic diversity within the group [85,104,105]. Mutagenesis studies in the GH 32 family enzymes have focused explicitly on interrogating the mechanisms behind the predominantly hydrolytic enzymes *versus* transferase enzymes, substrate specificity and the production of novel FOS species [98,104,105].

2.8.1 Changing of an invertase to a fructan exohydrolyse through a point mutation

Plants usually have multiple GH 32 enzymes present, with some of the enzymes playing a role in the storage of fructose in the form of fructans, while other enzymes play a role in degrading stored FOS [105]. For example, plant GH 32 enzymes include cell wall invertases and fructan exohydrolases, which are closely related on a molecular and structural level. Cell wall invertases prefer hydrolysing β-(2→1) glycosidic bonds between the glucose and fructose of sucrose, while fructan exohydrolases favours the hydrolysis of β-(2→1) glycosidic bonds between the two fructose units of fructans [106]. Le Roy *et al.,* [105] made use of sequence and structural data combined with mutagenesis studies to understand the differences between

a plant cell wall invertase and fructan exohydrolase enzymes based on changes to key amino acids. During their search, they identified and substituted the following residues D239A, K242L, W20L, W47L and W82L (numbering based on plant cell wall invertase). The D239A change affected the invertase in that it lost the ability to hydrolyse sucrose but still retained the ability to hydrolyse 1-kestose [105]. The change of D239A did not increase the enzyme's ability to hydrolyse higher DP fructans thus did not increase the space to accommodate the higher DP fructans [105]. The K242L substitution variant retained its substrate specificity similar to the wild-type, and it was hypothesised that this change to lysine led to an interaction between lysine and D239 to keep its orientation [105].

Significant effects were observed with changes to the tryptophans (W20L, W47L and W82L), with W20L resulting in an increase in K_m of 200 times for sucrose compared to the wild-type and 120 times increase for W82L [105]. The change to W47L lead to an increase in K_m of 600 times compared to the wild-type. These effects of the tryptophan residues were not surprising as these residues are located near the active site indicating that these aromatic amino acids' hydrophobic rings were orientated to interact with the substrate [105]. Polar forces are essential to carbohydrate recognition in the active sites of various enzymes. Depending on the stereochemistry of the carbohydrate monomer, the presence of aromatic residues can play an essential role in stabilising the sugar-enzyme complex. The aromatic ring of aromatic amino acids interacts with the sugar ring through Van Der Waals contacts and CH-π stacking [105]. The authors concluded that within the plant cell wall invertases the presence of these tryptophan residues (W20, W47 and W82) forms an aromatic zone which plays a critical role in substrate binding. Changes made to these residues led to potentially unstable substrate binding in all the variants with a concomitant increase in the rate of release for both fructose and glucose as seen in the increase in K_{cat}. This aromatic zone of tryptophan residues is conserved among most invertases in the **W**MNDPN, **W**GN, and **W**SGSAT regions [105]. Le Roy *et al.,* [105] elaborated further suggesting that the aromatic zone in the β-fructofuranosidase active site is important for optimal and stable binding of sucrose. The changes to D239A led to the decrease in the hydrolysis of sucrose but retained activity on 1-kestose. It was hypothesised by Le Roy *et al.,* [105] the reason for retaining the activity on 1-kestose and not higher DP fructans was due to the enzymes catalytic pocket being optimised to accommodate sucrose and 1-kestose. The observation that one amino acid change, D239A, led to a significant decrease in affinity for sucrose also supports the theory that cell wall invertase and fructan exohydrolases share a common ancestor as one-point mutation led to cell wall invertases behaving like a fructan exohydrolases [105]. This study highlighted the similarities between the members of the GH 32 family and emphasised the importance of key residues and their effects on the functioning of β-fructofuranosidases.

2.8.2 Mutagenesis studies in yeast β-fructofuranosidases

The GH 32 family consists of different enzymes from different organisms such as plants and fungi. The latter has industrial and agricultural importance and therefore, has the most functional information available within this family. The yeast *S. cerevisiae*'s invertase (Suc2) is one of the enzymes studied most often, although there is still little information about the fructosyl transferase activity of this enzyme [104]. Lafraya' study [104] employed a semi-rational design to substitute residues in the Suc2 protein for gathering a better understanding of the transferase/hydrolysis activity and the amino acids that confer enzyme specificity. In 2013 the crystal structure of Suc2 was solved by Sainz-Polo *et al.,* [107] which led to the identification of key elements for the enzyme. These key elements include the dimeric active form of the enzyme as well as the multimeric configuration of the enzyme. The solved crystal structure provides structural insight that will improve rational and semi-rational engineering approaches in this enzyme. Other yeast β-fructofuranosidases that have been crystallised and characterised include the β-fructofuranosidase from *S. occidentalis* and *Xanthophyllomyces dendrorhous* [108,109]. Following the solved crystal structures of this enzyme, mutagenesis studies ensued creating variants that could be characterised to understand better the effect specific residues have on the enzyme's function [83,104,108,110]. Below these changes together with the potential use of this information will be discussed in detail.

2.8.2.1 Mutagenesis of *Schwanniomyces occidentalis* β-fructofuranosidase

S. occidentalis is a yeast of industrial interest because it can utilise inexpensive substrates such as starch and inulin, and is also efficient at secreting high molecular weight extracellular proteins [109]. *S. occidentalis*'s β-fructofuranosidase received research attention as it could hydrolyse sucrose, 1-kestose, nystose and raffinose and produce FOS [109]. The *S. occidentalis* β-fructofuranosidase predominantly produces the levan-type 6-kestose with a smaller amount of inulin-type 1-kestose [111]. The β-fructofuranosidase of *S. occidentalis* shares sequence similarity with other yeast invertases, inulinases, levanases, and fructosyltransferases all of which are included in the GH 32 family [109]. Álvaro-Benito *et al.,* [109] were the first to crystallise the *S. occidentalis* β-fructofuranosidase and mutated positions within the enzyme that were suspected of being involved in substrate binding. Table 2.1 explains the mutations used by Álvaro-Benito *et al.,* [109,112] which is relevant in this study.

Table 2.1. Rational mutations used in *S. occidentalis* β-fructofuranosidase.

Amino acid positions of interest	Amino acid substitution	Effect of amino acid substitution	The contribution of functional understanding of β-fructofuranosidase
D50	A	Inactive enzyme	Nucleophile
E230	A	Inactive enzyme	Acid/base catalyst
N142	Y	From the crystal structure, the change to N142Y led to the rearrangement of E107 and Q147.	N142 is located at the end of strand B2 and interacts with E107 and Q147 its replacement by a Tyr probably rearrange the active site residues, lowering the affinity for the sucrose, nystose and inulin
Q176			The position does not form part of the acceptor site for sucrose in transfructosylation
	E	No influence on enzyme hydrolytic or transferase activity	
	S	No influence on enzyme hydrolytic or transferase activity	
	N	No influence on enzyme hydrolytic or transferase activity	
Q228		Deleteriously influenced the transferase/hydrolysis activity as well as affinity and specificity	Variable loop known as loop 3 with Gln-228 shaping subsites +1 and serves donor substrate binding site
	V	Reducing catalytic efficiency for sucrose and inulin. The catalytic efficiency of the enzyme, especially towards the long and medium chain FOS substrates. Affected specificity nearly exclusively producing 6-kestose,	The more significant reduction in the enzyme activity in the mutation with T and V indicates that amino-acid side chain length plays a role in the hydrolytic activity of the enzyme.
	T	Decreased catalytic efficiency of the enzyme, especially towards the long and medium chain FOS substrates. Affected specificity with the production of uniform amounts of neokestose, 1-kestose and 6-kestose.	
	N	A smaller reduction in catalytic efficiency of the enzyme, especially towards the long and medium chain FOS substrates. Affected specificity with the majority of FOS production being	The smaller reduction in the enzyme activity in the mutation with N and E indicates that amino-acid side chain length plays a role in the hydrolytic activity of the enzyme
	E		

		6-kestose with some 1-kestose	
N254			Located in the variable loop region, also known as turn 4 (TIV). Asn-254 forms part of the active pocket and shapes subsites +2 for donor substrate binding.
	A	Decrease in catalytic efficiency for all substrates	Plays a role in binding medium chain FOS or ligands
	D	Decrease in catalytic efficiency for all substrates and a decrease in total FOS production	Confirmed that they contributed to the recognition/hydrolysis of medium and long chain FOS
	T	Increase the production of Neokestose indicating an effect on specificity	
	E	Decrease in FOS production	
Q435	A	Large reduction in catalytic efficiency of the two enzymes as a result of a decrease in affinity of sucrose, nystose and inulin.	Located at the β-sandwich domain of the B molecule and participate in the polar interaction network that shapes the interface and the active site pocket
Y462	A		
S281	I	Large reduction in enzyme efficiency for sucrose, nystose and inulin.	Seem to be related to the interface within the dimer or the conformation of the active site cavity. Contributing to the idea that oligomerisation is one of the features delineating *S. occidentalis* β-fructofuranosidase functionality

Álvaro-Benito *et al.,* [109] also investigated non-conserved residues N142, Q176, Q228, and N254 and hypothesised that these residues play a role in the enzyme's specificity. Furthermore, the location of residues Q435 and Y462 implicated these residues' putative involvement in substrate recognition [109]. The residue Q176 is located on Loop 2 (L2) in *S. occidentalis* β-fructofuranosidases and three substitutions were investigated (Q176S, Q176N, and Q176E) with the variants having no apparent effect on the hydrolytic or transferase activity of the enzyme. They concluded that this residue located in L2 did not play a significant role in determining the acceptor site for sucrose in the transfructosylation reaction [112].

β-fructofuranosidases have a highly variable loop region, loop 3 (L3). In the *Thermotoga maritima* β-fructofuranosidase residue Q188, present in this loop, is involved in glucose binding and can explain its disruptive effect [109,113]. Q228 is located at the same position

as Q188 in L3, and Álvaro-Benito *et al.*, [112] thus used this information to investigate its role in *S. occidentalis* β-fructofuranosidases. Álvaro-Benito *et al.*, [112] investigated the significance of Q228 substitution's effect on the hydrolysis and transferase activity of the enzyme. The authors substituted Q228 with four different residues (Q228V, Q228T, Q228N, and Q228E). These substitutions of Q228 deleteriously influenced the transferase/hydrolysis activity as well as the affinity and catalytic efficiency of the enzyme [112]. The variants Q228V and Q228T reduced the catalytic efficiency of the enzyme, especially towards the long and medium chain FOS substrates. This finding was in stark contrast to the effects of Q228N and Q228E which had a smaller reduction in the catalytic efficiency [112]. This differential effect indicated that the size of the amino acid side chain played an important role in the hydrolytic activity of the enzyme at position Q288 [112]. These substitutions influenced the transferase activity with all four variants leading to a decrease in total FOS production compared to the wild-type within 24h [112].

Furthermore, these substitutions influenced the specificity of the enzyme, with Q228V nearly exclusively producing 6-kestose, Q228N and Q228E also producing primarily 6-kestose with some 1-kestose and Q228T producing low amounts of FOS albeit a uniform quantity of neokestose, 1-kestose and 6-kestose [112]. The lower FOS production illustrates Q228T was the most disruptive of all the substitutions as it decreased the catalytic efficiency, total FOS production and lastly produced considerably less 6-kestose [112]. The latter highlights the importance of L3 in the *S. occidentalis* β-fructofuranosidases, the substitutions of residue Q228 within this loop leading to adverse effects on the transferase activity as well as contributing to the specificity of the enzyme.

The position of N254 is in the variable loop region, also known as turn 4 (T_{IV}), located in the active pocket of *S. occidentalis*'s β-fructofuranosidase [109]. To test the significance of N254, three substitutions N254D, N254T, and N254E were introduced. All of these substitutions confirmed that they contributed to the recognition/hydrolysis of medium and long chain FOS [112]. The N254E substitution was highly disruptive as it decreased the catalytic efficiency, transferase activity as well as FOS production. Álvaro-Benito *et al.*, [112] proposed this disruption was N254E causing a rearrangement of W314's orientation and a subsequent change to subsite +3 numbering according to Davies *et al.,* [114]. This numbering system was developed by Davies *et al.,* [114] for carbohydrate processing enzymes and proposed that negative numbers (e.g., -1, -2, -3 etc.) characterise subsites on the non-reducing terminus from the point of cleavage while the reducing end from the point of cleavage is denoted by positive numbers. It was noticed that substitutions of N254 seemed to be involved in substrate specificity as N254T increased neokestose production to 11.8 g/l over 24h compared to the wild-type which produced no detectable levels [112]. Álvaro-Benito *et al.*, [112] thus found that

both Q228 and N254 had an impact on substrate binding as well as an influence on the specificity of the inulin/levan type FOS produced by the *S. occidentalis*'s β-fructofuranosidase. The T_{IV} is important in *S. occidentalis*'s β-fructofuranosidase as the substitutions of position N254 led to a reduction in transferase activity which in turn led to decreased production of FOS.

2.8.2.2 Mutagenesis of *Xanthophyllomyces* dendrorhous β-fructofuranosidase

The β-fructofuranosidase from a basidiomycete yeast *Xanthophyllomyces dendrorhous* is another enzyme that has gained interest over the past few years [108]. As with most β-fructofuranosidases, *X. dendrorhous* catalyses the release of fructose for the non-reducing end of substrates such as sucrose, 1-kestose or nystose and can synthesis FOS [108]. The *X. dendrorhous* β-fructofuranosidase is unique as the enzyme can transfer fructose (acceptor) to a glucose unit of sucrose forming neokestose and neonystose, producing neo-FOS (6G-FOS series) [96,108]. The *X. dendrorhous* β-fructofuranosidases is a promiscuous enzyme and can fructosylate glucose-related substrates, i.e. maltose, isomaltose and trehalose [83,94]. The β-fructofuranosidase from *X. dendrorhous* is classified in the GH 32 family and is part of the GH-J clan. This clan shares a five-blade β-propeller N-terminal domain and a C-terminal β-sandwich domain [108]. The β-fructofuranosidase of *X. dendrorhous* is a multimeric protein (homodimeric in the active form) and is highly glycosylated with 60-70% (w/w) of the enzymes mass composed of N-linked carbohydrate, similar to other yeast β-fructofuranosidases [108,115]. As with the majority of GH 32 enzymes, the *X. dendrorhous* β-fructofuranosidase have both hydrolytic and transferase activity. The activity is based on the acceptor substrate which can be either water (hydrolysis) or a sugar acceptor (transferase). The enzyme also catalyses the hydrolysis of fructans releasing free fructose [108]. Ramírez-Escudero *et al.,* [108] crystallised the *X. dendrorhous* β-fructofuranosidase and mutated specific residues which were identified to have an impact on enzyme activity or specificity. Table 2.2 indicates the mutations used by Ramírez-Escudero *et al.,* [108] which are relevant to this study.

Table 2.2. Rational mutations use in *X. dendrorhous* β-fructofuranosidase.

Amino acid positions of interest	**Amino acid substitution**	**Effect of amino acid substitution**	**The contribution of functional understanding of β-fructofuranosidase**
D80	A	Inactive enzyme	Nucleophile
E303	A	Inactive enzyme	Acid/base catalyst
N58	S	Decrease thermostability and reduction in enzyme efficiency with an increase in K_m	

N107	S	Inactive enzyme (improper protein folding)	The improper folding can indicate that N107 is essential for the recognition in the secretory pathway as the mutation N107S lead to no protein being expressed.
N471	S	No effect was witnessed with the enzyme behaving like the wild-type.	N471 is the amino acid linking the β-structure to its catalytic domain
E334			The E334 position forms part of the T_{IV} loop; this loop shapes part of the active pocket. The E334 position is required for sucrose binding as the substitution with S and V lead to inactive enzyme. The substitution of Q led to detectable levels of activity on sucrose.
	S	Fully inactive enzyme	
	V		
	Q	The substitution had detectable levels of activity on sucrose	
H343			The H343 position forms part of the T_{IV} loop; this loop shapes part of the active pocket. H343 interact directly with the terminal glucose of nystose thus shaping position +3.
	A	The substitutions of H343 with either A/T led to the reduction in activity on the assayed substrates. H343T had an effect on the specificity with the variant producing an altered profile of FOS.	
	T		
660 GLFAGY	A truncated enzyme with GLFAGY being removed	Produced an inactive enzyme by distorting the active site architecture. The truncated enzyme was not detected on the SDS-PAGE	This segment is an extended arm following the small helix at the C-terminal. The extension forms part of the active pocket and harbours G660 and L661 which shapes subsite +2.

Positions E334 and H343 were selected as these residues were distorted in the crystal of the free enzyme but fixed tightly to glucose at position +1, when in a complex with sucrose, neokestose and raffinose [108]. The residues E334 and H343 were investigated by mutagenesis studies to examine if they contributed to regulating enzyme specificity [108]. Q341 was investigated within the loop, as it interacted with 1-kestose and nystose, i.e., these substrates were oriented towards this residue [108]. The turn region T_{IV} together with residues Q341, E334, H343 and N342 served as subsite +3 were required for the stabilisation of the inulin-type FOS [108]. The residue E334 was found to be necessary for enzyme activity because substitution with amino acids S, V or Q produced a deleterious effect on sucrose hydrolysis and led to a fully inactivated enzyme. Thus, supporting the hypothesis that this residue is crucial for substrate binding [108]. It was only the E334Q substitution that had detectable levels of activity on sucrose and significantly retained its activity on nystose [108]. The variants H343A/H343T showed a decreased activity on the assayed substrates supporting the hypothesis of the T_{IV} loop and that changes to H343A/H343T result in a decreased substrate affinity as confirmed by an increase in the K_m.

The *X. dendrorhous* β-fructofuranosidase has an unusual segment characterised by an extended arm at the C-terminus following an α-helix. The segment extends to surround the

active cavity and harbours G660 and L661 [108]. Another variant created by Ramírez-Escudero *et al.,* [108] was a truncated *X. dendrorhous* β-fructofuranosidase enzyme with the deletion Y659STOP lacking the last residues, which were comprised of 660 GLFAGY sequences, this change was selected because of Gly-660 and Leu-661 shaping subsite +2 [108]. The substitutions made to this segment produced an inactive enzyme, suggesting that these amino acid positions lead to a distortion of the active site [108].

2.8.2.3 Mutagenesis of *Saccharomyces cerevisiae* β-fructofuranosidase

The *S. cerevisiae* β-fructofuranosidase is a predominantly hydrolytic enzyme belonging to the GH 32 family. Suc2 produces FOS under high sucrose concentrations and produces higher levels of 6-kestose with lower levels of 1-kestose [116]. The interest in the transferase activity of Suc2 is due to its preference for 6-kestose production, which is gaining further interest as it is said to have an added prebiotic effect [117]. Suc2 produces FOS at low concentrations relative to the *Aspergillus sp.* fructosyltransferase [118,119]. The difference between high and low FOS producing enzymes is one of fundamental and commercial interest within the GH 32 family. Although the GH 32 enzymes that produce FOS are similar in sequence and structure, FOS is produced at different concentrations [104]. The underlying mechanism for this effect remains unknown. Lafraya *et al.* [104] investigated which amino acids positions affect transferase/hydrolyse activity in Suc2 as well as the structural features which determine the transferase efficiency and product specificity. In Table 2.3, the mutations relevant to this study which were introduced by Lafraya *et al.,* [104] are discussed.

Table 2.3. Rational mutations use in *S. cerevisiae* β-fructofuranosidase.

Amino acid positions of interest	**Amino acid substitution**	**Effect of amino acid substitution**	**The contribution of functional understanding of β-fructofuranosidase**
W19	Y	The substitution did not increase transferase activity although the substitution did increase the global activity of the enzyme.	W19 forms part of the β-fructofuranosidase motif (**W**MNDPNG) which is near the nucleophile. The proximity of W19 to the nucleophile could explain the potential structural changes which have led to an increase in the global activity.
F82	W	Decrease in 6-kestose production and a small reduction in global activity	The change to F82W which is located in the hydrophobic pocket was not sufficient to change

			the hydrolase/transferase ratio.
N21	S	A 6-fold increase in the production of 6-kestose	The effect of the substitution of N21S on the enzyme can be explained by the hydrogen bond formed between N21 and fructose.
N24	S	The substitution increased the production of 6-kestose by 6-fold. The substitution led to a significant decrease in global activity.	The substitution of N24S on the enzyme can be explained because N24 forms a bridge between the two β-sheets which effects the amino acids interacting with the substrate. This substitution thus leads to a flexible catalytic site.
W19Y/N21S		The substitution increased the synthesis of 6-kestose by 15%. The substitution led to a 50% reduction in the global activity of the enzyme.	The double and triple variants of the single variants W19Y, N21S and N24S are the same as the single variants although the synergistic effects need to be further studied.
W19Y/N24S		The substitution increased the synthesis of 6-kestose by 10%. The substitution led to a 10-fold reduction in global activity.	
W19Y/N21S/N24S		The triple variant produced similar amounts of 6-kestose as W19Y/N21S. The triple variant saw one of the largest reductions with a 63-fold reduction in global activity.	
W287	Y	No expression due to improper folding	Due to the position of W287, the substitution of the residue led to the disruption of the protein structure. This can be explained by W287 being compactly packed inside the enzyme; thus the disruption led to incomplete folding.
W291	Y	Significant decrease the global activity	The substitution of W291 which is located in the hydrophobic pocket was not sufficient to change the hydrolase/transferase ratio of the enzyme.
P205	V	Global activity was similar to the wild-type enzyme, but the variant produced fivefold more 6-kestose.	P205 is associated with the “EC” motif from “EC**P**”. The substitution of P205V might have led to an increase in the flexibility of the active site.
P205V/ W19Y/N21S		The triple variant increased 6-kestose production by 10-fold producing a similar amount of 6-kestose to the double variant W19Y/N21S.	
D22	N	Inactive enzyme	Residue acted as the nucleophile

N228		Substantial reduction in global activity	The results obtained by Lafraya *et al.* [104] implicated N228 as a key component to fix sucrose in a specific orientation at the acceptor binding site
	R	No FOS production	The substitution of R interfered with the binding of sucrose.
	A	Higher global activity than wild-type with similar transferase activity	The substitution with A led to the interaction of W291 with the substrate which means hydrolysis and transferase activity were still intact.
N228R/N21S		The double variant saw a significant reduction in both the hydrolytic activity and transferase activity.	The double substitution has not overcome the adverse effects conferred by the substitution of N228R.
N228A/N21S		The double variant increased the FOS production significantly by 16.4% as well as producing equimolar 1-kestose and 6-kestose.	The synergistic effect of the double mutation is attributed to a more flexible active site which was less selective to a specific glycosidic bond, i.e. β-(2→1) glycosidic bond or β-(2→6) glycosidic bond

Based on previous work on GH 32 and GH 68 enzymes, Lafraya *et al.* [104] selected W19 and F82 (*S. cerevisiae* enzyme numbering), these residues form part of the hydrophobic pocket of the active site and are involved in substrate binding [104]. W19 is situated within the β-fructofuranosidase motif ("**W**MN**D**PNG") containing the D23 acting as the nucleophile. Within this motif ("**W**M**N**DP**N**G") the residues corresponding to W19, N21 and N24 are conserved among hydrolases in the GH 32 family [104]. Furthermore, these residues listed above are conserved among hydrolases, but they are often replaced by other residues in transferase enzymes [104]. Guided by prior knowledge Lafraya *et al.* [104] used the following substitutions: W19Y, N21S and N24S. These substitutions were further expanded to double (W19Y/N21S, W19Y/) and triple substitutions (W19Y/N21S/N24S).

Plant invertases have a W instead of an F82 present in Suc2 within the "WSGSAT" motif (in plants), although in both the GH 32 and GH 68 families with transferase activity show the preference for W [104]. Lafraya *et al.* [104] used the information regarding F82 to create substitution F82W, to asses a change in the transferase/hydrolase ratio of Suc2. Furthermore, Lafraya *et al.* [104] investigated W287 based on information from a substitution of W440 with Y in the *Allium cepa* invertase which led to the doubling in transferase capacity [120]. The W287 residue is located within the motive "WG**W**/Y" which is relatively conserved among GH

32 enzymes with plant invertases and transferase enzymes containing W and Y, respectively, in the third position [104]. Structural modelling of Suc2 was superimposed on the experimentally solved structure of the *Arabidopsis thaliana* invertase, highlighting that position W287 in Suc2 corresponded to W440 in *A. cepa* invertase and W340 in *A. thaliana* invertase [104]. This position was further investigated by Lafraya *et al.* [104] with structural alignments of transferases in the GH 68 family, and the homologous residue was a Y which agreed with the hypothesis that replacing W with Y favours the transfructosylating activity. *S. cerevisiae* β-fructofuranosidase W287 is structurally located at the same position as W340 from *A. thaliana* invertase. However, W340 is located at the surface in the *A. thaliana* invertase whereas W287 in the *S. cerevisiae* β-fructofuranosidase is part of a loop and hidden from the solvent [104]. In Suc2 W291, located at the entrance of the active site, was also investigated by Lafraya *et al.* [104] as an alternative target for mutagenesis within the hydrophobic pocket of Suc2, the substitutions W287Y and W291Y were induced. The "EC" motif has been investigated as this motif is conserved among the GH 32 enzymes family. In this motif, the E is the residue acting as the general acid/base catalyst (E204 in *S. cerevisiae* β-fructofuranosidase) in the reaction mechanism [121]. There is some divergence among the GH 32 enzymes with vascular plant invertases and transferases preferring a V at this "EC**V**/P" position while cell-wall invertases from bacteria, fungi and plants prefer a P [104]. Using this information, the "ECP" (as in *S. cerevisiae* β-fructofuranosidase*)* was substituted as P205V, this was then further combined with W19Y to form the variant W19Y/N21S to assess the impact of the modifications on the transfructosylating capacity of this enzyme. The outcome from the experimental data obtained by Lafraya *et al.* [104] identified the amino acids and their positions which contribute to a change in enzyme specificity by changing the ratio of 6-kestose:1-kestose. They concluded that the product specificity was determined by the orientation of the sucrose acceptor molecule in the catalytic site with respect to the reaction intermediate fructose-enzyme complex [104]. To determine which interactions theoretically stabilised the acceptor molecule in the active site, Lafraya *et al.* [104] performed a molecular docking simulation with Suc2 and the ligand 6-kestose. Suc2 subsite +1 (named according to nomenclature of Davies *et al.,* [114]) was suspected of being involved in the binding of sucrose as both the donor and acceptor molecule and the residue at site +2 (N228) was selected for mutagenesis as this residue was theoretically needed for the transferase reaction [104].

