Engineering a fungal β-fructofuranosidase

by

Kim Mary Trollope

Dissertation presented for the degree of
Doctor of Philosophy (Science)

at
Stellenbosch University

Supervisor: Dr Heinrich Volschenk

Co-supervisor: Prof Johann F. Görgens

March 2015

DECLARATION

By submitting this thesis/dissertation 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.

This dissertation includes one original paper published in a peer-reviewed journal and three unpublished publications. The development and writing of the papers (published and unpublished) were my principal responsibility and, for each of the cases where this is not the case, a declaration is included in the dissertation indicating the nature and extent of the contributions of co-authors.

Date: December 2014 Kim Mary Trollope

SUMMARY

β-fructofuranosidases are hydrolytic enzymes that act on sucrose to yield the products glucose and fructose. Under high substrate conditions these enzymes display fructosyltransferase activity which results in the synthesis of fructooligosaccharides (FOS). Some enzymes display higher propensities for FOS synthesis than others, with the determinants of this activity remaining unclear. The consumption of FOS produces a prebiotic effect that positively alters the composition of the colonic microflora, and as a result is linked to improved human and animal health. The increased demand for FOS has necessitated the industrial production of these nutraceuticals. In enzymatic sucrose biotransformation processes operating at high substrate loading and temperatures between 50 and 60°C, β-fructofuranosidase activity is negatively influenced by glucose product inhibition and thermal instability.

The aim of this study was therefore to engineer the *Aspergillus japonicus* β-fructofuranosidase, FopA, to improve a FOS synthesis bioprocess. A dual approach was employed to engineer FopA so as to increase the probability of obtaining an improved enzyme variant(s). A random mutagenesis approach was applied to harness the potential of the randomness of introduced mutations as precise structural knowledge of the enzyme regions involved in the phenotypic presentation of product inhibition, specific activity and thermal stability was unavailable. A semi-rational approach afforded the additional opportunity to reduce the number of variants to be screened, yet theoretically increased the functional content of the library. This study details the development of a method to rapidly quantify FOS using Fourier transform mid infrared attenuated total reflectance spectroscopy and multivariate data analysis. The method offers improvements over conventionally used high performance liquid chromatography in terms of reduced sample analysis times and the absence of toxic waste products. This is the first report on the direct screening of an enzyme variant library for FOS synthesis to identify improved variants and will significantly support future engineering of β-fructofuranosidases using random mutagenesis approaches. The random mutagenesis approach yielded a variant displaying limited relief from glucose inhibition. At the peak difference in performance, the variant produced 28% more FOS from the same amount of sucrose, when compared to the parent. The semi-rational approach, using a combined crystal structure and evolutionary-guided approach, yielded a four amino acid combination variant displaying improved specific activity and thermostability that was able to reduce the time to completion of an industrial-like FOS synthesis reaction by 26%. The positive outcome of the semi-rational approach showed that engineering loops regions in an enzyme is a feasible strategy to improve both specific activity and thermostability, most probably due to the modification of enzyme structural flexibility.

A bioinformatic tool that enables the identification of β-fructofuranosidases displaying high-level FOS synthesis from protein sequence alone was also developed during the study. These investigations revealed conserved sequence motifs characteristic of enzymes displaying low- and high-level FOS synthesis and a structural loop, unique to the latter group, that were readily applicable identifiers of FOS synthesis capacity. The tool presented may also be useful to improve the understanding of the structure-function relationships of β-fructofuranosidases by facilitating the identification of variations in groups of enzymes that have been functionally sub-classified.

OPSOMMING

β-fruktofuranosidases is hidrolitiese ensieme wat op sukrose inwerk en glukose en fruktose as produkte vorm. Onder toestande met hoë substraatkondisies vertoon hierdie ensieme fruktosieltransferase-aktiwiteit wat tot die sintese van frukto-oligosakkariede (FOS) lei. Sommige ensieme neig na 'n hoër FOS-sintese as ander, maar die bepalende faktore vir hierdie aktiwiteit is nog onbekend. Die verbruik van FOS veroorsaak 'n prebiotiese effek wat die samestelling van kolon mikroflora positief beïnvloed en met verhoogde mens- en dieregesondheid verbind word. Die verhoogde aanvraag vir FOS het die industriële produksie van hierdie nutraseutiese middel genoodsaak. Tydens ensiemgedrewe sukrose-biotransformasieprosesse by hoë substraatladings en temperature tussen 50 en 60 °C, word β-fruktofuranosidase-aktiwiteit negatief deur glukose produkonderdrukking en termiese onstabiliteit beïnvloed.

Die doel van hierdie studie was dus om die *Aspergillus japonicus* β-fruktofuranosidase, FopA, vir 'n verbeterde FOS-sintese bioproses te manipuleer. 'n Tweeledige benadering is vir FopA manipulasie gevolg om die waarskynlikheid van verbeterde variant(e) te verhoog. 'n Lukrake mutagenese benadering, wat die potensiaal van ingevoegde mutasie ewekansigheid inspan, is in die lig van onvoldoende akkurate kennis van die strukturele gedeeltes betrokke by produkinhibisie-, spesifieke aktiwiteit- en termiese stabiliteit fenotipes gevolg. Die toepassing van 'n semi-rasionele benadering het ook geleentheid vir die sifting van 'n kleiner variantbibioloteek geskep, terwyl die funksionele inhoud teoreties verhoog word. Die studie beskryf die ontwikkeling van 'n metode vir die vinnige kwantifisering van FOS, gebaseer op Fourier transform middel infrarooi geattenueerde totale refleksie spektroskopie en meerveranderlike data-analise. Dit is die eerste melding van 'n direkte sifting van 'n ensiemvariantversameling vir FOS-sintese om verbeterde variante te identifiseer, en kan die toekomstige manipulasie van β-fruktofuranosidases deur middel van lukrake mutagenese-benaderings beduidend ondersteun. Die lukrake mutagenese-benadering het 'n variant met beperkte opheffing van glukose-onderdrukking gelewer. By die punt waar die prestasie die meeste verskil, het die variant 28% meer FOS vanaf dieselfde hoeveelheid sukrose geproduseer in vergelyking met die ouer-ensiem. Die semi-rasionele benadering, gegrond op 'n kombinasie van kristalstruktuur en evolusionêre-geleide benaderings, het 'n vier-aminosuurkombinasie variant met hoër spesifieke aktiwiteit en termostabiliteit gelewer wat die voltooiingstyd van 'n tipiese industriële FOS sintesereaksie met 26% kon verkort. Die positiewe uitkoms van die semi-rasionele benadering het aangedui dat manipulasie van die lusgedeeltes in 'n ensiem 'n lewensvatbare strategie is om beide spesifieke aktiwiteit en termostabiliteit te verbeter, moontlik as gevolg van wysigings in die buigsaamheid van die ensiemstruktuur.

’n Bioïnformatika-hulpmiddel vir die identifikasie van β-fruktofuranosidases met hoë vlakke van FOS-sintese op grond van proteïenvolgordes is ook tydens die studie ontwikkel. Motiewe met gekonserveerde volgordes kenmerkend van lae- en hoë-vlak FOS-produserende ensieme en ’n strukturele lus, uniek tot die laasgenoemde groep, is tydens die ondersoek onthul wat as maklike identifiseerders van FOS-sintesekapasiteit kan dien. Die voorgestelde hulpmiddel kan ook nuttig wees om die struktuur-funksie-verwantskap van β-fruktofuranosidases beter te verstaan deur die identifikasie van variasie in ensiemgroepe wat funksioneel gesubklassifiseer is.

BIOGRAPHICAL SKETCH

Kim Mary Trollope was born in Alice, South Africa on 9 May 1977. She matriculated from Victoria Girls' High School in Grahamstown in 1994. She enrolled at Stellenbosch University in 2000 and obtained a BSc degree (Biochemistry, Microbiology and Physiology) in 2002 and a BSc Hons (Wine Biotechnology) degree *cum laude* in 2003. In 2006 she obtained her MSc *cum laude* from the Institute for Wine Biotechnology, Stellenbosch University. Her thesis was entitled 'Investigation of resveratrol production by genetically engineered *Saccharomyces cerevisiae* strains'. In 2010 she enrolled for her PhD at the Department of Microbiology, Stellenbosch University.

ACKNOWLEDGEMENTS

Without the contributions and support of the following people and institutions the completion of this study would not have been possible and I would like to thank:

Dr Heinrich Volschenk, Department of Microbiology, for acting as my supervisor and providing me with guidance, support and many valuable discussions;

Prof Johann Görgens, Department of Process Engineering, for his practical recommendations and resources which greatly facilitated this study;

Dr Hélène Nieuwoudt, Institute for Wine Biotechnology, whose dedication, expertise and enthusiasm were invaluable;

My colleagues and friends at the Department of Microbiology and Process Engineering, especially Niël van Wyk, María García-Aparicio, Soraya Omardien, Jamie Gallant, Heinrich Kroukamp, Gerhardt Coetzee, Momo Kotjomela and Eugene van Rensburg;

The L'Oreal-UNESCO 'For Women In Science' programme which is providing inspiration and support for women in all fields of science;

Antoinette van Zyl (Anatech) and the Eigen Vector Research helpdesk for extremely valuable technical assistance;

Michael Bester;

My family, Winston, Lynne, Penelope and Brett;

Technology Innovation Agency, National Research Foundation and Stellenbosch University for financial support.

DECLARATION BY THE CANDIDATE

This dissertation is presented as a compilation of 7 chapters. Each chapter is introduced separately and is written according to the style of the journal *Applied Microbiology and Biotechnology*.

The contributions of all authors are tabulated below. The overall contributions of the promoters to the study are not quantifiable.

Chapter 1 **General Introduction and project aims**

Authors K.M. Trollope

Nature of contribution	Name	email address	Extent of contribution
Drafting	K.M. Trollope	kim@sun.ac.za	85%
Editing	M. García-Aparicio	garcia@sun.ac.za	5%
	J.F. Görgens	jgorgens@sun.ac.za	5%
	H. Volschenk	volschenkh@sun.ac.za	5%

Chapter 2 **Literature review**

Title The origins and significance of fructooligosaccharides

Authors K.M. Trollope

Nature of contribution	Name	email address	Extent of contribution
Drafting	K.M. Trollope	kim@sun.ac.za	90%
Editing	M. García-Aparicio	garcia@sun.ac.za	2.5%
	J.F. Görgens	jgorgens@sun.ac.za	2.5%
	H. Volschenk	volschenkh@sun.ac.za	5%

Chapter 3 **Research results I**

Title Screening a random mutagenesis library of a fungal β-fructofuranosidase using FT-MIR ATR spectroscopy and multivariate analysis

Authors K.M. Trollope, H.H. Nieuwoudt, J.F. Görgens, H. Volschenk

Nature of contribution	Name	email address	Extent of contribution
Designed & performed experiments, data analysis, manuscript drafting	K.M. Trollope	kim@sun.ac.za	80%
Manuscript editing	H.H. Nieuwoudt	hhn@sun.ac.za	7.5%
	J.F. Görgens	jgorgens@sun.ac.za	5%
	H. Volschenk	volschenkh@sun.ac.za	7.5%

Chapter 4	**Research results II**		
Title	Direct, simultaneous quantification of fructooligosaccharides by FT-MIR ATR spectroscopy and chemometrics for rapid identification of superior, engineered β-fructofuranosidases		
Authors	K.M. Trollope, J.F. Görgens, R. Bro, H. Volschenk, H.H. Nieuwoudt		
Nature of contribution	**Name**	**email address**	**Extent of contribution**
Designed & performed experiments, data analysis, manuscript drafting	K.M. Trollope	kim@sun.ac.za	75%
Manuscript editing	R. Bro	rb@life.ku.dk	6.5%
	H. Volschenk	volschenkh@sun.ac.za	2.5%
	J.F. Görgens	jgorgens@sun.ac.za	1%
	H.H. Nieuwoudt	hhn@sun.ac.za	15%

Chapter 5	**Research results III**		
Title	Directed evolution of loop regions in a fungal β-fructofuranosidase for improved fructooligosaccharide production		
Authors	K.M. Trollope, J.F. Görgens, H. Volschenk		
Nature of contribution	**Name**	**email address**	**Extent of contribution**
Designed & performed experiments, data analysis, manuscript drafting	K.M. Trollope	kim@sun.ac.za	75%
Manuscript editing	J.F. Görgens	jgorgens@sun.ac.za	10%
	H. Volschenk	volschenkh@sun.ac.za	15%

Chapter 6	**Research results IV**		
Title	Sequence-based functional sub-classification of sucrolytic enzymes in glycosyl hydrolase family 32 for the identification of high-level fructooligosaccharide synthesis		
Authors	K.M. Trollope, N. van Wyk, M.A. Kotjomela, J.F. Görgens, H. Volschenk		
Nature of contribution	**Name**	**email address**	**Extent of contribution**
Designed & performed experiments, data analysis	K.M. Trollope	kim@sun.ac.za	70%
	N. van Wyk	nielvw@sun.ac.za	15%
	M.A. Kotjomela	kotjomela@gmail.com	9%
Manuscript drafting	K.M. Trollope	kim@sun.ac.za	
Manuscript editing	J.F. Görgens	jgorgens@sun.ac.za	1%
	H. Volschenk	volschenkh@sun.ac.za	5%

Chapter 7	**General discussion and conclusions**		
Authors	K.M. Trollope		
Nature of contribution	**Name**	**email address**	**Extent of contribution**
Drafting	K.M. Trollope	kim@sun.ac.za	90%
Editing	J.F. Görgens	jgorgens@sun.ac.za	5%
	H. Volschenk	volschenkh@sun.ac.za	5%

Signature of candidate: ……

Date: ……

DECLARATION BY CO-AUTHORS

The undersigned hereby confirm that:

1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to Chapters 1 to 7,
2. no other authors contributed to Chapters 1 to 7, besides those specified above, and
3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapters 1 to 7 of this dissertation.

Signature	Institutional affiliation	Date
M. García-Aparicio	Stellenbosch University	
R. Bro	University of Copenhagen	
H.H. Nieuwoudt	Institute for Wine Biotechnology	
N. van Wyk	Stellenbosch University	
M.A. Kotjomela	Stellenbosch University	
J.F. Görgens	Stellenbosch University	
H. Volschenk	Stellenbosch University	

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 1

INTRODUCTION AND PROJECT AIMS

1 GENERAL INTRODUCTION AND PROJECT AIMS

1.1 INTRODUCTION

The sugar industry is an important contributor to the national economy of South Africa, being a major employment provider in the agricultural sector and source of foreign exchange earnings. It produces 2.2 million tons of sugar per season and generates an annual estimated average direct income of over R12 billion. It is estimated that the livelihoods of approximately 2% of South Africa's population are dependent on the sugar industry (South African Sugar Association, no date). The export market is a residual market for the South African sugar industry after local demand in the Southern African Customs Union market has been satisfied. However, due to large international market price fluctuations, export sugar returns often become unprofitable for the South African sugar producers. The conversion of sucrose to fructooligosaccharides (FOS) presents a strategic opportunity to add value to this commodity product, which could potentially buffer the exposure of the local industry to fluctuating international sugar prices. The retail price of FOS is approximately ten times that of raw sugar and presents an attractive alternative to exporting sucrose (Shedlock, 2014). Based on an estimated €150/kg FOS market price (Dominguez et al. 2014), the added value to South African sucrose could reach as much as 400 times.

FOS are short to medium length chains of β-D-fructans which have the general structure of a glucose unit linked to multiple fructose units. The degree of polymerisation in fructans can range from as little as two to thousands of fructose units, depending on the biological source. Different fructans are distinguished based on the type of glycosidic linkages between the fructose units. Inulins are linear fructans where the fructose units are almost exclusively linked via β(2→1) linkages (Pontis and del Campillo 1985). Due to the nature of the linkages between monomers, inulin-type fructans are resistant to hydrolysis by human digestive enzymes and extreme pH levels in the gastrointestinal tract and are therefore considered non-digestible oligosaccharides (Delzenne and Roberfroid 1994). Reaching the colon intact, FOS selectively stimulate the growth of beneficial bifidobacteria and thus meet all the requirements to be classified as a prebiotic (Gibson and Roberfroid 1995). Sustained consumption of inulin-type fructans is responsible for an altered gut microbial composition, leading to biochemical, genetic and physiological changes in the gut linked to improved human health. Inulin-type fructans are associated with improved immune function and reduced risks of developing osteoporosis, acute and chronic inflammation of the bowel, colorectal cancer, and some of the metabolic disorders associated with obesity (Gibson and Delzenne 2008). Numerous human intervention studies have been conducted with inulin-type fructans that have provided relevant scientific data to substantiate health claims, which is not the case for many other non-digestible oligosaccharides (Roberfroid 2007). Global health estimates by the World Health Organisation (WHO) project

increases in the number of deaths caused by non-communicable diseases from 68.4% in 2015 to 73.9% by 2030. Cancers, including colorectal cancers, diabetes mellitus and cardiovascular diseases contribute substantially to the causes of death among the non-communicable diseases (WHO, 2014). Diet and lifestyle are argued to contribute to the development of the so-called lifestyle diseases, which include the aforementioned non-communicable diseases. The intake of functional foods, dietary components that provide health benefits beyond basic nutrition (Diplock et al. 1999), provides an option to increase the consumption of health promoting constituents which may lower the risks of developing certain diseases. The prebiotic effect of FOS has also been observed in fish and chickens, and FOS, together with other oligosaccharides, are widely applied in animal feeds (Ao & Choct, 2013; Grisdale-Helland, Helland, & Gatlin, 2008; Kim, Seo, Kim, & Paik, 2011). In an attempt to clamp down on the rise of drug resistant microorganisms, the European Union has prohibited the use of antibiotics for non-therapeutic, growth promoting purposes in aquaculture and livestock farming (European Commission, 2005). Prebiotic supplemented animal feeds are therefore used in many European Union member states to improve animal growth performances.

The natural occurrence of FOS in foods extends to onions, asparagus roots, Jerusalem artichoke tubers, garlic, salsify and leeks, some of which are uncommon ingredients in human diets. Owing to their physicochemical properties and sweet taste, FOS from exogenous sources are added to pastry, confectionery and dairy products and their consumption accounts for a major source of FOS in human diets (Molis et al. 1996). The supplementation of foods drives demand for large scale FOS production. The extraction of FOS from plants is a low yielding process, limited by seasonal conditions affecting plant growth (Yun and Song 1999) and for this reason whole cell or *ex vivo* enzyme-based processes are the preferred modes for large scale FOS production. Enzymatic synthesis of FOS from sucrose involves the transfer of fructose units to an acceptor molecule, which is initially sucrose (Edelman and Bacon 1951; Fischer et al. 1951; White and Secor 1952). This reaction is performed by β-fructofuranosidases and fructosyltransferases with 1-kestose, nystose and β-fructofuranosylnystose being the main FOS products. Some enzymes display greater propensities for FOS synthesis than others, with the specific determinants of this ability remaining unclear. Synthesis reactions may be carried out via whole cell biotransformation by FOS-producing organisms such as *Aspergillus* spp. and *Aureobasidium pullulans* (Cruz et al. 1998; Dominguez et al. 2012; Sánchez et al. 2008; Sangeetha et al. 2004). Alternatively, enzyme production by host organisms and FOS synthesis from sucrose may be separated, which allows for discrete optimisation of both steps in a two-stage FOS synthesis process (Maiorano et al. 2008). Although enzymes responsible for the synthesis of FOS occur in plants, fungi and bacteria, fungal enzymes are the preferred industrial biocatalysts owing to their

relatively high specific activities, productivities, FOS yields and thermal stability characteristics (Yun 1996).

Optimisation of a two-stage enzyme-based sucrose-to-FOS bioprocess can be targeted to three main aspects, namely improvement of the enzyme catalyst itself, the production/preparation of the enzyme and the reaction conditions during application of the enzyme. Bioprospecting by screening culturable microbes for novel enzymes with desired properties has been the main route to identifying good catalysts for FOS production from sucrose (Hidaka et al. 1988; Maugeri and Hernalsteens 2007). Heterologous production of novel enzymes in a β-fructofuranosidase-deficient expression host presents a flexible alternative to native hosts for the optimisation of enzyme production owing to the vast array of recombinant DNA and synthetic biology tools available. Extensive knowledge on heterologous host cultivation also exists to maximise enzyme production. Finally, reaction conditions for the *in vitro* synthesis of FOS can easily be manipulated in order to maximise FOS yields and process productivity when using enzyme-based rather than whole-cell biocatalysts. As the latter two aspects of process optimisation, i.e. enzyme production and application of the enzyme, have generally been optimised in most FOS production processes (Dominguez et al. 2014), exploration of new genomes for novel enzymes or improving existing enzyme catalysts presents opportunities with the most scope for improvement of FOS production.

Improving enzyme traits by protein engineering presents a specialist alternative for obtaining novel and more efficient enzyme catalysts for FOS production from sucrose. Strategies for engineering proteins range from completely random mutagenesis of the coding gene to the rational design of variants. In addition to simple implementation, a random mutagenesis strategy has the advantage that little knowledge of the structure-function relationship of the enzyme is required. However, the large enzyme variant library size associated with this strategy is often prohibitive as a rapid, high throughput screening method to identify improved variants is a prerequisite. Based on the available research literature and patent landscape a feasible rapid screening method that directly quantifies FOS to enable the screening of large enzyme variant libraries is unavailable. Only a single β-fructofuranosidase has been engineered using a random, directed evolution approach (de Abreu et al. 2013; de Abreu et al. 2011). Although effective, the screen used to identify improved variants was based on the quantification of reaction products that gave indirect estimates of FOS synthesis, not actual FOS products. In contrast, rational design of variants does not generate large variant libraries but requires a good understanding of the enzyme structure-function relationship to target specific amino acid substitutions that will give the required results. Being a complex matter, clear routes for mutagenesis to predictably improve FOS synthesis without compromising other components of enzyme function have not been resolved. Semi-rational engineering

approaches that offer a mid-way between random and rational approaches are good alternatives to consider for improving β-fructofuranosidases due to the screening and structure-function knowledge limitations.

As opposed to traditional bioprospecting by laborious screening of culturable microbes for FOS synthesis and subsequent gene hunting, *in silico* protein homology-based bioprospecting coupled to heterologous expression of synthetic genes is a time- and labour-saving alternative. With large sequence repositories available in the post-genomics era, obtaining homologous coding sequences does not generally present an inordinate challenge. The β-fructofuranosidase from *A. japonicus* (previously reported as *A. niger* ATCC 20611 by Hidaka et al. (1988)), is currently used for the commercial production of FOS by Meiji Seika Kaisha, one of the first companies to supply FOS. The available scientific literature indicates that this patent-protected enzyme possesses superior FOS synthesis properties compared to the majority of characterised β-fructofuranosidases which are essentially unsuitable for industrial FOS production (Hirayama et al. 1989). Additionally, the sequence of the coding gene is known (Yanai et al. 2001). In order to obtain novel β-fructofuranosidase sequences for FOS synthesis it would appear intuitive to BLAST search the databases using the sequence for the most efficient characterised inulin-type FOS synthesising enzyme, i.e. the *A. japonicus* β-fructofuranosidase sequence (Altschul, Gish, Miller, Myers, & Lipman, 1990). However, FOS-producing fungi often possess multiple fructan-active enzymes, up to 12 in some cases, all classified in the same family as β-fructofuransidases, namely glycosyl hydrolase (GH) family 32 in the sequence-based classification of carbohydrate-active enzymes (Parrent, James, Vasaitis, & Taylor, 2009; Cantarel et al., 2009). Hits obtained from a homology-based search using the *A. japonicus* sequence would thus encompass enzymes with activities ranging between inulinase, levanase, β-fructofuranosidase/invertase and fructosyltransferase (Cantarel et al. 2009; Henrissat and Bairoch 1993). Specific functional classification of enzyme activity for a GH32 protein sequence relies on phylogenetic analysis of functionally characterised enzymes and its contextual position in the associated phylogenetic tree. Putative identification of a sequence to the level of β-fructofuranosidase is possible, however sequence-based discrimination between enzymes that predominantly hydrolyse sucrose or predominantly synthesise FOS has not been reported.

1.2 AIMS OF THE STUDY

The aim of this study was to engineer the *A. japonicus* β-fructofuranosidase (FopA) for improved synthesis of inulin-type FOS from sucrose. It was hypothesised that this process would create new avenues to investigate the determinants of fructosyltransferase activity which

are responsible for the high level FOS synthesis capabilities displayed by this fungal enzyme. The specific objectives for the study were as follows:

i. Enhance the activity of the current industry standard β-fructofuranosidase, FopA, employing random and semi-rational protein engineering approaches. Superior activity included any property, such as increased thermostability, relief from product inhibition and/or higher specific activity. The enzyme was to possess industrially relevant activity levels in order to maximise the bioprocess economics of a two-stage FOS production facility.
ii. Develop a rapid variant library screening method to facilitate directed evolution of FOS synthesising enzymes by random mutagenesis approaches, for which large libraries are a prerequisite. The preferred screening method was to directly quantify the FOS products of enzyme activity.
iii. Develop a protein sequence-based strategy to identify novel GH32 enzymes displaying high-level FOS synthesis.

This body of work contributed to a project funded by the Technology Innovation Agency, a South African government funding agency, to develop a technology package for the commercial production of FOS. Freedom to operate based on a survey of the patent landscape was established. The sequence of the wild type *A. japonicus* β-fructofuranosidase (FopA) is protected by patent US6337201 and equivalent versions of this patent have been filed, and are in effect, in Japan, Taiwan and South Korea. No equivalent versions have been granted in Europe or South Africa. The research done in this study on the FopA β-fructofuranosidase does not infringe on the intellectual property of the owners of patent US6337201 and its equivalents.

This dissertation is presented as a number of chapters covering a review of the scientific literature relevant to the study topic (Chapter 2) and the research performed to meet the aim of the study (Chapters 3 to 6). A general discussion and conclusions are covered in Chapter 7.

The first objective was addressed in Chapters 3 and 5. At the outset of the study two strategies were identified that could be applied to engineer the *A. japonicus* β-fructofuranosidase, using either random mutagenesis or a semi-rational approach. Chapter 3 details the work that combined a random mutagenesis strategy with the development of a potentially high throughput library screening method. It was published as *Screening a random mutagenesis library of a fungal β-fructofuranosidase using FT-MIR ATR spectroscopy and multivariate analysis* in the journal *Applied Microbiology and Biotechnology*. Chapter 5 details the generation of a significantly improved FopA variant using a semi-rational protein engineering approach that combined the aforementioned advantages of random mutagenesis with the small library size

usually associated with the rational design of variants. A manageable library was generated using a combined crystal structure- and evolutionary-guided approach that permitted variant screening to be performed using conventional methods.

The second objective was addressed in Chapters 3 and 4. Chapter 4 extended the screening concept developed in Chapter 3 to encompass simultaneous quantification of residual sucrose substrate, glucose, fructose, 1-kestose and nystose, all sugars relevant to β-fructofuranosidase activity. The capacity to quantify multiple reaction products improved the quality of the screening method that could be applied to identify improved variants and is the first report of a rapid, direct screen for the quantification of FOS.

Chapter 6 describes the work done to address objective three. An *in silico* strategy to identify high-level FOS synthesising GH32 enzymes was developed based on bioinformatic analyses of publicly available sequences and published experimental data on GH32 enzymes. In excess of 30 novel, potentially high-level FOS synthesising enzymes were identified. The described method will facilitate a targeted bioprospecting approach for identifying novel FOS synthesising enzymes, as the departure point is narrowed down substantially from over a few hundred GH32 sequences.

1.3 OUTCOMES

The work presented in this dissertation demonstrates the potential of gaining fundamental knowledge on a topic despite the aim of the work being very much applied. In order to be able to manipulate a system an understanding of its functioning is required. This study has contributed tools that will facilitate the further study of the structure-function relationship of β-fructofuranosidases and their engineering for improved traits. The outcomes of the study included:

i. The generation of new intellectual property pertaining to the construction of a novel FopA variant with improved thermostability, specific activity and altered substrate binding using a semi-rational engineering approach.
ii. A rapid variant library screening method that provides a means to tailor the enzymes by random mutagenesis-based directed evolution or semi-rational approaches that generate large libraries.
iii. A semi-rational engineering strategy, targeted at engineering loop regions, that provides an alternative to random mutagenesis and reduces the variant library screening burden significantly.

iv. A sequence-based method to functionally annotate and identify novel β-fructofuranosidases that may be sufficiently active as is, or can be further engineered to enhance activity for FOS synthesis applications.

Principles developed in this research may be readily applied to the study of other enzyme families to accelerate the identification of specific catalysts for a broad range of applications.

1.4 LITERATURE CITED

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410

Ao, Z., & Choct, M. (2013). Oligosaccharides affect performance and gut development of broiler chickens. *Asian-Australasian Journal of Animal Sciences*, *26*(1), 116–121

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 37:D233–238

Cruz R, Cruz VD, Belini MZ, Belote JG, Vieira CR (1998) Production of fructooligosaccharides by the mycelia of *Aspergillus japonicus* immobilized in calcium alginate. Bioresour Technol 65:139–143

De Abreu M, Alvaro-Benito M, Sanz-Aparicio J, Plou FJ, Fernandez-Lobato M, Alcalde M (2013) Synthesis of 6-Kestose using an Efficient β-Fructofuranosidase Engineered by Directed Evolution. Adv Synth Catal 355:1698–1702

De Abreu MA, Alvaro-Benito M, Plou FJ, Fernández-Lobato M, Alcalde M (2011) Screening β-fructofuranosidases mutant libraries to enhance the transglycosylation rates of β-(2→6) fructooligosaccharides. Comb Chem High Throughput Screen 14:730–738

Delzenne NM, Roberfroid MB (1994) Physiological effects of non-digestible oligosaccharides. Leb Wiss Technol 27:1–6

Diplock AT, Aggett PJ, Ashwell M, Bornet F, Fern EB, Roberfroid MB (1999) Scientific concepts of functional foods in Europe. Consensus document. Br J Nutr 81 Suppl 1:S1–27

Dominguez A, Nobre C, Rodrigues LR, Peres AM, Torres D, Rocha I, Lima N, Teixeira J (2012) New improved method for fructooligosaccharides production by *Aureobasidium pullulans*. Carbohydr Polym 89:1174–1179

Dominguez AL, Rodrigues LR, Lima NM, Teixeira JA (2014) An Overview of the Recent Developments on Fructooligosaccharide Production and Applications. Food Bioprocess Technol 7:324–337

Edelman J, Bacon JSD (1951) Transfructosidation in extracts of the tubers of *Helianthus tuberosus* L. Biochem J 49:529–540

European Commission (2005) Ban on antibiotics as growth promoters in animal feed enters into effect. http://europa.eu/rapid/press-release_IP-05-1687_en.htm#fnB1. Accessed 2 September 2014

Fischer EH, Kohtès L, Fellig J (1951) Propriétés de l'invertase purifiée. Helv Chim Acta 34:1132–1138

Gibson G, Delzenne N (2008) Inulin and oligofructose: New scientific developments. Nutr Today 43:54–59

Gibson GR, Roberfroid MB (1995) Dietary Modulation of the Human Colonic Microbiota: Introducing the Concept of Prebiotics. J Nutr 125:1401–1412

Grisdale-Helland, B., Helland, S. J., & Gatlin, D. M. (2008). The effects of dietary supplementation with mannanoligosaccharide, fructooligosaccharide or galactooligosaccharide on the growth and feed utilization of Atlantic salmon (*Salmo salar*). *Aquaculture*, *283*(1-4), 163–167

Henrissat B, Bairoch A (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 293 (Pt 3):781–788

Hidaka H, Hirayama M, Sumi N (1988) A fructooligosaccharide-producing enzyme from *Aspergillus niger* ATCC 20611. Agric Biol Chem 52:1181–1187

Hirayama M, Sumi N, Hidaka H (1989) Purification and properties of a Fructoooligosaccharide-producing β-fructofuranosidase from *Aspergillus niger* ATCC 20611. Agric Biol Chem 53:667–673

Kim, G.-B., Seo, Y. M., Kim, C. H., & Paik, I. K. (2011). Effect of dietary prebiotic supplementation on the performance, intestinal microflora, and immune response of broilers. *Poultry Science*, *90*(1), 75–82

Maiorano AE, Piccoli RM, da Silva ES, de Andrade Rodrigues MF (2008) Microbial production of fructosyltransferases for synthesis of prebiotics. Biotechnol Lett 30:1867–1877

Maugeri F, Hernalsteens S (2007) Screening of yeast strains for transfructosylating activity. J Mol Catal B Enzym 49:43–49

Molis C, Flourie B, Ouarne F, Gailing MF, Lartigue S, Guibert A, Bornet F, Galmiche JP (1996) Digestion, excretion, and energy value of fructooligosaccharides in healthy humans. Am J Clin Nutr 64:324–328

Parrent J, James T, Vasaitis R, Taylor A (2009) Friend or foe? Evolutionary history of glycoside hydrolase family 32 genes encoding for sucrolytic activity in fungi and its implications for plant-fungal symbioses. BMC Evol Biol 9:148–164

Pontis HG, del Campillo E (1985) Fructans. In: Dey PM, Dixon RA (eds) Biochem. Storage Carbohydrates Green Plants. Academic Press, London, pp 205–227

Roberfroid M (2007) Prebiotics: the concept revisited. J Nutr 137:830S–837S

Sánchez O, Guio F, Garcia D, Silva E, Caicedo L (2008) Fructooligosaccharides production by *Aspergillus* sp. N74 in a mechanically agitated airlift reactor. Food Bioprod Process 86:109–115

Sangeetha PT, Ramesh MN, Prapulla SG (2004) Production of fructo-oligosaccharides by fructosyl transferase from *Aspergillus oryzae* CFR 202 and *Aureobasidium pullulans* CFR 77. Process Biochem 39:755–760

Shedlock MP (2014) Techno-economics of Industrial Scale β-D-fructofuranosidase and Short-chain Fructooligosaccharides Production. Stellenbosch University

South African Sugar Association (SASA) (n.d.) Industry Overview. www.sasa.org.za. Accessed 2 September 2014

White LM, Secor GE (1952) The oligosaccharides formed during the sucrose-invertase reaction. Arch Biochem Biophys 36:490–491

World Health Organization (2014) Health statistics and informations systems. http://www.who.int/healthinfo/global_burden_disease/projections/en. Accessed 2 September 2014

Yanai K, Nakane A, Kawate A, Hirayama M (2001) Molecular cloning and characterization of the fructooligosaccharide-producing β-fructofuranosidase gene from *Aspergillus niger* ATCC 20611. Biosci Biotechnol Biochem 65:766–773

Yun JW (1996) Fructooligosaccharides—Occurrence, preparation, and application. Enzyme Microb Technol 19:107–117

Yun JW, Song SK (1999) Enzymatic Production of Fructooligosaccharides from Sucrose. In: Bucke C (ed) Carbohydr. Biotechnol. Protoc. Humana Press, Totowa, N.J., pp 141–151

CHAPTER 2

LITERATURE REVIEW

The origins and significance of fructooligosaccharides

2 LITERATURE REVIEW:

The origins and significance of fructooligosaccharides

2.1 INTRODUCTION

Fructooligosaccharides (FOS) are short fructose polymers produced by the enzymatic transfer of fructosyl moieties to a sugar (sucrose) acceptor molecule. In essence, this is a glycosylation reaction. Glycosylation of proteins, lipids and small organic molecules can influence their physicochemical and biological properties. Efficient processes to produce these glycosides will find a wide range of applications in the pharmaceutical, cosmetic, tissue culture and food industries (Desmet et al. 2012). Conventional chemical synthesis of glycosides based on organic chemistry principles is not a trivial process and involves multiple steps to obtain the target molecules. Alternatively, enzymatic synthesis can be utilised to achieve similar results with numerous benefits. Biotransformation of substrates by enzymes occurs with high catalytic efficiency and selectivity under mild and environmentally friendly conditions. Combining chemical synthesis with enzymatic catalysis is seen to be the future for optimised glycoside synthesis and benefits from the strengths of both disciplines (Wallace and Balskus 2014). The industrial synthesis of various oligosaccharides by enzymatic means harnesses the potential of enzymes to transfer glycosyl moieties to a variety of natural substrates, for example the industrial biotransformation of sucrose to FOS by β-fructofuranosidase. The enzymatic transfer of glycosyl moieties to alternative substrates such as proteins, lipids and small molecules could require the engineering of the transferase enzymes to accept novel substrates, as well as other properties that influence the efficiency of the process such as specific activity, thermostability and tolerance to organic solvents. Knowledge of the enzyme structure-function relationship will facilitate such endeavours and one of the logical places to realise the acquisition of such an understanding is the engineering of the enzyme and evaluation of the effects on its native activity.

The aim of this study was to engineer the *Aspergillus japonicus* β-fructofuranosidase, FopA, to improve a FOS synthesis bioprocess. In order to engineer this fungal β-fructofuranosidase pertinent information is required to evaluate the relevance and appreciate the necessity for the undertaking. The purpose of this literature review is to provide the relevant background to contextualise the experimental work carried out in this study. Questions raised by the dissertation title such as ‘What is a β-fructofuranosidase?’, ‘Why is it necessary to engineer it?’, ‘Why specifically a fungal candidate?’ and ‘How does one engineer a fungal β-fructofuranosidase?’ will be addressed.

2.2 HYDROLYTIC ENZYMES WITH GLYCOSYLTRANSFERASE ACTIVITY

In the enzyme nomenclature system based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) β-fructofuranosidases are assigned the Enzyme Commission (EC) number 3.2.1.26. By implication they are hydrolases (3), more specifically glycosylases (3.2) hydrolysing *O*- and *S*-glycosyl compounds (3.2.1). Ultimately, they catalyse the hydrolysis of terminal non-reducing β-D-fructofuranoside residues in β-D-fructofuranosides (3.2.1.26) (Myrbäck 1960; Neumann and Lampen 1967). As just detailed, this enzyme classification system is based on the type of reaction catalysed as well as substrate specificity and relies on experimental functional data. The primary substrate for β-fructofuranosidases is sucrose, with glucose and fructose as resultant products of activity. It is one of the earliest studied enzymes with the first report of its isolation being in 1860 by Berthelot (Barnett 2003; Berthelot 1860). β-fructofuranosidases are commonly known as invertases, the name given to them by early researchers, as the enzyme is responsible for the conversion of cane sugar to *invert* sugar consisting of equal parts of dextrorotatory D-glucose and levorotatory D-fructose (O'Sullivan and Tompson 1890). The alternative enzyme classification system, namely the structural- and sequence-based carbohydrate-active enzymes (CAZy) system (Henrissat 1991), will be discussed in section 2.4.

β-fructofuranosidases also possess transfructosidase activity whereby fructooligosaccharides (FOS) are synthesised from sucrose (Edelman and Bacon 1951; Fischer et al. 1951; White and Secor 1952). Other terms used for this activity are *fructotransferase*, *transfructosylation* and *fructosyltransferase*. In general terms an oligosaccharide is a carbohydrate polymer consisting of between two and 10 monosaccharide residues linked by *O*-glycosidic bonds (McNaught 1997). Enzymatic synthesis of glycosidic bonds is performed by the activity of glycosyltransferases whereby a sugar residue from a glycosyl donor is transferred to an acceptor molecule. This route of oligosaccharide synthesis uses the innate biosynthetic machinery of living organisms. Glycosyltransferases are categorised as hexosyltransferases (EC 2.4.1), pentosyltransferases (EC 2.4.2) and those transferring other glycosyl groups (EC 2.4.99), depending on the nature of the sugar residue being transferred. Three main mechanistic groups are derived from the nature of the donor molecule, namely: Leloir-type glycosyltransferases, which utilise sugar nucleotides (e.g. UDP-glucosyltransferases); non-Leloir glycosyltransferases which use sugar-1-phosphates (e.g. phosphorylases); and transglycosidases, which act on non-activated sugars such as sucrose, lactose or starch (Plou et al. 2007). The latter group includes β-fructofuranosidases which usually catalyse the transfer of an enzyme-bound fructose residue to water, giving the hydrolysis products glucose and fructose. *In vitro* glycosidase-catalysed (glycosylase-/glycosyl hydrolase-catalysed) synthesis of oligosaccharides makes use of exo-glycosidases whereby glycosyl transfer occurs only to the

non-reducing terminal monosaccharide unit of substrates. Instead of water, a glycosyl acceptor molecule receives the glycosyl residue from the glycosyl-enzyme intermediate. The synthetic reaction is brought about using the thermodynamic approach, also called reverse hydrolysis, whereby a high concentration of substrate drives the reaction towards equilibrium (Crout and Vic 1998). In the case of FOS synthesis, β-fructofuranosidase is incubated with high concentrations of sucrose (~600 g/l) to favour reverse hydrolysis.

2.3 BIOLOGICAL OCCURRENCE OF FRUCTAN SYNTHESISING ENZYMES

Fructans are polysaccharides where one or more fructosyl-fructose linkages constitute a majority of the linkages in the molecule. Fructans have a general structure of a glucose linked to multiple fructose units. The biosynthesis of these polysaccharides starts with the incorporation of a fructose moiety to one of the three primary hydroxyl groups of sucrose — that is C-1 of the fructose moiety to give 1-kestose; C-6 of the fructose moiety to give 6-kestose; or C-6 of the glucose moiety to give neokestose (Fig. 2.1). Fructans can thus be linear or branched polymers (Taiz and Zeiger 2010). Fructans are naturally produced by bacteria, fungi and plants. Animals have not been found to synthesise fructans (Alméciga-Díaz et al. 2011). The fructan chemical structures, the enzymes responsible for their synthesis and their biological roles differ between the aforementioned groups of organisms and will be discussed separately. Fungal fructan synthesis is the focus of this dissertation and will therefore be discussed in depth. The reason for this focus will also be addressed.

2.3.1 Fructan synthesis in bacteria

2.3.1.1 Types of fructan and responsible genera

Bacteria produce fructans in which the majority of fructose units are β(2→6) linked but some branches may contain β(2→1) linkages (Dedonder 1966). These polymers are known as levans. Many Gram-positive and Gram-negative bacteria produce levan, while inulin synthesis with β(2→1) linkages has been reported so far only in the Gram-positive species *Streptococcus mutans*, *Lactobacillus reuteri* and *Leuconostoc citreum*. Bacterial levan and inulin are the largest fructans in nature, with a degree of polymerisation (DP) reaching up to 100 000 fructose units (Banguela and Hernández 2006).

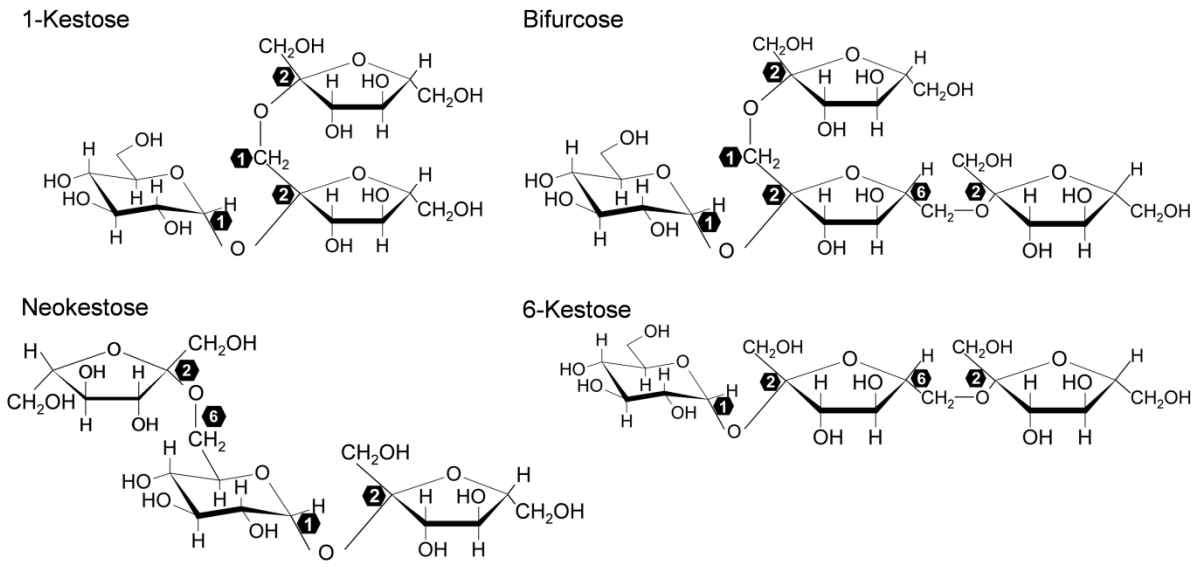

Fig. 2.1 Chemical structures of fructans. Redrawn from Banguela et al. (2006) and Vijn and Smeekens (1999).

Levan producing bacteria include strains belonging to the genera *Bacillus*, *Streptococcus*, *Pseudomonas*, *Erwinia* and *Actinomyces* (Hendry and Wallace 1993). Bacterial levan is produced extracellularly from sucrose by the action of a single enzyme, levansucrase. Levansucrase activity has been assigned as EC 2.4.1.10 which implies glycosyltransferase status (Webb 1992). Acceptor molecules for the transfructosidase reaction include water (sucrose hydrolysis), glucose, fructans and sucrose (Hernandez et al. 1995). Thus the production of low DP FOS and high DP levan are mediated by a single enzyme. Generally, levansucrases from Gram-positive bacteria catalyse the formation of high DP levan without transient accumulation of FOS, whereas the enzymes from Gram-negative species render high levels of inulin type FOS (1-kestose and nystose) with lower yields of high DP levan (Banguela and Hernández 2006). Similarly to β-fructofuranosidases, levansucrases are classified as glycosyl hydrolases (GH) according to the sequence-based classification of carbohydrate-active enzymes (CAZy), although in family GH68 (Cantarel et al. 2009). In addition to synthesis of levan, bacteria possess endo- and exolevanases as well as non-specific β-fructosidases for the hydrolysis of levan (Vijn and Smeekens 1999).

2.3.1.2 Biological role

The biological role of bacterial fructans has been linked to their interactions with eukaryotes as fructan synthesising bacteria are either plant pathogens or symbionts or are associated with oral and gut floras of humans and animals (Velázquez-Hernández et al. 2009; Vijn and Smeekens 1999). Studies involving gene disruption of fructan synthesising enzymes have contributed to the understanding of the role of microbial fructans and are reviewed extensively by Velázquez-

Hernández et al (2009). Fructans contribute to pathogen virulence by supplying energy stored in reserve carbohydrates and influencing biofilm formation. Levans impose high viscosities on aqueous solutions, and their presumed functions involve protecting cells from desiccation and facilitating adherence to surfaces (Kiska and Macrina 1994a), which supports the idea of fructan involvement in biofilm formation. In agreement with previous studies, it has recently been shown that although levan is not an absolute requirement for the formation of *B. subtilis* biofilms, it can be a structural and stabilising component of floating biofilms. Additionally it was proposed that when levan was included in the exopolymeric substances (EPS), it could serve as a nutritional reserve (Dogsa et al. 2013). Fructans contribute to biofilm formation by *Lactobacillus* spp. in the gastrointestinal tract of birds and mammals. Defects in inulosucrase activity, which is responsible for the synthesis of polymeric inulin, resulted in reduced competitiveness relative to wild type stains and impacted on colonisation of the mouse gastrointestinal tract by *L. reuteri* LTH5448 (Walter et al. 2008).

Fructans are also proposed to function in nutrient signalling. The soil and rhizosphere associated *B. subtilis* is proposed to generate the disaccharide, levanbiose, by the action of extracellular levanase. No cellular transporters have been identified and levanbiose is proposed to accumulate in the extracellular environment. Drawing parallels with sugars implicated as signalling molecules regulating carbon metabolism, growth and development in plant cells, Daguer et al. (2004) proposed that levanbiose could act as a signalling molecule between *B. subtilis* cells in soil to control survival functions, or between *B. subtilis* and other microorganisms or plants as part of a molecular dialog to establish corresponding interactions.

The biological functions mediated by fructans in bacteria and other organisms reflect an interaction of processes, triggered by the presence of sucrose-related carbohydrates in the cellular environment, that result in the synthesis and/or hydrolysis of the fructans. Transcriptional regulation of responsible coding genes for enzymes involved in the relevant modifications is reasonably well studied, although further research is required to obtain a clearer indication of the global *in vivo* functioning. In the Gram-negative endophyte, *Gluconacetobacter diazotrophicus*, genes involved in fructan metabolism form an operon. The gene encoding the levansucrase was found to be constitutively expressed while the levanase was induced at low fructose concentrations but repressed in the presence of glucose. The fine regulation of the accumulation of the respective gene products allows the organism to access plant sucrose efficiently (Menéndez et al. 2009). In the Gram-positive *B. subtilis*, the genes for levan synthesis (*sacB*, encoding levansucrase) and degradation (*levB*, encoding endolevanase) are co-transcribed via a sucrose-inducible antitermination mechanism (Crutz et al. 1990; Pereira et al. 2001). Fructose induces exolevanase (*sacC*) transcription which is controlled by catabolite control protein A (CcpA) mediated carbon catabolite repression (CCR) (Martin-

Verstraete et al. 1999; Martin-Verstraete et al. 1995). Although much is known about the mechanisms involved in the regulation of the levanase operon it is still not clear how this relates to *B. subtilis* survival in the soil environment. The high level expression of levansucrase in *B. subtilis* is thought to be related to its extracellular location as the activity may be rapidly diluted by diffusion away from the cells and thus higher amounts of the enzyme are required (Pereira et al. 2001). The high molecular weight of levan limits its diffusion away from cells and access to the storage polymer is retained. Levanases are usually cell-associated and hence are required in lower concentrations. The resulting products from their activity are in close proximity to the cells are thus available for assimilation (Martin-Verstraete et al. 1995).

Streptococcus mutans and *Actinomyces naeslundii*, which inhabit oral environments and produce dental caries, produce inulin or levan via constitutively expressed fructosyltransferases (Bergeron et al. 2000; Kiska and Macrina 1994b). This may contribute to biofilm formation, persistence in the oral cavity and aid attachment to teeth. Similarly to *G. diazatrophicus*, fructose activates *S. mutans* exofructanase transcription but it is suppressed by glucose via a CcpA-independent CCR mechanism (Wen and Burne 2002; Zeng et al. 2006). *A. naeslundii* secretes a sucrose-inducible exolevanase and a fructose-inducible fructanase (Bergeron and Burne 2001).

2.3.2 Fructan synthesis in plants

2.3.2.1 Types of fructan and biosynthesis

In comparison to bacteria, plant fructans display a higher degree of structural diversity. They generally have lower DPs than bacterial fructans with an average size ranging between DP 30–50. Some plant fructans with a DP >200 have been isolated (Altenbach and Ritsema 2007). Owing to the nature of the glycosidic linkages between monomers, five classes of fructans can be distinguished, namely inulin, levan, mixed levan, inulin neoseries and levan neoseries (Vijn and Smeekens 1999). Inulins are linear fructans with fructose units almost exclusively linked via $\beta(2\rightarrow1)$ linkages (Pontis and del Campillo 1985), and are usually found in dicotyledonous plant species belonging to the order Asterales, such as chicory and Jerusalem artichoke (Bonnett et al. 1994; Koops and Jonker 1996). The shortest fructan of the inulin type is 1-kestose. In fructans with higher degrees of polymerisation, the coupling of fructose units proceeds similarly to the bonding of the second fructose to the first (see Fig. 2.1). Levans, are also linear fructans but the fructose units are (mostly) linked via $\beta(2\rightarrow6)$ linkages (Suzuki and Pollock 1986). They are also known as phleins in plants. Fig. 2.1 shows 6-kestose which is the shortest fructan of the levan type. Fructans of the mixed type, which are referred to as graminans (Carpita et al. 1989), have both $\beta(2\rightarrow1)$ and $\beta(2\rightarrow6)$ linkages between the fructose

units and thus contain branches. Levan and graminan fructans have been mainly associated with monocotyledonous plants belonging to families Poaceae, Alliaceae, Asparagaceae, Agavaceae, Amaryllidaceae, Haemodoraceae, and Iridaceae (Banguela and Hernández 2006). However, recently these fructans have been isolated from the eudicot, *Pachysandra terminalis* (Van den Ende et al. 2011). Bifurcose is an example of a graminan fructan (Fig. 2.1). The neoseries fructans contain either linear (2→1)-linked β-D-fructosyl units or linear (2→6)-linked β-D-fructosyl units attached to both C1 and C6 of the glucose moiety of a sucrose molecule. The smallest inulin neoseries molecule is neokestose, shown in Fig. 2.1 (Vijn and Smeekens 1999).

The biosynthesis of fructans is well studied in plants and the synthesis of fructans of different DPs can occur as a result of the activity of three distinct enzymes, namely sucrose: sucrose 1-fructosyltransferase (1-SST, EC 2.4.1.99), fructan:fructan fructosyltransferase (FFT, EC 2.4.1.x) and fructan exohydrolases (FEH; EC 3.2.1.80). The high degree of structural diversity of plant fructans is the result of the action of different FFT enzymes. 1-SST is found in all fructan producing plants and catalyses the transfer of a fructosyl moiety from one vacuolar sucrose molecule to another, producing the trisaccharide, 1-kestose, and a glucose molecule (Van den Ende et al. 2006; Hellwege et al. 1997; Lüscher et al. 2000; Shiomi and Izawa 1980). Trisaccharides may also arise by the reversible transfer of a fructosyl moiety from higher oligomers to sucrose. The polymeric chain is lengthened by the action of FFT - it uses a fructosyl residue from the trisaccharide formed by 1-SST for further chain elongation of sucrose or a fructan polymer, releasing sucrose (Edelman and Jefford 1968; Van Laere and Van Den Ende 2002). In dicotyledonous plants, 1-FFT is responsible for the elongation of 1-kestose to higher DP inulin (Van den Ende et al. 2006). Different FFTs are required to produce the wide variety of fructans found in monocotyledonous plants. Combinations of 1-SST, 1-FFT, 6-SFT (sucrose:fructan 6-fructosyltransferase) and 6G-FFT (fructan:fructan 6G-fructosyltransferase) activities give rise to complex mixtures of inulin, levan, graminan and neoseries fructans. Fructan hydrolysis is catalysed by FEH and acid invertase (Banguela and Hernández 2006; Vijn and Smeekens 1999). A schematic showing a model of fructan biosynthesis in plants is shown in Fig. 2.2. The plant fructan-active enzymes are classified in CAZy family GH32 (Cantarel et al. 2009).