The substitution of W19Y in Suc2 deviated from the observations in plants and other yeast as it did not affect the transferase activity of Suc2 [104]. When W19Y was combined with N21S, there was an enhanced synthesis of 6-kestose, although this is not observed for the W19Y/N24S double mutant [104]. The synergistic effect was suggested to result from the structural proximity of W19 and N21 [104]. The mutation W19Y did not affect the transferase

activity although it had a significant effect on the global activity of the enzyme, increasing the global activity by 2020 units which implies an increase in the hydrolytic activity of the enzyme [104].

Individually, both N21S and N24S substitutions increased the synthesis of 6-kestose (6-fold increase), with the total FOS yield increasing from 3% to 11% [104]. The substitutions N21S and N24S were combined with W19Y and a synergistic effect led to an increase in the transfructosylation activity of Suc2 with W19Y/N24S increasing FOS yield from 3% to 13% [104]. This effect was further improved for the double and triple variants containing W19Y/N21S as the total FOS yield increase to ~18% of total sugars [104]. The maximum 6-kestose production was 270 mM for N21S and 200 mM for N24S after 20.5 h of incubation. The half-life of 6-kestose was observed at 80 h for N21S and 15 h for the N24S, indicating that N21S had a decrease in affinity for 6-kestose. The lower K_m of N21S confirmed this observation for 6-kestose [104]. The maximum 6-kestose production for W19Y/N21S was 370 mM at 20.5 h with a half-life of 60 h which showed an increase in affinity for 6-kestose opposed to N21S [104]. The double variant W19Y/N24S was similar regarding the synthesis of 6-kestose compared to the single mutant N24S, but when comparing their global activities, a two-fold decrease was noted [104]. The triple variant W19Y/N21S/N24S had a synergistic effect and produced similar concentrations of 6-kestose to variant W19Y/N21S [104]. There was a 5-time reduction in global activity when comparing W19Y/N21S/N24S with W19Y/N24S and a 26-fold reduction when compared to W19Y/N21S highlighting the synergistic effect of amino acids on each other [104]. The triple mutant produced a 6-kestose concentration of 380 mM at 20.5 h with a half-life of 60 h, as with the rest of the variants listed above there was no increase in the 1-kestose production compared to the wild-type [104].

The mutant P205V had the same activity as the wild-type which was a stark contrast to the triple mutation of P205V/W19Y/N21S that resulted in a substantial reduction in activity [104]. With similar global activity to the wild-type the transferase activity was enhanced with P205V increasing the 6-kestose yield 5-fold and generating 10% total FOS, the triple mutant P205V/W19Y/N21S synthesised 10-fold more 6-kestose than the wild-type and produced 18% total FOS [104]. The triple mutant P205V/W19Y/N21S had similar transfructosylating activity than the double variant W19Y/N21S although there was a significant decrease in the triple variant's global activity [104]. The explanation proposed by Lafraya *et al.,* [104] for the increase in transferase activity by P205V was the reduction in the structural rigidity conferred by the P. This change could affect C204 which forms a hydrogen bond with the transition state stabiliser and E203 which is the acid/base catalyst [104]. Lafraya *et al.,* [104] expanded on this explanation and suggested that the reduced rigidity influenced the orientation of the residues, affecting the pKa as well as the optimum pH of the enzyme.

The position N228 were substituted as follows N228R, N228A, N228R/N21S, N228A/N21S. All the N228 substitutions resulted in a substantial reduction in global activity except for N228A which had an increase in global activity compared to the wild-type [104]. The transferase activity of these mutants differed from each other with N228R producing no FOS and N228R/N21S producing low amounts of FOS [104]. The other N228 mutations were different, with N228A producing slightly higher FOS than the wild-type while N228A/N21S showed significantly increased FOS production (16.4% total FOS) [104]. The substitution N228A/N21S was the only mutation that changed the specificity of the enzyme producing equimolar 1-kestose and 6-kestose concentrations [104]. The specificity of N21S for 6-kestose was reversed when it was combined with N228A highlighting the intricate interaction between residues within the active site [104]. The substitutions of N228 highlighted this residue as a key amino acid in fixing sucrose in the specific orientation at the acceptor-substrate binding site as predicted by the docking studies [104]. The substitution of N228R severely interfered with the binding of sucrose and significantly affected the transferase activity. Lafraya *et al.,* [104] attributed the N228R variant's reduced transferase activity to the longer side chain of R. The substitution of N228A produced a different result than N228R, this observation was attributed to the smaller side-chain of A which meant the binding of the substrate could still take place. Lafraya *et al.* [104] highlighted the interaction of the substrate with W291 which served as a hydrophobic platform allowing the sugar ring to have broader orientation. Lafraya *et al.,* [104] hypothesised that this broader orientation of the sucrose resulted in the production of equimolar 1-kestose (β-(2→1) glycosidic bond) and 6-kestose (β-(2→6) glycosidic bond).

The above overview of the rational engineering research in yeast β-fructofuranosidases highlights the complexity of understanding what governs transferase/hydrolytic activity as well as the enzyme's selectivity. This review illustrates that although these enzymes are seemingly similar in structure and sequence, it does not necessarily correlate directly to the enzyme's activity or specificity. This overview of yeast β-fructofuranosidases highlighted some structural features prevalent in governing the activity and selectivity of the β-fructofuranosidases.

2.8.3 A bioinformatics tool to rapidly distinguish between high-level FOS synthesising and predominantly sucrose hydrolysing enzymes from fungal genomic data

Trollope *et al.,* [40] made use of a systematic review to gather information on the family of GH 32 enzymes, more specifically the fungal members within this family. The information consisted of functional, sequence and structural data. This data was used to investigate the differences/similarities within the fungal GH 32 family. The purpose was to differentiate

between fungal GH 32 enzyme that had high transferase activity and ones with low transferase activity.

The first phase was the systematic review which entailed the collection of functional data from fungal GH 32 enzymes acting only on sucrose which limited the enzymes to β-fructofuranosidases and fructosyltransferases. The second phase involved multiple sequence alignments and phylogenetic tree analyses that lead to distinct sequence groupings. These distinct groupings were divided into two groups: those enzymes that produced more than ~20% FOS and those enzymes producing limited amounts of FOS (less than 20%). The 20% threshold was selected as an arbitrary number to differentiate between high and low FOS producing enzymes.

Further analysis of the multiple sequence alignments highlighted that the high FOS producing group showed unique conserved motifs (GQIG**D**PC and F**E**T) whereas the low FOS producing enzymes contained the WMN**D**PNG and **E**CP motifs. The sequence alignment was extrapolated to a larger group that consisted of 258 fungal GH 32 enzymes to ensure that the difference witnessed in the smaller group was not an artefact. After the analysis, 32 protein sequences were identified and grouped, all containing the GQIG**D**P and F**E**T motif. These 32 sequences were realigned and using Hidden Markov Model and skylign to generate sequence logo's it showed that there were differences between the rest of the GH 32 family and the high FOS producing enzymes. There were only two motifs, the A and E, that were different, with motif D being similar. This meant that high FOS producers had a motif combination of GQIG**D**PC and F**E**T and low FOS producers motifs WMN**D**PNG and **E**CP. By using the motifs GQIG**D**PC and F**E**T the previously uncharacterised high FOS producer (*Aspergillus niger* Suc1 and SucA β-fructofuranosidase) was predicted and selected. The uncharacterised low FOS producer (*Aureobasidium pullulans* β-fructofuranosidase) was predicted and selected using the motifs WMN**D**PNG and **E**CP. These enzymes were assessed using various molecular techniques to determine their ability to form FOS. This prediction tool was successful in differentiating between high FOS and low FOS producing enzymes with *A. niger* Suc1 and SucA producing high levels of FOS and *A. pullulans* producing low levels of FOS.

Using only the primary structure of the enzymes Trollope *et al.,* [40] successfully predicted high and low FOS producing enzymes. Additionally, a physical clash was observed when the *Aspergillus japonicus* fructosyltransferase 3D structure co-crystallised with nystose (3LEM) [122] was superimposed with *S. cerevisiae* β-fructofuranosidase (4eqv) [6]. This physical clash between nystose from *A. japonicus* and a loop in Suc2 was experimentally confirmed as Suc2 could not produce nystose at detectable levels [40]. This loop was identified in the sequence of Suc2 from N42 to L52 and later became known as TI [40,108]. The loop identified in the

structural alignment was confirmed in the sequence alignment data, highlighting that the fungal GH 32 enzymes with low FOS production had a more extended sequence for this region conferring a longer loop.

In contrast, the fungal GH 32 enzymes with higher FOS production had a shorter sequence hence a shorter loop [40]. The identification of the difference between the two loop lengths led to the generation of the hypothesis for this thesis. It was hypothesised that this loop is responsible for the higher degree of polymerisation in the β-fructofuranosidases and their subsequent FOS production. This study set out to further investigate the role of the TI loop on the degree of FOS polymerisation by β-fructofuranosidases. We hypothesised the underlying reason why some β-fructofuranosidases are unable to form high DP FOS compared to others might be due to some spatial constraint imposed by the TI loop region. To prove or disprove the hypothesis, a rational design approach was employed to produce changes within the TI loop of *S. cerevisae*'s β-fructofuranosidases using both site-directed mutagenesis and homologous recombination. Site-directed mutagenesis was used to introduce a W48A amino acid substitution within *S. cerevisae*'s β-fructofuranosidases TI loop region. Furthermore, homologous recombination was used to shorten the longer TI loop in *S. cerevisae*'s β-fructofuranosidase by several amino acids or replaced the loop with the shorter loop sequence from the high FOS producing *A. japonicus* fopA enzyme.

2.9 References

[1] J. Wisniak, The History of Catalysis. From the beginning to Nobel prizes, Educ. Química. 21 (2010) 60–69. doi:10.1016/S0187-893X(18)30074-0.

[2] J.N. Armor, A history of industrial catalysis, Catal. Today. 163 (2011) 3–9. doi:10.1016/j.cattod.2009.11.019.

[3] U.T. Bornscheuer, G.W. Huisman, R.J. Kazlauskas, S. Lutz, J.C. Moore, K. Robins, Engineering the third wave of biocatalysis, Nature. 485 (2012) 185–194. doi:10.1038/nature11117.

[4] Y.H.P. Zhang, J. Sun, Y. Ma, Biomanufacturing: history and perspective, J. Ind. Microbiol. Biotechnol. 44 (2017) 773–784. doi:10.1007/s10295-016-1863-2.

[5] J. Steensels, K.J. Verstrepen, Taming wild yeast: Potential of conventional and nonconventional yeasts in industrial fermentations., Annu. Rev. Microbiol. (2014) 61–80. doi:10.1146/annurev-micro-091213-113025.

[6] C. Weizmann, GB191504845A.pdf, 1915.

[7] Y. Ni, Z. Sun, Recent progress on industrial fermentative production of acetone-butanol-ethanol by *Clostridium acetobutylicum* in China, Appl. Microbiol. Biotechnol. 83 (2009) 415–423. doi:10.1007/s00253-009-2003-y.

[8] G. Walsh, G. Walsh, Industrial Enzymes: Proteases and Carbrohydrases, Proteins. (2015) 327–369.

[9] J.M. Choi, S.S. Han, H.S. Kim, Industrial applications of enzyme biocatalysis: Current status and

future aspects, Biotechnol. Adv. 33 (2015) 1443–1454. doi:10.1016/j.biotechadv.2015.02.014.

[10] Y.P. Zhang, Production of biocommodities and bioelectricity by cell-free synthetic enzymatic pathway biotransformations : Challenges and Opportunities, 105 (2010) 663–677. doi:10.1002/bit.22630.

[11] H.E. Schoemaker, D. Mink, M.G. Wubbolts, Dispelling the Myths — Biocatalysis in Industrial Synthesis, 299 (2003) 1694–1698.

[12] R.A. Sheldon, J.M. Woodley, Role of biocatalysis in sustainable chemistry, Chem. Rev. (2017) acs.chemrev.7b00203. doi:10.1021/acs.chemrev.7b00203.

[13] B. Lindström, L.J. Pettersson, A brief history of catalysis, Cattech. 7 (2003) 130–138. doi:10.1023/A:1025001809516.

[14] P. Anastas, J.C. Warner, Green Chemistry: Theory and Practice, Oxford University Press: Oxford, U.K., 1998.

[15] J.H. Clark, T. Farmer, L. Herrero-Davila, J. Sherwood, Circular economy design considerations for research and development in the chemical sciences, Green Chem. 18 (2016) 3914–3934.

[16] U.T. Bornscheuer, The fourth wave of biocatalysis is approaching, Philos. Trans. A. Math. Phys. Eng. Sci. 376 (2018) 20170063. doi:10.1098/rsta.2017.0063.

[17] J.D. Tyzack, N. Furnham, I. Sillitoe, C.M. Orengo, J.M. Thornton, Understanding enzyme function evolution from a computational perspective, Curr. Opin. Struct. Biol. 47 (2017) 131–139. doi:10.1016/j.sbi.2017.08.003.

[18] J. Bertrán, A DARWINIAN PROCESS : THE MOLECULAR EVOLUTION OF ENZYMES, 64 (2008) 771–782.

[19] S.D. Copley, Shining a light on enzyme promiscuity, Curr. Opin. Struct. Biol. 47 (2017) 167–175. doi:10.1016/j.sbi.2017.11.001.

[20] H. Huang, C. Pandya, C. Liu, N.F. Al-Obaidi, M. Wang, L. Zheng, S. Toews Keating, M. Aono, J.D. Love, B. Evans, R.D. Seidel, B.S. Hillerich, S.J. Garforth, S.C. Almo, P.S. Mariano, D. Dunaway-Mariano, K.N. Allen, J.D. Farelli, Panoramic view of a superfamily of phosphatases through substrate profiling, Proc. Natl. Acad. Sci. 112 (2015) E1974–E1983. doi:10.1073/pnas.1423570112.

[21] C. Darwin, On the origin of species- by means of natural selection, Orig. Species. (1859) 440. doi:10.4324/9780203509104.

[22] F. Baier, J.N. Copp, N. Tokuriki, Evolution of enzyme superfamilies: Comprehensive exploration of sequence-function relationships, Biochemistry. 55 (2016) 6375–6388. doi:10.1021/acs.biochem.6b00723.

[23] S.D. Brown, P.C. Babbitt, New insights about enzyme evolution from large scale studies of sequence and structure relationships, J. Biol. Chem. 289 (2014) 30221–30228. doi:10.1074/jbc.R114.569350.

[24] N. Tokuriki, F. Stricher, L. Serrano, D.S. Tawfik, How protein stability and new functions trade off, PLoS Comput. Biol. 4 (2008) 35–37. doi:10.1371/journal.pcbi.1000002.

[25] M.A. DePristo, D.M. Weinreich, D.L. Hartl, Missense meanderings in sequence space: A biophysical view of protein evolution, Nat. Rev. Genet. 6 (2005) 678–687. doi:10.1038/nrg1672.

[26] J.B. Mitchell, Enzyme function and its evolution, Curr. Opin. Struct. Biol. 47 (2017) 151–156. doi:10.1016/j.sbi.2017.10.004.

[27] M.A. Siddiq, G.K. Hochberg, J.W. Thornton, Evolution of protein specificity: insights from ancestral protein reconstruction, Curr. Opin. Struct. Biol. 47 (2017) 113–122. doi:10.1016/j.sbi.2017.07.003.

[28] J. Damborsky, J. Brezovsky, Computational tools for designing and engineering enzymes, Curr. Opin. Chem. Biol. 19 (2014) 8–16. doi:10.1016/j.cbpa.2013.12.003.

[29] M.C. Ebert, J.N. Pelletier, Computational tools for enzyme improvement: why everyone can – and should – use them, Curr. Opin. Chem. Biol. 37 (2017) 89–96. doi:10.1016/j.cbpa.2017.01.021.

[30] N. Gurung, S. Ray, S. Bose, V. Rai, W.F. K, A Broader View : Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond, 2013 (2013).

[31] N.S. Scrutton, Speeding up enzyme engineering computationally, IUCrJ. 4 (2017) 5–6. doi:10.1107/S2052252516019692.

[32] K. Steiner, H. Schwab, Recent Advances in rational approaches for enzyme engineering, Comput. Struct. Biotechnol. J. 2 (2012) e201209010. doi:10.5936/csbj.201209010.

[33] Q. Liu, G. Xun, Y. Feng, The state-of-the-art strategies of protein engineering for enzyme stabilization, Biotechnol. Adv. (2018) #pagerange#. doi:10.1016/j.biotechadv.2018.10.011.

[34] R. Chen, Enzyme engineering: Rational redesign versus directed evolution, Trends Biotechnol. 19 (2001) 13–14. doi:10.1016/S0167-7799(00)01522-5.

[35] C. Wilson, D.A. Agard, Engineering substrate specificity, Curr. Opin. Struct. Biol. 1 (1991) 617–623. doi:10.1016/S0959-440X(05)80086-7.

[36] M. Lehmann, L. Pasamontes, S.F. Lassen, M. Wyss, The consensus concept for thermostability engineering of proteins, 1543 (2000) 408–415.

[37] J. Bendl, J. Stourac, E. Sebestova, O. Vavra, M. Musil, J. Brezovsky, J. Damborsky, HotSpot Wizard 2.0: automated design of site-specific mutations and smart libraries in protein engineering, Nucleic Acids Res. 44 (2016) W479–W487. doi:10.1093/nar/gkw416.

[38] M.S. Newton, V.L. Arcus, M.L. Gerth, W.M. Patrick, Enzyme evolution: innovation is easy, optimization is complicated, Curr. Opin. Struct. Biol. 48 (2018) 110–116. doi:10.1016/j.sbi.2017.11.007.

[39] Z. Xiao, H. Bergeron, S. Grosse, M. Beauchemin, M.L. Garron, D. Shaya, T. Sulea, M. Cygler, P.C.K. Lau, Improvement of the thermostability and activity of a pectate lyase by single amino acid substitutions, using a strategy based on melting-temperature- guided sequence alignment, Appl. Environ. Microbiol. 74 (2008) 1183–1189. doi:10.1128/AEM.02220-07.

[40] K.M. Trollope, N. van Wyk, M.A. Kotjomela, H. Volschenk, Sequence and structure-based prediction of fructosyltransferase activity for functional sub-classification of fungal GH32 enzymes., FEBS J. (2015) 1–15. doi:10.1111/febs.13536.

[41] S. Govindarajan, R. Recabarren, R.A. Goldstein, Estimating the total number of protein folds, Proteins Struct. Funct. Bioinforma. 35 (1999) 408–14. doi:10.1002/(SICI)1097-0134(19990601)35:4<408::AID-PROT4>3.0.CO;2-A.

[42] L.A. Mirny, E.I. Shakhnovich, Universally conserved positions in protein folds: Reading evolutionary signals about stability, folding kinetics and function, J. Mol. Biol. 291 (1999) 177–196. doi:10.1006/jmbi.1999.2911.

[43] A.S. Bommarius, J.K. Blum, M.J. Abrahamson, Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst, Curr. Opin. Chem. Biol. 15 (2011) 194–200. doi:10.1016/j.cbpa.2010.11.011.

[44] R. Rouet, D.B. Langley, P. Schofield, M. Christie, B. Roome, B.T. Porebski, Structural reconstruction of protein ancestry, (2017). doi:10.1073/pnas.1613477114.

[45] A. Cerdobbel, K. De Winter, D. Aerts, R. Kuipers, H.J. Joosten, W. Soetaert, T. Desmet, Increasing the thermostability of sucrose phosphorylase by a combination of sequence- and structure-based mutagenesis, Protein Eng. Des. Sel. 24 (2011) 829–834. doi:10.1093/protein/gzr042.

[46] A. Kukol, Molecular Modeling of Proteins, 2008. doi:10.1007/978-1-59745-177-2.

[47] Q. Zhang, J. Yang, K. Liang, L. Feng, S. Li, J. Wan, Binding interaction analysis of the active site and its inhibitors for neuraminidase (N1 Subtype) of human influenza virus by the integration

of molecular docking , FMO Calculation and 3D-QSAR CoMFA Modeling, (2008) 1802–1812.

[48] C.A.F. George A. Khoury, James Smadbeck, Chris A. Kieslich, Protein folding and de novo protein design for biotechnological applications, Trends Biotechnol. 32 (2015) 99–109. doi:10.1016/j.tibtech.2013.10.008.Protein.

[49] O. Trott, A. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, J. Comput. Chem. 31 (2010) 455–461. doi:10.1002/jcc.21334.AutoDock.

[50] L. Kang, Q. Guo, X. Wang, A hierarchical method for molecular docking using cloud computing, Bioorganic Med. Chem. Lett. 22 (2012) 6568–6572. doi:10.1016/j.bmcl.2012.09.016.

[51] D.A. Ostrov, J.A. Hernández Prada, P.E. Corsino, K.A. Finton, N. Le, T.C. Rowe, Discovery of novel DNA gyrase inhibitors by high-throughput virtual screening, Antimicrob. Agents Chemother. 51 (2007) 3688–3698. doi:10.1128/AAC.00392-07.

[52] M. Karplus, G.A. Petsko, Molecular dynamics simulations in biology, Nature. 347 (1990) 631.

[53] M.C. Childers, V. Daggett, Validating molecular dynamics simulations against experimental observables in light of underlying conformational ensembles, J. Phys. Chem. B. 122 (2018) 6673–6689. doi:10.1021/acs.jpcb.8b02144.

[54] M.C. Childers, V. Daggett, Insights from molecular dynamics simulations for computational protein design, Mol. Syst. Des. Eng. 2 (2017) 9–33. doi:10.1039/C6ME00083E.

[55] H.-L. Liu, W.-C. Wang, Protein engineering to improve the thermostability of glucoamylase from *Aspergillus awamori* based on molecular dynamics simulations, Protein Eng. Des. Sel. 16 (2003) 19–25. doi:10.1093/proeng/gzg007.

[56] A.M. Razavi, W.M. Wuest, V.A. Voelz, Computational screening and selection of cyclic peptide hairpin mimetics by molecular simulation and kinetic network models, J. Chem. Inf. Model. 54 (2014) 1425-1432 doi: 10.1021/ci500102y.

[57] S. Lutz, Beyond directed evolution-semi-rational protein engineering and design, Curr. Opin. Biotechnol. (2010). doi:10.1016/j.copbio.2010.08.011.

[58] T. Davids, M. Schmidt, D. Böttcher, U.T. Bornscheuer, Strategies for the discovery and engineering of enzymes for biocatalysis, Curr. Opin. Chem. Biol. 17 (2013) 215–220. doi:10.1016/j.cbpa.2013.02.022.

[59] L.G. Otten, F. Hollmann, I.W.C.E. Arends, Enzyme engineering for enantioselectivity: from trial-and-error to rational design?, Trends Biotechnol. 28 (2010) 46–54. doi:10.1016/j.tibtech.2009.10.001.

[60] H. Kries, R. Blomberg, D. Hilvert, De novo enzymes by computational design, Curr. Opin. Chem. Biol. 17 (2013) 221–228. doi:10.1016/j.cbpa.2013.02.012.

[61] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J.E. Sternberg, Europe PMC Funders Group The Phyre2 web portal for protein modelling, prediction and analysis, Nat. Protoc. 10 (2015) 845–858. doi:10.1038/nprot.2015.053.The.

[62] B.I. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics, Nucleic Acids Res. 37 (2009) 233–238. doi:10.1093/nar/gkn663.

[63] P.A. Romero, F.H. Arnold, Exploring protein fitness landscapes by directed evolution, Nat. Rev. Cell Biol. 10 (2009) 866–876. doi:doi:10.1038/nrm2805.

[64] R.A. Chica, N. Doucet, J.N. Pelletier, Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design, Curr. Opin. Biotechnol. (2005). doi:10.1016/j.copbio.2005.06.004.

[65] D.N. Bolon, C.A. Voigt, S.L. Mayo, De novo design of biocatalysts, Curr. Opin. Chem. Biol. 6 (2002) 125–129. doi:10.1016/S1367-5931(02)00303-4.

[66] N.J. Turner, Directed evolution drives the next generation of biocatalysts, Nat. Chem. Biol. 5 (2009) 567–573. doi:10.1038/nchembio.203.

[67] H.M. Li, L.H. Mei, V.B. Urlacher, R.D. Schmid, Cytochrome P450 BM-3 evolved by random and saturation mutagenesis as an effective indole-hydroxylating catalyst, Appl. Biochem. Biotechnol. 144 (2008) 27–36. doi:10.1007/s12010-007-8002-5.

[68] J.F. Chaparro-Riggers, K.M. Polizzi, A.S. Bommarius, Better library design: Data-driven protein engineering, Biotechnol. J. 2 (2007) 180–191. doi:10.1002/biot.200600170.

[69] W.P. Stemmer, Rapid evolution of a protein in vitro by DNA shuffling, Nature. (1994). doi:10.1038/370389a0.

[70] M.S. Packer, D.R. Liu, Methods for the directed evolution of proteins, Nat. Rev. Genet. 16 (2015) 379–394. doi:10.1038/nrg3927.

[71] A. Currin, N. Swainston, P.J. Day, D.B. Kell, Synthetic biology for the directed evolution of protein biocatalysts: Navigating sequence space intelligently, Chem. Soc. Rev. 44 (2015) 1172–1239. doi:10.1039/c4cs00351a.

[72] T.S. Wong, D. Zhurina, U. Schwaneberg, The diversity challenge in directed protein evolution, Comb. Chem. High Throughput Screen. 9 (2006) 271–88. doi:10.2174/138620706776843192.