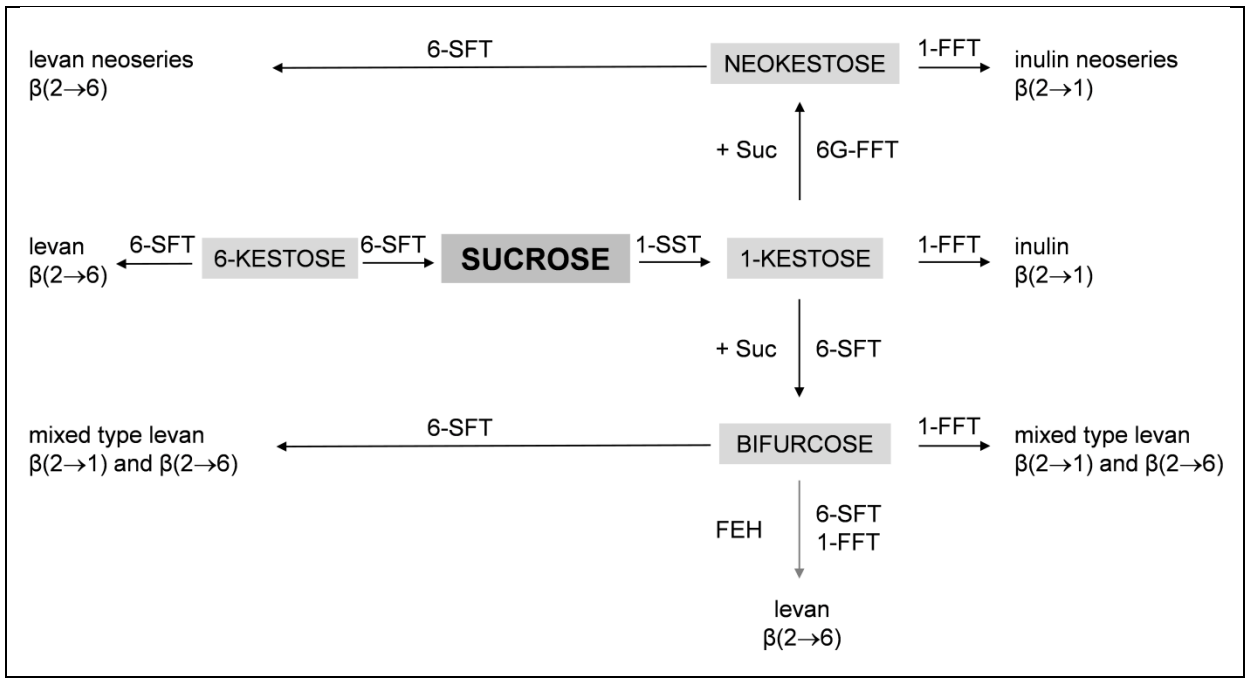

Fig. 2.2 Model of fructan biosynthesis in plants. Starting from sucrose (Suc), structurally different fructan molecules can be produced by the concerted action of different fructosyltransferases. For a detailed description, see text. The light grey arrow shows an alternative route for the production of levan, as suggested by Wiemken et al. (1995). 1-FFT, fructan:fructan fructosyltransferase; 6-SFT, sucrose:fructan 6-fructosyltransferase; 6G-FFT, fructan:fructan 6G-fructosyltransferase; 1-SST, sucrose: sucrose 1-fructosyltransferase; FEH, Fructan exohydrolase. Redrawn from Vijn and Smeekens (1999).

2.3.2.2 Biological role of plant fructans

The least contentious role attributed to fructan synthesis in plants is as an alternative reserve carbohydrate to starch. Fructans differ from starch in that they are soluble polymers produced in vacuoles while insoluble starch is synthesised in plastids. Storage capacity of vacuoles is larger than plastids and capacity may be further increased by the formation of specialised tubers, bulbs or succulent stems. In some plants fructan storage exceeds starch (Brocklebank and Hendry 1989). Fructans accumulate during growth if carbon production exceeds demands and are hydrolysed when energy is required. Vacuolar fructan synthesis results in lowered cellular sucrose concentration and prevents sugar-induced feedback inhibition of photosynthesis (Vijn and Smeekens 1999). The solubility of fructans further relates to osmoregulatory functions in response to changing environmental conditions. An example of osmotic adaptation via fructans is the hydrolysis of high DP fructans by FEHs to release fructose, lowering the osmotic potential which causes water inflow and results in flower opening (Bieleski 1993; Vergauwen et al. 2000).

A further biological role of fructans in plants is linked to drought and cold stress tolerance. As environmental stress responses in plants are complex study topics, direct correlations between stress and fructan accumulation have been difficult to show (Vijn and Smeekens 1999). Further

support for their role in environmental stress tolerance is the global distribution of fructan-accumulating plants. They are especially plentiful in temperate and arid climate zones with seasonal frost or drought periods, while they are virtually absent in tropical regions (Hendry 1993). It has been shown in natural fructan accumulating and transgenic plants that drought and cold stresses result in fructan accumulation. Cellular membrane stabilisation and free radical scavenging by fructans have been demonstrated by *in vitro* studies to provide evidence for the mechanisms of stress tolerance (Livingston et al. 2009; Van den Ende 2013 plus references within). That fructans, raffinose family oligosaccharides (RFOs) and anthocyanins accumulate when sucrose levels are high and sucrose signalling pathway(s) activate plant defence pathways, has led to speculation that small fructan molecules and RFOs are phloem-mobile stress signals and form the basis for 'sugar-based resistance'. These ideas are in their infancy and require further study (Van den Ende 2013).

2.3.3 Fructan synthesis in fungi

2.3.3.1 Types of fructans and fungal enzyme nomenclature

Production of low DP fructans by fungi is well documented. Only three reports are available on the synthesis of high DP inulin (similar molecular weight to bacterial levan) by *Penicillium chrysogenum* and *Aspergillus sydowii* (Heyer and Wendenburg 2001; Kawai et al. 1973; Oláh et al. 1993). It is unclear whether high DP fructan synthesis is uncommon in fungi or if it is not well documented. Strains belonging to many fungal genera have been found to produce FOS and include *Aspergillus*, *Penicillium*, *Aureobasidium*, *Claviceps*, *Fusarium* and *Scopulariopsis* spp. (Yun 1996). Fungi are able to synthesise fructans of the inulin, levan and neoseries types (Alvaro-Benito et al. 2010; Hidaka et al. 1988; Lafraya et al. 2011; Linde et al. 2009). Depending on the enzyme and its fungal source, fructans of up to DP4 can by synthesised from sucrose. Fructans with higher DPs 'derived' from fungi are not usually synthesised from sucrose but rather are produced by the hydrolysis of plant inulin by fungal inulinases (Roberfroid 2007). Low DP fructans (DP<9) are termed oligofructosides, fructooligosaccharides (FOS) or short chain FOS (scFOS) (Velázquez-Hernández et al. 2009). In this review low DP fructans synthesised by fungi will be referred to as FOS.

In contrast to plants and similar to bacteria, FOS are produced from sucrose by the action of a single fungal enzyme. EC activities attributed to the enzymes responsible for FOS synthesis are 3.2.1.26 (β-fructofuranosidase) and 2.4.1.9 (fructosyltransferase). Nomenclature inconsistencies for FOS synthesising enzymes are found throughout the scientific literature. It is likely related to the fact that the primary activity of β-fructofuranosidases is sucrose hydrolysis, with transfructosidase activity usually only dominant under reverse hydrolysis

conditions. Furthermore, transglycosidase activity is a type of glycosyltransferase activity, which most likely gives rise to the term *fructosyltransferase*, and variations thereof. It is supposed that fructosyltransferases do not exhibit hydrolytic activity (water acceptor molecule) but this is inaccurate as glycosyltransferases of the transglycosidase type have been shown to utilise water as an acceptor molecule (Plou et al. 2007). Further complexity is introduced by the enzyme nomenclature schemes of the sequence-based CAZy system and the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). In the CAZy scheme all fructan synthesising enzymes are classified in *glycosyl hydrolases* families (GH32 and GH68) while the NC-IUBMB nomenclature recommendations places them in both *transferase* and *hydrolase* classes. Currently, besides chemical assays of the enzyme activity, no other method exists to distinguish between fungal GH32 enzymes that predominantly hydrolyse sucrose or those that display high level FOS synthesis. Furthermore, even if the enzyme is known to display high level FOS synthesis, it is not clear what aspect of enzyme structure or sequence is responsible for the phenotype.

2.3.3.2 Identification of potential FOS synthesising fungi

The screening of culturable FOS producing fungi has given rise to a technique used to identify strains/enzymes which display superior FOS synthesis abilities. A generalised reaction mechanism for FOS synthesis by fungal enzymes is (Jung et al. 1989):

$$GF_n + GF_n \rightarrow GF_{n-1} + GF_{n+1} \qquad (n = 1\text{-}3)$$

In this reaction the enzyme acts on sucrose in a disproportionation type reaction where one sucrose molecule acts as a fructosyl donor while another acts as the acceptor. If water is the acceptor molecule, sucrose or FOS hydrolysis is the result with free fructose and glucose as products (Yun 1996). Transfructosylating (U_t) and hydrolysing (U_h) activity is calculated from the amount of trisaccharides (GF_2) and fructose, respectively. A ratio of the two activities (U_t:U_h) gives an indication of FOS synthesis performance, with high values being superior (Hidaka et al. 1988). The technique has proven effective but its dependence on reaction conditions and labour intensity are shortfalls. FOS are mainly quantified by high performance liquid chromatography (HPLC) which is sensitive and reasonably accurate but the technique lacks speed and is dependent on expensive, specialised equipment and consumables. As the Ut:Uh technique has been applied for fungal screening, whole cell sucrose biotransformation to FOS under reverse hydrolysis assay conditions are routinely used. Enzyme dosage relative to substrate concentration is an important factor for optimal FOS synthesis and if the enzyme activity is too high, sucrose hydrolysis dominates (Hidaka et al. 1988). The growth and concomitant enzyme production in standardised media by different microbes varies and hence the U_t:U_h ratio may not be a proper reflection of the organism's FOS synthesis potential. The identification of fungi with high level FOS synthesis potential is further complicated by the

quantification of only a single FOS species (GF_2) as an indicator of U_t. Time course studies of enzymatic FOS synthesis indicate that the initial FOS species is GF_2 and as the reaction progresses, it serves as fructosyl acceptor and is converted to the tetrasaccharide (GF_3). A decrease in GF_2 levels thus accompanies GF_3 production (Nishizawa et al. 2001). An enzyme with superior kinetics will generate GF_3 faster than others and the fungal host(s) will be overlooked if only GF_2 is quantified for U_t determination. The principles described above can be applied interchangeably to FOS producing fungi or specific enzymes, as ultimately it is the enzyme that imparts the FOS production phenotype to the fungus. It should be made clear that quantification of multiple FOS synthesis reaction products is important for high confidence in screening data that support the identification of superior enzymes.

2.3.3.3 Biological role of FOS in fungi

The role of FOS in fungal physiology remains unclear. The greater proportion of studies on fungal FOS production are those with a biotechnological focus on the enzyme characteristics and the physiological role of FOS is not examined. FOS are proposed to serve as storage carbohydrates as evidenced by FOS synthesis when fungi are grown in sucrose-containing media. FOS are used as an energy source when the sucrose becomes exhausted (Gupta and Bhatia 1982; Yun 1996). Sucrose serves most plants as a means to transport carbon throughout their tissues, and its abundance within plants makes it a valuable carbon source for many fungi that are associated with plants. In studies of plant-fungal associations, the invertase action of the plant pathogen *Uromyces fabae* β-fructofuranosidase is proposed to assist fungal acquisition of plant-derived sucrose during infection (Voegele et al. 2006). Studies on the biological role of FOS are complicated as the source of FOS species is not readily distinguishable as both the plant and fungus produce similar molecules. Studies that combine analyses of plant and fungal gene transcription data with quantitative FOS data may provide more information on the biological role played by FOS in fungal existence. It is tempting to speculate that FOS are synthesised by fungi in order to interfere with plant signalling and induction of defence mechanisms. Sucrose, glucose and fructose are important signalling molecules for various responses in plants (Tauzin and Giardina 2014) and the alteration of their levels by the activity of fungal β-fructofuranosidases may assist fungal infection. It has been shown that GH32 gene abundance is positively correlated with plant biotrophic fungi which include plant pathogens, endophytic and lichenic fungi but not mycorrhizal fungi. The gene family size has also been shown to be expanding amongst plant pathogens, which implies gene duplication and differentiation of gene product function. For example, twelve GH32 genes were identified in the F*usarium oxysporum* genome (Parrent et al. 2009). A method to functionally characterise GH32 enzymes based on sequence would significantly facilitate the design of

experiments (and interpretation of data) aimed at relating GH32 enzyme activity to fungal adaptation under different nutritional modes.

2.3.3.4 Characterisation of fungal β-fructofuranosidases

The β-fructofuranosidase (invertase) from the yeast *Saccharomyces cerevisiae* is one of the earliest enzymes to be isolated and characterised (Berthelot 1860). The work of Michaelis-Menten on enzyme kinetics was carried out using yeast invertase and the contribution of this enzyme to the study of biochemistry is significant (Michaelis et al. 2011; Michaelis and Menten 1913). Much is known about *Saccharomyces cerevisiae* β-fructofuranosidase but it is intriguing to note that the crystal structure of Suc2p was only recently solved. The work revealed a biological assembly consisting of a 'tetramer of dimers' (Sainz-Polo et al. 2013). The gene sequence for the enzyme was first determined by Taussig and Carlson (1983) and it has been shown that the same gene (*SUC2*) encodes both the intracellular and secreted forms, which are differentially regulated by glucose (Carlson and Botstein 1982). Suc2p is the quintessential β-fructofuranosidase known to hydrolyse sucrose to glucose and fructose. That this enzyme is able to produce FOS is a lesser-known fact (Bacon 1954). It has been shown to produce both inulin- and levan-type trisaccharides, 1-kestose and 6-kestose, respectively from sucrose. Suc2p has also been reported to produce neokestose (Hidaka et al. 1988). No higher DP oligosaccharides were identified following incubation of the enzyme with sucrose (Hidaka et al. 1988; Lafraya et al. 2011). Although Suc2p was incubated under conditions favouring FOS synthesis (reverse hydrolysis conditions with high substrate concentration; 600 g/l sucrose), the yields of FOS were very low, comprising less than 3% of total sugar. The report of Lafraya et al. (2011) is one of a few investigating the structural determinants of fructosyltransferase efficiency and product specificity of a fungal β-fructofuranosidase. Amino acid substitutions were made following identification of variable positions from alignments of fungal, bacterial and plant GH32 hydrolases and transferases, as well as bacterial GH68 transferases. Modifications to the β-fructosidase (WMNDPNG) and ECP motifs surrounding active site residues (underlined) affected increases in fructosyltransferase activity, raising the percent FOS of total sugar from 2.5 to 18.3% after an incubation period of 24-48 h. A decrease in structural rigidity of the active site was proposed to be key to modifying transfructosylating efficiency. Clear determinants of product specificity were not obtained but residues possibly involved were identified.

A fungal β-fructofuranosidase displaying high level FOS synthesis has been isolated from *A. japonicus*. The fungal strain producing this enzyme was initially classified as *A. niger* ATCC 20611 and reports on this enzyme, named FopA, refer to it as an *A. niger* enzyme (Hidaka et al. 1988; Hirayama et al. 1989; Nishizawa et al. 2001; Yanai et al. 2001). The strain was later reclassified as *A. japonicus* ATCC 20611 (ATCC culture collection database;

http://www.lgcpromochem-atcc.com, Yuan et al. 2006). FopA is very well studied and it has been shown to be one of the most efficient FOS synthesising β-fructofuranosidases. The enzyme has been purified and the coding gene sequence is known (Yanai et al. 2001). It produces inulin-type FOS from sucrose with 1-kestose (GF_2), nystose (GF_3) and β-fructofuranosylnystose (GF_4) as major products. By virtue of a U_t:U_h ratio of 14.2, *A. japonicus* ATCC 20611 was initially identified as harbouring an enzyme displaying good FOS synthesis potential. The second best candidate in the screen for FOS synthesising microbes/enzymes was *Aureobasidium pullulans* with a U_t:U_h ratio of 9.8 (Hidaka et al. 1988). The fructosyltransferase activity of FopA has been shown to dominate over hydrolytic activity when tested at high substrate concentration (reverse hydrolysis conditions; 500 g/l sucrose) and low concentrations (50 g/l sucrose). Even at 5 g/l sucrose, FOS synthesis is still observed while under the same conditions no FOS are produced by Suc2p of *S. cerevisiae* (Hidaka et al. 1988; Hirayama et al. 1989). Purified FopA catalyses the synthesis of FOS from sucrose with a yield of ~60% FOS of total sugar within 3 h (Nishizawa et al. 2001). It is this capacity that distinguishes FopA as an enzyme capable of high level FOS synthesis. Both FOS yield and the time span of a few hours to reach given yields are important distinguishing traits of enzymes displaying high level FOS synthesis. What sets FopA apart from other β-fructofuranosidases displaying dominant sucrose hydrolysis activity is not clear. Although strategic substitution of certain amino acids may improve fructosyltransferase activity of β-fructofuranosidases with dominant hydrolytic activity, it has not yet been demonstrated that it is possible to transform such a fungal enzyme into one displaying high level FOS synthesis of the same order as FopA.

Although studies investigating the latter possibility in fungal enzymes are scarce, it is proposed that the determinants of high level FOS synthesis do not only rely on single amino acids at certain positions in the three dimensional enzyme structure. Rather, a particular spatial positioning of amino acids in the folded enzyme, as well as particular key amino acids are responsible for an observed enzyme activity. These determinants are at particular sub-localities within the enzyme's native structure. As protein sequence is an important driver of three dimensional enzyme structure, it is theoretically possible to identify characteristic sequence motifs that give rise to the required positioning of amino acids that produce a certain enzyme activity. The three dimensional structure of *A. japonicus* fructosyltransferase has been solved and is the only crystal structure of an enzyme displaying high level FOS synthesis (Chuankhayan et al. 2010). The extracellular enzyme exists as a monomer. Superimposition of crystal structure models of closely related GH32 and GH68 enzymes with the structure of FopA revealed differentiation of the shape and size of the active site pockets. Despite different residues surrounding the active site pockets with varied shapes between the enzymes, sucrose substrate bound to the different enzymes could be well aligned. Chuankhayan et al. (2010) proposed that amino acid residues surrounding sugar moieties at the +2 subsite, i.e. further

away from active site residues, were responsible for the varied substrate specificities of the different enzymes. The β-fructosidase motif (WMNDPNG) was not conserved in FopA and other GH32 transferases. The motif sequence was QIGDPC in FopA which altered the shape of the active-site pocket due to interactions with other surrounding residues. Mutations at equivalent positions in this region of plant enzymes alter transfructosylation activity and product specificity (Ritsema et al. 2006; Ritsema et al. 2005; Schroeven et al. 2008). Thus indications are there that the β-fructosidase motif is involved in determination of transfructosylation activity.

The synthesis of FOS by the *Schwanniomyces occidentalis* β-fructofuranosidase is also well studied. Although the enzyme displays predominantly sucrose hydrolysis activity, under reverse hydrolysis conditions it synthesises FOS, mostly the levan-type 6-kestose and smaller amounts of 1-kestose (Alvaro-Benito et al. 2007). Particular interest in the *S. occidentalis* β-fructofuranosidase relates to the observation that the 6-kestose yields at 16.9% of total sugars after approximately 20 h are the highest reported for enzymatic synthesis of 6-kestose. The three dimensional structure of the enzyme has been solved and it was revealed to exist as a homodimer (Alvaro-Benito et al. 2010b). Further study of the enzyme revealed that non-universal decoding of the leucine CUG codon by *S. occidentalis* results in a serine at position 196 of the wild type enzyme (Alvaro-Benito et al. 2010a). Both versions of the enzyme, FFase-Ser196 and FFase-Leu196 were compared by heterologous expression in *S. cerevisiae.* The catalytic efficiency of FFase-Leu196 was severely compromised, being 1000 times lower than that of FFase-Ser196. The temperature and pH optima were also decreased. However, FFase-Leu196 showed a 3-fold improved transferase activity and maximum yields of 21.1% FOS of total sugar were attained after 72 h. The synthesis of 1-kestose was also reduced with the 6-kestose:1-kestose ratio increased from 3:1 to 15:1. Mapping the leucine substitution to the crystal structure proposed that a local rearrangement was caused which influenced the positioning of residues impacting on two of the three catalytic triad residues. A further mutation introduced to the β-fructosidase motif caused significant improvements in production levels of both types of trisaccharides. Four rounds of directed evolution of the FFase-Leu196 protein, employing a random mutagenesis strategy, have improved the transfructosylation activity of the *S. occidentalis* β-fructofuranosidase (de Abreu et al. 2013). High selectivity for 6-kestose synthesis remained fairly constant and FOS levels were approximately doubled from 70.7 g/l FOS to 168.3 g/l and 158.2 g/l FOS for two variants, respectively. The synthesis of neokestose and a tetrasaccharide by the variants was also noted. The four mutations responsible for each variant's activity could not be rationalised or were proposed to impact on active site residues. Further work on this enzyme has provided direct evidence for determinants of the ability to degrade long substrates (high DP inulin as well as longer FOS) relating to its dimeric assembly and the determinants of product specificity (Álvaro-Benito et al. 2012).

2.4 EVOLUTIONARY CLASSIFICATION OF FRUCTAN SYNTHESISING ENZYMES

The sequence-based Carbohydrate-Active Enzymes (CAZy) classification system was designed in such a way as to reflect structural features of enzymes better than their single substrate specificity, reveal the evolutionary relationships between groups of enzymes and provide a convenient framework to understand mechanistic properties (Cantarel et al. 2009; www.cazy.org). The traditional biochemical IUBMB classification system for enzymes is based on experimentally determined substrate specificities, which are referred to by Cantarel et al. (2009). Both systems serve users differently but the CAZy system proves a very valuable resource for exploring the evolution of fructan synthesising enzymes. As previously mentioned, bacterial enzymes displaying activity toward fructans are classified in family GH68 while fungal and plant enzymes are classified in family GH32 (Lombard et al. 2014). Together, the families are further grouped into clan GH-J as all members possess a 5-fold β-propeller domain as shown in Fig. 2.3 (Alberto et al. 2004; Davies and Henrissat 1995). Members of a clan display common evolutionary origin, three dimensional structure, homologous positions of catalytic residues and the same stereochemical outcome of glycosidic bond hydrolysis (Henrissat et al. 1995; Naumoff 2011). Clan GH-J members have an aspartate residue acting as catalytic nucleophile while a glutamate residue serves as a catalytic proton donor, and both are situated in the N-terminal β-propeller domain (Fig. 2.3) (Pons et al. 2004). Furthermore GH-J enzymes retain the configuration of the anomeric carbon in the substrate (Koshland and Stein 1954).

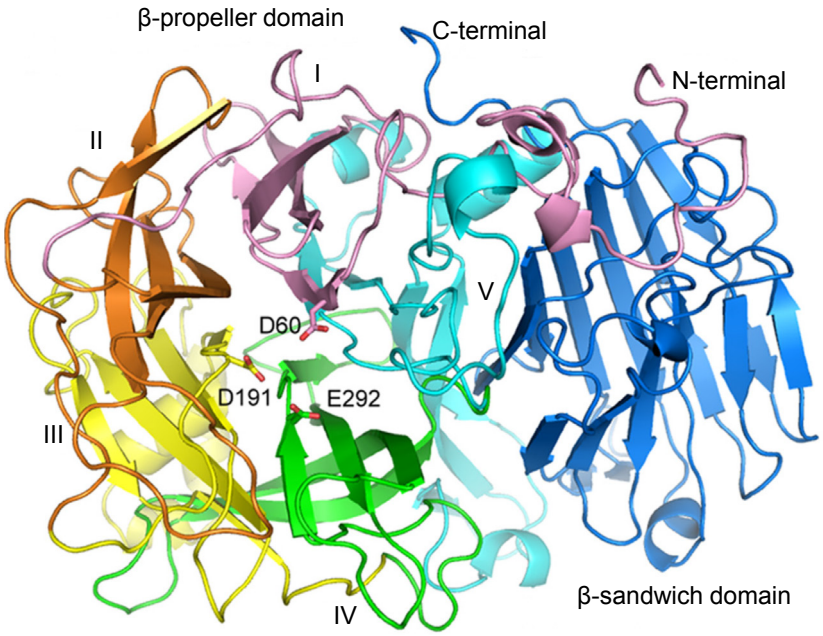

Fig. 2.3 Three dimensional structure of the *Aspergillus japonicus* fructosyltransferase showing the characteristic structural features of a glycosyl hydrolase family 32 (GH32) enzyme. The N-terminal β-propeller domain consists of five blades numbered and coloured in pink, orange, yellow, green, and cyan colours, respectively. The β-sandwich domain is coloured blue and the N- and C-termini are labelled. The catalytic triad residues are shown in *stick* and are also labelled. Reproduced from Chuankhayan et al. (2010)

The main difference between family GH32 and GH68 enzymes is that enzymes belonging to the latter do not possess the C-terminal β-sandwich domain (Fig. 2.3) (Sainz-Polo et al. 2013). Families GH32 and GH68 together with families GH43 (α-L-arabinases) and GH62 (β-xylosidases) constitute the β-fructosidase superfamily as they display related tertiary structures (Naumoff 2001). This classification of fructan-active enzymes reflects an ancient divergence from a common ancestor to acquire novel substrate specificities (Henrissat et al. 1995).

An extensive phylogenetic analysis of fructosyltransferase enzymes in bacteria, plants and fungi revealed that sequences grouped in seven clades, five for plants, one for fungi, and one for bacteria (Alméciga-Díaz et al. 2011). Enzymes from mono- and dicotyledonous plants grouped separately with monocotyledonous enzymes distributed between four clades. Fungal enzymes were splits in two subgroups. The analyses suggested that fructosyltransferases from fungi and bacteria probably evolved from dicotyledonous fructosyltransferases.

2.5 PHYSIOLOGICAL EFFECTS OF FOS CONSUMPTION

The human microbiome project, an international effort to study the interaction between humans and their resident microorganisms and its role in human health and disease (Peterson et al. 2009), is increasing the credibility of past studies that have implicated this interaction in human health. Host-microbe interactions occur mainly along mucosal surfaces with the human intestinal mucosa representing one of the major interfaces. Resident bacteria are tenfold more numerous than human somatic and germ cells and represent a combined microbial genome much larger than the human genome (Shanahan 2002). Collectively, the flora has a metabolic activity equal to a virtual organ within an organ (Bocci 1992). The gut flora has been referred to as the *forgotten organ* (O'Hara and Shanahan 2006) as its presence has historically been largely overlooked. By the 1980s Japanese researchers had demonstrated the influence of FOS on intestinal bifidobacteria (Mitsuoka et al. 1987; Yazawa et al. 1978). The *prebiotic* concept was introduced by Gibson and Roberfroid (1995), the definition of which has been revised to: *A dietary prebiotic is a selectively fermented ingredient that results in specific changes, in the composition and/or activity of the gastrointestinal microbiota, thus conferring benefit(s) upon host health* (Roberfroid et al. 2010). FOS fit the prebiotic definition as they are non-digestible oligosaccharides that reach the colon intact and selectively stimulate the growth of bifidobacteria, which have been linked to improved human health (Gibson and Roberfroid 1995). Mammalian digestive enzymes are unable to degrade the $\beta(2\rightarrow1)$ linkages present in FOS (Rumessen et al. 1990; Stone-Dorshow and Levitt 1987). Sucrose cleavage proceeds by the activity of α-glucosidase which clarifies why FOS are not digested but sucrose is (Conklin et al. 1975; Takesue 1969).