[73] A. Crameri, S.A. Raillard, E. Bermudez, W.P.C. Stemmer, DNA shuffling of a family of genes from diverse species accelerates directed evolution, Nature. 391 (1998) 288–291. doi:10.1038/34663.

[74] M.T. Reetz, Biocatalysis in organic chemistry and biotechnology: Past, present, and future, J. Am. Chem. Soc. 135 (2013) 12480–12496. doi:10.1021/ja405051f.

[75] M.T. Reetz, M. Bocola, J.D. Carballeira, D. Zha, Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test, Angew. Chem. 117 (2005) 4264–4268. doi:10.1002/ange.200500767.

[76] Y.L. Boersma, T. Pijning, M.S. Bosma, A.M. van der Sloot, L.F. Godinho, M.J. Dröge, R.T. Winter, G. van Pouderoyen, B.W. Dijkstra, W.J. Quax, Loop grafting of *Bacillus subtilis* lipase A: Inversion of enantioselectivity, Chem. Biol. 15 (2008) 782–789. doi:10.1016/j.chembiol.2008.06.009.

[77] D. Quaglia, M.C.C.J.C. Ebert, P.F. Mugford, J.N. Pelletier, Enzyme engineering: A synthetic biology approach for more effective library generation and automated high-throughput screening, PLoS One. 12 (2017) 1–14. doi:10.1371/journal.pone.0171741.

[78] F. Sites, Computational Drug Discovery and Design, 819 (2012) 29–42. doi:10.1007/978-1-61779-465-0.

[79] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. 42 (2014) 490–495. doi:10.1093/nar/gkt1178.

[80] W. Lammens, K. Le Roy, L. Schroeven, A. Van Laere, A. Rabijns, W. Van Den Ende, Structural insights into glycoside hydrolase family 32 and 68 enzymes: Functional implications, J. Exp. Bot. 60 (2009) 727–740. doi:10.1093/jxb/ern333.

[81] J.C. Espín, M.T. García-Conesa, F.A. Tomás-Barberán, Nutraceuticals: Facts and fiction, Phytochemistry. 68 (2007) 2986–3008. doi:10.1016/j.phytochem.2007.09.014.

[82] C.G. Kumar, S. Sripada, Y. Poornachandra, Chapter 14 – Status and Future Prospects of Fructooligosaccharides as Nutraceuticals, Elsevier Inc., 2018. doi:10.1016/B978-0-12-811448-3.00014-0.

[83] D. Piedrabuena, N. Míguez, A. Poveda, F.J. Plou, M. Fernández-Lobato, Exploring the transferase activity of Ffase from *Schwanniomyces occidentalis*, a β-fructofuranosidase showing high fructosyl-acceptor promiscuity, Appl. Microbiol. Biotechnol. 100 (2016) 8769–8778. doi:10.1007/s00253-016-7628-z.

[84] M.E. Ortiz-Soto, C. Possiel, J. Görl, A. Vogel, R. Schmiedel, J. Seibel, Impaired coordination of nucleophile and increased hydrophobicity in the +1 subsite shift levansucrase activity towards transfructosylation, Glycobiology. 27 (2017) 755–765. doi:10.1093/glycob/cwx050.

[85] S. Cimini, L. Di Paola, A. Giuliani, A. Ridolfi, L. De Gara, GH32 family activity: a topological approach through protein contact networks, Plant Mol. Biol. 92 (2016) 401–410. doi:10.1007/s11103-016-0515-2.

[86] T. Pons, D.G. Naumoff, C. Martínez-Fleites, L. Hernández, Three acidic residues are at the active site of a β-propeller architecture in Glycoside Hydrolase families 32, 43, 62, and 68, Proteins Struct. Funct. Genet. 54 (2004) 424–432. doi:10.1002/prot.10604.

[87] L. Michaelis, M.L. Menten, Die Kinetik der Invertinwirkung, Biochem Z. 49 (1913) 333–369. doi:10.1021/bi201284u.

[88] D.E. Koshland, S. Stein, Correlation of bond breaking with enzyme specificity. Cleavage point of invertase, J. Biol. Chem. 139 (1954).

[89] V.A. Reddy, F. Maley, Identification of an Active-site residue in yeast invertase by affinity labeling and site-directed mutagenesis, J. Biol. Chem. 265 (1990) 10817–10821.

[90] S. Yuan, K. Le Roy, T. Venken, W. Lammens, W. van den Ende, M. de Maeyer, pKa modulation of the acid/base catalyst within GH32 and GH68: A role in substrate/inhibitor specificity?, PLoS One. 7 (2012) 1–12. doi:10.1371/journal.pone.0037453.

[91] A. Bujacz, M. Jedrzejczak-Krzepkowska, S. Bielecki, I. Redzynia, G. Bujacz, Crystal structures of the apo form of β-fructofuranosidase from *Bifidobacterium longum* and its complex with fructose, FEBS J. 278 (2011) 1728–1744. doi:10.1111/j.1742-4658.2011.08098.x.

[92] M. Antošová, M. Polakovic, Fructosyltransferases : The enzymes catalyzing production, Chem. Pap. 55 (2001) 350–358.

[93] C.M. Hasler, Functional foods: benefits, concerns and challenges-a position paper from the American Council on Science and Health., J. Nutr. 132 (2002) 3772–81. doi:10.1002/mus.20330.

[94] M. Gimeno-Perez, P. Santos-Moriano, L. Fernandez-Arrojo, A. Poveda, J. Jimenez-Barbero, A.O. Ballesteros, M. Fernandez-Lobato, F.J. Plou, Regioselective synthesis of neo-erlose by the β-fructofuranosidase from *Xanthophyllomyces dendrorhous*, 2014. doi:10.1016/j.procbio.2013.12.018.

[95] I. Siró, E. Kápolna, B. Kápolna, A. Lugasi, Functional food. Product development, marketing and consumer acceptance-A review, Appetite. 51 (2008) 456–467. doi:10.1016/j.appet.2008.05.060.

[96] I. Vijn, S. Smeekens, Fructan: More than a reserve carbohydrate?, Plant Physiol. 120 (1999) 351–360. doi:10.1104/pp.120.2.351.

[97] B. Lasseur, J. Lothier, A. Djoumad, B. De Coninck, S. Smeekens, A. Van Laere, A. Morvan-Bertrand, W. Van Den Ende, M.P. Prud'homme, Molecular and functional characterization of a cDNA encoding fructan:fructan 6G-fructosyltransferase (6G-FFT)/fructan:fructan 1-fructosyltransferase (1-FFT) from perennial ryegrass (*Lolium perenne* L.), J. Exp. Bot. 57 (2006) 2719–2734. doi:10.1093/jxb/erl034.

[98] F.R. Bornet, F. Brouns, Y. Tashiro, V. Duvillier, Nutritional aspects of short-chain fructooligosaccharides: natural occurrence, chemistry, physiology and health implications, Dig Liver Dis. 34 Suppl 2 (2002) S111-20. doi:10.1016/S1590-8658(02)80177-3.

[99] J. Marín-Navarro, D. Talens-Perales, J. Polaina, One-pot production of fructooligosaccharides by a *Saccharomyces cerevisiae* strain expressing an engineered invertase, Appl. Microbiol. Biotechnol. 99 (2015) 2549–55. doi:10.1007/s00253-014-6312-4.

[100] G.L.E.Y.Y.R. Gibson, M.B. Roberfroid, Dietary modulation of the human colonic microbiota: Introducing the concept of prebiotics, J. Nutr. 125 (1995) 1401–12.

[101] S. Kilian, S. Kritzinger, C. Rycroft, G. Gibson, J. Preez, The effects of the novel bifidogenic

trisaccharide, neokestose, on the human colonic microbiota, (2002) 637–644.

[102] W. Li, K. Wang, Y. Sun, H. Ye, B. Hu, X. Zeng, Influences of structures of galactooligosaccharides and fructooligosaccharides on the fermentation in vitro by human intestinal microbiota, J. Funct. Foods. 13 (2015) 158–168. doi:10.1016/j.jff.2014.12.044.

[103] W. Li, S. Yu, T. Zhang, B. Jiang, W. Mu, Recent novel applications of levansucrases, Appl. Microbiol. Biotechnol. 99 (2015) 6959–6969. doi:10.1007/s00253-015-6797-5.

[104] A. Lafraya, J. Sanz-Aparicio, J. Polaina, J. Marin-Navarro, Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase, Appl. Environ. Microbiol. 77 (2011) 6148–6157. doi:10.1128/AEM.05032-11.

[105] K. Le Roy, W. Lammens, M. Verhaest, B. De Coninck, A. Rabijns, A. Van Laere, W. Van den Ende, Unraveling the difference between invertases and fructan exohydrolases: A single amino acid (Asp-239) substitution transforms *Arabidopsis* cell wall Invertase1 into a fructan 1-exohydrolase, Plant Physiol. 145 (2007) 616–625. doi:10.1104/pp.107.105049.

[106] J. De Roover, A. Van Laere, M. De Winter, J.W. Timmermans, W. Van Den Ende, Purification and properties of a second fructan exohydrolase from the roots of *Cichorium intybus*, Physiol. Plant. 106 (1999) 28–34. doi:10.1034/j.1399-3054.1999.106104.x.

[107] M.A. Sainz-Polo, M. Ramírez-Escudero, A. Lafraya, B. Gonza, J. Marín-Navarro, J. Polaina, J. Sanz-aparicio, Three-dimensional structure of *Saccharomyces* Invertase: Role of a non-catalytic domain in oligomerization and substrate specificity, J. Biol. Chem. 288 (2013) 9755–9766. doi:10.1074/jbc.M112.446435.

[108] M. Ramírez-Escudero, M. Gimeno-Perez, B.G. Lez, D. Linde, Z. Merdzo, M. Fernández-Lobato, J. Sanz-Aparicio, Structural analysis of β-fructofuranosidase from *Xanthophyllomyces dendrorhous* reveals unique features and the crucial role of N-Glycosylation in oligomerization and activity, J. Biol. Chem. 291 (2016) 6843–6857. doi:10.1074/jbc.M115.708495.

[109] M. Álvaro-Benito, A. Polo, B. González, M. Fernández-Lobato, J. Sanz-Aparicio, Structural and kinetic analysis of *Schwanniomyces occidentalis* invertase reveals a new oligomerization pattern and the role of its supplementary domain in substrate binding, J. Biol. Chem. 285 (2010) 13930–13941. doi:10.1074/jbc.M109.095430.

[110] M. Jaiswal, In silico mining of simple sequence repeats in whole genome of *Xanthomonas* sp., J. Comput. Sci. Syst. Biol. 7 (2014) 203–208. doi:10.4172/jcsb.1000157.

[111] M. Álvaro-Benito, M. de Abreu, L. Fernández-Arrojo, F.J. Plou, J. Jiménez-Barbero, A. Ballesteros, J. Polaina, M. Fernández-Lobato, Characterization of a β-fructofuranosidase from *Schwanniomyces occidentalis* with transfructosylating activity yielding the prebiotic 6-kestose, J. Biotechnol. 132 (2007) 75–81. doi:10.1016/j.jbiotec.2007.07.939.

[112] M. Álvaro-Benito, M.A. Sainz-Polo, D. González-Pérez, B. González, F.J. Plou, M. Fernández-Lobato, J. Sanz-Aparicio, Structural and kinetic insights reveal that the amino acid pair Gln-228/Asn-254 modulates the transfructosylating specificity of *Schwanniomyces occidentalis* β-fructofuranosidase, an enzyme that produces prebiotics, J. Biol. Chem. 287 (2012) 19674–19686. doi:10.1074/jbc.M112.355503.

[113] F. Alberto, E. Jordi, B. Henrissat, M. Czjzek, Crystal structure of inactivated *Thermotoga maritima* invertase in complex with the trisaccharide substrate raffinose, Biochem. J. 395 (2006) 457–462. doi:10.1042/BJ20051936.

[114] G.J. Davies, K.S. Wilson, B. Henrissat, Nomenclature for sugar-binding subsites in glycosyl hydrolases, Biochem. J. 321 (1997) 557–559. doi:10.1007/s007920050009.

[115] V.A. Reddy, R.S. Johnson, K. Biemann, R.S. Williams, F.D. Ziegler, R.B. Trimble, F. Maley, Characterization of the glycosylation sites in yeast external invertase. I. N-linked oligosaccharide content of the individual sequons, J. Biol. Chem. 263 (1988) 6978–6985.

[116] A.J.J. Straathof, A. Kieboom, H. Van Bekkum, Invertase-catalysed fructosyl transfer in concentrated solutions of sucrose, Carbohydr. Res. 146 (1986) 154–159.

[117] M. Batista, K. Simo, C. De Almeida, R. Aparecida, B. Pessoni, M. Regina, L. Figueiredo-Ribeiro, R. De Ca, Production of 6-kestose by the filamentous fungus *Gliocladium virens* as affected by sucrose concentration, 4 (2012) 2–9. doi:10.1016/j.myc.2012.09.012.

[118] S.I. Mussatto, J. a. Teixeira, Increase in the fructooligosaccharides yield and productivity by solid-state fermentation with *Aspergillus japonicus* using agro-industrial residues as support and nutrient source, Biochem. Eng. J. 53 (2010) 154–157. doi:10.1016/j.bej.2010.09.012.

[119] J.W. Yun, Fructooligosaccharides—Occurrence, preparation, and application, Enzyme Microb. Technol. 19 (1996) 107–117. doi:10.1016/0141-0229(95)00188-3.

[120] D. Altenbach, E. Rudiño-Pinera, C. Olvera, T. Boller, A. Wiemken, T. Ritsema, An acceptor-substrate binding site determining glycosyl transfer emerges from mutant analysis of a plant vacuolar invertase and a fructosyltransferase, Plant Mol. Biol. 69 (2009) 47–56. doi:10.1007/s11103-008-9404-7.

[121] A.M. Vandamme, C. Michaux, A. Mayard, I. Housen, Asparagine 42 of the conserved endo-inulinase INU2 motif WMNDPN from *Aspergillus ficuum* plays a role in activity specificity, FEBS Open Bio. 3 (2013) 467–472. doi:10.1016/j.fob.2013.10.009.

[122] P. Chuankhayan, C.-Y. Hsieh, Y.-C. Huang, Y.-Y. Hsieh, H.-H. Guan, Y.-C. Hsieh, Y.-C. Tien, C.-D. Chen, C.-M. Chiang, C.-J. Chen, Crystal structures of *Aspergillus japonicus* fructosyltransferase complex with donor/acceptor substrates reveal complete subsites in the active site for catalysis, J. Biol. Chem. 285 (2010) 23251–23264. doi:10.1074/jbc.M110.113027.

CHAPTER 3

RESEARCH RESULTS

The rational redesign of a loop region of *Saccharomyces cerevisiae* β-fructofuranosidase

Chapter 3: The rational redesign of a loop region of *Saccharomyces cerevisiae* β-fructofuranosidase

J. Gibbon[a], K.M. Trollope[a], H. Volschenk[a]

[a]*Department of Microbiology, University of Stellenbosch, Private Bag X1, Matieland, 7602, South Africa*

3.1 Abstract

The *Saccharomyces cerevisiae* β-fructofuranosidase (Suc2), also known as invertase, catalyses the hydrolysis of sucrose and fructooligosaccharides (FOS). At a high sucrose concentration, Suc2 displays transfructosylating activity which results in the synthesis of FOS. Suc2 produces predominantly 6-kestose (DP 3) and to a lesser extent 1-kestose (DP 3). Suc2 has a regiospecificity towards the formation of a β-(2-6)-glycosidic bond with the fructose of sucrose. During this study the TI loop region in Suc2 was mutated using homologous recombination to introduce changes to the TI loop region, producing the variants W48A, PGPL and NFS. The changes in the enzyme variants allowed them to produce higher DP FOS including nystose (DP4) and 1^F-fructofuranosyl nystose (DP5). The activities of the enzyme variants were changed from being predominantly hydrolytic to being predominantly tranfructosylating, resulting in an increase in total FOS production by 3 to 4 times compared to Suc2. The mutation to the TI loop influenced the regiospecificity of the variants, as the variants produced more than 3-times neokestose, with the W48A variant producing 10-times more neokestose than Suc2. Results from this study confirm that the TI loop is one of the contributing factors that differentiate between enzymes that have predominantly high transfructosylating activity compared to those that do not present this activity. The changes to the TI loop resulted in the Suc2 enzyme changing the prevailing activity from hydrolytic to transfructosylating. Furthermore, this loop plays a crucial role in determining product formation as it affected both the DP of FOS and the regiospecificity.

3.2 Introduction

β-fructofuranosidases, also known as invertases (EC 3.2.1.26), catalyse the hydrolysis of sucrose but under high substrate conditions display transfructosylating activity which results in the synthesis of fructooligosaccharides (FOS) [1,2]. Combined with a degree of catalytic promiscuity displayed by β-fructofuranosidases, i.e. the ability to bind and fructosylate a broad range of substrates, these enzymes are of particular biotechnological importance with applications in the functional food and pharmaceutical industries [3].

β-fructofuranosidases are classified within the GH 32 family of Carbohydrate-Active enZYmes (CAZy) based on amino acid sequence together with several other invertases, inulinases and levanases [4]. Based on enzyme structural features, β-fructofuranosidases belong to the GH-J clan which share a β-propeller domain consisting of four anti-parallel β-strands and a central negatively charged cavity with three conserved amino acids [1]. These three amino acids are referred to as the catalytic triad which consists of a nucleophile Asp, an acid-base catalyst Glu and a transition state stabiliser Asp located in the conserved motifs WMN**D**PNG, **E**C and R**D**P, respectively [1]. Catalysis by β-fructofuranosidases follows a double displacement mechanism that consists of two steps. Firstly, the enzyme binds the donor substrate, either sucrose or fructan, in the active site resulting in the protonation of the donor substrate by Glu followed by a nucleophilic attack from Asp which forms a fructosyl-enzyme intermediate. The second step involves the binding of an acceptor molecule, sucrose or water, which is protonated and acts as a nucleophile that binds to the fructosyl-enzyme intermediate thus releasing either free fructose or fructooligosaccharides, respectively [5]. Due to this mechanism, under high sucrose concentrations, the fructosyltransferase reaction is favoured as a result of the lower water activity and due to water and sucrose competing for the same binding site [6].

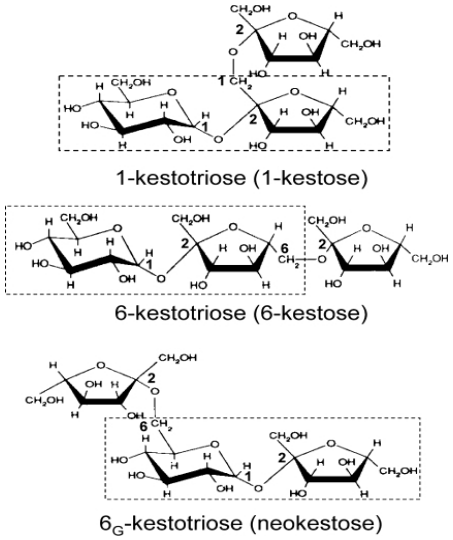

Figure 3.1. The three predominant types of FOS being produced by GH 32 and GH 64 enzymes. From the top inulin-type FOS with 1-kestose being the first to from consisting of fructose bound to fructose with a β-(2-1) glycosidic bond. The second is a levan-type FOS known as 6-kestose which is fructose bound to fructose with a β-(2-6) glycosidic bond. At the bottom is the neo-series FOS with neokestose which is fructose bound to the glucose with a β-(2-6) glycosidic bond. The image is adapted from [7].

The resulting FOS consist of a fructosyl residue bound to sucrose [8]. FOS can differ structurally in several ways, one of which is the difference in the binding orientation of the fructan to the sucrose [6]. The main glycosidic bonds include β-(2-1)-linked fructose (1-kestose or inulin-type), β-(2-6)-linked fructose (6-kestose or levan-type) and β-(2-6)-linked glucose (neokestose or neo-FOS) all of which are isomers (Fig. 3.1). The formation of these isomers is determined by the orientation of the acceptor, which in turn is dependent on the interactions of specific amino acids surrounding the active site [9]. The enzymes produce mixtures of these isomers favouring a particular type of FOS, i.e., favouring the formation of levan-type FOS, inulin-type FOS or neo-type FOS [10–12]. Neo-FOS (neokestose and neonystose) is formed when the glucose of the acceptor molecule binds to the donor fructose via a β-(2-6)-linked glucose [13].

Xanthophyllomyces dendrorhous is a well-known producer of neokestose and in 2016 the crystal structure of the *X. dendrorhous* β-fructofuranosidase was resolved. The mode of action of the enzyme was also proposed with amino acid W105 (5ANN amino acid numbering) involved in the stacking and thus the orientation of the acceptor sucrose [14]. The formation of β-(2-6)-glycosidic bond is common for β-fructofuranosidases from the yeasts *Saccharomyces cerevisiae*, *Schwanniomyces occidentalis* and *X. dendrorhous* [15]. These β-fructofuranosidase enzymes have a regiospecificity for producing 6-kestose or neokestose, thus favouring the orientation of the acceptor sucrose to form a β-(2-6)-link with fructose/glucose. The amino acid N228 (numbering according to *S. occidentalis*) plays an important role in enzyme functionality in that an N228R substitution had a detrimental effect on transfructosylating activity in β-fructofuranosidases favouring the production of 6-kestose. However, with an N228A substitution combined with N21S, the enzyme produces equimolar concentrations of 1-kestose and 6-kestose. The latter highlights that amino acid N228 plays a critical role in orientating the sucrose acceptor molecule in the binding site to form the β-(2-6)-linked bond [10].

β-fructofuranosidases do not terminate their reaction after the formation of kestose (GF_2), but can add additional fructose molecules to the fructans increasing the degree of polymerisation (DP) [16]. The increase in DP occurs when the acceptor molecule 1-kestose (DP3) is used forming nystose (DP4), which then proceeds to form 1^F-fructofuranosyl nystose (DP5) using nystose as acceptor molecule. An increase in DP does not only refer to inulin-type FOS but can also occur in the other FOS isomers such as neonystose. Different β-fructofuranosidases produce a range of FOS of different DP. The fructosyltransferase from *Aspergillus japonicus* fructosyltransferase (AjFFase) produces a range of FOS from DP 3 – 6, while other fungal enzymes can produce DP 3 – 13 [17]. Yeast β-fructofuranosidases have only a limited ability

to produce DP 4 and 5 with *X. dendrorhous* having the ability to produce neonystose (DP4) [18]. Detection or quantification of FOS with a DP higher than 3 has not been shown for the *S. cerevisiae* β-fructofuranosidase, i.e. no formation of nystose (DP4) or 1^F-fructofuranosyl nystose (DP5) [19].

In the GH 32 family, there are β-fructofuranosidases (e.g. fructosyltransferase from *A. japonicus*) with a relatively high transferase activity with low hydrolytic activity as well as enzymes where the hydrolytic reaction is favoured (β-fructofuranosidases such as Suc2, from *S. cerevisiae*). Trollope *et al.,* [20] developed a sequence/structural based sub-classification system for fungal β-fructofuranosidases and identified a loop region, known as TI [14], between β-sheet B1 and β-sheet C1 within the β-propeller domain, that plays a potential role in determining the fructosyltransferase activity of β-fructofuranosidases. In β-fructofuranosidases which favour sucrose hydrolysis the TI loop is longer in amino acid length, compared to a shorter loop region in β-fructofuranosidases displaying higher transfructosylating activity [20].

This study set out to further investigate the role of the TI loop on the degree of FOS polymerisation by β-fructofuranosidases. We hypothesised the underlying reason for some β-fructofuranosidases being unable to form high DP FOS might be due to some spatial constraint imposed by the TI loop region. Firstly, employing a rational design approach, a W48A amino acid change within Suc2 loop region using site-directed mutagenesis was made, i.e. a bulky amino acid W was substituted with a less bulky amino acid A in the TI loop region. Secondly, using homologous recombination the longer TI loop in Suc2 was shortened by several amino acids or replaced with the shorter loop sequence from AjFFase. The effects of these changes on the transferase activity, thermostability, regiospecificity and substrate affinity of the enzyme were evaluated.

3.3 Materials and Methods

3.3.1 Culturing conditions, microbial strains and plasmids

All microbial strains with their respective genotypes used in this study are listed in Table 3.1. *Escherichia coli* DH5α was used for cloning and amplification of plasmids. *E. coli* cells were grown at 37°C in Luria Bertani (LB) broth (New England Biolabs, Midrand, South Africa) supplemented with either 100 µg/ml ampicillin or 50 µg/ml kanamycin. *S. cerevisiae* EUROSCARF Y12321 (*SUC2* gene knockout) was used as the expression and homologous recombination host. *S. cerevisiae* strains were grown in Yeast Peptone Dextrose (YPD) medium (10 g/l yeast extract, 20 g/l bactopeptone and 20 g/l glucose) or in Synthetic

Complete (SC) medium containing 20 g/l glucose, 6.7 g/l Yeast Nitrogen Base without amino acids (Difco Laboratories, Detroit, USA) and 1.3 g/l amino acid pool without uracil [21]. *S. cerevisiae* transformants were selected on solid SC^{-ura} medium plates after 72h of growth. Solid media for the growth of bacteria and yeast contained 15 g/l bacteriological agar (New England Biolabs, Midrand, South Africa). For maximal enzyme production, *S. cerevisiae* EUROSCARF Y02321 was cultivated for 72h in double-strength SC^{-ura} medium containing 20 g/l glucose, 13.4 g/l yeast nitrogen base (YNB) without amino acids and 2.6 g/l amino acids pool without uracil. The media were buffered with 20 g/l succinic acid and the pH adjusted to 6.0 using 10 N KOH. *X. dendrorhous* ATCC 24230 was grown on Maltose Minimal Media (MMM) (0.7 g/l Yeast Nitrogen Base without amino acids (Difco Laboratories, Detroit, USA), 2 g/l maltose).