The majority of scientific data on the prebiotic effect, including experimental and human data, have been obtained using food ingredients/supplements containing inulin-type fructans and galacto-oligosaccharides (Roberfroid et al. 2010). This is likely due to the consistent availability of sufficient quantities of the supplements of which the compositions are known - practical considerations impacting on the completion of studies and the reliability of scientific data. It should be noted that numerous claims relating to the prebiotic effect of other oligosaccharides and polysaccharides are yet to be substantiated by human studies to establish direct causal links between their consumption, alteration of gut microbiota and improved human health. All these criteria need to be demonstrated for a substance to be classified as a prebiotic.

The prebiotic effect of inulin-type fructans has been attributed to native plant-derived inulin (DP 2–60, average DP = 12), oligofructose (DP 2–8, average DP = 4), and long-chain inulin known as inulin HP (DP 10–60, average DP = 25) (Roberfroid 2007). The altered composition and/or activity of the colonic microbiota as a result of the consumption of inulin-type fructans can influence host health by: inhibiting the growth of exogeneous and/or harmful bacteria; stimulating immune functions; aiding in digestion and/or absorption of food ingredients/minerals; and synthesising vitamins. Inulin-type fructans are fermented by bifidobacteria to produce short chain fatty acids (SCFA) and lactate which result in decreased pH conditions that exert an antibacterial effect (Gibson and Roberfroid 1995). The low pH as a result of SCFA production is also proposed to be responsible for the improved absorption of calcium and magnesium associated with the prebiotic effect via altered solubility and cation exchange mechanisms (Heijnen et al. 1993). The effectiveness of inulin-type fructans may be linked to the DP of the preparations but generalisations are difficult to make. It has been shown that Synergy 1, a mixture of oligofructose and inulin HP, is more effective than either two of the separate components at improving mineral absorption in rats (Coudray et al. 2003). In humans this effect is not observed in the small intestine and altered mineral absorption is thought to be mediated via altered colonic microflora rather than direct effects on the absorption process in human cells (Roberfroid 2007). Absorbed colonic acetate has recently been shown to influence central appetite regulation which has major implications for combatting obesity (Frost et al. 2014). Butyrate is metabolised by colonic epithelium and it is acknowledged that it may reduce the risk of colon cancer by inducing apoptosis (Perrin et al. 2001). Further pathologies that may be alleviated by the prebiotic effect are type 2 diabetes, metabolic syndrome and irritable bowel syndrome which are extensively reviewed by Roberfroid et al. (2010).

There is an overwhelming body of evidence that supports the concept of modulating human health by the intake of functional foods, including inulin-type FOS, which may conceivably prolong and enhance the quality of life of many individuals. Prebiotics have the advantage that they are stable and not significantly affected by transit through the variable pH conditions of the

stomach and small intestine and reach the colon intact where they can exert their effects on the colonic microbiota. Probiotics (live microorganisms) are more susceptible to long term storage and the transit through the gut, which requires additional measures to ensure their efficacy (Gibson and Roberfroid 1995).

Although much of the focus of the prebiotic effect has been studied in humans, animals also benefit from inulin-type fructan diet supplementation. In young animals diet supplementation with FOS aids in the recovery from digestive tract infections (Apanavicius et al. 2007; Correa-Matos et al. 2003). Additionally supplementation of animal diets improves growth performance of chickens and fish (Akrami et al. 2013; Saleh et al. 2014; Soleimani et al. 2012). With food security becoming an increasingly important consideration, the impact of fructans on the agricultural sector may increase.

2.6 INDUSTRIAL PRODUCTION OF FOS

A FOS market analysis was performed by our group with accurate numbers being difficult to come by. FOS are grouped in the prebiotic market which includes a range of oligosaccharides and sugars. Based on a market price of USD 10/kg and projections of published figures, Shedlock (2014) places the global FOS market between 50 000 and 100 000 tonnes.

2.6.1 Industrial FOS preparations derived from inulin

The compositions of industrial inulin-type fructan preparations are quite varied and contain low DP FOS, high DP inulin or combinations of both. Table 2.1 details the compositions of some commercial fructan preparations. Inulin is derived from chicory roots and the DP varies from 2 to ~60 units with an average DP = 12. The partial hydrolysis of inulin using endoinulinases yields oligofructose in which the DP varies from 2 to 7 with an average DP = 4. It is composed mostly of lower-DP oligosaccharides, namely, 1-kestose (GF_2), nystose (GF_3) and β-fructofuranosylnystose (GF_4), as well as inulobiose (F_2), inulotriose (F_3), and inulotetraose (F_4). Oligomers with DP < 10 may be removed by applying chromatographic separation techniques to produce a high molecular weight inulin-type fructan or inulin HP (Raftiline HP in Table 2.1). A mixture of 2 distinct populations of the low molecular weight oligofructose and the high molecular weight inulin is known as oligofructose-enriched inulin or Synergy (Roberfroid 2007). The preparations produced by Orafti are made from chicory inulin (BENEO GmbH 2013).

2.6.2 Industrial FOS preparations from sucrose

The synthesis of FOS from sucrose is the method used by the remaining companies listed in Table 2.1. The products contain FOS with a DP ranging from 2 to 4 and average DP = 3.6 (similar to oligofructose) (Roberfroid 2007).

Although bacteria, plants and fungi produce enzymes for FOS synthesis, fungal enzymes are preferred for industrial FOS production. Typical reaction conditions are ~50-70% sucrose as substrate, pH5-6.5 and reaction temperatures at 50-60 °C (Yun 1996, Nishizawa et al. 2001). The use of plant enzymes is limited by the requirement of two enzymes for FOS synthesis; the production of two enzymes being uneconomical compared to a single enzyme. Microbial enzymes also display better thermal stabilities than plant enzymes (Yun 1996). The FOS yields of fungal enzymes range between 50-60% of total sugars (Nishizawa et al. 2001) whereas bacterial inulosucrases and levansucrases produce significantly lower yields. Furthermore, levansucrases generally yield high DP levan which has many industrial uses but is not the target fructan of industrial FOS synthesis (Bittencourt et al. 2014; Ozimek et al. 2006). Thus fungal β-fructofuranosidases/fructosyltransferases possess a number of traits that render them more suitable to industrial conditions than their bacterial- and plant-derived counterparts.

There are two main industrial processes for FOS synthesis – a batch system using soluble enzyme or a continuous process using immobilised enzyme or whole cells (Yun 1996). The complete FOS synthesis process is a two-stage operation with initial production of enzyme or microbial biomass under specific conditions followed by the synthesis of FOS in a batch or continuous process using the produced biocatalyst (Fig. 2.4). The high substrate concentrations are used to promote fructosyltransferase enzyme activity and limit downstream processing costs, but together with the relatively high temperatures also limit unwanted microbial contamination of reaction vessels. Immobilisation increases the stability of the enzyme or cells and facilitates easy removal of the catalyst from the final reaction mixture. However, it also limits substrate and product diffusion capacities in column reactors. A batch process requires an additional step to remove enzyme which is not a factor in a continuous process but the latter is limited by lower flow rates required to maximise diffusion of substrates and products in gel matrices (Yun 1996). The productivities of continuous processes using immobilised enzyme are superior to immobilised cells, with values of 1190 g/l h reported for the former (Hayashi et al. 1991) and 180 g/l h for the latter (Yun et al. 1992). The design of a FOS production process therefore must consider catalyst stability, dosage, yields and productivity factors.

Table 2.1 Commercial preparations of inulin-type fructans (Andersen et al. 1999; Nishizawa et al. 2001)

Trade name	FOS content	DP Range	DP average	Free glucose content	Free fructose content	Free sucrose content	Producer
Raftilose L30	28-32% oligofructose, syrup	2-8	NR	6-10%	57-61%	1-5%	Orafti
Raftilose L60	58-62% oligofructose, syrup	2-8	NR	8-12%		28-32%	Orafti
Raftilose L85	83-87% oligofructose, syrup	2-8	NR	6-10%		5-9%	Orafti
Raftilose L95	93.2-95% oligofructose, syrup	2-8	NR	4.2-6.8%			Orafti
Raftilose P95	93-97% oligofructose, powder	2-8	NR	3-7%			Orafti
Raftiline ST	90-94% chicory inulin, powder	2-60	10-12	0-4%		4-8%	Orafti
Raftiline GR	90-94% chicory inulin,granulated powder	2-60	10-12	0-4%		4-8%	Orafti
Raftiline LS	99% chicory inulin	2-60	10-12	1%			Orafti
Raftiline HP	99% chicory inulin	2-60	22-25	0.5%			Orafti
Neosugar G	27% GF_2 + 27% GF_3 + 6% GF_4	3-5	NR	30%	-	10%	Meiji Seika Kaisha
Neosugar P	36% GF_2 + 50% GF_3 + 9% GF_4	3-5	NR	2%	-	3%	Meiji Seika Kaisha
Meioligo P	42% GF_2 + 46% GF_3 +9% GF_4	NR	NR	NR	NR	NR	Meiji Seika Kaisha
Actilight 550S	45±3% GF_2 + 45±3% GF_3 + 10±3% GF_4, FOS 55%, syrup	3-5	NR	45%			Beghin-Meiji
Actilight 950S	37±6% GF_2 + 53±6% GF_3 + 10±4% GF_4, FOS 95%, syrup	3-5	NR	5%			Beghin-Meiji
Actilight 950P	37±6% GF_2 + 53±6% GF_3 + 10±4% GF_4, FOS 95%, powder	3-5	NR	5%			Beghin-Meiji
NutraFlora	95% FOS	NR	NR	NR	NR	NR	Golden Technologies
Oligo-Sugar	60% FOS	NR	NR	NR	NR	NR	Cheil Foods and Chemicals
Fructafit HD	90% inulin	NR	NR	1.5%	4.0%	4.5%	Fructall
Fibruline	NR	2-50	9	NR	NR	NR	Cosucra

NR – not reported
Beghin-Meiji, France
Orafti, Belgium
Cosucra, Belgium
Meiji Seika Kaisha, Japan
Cheil Foods and Chemicals, Korea

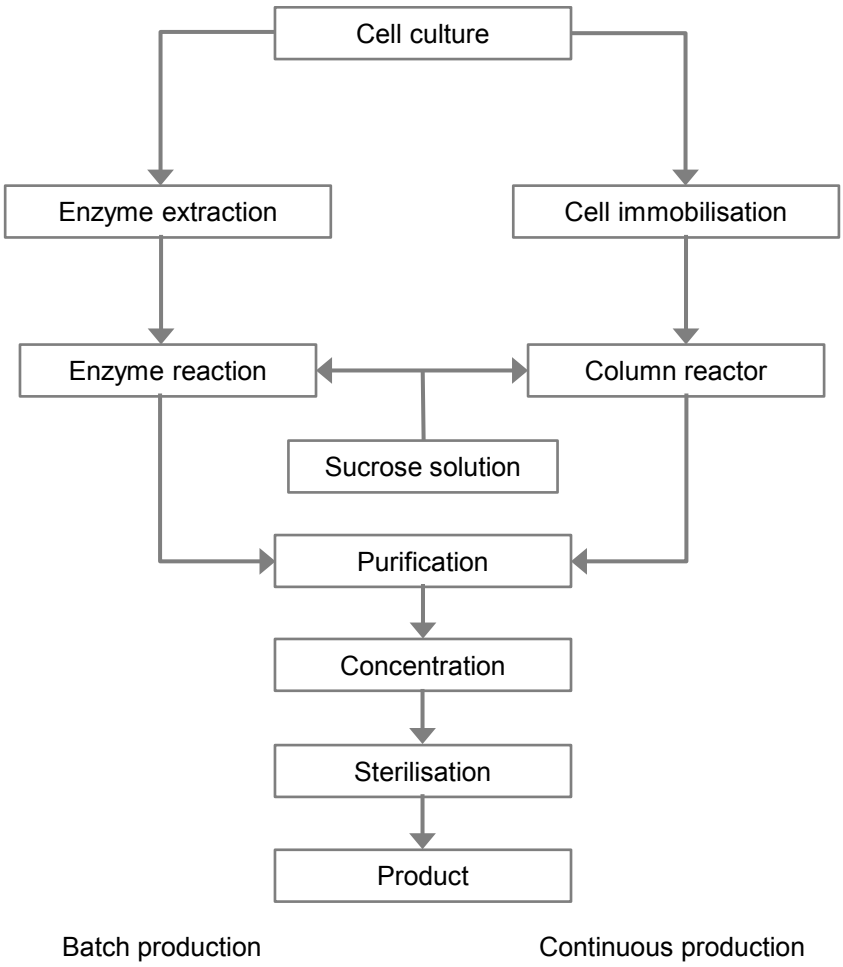

Fig. 2.4 Schematic of the industrial processes for FOS production. Redrawn from Yun et al. (1996).

2.6.3 Opportunities to improve enzymatic FOS synthesis from sucrose

A major limitation of the enzymatic synthesis of FOS from sucrose is enzyme inhibition by the glucose product which limits maximum conversion yields to 55-60% (Dominguez et al. 2013; Jong and Seung 1993; Jung et al. 1989; Sangeetha et al. 2005). The enzyme from *A. pullulans* is competitively inhibited from glucose while that of *A. japonicus* displays non-competitive inhibition (Jung et al. 1989; Nishizawa et al. 2001). The residual sucrose, glucose and fructose (if any) pose problems for certain applications of FOS preparations, for example in products tailored for diabetics. Methods to remove the unwanted sugars to yield high-content FOS preparations include: simulated moving-bed chromatographic separation for the removal of sugars from the final reaction mixture; a mixed enzyme system for FOS synthesis in which glucose oxidase is included to convert glucose to gluconic acid; application of a membrane reactor system using a nano-filtration membrane through which glucose permeates but sucrose and FOS do not (Jong and Seung 1993; Nishizawa et al. 2001; Yun 1996). Only the latter two methods can potentially relieve product inhibition by glucose but they will increase process costs. Additionally, in the mixed enzyme system

optimal temperatures for the two enzymes differ. Also, glucose oxidase requires oxygen, the solubility of which is decreased at the higher temperatures required for optimal β-fructofuranosidase activity. Thus operational parameters have to be adjusted to suit both enzymes, often resulting in a compromise at the expense of productivity (Yun and Song 1999). Engineering the enzyme to relieve product inhibition offers a way to increase FOS yields yet limit additional input costs.

The stability of the enzyme used in the process is affected by relatively high incubation temperatures, typically between 50 and 60 °C. Immobilised cells remain stable for approximately 100 days while immobilised enzymes only remain stable for 35 days, however at lower operational temperatures (Yun and Song 1999). To be able to exploit the higher productivities of immobilised enzyme systems as opposed to cells, improved thermal stability of the enzymes would be advantageous. A batch process would also benefit from higher operational temperatures, permitted by a thermostable enzyme as follows. An increase in temperature is known to affect enzyme reactions on two levels, namely the catalysed reaction itself and the thermal inactivation of the enzyme. At lower temperatures enzyme inactivation is very slow and has no noticeable effect on the rate of the catalysed reaction. Thus the overall reaction rate therefore rises with increases in temperature, as do all chemical reactions. At higher temperatures enzyme inactivation becomes more prominent, causing the effective concentration of active enzyme to decrease during the course of the reaction, accompanied by a reduced reaction rate (Laidler and Peterman 2009). An enzyme displaying higher thermal stability would delay the effect of enzyme inactivation and allow for reactions to be conducted at higher temperatures, reducing the time to completion of the reaction. Again, protein engineering offers a means to achieve such improvements.

The cost of enzymes remains an important factor in the economics of enzyme-catalysed bioprocesses. The higher the specific activity of the enzyme the less is required to complete a given reaction. This impacts the scale of the required enzyme production stage as well as the required enzyme dosage in the FOS synthesis stage. Alternatively, the turnaround time of the FOS synthesis stage can be reduced by an enzyme with higher specific activity, while maintaining set enzyme dosages. Enzyme activity has an impact on the amount of enzyme required per gram of FOS produced and ultimately dictates the design and scale of the facility as well as the operational procedures, all with cost implications (Newman et al. 2013).

2.7 PROTEIN ENGINEERING

2.7.1 Brief overview

Protein engineering is a molecular biology technique that makes use of recombinant DNA technologies and heterologous gene expression. A typical workflow for engineering a protein involves the following steps, all of which have multiple options for their implementation:

i. Mutate and clone the coding gene to create a variant library
ii. Heterologously express the variant library
iii. Quantify (screen) the activities of the enzyme variants to select improved variants.

The approach to pinpoint mutations can be random, semi-rational or rational. The required outcome of the experiment is an important consideration for selecting an engineering approach. Although this seems obvious, it should be stated that amino acid substitution, via gene mutagenesis, is performed either to investigate structure-function relationships in a protein or to engineer improved protein (usually enzyme) function. The former requires targeted amino acid substitutions, the effects of which can be rationalised, hence this approached being termed a rational one. Selecting a random mutagenesis approach to study structure-function relationships would be counterintuitive. That stated it is possible to derive structure-function information from random mutations. Often mutations that could not have been rationally identified cause beneficial changes in enzyme activity (de Abreu et al. 2013; Shimotohno et al. 2001). How the improvement is mediated however, sometimes remains unclear. If the purpose of the work is to improve enzyme specific activity, alter substrate or product specificity and/or increase enzyme robustness in selected process conditions, directed evolution via a random or semi-rational approach is often used. The capacity of the library screening method and prior knowledge of structure and function generally dictate which mutagenesis approach will be used. A summary of the application of different approaches for protein engineering is given in Fig. 2.5.

2.7.2 Advantages and disadvantages of protein engineering approaches

2.7.2.1 Random mutagenesis

Error prone polymerase chain reaction (epPCR) is the pillar on which the random mutagenesis approach is built and has been used to generate mutations in several studies (Chen and Arnold 1993; Cirino et al. 2003; Henke and Bornscheuer 1999; Loo et al. 2004). It is often combined with *in vitro* DNA shuffling to increase sequence diversity (Stemmer 1994).

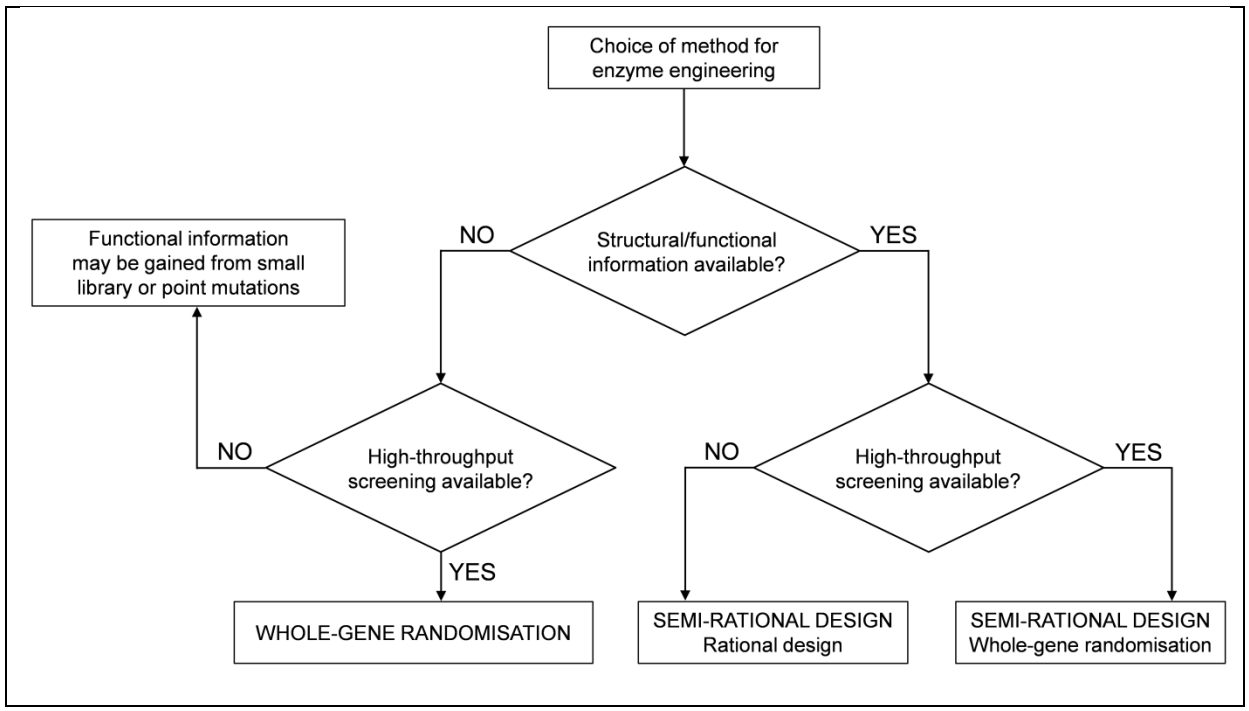

Fig. 2.5 Options for selecting protein engineering approaches are dependent on prior knowledge of structure and function as well as library screening capacity. Redrawn from Chica et al. (2005).

Other techniques to introduce random mutations include chemical mutagenesis and mutator strains (Labrou 2010). The main advantage of a random mutational approach is that it does not require extensive knowledge of the relationships between sequence, structure and function of proteins to execute (Kuchner and Arnold 1997). Additionally, the implementation of epPCR is fairly simple. The method requires that the number of introduced mutations should be kept low as most mutations are deleterious (Suzuki et al. 1996). However, this necessitates very large libraries in order to sample a sufficient number of unique amino acid variants that could significantly improve activity. For example, simultaneously combining up to five amino acid changes in a sequence of 200 amino acids, theoretically generates a library of 6.3×10^{15} variants, given the accessibility of every possible amino acid substitution (Kuchner and Arnold 1997). Due to the degeneracy of the amino acid code not all substitutions are accessible by single nucleotide changes so the aforementioned library size is somewhat reduced (10^{6-9}), unavoidably at the cost of sequence diversity. However, it still remains substantial. Polymerases display mutational biases towards certain nucleotides which also reduces possible sequence diversity (Cline et al. 1996; Tindall and Kunkel 1988). Nevertheless, given the large libraries it is advantageous to include a selection system with random mutagenesis approaches to eliminate inactive variants during the heterologous host transformation step (Tao and Cornish 2002). This reduces the library size to feasible proportions for screening, using a prerequisite high throughput screening (HTS) method. Costly robotic equipment forms an integral part of such methods. Applying a selection system is not always possible and not all enzyme activities are amenable to developing a

HTS method. Neither are all screening methods easy to implement at the required scale, limiting the application of random mutagenesis approaches.

2.7.2.2 Rational design

This was the earliest approach used to engineer proteins. As it relies on an in-depth knowledge of the structural features of an enzyme's active site and their contribution to function (Chica et al. 2005), the approach requires a crystal structure or at least a high quality homology model to identify substitutions to be tested. Such knowledge is not always readily available and is a limiting factor for the application of a rational approach. The clear advantage is that few variants are generated with a very low screening burden, which often increases the feasibility of more complex variant analyses.

2.7.2.3 Semi-rational

A semi-rational approach combines elements of random and rational approaches and bypasses certain limitations of both strategies. Libraries generated in this way are 'smarter' than random libraries as they preselect promising target sites and limit amino acid diversity identified from functional information (from point mutations, random mutagenesis or deduction by sequence alignment) and/or structural information, should it exist (Chica et al. 2005). Computational methods using quantitative mechanics, molecular dynamic calculations as well as machine-learning algorithms are also utilised to explore the impact of amino acid substitutions on protein structure and stability, so aiding in the preselection of mutations (Lutz 2010). Results from enzyme engineering data indicate that most mutations that beneficially affect certain enzyme properties (enantioselectivity, substrate specificity and new catalytic activities) are situated in or near the active site, and more specifically near residues that are involved in ligand binding or catalysis (Dalby 2003; Park et al. 2005; Schallmey et al. 2013; Strausberg et al. 2005). Stability and specific activity can be improved by mutations either near or far from the active site (Chica et al. 2005). Studies in many protein systems indicate that long-range interactions of amino acids are important in binding (and catalytic) specificity. Important biological functions such as substrate recognition in the chymotrypsin family of serine proteases (Hedstrom 1996; Hedstrom et al. 1992), the tuning of antibody specificity through B-cell maturation (Patten et al. 1996) and the cooperativity of oxygen binding in hemoglobin (Paoli et al. 1996; Perutz et al. 1998; Perutz et al. 1987; Perutz 1970) all depend on residues both in direct contact with the substrate and also distant residues located in loops and other secondary structural elements. Crystallographic studies (Paoli et al. 1996; Patten et al. 1996; Williams et al. 1996) have indicated that the distant

residues participating in substrate recognition do so by acting via linked residue positions to control the structure of the substrate-binding site (Süel et al. 2003). The importance of loops to engineer proteins for improved thermal stability (Hardy et al. 1994), improved activity in harsh polar organic solvents (Chen and Arnold 1993; Yedavalli and Rao 2013) and specific activity (Okuda et al. 2013; Ramasubbu et al. 2003) has been demonstrated on numerous occasions. The focus on specific amino acid positions translates into the design of smaller libraries so that a HTS method is no longer a prerequisite. Consideration of evolutionary variability, topological constraints and mechanistic features to select specific amino acids can result in functionally enriched libraries (Lutz 2010). However, the availability of a HTS facilitates the exploration of extended sequence space by methods such as site-saturation mutagenesis which substitutes all 20 amino acids at selected positions (Kotzia and Labrou 2009; Savile et al. 2010). A semi-rational approach allows the level of rational or random elements to be adjusted according to the availability of prior knowledge, giving a flexible option to effectively tailor enzymes for many applications. Commercial gene synthesis has greatly facilitated the manufacture of customised variant libraries and offers a rapid way to access multiple specifically designed mutations, thereby significantly decreasing the labour intensity of library generation (Chen et al. 2010; Govindarajan et al. 2014; Heinzelman et al. 2009).

2.8 CONCLUDING REMARKS

An assessment of the literature led to the conclusion that a sound way to obtain an enzyme with improved activity is to engineer an enzyme with a proven high performance, to further increase its activity. The β-fructofuranosidase from *A. japonicus* was selected as the candidate for further engineering based on the fact that it was a very well-studied enzyme. This prior knowledge created latitude to attempt multiple avenues to improve the enzyme. Furthermore, FopA had a proven ability to efficiently produce FOS at a high level and of the linkage type that has been proven to possess health benefits for its consumers.

The only work done on engineering FopA has used what appeared to be a rational (or semi-rational) approach to generate variants that accumulate either 1-kestose with limited nystose byproduct or nystose with limited β-fructofuranosylnystose byproduct (Nakamura et al. 2010). This patent did not reveal the methodology used to select positions for mutation but only seven single amino acid substitution variants were listed. The use of the variants was in the production of high purity crystals of 1-kestose or nystose which are purported to have ‘excellent characteristics in terms of properties and workability while maintaining the physiological functions of the fructooligosaccharide’. Little could be learnt from this work regarding new engineering strategies for β-fructofuranosidases but the results from improved

variants revealed some information on amino acid residues involved in determining product DP.

There is a paucity of literature on computational methods to identify novel β-fructofuranosidases from genomic data that display high propensities for FOS synthesis. Studies on cellulases have demonstrated the principle of identifying conserved sequence and structural features to differentiate cellulases from non-cellulases in genomic data sets (Sukharnikov et al. 2012; Sukharnikov et al. 2011). Family GH32 enzymes present a similar scenario with multiple enzyme activities (EC classes) grouped in one family. To distinguish, from sequence and structure, between β-fructofuranosidases acting on the same substrate but displaying different levels of activity represents the next level of such computational predictions.