Table 3.1. Microbial Strains and Plasmids

Strain or plasmid	Relevant genotype or construct	Reference or source
Bacterial strains		
Escherichia coli DH5α	F$^-$ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ$^-$	New England Biolabs, Midrand, South Africa
Yeast strains		
Saccharomyces cerevisiae EUROSCARF Y12321	BY4742; *MATα; ura3Δ0; leu2Δ0; his3Δ1; lys2Δ0; YIL162w::kanMX4*	[22]
Saccharomyces cerevisiae BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	[22]
Xanthophyllomyces dendrorhous: UCD 67-385	*ATCC 24230, wild-type*	ATCC 24230 [23]
Plasmids		
pJC1	*bla URA3 PGK1*$_P$*-PGK1*$_T$	[24] Invitrogen, Life Technologies
pJET1.2	Ap^R, P$_{lacUV5}$,*eco47IR*, ColE1-replicon	Thermo Fisher Scientific
pJET*SUC2*	Ap^R, P$_{lacUV5}$,*eco47IR*, ColE1-replicon *SUC2*	This study
pJC1*SUC2*	*bla URA3 PGK1*$_P$*-xln2-SUC2-PGK1*$_T$	This study
pET-24d(+)*suc2NFS*(N42-L52/ HDGDG)	*kanMX T7*$_P$*-xln2-suc2-NFS-T7*$_T$	This study
pJC1*suc2PGPL (N42-L52/ NPGPL)*	*bla URA3 PGK1*$_P$*-xln2-suc2*-NPGPL-*PGK1*$_T$	This study
pJC1*suc2FS (N42-L52/ VGFLHDGDGIAG)*	*bla URA3 PGK1*$_P$*-xln2-suc2*-V75-G85-*PGK1*$_T$	This study
pJC1*suc2AW (W48A)*	*bla URA3 PGK1*$_P$*-xln2-suc2*-W48A-*PGK1*$_T$	This study
pJC1*suc2NFS*(N42-L52/HDGDG)	*bla URA3 PGK1*$_P$*-xln2-suc2*-HDGDG-*PGK1*$_T$	This study
pJC1*fopA*	*bla URA3 PGK1*$_P$*-xln2-fopA- PGK1*$_T$	This laboratory

3.3.2 Cloning of the β-fructofuranosidase gene

3.3.2.1 PCR and DNA manipulation

All PCR reactions were set up using standard protocols as recommended by the Kapa Biosystems (Boston, US). The conditions for all the PCR reactions (KAPA HiFi and KAPA2G) were kept standard according to the manufacturer's instructions with the optimised annealing temperature of 62.5°C for 30 cycles (Kapa Biosystems, Boston, US). For cloning purposes, PCR reactions were scaled to 50 µl reaction whereas 12.5 µl PCR reactions were used for diagnostic gene detection. All amplified genes were analysed on 1% agarose gel with 1x TAE buffer and running at 100V.

3.3.2.2 Restriction enzyme-based cloning

The *SUC2* gene (NCBI Gene ID 854644 and Genbank: NC_001141.2) was isolated from *S. cerevisiae* BY4742, and the mutated *SUC2* gene (NFS) was synthesised (Inqaba Biotechnology, Pretoria, South Africa). *S. cerevisiae* BY4742 gDNA served as a template to amplify the wild-type *SUC2* gene (SGD Systematic Name: YIL162W; NCBI Gene ID: 854644; SGD ID: S000001424; Genbank: NC_001141.2) without its native secretion signal using primers F'*SUC2* and R'*SUC2*. The 1599 bp *SUC2* PCR amplicon was purified after gel electrophoresis using the Zymoclean gel isolation kit (Zymo Research Corp, Irvine, CA). The *SUC2* fragment was blunt-end cloned into the pJET1.2 cloning vector (Fermentas, MA, USA) to give pJET*SUC2*. pJET*SUC2* was digested with *Bgl*II and *Xho*I to obtain the *SUC2* gene fragment for ligation into *BglII-XhoI* pJC1*fopA* [25] using T4 DNA ligase (Fermentas, Waltham, USA) producing pJC*SUC2*. The synthetic gene was supplied cloned into pET-24d(+) and is listed in Table 3.1. Using the restriction enzymes *Eco*RI and *Xho*I the *SUC2*-NFS gene with the *xln*2 secretion signal and poly-HIS-tag (6x) was isolated from the pET-24d(+)*suc2*(N42-L52/HDGDG) plasmid and ligated into the pJC1 plasmid to produce pJC1*suc2(N42-L52/HDGDG)*.

3.3.3 Mutagenesis of β-fructofuranosidase and cloning by *in vivo* homologous recombination

Constructing the three variants of *SUC2* (AW, FS and PGPL) a PCR-based method with the primers (R'*AW*, F'*PGPL*, F'*FS*) listed in Table 3.2 was used to introduce the mutations into the *SUC2* gene. The primers in Table 3.2 are annotated as follows: the underlined sequences represent the homologous sequence needed for recombination between the PCR product and expression vector pJC1. The sequence in bold represents the specific sequence modifications. The primers included restriction enzyme sites, these sites are depicted with the

symbol: |, and is unique to each variant. The sequence from the 3'-prime end to the start of the mutation (shown in italics) provided homology to *SUC2* gene for the first cycle of the PCR reaction. The method entailed using the forward primer F'PgpI or F'Fs in conjunction with the R'*SUC2* primer to amplify the linear product needed for homologous recombination [26]. As for AW, to generate the linear product needed for homologous recombination the pMBRE4-1'F forward primer in conjunction with the R'*AW* reverse primer was used. After the amplification of FS and AW, the reactions were digested with *Dpn*I to eliminate plasmid template DNA used in the PCR step. Following the *Dpn*I digestion, the PCR products were checked with gel electrophoresis, followed by gel extraction with the High Pure PCR purification kit (Roche, Mannheim, Germany).

The vector pJC1*SUC2* was linearised with the combination of either *Age*I and *Bcl*I or *EcoR*I and *Bgl*II. The former restriction enzyme combination was used for the PGPL and FS, and the latter was used for AW. Using electroporation [27], the linearised pJC1*SUC2* was co-transformed with the amplified PCR products listed above with a ratio of vector: insert of 1:5 into *S. cerevisiae* EUROSCARF Y02321. After the transformation, the cells were plated on SC^{-ura}. After 72 h incubation, transformants were restreaked on SC^{-ura} and incubated for 72h.

3.3.3.1 Mutation confirmation and sequencing

Single transformed *S. cerevisiae* colonies were inoculated in 5 ml SC^{-ura} media and grown overnight, plasmids were extracted using the rapid isolation of yeast chromosomal DNA method [28]. After extraction, plasmids were amplified in *E. coli* transformed using high-efficiency transformation by electroporation [29]. *E. coli* extracted plasmids were digested with *Sma*I (PGPL variant), *Nae*I (AW and FS variant) or *Sal*I (NFS variant) to confirm the presence of mutations. Sanger sequencing of the plasmids was used to confirm the accuracy of PCR and homologous recombination [30]. Sequence results were analysed using CLC Main Workbench Version 7.6.2 (Qiagen, Venlo, Netherlands).

Table 3.2. Primers for the homologous recombination and *SUC2* amplification

Primer name	Restriction enzyme site	Template	Primer sequence
Primers for *SUC2* amplification			
F'SUC2	*Bgl*II	SUC2 Suc2 gDNA	5'-CGA GAT CTA TGA CAA ACG AAA CTA GCG ATA GACC-3'
R'SUC2	*Xho*I	SUC2 Suc2 gDNA	5'-GCC ACT CGA GCT ATT TTA CTT CCC TTA CTTG-3'
Primers for homologous recombination			
R'AW W48A-pJC1- reverse	*Nae*I	pJC1 SUC2-SUC2	5'-CCA AAA CAA TGG CGT (GCC|**GGC)** TAC GGT GTC ATT-3'

F'PgpI N42-L52/NPGPL-pJC1-forward	*Sma*I	pJC1 SUC2-SUC2	5'-GAA AAA GAT GCC AAA TGG CAT CTG TAC TTT CAA TAC AAC (**CCC**|**GGG) CCA TTG** *TTT TGG GGC CAT GCT ACT TCC*-3'
F'Fs N42-L52/V75-G85-pJC1-forward	*Nae*I	pJC1 SUC2-SUC2	5'-GAA AAA GAT GCC AAA TGG CAT CTG TAC TTT CAA TAC AAC **GTG GGC TTT CTG CAT GAT GGC GAT GGC ATT (GCC**|*GGC) CAT GCT ACT TCC GAT GAT TTG-3*'
Primer for sequencing			
pMBRE4-1'F			5'-gtt tag tag aac ctc gtg aaa ctt a -3'
pMBRE4-3'R			5'-act taa aat acg ctg aac ccg aac at-3'

Underline = homology sequence to Suc2*SUC2*. **Bold** = New sequence (changed sequence) (|) = restriction sites

3.3.4 Expression and purification of β-fructofuranosidase enzyme variants

After plasmids were sequenced, *S. cerevisiae* EUROSCARF Y02321 was transformed with one of the following plasmids pJC1*SUC2*, pJC1suc2PG, pJC1*suc2*FS, pJC1*suc2*WA or pJC1*suc2*NFS using the lithium acetate method [31]. *S. cerevisiae* transformants were selected on SC^{-ura}, and the successful transformation was confirmed with colony PCR [32]. Selected transformants were streaked on SC^{-ura} was grown at 30°C for 72 h. Following their growth, a single colony was selected and inoculated in 5 ml of SC^{-ura} media and grown overnight. After that, 500 ml of 2x SC^{-ura} buffered with 20 g/l succinic acid was inoculated with the culture and incubated for 72h at 30°C.

3.3.4.1 Protein purification

Following 72h of recombinant protein expression, the cultures were harvested by centrifugation at 6 000 rpm, at 4°C for 10min. The supernatant was filtered through a 0.45 µm cellulose acetate filter (Sartorius, Goettingen, Germany). The following steps were performed on ice. The flow-through was concentrated using a Millipore 10 000 MWCO membrane coupled to an Amicon (model 8050) ultrafiltration cell (Millipore Co., Bedford, USA). With the total volume passed through the membrane the protein was concentrated 50x and resuspended in 10 ml of 20 mM sodium phosphate and 50 mM NaCl, pH 8 buffer. Complete EDTA-free Protease Inhibitor Cocktail (Roche, Mannheim, Germany) was added to 10 ml of concentrated crude protein sample and was stored at 4°C. The 10 ml of concentrated protein was loaded onto a Qiagen 1 ml superflow IMAC column (Qiagen, MD, USA) pre-equilibrated with NPI 0 (sodium phosphate 50 mM, sodium chloride 300 mM, and imidazole 0 mM, pH 8) at a flow rate of 1 ml/min. The column was placed on the ÄKTA (ÄKTA Purifier Explorer,

Pharmacia, Uppsala, Sweden) purifier and the following program was executed: 10 Column Volumes of NPI 0 at 1 ml/min was followed by a wash step of 10 column volumes (CV) of 10% (v/v) NPI 250 (sodium phosphate 50 mM, sodium chloride 300 mM and imidazole 250 mM, pH 8) and 90% (v/v) NPI 0 at 1 ml/min. The His-tag protein was eluted with 100% (v/v) NPI 250 starting at a block volume of 0.6 CV and ending at 2.6 CV that equates to a final elution volume of 2 ml at 1 ml/min. This process was repeated twice with each of the crude protein preparations (Suc2, PGPL, AW, FS, and NFS) but between the different proteins the columns were stripped and recharged with nickel (II) chloride (Sigma-Aldrich, St. Louis, USA) to prevent cross-contamination. The eluted proteins were subjected to SDS-PAGE and stained with Coomassie (10% (v/v) acetic acid, 0.006% (w/v) Coomassie brilliant blue G-250 and 90% (v/v) H_2O) [33]. After verifying that the samples were pure, the two fractions were pooled and transferred to Spectra/Por dialysis membrane 3500 MWCO (Spectrum Inc., CA, USA). The pooled protein samples were dialysed overnight in 10 mM sodium phosphate buffer (pH 6) with 50 mM NaCl and stored at 4°C until further analysis.

3.3.4.2 SDS-PAGE analysis

Samples were mixed with 2x loading buffer (60 mM Tris-HCl (pH 6.8), 25% (v/v) glycerol, 2% (v/v) SDS, 14 mM β-mercaptoethanol and bromophenol blue) and boiled at 100°C for 5 min. SDS-PAGE analysis was performed with 10% (w/v) polyacrylamide resolving gel and 5% (w/v) stacking gels in a Tris-glycine buffer system (25 mM Tris-HCl, 250 mM glycine, 0.1% (v/v) SDS) [34]. Gels were run at 120 V for 1 h and either silver-stained or Coomassie stained [33].

Gels for zymogram analysis deviated from the polyacrylamide gels detailed above in that the SDS concentration was reduced to 0.1% (v/v), the resolving gels final polyacrylamide concentration was 6% (w/v), and the protein samples were not boiled. Following electrophoresis, the gels were washed twice in 50 mM sodium acetate buffer containing 0.5% (v/v) Triton X-100 (pH 5.6) for 30 min to remove the SDS. Gels were then incubated at 50°C in 1 M sucrose dissolved in 50 mM sodium acetate (pH 4.8) buffer for 1 h. After the removal of the sucrose solution, the gels were washed 3 times using 18.2 MΩ MilliQ water. The gels were then incubated in boiling 250 mM NaOH solution containing 0.2% (w/v) triphenyltetrazolium chloride (TTC) (Sigma-Aldrich, St. Louis, USA). The gels were left until a red band formed after which the TTC was discarded, and the reaction was stopped with 5% (v/v) acetic acid [35].

3.3.5 Protein quantification and protein assays

All purified protein samples were quantified using the Pierce BCA protein assay kit (Thermo Fisher Scientific). After quantification, a standard assay was performed to assess the global activity of the wild-type (Suc2) and the variants (AW, NFS, FS, PGPL). The standard assay contained the following ranges of protein concentrations Suc2 (4 ng, 10 ng, 20 ng, 50 ng and 100 ng), AW (4 ng, 10 ng, 50 ng and 100 ng), NFS (100 ng, 200 ng) and PGPL (200 ng and 400 ng) The reaction mix consisted of 100 g/l sucrose dissolved in 100 mM sodium acetate buffer (pH 4.8) and was incubated for 1h at 50°C [10]. One global unit of enzyme activity was defined as the amount of enzyme that releases 1 µmole of glucose per minute under the conditions above [10]. For FOS production the assay reaction consisted of 600 g/l sucrose dissolved in 100 mM sodium acetate (pH 4.8). Based on the results from the standard assay, the enzyme dosages for the variants (AW, NFS, FS, PGPL) were standardised on a protein mass based on that of Suc2 (Suc2 protein correlating to 10 U/g of sucrose) [25]. In other words, different volumes of enzymes would be added due to the difference in protein concentration of the variant proteins. Due to the low activity of variants PGPL and NFS, the amount of protein was increased 4x. Samples (140 µl) were taken at 0 h, 1 h, 4 h, 8 h, 12 h, 24 h, 48 h and 72 h for the 10 U/g experiment, and for the 40 U/g experiment it was 0 h, 12 h, 24 h, 48 h, 72 h and 120 h. The positive control was wild-type Suc2 and the negative control was the reaction mix containing the enzyme substituted with buffer.

3.3.5.1 High-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD)

The enzyme reaction in the samples taken in 3.3.5 was terminated with 8 µl 35% (v/v) perchloric acid (PCA) and incubated for 10 min on ice. This was followed by the addition 8 µl 7N KOH with a further 10 min incubation on ice. The samples were kept at 4 °C. Before HPLC analysis, the dilutions were prepared by centrifuging samples for 15 min at 15 000rpm. The dilution series included the following: 200x, 500x, 1000x, 10 000x dilutions with 18.2 MΩ milliQ water. All the sample dilutions were analysed with CarboPac PA-1 column attached to a high-performance anion-exchange chromatography linked to pulsed amperometric detection (HPAEC-PAD) as detailed in Trollope *et al.*, [36] with modifications to the gradient elution programme. Using solutions A (250 mM NaOH), C (Water) and D (130 mM NaOH and 500 mM NaOAc), with the following elution profile was applied: (0 min) 85% C; (15 min) 85% C; (15.1 min) 42% C; (19 min) 10% C, 20% D; (23 min) 10% C, 20% D; (23.1 min) 85% C; (23.2 min) 85% C; (34.2 min) 85% C; (34.3 min) 85% C.

To prepare an external standard curve the HPLC grade standards (99% pure) at the following concentrations were used: 10 µg/ml, 15 µg/ml, 25 µg/ml, 30 µg/ml, 60 µg/ml and 100 µg/ml. Glucose, fructose, sucrose (Sigma-Aldrich, St. Louis, Missouri, United States), 1-kestose, nystose, 1^F-fructofuranosyl nystose (GF_4) (Wako Chemicals GmbH, Neuss, Germany) and 6-kestose [37] were used as standards. To identify neokestose, the oligosaccharide was produced in-house as there is no commercially available standard. A literature search identified the yeast *X. dendrorhous ATCC 24230* as an organism that produces the majority of its FOS in the neo-series G^6 with a FOS profile consisting of 1-kestose, neokestose and neonystose [15]. *X. dendrorhous ATCC 24230*, kindly provided by Prof A. Botha (Stellenbosch University) was grown on MMM for 72h after which the cells were pelleted by centrifugation at 8000 rpm for 5 min. The pelleted cells were resuspended in 500 µl 18.2 MΩ milliQ water and inoculated with 400 g/l sucrose, 100 mM citrate-phosphate buffer (pH 7) at 20°C. After 24 h of incubation, the sample was prepared for HPLC analysis as previously described. To identify the neokestose peak by retention time all the sugars were analysed by HPLC to ensure that the peaks in the sample correlated with chromatograms obtained by Shiomi *et al.,* [38]. To calculate the concentration of neokestose the areas of 6-kestose standards were used as a proxy to generate a standard curve. The area data from peaks with the same retention time as neokestose were used to calculate neokestose concentrations in unknown samples. This method allowed us to obtain relative neokestose values for samples and it was accepted that the values calculated were not true concentrations as the detector responses for different sugars are known to differ.

3.3.6 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) was employed to characterise the Suc2 variants by making use of protein thermal denaturation assays as described by Niesen *et al.,* [39]. The working concentration of 8x SYPRO orange (Sigma-Aldrich, St. Louis, MO, USA) was used in combination with 50 ng/µl of protein per reaction for all reactions listed below. Samples were incubated on a StepOnePlus real-time PCR machine (Applied Biosystems) with a step gradient which increased the temperature by 1 °C per min from 25-95 °C. To determine the influence of pH, Suc2 and variants were subjected to a pH range of 3, 4, 4.8, 5, 6 and 7 using 100 mM citrate-phosphate buffer. To test the stabilisation effect of sucrose, Suc2 and variant proteins were incubated with sucrose concentrations ranging from 100 mM to 500 mM increasing with an increment of 100 mM. The reaction mixture for the sucrose assay listed above consisted of 99.9% pure Sucrose (Sigma-Aldrich, St. Louis, MO, USA) and 0.1 M citrate-phosphate buffer at pH 4.8. Furthermore, trehalose (0.300 mM), D-(+)-Raffinose

pentahydrate (200 mM), cellobiose (100 mM), lactulose (300 mM) and maltose (300 mM) were also tested. Control reactions included all the reaction components but lacking either the enzyme or the substrate. The multicomponent data was exported from the StepOne software to Excel 2013. Data was sorted, and any necessary text editing was performed after which it was exported to GraphPad prism 7.03. The data was trimmed to two points above minimum and 2 points below the maximum fluorescence. The Boltzmann function was used to fit the formatted data and V_{50} (https://www.graphpad.com/guides/prism/7/curvefitting/index.htm?reg_classic_boltzmann.htm) was used with the Boltzmann equation which is the midpoint between the top and bottom temperature. The V_{50} was used to determine the melting temperature (T_m) of the proteins. Proteins decrease in stability with an increase in temperature as the temperature increase the proteins unfold, and T_m is the midpoint where folded and unfolded proteins are equal [39]. A one-way analysis of variance (ANOVA) with Dunnett's post hoc test was used to test for significant differences at the 95% confidence level between the T_m data for the different treatments.

3.3.7 Computer analysis and molecular modelling

All the sequence alignments, plasmids and assembly of sequencing results were analysed using CLC Main Workbench Version 7.6.2 (Qiagen, Venlo, Netherlands)

3.3.7.1 Homology Modelling

Homology models of all the variants were generated using the SWISS-MODEL web server [40]. The crystal structure of *S. cerevisiae* 4EQV from Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB Protein Data Bank) (https://www.rcsb.org/) served as the template for all the variants. The sequence similarities for the different models was PGPL (99.8%), AW (99.8%) and for NFS (99.21%).

3.3.7.2 Computer-aided design and sequence identification

The 3D structures of *A. japonicus* fructosyltransferase (AjFFase) 3LEM, *S. occidentalis* β-fructofuranosidase 3KF3 and *S. cerevisiae* β-fructofuranosidase 4EQV were obtained from RCSB Protein Data Bank (https://www.rcsb.org/). Structural alignment was performed using USCF chimera software version 1.10.2 (http://www.cgl.ucsf.edu/chimera/index.html) [41]. All the 3D structures were aligned using the best-alignment pair of the chain between the reference and the matched structures using the Needleman-Wunsch alignment algorithm and the BLOSUM-62 matrix. Secondary structure scores were weighted to 30%. With the variants

AW, PGPL and NFS modelled they could be analysed using Chimera software [41], within the software the models of the variants were superimposed with 4EQV, matching to the Suc2 crystal structure [42]. The crystal structure of Suc2 (4EQV) was not crystallised with a ligand thus had to be docked with the ligands of interest, including sucrose, 1-kestose, 6-kestose, neokestose and nystose. The docking of all the ligands listed was performed for all the variants homology models to compare the differences/similarities.

3.3.7.3 Ligand preparation

All the ligands (sucrose, 1-kestose, 6-kestose, neokestose and nystose) were downloaded from PubChem (http://pubchem.ncbi.nlm.nih.gov/) in 3D SDF format. The SDF files were converted to PDB format using Open Babel converter (http://openbabel.org/wiki/Main_page) which was then converted to a pdbqt file using Autodock tools 4.2 [43].

3.3.7.4 Protein preparation

The homology models for all the variants needed to be processed before docking using Autodock tools 4.2 [43]. Processing the homology models of the variants and Suc2 (4EQV) involved removing sub-units C-H which only left sub-unit A and B, i.e., the dimer configuration of the protein structure which was solved as a tetramer of dimers [42]. Using the dimer configuration all the water molecules were removed, the hydrogens were added, and Gaussian charges were calculated after which the dimer was then saved as a pdbqt file. The pdbqt files of Suc2 (4EQV) and all the variants were used in the molecular docking.

3.3.7.5 Molecular docking

The structure of Suc2 and all variants were not crystallised with a ligand thus had to be docked with the ligands of interest, including sucrose, 1-kestose, 6-kestose, neokestose and nystose. The docking process started by identifying the coordinates of the grid box using literature and NCBI CDD [44]. The catalytic triad was identified and used to centre the grid box. Employing Autodock tools 4.2 the grid box centre coordinates were identified and used in the file for Autodock vina [45]. The optimal grid box size was calculated using Feinstein & Brylinski [46] script (www.brylinski.org/content/docking-box-size) for each of the ligands. With all the needed parameters for Autodock Vina, each ligand was run for 20 times concurrently to the search space that was described above. The results for docking was the protein-ligand interaction and is reported as the affinity energy (kcal mol^{-1}). To identify the protein-ligand interactions

LigPlot+v.1.4 [47] was used. To assess the spatial deference between Suc2 and variant proteins, the Castp server [48] was used to calculate the volume differences between the various protein models.

3.4 Results

3.4.1 The rational mutation strategies

The length of the TI loop provided a possible reason for the lack of higher DP FOS (DP 4,5 etc.) production in Suc2 due to limited space within the active site [25,42]. It was hypothesised that there is a steric hindrance limiting space; thus Suc2 is not able to form higher DP FOS. All the changes made in the variants aimed at creating space which would theoretically allow the synthesis of higher DP FOS. The loop region targeted for mutagenesis in this study was identified in a previous study [25]. It was proposed that the structure of the TI loop (N42-L52 in Suc2) may play a role in a β-fructofuranosidase's transferase activity which is intricately linked to the ability to produce FOS species with degrees of polymerisation higher than three. In this study, further support for this hypothesis was obtained by adding the *Schwanniomyces occidentalis* β-fructofuranosidase (SoFFase) to the 3D structure superimposition of *A. japonicus* and *S. cerevisiae* enzymes. SoFFase has similar hydrolysis and transfructosylating activities to the Suc2 – it favours the hydrolysis of sucrose, 1-kestose or raffinose [49]. The SoFFase is capable of catalysing the synthesis of 6-kestose in larger quantities than 1-kestose, thus having a higher selectivity for the formation of the β-(2-6)-linked fructose FOS (levan-type FOS) [37]. Using the data from a sequence alignment (Fig. 3.2) of *A. japonicus* fructosyltransferase (AjFFase), Suc2 and SoFFase coupled with the structural analysis; four mutation strategies were formulated. First the W48A substitution, second the deletion of N42-L52 (in the TI loop) leaving only the 5 residues NPGPL, replacement of N42-L52 (TI loop) with the sequence from AjFFase, VGFLHDGDGIAG (V75-G85), and finally the replacement of N42-L52 with a loop from AjFFase HDGDG.

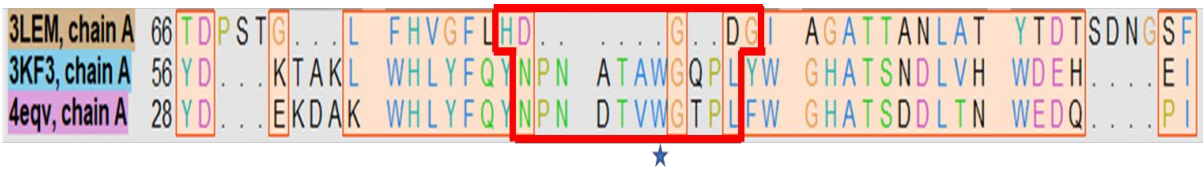

Figure 3.2. Sequence alignment of a segment of the β-fructofuranosidase's from *S. cerevisiae* (4EQV), *S. occidentalis* (3KF3) and *A. japonicus* (3LEM). The red box indicates the TI loop region within each of the respective enzymes highlighting the length difference. The "." indicates the gaps added to the sequence during sequence alignment.

The three superimposed 3D structures listed above (data not shown) highlighted that there was an obstruction between the W in both the Suc2 (W48) and SoFFase (W76) with nystose (DP4) which was co-crystallised with AjFFase (3LEM). Based on our hypothesis, this residue was a potential engineering target as it theoretically hinders the binding of higher DP FOS. This variant was named *AW* and consisted of substituting tryptophan (W) with an alanine (A) (W48A). A was selected as it has a shorter side-chain. This substitution would theoretically create space in the catalytic pocket of Suc2 but still stabilise the substrate through a hydrogen bond with the methyl group. The second mutation was the shortening of the TI loop N42-L52 (4EQV numbering) to the minimum loop size of five amino acid residues (NPGPL) according to Koga *et al.*, [50]. This change would theoretically open the catalytic pocket and increase the space within the pocket to accommodate higher DP FOS. The minimal loop retained amino acids P and G as they are known for inducing a β-turn [51] thus shorting the loop by 5 amino acids. This variant was abbreviated as *PGPL*. As stated in the study of Trollope *et al.,* [20] the loop region within the β-fructofuranosidases has the potential to limit the length of the FOS being produced by these enzymes. The AjFFase loop is shorter than in Suc2, and it was hypothesised to play a role in this enzyme's ability to produce higher DP FOS. Using this knowledge, the third mutation was the substitution of NDTVWGTPL (N42-L52 4EQV numbering) with VGFLHDGDGIAG (V75-G85 3LEM numbering), i.e. replacing the Suc2 TI loop with the shorter AjFFase TI loop. This variant was abbreviated as *FS*. The FS variant was designed and aimed at determining whether higher transfructosylating activity could be unlocked. The final mutation was another minimal loop named *NFS* but replacing NDTVWGTPL (N42-L52) in 4EQV with the HDHDG sequence of the 3LEM TI loop in between the two β-sheets.

3.4.2 Predicted results of the mutagenesis strategies

The docking of all the ligands listed in section 3.3.7.3 was performed for all the variant homology models to compare the differences/similarities. In Fig. 3.3, 4EQV docked with 6-kestose are superimposed with PGPL docked with 6-kestose. The loop in Fig. 3.3 (a) coloured red is the longer loop of Suc2 which obstructs the visualisation of the theoretical space created by the shortened loop. The loop was made transparent in Fig. 3.3 (b) which exposed the amount of space created by this mutation. Similarly, when comparing NFS with the Suc2, it is evident when contrasting Fig. 3.4 (a) and (b) the shortened loop in Fig. 3.4 (a) theoretically created space within the catalytic pocket. The effect of the change AW is smaller than the other two variants. The substitution of one amino acid when comparing Fig. 3.5 (a) and (b) is still evident with the shape of the active site entrance still being affected.

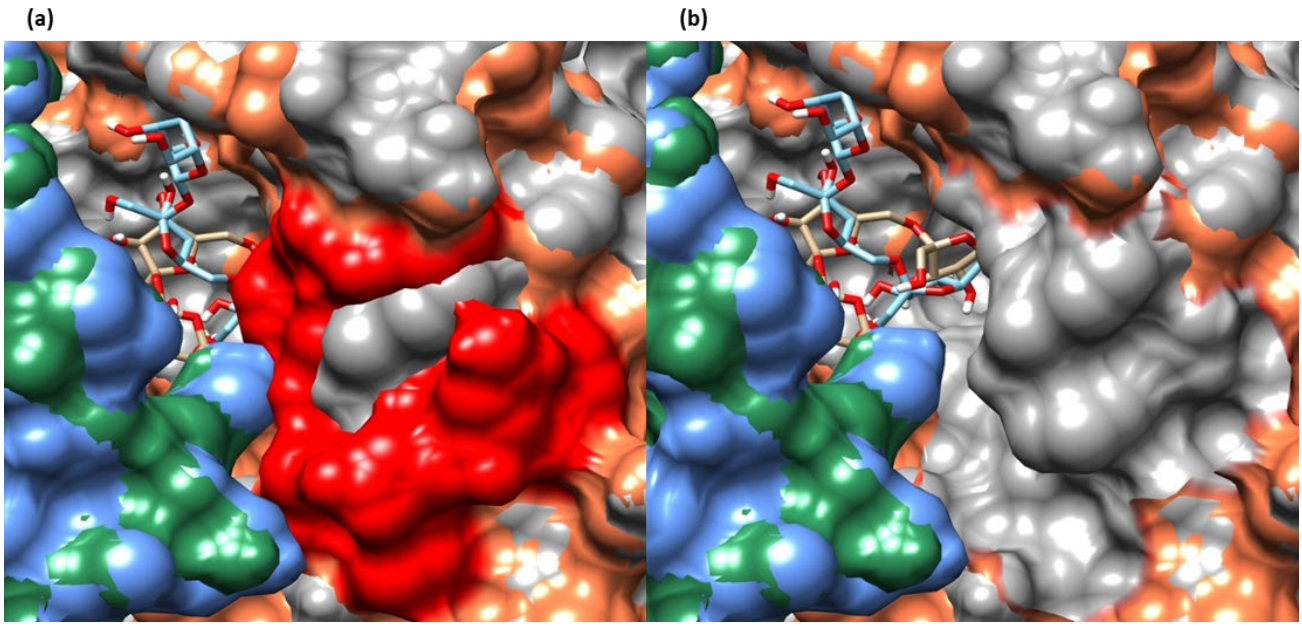

Figure 3.3. Superimposed PGPL and Suc2 docked with 6-kestose**.** PGPL chain A (grey) and chain B (blue), Suc2 chain A (brown) and chain B (green) with the brown and blue stick (6-kestose). **(a)** The loop of Suc2 (in red) at the catalytic pocket compared to **(b)** TI loop replaced by NPGPL highlighting the created space in the catalytic pocket.

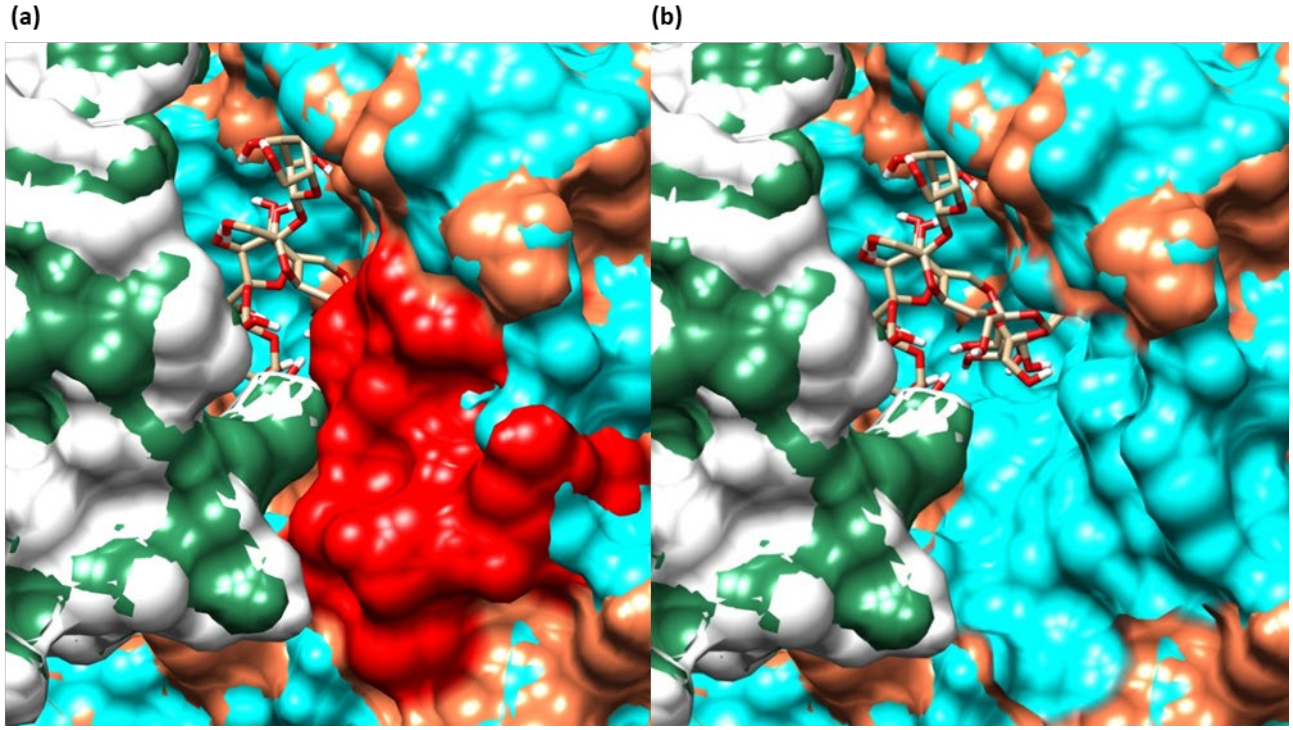

Figure 3.4. Superimposed NFS and Suc2 docked with 6-kestose**.** NFS chain A (cyan) and chain b (mustard), Suc2 chain A (brown) and chain B (green) with the brown stick (6-kestose). **(a)** NFS with the loop (red) in Suc2 **(b)** which is substituted with the loop of AjFFase.

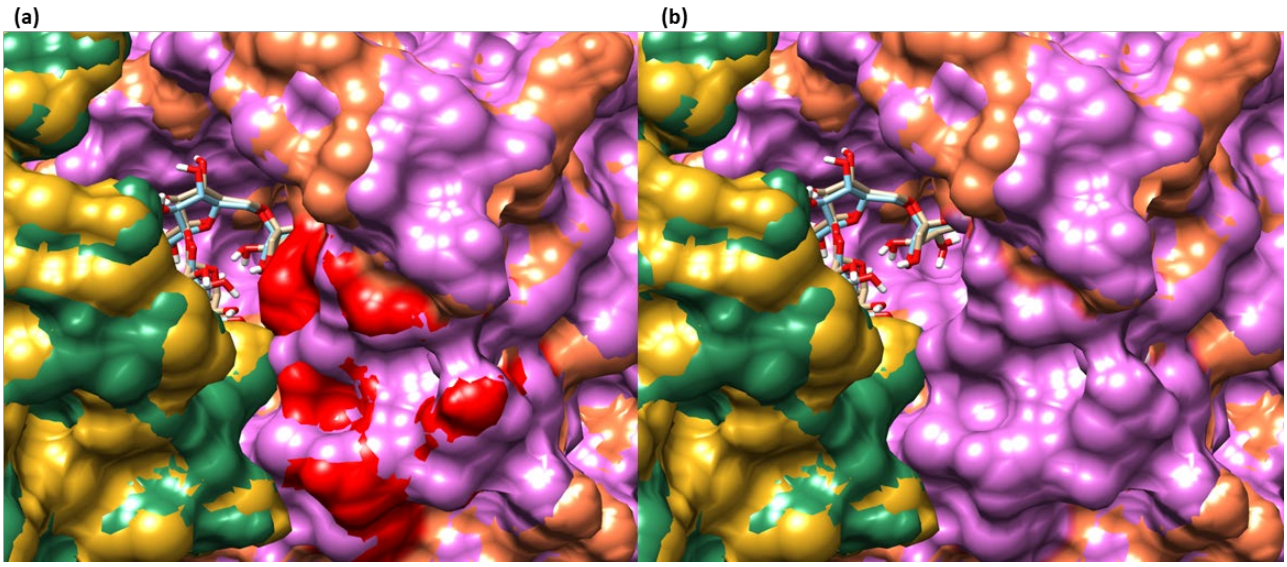

Figure 3.5. Superimposed AW and Suc2 docked with 6-kestose. AW chain A (pink) and chain B (gold), Suc2 chain A (brown), chain B (green), loop (red) with the brown and blue stick (6-kestose). **(a)** is the loop (red) in AW and **(b)** the space created by substituting the Trp with Ala.

3.4.3 Expression and purification of the β-fructofuranosidase variant proteins

The proteins used in all of the experiments were expressed in *S. cerevisiae* EUROSCARF Y12321 (BY4742; *MATα; ura3Δ0; leu2Δ0; his3Δ1; lys2Δ0; YIL162w::kanMX4*). Suc2 is a highly glycosylated protein, and the glycosylation (mannan) can contribute up to 50% (w/w) of the weight of Suc2 [52]. To ensure the proteins that were expressed were the correct size the proteins were deglycosylated with endo H (Fig. 3.6). The deglycosylated Suc2 protein has a theoretical weight of 58.5 kDa, however, due to the addition of the 6x His tag, the theoretical weight for variants and the Suc2 were calculated to be between 59-60.8 kDa. Fig. 3.6 (a) shows AW, PGPL and SUC2 after purification (lane 3-5) and after dialysis (lanes 6-8). Fig. 3.6 (b) shows NFS and SUC2 (lane 3 and 4) after purification and after dialysis (lane 6 and 7). It is evident from Fig. 3.6 that the Suc2 and protein variants were pure (≥95%) and that it was of the expected size (59-60.8 KDa), in accordance with Sainz-Polo *et al.,* [42]. Although FS was expressed in *S. cerevisiae* EUROSCARF and visualised on SDS-PAGE at the correct size, the enzyme activity was very low as there was a negative result in both TTC (no colourimetric change in the assay) and below the limit of detection in the standard assays (data not shown). As a result, FS was excluded from the rest of the study.

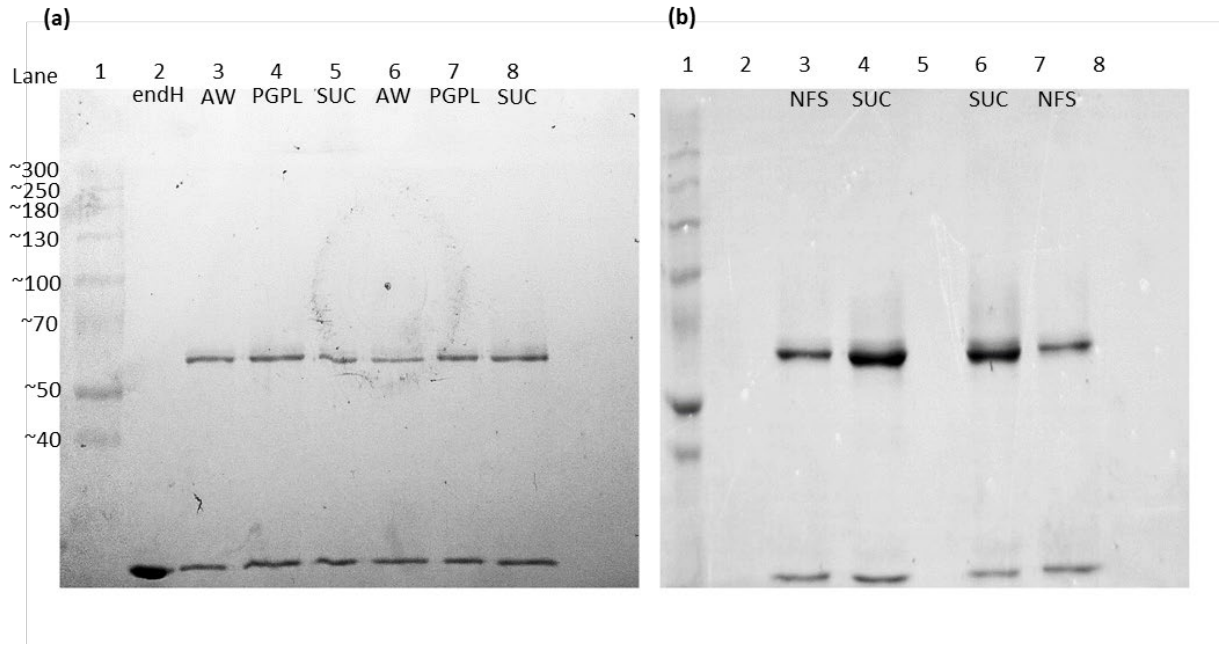

Figure 3.6. Coomassie blue-stained SDS-PAGE analysis of purified Suc2 and variant proteins heterologously expressed in *S. cerevisiae* EUROSCARF. Lane 1: Thermo Scientific Spectra Multicolour High Range Protein Ladder Lane 2: NEB Endo H deglycosylation enzyme **(a)** Lane 3-4 are the variants and lane 5 is a Suc2 protein sample after purification with lanes 6-8 is the same samples after dialysis **(b)** Lane 3 is the variant and lane 4 is Suc2 protein after purification and lane 6 and 7 are the same samples after dialysis.

3.4.4 The loop changes and their effects on the formation of higher DP FOS

To assess the impact of the loop engineering on the enzyme's activity, the enzymes were assayed with 600 g/l sucrose which is required for maximum transfructosylating activity [6]. Under these conditions, Suc2 was unable to produce higher DP FOS (DP >4), although under similar conditions all the variants were able to produce higher DP FOS. The higher DP FOS produced by all the variants included but were not limited to nystose (DP4) and GF_4 (DP5) (Fig. 3.7). Sugars elute from the CarboPac PA1 column in order of increasing length of FOS [53]. Also evident from Fig. 3.7 is that there was other high DP FOS species present, although these were not identified due to a lack of available standards. The variants produced quantifiable amounts of the higher DP FOS, with the PGPL variant producing 2.99 g/l GF_4, the variant NFS was able to amass 2.05 g/l GF_4, while the AW variant produced 1.19 g/l GF_4. The variants produced quantifiable amounts of nystose, with the PGPL variant producing 3.63 g/l nystose, NFS variant produced 1.8 g/l nystose and AW producing nystose although it was below the limits of quantification. In Fig. 3.7 both peak 4 (nystose) and 5 (GF_4) were not detected for the 150x and 200x dilutions of these two products, indicating that Suc2 did not produce them or they were at levels below the level of detection for HPLC detector.

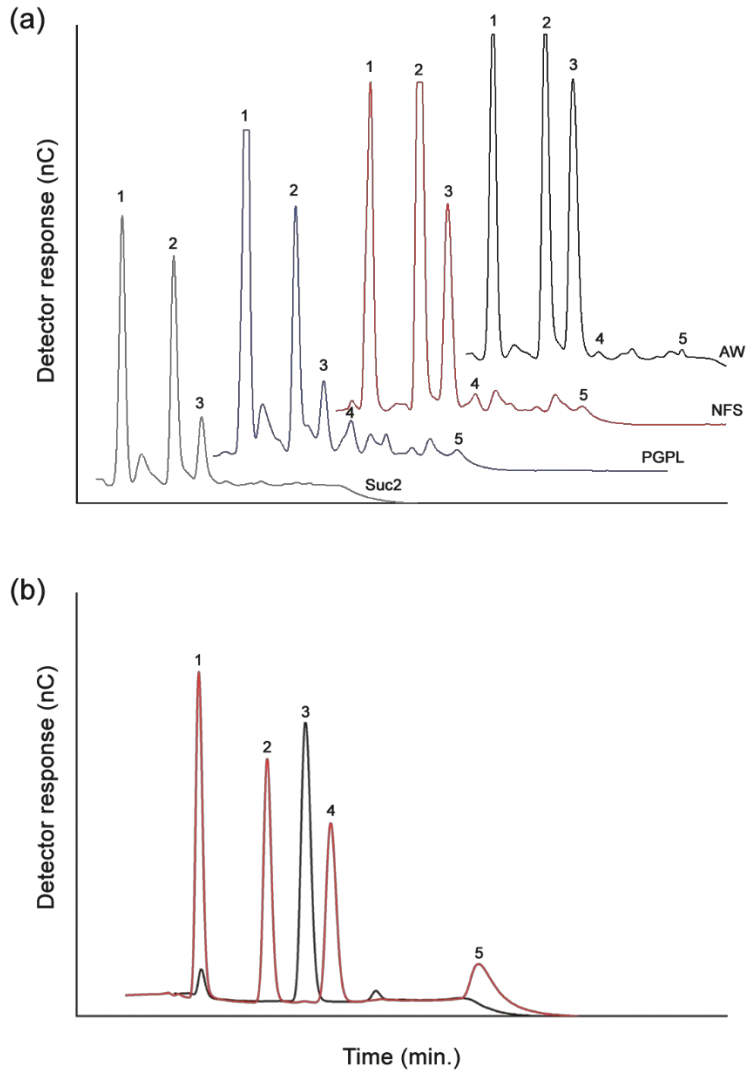

Figure 3.7. Compilation of the variants and Suc2 chromatograms at peak FOS production. **(a)** Suc2 at 8 h (grey), PGPL at 120 h (blue), NFS at 120 h (red) and AW at 8 h (black). Peak 1 (1-kestose), peak 2 (6-kestose), peak 3 (neokestose), peak 4 (nystose) and peak 5 (GF_4) were identified based on retention times of authentic standards. **(b)** The standards used in the identification and quantification of the sugars (red) the standards of peak 1 (1-kestose), peak 2 (6-kestose), peak 4 (nystose) and peak 5 (GF_4). The in-house standard for neokestose (black) peak 3 (neokestose) although not a pure standard as it contained 1-kestose and most likely neonystose (peaks 1 and 5, respectively).

3.4.5 The effect of the mutations on global activity and total FOS production

To test the effect of the different mutation strategies on the global activity of the variant enzyme a standard assay was conducted. Listed in Table 3.3 is all the global activity of the respective variants. Suc2 had the highest specific activity followed by AW which had the highest specific activity of the variants at 3300.88 U/mg. The loop replacement variants PGPL, NFS and FS, saw a drastic decrease in global activity, with 43.96 U/mg, 1.33 U/mg for PGPL and NSF, respectively.

Table 3.3. The global activity of Suc2 and the variant enzymes.

Enzyme	Specific activity (U/mg)	Standard error
Suc2	5107.08	80.84
NFS	1.33	0.41
AW	3300.88	111.17
PGPL	43.94	3.35
FS	Below limit of detection	-

The global activity can be assessed by measuring the glucose concentration as it is released during both the hydrolysis and the transfructosylating reactions. The hydrolytic activity was based on the concentration of free fructose as it is released in the hydrolysis of both FOS and sucrose. The relative transfructosylating activity was calculated using the glucose concentration divided by the fructose concentration as well as assessing the final concentration of total FOS [10]. The differences in the catalytic rate between the variants and Suc2 meant the maximum FOS production was compared at different time points as can be seen in Figs. 3.8 and 3.9.

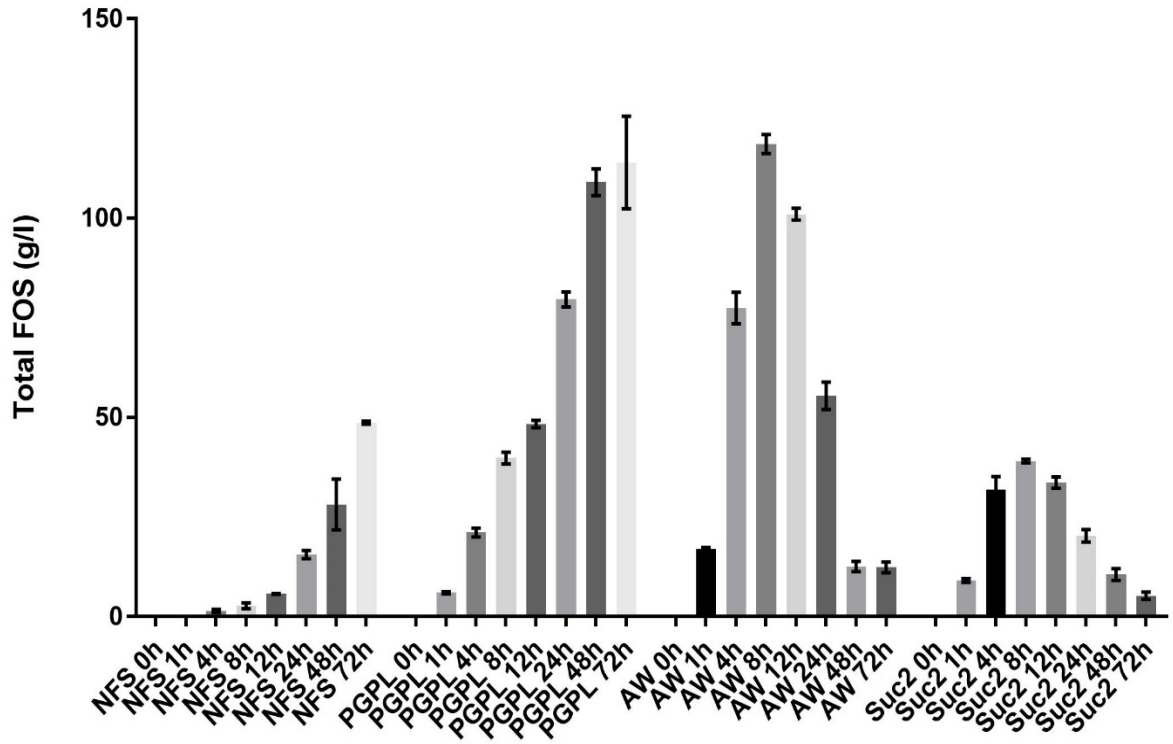

Figure 3.8. Total FOS production within 72h at 10U enzyme loading. The experimental conditions for this experiment were as follows: 600 g/l sucrose in 0.1 M sodium acetate pH 4.8 incubated with 10 U/g sucrose enzyme at 50°C. The sampling points are illustrated in the figure with 140 µl samples taken at 0 h, 1 h, 4 h, 8 h, 12 h, 24 h, 48 h and 72 h. The bars represent the average of triplicates at each sampling point, and the error bars represent the standard error of the mean.