2.9 LITERATURE CITED

De Abreu M, Alvaro-Benito M, Sanz-Aparicio J, Plou FJ, Fernandez-Lobato M, Alcalde M (2013) Synthesis of 6-Kestose using an Efficient β-Fructofuranosidase Engineered by Directed Evolution. Adv Synth Catal 355:1698–1702

Akrami R, Iri Y, Rostami HK, Razeghi Mansour M (2013) Effect of dietary supplementation of fructooligosaccharide (FOS) on growth performance, survival, lactobacillus bacterial population and hemato-immunological parameters of stellate sturgeon (*Acipenser stellatus*) juvenile. Fish Shellfish Immunol 35:1235–1239

Alberto F, Bignon C, Sulzenbacher G, Henrissat B, Czjzek M (2004) The three-dimensional structure of invertase (beta-fructosidase) from Thermotoga maritima reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. J Biol Chem 279:18903–18910

Alméciga-Díaz CJ, Gutierrez AM, Bahamon I, Rodríguez A, Rodríguez M a, Sánchez OF (2011) Computational analysis of the fructosyltransferase enzymes in plants, fungi and bacteria. Gene 484:26–34

Altenbach D, Ritsema T (2007) Structure–function relations and evolution of fructosyltransferases. In: Norio S, Noureddine B, Shuichi O (eds) Recent Advances in Fructooligosaccharides Research. Signpost, Kerala, pp 135–156

Alvaro-Benito M, de Abreu M, Fernández-Arrojo L, Plou FJ, Jiménez-Barbero J, Ballesteros A, Polaina J, Fernández-Lobato M (2007) Characterization of a beta-fructofuranosidase from *Schwanniomyces occidentalis* with transfructosylating activity yielding the prebiotic 6-kestose. J Biotechnol 132:75–81

Alvaro-Benito M, de Abreu M, Portillo F, Sanz-Aparicio J, Fernández-Lobato M (2010a) New insights into the fructosyltransferase activity of *Schwanniomyces occidentalis* ß-fructofuranosidase, emerging from nonconventional codon usage and directed mutation. Appl Environ Microbiol 76:7491–7499

Alvaro-Benito M, Polo A, González B, Fernández-Lobato M, Sanz-Aparicio J (2010b) 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 285:13930–13941

Álvaro-Benito M, Sainz-Polo MA, González-Pérez D, González B, Plou FJ, Fernández-Lobato M, Sanz-Aparicio J (2012) 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 :19674–19686

Andersen R, Andersson C, Gudmundsdóttir E, Hallikainen A, Kudesen I, Mejborn H, Mølck A-M, Paulsen JE, Poulsen M (1999) Safety Evaluation of Fructans. 27–29

Apanavicius CJ, Powell KL, Vester BM, Karr-Lilienthal LK, Pope LL, Fastinger ND, Wallig MA, Tappenden KA, Swanson KS (2007) Fructan supplementation and infection affect food intake, fever, and epithelial sloughing from *Salmonella* challenge in weanling puppies. J Nutr 137:1923–1930

Bacon JS (1954) The oligosaccharides produced by the action of yeast invertase preparations on sucrose. Biochem J 57:320–328

Banguela A, Hernández L (2006) Fructans: from natural sources to transgenic plants. Biotecnol Apl 23:202–210

Barnett JA (2003) Beginnings of microbiology and biochemistry: the contribution of yeast research. Microbiology 149:557–567

BENEO GmbH (2013) Ingredients & Benefits, Human Nutrition, Functional Fibres. http://www.beneo.com/Ingredients_Benefits/Human_Nutrition/Functional_Fibres_1/. Accessed 2 Sep 2014

Bergeron LJ, Burne RA (2001) Roles of Fructosyltransferase and Levanase-Sucrase of *Actinomyces naeslundii* in Fructan and Sucrose Metabolism. Infect Immun 69:5395–5402

Bergeron LJ, Morou-Bermudez E, Burne RA (2000) Characterization of the fructosyltransferase gene of *Actinomyces naeslundii* WVU45. J Bacteriol 182:3649–3654

Berthelot PEM (1860) Sur la fermentation glucosique du sucre de canne. Comptes rendus l'Académie des Sci 50:980–984

Bieleski RL (1993) Fructan Hydrolysis Drives Petal Expansion in the Ephemeral Daylily Flower. Plant Physiol 103:213–219

Bittencourt P, Borsato D, Antonia M, Colabone P (2014) High production of fructooligosaccharides by levansucrase from *Bacillus subtilis* natto CCT 7712. Afr J Biotechnol 13:2734–2740

Bocci V (1992) The neglected organ: bacterial flora has a crucial immunostimulatory role. Perspect Biol Meoid 35:251–260

Bonnett GD, Sims IM, St. John JA, Simpson RJ (1994) Purification and characterization of fructans with beta-2, 1- and beta-2, 6-glycosidic linkages suitable for enzyme studies. New Phytol 127:261–269

Brocklebank KJ, Hendry GAF (1989) Characteristics of plant species which store different types of reserve carbohydrates. New Phytol 112:255–260

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 37:D233–238

Carlson M, Botstein D (1982) Two differentially regulated mRNAs with different 5′ ends encode secreted and intracellular forms of yeast invertase. Cell 28:145–154

Carpita NC, Kanabus J, Housley TL (1989) Linkage Structure of Fructans and Fructan Oligomers from *Triticum aestivum* and *Festuca arundinacea* Leaves. J Plant Physiol 134:162–168

Chen F, Gaucher EA, Leal NA, Hutter D, Havemann SA, Govindarajan S, Ortlund EA, Benner SA (2010) Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. Proc Natl Acad Sci U S A 107:1948–1953

Chen K, Arnold FH (1993) Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc Natl Acad Sci 90:5618–5622

Chica RA, Doucet N, Pelletier JN (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr Opin Biotechnol 16:378–384

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

Cirino P, Mayer K, Umeno D (2003) Generating Mutant Libraries Using Error-Prone PCR. In: Arnold F, Georgiou G (eds) Dir. Evol. Libr. Creat. SE - 1. Humana Press, pp 3–9

Cline J, Braman JC, Hogrefe HH (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res 24:3546–3551

Conklin KA, Yamashiro KM, Gray GM (1975) Human intestinal sucrase-isomaltase. Identification of free sucrase and isomaltase and cleavage of the hybrid into active distinct subunits. J Biol Chem 250:5735–5741

Correa-Matos NJ, Donovan SM, Isaacson RE, Gaskins HR, White BA, Tappenden KA (2003) Fermentable fiber reduces recovery time and improves intestinal function in piglets following *Salmonella typhimurium* infection. J Nutr 133:1845–1852

Coudray C, Tressol JC, Gueux E, Rayssiguier Y (2003) Effects of inulin-type fructans of different chain length and type of branching on intestinal absorption and balance of calcium and magnesium in rats. Eur J Nutr 42:91–98

Crout D, Vic G (1998) Glycosidases and glycosyl transferases in glycoside and oligosaccharide synthesis. Curr Opin Chem Biol 2: 98–111

Crutz A, Steinmetz M, Aymerich S, Richter R, Le Coq D (1990) Induction of levansucrase in *Bacillus subtilis*: an antitermination mechanism negatively controlled by the phosphotransferase system. J Bacteriol 172:1043–1050

Daguer J-P, Geissmann T, Petit-Glatron M-F, Chambert R (2004) Autogenous modulation of the *Bacillus subtilis* sacB–levB–yveA levansucrase operon by the levB transcript. Microbiology 150 :3669–3679

Dalby PA (2003) Optimising enzyme function by directed evolution. Curr Opin Struct Biol 13:500–505

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

Dedonder R (1966) Levansucrase from *Bacillus subtilis*. In: Neufeld EF, Ginsburg V (eds) Methods Enzymol. Academic Press, New York, pp 500–505

Desmet T, Soetaert W, Bojarová P, Křen V, Dijkhuizen L, Eastwick-Field V, Schiller A (2012) Enzymatic glycosylation of small molecules: challenging substrates require tailored catalysts. Chemistry 18:10786–10801

Dogsa I, Brloznik M, Stopar D, Mandic-Mulec I (2013) Exopolymer Diversity and the Role of Levan in *Bacillus subtilis* Biofilms. PLoS One 8:e62044

Dominguez AL, Rodrigues LR, Lima NM, Teixeira JA (2013) An Overview of the Recent Developments on Fructooligosaccharide Production and Applications. Food Bioprocess Technol 7:324–337

Edelman J, Bacon J (1951) Transfructosidation in extracts of the tubers of *Helianthus tuberosus* L. Biochem J 49:529–540

Edelman J, Jefford TG (1968) The mechanisim of fructosan metabolism in higher plants as exemplified in *Helianthus tuberosus*. New Phytol 67:517–531

Fischer EH, Kohtès L, Fellig J (1951) Propriétés de l'invertase purifiée. Helv Chim Acta 34:1132–1138

Frost G, Sleeth ML, Sahuri-Arisoylu M, Lizarbe B, Cerdan S, Brody L, Anastasovska J, Ghourab S, Hankir M, Zhang S, Carling D, Swann JR, Gibson G, Viardot A, Morrison D, Louise Thomas E, Bell JD (2014) The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. Nat Commun 5:3611–3622

Gibson GR, Roberfroid MB (1995) Dietary Modulation of the Human Colonic Microbiota: Introducing the Concept of Prebiotics. J Nutr 125:1401–1412

Govindarajan S, Mannervik B, Silverman JA, Wright K, Regitsky D, Hegazy U, Purcell TJ, Welch M, Minshull J, Gustafsson C (2014) Mapping of Amino Acid Substitutions Conferring Herbicide Resistance in Wheat Glutathione Transferase. ACS Synth Biol. Epub ahead of print

Gupta AK, Bhatia IS (1982) Glucofructosan biosynthesis in *Fusarium oxysporum*: Regulation and substrate specificity of fructosyl transferase and invertase. Phytochemistry 21:1249–1253

Hardy F, Vriend G, Vinne B van der, Frigerio F, Grandi G, Venema G, Ejjsink VGH (1994) The effect of engineering surface loops on the thermal stability of Bacillus subtilis neutral protease. "Protein Eng Des Sel 7:425–430. doi: 10.1093/protein/7.3.425

Hedstrom L (1996) Trypsin: a case study in the structural determinants of enzyme specificity. Biol Chem 377:465–70.

Hedstrom L, Szilagyi L, Rutter WJ (1992) Converting trypsin to chymotrypsin: the role of surface loops. Science 255:1249–53.

Heijnen AM, Brink EJ, Lemmens AG, Beynen AC (1993) Ileal pH and apparent absorption of magnesium in rats fed on diets containing either lactose or lactulose. Br J Nutr 70:747–56

Heinzelman P, Snow CD, Wu I, Nguyen C, Villalobos A, Govindarajan S, Minshull J, Arnold FH (2009) A family of thermostable fungal cellulases created by structure-guided recombination. Proc Natl Acad Sci U S A 106:5610–5615

Hellwege EM, Gritscher D, Willmitzer L, Heyer AG (1997) Transgenic potato tubers accumulate high levels of 1-kestose and nystose: functional identification of a sucrose sucrose 1-fructosyltransferase of artichoke (*Cynara scolymus*) blossom discs. Plant J 12:1057–1065

Hendry G, Wallace R (1993) The origin, distribution, and evolutionary significance of fructans. In: Suzuki M, Chatterton NJ (eds) Sci. Technol. Fruct. CRC Press, Boca Raton, FL, pp 119–139

Hendry GAF (1993) Evolutionary origins and natural functions of fructans – a climatological, biogeographic and mechanistic appraisal. New Phytol 123:3–14

Henke E, Bornscheuer UT (1999) Directed evolution of an esterase from *Pseudomonas fluorescens*. Random mutagenesis by error-prone PCR or a mutator strain and identification of mutants showing enhanced enantioselectivity by a resorufin-based fluorescence assay. Biol Chem 380:1029–1033

Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP, Davies G (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc Natl Acad Sci 92 :7090–7094

Hernandez L, Arrieta J, Menendez C, Vazquez R, Coego A, Suarez V, Selman G, Petit-Glatron MF, Chambert R (1995) Isolation and enzymic properties of levansucrase secreted by *Acetobacter diazotrophicus* SRT4, a bacterium associated with sugar cane. Biochem J 309:113–118

Heyer AG, Wendenburg R (2001) Gene cloning and functional characterization by heterologous expression of the fructosyltransferase of *Aspergillus sydowi* IAM 2544. Appl Env Microbiol 67:363–370

Hidaka H, Hirayama M, Sumi N (1988) A fructooligosaccharide-producing enzyme from *Aspergillus niger* ATCC 20611. Agric Biol Chem 52:1181–1187

Hirayama M, Sumi N, Hidaka H (1989) Purification and properties of a Fructoooligosaccharide-producing β-fructofuranosidase from *Aspergillus niger* ATCC 20611. Agric Biol Chem 53:667–673

Hayashi S, Kinoshita J, Nonoguchi M, Takasaki Y, Imada K (1991) Continous production of 1-kestose by β-fructofuranosidase immobilized on *Shirasu* porous glass. Biotechnol Lett 13:395–398

Jong WY, Seung KS (1993) The production of high-content fructo-oligosaccharides from sucrose by the mixed-enzyme system of fructosyltransferase and glucose oxidase. Biotechnol Lett 15:573–576

Jung KH, Yun JW, Kang KR, Lim JY, Lee JH (1989) Mathematical model for enzymatic production of fructo-oligosaccharides from sucrose. Enzyme Microb Technol 11:491–494

Kawai G, Taniguchi H, Nakamura M (1973) Polyfructan and Oligofructans Synthesized from Sucrose by Conidia of *Aspergillus sydowi* IAM 2544. Agric Biol Chem 37:2111–2119

Kiska DL, Macrina FL (1994a) Genetic analysis of fructan-hyperproducing strains of *Streptococcus mutans*. Infect Immun 62:2679–2686

Kiska DL, Macrina FL (1994b) Genetic regulation of fructosyltransferase in *Streptococcus mutans*. Infect Immun 62:1241–1251

Koops AJ, Jonker HH (1996) Purification and Characterization of the Enzymes of Fructan Biosynthesis in Tubers of *Helianthus tuberosus* Colombia (II. Purification of Sucrose:Sucrose 1-Fructosyltransferase and Reconstitution of Fructan Synthesis in Vitro with Purified Sucrose:Sucrose 1-Fructosyltransferase and Fructan:Fructan 1-Fructosyltransferase). Plant Physiol 110 :1167–1175

Koshland DE, Stein SS (1954) Correlation of bond breaking with enzyme specificity; cleavage point of invertase. J Biol Chem 208:139–148

Kotzia GA, Labrou NE (2009) Engineering thermal stability of L-asparaginase by in vitro directed evolution. FEBS J 276:1750–61

Kuchner O, Arnold FH (1997) Directed evolution of enzyme catalysts. Trends Biotechnol 15:523–30

Labrou NE (2010) Random mutagenesis methods for in vitro directed enzyme evolution. Curr Protein Pept Sci 11:91–100

Lafraya Á, Sanz-Aparicio J, Polaina J, Marín-Navarro J (2011) Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase. Appl Environ Microbiol 77:6148–6157

Laidler KJ, Peterman BF (2009) Temperature effects in enzyme kinetics. In: Purich D (ed) Contemp. Enzym. Kinet. Mech. Third Ed., Third Edit. Academic Press, pp 177–197

Linde D, Macias I, Fernández-Arrojo L, Plou FJ, Jiménez A, Fernández-Lobato M (2009) Molecular and biochemical characterization of a beta-fructofuranosidase from *Xanthophyllomyces dendrorhous*. Appl Environ Microbiol 75:1065–1073

Livingston DP, Hincha DK, Heyer AG (2009) Fructan and its relationship to abiotic stress tolerance in plants. Cell Mol Life Sci 66:2007–2023

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–495

Loo B Van, Spelberg J, Kingma J (2004) Directed Evolution of Epoxide Hydrolase from A. radiobacter toward Higher Enantioselectivity by Error-Prone PCR and DNA Shuffling. Chem Biol 11:981–990

Lüscher M, Hochstrasser U, Vogel G, Aeschbacher R, Galati V, Nelson CJ, Boller T, Wiemken A (2000) Cloning and functional analysis of sucrose:sucrose 1-fructosyltransferase from tall fescue. Plant Physiol 124:1217–2128

Lutz S (2010) Beyond directed evolution – semi-rational protein engineering and design. Curr Opin Biotechnol 21:734–743

Martin-Verstraete I, Deutscher J, Galinier A (1999) Phosphorylation of HPr and Crh by HprK, early steps in the catabolite repression signalling pathway for the *Bacillus subtilis* levanase operon. J Bacteriol 181:2966–2969

Martin-Verstraete I, Stülke J, Klier A, Rapoport G (1995) Two different mechanisms mediate catabolite repression of the *Bacillus subtilis* levanase operon. J Bacteriol 177:6919–6927

McNaught AD (1997) International Union of Pure and Applied Chemistry and International Union of Biochemistry and Molecular Biology - Joint Commission on Biochemical Nomenclature - Nomenclature of carbohydrates (Recommendations 1996). (Reprinted from Pure Appl Chem, vol 68, pp. 1919–2008, 1996) San Diego: Academic Press Inc. – oligosaccharide nomenclature)

Menéndez C, Banguela A, Caballero-Mellado J, Hernández L (2009) Transcriptional regulation and signal-peptide-dependent secretion of exolevanase (LsdB) in the endophyte *Gluconacetobacter diazotrophicus*. Appl Environ Microbiol 75:1782–1785

Michaelis L, Menten M (1913) Die Kinetik der Invertinwirkung. Biochem Z 49:333–369

Michaelis L, Menten ML, Johnson K a, Goody RS (2011) The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. Biochemistry 50:8264–8269

Mitsuoka T, Hidaka H, Eida T (1987) Effect of fructo-oligosaccharides on intestinal microflora. Nahrung 31:427–436

Myrbäck K (1960) Invertases. In: Boyer PD, Lardy H, Myrbäck K (eds) Enzym., 2nd ed. Academic Press, New York, pp 379–396

Nakamura H, Nakane A, Kubota H (2010) β-fructofuranosidase variants

Naumoff DG (2011) Hierarchical classification of glycoside hydrolases. Biochem Biokhimiia 76:622–635

Naumoff DG (2001) β-fructosidase superfamily: homology with some alpha-L-arabinases and beta-D-xylosidases. Proteins 42:66–76

Neumann NP, Lampen JO (1967) Purification and properties of yeast invertase. Biochemistry 6:468–475

Newman RH, Vaidya AA, Sohel MI, Jack MW (2013) Optimizing the enzyme loading and incubation time in enzymatic hydrolysis of lignocellulosic substrates. Bioresour Technol 129:33–38

Nishizawa K, Nakajima M, Nabetani H (2001) Kinetic study on transfructosylation by β-fructofuranosidase from *Aspergillus niger* ATCC 20611 and availability of a membrane reactor for fructooligosaccharide production. Food Sci Technol Res 7:39–44

O'Hara AM, Shanahan F (2006) The gut flora as a forgotten organ. EMBO Rep 7:688–93

O'Sullivan C, Tompson FW (1890) LX. - Invertase: a contribution to the history of an enzyme or unorganised ferment. J Chem Soc Trans 57:834

Okuda M, Ozawa T, Tohata M, Sato T, Saeki K, Ozaki K (2013) A single mutation within a Ca(2+) binding loop increases proteolytic activity, thermal stability, and surfactant stability. Biochim Biophys Acta 1834:634–641

Oláh A, Papp Z, Szentirmai A (1993) Inulin formation of penicillin producing industrial *Penicillium chrysogenum* strains. Acta Microbiol Hung 40:379–386

Ozimek LK, Kralj S, Maarel MJEC Van Der, Dijkhuizen L (2006) The levansucrase and inulosucrase enzymes of *Lactobacillus reuteri* 121 catalyse processive and non-processive transglycosylation reactions. Microbiology 1187–1196

Paoli M, Liddington R, Tame J, Wilkinson A, Dodson G (1996) Crystal structure of T state haemoglobin with oxygen bound at all four haems. J Mol Biol 256:775–792

Park S, Morley KL, Horsman GP, Holmquist M, Hult K, Kazlauskas RJ (2005) Focusing mutations into the *P. fluorescens* esterase binding site increases enantioselectivity more effectively than distant mutations. Chem Biol 12:45–54

Parrent JL, James TY, Vasaitis R, Taylor AF (2009) Friend or foe? Evolutionary history of glycoside hydrolase family 32 genes encoding for sucrolytic activity in fungi and its implications for plant-fungal symbioses. BMC Evol Biol 9:148–164

Patten PA, Gray NS, Yang PL, Marks CB, Wedemayer GJ, Boniface JJ, Stevens RC, Schultz PG (1996) The immunological evolution of catalysis. Science 271:1086–1091

Pereira Y, Petit-Glatron MF, Chambert R (2001) *yveB*, Encoding endolevanase LevB, is part of the *sacB-yveB-yveA* levansucrase tricistronic operon in *Bacillus subtilis*. Microbiology 147:3413–3419

Perrin P, Pierre F, Patry Y, Champ M, Berreur M, Pradal G, Bornet F, Meflah K, Menanteau J (2001) Only fibres promoting a stable butyrate producing colonic ecosystem decrease the rate of aberrant crypt foci in rats. Gut 48:53–61

Perutz MF (1970) Stereochemistry of cooperative effects in haemoglobin. Nature 228:726–739

Perutz MF, Fermi G, Luisi B, Shaanan B, Liddington RC (1987) Stereochemistry of cooperative mechanisms in hemoglobin. Cold Spring Harb Symp Quant Biol 52:555–565

Perutz MF, Wilkinson AJ, Paoli M, Dodson GG (1998) The stereochemical mechanism of the cooperative effects in hemoglobin revisited. Annu Rev Biophys Biomol Struct 27:1–34

Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M (2009) The NIH Human Microbiome Project. Genome Res 19:2317–2323

Plou FJ, Gómez de Segura A, Ballesteros A (2007) Application of Glycosidases and Transglycosidases in the Synthesis of Oligosaccharides. In: Polaina J, MacCabe AP (eds) Ind. Enzym. Springer Netherlands, Dordrecht, pp 141–157

Pons T, Naumoff DG, Martínez-Fleites C, Hernández L (2004) Three acidic residues are at the active site of a beta-propeller architecture in glycoside hydrolase families 32, 43, 62, and 68. Proteins 54:424–432

Pontis HG, del Campillo E (1985) Fructans. In: Dey PM, Dixon RA (eds) Biochem. Storage Carbohydrates Green Plants. Academic Press, London, pp 205–227

Ramasubbu N, Ragunath C, Mishra PJ (2003) Probing the role of a mobile loop in substrate binding and enzyme activity of human salivary amylase. J Mol Biol 325:1061–1076

Ritsema T, Hernández L, Verhaar A, Altenbach D, Boller T, Wiemken A, Smeekens S (2006) Developing fructan-synthesizing capability in a plant invertase via mutations in the sucrose-binding box. Plant J 48:228–237

Ritsema T, Verhaar A, Vijn I, Smeekens S (2005) Using natural variation to investigate the function of individual amino acids in the sucrose-binding box of fructan:fructan 6G-fructosyltransferase (6G-FFT) in product formation. Plant Mol Biol 58:597–607

Roberfroid M, Gibson GR, Hoyles L, McCartney AL, Rastall R, Rowland I, Wolvers D, Watzl B, Szajewska H, Stahl B, Guarner F, Respondek F, Whelan K, Coxam V, Davicco M-J, Léotoing L, Wittrant Y, Delzenne NM, Cani PD, Neyrinck AM, Meheust A (2010) Prebiotic effects: metabolic and health benefits. Br J Nutr 104:S1–S63

Roberfroid MB (2007) Inulin-type fructans: functional food ingredients. J Nutr 137:2493S–2502S

Rumessen JJ, Bodé S, Hamberg O, Gudmand-Høyer E (1990) Fructans of Jerusalem artichokes: intestinal transport, absorption, fermentation, and influence on blood glucose, insulin, and C-peptide responses in healthy subjects. Am J Clin Nutr 52:675–681

Sainz-Polo MA, Ramírez-Escudero M, Lafraya A, González B, Marín-Navarro J, Polaina J, Sanz-Aparicio J (2013a) Three-dimensional structure of *Saccharomyces* invertase: role of a non-catalytic domain in oligomerization and substrate specificity. J Biol Chem 288:9755–9766

Saleh AA, Amber K, El-Magd MA, Atta MS, Mohammed AA, Ragab MM, Abd El-Kader H (2014) Integrative effects of feeding *Aspergillus awamori* and fructooligosaccharide on growth performance and digestibility in broilers: promotion muscle protein metabolism. Biomed Res Int 2014:946859

Sangeetha PT, Ramesh MN, Prapulla SG (2005) Maximization of fructooligosaccharide production by two stage continuous process and its scale up. J Food Eng 68:57–64

Savile CK, Janey JM, Mundorff EC, Moore JC, Tam S, Jarvis WR, Colbeck JC, Krebber A, Fleitz FJ, Brands J, Devine PN, Huisman GW, Hughes GJ (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. Science 329:305–309

Schallmey M, Floor RJ, Hauer B, Breuer M, Jekel P a, Wijma HJ, Dijkstra BW, Janssen DB (2013) Biocatalytic and structural properties of a highly engineered halohydrin dehalogenase. ChemBiochem 14:870–881

Schroeven L, Lammens W, Van Laere A, Van den Ende W (2008) Transforming wheat vacuolar invertase into a high affinity sucrose:sucrose 1-fructosyltransferase. New Phytol 180:822–831

Shanahan F (2002) The host-microbe interface within the gut. Best Pract Res Clin Gastroenterol 16:915–931

Shedlock, M. P. (2014). *Techno-economics of Industrial Scale β-D-fructofuranosidase and Short-chain Fructooligosaccharides Production*. Dissertation, Stellenbosch University.