The total FOS in this study refers to the combination of 1-kestose, 6-kestose and neokestose as they were the prolific FOS species. As NFS and PGPL had a significantly lower global activity, the assay was conducted at 10 U per gram sucrose as well as 40 U per gram sucrose. The total FOS production of the 10 U assay is represented in Fig. 3.8, which highlights the difference between the variants and Suc2. Suc2 and AW had a similar specific activity (Table 3.3) thus produced FOS in a shorter time frame. NFS and PGPL had a lower specific activity for FOS production. In the 10 U assay, NFS variant converted 21.95% (w/w) of the sucrose at maximum FOS production after 72h of incubation. In contrast, Suc2 converted 67.45% (w/w) of sucrose at maximum FOS production after 8h of incubation. Fig. 3.8 illustrates the differences between NFS and Suc2, i.e., the NFS variant converted 8.4% (w/w) of the sucrose to total FOS within 72 h where Suc2 converted 6.08% (w/w) of the sucrose to FOS within 8 h. To assess the transfructosylation activity at the maximum level of FOS production, the ratio of glucose/fructose was used. The NFS variant had a glucose/fructose ratio of 2.4 compared to the Suc2 enzyme ratio of 1.2, i.e. the NFS variant displayed double the transfructosylating activity compared to Suc2. The increase in transferase activity is further supported by the final concentration in total FOS with NFS producing 48.68 g/l total FOS in 72 h compared to Suc2 that produced 38.06 g/l total FOS in 8 h.

Evaluating the activity of AW, PGPL and Suc2 enzymes in Fig. 3.8, the 10 U assay showed the following at maximum FOS production: AW converted 23.21% (w/w) of the sucrose to FOS in 8h compared to PGPL which converted 21.25% (w/w) of the sucrose to FOS in 72h (although the 48h data was not significantly less), while Suc2 converted 5.58% (w/w) sucrose to FOS in 8 h. At maximum FOS production, AW converted 46.95% (w/w) of the total sucrose, PGPL converted 62.13% (w/w) (54.01% conversion at 48h) of the sucrose, and Suc2 enzyme converted 67.24% (w/w) of the sucrose. Evaluating the results listed above, both AW and Suc2 enzymes have a higher catalytic rate due to similar specific activities and kinetics, producing their maximum FOS at 8 h. PGPL has a lower specific activity and thus had a slower catalytic rate than Suc2. If the glucose/fructose ratio is assessed at maximum FOS production, it is clear that AW had a ratio of 1.5 compared to Suc2 which had a ratio of 1.2 and PGPL had a ratio 2.01 after 48h and 1.5 after 72h.

Due to the lower specific activity of the variants NFS and PGPL a second experiment was executed in which the enzyme concentration was increased to 40 U per gram sucrose. The purpose of this experiment was to determine the maximum FOS for NFS and PGPL as the FOS levels had not plateaued in the 10 U experiment in the tested time frame. With a 4x increase in enzyme concentration (Fig. 3.9), Suc2 hydrolysed most of the sucrose within 12 h, as expected. Maximal FOS production for NFS at 40 U was from 72h to 120h as there was no significant difference between the total FOS production of 21.01% (w/w) of converted

sucrose from the time point of 72 to 120h. Maximal FOS production for PGPL was at 24 h converting 20% (w/w) of sucrose to FOS. Assessing the sucrose concentration for NFS at maximal FOS production, 50.01% (w/w) of the sucrose was converted in 72h and PGPL converted 65.28% (w/w) of the sucrose in 24h. To calculate the relative transferase activity at maximum FOS production the equation described earlier was used. NFS had a ratio of 2.1 glucose to fructose while PGPL had a ratio of 1.4. This illustrates a higher transferase activity for NFS at the maximum FOS production compared to PGPL. PGPL started to hydrolyse the FOS as soon as the concentration was at the maximum of 151.8 g/l and can also be supported by an increase in the concentration of fructose which indicated a shift from transferase to hydrolysis (data not shown). In figures 3.8 and 3.9 it is striking how different NFS is to the rest of the enzyme. NFS did not start hydrolysing FOS as with the rest of the enzymes and can be seen in Fig. 3.9 as the concentration of FOS stayed constant from 72 h to 120 h. This is an indication of a lower affinity for FOS which can be the reason for the accumulation of FOS. This lower affinity is supported by the fructose data where there was a delay in the hydrolysis of FOS with the first substantial increase in free fructose between 48 to 72 h and the second between 96 to 120 h (data not shown). Combining the fructose and sucrose data it was noted that the sucrose did not decrease sharply, this suggests the transferase and hydrolysis are at equilibrium indicating the total FOS concentration was maintained.

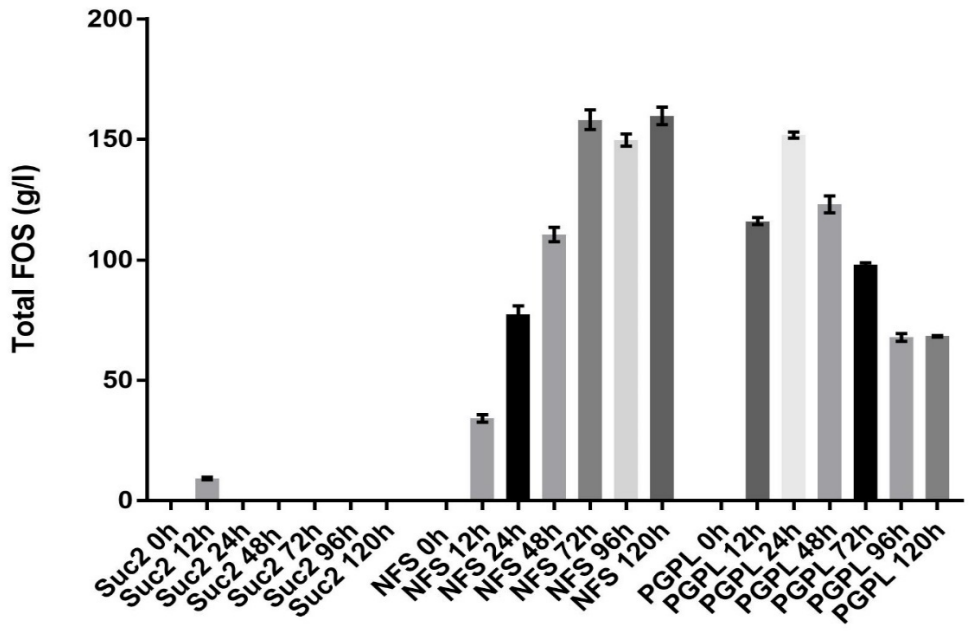

Figure 3.9. Total FOS produced within 72 h in 40 U experiment. The experimental conditions for this experiment were as follow, 600 g/l sucrose in 0.1 M sodium acetate pH 4.8 incubated with 40 U/g sucrose enzyme at 50°C. The sampling points are illustrated in the figure with 140 µl samples taken at 0 h, 12 h, 24 h, 48 h, 72 h, 96 h and 120 h. The bars consist of the average of the triplicates at each sampling point, and the error bars represent the standard error of the mean.

The stark differences between global activity in the standard assay (100 g/l) were not reflected in the 600 g/l assay due to the higher concentration of sucrose leading to an increase in glucose production. Glucose causes product inhibition as it competes for the binding site in these enzymes; this would affect the global activity of the enzyme [54]. The transfructosylating activities of all the variants were higher than Suc2, although this was not apparent in the relative transferase activity but was reflected in the total FOS production. With the higher enzyme dosage for PGPL and NFS, there was both an increase in the catalytic rate as expected, but there was an increase in the total FOS produced by these two variants.

In Fig. 3.8 and 3.9 the difference in total FOS production at maximum FOS production at 10 U/g sucrose and 40 U/g sucrose are shown. PGPL at an enzyme concentration of 40 U/g sucrose produced 151.8 g/l total FOS, NFS at an enzyme concentration of 40 U/g sucrose produced 158.1 g/l total FOS, AW at an enzyme concentration of 10 U/g sucrose produced 113.5 g/l total FOS and at an enzyme concentration of 10 U/g sucrose Suc2 produced 37.38 g/l total FOS. Fig. 3.11 shows the FOS production for all the variants and Suc2 at an enzyme concentration of 10 U/g sucrose at maximum FOS production. PGPL did not produce significantly more FOS than AW in the 10 U/g sucrose assay. The difference between AW and PGPL in the 10 U/g sucrose assay was AW produced its maximum FOS within 8 h were PGPL did it within 72 h. Due to the low catalytic rate of NFS, it produced 48.68 g/l total FOS at an enzyme concentration of 10 U/g sucrose. Although both NFS and PGPL produced more total FOS than AW in the 40 U/g sucrose assay. The reason being there is a dose-dependent increase in total FOS production.

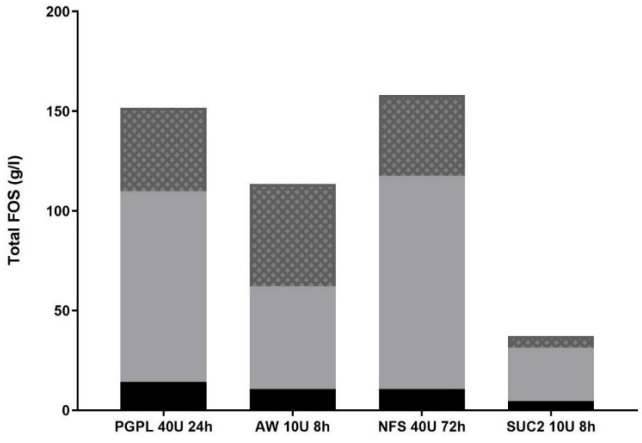

Figure 3.10. Total FOS production at maximum FOS production for all variants and Suc2 with the distribution of the individual FOS. Total FOS comparison for each of the variants and Suc2 is shown at their maximum FOS production. The concentrations of enzymes differ with AW and Suc2 production in the 10 U experiment and PGPL and NFS at 40 U. The total FOS is broken down into the individual FOS's that contributed to the total FOS. These include 1-kestose (black), 6-kestose (grey) and neokestose (dark grey with pattern). The data consisted of the triplicate for each component taken at different time points as per their maximum FOS production.

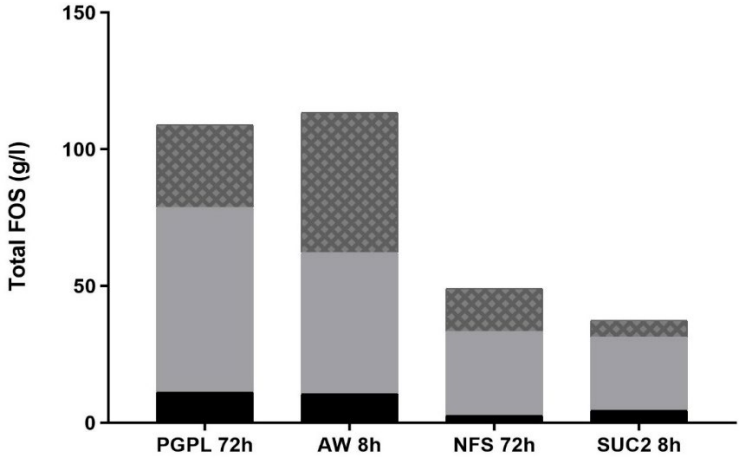

Figure 3.11. Total FOS production for 10 U at maximum FOS productions. Total FOS comparison of the variants and Suc2 at their maximum FOS production at an enzyme concentration of 10 U/g sucrose. The total FOS is broken down into the individual FOS's that contributed to the total FOS, i.e., 1-kestose (black), 6-kestose (grey) and neokestose (dark grey with pattern). The data consisted of the triplicate for each component and was taken at a time point where maximum FOS production was established.

3.4.6 The effect of the mutations on the regiospecificity of the enzyme

The total FOS consisted of the key products produced by Suc2 consisting of 1-kestose, 6-kestose and neokestose. The total FOS does not indicate the full extent of the changes in variant enzymes. The mutations increased the production of total FOS but also changed the regiospecificity of the enzymes. Suc2 is known to favour the production of 6-kestose and produce to a lesser degree 1-kestose and neokestose [6,10]. In the 10 U experiment, Suc2 produced predominantly 6-kestose with a final concentration of 27.97 g/l followed by the production of 5.14 g/l neokestose and 4.82 g/l 1-kestose yielding a ratio of 1:5.8:1.3. Illustrated in Fig. 3.11 is the respective variants at maximum FOS production at different time points and the proportions of each of the FOS contributing to the total FOS production. The 1-kestose production by AW was 11.08 g/l which was more than double compared to Suc2 (4.82 g/l). AW produced 53.92 g/l 6-kestose which was twice the amount of 6-kestose produced by Suc2. AW was able to produce 53.58 g/l neokestose which is ten times more than Suc2. NFS produced 2.61 g/l 1-kestose in the 10 U/g sucrose assay but produced 30.76 g/l 6-kestose and 15.76 g/l neokestose. PGPL produced 11.2 g/l 1-kestose in the 10 U/g sucrose experiment, 67.55 g/l 6-kestose and 30.3 g/l neokestose. All the variants produced quantifiable amounts of 1-kestose with NFS producing the least and AW and PGPL producing more than double the amount of 1-kestose compared to Suc2. The 6-kestose production in all the variants was significantly more than that of Suc2 with all the variant producing double the amount of 6-kestose. A striking observation was the increase in neokestose production as all the variants produced five times or more neokestose than Suc2. The effect of the mutations

on the variants regioselectivity was determined by using the ratio of 1-kestose, 6-kestose and neokestose respective to each other. Suc2 produced a ratio of 1:5.8:1.3, so for every 1 gram of 1-kestose Suc2 produced 5.8 grams 6-kestose and 1.3 grams of neokestose. The variants had the following ratio for the 10 U/g sucrose assay: AW produced a ratio of 1:5:5, PGPL 1:6:3 and NFS 1:11:6. The regiospecificity of PGPL and NFS did not change significantly in the 40 U assay compared to the 10 U assay.

3.4.7 Effect of the mutations on the stabilisation of the enzyme

To evaluate the stability of the mutated enzymes, differential scanning fluorimetry was used to determine the enzyme melting temperatures (T_m). The T_m is the temperature at which half of the proteins are unfolded, as reflected in fluorescence which will have a transitional midpoint on the fluorescent/temperature curve [39]. Using a Boltzmann fitting, the T_m can be calculated as it corresponds to the midpoint on the fluorescent/temperature curve. In the DSF experiments to differentiate between the Suc2 controls, Suc1 was used in the NFS experiment, and Suc2 was used in both AW and PGPL experiments. The T_m of all the variants and Suc2 is illustrated in Fig. 3.12 at a pH of 4.8. The T_m of Suc2 ranged between 68.6°C and 69.1°C which was higher than all the variants showing that all the mutations had a destabilising effect on the enzyme variants. AW's T_m was 65.99°C and PGPL's T_m was 61.01°C thus the AW T_m decreased with 3.11°C and the T_m of PGPL decreased with 8.09°C, respectively, compared to Suc2. The NFS T_m was 62.3°C thus decreasing with 6.3°C compared to Suc1. This is a significant reduction in the thermostability of all the variant enzymes compared to Suc2. The pH optima did not significantly change, with the optimal pH still being pH 4.8 for all variants and Suc2 (data not shown).

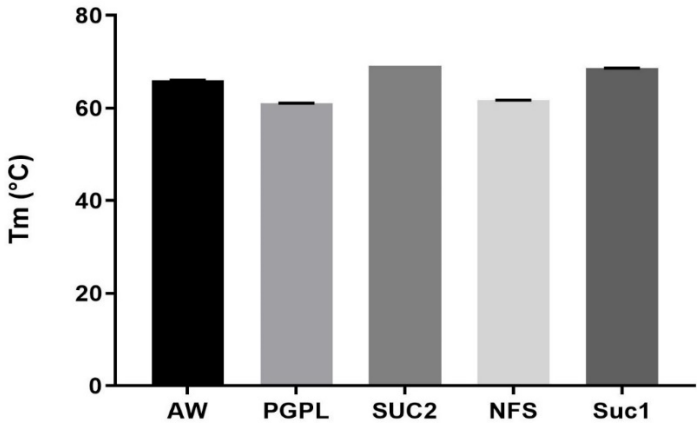

Figure 3.12. T_m of all the variants and Suc2 at optimum pH. This data represents the T_m of the variants and Suc2's at pH 4.8 without ligand. The data consists of triplicates produce by using GraphPad prism 7 Boltzmann fitting (V_{50}) and the error bars is the standard error of the mean.

3.4.8 The effect of the mutations on the enzyme's affinity for sucrose using differential scanning fluorimetry

Differential Scanning Fluorimetry (DSF) was employed in conjunction with different sucrose concentrations to assess the affinity of the enzymes towards sucrose. The change in T_m for each of the samples was compared to their own control (respective enzyme and reaction mix without sucrose). To assess the stabilising effect of the ligand, the enzyme's increase in T_m was deducted from the positive control of each of the variants and Suc2. This change in the T_m (ΔT_m) was used to assess the binding affinity of the ligand [55].

Fig. 3.13 and 3.14 represent the ΔT_m of the variants and Suc2 with respect to the increase in sucrose concentration. At a sucrose concentration of 200 mM and higher AW showed a significant increase in the T_m when compared to its control, while there was no significant difference at 100 mM. It was noted that PGPL had a significant increase in T_m compared to its control at a sucrose concentration of 300 mM and above thus showing no significant difference for 100 and 200 mM. NFS had a significant increase in its T_m at a sucrose concentration of 300 mM and above when compared to its control. An initial concentration of 100 mM Suc2 had a significant increase in T_m when compared to its controls.

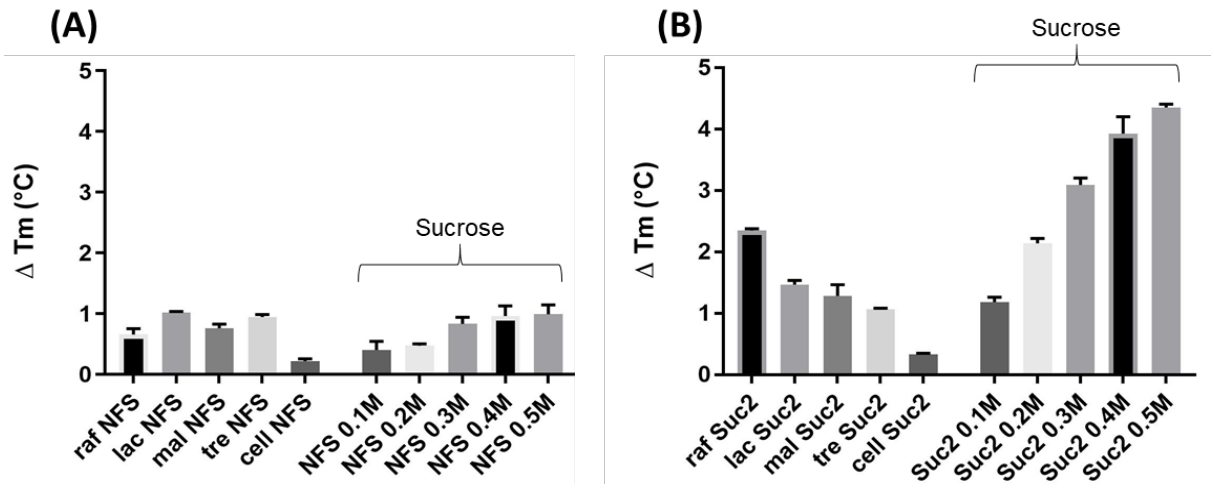

Figure 3.13. The comparison between the affinity of NFS and Suc2 for sucrose and other sugars. (A) The ΔT_m data for NFS normalised with the DSF reaction mixture and the enzyme (NFS) without the presence of any ligand. The sugar concentrations used in the experiment were 200 mM raffinose (raf), 300 mM lactulose (lac), 300 mM maltose (mal), 300 mM trehalose (tre) and 100 mM cellobiose (cell). (B) the ΔT_m data for Suc2 normalised with the DSF reaction mixture and the enzyme (Suc2) without the presence of any ligand.

3.4.9 The effect of the mutations on ligand selectivity using differential scanning fluorimetry

The binding affinity of different ligands was tested to potentially identify if there were a preference for specific bonds or orientation of glucose and fructose. These sugars included raffinose (200 mM), lactulose (300 mM), maltose (300 mM), trehalose (300 mM) and cellobiose (100 mM). Cellobiose and raffinose concentrations were decreased compared to the other sugars due to solubility issues.

In Fig. 3.13 (a) it is noted that NFS had a higher affinity for raffinose compared to sucrose as there was a significant increase in T_m for raffinose but no significant change with sucrose. All sugars except for cellobiose had a significantly higher T_m than the NFS variant control (Fig. 3.13). Lactulose and maltose had significantly higher T_m compared to their controls although there was no difference between these two and sucrose. Highlighting NFS had similar affinities for lactulose and maltose as for sucrose. Trehalose increased NFS T_m with 0.94°C this is a significant increase in T_m when compare to the control as well as the sucrose data. In the NFS experiment, Suc2 was the wild-type control and is illustrated in Fig. 3.13 (b). Fig. 13 (a) and (b) illustrate the difference in ΔT_m between NFS and Suc2 with the latter having higher T_m changes compared to the former. Overall Suc2 had a higher affinity for sucrose compared to lactulose, maltose and trehalose with raffinose having a similar increase in T_m than sucrose.

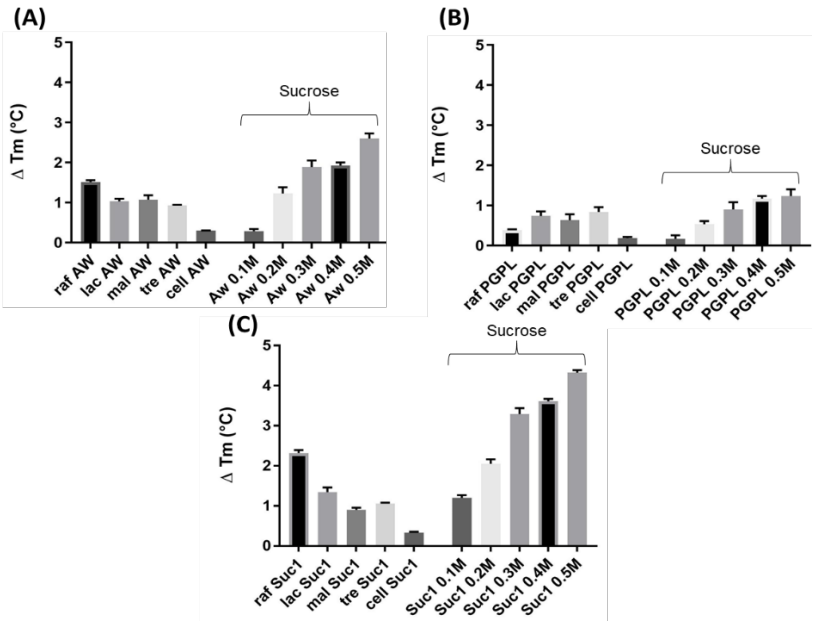

Figure 3.14. The comparison between the affinity of sucrose and other sugars. (a) The ΔT_m data for AW normalised with the DSF reaction mixture and the enzyme (AW) without the presence of any ligand. The sugars concentration use in the experiment was 200 mM raffinose (raf), 300 mM lactulose (lac), 300 mM maltose (mal), 300 mM trehalose (tre) and 100 mM cellobiose (cell). (b) The ΔT_m data for PGPL normalised with the DSF reaction mixture and the enzyme (PGPL) without the presence of any ligand. (c) Suc1 normalised with the DSF reaction mixture and the enzyme (Suc1) without the presence of any ligand.

In Fig. 3.14 illustrates the ΔT_m of AW, PGPL and Suc1 (Suc1 is only a naming convention to differentiate between the two experiments wild-types) for both the sugars (cellobiose, lactulose, maltose, raffinose, and trehalose) and sucrose. The ΔT_m was used to determine the binding affinity of the respective sugars. AW had a significant increase in the T_m for lactulose, maltose and trehalose when compared to the AW control (no ligand). AW saw a significant reduction in T_m with these sugars compared to sucrose. Fig. 3.14 (b) illustrates the binding affinity of PGPL to the different sugar analogues and sucrose. PGPL had a significant increase in its T_m for lactulose, maltose and trehalose. However, when compared to the sucrose data there was no significance in the T_m. Fig. 3.14 (c) represents the ΔT_m data for Suc1 (wild-type) which was used in the experiments with AW and PGPL. The Suc1, as with Suc2 in the NFS experiment, showed a comparatively higher response to the addition of ligands as their T_m's increased comparatively.

3.4.10 The effects of mutations on the hypothetical 3D structures and its implication on substrate binding

The variants were created *in-silico* using homology modelling software SWISS-MODEL with the sequence created in CLC-Main Workbench (Qiagen, Venlo, Netherlands). The model's template was 4EQV from the RCBS PDB website (https://www.rcsb.org/pdb/) and thus produced a homo-8-mer (octamer) homology models with the respective sequences. Most yeast β-fructofuranosidases, including the *S. cerevisiae* β-fructofuranosidase, occur as dimers which dimerise to form a tetramer and then the tetramers dimerise to form an octamer [42]. The octamer was edited using Autodock tools 4.2 [43] to remove the sub-units C-H, leaving only chains A and B or the AB dimer. The dimer was used in both molecular docking and to calculate the area of the catalytic pocket using the CastP server [48]. The CastP server makes use of computational geometry to identify and calculate the area and volume of cavities [56]. Using the CastP server the area was calculated and is listed in Table 3.4. As expected, the modifications of the loop did theoretically change the area of the catalytic pocket, increasing the area of AW by 2.9 %, PGPL by 10.7% and NFS by 23.6% compared to the Suc2 enzyme.

Table 3.4. Area of the catalytic pocket and the interface between the dimers

Enzymes	**Area (Å^3)**
Suc2	1482.6
NFS	1832.6
AW	1526.9
PGPL	1641.2

3.4.11 Analysis of structural features of the models combined with molecular docking to visualise topological similarities/differences.

The interest in β-fructofuranosidases has led to the availability of multiple crystal structures for this enzyme family [14,42,57,58]. Availability of these 3D models and published enzyme functional data allow for enzyme structure-function hypothesis testing. The topology of the catalytic pockets is claimed to play an important role in the function of β-fructofuranosidases [42,57,59]. Fig. 3.15 shows the 3D structures of three β-fructofuranosidases, AjFFase (3LEM) co-crystallised with nystose (turquoise), the NFS variant homology model docked with nystose (cream) and *S. cerevisiae*'s Suc2 enzyme, (4EQV) in native form (grey). The red circle illustrates the location of the catalytic pocket and the yellow areas indicates the TI loop in their respective proteins. There is a size difference between the catalytic pocket of Suc2, NFS and AjFFase. Fig. 3.15 illustrates the difference between the catalytic pocket (indicated in red circle) of AjFFase and Suc2 with the former having a deep and wider pocket than the latter. The catalytic pocket of NFS is as wide as that of AjFFase but is not as deep; therefore more of a cleft type pocket rather than a tube pocket. Suc2 is a narrow and relative shallow pocket when compared to AjFFase and NFS. The topology of NFS resembles that of AjFFase as it contains a reduced loop length.

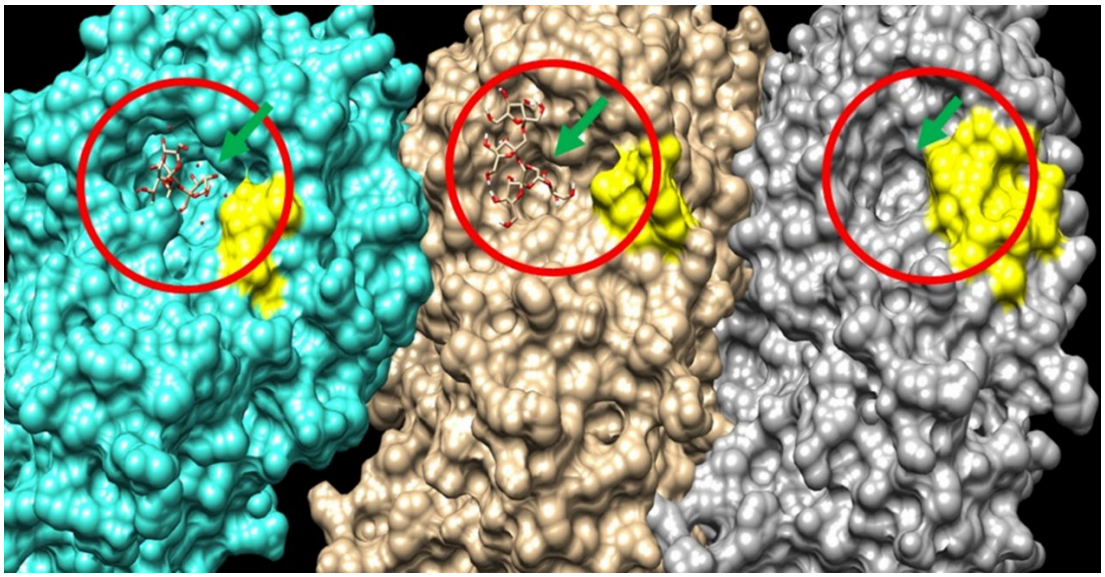

Figure 3.15. Enzyme topological comparison between a high FOS producer, a low FOS producer and a variant of a low FOS producer. Side by side comparison of AjFFAse (turquoise), NFS (cream) and Suc2 (grey) highlights in yellow the respective loops and the red circle indicates the catalytic cavity with the green arrow indicating the entrance to the active site. The entrance of the Suc2 catalytic site is narrow compared to AjFFAse and NFS. The catalytic cavity of NFS resembles that of AjFFAse rather than Suc2. There are however differences as AjFFAse has more structure around the cavity whereas NFS has none.

Using Autodock vina [45] the variants and Suc2 enzyme were docked with the ligands 1-kestose, 6-kestose, neokestose, nystose and sucrose. The results of the docking simulations were analysed with Ligplot+ to identify the orientation and the amino acids interacting with the ligands. Using data from Reddy & Maley [5] for the Suc2 β-fructofuranosidase the following amino acids forming part of the catalytic triad should theoretically bind to the ligand during docking: in Suc2 and AW Asp 22 (nucleophile), Asp 151 (transition state stabiliser) and Glu 203 (acid/base catalyst). Due to the variants having changes to their sequence, the amino acid numbering changed for the PGPL (Asp 22, Asp 145 and Glu 197) and the NFS variant (Asp 23, Asp 146 and Glu 198). The critical part in the docking was the binding of Asp 22/23 and Glu 203/204 as these are required for the substrate to bind to the protein [5]. Docking the variants and Suc2 enzyme with the ligands showed that the success rate of the ligand docking was low with only 6-kestose docking with the aforementioned active residues. Quality control was not possible as there are no ligand-docked models for 4EQV to measure the quality of the final docked complexes. Therefore, the docking results obtained could only be used as a suggestive tool for understanding the effect that the changes had on the regiospecificity. Furthermore, docking the variants and Suc2 enzyme with nystose was unsuccessful as the predicted orientation of the ligand was incorrect with the -1 position being occupied by glucose, whereas the correct orientation should be fructose occupying the -1 position with glucose facing towards the outside and occupying the +1 and more positions [42,60]. Fig. 3.16 (a) and (b) illustrates the successful docking of AW and Suc2, with the catalytic triad interacting with 6-kestose. Contrasting Fig. 3.16 (a) with (b) there is a difference in the amino acids interacting with 6-kestose between AW and Suc2.

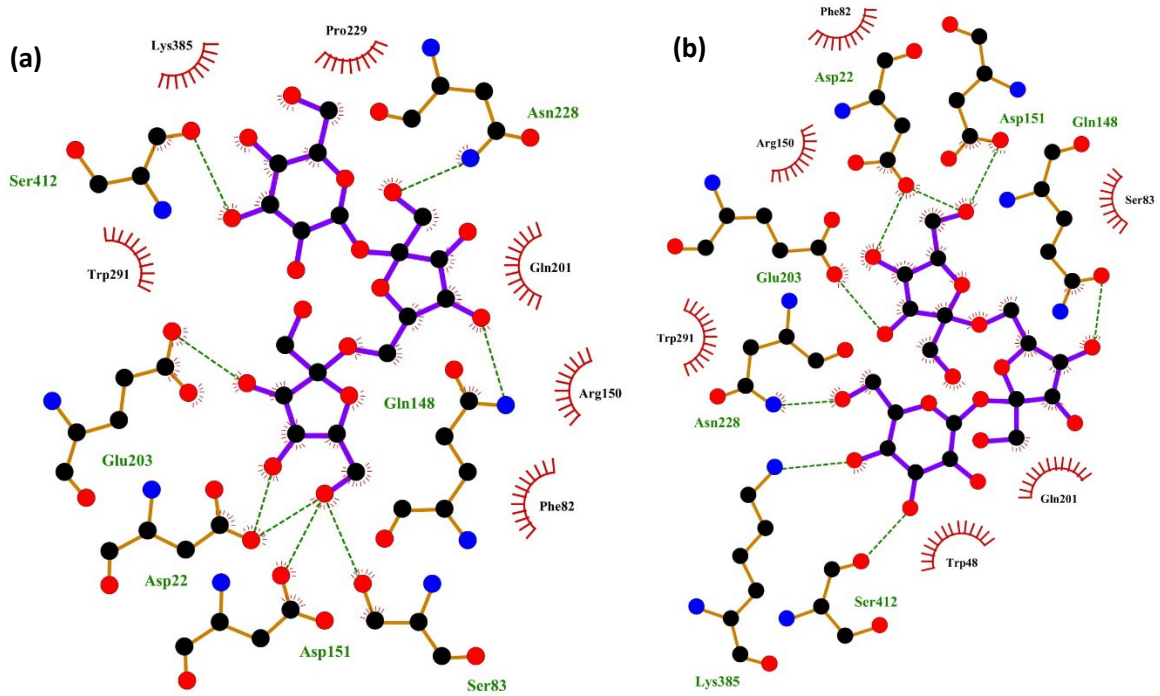

Figure 3.16. The 2D structures of the active site of Suc2 and AW docked with 6-kestose. **(a)** 6-kestose docked with Suc2 using Autodock vina. The 2D image was generated using ligplot+, and it highlights the interaction between the residues and the ligand (purple). The red residues are hydrophobic interactions with the green dotted lines being hydrogen interactions. **(b)** AW docked with 6-kestose using autodock vina. The 2D image was generated using ligplot+, and it highlights the interaction between the residues and the ligand (purple).

3.5 Discussion

In this study, the TI loop located between β-sheet B1 and C1 had the potential to explain why certain β-fructofuranosidases could produce higher DP FOS than others. Sainz-Polo *et al.,* [42] proposed that the limiting factor for binding three to four sugar units (DP4 and DP5) was a space limitation imposed by the small catalytic pocket of some β-fructofuranosidases and thus the reason why higher DP FOS cannot be produced. Our results implicate the TI loop as a further contributing factor for the limited formation of higher DP FOS in β-fructofuranosidases. The TI loop is known as a hydrophobic wall or zone important for substrate stabilisation and orientation within the catalytic site of β-fructofuranosidase [42,59]. The shortening of the TI loop increased access of substrates to the catalytic pocket of the enzyme by decreasing the size of the hydrophobic zone which has many bulky aromatic amino acids in the Suc2 enzyme. TI loop variants of Suc2 constructed in this study (AW, NFS and PGPL) led to detectable and quantifiable concentrations of nystose and GF_4. To our knowledge, no evidence of GF_4 production by Suc2 has been published before.

The tryptophan residue at position 48 (W48) in Suc2 is known for phi-stacking which implicates substrate stabilisation [42]. W48 has been shown to have an important role in multiple β-fructofuranosidases as it is highly conserved, especially in yeast β-fructofuranosidases such as *S. occidentalis* and *X. dendrorhous*. Previously, the position of W76 in *S. occidentalis* β-fructofuranosidase was identified to be a key residue in the active site [61]. Positionally W76 of the *S. occidentalis* β-fructofuranosidase and W48 in Suc2 occupy the same position. In this study, W48A substitution led to an increase in transferase activity, most likely due to the space created (illustrated in section 3.4.2) by the small sidechain of Ala facilitating the entrance of acceptor substrate molecules. In a similar study Alvaro-Benito *et al.,* [62] highlighted that this position is highly conserved in the GH 32 family further supporting the importance of this position. The importance was highlighted in Ramirez-Escudero *et al.,* [14] as W105 forms +2 position in *X. dendrorhous* which is positionally similar to W48 an W76. Mutations in these enzymes can be tested to test if similar outcomes will be obtained and further support the importance of the position of W48, W76 and W105 in the yeast β-fructofuranosidases.

Shortening of the TI loop region of a β-fructofuranosidase has not been previously described. Our results in combination with the *in-silico* methods (cavity area calculation), suggests that the shorter TI loop region creates space in the catalytic pocket. As the TI loop forms part of the entrance to the active site, it could be argued that the removal of these residues allows the acceptor sucrose to enter more readily and thus increase in transferase activity. However, the variants created in our study did not produce high levels of higher DP FOS (DP > 4) as found in the *A. japonicus* enzyme, and this might be due to the absence of the stabilising amino acids, which help substrate binding. It is also worth noting that β-fructofuranosidase enzymes are structurally similar although topologically different. The topological differences of the catalytic pockets are evident with the high transferase AjFFase having a deep catalytic pocket [57], whereas Suc2 has a cleft-like catalytic pocket [42]. The former has previously been suggested as the reason why AjFFase might have higher transferase activity. Chuankhayan *et al.,* [57] claimed that glucose might prohibit the release of fructose by inhibiting the release of fructose from the active site after the first stage transfructosylation or hydrolysis. Due to the mechanism of these enzymes, the release or retaining of fructose is crucial for transferase activity with the release of fructose leading to hydrolysis. Therefore, we suggest that the cleft-like catalytic pocket of Suc2 could be the reason for the quick release of fructose leading to high hydrolysis activity. The retention of glucose can be the reason for the increase in transferase activity with a decrease in global activity in the Suc2 variants in this study. As Suc2 retained glucose, to prohibit the release of fructose, it allows for a longer time to elapse and to increase the likelihood of sucrose binding to fructose [59], although there can be other explanations for the increase in transfructosylating activity.

As the substitution of W48A leads to an increase of the transferase activity by producing five times more FOS while retaining 65.19% of its global activity. Our data suggest that the W48 residue is important in the interaction of the acceptor molecule (water or sucrose), thus potentially decreasing the affinity for water and increasing the affinity for sucrose. FS was not discussed in the results as the amino acids residues FW were accidentally excluded from the β-sheet (C1), adjacent to the TI loop during the genetic engineering stage. The protein was expressed and was visualised on SDS-PAGE gel, although its activity was not detectable. Due to the latter and the fact that the mutation strategy did not only change the TI loop but extended into the β-sheet, the FS variant was excluded as the rationalisation of these changes would have added more speculation. It is still worth noting that employing our rational enzyme engineering strategy, constituting drastic changes, we achieved a success rate of 75%, with only FS decreasing the success rate due to human error. The success rate of the mutations is based on the hypothesis that the shortening of the TI loop will lead to an increase in DP of

FOS and transferase activity. The latter highlights the importance of loops in enzymes as the changes made to the TI loop led to an increase in DP and transferase activity.

The TI loop variants of Suc2 constructed in this study (AW, NFS and PGPL) resulted in a reduction of the enzyme's hydrolytic activity as well as the global enzyme activity with a corresponding increase in transferase activity for all the variants compared to the Suc2 enzyme. The same observation was previously reported as a general observation for mutated β-fructofuranosidase enzymes [10]. The changes introduced to the enzyme affected not only the transferase activity but also changed the enzyme's substrate regiospecificity. The regiospecificity was assessed using the ratio of 1-kestose to 6-kestose to neokestose Suc2 produced a ratio of 1: 5.8: 1.3, AW produced a ratio of 1: 5: 5, PGPL produced a ratio of 1: 6: 3 and NFS produced a ratio of 1: 11: 6. The change in selectivity of the enzymes can be attributed to the TI loop having a critical role in the orientation of the sucrose molecule (acceptor substrate) upon entering the catalytic pocket of the enzyme. The concentration of 1-kestose doubled in the variants compared to Suc2 which can be explained by a higher transferase activity in the variants. The increase in products containing a β-(2-6)-glycosidic bond, i.e. 6-kestose and neokestose, indicates that the TI loop possibly plays a role in the orientation of the acceptor sucrose. In the case of neokestose sucrose is reversed and the glucose of the sucrose is bound to the protein-fructose complex. AW produced neokestose and 6-kestose in equimolar amounts, which indicates that the enzyme was more flexible regarding the orientation of the sucrose molecule. Suc2 has a selectivity towards the β-(2-6)-glycosidic bond although the mechanism is still unknown [10]. It was proposed that W105 residue plays a critical role in the formation of neokestose as per the proposed mechanism in the *X. dendrorhous* β-fructofuranosidase [14]. W105 in *X. dendrorhous* β-fructofuranosidase enzymes is in the same position as W48 in Suc2. Replacing of W with A in the AW variant led to a change in the orientation of the acceptor sucrose entering into the catalytic pocket. It is proposed that W48 is part of the catalytic pocket entrance in Suc2 as it can interact with the accepting sucrose thus changing the orientation. Our data support this idea as the AW variant produced equimolar 6-kestose and neokestose.

The NFS and PGPL β-fructofuranosidase variants showed an increase in total FOS production which is attributed to the increase in the transferase activity of these two enzyme variants. As proposed above, this increase in transferase activity can be credited to the catalytic pocket being more open allowing the acceptor sucrose to enter more readily. This conclusion is however speculative and should be confirmed by further investigation. The PGPL variant showed an increase in total FOS production with an increase in 1-kestose, 6-kestose and neokestose. The selectivity of PGPL was closer to that of Suc2, but there was more

neokestose produced by PGPL. The NFS variant produced increased total FOS levels but with an increase in selectivity for 6-kestose when compared to AW, PGPL and Suc2. The NFS variant showed selectivity towards neokestose compared to Suc2 although not as high as in AW variant. The selectivity towards the formation of neokestose in both NFS and PGPL can be attributed to the removal of W48. The loop seems to have no effect on the selectivity for β-(2-1)-glycosidic bonds and β-(2-6)-glycosidic bonds as the formation of β-(2-6)-glycosidic bonds were still favoured in all of the variants. The changes to the TI loop affected the regiospecificity as the variants had higher regioselectivity towards the 6-OH hydroxyl group of glucose.

In summary, changes to the TI loop in AW, PGPL and NFS affected the activity of the *S. cerevisiae* β-fructofuranosidase enzyme by shifting the activity from hydrolysis to transferase. The W48A showed that the W residue plays a vital role in governing selectivity of 6-kestose and neokestose. PGPL and NFS had an increase in transferase activity but a decrease in hydrolytic activity. NFS had a selectivity toward 6-kestose and PGPL had a similar selectivity as Suc2. NFS and PGPL had an increased selectivity towards neokestose when compared to Suc2. Assessing the total FOS data together with the individual FOS components, it is worth noting that both the transfructosylating activity and regioselectivity have been affected in the variants. The latter highlights the disruptive effect the mutations had on the activity of the enzymes, i.e., not only leading to the change in enzyme specific activity but also the binding of sucrose.

A general observation that can be made from the DSF analysis was that all variants showed a reduction in the thermostability compared to Suc2. The decrease was less severe for the AW variant compared to the other variants, which could probably be explained by the latter two mutations being more severe (six amino acids was removed in total). DSF is a very sensitive method and leads to some experimental difficulties with data interpretation. The destabilising effect was confirmed in the pH/T_m experiment which highlighted that there was a statistically significant reduction in the T_m of all variants when compared to the Suc2 enzyme. The results in the DSF experiment illustrated the destabilising effect of shortening or altering the loop in Suc2. The decrease in the thermostability might contribute to the increase of regiospecificity of the variants as the mutations could have led to a more flexible active site [63].

The data from the experiment using different sucrose concentrations was used to assess the effect the mutations had on the sucrose affinity in the variant enzymes. The sucrose affinity of Suc2 enzyme was higher with the T_m being significantly different at a sucrose concentration of 100 mM. The AW variant followed this result with a T_m increase which was statistically

significant at a concentration of 200 mM sucrose. NFS and PGPL variants showed a statistically significant increase in T_m at a concentration of 300 mM sucrose. The latter indicates that these variants' affinity for sucrose decreased as a consequence of changes to the TI loop. These results highlight the powerful stabilising effect conferred by sucrose when binding in the catalytic site of Suc2 as opposed to a very weak interaction in the variants. These results illustrate the high affinity of Suc2 towards sucrose with the variants having a lower affinity for sucrose. This can be one of the reasons for the increase in the formation of neokestose in the variants. The lower affinity for sucrose correlates with a flexible catalytic pocket and a weaker interaction with sucrose. This flexible catalytic pocket was further supported by the thermostability assay which implicated the TI loop as being crucial in stabilising the enzyme, as there was a significant decrease in T_m for all variants.

β-fructofuranosidases are promiscuous enzymes and these enzymes can fructosylate multiple substrates, especially sugar alcohols [3]. Although they are promiscuous enzymes, they favour either the β-(2-1)-glycosidic bond or the β-(2-6)-glycosidic bond when synthesising FOS [37]. Constraints of higher DP FOS production might be due to physical limitations as the shape of the catalytic pocket cannot facilitate the production of GF_4 or higher DP. Various substrates with different glycosidic bonds were used to assess the ability of the enzymes to bind other sugars. Cellobiose is a dimer of glucose that is joined by a β-(1-4)-glycosidic bond. We know β-fructofuranosidase can bind glucose although there was no stabilising effect conferred by cellobiose for any of the variants or Suc2. The lower concentration of 100 mM cellobiose can be the reason for the variants not showing any stabilising effect. In Suc2 the concentration of cellobiose cannot be the reason for the insignificant change in T_m as sucrose at the same concentration yielded a significant increase in T_m. The latter can be explained by the enzyme being less able to recognise the β-(1-4)-glycosidic bond thus discriminating based on the bond [64]. Maltose, which is composed of two glucose molecules that are bound by an α-(1-4)-glycosidic bond, had a stabilising effect on the variants and Suc2, which was in contrast to cellobiose which also contains two molecules of glucose but the bond differed as well as the concentration. NFS and PGPL were different as there was no statistical difference between 300 mM maltose and 300 mM sucrose. Both Suc2 and AW showed a decrease in the stabilising effect of maltose when compared to sucrose. Based on the results there must be an enzyme-substrate interaction in NFS and PGPL that is not present in AW or Suc2. Trehalose (glucose α,1-1 glucose) affected Suc2 and AW in that it did stabilise the enzyme when compared to the control although when compared to the sucrose of similar concentration lead to a decrease in T_m. Variants PGPL and NFS differed from Suc2 and AW as there was no difference between the stabilisation effect of sucrose compared to the trehalose. A similar result was observed for maltose with NFS and PGPL being different from AW and Suc2.

Lactulose (galactose β, 1-4 fructose) has the same β-(1-4)-glycosidic bond which is also present in cellobiose. Cellobiose did not affect the stability of the variants and Suc2, although lactulose did. This effect can be attributed to the higher concentration of the lactulose at 300 mM compared to 100 mM cellobiose.

NFS and PGPL appear to have a similar affinity for both lactulose and sucrose as there is no difference at 300 mM sucrose and 300 mM lactulose. The reason for this can be the fructose in lactulose that might lead to a similar stabilising effect as sucrose. Raffinose (galactose α,1-6 glucose α, 1-2 β-fructose) is the only trisaccharide tested in this experiment with a terminal fructose in a β-(2-1)-glycosidic bond similar to 1-kestose. Raffinose is a substrate that Suc2 can hydrolyse forming fructose and melibiose [65]. Raffinose did increase the T_m for PGPL, AW and Suc2 but not for NFS. Raffinose increased the T_m of PGPL, while sucrose at the same concentration had no effect. DSF is a screening method, and more research will be needed to determine whether these effects translate into useful data given the high sensitivity shortcoming of this method. As the relative increases of the T_m shifts were small in the variants compared to the relative increase of Suc2, it is difficult to compare the variants to Suc2. To truly test the extent of these ligands and their effect on the variants and Suc2 further research is needed. Additional experiments can be done to see if the variants are binding these ligands and to assess if the changes to the TI loop lead to broader substrate binding.

3.6 Conclusion

There have been multiple studies trying to increase the transferase activity of the GH 32 family with varying success [3,9–11,14,25,59,66]. The question that most of these studies tried to answer is the reason for the transfructosylating activity and what governs this property in the GH 32 family enzymes. During this study, we mutated the *S. cerevisiae* β-fructofuranosidase, a predominantly hydrolytic enzyme, shifting it to a transferase favouring enzyme. The variants in this study displayed a significantly decreased hydrolytic ability and increased the FOS production from 4% in Suc2 to over 20% in the variants. With this increase in the yield of FOS, the variants also synthesised quantifiable amounts of nystose and GF_4 which have not been detected previously in the *S. cerevisiae* enzyme.

Additionally, these mutations changed the regiospecificity of the FOS with the variants producing levan-, inulin- and neotype FOS. The mutations had a significant effect on the neo-type FOS with the variants producing neokestose at a level only seen in a β-fructofuranosidase that have a preference towards producing this type of FOS. These mutations did come at a cost as it significantly decreased the thermostability and global activity of the enzymes. This

decrease in stability was not in all of the variants as AW kept a considerable amount of its global activity and still increased the yield of FOS to more than 20%. The W48 position in *S. cerevisiae* β-fructofuranosidase has highlighted its importance in the hydrolytic/transferase activity of the enzyme. W48 is one of the key amino acids which contributes to *S. cerevisiae*'s β-fructofuranosidase being a predominantly hydrolytic enzyme. Furthermore, positionally the W48 has been implicated as being a crucial amino acid in multiple yeast β-fructofuranosidases. These variants can serve as a baseline for future studies by building on the increased transferase activity and finding mutations that will increase the catalytic rate.

3.7 References

[1] T. Pons, D.G. Naumoff, C. Martínez-Fleites, L. Hernández, Three acidic residues are at the active site of a β-propeller architecture in Glycoside Hydrolase families 32, 43, 62, and 68, Proteins Struct. Funct. Genet. 54 (2004) 424–432. doi:10.1002/prot.10604.

[2] J.-M. Choi, S.-S. Han, H.-S. Kim, Industrial applications of enzyme biocatalysis: Current status and future aspects, Biotechnol. Adv. 33 (2015) 1–12. doi:10.1016/j.biotechadv.2015.02.014.

[3] D. Piedrabuena, N. Míguez, A. Poveda, F.J. Plou, M. Fernández-Lobato, Exploring the transferase activity of Ffase from *Schwanniomyces occidentalis*, a β-fructofuranosidase showing high fructosyl-acceptor promiscuity, Appl. Microbiol. Biotechnol. 100 (2016) 8769–8778. doi:10.1007/s00253-016-7628-z.

[4] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. 42 (2014) 490–495. doi:10.1093/nar/gkt1178.

[5] A. Reddy, F. Maley, Studies on identifying the catalytic role of Glu-204 in the active site of yeast invertase, J. Biol. Chem. 271 (1996) 13953–13958. doi:10.1074/jbc.271.24.13953.

[6] A.J.J. Straathof, A. Kieboom, H. Van Bekkum, Invertase-catalysed fructosyl transfer in concentrated solutions of sucrose, Carbohydr. Res. 146 (1986) 154–159. doi.org/10.1016/0008-6215(86)85033-9

[7] A.G. Heyer, Fructan and its relationship to abiotic stress tolerance in plants, Cell. Mol. Life Sci. 66 (2009) 2007–2023. doi:10.1007/s00018-009-0002-x.

[8] T. Tokunaga, T. Oku, N. Hosoya, Influence of chronic intake of new sweetener fructooligosaccharide (Neosugar) on growth and gastrointestinal function of the rat, J. Nutr. Sci. Vitaminol. (Tokyo). 32 (1986) 111–21. doi:10.3177/jnsv.32.111.