Shimotohno A, Oue S, Yano T, Kuramitsu S, Kagamiyama H (2001) Demonstration of the importance and usefulness of manipulating non-active-site residues in protein design. J Biochem 129:943–948

Shiomi N, Izawa M (1980) Purification and Characterization of Sucrose: sucrose l-fructosyltransferase from the Roots of Asparagus (*Asparagus officinalis* L.). Agric Biol Chem 44:603–614

Soleimani N, Hoseinifar SH, Merrifield DL, Barati M, Abadi ZH (2012) Dietary supplementation of fructooligosaccharide (FOS) improves the innate immune response, stress resistance, digestive enzyme activities and growth performance of Caspian roach (*Rutilus rutilus*) fry. Fish Shellfish Immunol 32:316–321

Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370:389–391

Stone-Dorshow T, Levitt MD (1987) Gaseous response to ingestion of a poorly absorbed fructo-oligosaccharide sweetener. Am J Clin Nutr 46:61–65

Strausberg SL, Ruan B, Fisher KE, Alexander PA, Bryan PN (2005) Directed coevolution of stability and catalytic activity in calcium-free subtilisin. Biochemistry 44:3272–3279

Süel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol 10:59–69

Sukharnikov LO, Alahuhta M, Brunecky R, Upadhyay A, Himmel ME, Lunin V V, Zhulin IB (2012) Sequence, structure, and evolution of cellulases in glycoside hydrolase family 48. J Biol Chem 287:41068–41077

Sukharnikov LO, Cantwell BJ, Podar M, Zhulin IB (2011) Cellulases: ambiguous nonhomologous enzymes in a genomic perspective. Trends Biotechnol 29:473–479

Suzuki M, Christians FC, Kim B, Skandalis A, Black ME, Loeb LA (1996) Tolerance of different proteins for amino acid diversity. Mol Divers 2:111–118

Suzuki M, Pollock CJ (1986) Extraction and characterization of the enzymes of fructan biosynthesis in timothy (*Phleum pratense*). Can J Bot 64:1884–1887

Taiz L, Zeiger E (2010) Topic 8.14: Fructans. In: Plant Physiol. Fifth Ed. Online. http://www.plantphys.net. Accessed 21 Sep 2014

Takesue Y (1969) Purification and properties of rabbit intestinal sucrase. J Biochem 65:545–552

Tao H, Cornish VW (2002) Milestones in directed enzyme evolution. Curr Opin Chem Biol 6:858–864

Taussig R, Carlson M (1983) Nucleotide sequence of the yeast *SUC2* gene for invertase. Nucleic Acids Res 11:1943–1954

Tauzin AS, Giardina T (2014) Sucrose and invertases, a part of the plant defense response to the biotic stresses. Front Plant Sci 5:293–301

Tindall KR, Kunkel TA (1988) Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. Biochemistry 27:6008–6013

Van den Ende W (2013) Multifunctional fructans and raffinose family oligosaccharides. Front Plant Sci 4:1–11

Van den Ende W, Clerens S, Vergauwen R, Boogaerts D, Le Roy K, Arckens L, Van Laere A (2006) Cloning and functional analysis of a high DP fructan:fructan 1-fructosyl transferase from *Echinops ritro* (Asteraceae): comparison of the native and recombinant enzymes. J Exp Bot 57:775–789

Van den Ende W, Coopman M, Clerens S, Vergauwen R, Le Roy K, Lammens W, Van Laere A (2011) Unexpected presence of graminan- and levan-type fructans in the evergreen frost-hardy eudicot *Pachysandra terminalis* (Buxaceae): purification, cloning, and functional analysis of a 6-SST/6-SFT enzyme. Plant Physiol 155:603–614

Van Laere A, Van Den Ende W (2002) Inulin metabolism in dicots: chicory as a model system. Plant Cell Environ 25:803–813

Velázquez-Hernández ML, Baizabal-Aguirre VM, Bravo-Patiño a, Cajero-Juárez M, Chávez-Moctezuma MP, Valdez-Alarcón JJ (2009) Microbial fructosyltransferases and the role of fructans. J Appl Microbiol 106:1763–1778

Vergauwen R, Van den Ende W, Van Laere A (2000) The role of fructan in flowering of *Campanula rapunculoides*. J Exp Bot 51 :1261–1266

Vijn I, Smeekens S (1999) Fructan: more than a reserve carbohydrate? Plant Physiol 120:351–359

Voegele RT, Wirsel S, Möll U, Lechner M, Mendgen K (2006) Cloning and characterization of a novel invertase from the obligate biotroph *Uromyces fabae* and analysis of expression patterns of host and pathogen invertases in the course of infection. Mol Plant Microbe Interact 19:625–634

Wallace S, Balskus EP (2014) Opportunities for merging chemical and biological synthesis. Curr Opin Biotechnol 30C:1–8

Walter J, Schwab C, Loach DM, Gänzle MG, Tannock GW (2008) Glucosyltransferase A (GtfA) and inulosucrase (Inu) of *Lactobacillus reuteri* TMW1.106 contribute to cell aggregation, in vitro biofilm formation, and colonization of the mouse gastrointestinal tract. Microbiology 154:72–80

Webb EC (1992) Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes. Biochem Educ 21:23–130

Wen ZT, Burne RA (2002) Analysis of cis- and trans-Acting Factors Involved in Regulation of the *Streptococcus mutans* Fructanase Gene (*fruA*). J Bacteriol 184 :126–133

White LM, Secor GE (1952) The oligosaccharides formed during the sucrose-invertase reaction. Arch Biochem Biophys 36:490–491

Wiemken A, Sprenger N, Boller T (1995) Fructan: an extension of sucrose by sucrose. In: Pontis H, Salerno G, Echeverria E (eds) Sucrose Metab. Biochem. Physiol. Mol. Biol. American Society of Plant Physiologists, Rockville, pp 178–189

Williams DC, Benjamin DC, Poljak RJ, Rule GS (1996) Global changes in amide hydrogen exchange rates for a protein antigen in complex with three different antibodies. J Mol Biol 257:866–876

Yanai K, Nakane A, Kawate A, Hirayama M (2001) Molecular cloning and characterization of the fructooligosaccharide-producing β-fructofuranosidase gene from *Aspergillus niger* ATCC 20611. Biosci Biotechnol Biochem 65:766–773

Yazawa K, Imai K, Tamura Z (1978) Oligosaccharides and polysaccharides specifically utilizable by bifidobacteria. Chem Pharm Bull (Tokyo) 26:3306–3311

Yedavalli P, Rao NM (2013) Engineering the loops in a lipase for stability in DMSO. Protein Eng Des Sel 26:317–324

Yuan X-L, Goosen C, Kools H, van der Maarel MJEC, van den Hondel C a MJJ, Dijkhuizen L, Ram AFJ (2006) Database mining and transcriptional analysis of genes encoding inulin-modifying enzymes of *Aspergillus niger*. Microbiology 152:3061–3073

Yun JW, Jung KH, Jeon YJ, Lee JH (1992) Continuous production of fructo-oligosaccharides by immobilized cells of *Aureobasidium pullulans*. J Microbiol Biotechnol 2:98–101

Yun JW (1996) Fructooligosaccharides—Occurrence, preparation, and application. Enzyme Microb Technol 19:107–117

Yun JW, Song SK (1999) Enzymatic Production of Fructooligosaccharides from Sucrose. In: Bucke C (ed) Carbohydr. Biotechnol. Protoc. Humana Press, Totowa, N.J., pp 141–151

Zeng L, Wen ZT, Burne RA (2006) A novel signal transduction system and feedback loop regulate fructan hydrolase gene expression in *Streptococcus mutans*. Mol Microbiol 62:187–200

CHAPTER 3

RESEARCH RESULTS I

Screening a random mutagenesis library of a fungal β-fructofuranosidase using FT-MIR ATR spectroscopy and multivariate analysis.

Published in Appl. Microbiol. Biotechnol. (2014) ***98****:4063–73*

3 RESEARCH RESULTS I:

Screening a random mutagenesis library of a fungal β-fructofuranosidase using FT-MIR ATR spectroscopy and multivariate analysis

K.M. Trollope[1], H.H. Nieuwoudt[2], J.F. Görgens[3], H. Volschenk[1]

[1] Department of Microbiology, Stellenbosch University, Private Bag X1, Stellenbosch, 7602, South Africa
[2] Institute for Wine Biotechnology, Department of Viticulture and Oenology, Stellenbosch University, Private Bag X1, Stellenbosch, 7602, South Africa
[3] Department of Process Engineering, Private Bag X1, Stellenbosch University, Stellenbosch, 7602, South Africa

ABSTRACT

Short chain fructooligosaccharides (scFOS) are valuable health promoting food additives. During the batch production of scFOS from sucrose the β-fructofuranosidase catalyst is subject to product inhibition by glucose. Engineering the enzyme for reduced sensitivity to glucose could improve product yields or process productivity while preserving the simple industrial batch design. Random mutagenesis is a useful technique for engineering proteins but should be coupled to a relevant high-throughput screen. Such a screen for sucrose and scFOS quantification remains elusive. This work presents the development of a screening method displaying potential high-throughput capacity for the evaluation of β-fructofuranosidase libraries using Fourier transform mid infrared attenuated total reflectance (FT-MIR ATR) spectroscopy and multivariate analysis. A calibration model for the quantification of sucrose in enzyme assay samples ranged from 5 to 200 g/l and the standard error of prediction was below 13 g/l. A library of the *Aspergillus japonicus fopA* gene was generated by error prone PCR and screened in *Saccharomyces cerevisiae*. Using FT-MIR ATR spectroscopy, potential hits were identified as those variants that converted more sucrose in the presence of the glucose inhibitor than the parent. Subsequent analysis of reaction products generated by top performers using high performance liquid chromatography identified a variant producing higher scFOS levels than the parent. At the peak difference in performance the variant produced 28% more scFOS from the same amount of sucrose. This study highlights the application of FT-MIR ATR spectroscopy to a variant discovery pipeline in the directed evolution of a β-fructofuranosidase for enhanced scFOS production.

3.1 INTRODUCTION

β-fructofuranosidases, also known as invertases (EC 3.2.1.26), are enzymes responsible for the synthesis of short chain fructooligosaccharides (scFOS) from sucrose. ScFOS serve as sweeteners in the food industry and have attracted much attention due to the human health-promoting effects associated with their prebiotic, bifidogenic properties. Extensive reviews are available on this topic (Gibson and Roberfroid 1995; Roberfroid and Slavin 2000; Roberfroid 2007). The β-fructofuranosidase from *Aspergillus niger* ATCC 20611 is used for the commercial production of linear, β(2→1) linked inulin-type scFOS from sucrose (Hidaka et al. 1988). Based on their amino acid sequence similarities, β-fructofuranosidases are classified in Glycoside Hydrolase (GH) families 32 and 68, together with enzymes possessing activities on fructose-containing substrates, including but not limited to exo-inulinase (EC 3.2.1.80), endo-inulinase (EC 3.2.1.7), sucrose:fructan 6-fructosyltransferase (EC 2.4.1.10), levansucrase (EC 2.4.1.10) and inulosucrase (EC 2.4.1.9) (Cantarel et al. 2009). GH 32 and 68 members are further classified into clan GH-J with all enzymes displaying a 5-bladed β-propeller catalytic core module. Substrate hydrolysis occurs via a retaining, double-displacement mechanism, whereby water serves as acceptor molecule for a fructosyl unit constituting an enzyme-fructosyl intermediate. Although classified as hydrolases, β-fructofuranosidases are able to perform transferase/transfructosylation reactions whereby a sugar molecule serves as acceptor molecule for a fructosyl group rather than water. In so doing, this transferase activity results in the synthesis of 1-kestose (GF2), nystose (GF3) and 1^F-β-frucofuranosyl nystose (GF4), depending on the acceptor sugar molecule (Edelman and Bacon 1951; Fischer et al. 1951; White and Secor 1952). Manipulation of reaction conditions allows transferase activity to be favoured over hydrolase activity, but the reaction mixture at any given point is a reflection of the combined activities. Glucose is produced as a product of both activities and accumulation of glucose has been shown to significantly inhibit enzyme activity (Nishizawa et al. 2001).

Improvement of protein characteristics by directed evolution is a powerful and commonly employed technique (Cobb et al. 2012; Dougherty and Arnold 2009; Ruff et al. 2013). Many variations on the technique are found but essentially they all involve altering the sequence of the coding gene, expression of altered genes and screening the gene products for altered performance in an iterative manner. Error prone PCR (epPCR) is one of the simplest ways to mutate a gene sequence and does not require mechanistic knowledge about the target protein to implement (Kuchner and Arnold 1997). Furthermore, the randomness of mutations generated using this technique allows for the modification of regions in proteins related to the trait of interest that may not have been rationally identified, thereby contributing to structure-function understanding of a protein (Shimotohno et al. 2001). The proviso associated with such a generalised approach to protein engineering is the ability to rapidly screen the large libraries generated for desired traits (Romero and Arnold 2009).

A study by de Abreu et al. (2011) reports on enhancing transglycosylation rates of a β-fructofuranosidase using random mutagenesis. They developed a high-throughput colorimetric library screening procedure based on the separate enzymatic quantification of glucose and fructose. Another report on engineering *Saccharomyces cerevisiae* invertase used site-directed mutagenesis (Lafraya et al. 2011), which did not require high-throughput screening to evaluate variants due to the relatively small numbers of variants generated. High performance liquid chromatography (HPLC) is routinely used to quantify reaction products, including oligosaccharides.

Fourier transform infrared (FT-IR) spectroscopy is used for routine analyses of wines (Kupina and Shrikhande 2003; Patz et al. 1999) and in the food industry for quality determination as well as authentication of raw ingredients (Duarte et al. 2002; Li-Chan et al. 2010; Wang et al. 2010). It is also used in specialised applications such as the tracking of fermentation by yeasts (Cozzolino and Curtin 2012; Nieuwoudt et al. 2006; Schenk et al. 2007) and evaluation of pre-treated biomass for bioethanol production (Sills and Gossett 2012). Furthermore, a FT-IR spectroscopic method was developed to evaluate enantiomeric purity of synthetic organic compounds and has potential application to screening libraries of asymmetric catalysts (Tielmann et al. 2003). To date this methodology has not yet been applied to actual library screening. The capacity of FT-MIR ATR spectroscopy to analyse multiple compounds simultaneously and the speed of analysis makes this an attractive analytical technique, especially for screening large numbers of samples. However, the generation of multivariate calibration models is often complex and time consuming, which limits widespread application of this technique.

In this study we aimed to develop and demonstrate a high-throughput screen based on FT-MIR ATR spectroscopy and its application to screening a β-fructofuranosidase variant library generated by random mutagenesis. Using FT-MIR ATR spectroscopy, variants were screened for sucrose conversion in the presence and absence of glucose in order to identify variants with reduced sensitivity to glucose inhibition. The improved functionality of selected mutants was subsequently confirmed by HPLC.

3.2 MATERIALS AND METHODS

3.2.1 Microbial strains

S. cerevisiae EUROSCARF Y02321 [BY4741; *Mat a; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YIL162w(SUC2)::kanMX4*] was used as host for the random library (Brachmann et al. 1998). *Escherichia coli* DH5α [*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*] (New England Biolabs) was used for cloning and amplification of

plasmids. *E. coli* cells were grown at 37°C in Luria Bertani broth supplemented with 100 μg/ml ampicillin or 50 μg/ml kanamycin, as appropriate.

3.2.2 DNA manipulations, plasmids and random mutagenesis

All DNA manipulations were performed according to standard methods (Sambrook et al. 1989). Restriction enzymes and T4 DNA ligase were used according the specifications of the supplier (ThermoScientific, Waltham, Massachusetts, USA). The *Aspergillus japonicus* [previously classified as *A. niger* by Yanai et al. (2001) but reclassified by the curators of the ATCC culture collection (http://www.lgcpromochem-atcc.com)] β-fructofuranosidase *fopA* gene (GenBank accession number AB046383) was codon optimised and synthesised by GeneArt (Regensburg, Germany). The native secretion signal was replaced by the *Trichoderma reesei* endoxylanase 2 (*xln2*) secretion signal (Saarelainen et al. 1993). The GenBank accession number assigned to this synthetic construct was KF536718. The construct was subcloned as a 2033-bp *Eco*RI-*Xho*I fragment from the pMK-RQfopA cloning vector (GeneArt, Regensburg, Germany) to the *Eco*RI-*Xho*I digested pJC1 *E. coli – S. cerevisiae* shuttle vector (Crous et al. 1995). The final construct comprised *fopA* under the control of the *S. cerevisiae* phosphoglycerate kinase (*PGK1*) promoter and terminator on a multicopy, episomal plasmid. For error prone PCR (epPCR) the pMK-RQfopA plasmid served as template. Random mutagenesis across the full open reading frame of *fopA* was performed using Mutazyme II DNA polymerase from the GeneMorph II Random Mutagenesis Kit (Agilent Technologies Inc., Santa Clara, California, USA). The primer pair (5' CGATGGTTTCTTTCACATCCTTGTTGGCTG 3' and 5' CGTTAGTTTCTCTCTGGCCAAGCGTTG 3') was used for epPCR. High and low mutation frequencies were obtained by using 50 pg and 100 ng of target DNA, respectively in combination with 20 thermal cycles. Mutagenised DNA was amplified using *Taq* polymerase (Thermoscientific, Waltham, Massachusetts, USA) in order to improve DNA yields following gel purification.

3.2.3 Yeast transformation

Yeast transformation was performed using a standard lithium acetate method (Hill et al. 1991). Recombination-mediated plasmid construction in yeast was achieved by co-transforming epPCR products and *Bst*EII linearised pJCfopA (Oldenburg et al. 1997). Following library screening, total DNA was extracted from candidate yeasts and transformed to *E. coli* by means of electroporation. Positive and negative control strains were obtained by transforming yeast with pJC1fopA and pJC1 plasmids, respectively.

3.2.4 Cultivation procedures

S. cerevisiae was cultivated at 30°C in YPD (1% yeast extract, 2% peptone and 2% glucose) and in synthetic medium, SC without uracil [2% carbon source, 0.67% yeast nitrogen base without amino acids (with ammonium sulphate) (Difco Laboratories, Detroit, Michigan, USA) and 0.13% amino acid dropout pool (Sherman et al. 1979)]. Glucose served as carbon source in solid media and galactose was used in liquid SC^{-ura} media. Solid media contained 2% agar (Difco Laboratories).

Yeast transformants were manually transferred from agar plates to 2 ml, round bottomed 96 deep well plates (Merck) containing 1.25 ml SC^{-ura} media. Each well contained a single 2mm glass bead (Merck) to facilitate mixing. Plates were sealed with sterile, breathable AeraSeal™ film (Excel Scientific Inc., Victorville, California, USA) and shaken at 200 rpm for 4 days. Fifty microliters of this culture were transferred to a new plate and cultivated for a further 4 days. Master plates were generated using a 96 well replicator (Applikon Biotechnology, Netherlands). Cells were removed via centrifugation at 3000 rpm and supernatants were used in assays as a source of enzyme.

For improved cultivation conditions in terms of mixing and aeration, yeast were inoculated to an optical density (OD_{600}) of 0.1 in 24 well plates (Greiner Bio-One, Kremsmuenster, Austria) containing 1 ml SC^{-ura} medium and shaken at 200 rpm for 48 h. Three replicates for each variant were included.

3.2.5 Enzyme assays

3.2.5.1 96 well format

Enzyme (50 µl culture supernatant) was added to sucrose (Fluka, Sigma-Aldrich, St. Louis, Missouri, USA) substrate in a 1:1 ratio to produce a working concentration of 200 g/l sucrose in 50 mM citrate phosphate buffer at pH5.5. For inhibition studies glucose was included at a working concentration of 54 g/l. Assays were performed in 96 well PCR microplates and samples were incubated at 55°C for 2 h. The reactions were stopped by adding perchloric acid to 1.72% (v/v) on ice. Samples were neutralised by adding KOH to 0.26 N. Inhibition was calculated as the difference between variant activity without and with inhibitor (glucose) divided by the activity without inhibitor. Data were normalised relative to the parental activity. The parent was cultivated and assayed with triplicate repeats while no replicates for variants were included.

3.2.5.2 Shake flask format

One unit of enzyme was defined as the amount of enzyme required to produce 1 μmol glucose per minute when reacted with 100 g/l sucrose in 50 mM citrate phosphate buffer, pH5.5 at 40°C for 1 h (Hidaka et al. 1988).

Assays were performed in a 40 ml working volume with 600 g/l sucrose in 50 mM citrate phosphate buffer, pH5.5 at 62°C for 7 h. Enzyme was dosed at 7 U per g sucrose and flasks were shaken at 120 rpm. The selected assay conditions were previously optimised in our laboratory (unpublished data).

3.2.6 FT-MIR ATR spectroscopy

3.2.6.1 Samples and reference data

Data used to develop partial least squares (PLS) regression calibration models were generated in two ways. Firstly, sugar reference data for enzyme assay samples generated in 96 well format (see above) were obtained using HPLC (methodology described below). Secondly, spiked samples were generated as follows. Yeast transformed with the control pJC1 plasmid was grown in 50 ml SC^{-ura} medium for 3 days. This supernatant was diluted with citrate phosphate buffer and spiked with appropriate concentrations of sugars (glucose, fructose, sucrose, 1-kestose and nystose) in varying combinations to span concentrations relevant to the required calibration ranges. Blank samples containing no sugars were generated by combining yeast supernatants and citrate phosphate buffer. Blank samples were used to assess the contribution to the FT-MIR ATR spectra of yeast metabolites in the assay samples. In total 192 samples were scanned.

3.2.6.2 FT-MIR ATR instrument

Samples were scanned using a Bruker Alpha FT-IR spectrometer fitted with a heating Alpha-P sampling module with single reflection diamond attenuated total reflectance (Bruker Optics GmbH, Ettlingen, Germany). Prior to scanning a reference background measurement was taken using air. The FT-MIR ATR spectra were recorded on OPUS software version 6.5 (Bruker Optics GmbH, Ettlingen, Germany). The absorbance spectrum of each sample was obtained at 26°C by taking the average of 64 scans at a resolution of 4 cm^{-1} and acquired between 4000 and 550 cm^{-1}, with a scanner velocity of 7.5 KHz.

3.2.6.3 Data analysis

Spectra were exported from OPUS software in data point format (*.dpt) and imported into Solo standalone toolbox version 7.0.3 (Eigenvector Research, Inc., Wenatchee, Washington, USA). Multivariate analysis was performed by principal component analysis (PCA) and partial least squares regression. The FT-MIR ATR absorbance data served as the X matrix while sucrose concentrations were used as the Y matrix. PCA of the X matrix of sugar-containing and blank samples was used to explore the structure of the data. Loadings from PCA and the known fingerprint region for sugars were used to guide variable selection. PLS regression was used to find the correlation between the X and Y matrices and generate a calibration model for the quantification of sucrose (n = 134). Preprocessing for the X block used mean centring while for the Y block, data were autoscaled to unit variance. No further spectral preprocessing was applied. Cross validation was performed using a venetian blinds design with 15 data splits. After exclusion of gross outliers the models were evaluated using independent test sets (n = 58). All samples were randomly split into calibration and test sets using the automated onion function in Solo such that both sets spanned the entire concentration range of reference values. The validation was repeated four times with 70:30% data splits and once with a 50:50% split. When root mean square error of prediction (RMSEP) values for the test sets did not deviate excessively from the root mean square error of cross validation (RMSECV) values, models were deemed to be acceptable.

3.2.7 96 well plate assay validation

To quantify the experimental error introduced by cultivation and assay procedures in well plate format, assay data were examined where only strains expressing parent enzyme were inoculated across the entire plate. Assays were performed with 200 g/l sucrose substrate for 2 h at 55°C. Sucrose was quantified using FT-MIR ATR spectroscopy. Descriptive statistics were generated for the data as well as a coefficient of variation (ratio of standard deviation to the sample mean expressed as a percentage). Spatial uniformity across the well plate was assessed from row or column data analysed separately.

3.2.8 HPLC: High performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD)

3.2.8.1 Materials

HPLC-grade standards for L-arabinose, D-glucose, D-fructose and D-sucrose were purchased from Sigma-Aldrich (Schnelldorf, Germany). L-arabinose served as an internal standard. The fructooligosaccharide set (1-kestose, 1-nystose and 1^F-β-fructofuranosylnystose) was

purchased from Wako Chemicals GmbH (Neuss, Germany). Mobile phases were made using sodium acetate (71180, Sigma-Aldrich) and 50% NaOH solution (71686, Fluka) according to Dionex Technical Note 20. Ultra-pure 18 MΩ deionized water generated by a Milli-Q UF Plus system (Millipore, Molsheim, France) was used to make all solutions and for dilution of samples.

3.2.8.2 Chromatography and analysis

Appropriately diluted samples (10 µl) were analysed using a CarboPac PA1 (4 x 250 mm) analytical column coupled to a PA1 (4 x 50 mm) guard column (Thermo Scientific Dionex, Waltham, Massachusetts, USA) on a Dionex Ultimate 3000 system. The system consisted of a WPS-3000T SL (analytical) autosampler, LPG-3400 AB pump, Chromeleon version 6.80 software and a Coulochem III electrochemical detector (Esa, Inc., Chelmsford, MA, USA) with working gold electrode. The pulsed amperometric detection settings were for a quadruple-potential waveform (Rocklin et al. 1998) with a minor modification of the E3 pulse set at 600 mV for 10 ms. Gradient elution of sugars was performed using the program described in Table 3.1. The different sugars in samples were identified and quantified by comparing the retention time and ratio of sugar peak area to internal standard peak area to those of external standards.

Table 3.1 The gradient elution programme that was applied to separate the sugars in assay samples for quantification using high performance liquid chromatography

Time (min)	Flow (ml/min)	Solvent (% by volume)		
		A	**B**	**C**
0	1.0	10	90	0
12	1.0	58	42	0
16	1.0	70	10	20
23	1.0	70	10	20
23.1	1.0	10	90	0
23.2	1.5	10	90	0
34.2	1.5	10	90	0
34.3	1.0	10	90	0
39	10	10	90	0

Solvent A: 250 mM NaOH, B: water and C: 100 mM NaOH with 500 mM sodium acetate

3.3 RESULTS

3.3.1 Quantification of sucrose using FT-MIR ATR spectroscopy

Although FT-MIR ATR spectra spanned wavenumbers from 3996 to 559 cm^{-1}, it was evident upon inspection of multiple overlaid spectra that the region showing the highest variation in absorbance was below 1500 cm^{-1} (Fig. 3.1). This region, known as the fingerprint region (Leopold et al. 2011), also contains the IR spectral annotations of aqueous sugar hydrates by Max and Chapados (2007; 2001).

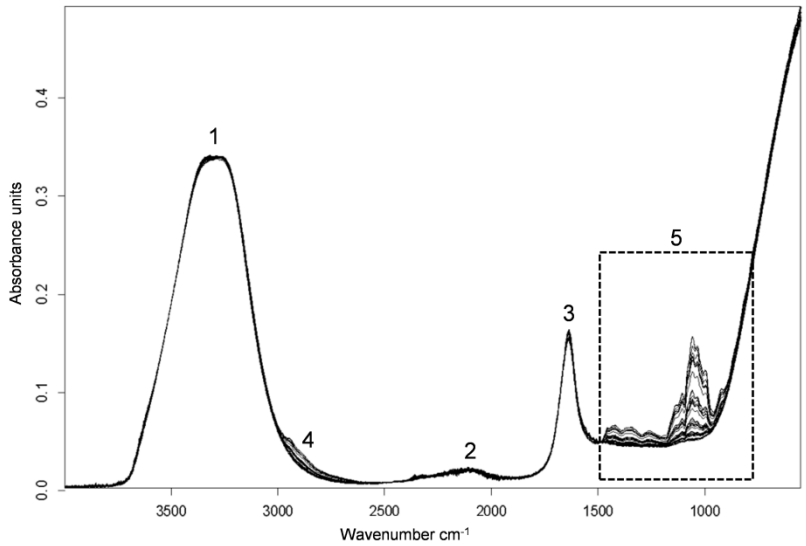

Fig. 3.1 Raw FT-MIR ATR spectra of samples. Peak annotations correspond to absorptions of: water at 3,600–3000 cm^{-1} (peak 1); water at ~2,100 cm^{-1} and nitrile and alkyne triple bonds at 2,260–2,210 cm^{-1} (peak 2); water at 1,645 cm^{-1} and protein at 1,650 cm^{-1} and 1,550 cm^{-1} (peak 3); carbohydrates at 3,000–2,700 cm^{-1} and 1,500–800 cm^{-1} (peaks 4 and 5) (Barth, 2007; Max and Chapados, 2001 and 2007). The fingerprint region of the sugars, highlighted by the *block*, was selected for sucrose model building

PCA of assay and blank (matrix without sugars) samples showed a clear grouping of the two types of samples on the scores plot of principal components (PC) 1 and 2. The loadings plot confirmed the importance of the wavenumbers below 1500 cm^{-1} for the grouping (Fig. 3.2). To reduce the complexity of modelling, wavenumbers ranging from 1600 to 800 cm^{-1} were selected *a priori* to develop a PLS calibration model for sucrose. Although it was apparent that wavenumbers below 1500 cm^{-1} were important for sugars, the *a priori* selection included wavenumbers from 1600 cm^{-1} to allow the mathematical modelling process to assign final importance to relevant wavenumbers.

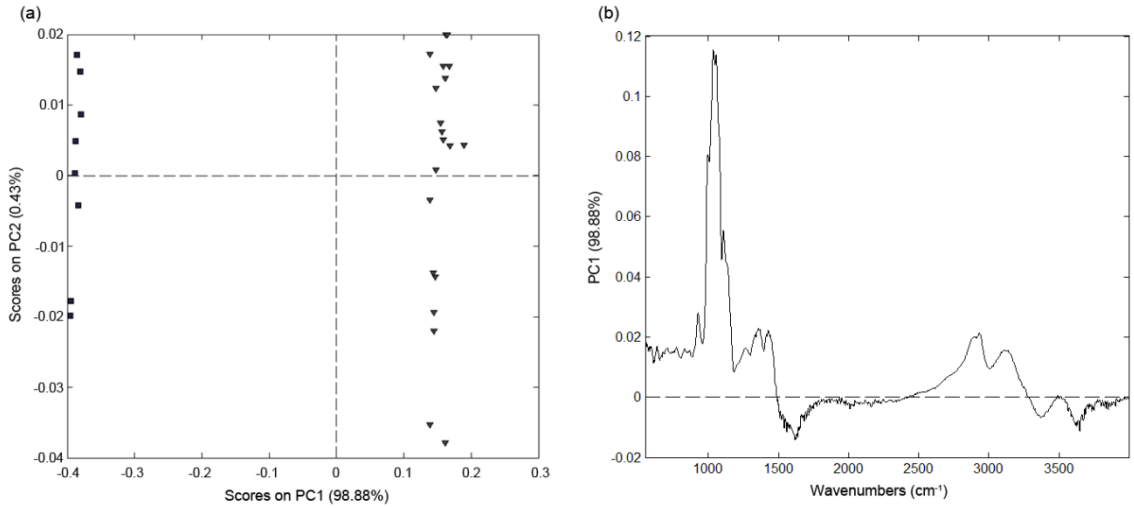

Fig. 3.2 Scores (a) and loadings (b) plots for principal component analysis of blank (*squares*) and sugar-containing (*inverted triangle*) samples. The scores plot shows the clear separation of the blank and sugar-containing samples. In the loadings plot it is evident that the wavenumbers ranging from approximately 900 to 1500 cm^{-1} are important for the grouping of the two kinds of samples, hence relevant to the sugars and the quantification thereof

The final model was validated using an independent test set and the regression plot is shown in Fig. 3.3. The regression statistics for calibration and validation samples are summarised in Table S1 and Table S2 (Supplementary Material). The coefficient of determination (R^2) was above 0.95 for five random combinations of calibration and independent test sets. The standard error of prediction ranged from 12.06 – 15.07 g/l sucrose, which was considered acceptable across a calibration ranging from approximately 5 to 200 g/l sucrose. The residual predictive deviation (RPD), which serves as an indicator of the model's predictive ability, was 4.07 which is above the value of 3 recommended for screening purposes (Williams 2001).

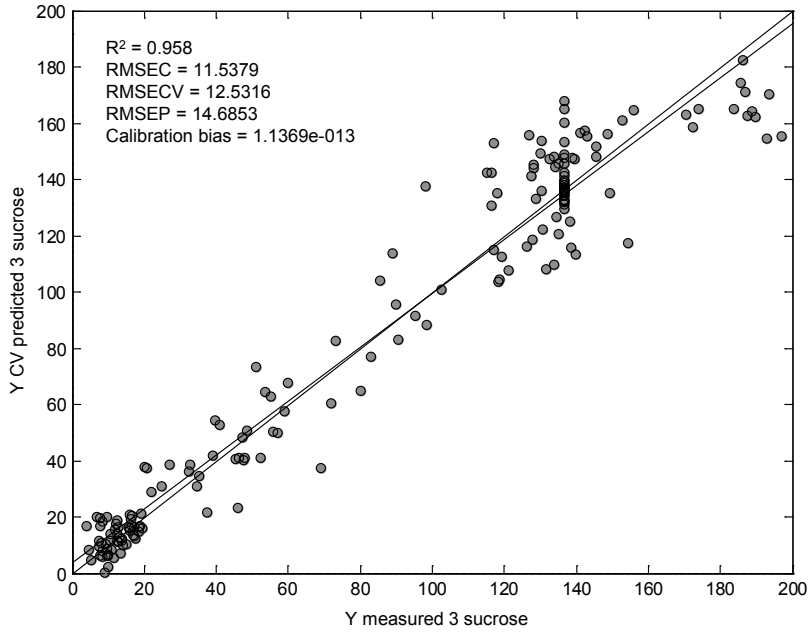

Fig. 3.3 Partial least squares regression model for the prediction of sucrose in assay samples. The coefficient of determination was 0.958, root mean square error of calibration 11.53 g/l, root mean square error of cross validation 12.53 g/l and the root mean square error of prediction 14.69 g/l

3.3.2 Library screening

Yeast supernatants, as source of variant FopA enzymes, were reacted with sucrose substrate and selected sugars (depending on the level of progress in the workflow) were quantified and used as indicators of variant performance. Details of the workflow are shown in Fig. 3.4. The enzyme variant library was screened for relief from glucose (product) inhibition and to this end assays were performed in the presence and absence of added glucose. This methodology was described for engineering a fungal β-galactosidase (Hu et al. 2010). The most severe inhibition conditions outlined in the study by Nishizawa et al. (2001) were used as departure point for testing glucose addition in our high-throughput experimental scenario in order to determine assay conditions under which to screen the library. The conditions selected were 200 g/l sucrose supplemented with 54 g/l glucose. The production of scFOS after 1 h was inhibited by 26% relative to conditions where no initial glucose was added (data not shown).

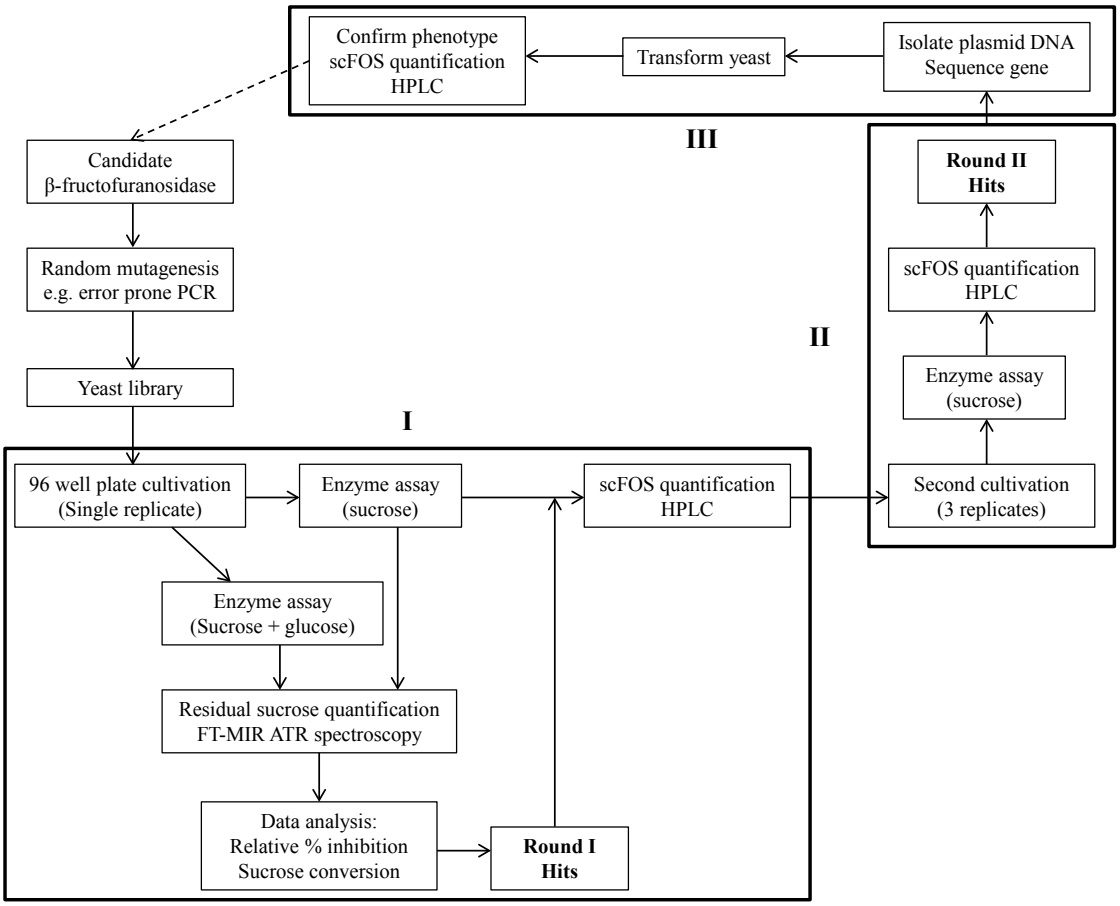

Fig. 3.4 A workflow for screening the β-fructofuranosidase library. Screen I was dependent on FT-MIR ATR spectroscopy for sucrose quantification, while screens II and III used HPLC for quantification of all sugars. Successive rounds of directed evolution can be applied by feeding a variant from screen III back into the pipeline

3.3.3 Validation of high-throughput screen

A summary of the validation results are given in Table 3.2. The coefficient of variation of the screen was calculated to be 13%. Systematic bias introduced by the cultivation and assay procedures did not appear. Using the data from the descriptive statistics (mean, range, standard deviation, standard error), it was calculated that for a 95% confidence level a cut off of the mean of parental replicates plus two standard errors would suffice to identify potential hits for further screening.

Table 3.2 A summary of the validation of the high-throughput screen. It reflects experimental error introduced by cultivation, assay and quantification procedures

Plate position		Average Residual sucrose concentration (g/l)	Standard deviation	Coefficient of variation
Column	1	34.2	2.4	7%
	2	40.3	6.7	17%
	4	41.3	4.2	10%
	6	41.3	0.05	0%
	8	35.3	3.2	9%
	10	38.8	10.8	28%
	12	33.8	0.8	2%
Column Average		37.8	4.0	10%
Row	A	35.1	4.0	11%
	B	39.1	4.4	11%
	D	35.5	3.4	10%
	F	40.0	8.7	22%
	H	41.0	5.5	14%
Row Average		38.1	5.2	13%
Plate Average		**37.2**	**4.7**	**13%**

3.3.4 Evaluating the random library: Screen I of all variants (Potential high-throughput stage using FT-MIR ATR spectroscopy)

Screen I served as a crude means to initially reduce the large number of variants to a manageable sample of improved variants that could be further scrutinised with the appropriate statistical and analytical techniques. Residual sucrose concentration, as determined by FT-MIR ATR spectroscopy, was used as an indicator of variant performance in the first round of library screening. A rapid assay for directly detecting transferase activity (scFOS) reaction products has not been reported; nor is there a direct colorimetric or fluorometric detection method for sucrose that does not involve prior hydrolysis steps. We reasoned that the lower the residual sucrose at the end of the enzyme reaction, the more active the variant represented by the particular supernatant. Samples with residual sucrose concentrations similar to the negative control strain (vector without insert) were deemed inactive. Of the 242 yeast supernatants evaluated, 109 retained activity. The active variants were further screened in the presence of glucose. Fig. 3.5 shows the results of the initial screen using the FT-MIR ATR spectroscopy data. A broad spectrum of activities was obtained as expected for a mutagenesis experiment. Numerous variants displayed either improved sucrose consumption or relative inhibition but only 23% displayed improvements in both criteria. Variants that consumed more sucrose in the presence of glucose displayed an improved relative inhibition (higher than one) as the difference in sucrose consumption under the two conditions became smaller. Improved variants are highlighted by the block in Fig. 3.5 and were selected for further screening with HPLC. To verify that hits from the sucrose conversion screen still retained their scFOS producing ability,

both 1-kestose and nystose were quantified using HPLC. Variants that produced more scFOS than the parent enzyme, and specifically nystose were further evaluated (data not shown). Higher nystose production is indicative of an enzyme that is turning over substrate more rapidly.

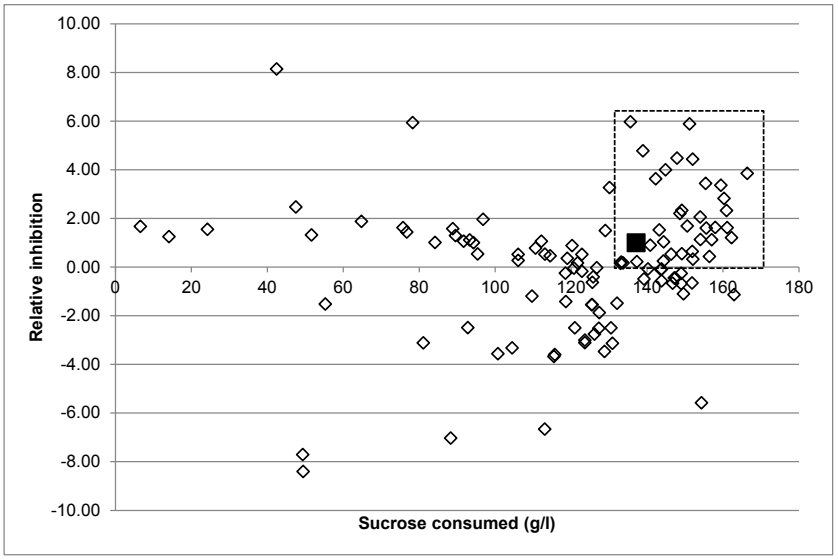

Fig. 3.5 Results of the initial screening of the random library using FT- MIR ATR spectroscopy. Data are shown for the 109 active variants. Variants, highlighted by the dashed rectangle, that displayed better sucrose consumption and relative inhibition than the parent (filled square), were further screened. Sucrose consumption was quantified under conditions with and without added glucose inhibitor. Relative inhibition was derived from the difference in sucrose consumption between the two conditions expressed as a percentage of uninhibited activity. Data were normalised to the parent

3.3.5 Evaluating the random library: Screen II of potentially improved variants

Eleven variants from screen I were selected for further scrutiny in the second screen where improved cultivation conditions with 3 repeats for all variants were included as well as scFOS quantification using HPLC. This sample represented 4.5% of the library, while assays did not include glucose supplementation. Screen I served to identify improved variants in terms of sucrose conversion which is however, not necessarily synonymous with improved scFOS production. Selected results of the HPAEC-PAD screen are shown in Fig. 3.6. Variants P127A5 and P127A8 were of particular interest as they produced 9.1% and 15.1% more nystose than the parent enzyme, respectively. Variant P127A8 showed the most promise, as the total scFOS production levels were significantly improved by 4.3% (p = 0.04), under the selected laboratory conditions. Sequencing of the coding gene revealed that the enzyme harboured a single A458V amino acid substitution [numbering as for Protein Data Bank entry 3LFI (Chuankhayan et al. 2010)]. The DNA sequence GenBank accession number for this variant is KF536719. Although P127A5 also produced more nystose than the parent enzyme, it used 1.7% more sucrose, which would translate into a lower yield of scFOS from sucrose. Of note were variants P8A2 and P8B12 that produced equivalent amount of total scFOS to the parent, yet utilised 4.8% and 4.4% less sucrose. P8A12 was noteworthy in the sense that it

was a 1-kestose accumulator. It produced only 5.8% less 1-kestose than the parent (52 vs 55 g/l) but produced 65% less nystose (8 vs 24 g/l). Furthermore, it utilised 11% less sucrose than the parent.

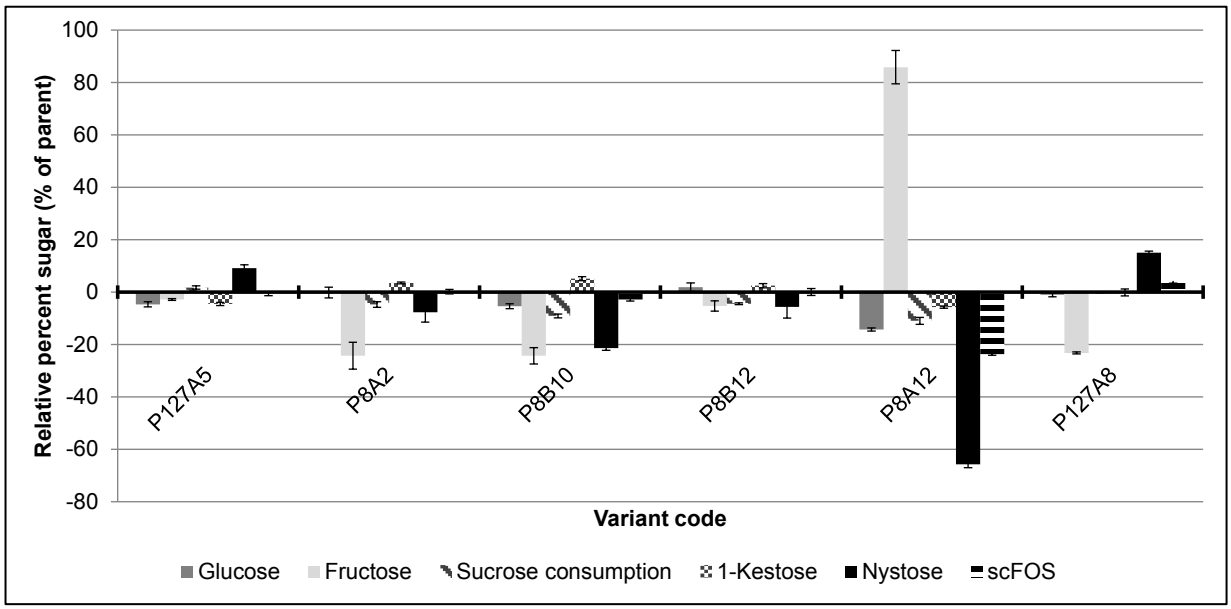

Fig. 3.6 Results of screen II showing percentages of sugars, as quantified by HPLC, for different variants normalised to the parent. Assays were performed in triplicate with 200 g/l sucrose at 55 °C for 2 h. Variant P127A8 was selected for further screening as it produced more scFOS than the parent. Error bars denote standard error (n = 3)

3.3.6 Evaluating the random library: Screen III of the best variant

The A458V (i.e. P127A8) variant enzyme expressed by a new yeast transformant displayed a similar phenotype to the original variant identified in screen II. These data confirmed that the altered phenotype observed was a result of the mutation introduced into *fopA* coding sequence rather than an artefact generated during yeast transformation. The performance of this variant was compared to the parental activity under conditions resembling industrial batch scFOS production. The results are shown in Fig. 3.7. The variant outperformed the parent enzyme in terms of 1-kestose and nystose production throughout the 7 h of incubation. However, the difference became less pronounced with time. The largest difference in 1-kestose levels (32 g/l) was observed at 4.5 h. For nystose the peak difference of 16.8 g/l was also observed at 4.5 h. The levels of GF4 production were similar for both enzymes across all time points with maximal production of 7.7 g/l recorded at 7 h. After 7 h variant A458V produced 228 g/l total scFOS compared to 217 g/l by the parent with a peak difference of 49 g/l scFOS observed at 4.5 h. The percentage composition of scFOS was similar for both enzymes after 7 h. The parent enzyme produced 1-kestose, nystose and GF4 in a 48:48:3.6 ratio (%), while variant A458V produced the oligosaccharides in a 49:48:3.4 ratio.

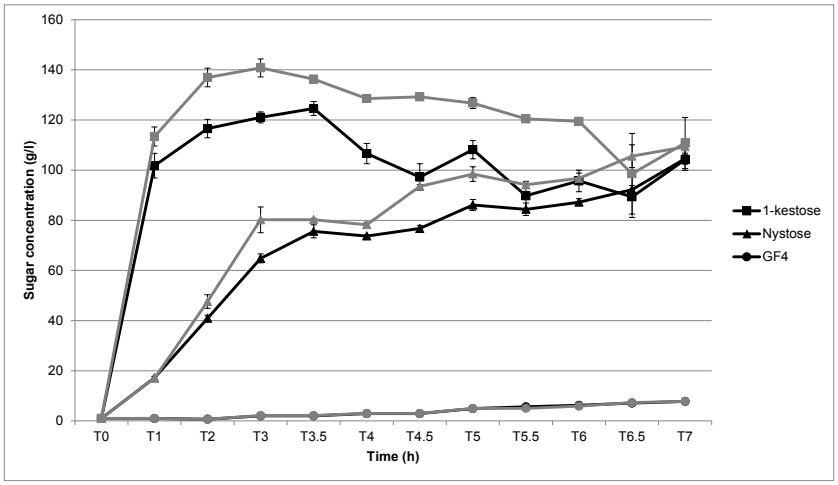

Fig. 3.7 Time course scFOS production by the parent (black lines)and variant A458V (grey lines). The assay was conducted at 62 °C for 7 h with 600 g/l sucrose and data points represent the average of four technical repeats. Equal units of each enzyme were dosed. Error bars denote standard error

3.4 DISCUSSION

To effectively engineer a protein by directed evolution requires optimisation of several steps involved in library construction and screening. The optimisations of error prone PCR mutagenesis methods, library cloning, host transformation and protein expression have been explored extensively in bacterial and yeast hosts, each with associated advantages and disadvantages. In this study our aim was to develop a simple, rapid and cost effective method for screening a β-fructofuranosidase library in yeast with the use of FT-MIR ATR spectroscopy. Only a few reports are available on engineering β-fructofuranosidases (de Abreu et al. 2011; Lafraya et al. 2011) with the study by de Abreu et al. (2011) being the first to report the screening of a sizeable random library. Although there is considerable interest in these enzymes for application in the production of fructooligosaccharides (Hirayama et al. 1989; Linde et al. 2009; Maiorano et al. 2008), we attribute the limited number of reports detailing engineering of β-fructofuranosidases to the limitations imposed by the absence of a direct, rapid and robust library screening method. The use of FT-MIR ATR spectroscopy proved to be a major step towards meeting the requirements for the screen, with sample analysis taking approximately 1 minute per sample to acquire raw data. The calibration for sucrose was linear between 5 and 200 g/l, which negated the need for sample dilution prior to analysis, further fast-tracking analysis. The coefficient of variation of the screen was within the acceptable level 10% (Arnold and Georgiou 2003) which endorsed the design of our screening method. To our knowledge, this is the first report of the direct quantification of sucrose in screens pertaining to a directed evolution study. The screen developed by de Abreu et al. (2011) was based on the dual colorimetric, enzyme-linked quantification of glucose and fructose and subsequent determination of the glucose to fructose ratio as indicator of variant performance. Still outstanding for both the enzyme-linked and FT-MIR ATR spectroscopy approaches is the direct

quantification of the products of the enzymes' transferase activity. Both screens still rely on the use of HPLC to confirm the production of scFOS by putative hits.

Library construction using homologous recombination in yeast (Oldenburg et al. 1997) greatly simplified the process. No cloning steps were required after epPCR and yeast transformation was performed using a routine heat shock method. Optimisation of yeast transformation by means of electroporation has been shown to increase library size significantly (Benatuil et al. 2010) and may be applied to the experimental approach used in this study to potentially generate highly productive yeast variant libraries in the order of 10^9 transformants. The yeast shuttle vector used in this study allowed for constitutive expression and secretion of variants, thereby simplifying production of enzyme, which involved a single sub-culturing step and centrifugation following the end of the yeast cultivation. At 242 transformants, the size of the library generated in this study by no means meets requirements to thoroughly investigate the available sequence space to engineer the β-fructofuranosidase at hand for major improvements in enzyme activity. Instead, we detail the development of a variant discovery pipeline that can be applied to β-fructofuranosidase libraries. With the employment of automated liquid handling robots and a microplate extension accessory to an infrared spectrometer, a high-throughput method for screening large β-fructofuranosidase libraries should prove feasible.

Nishizawa et al. (2001) studied the kinetics of scFOS synthesis by the β-fructofuranosidase of *A. niger* ATCC 20611 and showed the enzyme to be non-competitively inhibited by glucose. They designed a nanofiltration system that selectively removed glucose from the reaction mixture and were able to increase the conversion of a further 5% of sucrose substrate. Other groups have attempted to relieve product (glucose) inhibition by the addition of glucose isomerase (Guio et al. 2012; Yoshikawa et al. 2008) or glucose oxidase (Jong and Seung 1993; Sheu et al. 2001) to the reaction to convert glucose to fructose and gluconic acid, respectively. Both strategies have proven effective, but negatively impact the economics of the scFOS production process.

In an attempt to minimise the effect of glucose inhibition on scFOS production in a batch process, we screened our library for variants with lower sensitivity to glucose. The A458V variant displayed the lowest sensitivity to glucose and displayed improved scFOS production activity relative to the parent. When assayed under conditions resembling an industrial process, the peak difference in total scFOS production between the variant and the parent was 28%. However, at the point in the reaction where 1-kestose and nystose levels were equal the variant produced 5% more scFOS. This is the point where the reaction is typically stopped in industrial scFOS production (Nishizawa et al. 2001).

The A458V single amino acid substitution was located in the beta-sheet designated βV4-2 by Chuankhayan et al. (2010) which forms part of the five-bladed β-propeller domain characteristic of enzymes belonging to the GH32 family. The blades of the propeller surround a deep central cavity where substrate as well as glucose binds. The protein sequences for the β-fructofuranosidase (FopA) isolated from the strain originally designated *A. niger* ATCC 20611 and the β-fructofuranosidase from *A. japonicus* have a 99% identity, which allows for direct comparison of the variant to the solved crystal structures for the *A. japonicus* β-fructofuranosidase. A valine for alanine substitution is classified as neutral and although the side chain is bulkier, it is not expected to disrupt the structure of a β-sheet (Betts and Russell 2003). Retention of enzyme activity attested to this generalisation. As the substitution was in a region of the protein that does not line the active site pocket nor directly interacts with substrates, it is not possible to definitively rationalise its effect on protein function. It is supposed that the local conformation in the blade of the β-propeller is altered which in turn impacts on enzyme activity. As the difference in activity between the parent and A458V became less pronounced with increased reaction time, we could speculate that the substitution resulted in an altered geometry of the active site pocket and hence, altered interaction of pocket residues with the substrate. In the early stages of the reaction sucrose is the only substrate but as the reaction progresses it is converted to 1-kestose and nystose. As these sugars accumulate they replace sucrose as dominant substrate molecules and the effect of the A458V mutation on enzyme activity is not observed as prominently in this scenario.

In summary, this study introduces a novel, robust FT-MIR ATR spectroscopy-based method for quantifying sucrose in crude assay samples. Without any sample preparation required, this rapid method was applied to screen 242 enzyme variants of a β-fructofuranosidase library. An improved variant was ultimately identified that may serve as starting material for further evolution of this enzyme. A workflow is in place that would enable screening large libraries for improved β-fructofuranosidase traits.

3.5 ACKNOWLEDGEMENTS

The authors would like to acknowledge Jamie Gallant for technical assistance with scanning FT-MIR ATR samples. Furthermore, Drs Steven Pelly and Eugene van Rensburg are also thanked for their valuable insights. This work forms part of the Technology Innovation Agency (TIA) funded project PB110/08. Kim Trollope was supported by a grant from TIA and Stellenbosch University. She also holds a L'Oreal-UNESCO For Women in Science Regional Fellowship in Sub-Saharan Africa.