[9] M. Álvaro-Benito, A. Polo, B. González, M. Fernández-Lobato, J. Sanz-Aparicio, Structural and kinetic analysis of *Schwanniomyces occidentalis* invertase reveals a new oligomerization pattern and the role of its supplementary domain in substrate binding, J. Biol. Chem. 285 (2010) 13930–13941. doi:10.1074/jbc.M109.095430.

[10] A. Lafraya, J. Sanz-Aparicio, J. Polaina, J. Marin-Navarro, Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase, Appl. Environ. Microbiol. 77 (2011) 6148–6157. doi:10.1128/AEM.05032-11.

[11] M. De Abreu, M. Alvaro-Benito, J. Sanz-Aparicio, F.J. Plou, M. Fernandez-Lobato, M. Alcalde, Synthesis of 6-kestose using an efficient β-fructofuranosidase engineered by directed evolution, Adv. Synth. Catal. 355 (2013) 1698–1702. doi:10.1002/adsc.201200769.

[12] M. Gimeno-pérez, D. Linde, L. Fernández-Arrojo, Heterologous overproduction of β-fructofuranosidase from yeast *Xanthophyllomyces dendrorhous*, an enzyme producing prebiotic

sugars, (2015) 3459–3467. doi:10.1007/s00253-014-6145-1.

[13] S.M. Kritzinger, S.G. Kilian, M.A. Potgieter, J.C. Du Preez, The effect of production parameters on the synthesis of the prebiotic trisaccharide, neokestose, by *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*), Enzyme Microb. Technol. 32 (2003) 728–737. doi:10.1016/S0141-0229(03)00035-8.

[14] M. Ramírez-Escudero, M. Gimeno-Perez, B.G. Lez, D. Linde, Z. Merdzo, M. Fernández-Lobato, J. Sanz-Aparicio, Structural analysis of β-fructofuranosidase from *Xanthophyllomyces dendrorhous* reveals unique features and the crucial role of N-Glycosylation in oligomerization and activity, J. Biol. Chem. 291 (2016) 6843–6857. doi:10.1074/jbc.M115.708495.

[15] D. Linde, I. Macias, L. Fernández-Arrojo, F.J. Plou, A. Jiménez, M. Fernández-Lobato, Molecular and biochemical characterization of a β-fructofuranosidase from *Xanthophyllomyces dendrorhous*, Appl. Environ. Microbiol. 75 (2009) 1065–1073. doi:10.1128/AEM.02061-08.

[16] B.A.G. Dickerson, P-D-Fructofuranosidase from *Claviceps purpurea*, Biochem. J. (1972) 263–272.

[17] J.W. Yun, Fructooligosaccharides—Occurrence, preparation, and application, Enzyme Microb. Technol. 19 (1996) 107–117. doi:10.1016/0141-0229(95)00188-3.

[18] T.-H. Wang, Synthesis of Neofructooligosaccharides, Org. Chem. Insights. (2015) 1. doi:10.4137/OCI.S13222.

[19] S. Farine, C. Versluis, P.J. Bonnici, A. Heck, C. L'Homme, A. Puigserver, A. Biagini, Application of high performance anion exchange chromatography to study invertase-catalysed hydrolysis of sucrose and formation of intermediate fructan products, Appl. Microbiol. Biotechnol. 55 (2001) 55–60. doi:10.1007/s002530000493.

[20] K.M. Trollope, N. van Wyk, M.A. Kotjomela, H. Volschenk, Sequence and structure-based prediction of fructosyltransferase activity for functional sub-classification of fungal GH32 enzymes, FEBS J. (2015) 1–15. doi:10.1111/febs.13536.

[21] D.A. Treco, V. Lundblad, BASIC TECHNIQUES OF YEAST GENETICS, Curr. Protoc. Mol. Biol. Chapter 13 (2001) Unit13.12. doi:10.1002/0471142727.mb1301s23.

[22] C.B. Brachmann, A. Davies, G.J. Cost, E. Caputo, J. Li, P. Hieter, J.D. Boeke, Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications, Yeast. 14 (1998) 115–132. doi:10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2.

[23] K. Leiva, N. Werner, D. Sepúlveda, S. Barahona, M. Baeza, V. Cifuentes, J. Alcaíno, Identification and functional characterization of the CYP51 gene from the yeast *Xanthophyllomyces dendrorhous* that is involved in ergosterol biosynthesis, BMC Microbiol. 15 (2015) 1–12. doi:10.1186/s12866-015-0428-2.

[24] J.M. Crous, I.S. Pretorius, W.H. van Zyl, Cloning and expression of an *Aspergillus kawachii* endo-1,4-β-xylanase gene in *Saccharomyces cerevisiae*, Curr. Genet. 28 (1995) 467–473. https://link.springer.com/article/10.1007/BF00310817.

[25] K.M. Trollope, J.F. Görgens, H. Volschenk, Semirational directed evolution of loop regions in *Aspergillus japonicus* β-fructofuranosidase for improved fructooligosaccharide production, Appl. Environ. Microbiol. 81 (2015) 7319–7329. doi:10.1128/AEM.02134-15.

[26] S.B. Hua, M. Qiu, E. Chan, L. Zhu, Y. Luo, Minimum length of sequence homology required for in vivo cloning by homologous recombination in yeast., Plasmid. 38 (1997) 91–96. doi:10.1006/plas.1997.1305.

[27] K.M. Cho, Y.J. Yoo, H.S. Kang, δ-Integration of endo/exo-glucanase and ß-glucosidase genes into the yeast chromosomes for direct conversion of cellulose to ethanol, Enzyme Microb. Technol. 25 (1999) 23–30. doi:10.1016/S0141-0229(99)00011-3.

[28] C.S. Hoffman, Preparation of yeast DNA, RNA and proteins, Curr. Protoc. Mol. Biol. 13 (1997) 1–4. doi:10.1002/0471142727.mb1312s23.

[29] C.E. Seidman, K. Struhl, Introduction of plasmid DNA into cells, Curr. Protoc. Protein Sci. Appendix 4 (2001) 4D. doi:10.1002/0471140864.psa04ds13.

[30] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. 74 (1977) 5463–5467. doi:10.1073/pnas.74.12.5463.

[31] J. Hill, K.A.G. Donald, D.E. Griffiths, DMSO-enhanced whole cell yeast transformation, Nucleic Acids Res. 19 (1991) 6688. doi:10.1093/nar/19.23.6688.

[32] M. Ling, F. Merante, B.H. Robinson, A rapid and reliable DNA preparation method for screening a large number of yeast clones by polymerase chain reaction, Nucleic Acids Res. 23 (1995) 4924–4925. doi:10.1093/nar/23.23.4924.

[33] J. Sasse, S.R. Gallagher, Staining Proteins in Gels, Curr. Protoc. Mol. Biol. (2009) 1–27. doi:10.1002/0471142727.mb1006s85.

[34] S.R. Gallagher, ELECTROPHORETIC SEPARATION OF PROTEINS, Curr. Protoc. Mol. Biol. (1996).

[35] J. Rehm, L. Willmitzer, A.G. Heyer, Production of 1-kestose in transgenic yeast expressing a fructosyltransferase from *Aspergillus foetidus*, J. Bacteriol. 180 (1998) 1305–1310.

[36] K.M. Trollope, H. Volschenk, J.F. Görgens, Bro R, H.H. Nieuwoudt, Direct, simultaneous quantification of fructooligosaccharides by FT-MIR ATR spectroscopy and chemometrics for rapid identification of superior, engineered β -fructofuranosidases, (2015). doi:10.1007/s00216-014-8417-0.

[37] M. Álvaro-Benito, M. de Abreu, L. Fernández-Arrojo, F.J. Plou, J. Jiménez-Barbero, A. Ballesteros, J. Polaina, M. Fernández-Lobato, Characterization of a β-fructofuranosidase from *Schwanniomyces occidentalis* with transfructosylating activity yielding the prebiotic 6-kestose, J. Biotechnol. 132 (2007) 75–81. doi:10.1016/j.jbiotec.2007.07.939.

[38] N. Shiomi, S. Onodera, N.J. Chatterton, P.A. Harrison, Separation of fructooligosaccharide isomers by anion-exchange chromatography, Agric. Biol. Chem. 55 (1991) 1427–1428. doi:10.1080/00021369.1991.10870771.

[39] F.H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, Nat. Protoc. 2 (2007) 2212–2221. doi:10.1038/nprot.2007.321.

[40] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T.G. Cassarino, M. Bertoni, L. Bordoli, T. Schwede, SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information, Nucleic Acids Res. 42 (2014) 252–258. doi:10.1093/nar/gku340.

[41] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera - A visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612. doi:10.1002/jcc.20084.

[42] M.A. Sainz-Polo, M. Ramírez-Escudero, A. Lafraya, B. Gonza, J. Marín-Navarro, J. Polaina, J. Sanz-Aparicio, Three-dimensional structure of *Saccharomyces* invertase: Role of a non-catalytic domian in oligomerization and substrate specificity, J. Biol. Chem. 288 (2013) 9755–9766. doi:10.1074/jbc.M112.446435.

[43] G. Morris, R. Huey, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791. doi:10.1002/jcc.21256.AutoDock4.

[44] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, L.Y. Geer, S.H. Bryant, CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures, Nucleic Acids Res. 45 (2017) D200–D203. doi:10.1093/nar/gkw1129.

[45] O. Trott, A. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, J. Comput. Chem. 31 (2010) 455–

461. doi:10.1002/jcc.21334.AutoDock.

[46] W.P. Feinstein, M. Brylinski, Calculating an optimal box size for ligand docking and virtual screening against experimental and predicted binding pockets, J. Cheminform. 7 (2015). doi:10.1186/s13321-015-0067-5.

[47] R.A. Laskowski, M.B. Swindells, LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery, J. Chem. Inf. Model. 51 (2011) 2778–2786. doi:10.1021/ci200227u.

[48] J. Liang, C. Woodward, H. Edelsbrunner, Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design, Protein Sci. 7 (1998) 1884–1897. doi:10.1002/pro.5560070905.

[49] S. Gascón, N.P. Neumann, J.O. Lampen, Comparative study of the properties of the purified internal and external invertases from yeast, J. Biol. Chem. 243 (1968) 1573–1577. http://www.jbc.org/content/243/7/1573.abstract.

[50] N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T.B. Acton, G.T. Montelione, D. Baker, Principles for designing ideal protein structures, Nature. 491 (2012) 222–7. doi:10.1038/nature11600.

[51] R.H. Garrett, C.M. Grisham, Biochemistry, 2010. doi:10.1111/j.1468-3156.1975.tb00170.x.

[52] N.P. Neumann, J. Oliver Lampen, Purification and properties of yeast invertase, Biochemistry. 6 (1967) 468–475. doi:10.1021/bi00854a015.

[53] J. Rohrer, Analysis of carbohydrates by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAE-PAD) anion-exchange chromatography, Thermo Fish. Sci. (2013) 1–12. https://tools.thermofisher.com/content/sfs/brochures/TN-20-Analysis-Carbohydrates-HPAE-PAD-TN70671-EN.pdf.

[54] K. Trollope, H. Volschenk, Screening a random mutagenesis library of a fungal β-fructofuranosidase using FT-MIR ATR spectroscopy and multivariate analysis, Appl. Microbiol. Biotechnol. 98 (2013) 4063-4073. doi:10.1007/s00253-013-5419-3.

[55] K.A. Uniewicz, A. Ori, R. Xu, Y. Ahmed, M.C. Wilkinson, D.G. Fernig, E.A. Yates, Differential scanning fluorimetry measurement of protein stability changes upon binding to glycosaminoglycans: A screening test for binding specificity, Anal. Chem. 82 (2010) 3796–3802. doi:10.1021/ac100188x.

[56] T.A. Binkowski, S. Naghibzadeh, J. Liang, CASTp: Computed atlas of surface Topography of proteins, Nucleic Acids Res. 31 (2003) 3352–3355. doi:10.1093/nar/gkg512.

[57] P. Chuankhayan, C.-Y. Hsieh, Y.-C. Huang, Y.-Y. Hsieh, H.-H. Guan, Y.-C. Hsieh, Y.-C. Tien, C.-D. Chen, C.-M. Chiang, C.-J. Chen, Crystal structures of *Aspergillus japonicus* fructosyltransferase complex with donor/acceptor substrates reveal complete subsites in the active site for catalysis, J. Biol. Chem. 285 (2010) 23251–23264. doi:10.1074/jbc.M110.113027.

[58] C. Bignon, G. Sulzenbacher, B. Henrissat, M. Czjzek, The three-dimensional structure of invertase (β-fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases, J. Biol. Chem. 279 (2004) 18903–18910. doi:10.1074/jbc.M313911200.

[59] M. Álvaro-Benito, M.A. Sainz-Polo, D. González-Pérez, B. González, F.J. Plou, M. Fernández-Lobato, J. Sanz-Aparicio, Structural and kinetic insights reveal that the amino acid pair Gln-228/Asn-254 modulates the transfructosylating specificity of *Schwanniomyces occidentalis* β-fructofuranosidase, an enzyme that produces prebiotics, J. Biol. Chem. 287 (2012) 19674–19686. doi:10.1074/jbc.M112.355503.

[60] G.J. Davies, K.S. Wilson, B. Henrissat, Nomenclature for sugar-binding subsites in glycosyl hydrolases, Biochem. J. 321 (1997) 557–559. doi:10.1007/s007920050009.

[61] M. Nagaya, M. Kimura, Y. Gozu, S. Sato, K. Hirano, T. Tochio, A. Nishikawa, T. Tonozuka, Crystal structure of a β-fructofuranosidase with high transfructosylation activity from *Aspergillus kawachii*, Biosci. Biotechnol. Biochem. 8451 (2017) 1–10. doi:10.1080/09168451.2017.1353405.

[62] M. Alvaro-Benito, M. De Abreu, Portillo Francisco, J. Sanz-Aparicio, M. Fernandez-Lobato, New insights into the fructosyltransferase activity of emerging from nonconventional codon usage and directed mutation, Appl. Environ. Microbiol. 76 (2010) 7491–7499. doi:10.1128/AEM.01614-10.

[63] S.Y. Lam, R.C.Y. Yeung, T. Yu, K. Sze, K. Wong, A rigidifying salt-bridge favors the activity of thermophilic enzyme at high temperatures at the expense of low-temperature activity, PLOS Biol. 9 (2011) e1001027. doi:10.1371/journal.pbio.1001027.

[64] M.C. Rubio, R. Runco, A.R. Navarro, Invertase from a strain of *Rhodotorula glutinis*, Phytochemistry. 61 (2002) 605–609. doi:10.1016/S0031-9422(02)00336-9.

[65] C.W. Ford, Simultaneous quantitative determination of sucrose, raffinose and stachyose by invertase hydrolysis and gas-liquid chromatography, J. Sci. Food Agric. 30 (1979) 853–858. doi:10.1002/jsfa.2740300904.

[66] M. Gimeno-Perez, P. Santos-Moriano, L. Fernandez-Arrojo, A. Poveda, J. Jimenez-Barbero, A.O. Ballesteros, M. Fernandez-Lobato, F.J. Plou, Regioselective synthesis of neo-erlose by the β-fructofuranosidase from *Xanthophyllomyces dendrorhous*, Process Biochem. 49 (2014) 423-429. doi:10.1016/j.procbio.2013.12.018.

Acknowledgements

This work was in part supported by research grants of the National Research Foundation of South Africa.

CHAPTER 4

GENERAL DISCUSSION AND CONCLUSIONS

Chapter 4: General Discussion and Conclusions

4.1 General Discussion

The understanding of what governs the hydrolysis and transfructosylating activity in the GH 32 family has been extensively studied [1–8]. This question regarding what governs hydrolysis and transfructosylating activity in the GH 32 family is prevalent as these enzymes are structurally very similar although the activities are vastly different [9]. In plants, the differentiation between the different hydrolytic enzymes and fucosyltransferase enzymes is controlled by compartmentalisation. The research highlights that these enzymes potentially share a common ancestor [10]. However, with the fungal enzymes in the GH 32 family, the exact determinants for hydrolytic and fucosyltransferase differences are still unknown as the underlying principles regarding the two activities is still being unravelled. The classification or differentiation between fungal GH 32 enzymes with predominantly hydrolytic activity and those with predominantly transfructosylating activity was achieved using *in silico* techniques by Trollope *et al.* [11]. Building on the study by Trollope *et al.* [11] the TI loop was identified and proposed to have a role in regulating the DP of FOS being formed by GH 32 enzymes.

This study aimed to mutate the TI loop of *S. cerevisiae* β-fructofuranosidase (Suc2) using a rational enzyme engineering approach to increase the synthesis of high DP FOS. During this study four mutations were used to change Suc2 to increase the synthesis of higher DP FOS; the substitutions were AW, NFS and PGPL. These substitutions led to all the variant enzymes producing higher DP FOS which include both nystose and 1^F-fructofuranosyl nystose (GF_4). The latter is a significant finding as these higher DP FOS are not produced in the wild type Suc2. Various other higher DP FOS products were also observed, but due to the lack of standards these products could not be identified or quantified. A relatively low concentration of higher DP FOS was produced (DP 4 and 5) by the Suc2 variants in contrast to true fructosyltransferase enzymes (such as FopA) where a shift in FOS concentration is observed from DP 3 (1/6-kestose) converted to DP4 (nystose) and then DP5 (GF_4). A possible explanation for most of the products being DP3 in the Suc2 variants is that the mutations introduced opened access to the active site. However, the active site of Suc2 is not conducive to the production of higher DP FOS as it is unable to stabilise nystose and GF_4. Another possible explanation for the predominance of DP 3 FOS products could be that the variants of Suc2 make use of reverse hydrolysis to from FOS. The latter observation is supported by a concentration shift witnessed in fructosyltransferase enzymes which are not observed in the variants as the majority of FOS products are DP 3 (1/6-kestose and neokestose) [9].

The loop variants of Suc2 resulted in a shift in these enzymes from being hydrolytic enzymes to transferase enzymes as evident from the increased production of total FOS. Compared to the wild type enzyme (Suc2), AW and PGPL produced higher total FOS (3-times more total FOS) while NFS produced similar concentrations of total FOS. Due to the destabilising effect of the mutations on variants NFS and PGPL, indicated by low specific activity, the experiment was repeated at a four-time higher enzyme loading. Under these conditions both NFS and PGPL produced 4-times more total FOS than Suc2.

This increase in the transfructosylating activity was associated with a lower affinity for FOS, as the hydrolysis of FOS was slower in the NFS and PGPL variants compared to Suc2 and AW. The increase in the synthesis of FOS was accompanied by a change in the regiospecificity of the variants with the variants changing the ratio of 1-kestose: 6-kestose: neokestose synthesis. Although Suc2 produced neokestose, AW increased the production by 10-times, PGPL produced 6-times more, and NFS increased the production by 3-times. Neokestose is a series of FOS with the β-(2-6)-glycosidic bond between the fructosyl unit and the glucose of sucrose as opposed to 6-kestose where the β-(2-6)-glycosidic bond is between the fructosyl unit and the fructose of sucrose [12]. The production of neokestose by the variants indicates that the introduction of loop mutations led to an active site with more flexibility in terms of the orientation with which the acceptor, sucrose, could loosely bind as indicated by the increase of neokestose production by all variants [8]. Our findings are supported by research in the *X. dendrorhous* β-fructofuranosidase by Ramírez-Escudero *et al.,* [6] who proposed a mechanism for the production of neokestose in this β-fructofuranosidase where the W105 played a crucial role in the orientation of the acceptor sucrose. Interestingly, the W105 residue of the *X. dendrorhous* enzyme correlates with the W48 amino acid in Suc2 [6]. Results of our study support the proposed mechanism by which an increased production of neokestose is observed in some β-fructofuranosidases, especially the W48A *S. cerevisiae* enzyme variant.

The mutations introduced into the loop region of the Suc2 enzyme had a destabilising effect on the variants as indicated by decreased thermostability, i.e., a decrease in the T_m of >4 °C for all the variants compared to Suc2. Although a reduction in the T_m was observed, all the variants' pH optima stayed the same at pH 4.8. As expected, the reduction in the thermostability was accompanied by a reduction in the affinity for sucrose by all the variants. The reduction in the affinity for sucrose is in contrast to the increased FOS production as the acceptor molecule required for FOS production is sucrose, thus indicating a potential disruption of the acceptor site. The lower sucrose affinity could also have contributed to a more

flexible orientation of sucrose leading to increased production of neokestose [8]. The mutation led to a decrease in the specific activity with Suc2's specific activity of 5107 U/mg decreasing to 3300.8 U/mg for AW, followed by PGPL with 43.9 U/mg and NFS with 1.3 U/mg. The decrease in the specific activity supports the decrease in the affinity for the sucrose by the variants.

The aim of the study was met as the changes made in the TI loop of Suc2's β-fructofuranosidase resulted in higher DP being formed by all the variants. Furthermore, changes to the enzyme's catalytic properties were influenced by changing the activity of the enzyme from predominantly hydrolytic to transfructosylating activity and affecting its regiospecificity. All of which when combined indicates that the TI loop plays an essential role within the Suc2 β-fructofuranosidase.

4.2 General conclusions

The variants had increased production of FOS, but this increase came at a cost with a decrease in the specific activity and a decrease in the stability of the enzyme. Furthermore, targeting the TI loop region of Suc2 led to a high success rate regarding the shift from hydrolytic enzyme to transfructosylating enzyme. The TI loop was implicated in the regiospecificity of the enzyme as well, highlighting the critical role the TI loop plays in Suc2. Future research should focus on tailoring the variants of Suc2 with mutations to increase the specific activity and thermostability of the enzymes, i.e., mutations which will increase the activity and stability of the enzymes. The variants can be used as a template with random mutagenesis to stabilise the enzyme and increase its activity. Ultimately, the latter will bring us closer to a deeper understanding of what governs the hydrolysis/transfructosylating activities and understanding the difference between β-fructofuranosidase and fructosyltransferase enzymes.

4.3 References

[1] M. Álvaro-Benito, A. Polo, B. González, M. Fernández-Lobato, J. Sanz-Aparicio, Structural and kinetic analysis of *Schwanniomyces occidentalis* invertase reveals a new oligomerization pattern and the role of its supplementary domain in substrate binding, J. Biol. Chem. 285 (2010) 13930–13941. doi:10.1074/jbc.M109.095430.

[2] M. Álvaro-Benito, M.A. Sainz-Polo, D. González-Pérez, B. González, F.J. Plou, M. Fernández-Lobato, J. Sanz-Aparicio, Structural and kinetic insights reveal that the amino acid pair Gln-228/Asn-254 modulates the transfructosylating specificity of *Schwanniomyces occidentalis* β-fructofuranosidase, an enzyme that produces prebiotics, J. Biol. Chem. 287 (2012) 19674–

19686. doi:10.1074/jbc.M112.355503.

[3] D. Piedrabuena, N. Míguez, A. Poveda, F.J. Plou, M. Fernández-Lobato, Exploring the transferase activity of Ffase from *Schwanniomyces occidentalis*, a β-fructofuranosidase showing high fructosyl-acceptor promiscuity, Appl. Microbiol. Biotechnol. 100 (2016) 8769–8778. doi:10.1007/s00253-016-7628-z.

[4] M. De Abreu, M. Alvaro-Benito, J. Sanz-Aparicio, F.J. Plou, M. Fernandez-Lobato, M. Alcalde, Synthesis of 6-kestose using an efficient β-fructofuranosidase engineered by directed evolution, Adv. Synth. Catal. 355 (2013) 1698–1702. doi:10.1002/adsc.201200769.

[5] M. Gimeno-Perez, P. Santos-Moriano, L. Fernandez-Arrojo, A. Poveda, J. Jimenez-Barbero, A.O. Ballesteros, M. Fernandez-Lobato, F.J. Plou, Regioselective synthesis of neo-erlose by the β-fructofuranosidase from *Xanthophyllomyces dendrorhous*, Process Biochem. 49 (2014) 423-429. doi:10.1016/j.procbio.2013.12.018.

[6] M. Ramírez-Escudero, M. Gimeno-Perez, B.G. Lez, D. Linde, Z. Merdzo, M. Fernández-Lobato, J. Sanz-Aparicio, Structural analysis of β-fructofuranosidase from *Xanthophyllomyces dendrorhous* reveals unique features and the crucial role of N-Glycosylation in oligomerization and activity, J. Biol. Chem. 291 (2016) 6843–6857. doi:10.1074/jbc.M115.708495.

[7] K.M. Trollope, J.F. Görgens, H. Volschenk, Semirational directed evolution of loop regions in *Aspergillus japonicus* β-fructofuranosidase for improved fructooligosaccharide production, Appl. Environ. Microbiol. 81 (2015) 7319–7329. doi:10.1128/AEM.02134-15.

[8] A. Lafraya, J. Sanz-Aparicio, J. Polaina, J. Marin-Navarro, Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase, Appl. Environ. Microbiol. 77 (2011) 6148–6157. doi:10.1128/AEM.05032-11.

[9] M. Antošová, M. Polakovic, Fructosyltransferases : The Enzymes Catalyzing Production, Chem. Pap. 55 (2001) 350–358.

[10] J. De Roover, A. Van Laere, M. De Winter, J.W. Timmermans, W. Van Den Ende, Purification and properties of a second fructan exohydrolase from the roots of *Cichorium intybus*, Physiol. Plant. 106 (1999) 28–34. doi:10.1034/j.1399-3054.1999.106104.x.

[11] K.M. Trollope, N. van Wyk, M.A. Kotjomela, H. Volschenk, Sequence and structure-based prediction of fructosyltransferase activity for functional sub-classification of fungal GH32 enzymes, FEBS J. 282 (2015) 4782-4796. doi:10.1111/febs.13536.

[12] J.W. Yun, Fructooligosaccharides—Occurrence, preparation, and application, Enzyme Microb. Technol. 19 (1996) 107–117. doi:10.1016/0141-0229(95)00188-3.