3.6 LITERATURE CITED

Arnold, FH, Georgiou G (2003) Directed enzyme evolution screening and selection methods. Humana Press, Totowa, NJ

Benatuil L, Perez JM, Belk J, Hsieh C (2010) An improved yeast transformation method for the generation of very large human antibody libraries. Protein Eng Des Sel 23:155–159

Betts MJ, Russell RB (2003) Amino acid properties and consequences of subsitutions. In Barnes MR, Gray IC (eds) Bioinformatics for Geneticists. Wiley, New York, NY

Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast 14:115–132

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 37:D233–D238

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

Cobb RE, Si T, Zhao H (2012) Directed evolution: an evolving and enabling synthetic biology tool. Curr Opin Chem Biol 16:285–291

Cozzolino D, Curtin C (2012) The use of attenuated total reflectance as tool to monitor the time course of fermentation in wild ferments. Food Control 26:241-246

Crous JM, Pretorius IS, Van Zyl WH (1995) Cloning and expression of an *Aspergillus kawachii* endo-1,4-beta-xylanase gene in *Saccharomyces cerevisiae*. Curr Genet 28:467–473

de Abreu MA, Álvaro-Benito M, Plou FJ, Fernández-Lobato M, Alcalde M (2011) Screening β-fructofuranosidases mutant libraries to enhance the transglycosylation rates of β-(2→6) fructooligosaccharides. Comb Chem High Throughput Screening 14:730–738

Dougherty MJ, Arnold FH (2009) Directed evolution: new parts and optimized function. Curr Opin Biotechnol 20:486–491

Duarte IF, Barros A, Delgadillo I, Almeida C, Gil AM (2002) Application of FTIR spectroscopy for the quantification of sugars in mango juice as a function of ripening. J Agric Food Chem 50:3104–3111

Edelman J, Bacon JSD (1951) Transfructosidation in extracts of the tubers of *Helianthus tuberosus* L. Biochem J 49:529–540

Fischer EH, Kohtès L, Fellig J (1951) Propriétés de l'invertase purifiée. Helv Chim Acta 34:1132–1138

Gibson GR, Roberfroid MB (1995) Dietary Modulation of the Human Colonic Microbiota: Introducing the Concept of Prebiotics. J Nutr 125:1401–1412

Guio F, Rugeles LD, Rojas SE, Palomino MP, Camargo MC, Sánchez OF (2012) Kinetic modeling of fructooligosaccharide production using *Aspergillus oryzae* N74. Appl Biochem Biotechnol 167:142–163

Hill J, Donald K, Griffiths D (1991) DMSO-enhanced whole cell yeast transformation. Nucleic Acids Res 19:5791

Hidaka H, Hirayama M, Sumi N (1988) A fructooligosaccharide-producing enzyme from *Aspergillus niger* ATCC 20611. Agric Biol Chem 52:1181–1187

Hirayama M, Sumi N, Hidaka H (1989) Purification and properties of a fructoooligosaccharide-producing β-fructofuranosidase from *Aspergillus niger* ATCC 20611. Agric Biol Chem 53:667–673

Hu X, Robin S, O'Connell S, Walsh G, Wall JG (2010) Engineering of a fungal β-galactosidase to remove product inhibition by galactose. Appl Microbiol Biotechnol 87:1773–1782

Jong WY, Seung KS (1993) The production of high-content fructo-oligosaccharides from sucrose by the mixed-enzyme system of fructosyltransferase and glucose oxidase. Biotechnol Lett 15:573–576

Kuchner O, Arnold FH (1997) Directed evolution of enzyme catalysts. Trends Biotechnol 15:523–530

Kupina SA, Shrikhande AJ (2003) Evaluation of a Fourier transform infrared instrument for rapid quality-control wine analyses. Am J Enol Vitic 54:131–134

Lafraya Á, Sanz-Aparicio J, Polaina J, Marín-Navarro J (2011) Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase. Appl Environ Microbiol 77:6148–6157

Leopold LF, Leopold N, Diehl H-A, Socaciu C (2011) Quantification of carbohydrates in fruit juices using FTIR spectroscopy and multivariate analysis. Spectrosc-Int J 26:93–104

Li-Chan E, Chalmers JM, Griffiths P (2010) Applications of vibrational spectroscopy in food science: Volumes I and II: Analysis of food, drink and related materials. John Wiley & Sons, Chichester

Linde D, Macias I, Fernández-Arrojo L, Plou FJ, Jiménez A, Fernández-Lobato M (2009) Molecular and biochemical characterization of a β-fructofuranosidase from *Xanthophyllomyces dendrorhous*. Appl Environ Microbiol 75:1065–1073

Maiorano AE, Piccoli RM, Da Silva ES, De Andrade Rodrigues MF (2008) Microbial production of fructosyltransferases for synthesis of prebiotics. Biotechnol Lett 30:1867–1877

Max J-J, Chapados C (2007) Glucose and fructose hydrates in aqueous solution by IR spectroscopy. J Phys Chem A 111:2679–2689

Max J-J, Chapados C (2001) Sucrose hydrates in aqueous dolution by IR spectroscopy. J Phys Chem A 105:10681–10688

Nieuwoudt HH, Pretorius IS, Bauer FF, Nel DG, Prior BA (2006) Rapid screening of the fermentation profiles of wine yeasts by Fourier transform infrared spectroscopy. J Microbiol Methods 67:248–256

Nishizawa K, Nakajima M, Nabetani H (2001) Kinetic Study on transfructosylation by β-fructofuranosidase from *Aspergillus niger* ATCC 20611 and availability of a membrane reactor for fructooligosaccharide production. Food Sci Technol Res 7:39–44

Oldenburg KR, Vo KT, Michaelis S, Paddon C (1997) Recombination-mediated PCR-directed plasmid construction in vivo in yeast. Nucleic Acids Res 25:451–542

Patz CD, David A, Thente K, Kürbel P, Dietrich H (1999) Wine analysis with FTIR spectrometry. Vitic Enol Sci 54:80–87

Roberfroid M, Slavin J (2000) Nondigestible oligosaccharides. Crit Rev Food Sci Nutr 40: 461-480

Roberfroid MB (2007) Inulin-Type Fructans: Functional Food Ingredients. J Nutr 137:2493S–2502

Rocklin RD, Clarke AP, Weitzhandler M (1998) Improved long-term reproducibility for pulsed amperometric detection of carbohydrates via a new quadruple-potential waveform. Anal Chem 70:1496–1501

Romero PA, Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol 10:866–876

Ruff AJ, Dennig A, Schwaneberg U (2013) To get what we aim for: progress in diversity generation methods. FEBS J 280:2961–2978

Saarelainen R, Paloheimo M, Fagerström R, Suominen PL, Nevalainen KM (1993) Cloning, sequencing and enhanced expression of the *Trichoderma reesei* endoxylanase II (pI 9) gene xln2. Mol Gen Genet 241:497–503

Sambrook J, Fritsch E, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schenk J, Marison IW, Von Stockar U (2007) A simple method to monitor and control methanol feeding of *Pichia pastoris* fermentations using mid-IR spectroscopy. J Biotechnol 128:344–353

Sherman F, Fink G, Lawrence CW (1979) Methods in Yeast Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sheu DC, Lio PJ, Chen ST, Lin CT, Duan KJ (2001) Production of fructooligosaccharides in high yield using a mixed enzyme system of β-fructofuranosidase and glucose oxidase. Biotechnol Lett 23:1499–1503

Shimotohno A, Oue S, Yano T, Kuramitsu S, Kagamiyama H (2001) Demonstration of the importance and usefulness of manipulating non-active-site residues in protein design. J Biochem 129:943–948

Sills DL, Gossett JM (2012) Using FTIR spectroscopy to model alkaline pretreatment and enzymatic saccharification of six lignocellulosic biomasses. Biotechnol Bioeng 109:894–903

Tielmann P, Boese M, Luft M, Reetz MT (2003) A practical high-throughput screening system for enantioselectivity by using FTIR spectroscopy. Chem Eur J 9:3882–3887

Wang J, Kliks MM, Jun S, Jackson M, Li QX (2010) Rapid analysis of glucose, fructose, sucrose, and maltose in honeys from different geographic regions using Fourier transform infrared spectroscopy and multivariate analysis. J Food Sci 75:C208–C214

White LM, Secor GE (1952) The oligosaccharides formed during the sucrose-invertase reaction. Arch Biochem Biophys 36:490–491

Williams PC (2001) Implementation of Near-Infrared technology, In: Williams PC, K.H. Norris (eds.) Near infrared technology in the agricultural and food industries. American Association of Cereal Chemists, St. Paul, MN, pp 145–169

Yanai K, Nakane A, Kawate A, Hirayama M (2001) Molecular cloning and characterization of the fructooligosaccharide-producing β-fructofuranosidase gene from *Aspergillus niger* ATCC 20611. Biosci Biotechnol Biochem 65:766–773

Yoshikawa J, Amachi S, Shinoyama H, Fujii T (2008) Production of fructooligosaccharides by crude enzyme preparations of β-fructofuranosidase from *Aureobasidium pullulans*. Biotechnol Lett 30:535–539

3.7 APPENDIX: SUPPLEMENTARY MATERIAL

Table S1 Summary of regression statistics for calibration and validation of the sucrose model generated from FT-MIR ATR spectroscopy data

Calibration set		Validation set	
Sample number	134	Sample number	58
Number of PLS factors	6	Calibration bias	1.14×10^{-13}
		R^2	0.958
SECV (g/l)	12.53	SEP (g/l)	14.69
Concentration range (g/l)	4.08 – 204.08	Concentration range (g/l)	6.74 – 204.93
Mean ± S.D. (g/l)	82.49 ± 62.15	Mean ± S.D. (g/l)	96.12 ± 59.80
SEL (%)	10%	RPD	4.07

RPD residual predictive deviation (standard deviation of validation samples/SEP)
PLS partial least squares
SECV standard error of cross validation
S.D. standard deviation
SEL standard error of laboratory obtained with HPLC
SEP standard error of prediction

Table S2 Summary of independent test set validation statistics for the sucrose regression model of FTIR spectroscopy data

Calibration:Validation split (%)	R^2	RMSEP (g/l)
70:30	0.958	14.69
70:30	0.960	12.99
70:30	0.971	12.25
70:30	0.964	12.06
50:50	0.955	15.07

CHAPTER 4

RESEARCH RESULTS II

Direct, simultaneous quantification of fructooligosaccharides by FT-MIR ATR spectroscopy and chemometrics for rapid identification of superior, engineered β-fructofuranosidases

4 RESEARCH RESULTS II:

Direct, simultaneous quantification of fructooligosaccharides by FT-MIR ATR spectroscopy and chemometrics for rapid identification of superior, engineered β-fructofuranosidases

K.M. Trollope[1], J.F. Görgens[2], R. Bro[3], H. Volschenk[1], H.H. Nieuwoudt[4]

[1] Department of Microbiology, Stellenbosch University, Private Bag X1, Stellenbosch, 7602, South Africa
[2] Department of Process Engineering, Private Bag X1, Stellenbosch University, Stellenbosch, 7602, South Africa
[3] Department of Food Science, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
[4] Institute for Wine Biotechnology, Department of Viticulture and Oenology, Stellenbosch University, Private Bag X1, Stellenbosch, 7602, South Africa

ABSTRACT

Fructooligosaccharides are popular components of functional foods produced by the enzymatic transfer of fructose units to sucrose. Improving β-fructofuranosidase traits by protein engineering is restricted by the absence of a rapid, direct screening method for the fructooligosaccharide products produced by enzyme variants. The use of standard high performance liquid chromatography methods involves time-consuming sample preparation, chromatographic and data analysis steps. To overcome these limitations, this work presents a rapid method for screening β-fructofuranosidase variant libraries using Fourier transform mid infrared attenuated total reflectance spectroscopy and calibration using partial least squares (PLS) regression. The method offers notable improvements in terms of sample analysis times and cost, with the added benefit of the absence of toxic eluents. Exploratory data analysis employed orthogonal PLS to investigate the prevalence and potential sources of analyte uncorrelated variation in the spectra. Several combinations of spectral preprocessing and wavenumber interval selection methods were tested to develop optimised PLS regression models that were successfully applied to quantify glucose, fructose, sucrose, 1-kestose and nystose, the substrates and products of β-fructofuranosidase activity. To the best of our knowledge this is the first report of the use of infrared spectroscopy and PLS calibration for the quantification of 1-kestose and nystose. Results indicated that optimal wavenumber selection by interval PLS served to provide the best models for all sugars tested. Independent test set validation of the models showed between 20 and 48% reduction in average root mean square error of prediction, relative to models built with raw, full spectra. Application of this screening method will facilitate the engineering of β-fructofuranosidases and other glycosyltransferase enzymes by random mutagenesis strategies, as it provides, for the first time, a rapid, direct assay for transferase products that may be adapted to a high throughput set-up.

4.1 INTRODUCTION

Inulin-type short chain fructooligosaccharides (FOS) are linear fructose polymers linked by β-(2→1) bonds to a terminal glucose moiety, with a degree of polymerisation ranging from three to five units (Waterhouse and Chatterton 1993). FOS-rich products are considered functional foods as their consumption promotes human health by selectively stimulating the growth of bifidobacteria in the gut (Buddington et al. 1996; Gibson and Roberfroid 1995). Global demand for FOS is increasing which generates the impetus for improving industrial production processes. Current methods to produce these oligosaccharides include the hydrolysis of the fructose polymer, inulin or the enzymatic transfer of successive fructose units to sucrose by β-fructofuranosidases (Hidaka et al. 1988; Sangeetha et al. 2005a; Singh and Singh 2010). β-fructofuranosidases from fungal strains of *Aspergillus niger* and *Aureobasidium pullulans* are particularly effective at converting sucrose to FOS (Hidaka et al. 1988; Yoshikawa et al. 2008). Strategies to improve the yields of enzymatic FOS synthesis from sucrose are multifaceted and generally involve the optimisation of enzymatic FOS synthesis conditions (Nemukula et al. 2009; Vega and Zúniga-Hansen 2011) and enzyme production by host microorganisms in the case of whole cell biotransformation (Dominguez et al. 2012; Sangeetha et al. 2005b). Further advances in this field would rely on obtaining superior enzyme catalysts by either bio-prospecting for novel enzymes from FOS-producing microbes or engineering the existing enzymes for improved traits. Mutagenesis of the gene encoding the relevant enzyme forms the basis for engineering the protein and, depending on the mutagenesis strategy, often generates large libraries (10^3 – 10^7 colony forming units) of heterologously expressed enzyme variants. The libraries need to be screened to assess the resultant phenotypes and identify superior enzymes (Kuchner and Arnold 1997). Thus a rapid, cost effective high-throughput analytical method is highly advantageous to screen enzymes produced by separate, individual and often repeated cultivation of the host microbes. Furthermore, a direct screen for the product of interest is the optimal scenario, though many indirect assays methods are acceptable and implemented for practical purposes. For example, glucose can be rapidly quantified in multiple samples simultaneously by an enzymatic method and serves as an indicator of global β-fructofuranosidase activity. The global activity includes both sucrose hydrolysis and fructosyltransferase activity (de Abreu et al. 2011). The assay for glucose is typically applied due to the lack of a direct, high-throughput screening method for FOS. At the screening stage, higher glucose concentration in the assay medium implies higher enzyme activity. However, enzymes with enhanced hydrolytic activity will also yield more glucose and hence lead to the identification of false positives, in terms of FOS production, by the indirect screening method. Reduction of false positives can be achieved additionally quantifying fructose (rapid, indirect method) however, at further expense and time investment. A direct assay for FOS would have

reduced the screening burden significantly by replacing the two indirect screening methods with a single screen.

The routine analytical methods to quantify FOS, and hence enzyme transferase activity, rely on high performance liquid chromatography (HPLC). These methods are effective, but often require protracted sample preparation, extended run times and lengthy data processing steps to accurately quantify the sugars (Corradini et al. 2013; L'homme et al. 2001; Maugeri and Hernalsteens 2007). Furthermore, waste generated by HPLC often is hazardous to the environment and requires specialised treatment and disposal. The few reports on engineering β-fructofuranosidases/fructosyltransferases detail library screens that either used enzyme-based quantification of glucose and fructose as indirect measures of fructooligosaccharide synthesis (de Abreu et al. 2011) or direct visualisation of oligosaccharides by thin layer chromatography (Beine et al. 2008). The aforementioned methods lack either speed, ease of implementation or do not directly quantify the products of interest.

Literature abounds on sugar quantification using Fourier transform mid infrared (FT-MIR) spectroscopy in food products including fruits (Bureau et al. 2009; Duarte et al. 2002), wine fermentations (Cozzolino and Curtin 2012) and honey (Wang et al. 2010). The emphasis often falls on quantifying glucose, fructose and sucrose and other simple sugars that predominate. Blanch *et al*. (Blanch et al. 2012) described the qualitative characterisation of fructooligosaccharides using FT-MIR attenuated total reflectance (ATR) spectroscopy as an indicator of strawberry fruit water status. However, no literature is available on the quantification of fructooligosaccharides using Fourier transform infrared spectroscopy (mid or near infrared). The use of FT-MIR spectroscopy to evaluate the kinetics of enzyme reactions, including sucrose hydrolysis by β-fructofuranosidase has been described previously.(Cadet et al. 1995; Schindler et al. 1998) In terms of the wider applicability to biotechnology research, the methods described are limited by their single batch source of enzyme or the requirement of prior costly, low yield enzyme purification which negate matrix effects. Between-batch matrix variation can severely compromise the power of partial least squares (PLS) prediction models if not accounted for in the design of calibration sample sets. Typically, experiments involving the heterologous production of enzymes evaluate numerous batches of enzymes generated from repeated cultivations of host microbes. The cell-free culture supernatants or cell extracts are used as sources of enzyme (i.e. unpurified) for subsequent investigations into enzyme activity. Although cultivation media and conditions are standardised, the intrinsic variability of biological processes results in inconsistent chemical compositions of the resulting cultures of different batches, and therefore also changes in the relationships between the X-variables (spectral wavenumbers). Riley *et al.* (Riley et al. 1998) describe hurdles to quantifying nutrients and by-products in insect-cell bioreactors using near infrared spectroscopy. They noted that calibration

models with the best predictive ability for unseen samples were built using calibration samples encompassing the greatest variation in background matrix composition.

Partial least squares (PLS) regression (Geladi and Kowalski 1986) is a well-established method for multivariate modelling and calibration. More recent developments aimed at improvement of PLS regression statistics include data preprocessing and variable selection methods, as discussed below. Spectroscopic data generated over extended time periods are subject to variation originating from changes in the sample composition, as well as instrumental performance. Minimisation of the variation not pertaining to analyte concentration will usually contribute towards the development of accurate quantitative prediction models for future samples. A range of spectral preprocessing methods has been developed to remove uncorrelated systematic variation in data (Rinnan et al. 2009) which can sometimes improve the predictive ability of PLS models, depending on the type of systematic variation in the data. In addition to these methods, the Y-block (sugars in this study) focused orthogonal projections to latent structures (O-PLS) method developed by Trygg and Wold (Trygg and Wold 2002) removes uncorrelated variation in the X-block (spectra) and separates it into orthogonal components. This allows for identification of some sources of unrelated variation and simplifies the interpretation of PLS models. O-PLS has been applied in several studies (Stenlund et al. 2009; Stenlund et al. 2008; Tapp and Kemsley 2009).

Selection of the best wavenumber intervals is another technique used to simplify and improve PLS modelling. Sections of the spectra can be selected *a priori* using chemical knowledge about infrared spectra of the analytes of interest, for example selecting the fingerprint region in the case of carbohydrate modelling (Trollope et al. 2014). Interval PLS (iPLS) is an automated method that builds local PLS models for smaller intervals of a spectrum and compares them to the global PLS model for the full spectrum. The influence of including selected intervals on the root mean square error of cross validation (RMSECV) of the local model is compared to that of the global model. In so doing variables that are important for predicting analyte concentration are identified and interfering variables are excluded, thereby improving the predictive power of the model (Norgaard et al. 2000). A further advantage of this technique is the graphical output of the results of the local PLS modelling in the original X variable space (i.e. spectral wavenumbers) that facilitates the interpretation of the models in terms of the chemistry of the analytes.

Previously, a potential high-throughput screen for β-fructofuranosidase libraries based on FT-MIR ATR spectroscopic quantification of sucrose as an indicator of enzyme activity (Trollope et al. 2014) was reported. As follow-on to the initial work, the scope was expanded to develop PLS regression calibration models of FT-MIR ATR spectra for the quantification of sugars

relevant to β-fructofuranosidase activity – glucose, fructose, sucrose, 1-kestose and nystose. Different preprocessing and interval selection methods were evaluated with the aim to obtain models with improved predictive performance over raw, full spectra models. Optimisation of sucrose modelling was therefore included in this work. The pure component spectra for the five sugars were presented and the interval selections from PLS modelling were interpreted in terms of the chemistry of these spectra. In addition, some sources of variability related to the spectra collected over different rounds of mutagenesis and between-batch variation were identified and could be interpreted, using O-PLS regression. Application of the calibration models established in this work, will allow for the simultaneous quantification of the five aforementioned sugars in under a minute per sample, which represents a significant reduction in analysis time relative to current methods. This analytical approach will enable comprehensive screening of β-fructofuranosidase variant libraries and greatly facilitate the engineering of these enzymes by providing, for the first time, a rapid, direct screen for the substrates and preferred oligosaccharide products of enzyme activity.

4.2 MATERIALS AND METHODS

4.2.1 Samples and reference data

A detailed description of the samples and the reference HPLC method has been published (Trollope et al. 2014). Briefly, liquid samples were products of different batches of enzyme assays which reacted β-fructofuranosidase enzymes present in yeast supernatants with buffered sucrose (200 g L^{-1}). Batches in the context of this work refer to sets of samples generated over time by separate yeast cultivation events and subsequent assays. Fresh yeast growth media and enzyme substrate were prepared as per standard formulation for each set of cultivations and assays. Reference data for the samples were obtained using HPLC or by spiking known amounts of sugars into samples comprised of reference yeast (transformed with empty vector) supernatant combined 1:1 with assay substrate. Spiked samples served to expand the calibration concentration range of the models. They contained random combinations of the 5 sugars and thus did not necessarily display the same correlations between sugars as did the samples obtained from enzyme assays. Most samples contained a mixture of glucose, fructose, sucrose, 1-kestose and nystose with certain samples containing only glucose, fructose and sucrose. A summary of the sample statistics of the datasets used for the development of PLS regression calibration models is given in Table 4.1.

HPLC grade standards for D-glucose, D-fructose and D-sucrose were obtained from Sigma Aldrich (Schnelldorf, Germany). 1-Kestose and nystose (>99% purity) were purchased from Wako Chemicals GmbH (Neuss, Germany). The dried powders were scanned to obtain pure component spectra.

Table 4.1 Sample statistics of datasets used in PLS modelling.

	Glucose	Fructose	Sucrose	1-Kestose	Nystose
Number of samples*	112 (8)	151 (5)	180 (9)	159 (8)	129(7)
Number of batches	4	4	8	9	7
Concentration range (g/l)	53.0 – 104.5	2.6 – 88.0	4.1 – 204.1	0 - 77.6	1.0 – 43.0
Mean (g/l) ± S.D.	80.1 ± 13.3	16.3 ± 13.6	81.8 ± 60.7	25.9 ± 18.4	11.5 ± 9.8

*Number of spiked samples is indicated in brackets

4.2.2 FT-MIR ATR procedure

The spectrometer used was an Alpha-P FT-MIR ATR instrument fitted with a heatable diamond crystal sample plate with single reflection (Alpha Bruker Optics GmbH, Ettlingen, Germany). The absorbance spectrum of each liquid sample (80 μL) was obtained at 26°C by averaging 64 repeat scans at a resolution of 4 cm^{-1} and in the range 3996 to 550 cm^{-1}, with a scanner velocity of 7.5 KHz. Prior to scanning a reference background measurement was taken against air. The instrument was operated using OPUS software version 6.5 for Windows XP (Bruker Optics GmbH, Ettlingen, Germany). Over the course of construction of the β-fructofuranosidase library, several different Alpha spectrometers were used, based on instrument availability. Crystalline powders of pure sugars were scanned under the same conditions on an Alpha-P instrument fitted with a clamping device on the sample plate, in order to ensure good contact between the powder and the diamond crystal.

4.2.3 Data analysis

Pure component spectra were subjected to baseline correction in OPUS by the rubber band method, using 64 baseline points. The OPUS band assignment function was also used to identify potential functional groups represented in the spectra. The chemical structures of the sugars were referenced to identify the annotated functional groups that were relevant to the carbohydrate chemistries. Software band assignment was generalised in the sense that spectral bands were included for bonds found in proteins, lipids and other organic molecules.

For samples used in calibrations, spectra were exported from OPUS software in data point format (*.dpt) and imported into Excel version 2010 (Microsoft, Redmond, Washington, USA) where reference data were included. A category variable was added to identify different batches. The datasets were then imported into Solo standalone Toolbox version 7.5 (Eigenvector Research, Inc., Wenatchee, Washington, USA) for PLS and iPLS, and into Simca version 13.0.3.0 (Umetrics AB, Umeå, Sweden) for O-PLS analyses. Calibration datasets for

the modelling of the different sugars did not all include the same samples and/or batches, although there was some overlap.

The structures of the datasets were explored using principal component analysis (PCA) (Jolliffe 2002)and PLS regression. Gross outliers were removed after closer scrutiny of the leverage versus Y studentised residuals plots. For each sugar, partial least squares regression models were generated using spectra ranging from 3996.461 to 901.135 cm^{-1}. The spectra were trimmed from the lower end reducing the range from 550 cm^{-1} to 901 cm^{-1} as the libration bands of water dominated in this region (Max and Chapados 2007). The FT-MIR absorbance data and sugar concentrations served as the X and Y matrices, respectively. The effects of spectral preprocessing on modelling were investigated by comparing regression statistics of models generated when a selection of spectral filters (Savitzky-Golay first derivative and smoothing, orthogonal signal correction)normalization (mean mulitiplicative signal correction (MSC)) as well as variable selection methods (forward and reverse iPLS) were applied. Default preprocessing for the X-block used mean centering (after possible application of other types of preprocessing) while for the Y-block, data were autoscaled to unit variance. A venetian blinds design with 10 data splits was used to cross validate the calibration models during the training stage. The preprocessing was also part of the cross-validation. In the test stage, the predictive capacities of the models were evaluated using independent test sets generated by the automated onion function in Solo. Briefly, the most unique samples, in terms of distance from the mean, were placed in the calibration set and the next layer of unique samples placed into the test set. Remaining samples were randomly split between calibration and test sets. Calibration and validation sets were selected such that both sets spanned the entire concentration range of reference values. Model validation was repeated three times with 70:30% (calibration: validation) data splits and once with a 50:50% split. The number of latent variables (LVs) used in each model was selected to give low RMSECV values while simultaneously avoiding overfitting. Models that consistently gave the lowest RMSECV values were considered best. Finally, independent test sets were used to determine the root mean square error of prediction (RMSEP) of the best models.

4.3 RESULTS AND DISCUSSION

4.3.1 Analysis of pure component and assay sample spectra

Clear differences between the sugars' spectra were apparent in the spectral regions of 3600–2800 cm^{-1} and below 1500 cm^{-1} (Fig. S1 – see appendix). In the context of the pure sugars, O-H and C-H stretching were responsible for the bands in the 3600–3000 cm^{-1} and 3000–2800 cm^{-1} regions, respectively. The ATR diamond exhibits broad bands in the 2700–1800 cm^{-1} regions due to carbon bond absorption while hydrogen and nitrogen content is responsible for

bands at 3100 cm^{-1} and 1500–1000 cm^{-1}, respectively (Thongnopkun and Ekgasit 2005). The peaks were not clearly visible as they overlapped with absorption bands of the sugars. Fig. 4.1 shows the line spectra between 1500–900 cm^{-1} for the pure sugar powders (A and B) and liquid samples (C) used in calibration modelling. This region showed the greatest amount of spectral variability between the sugars.

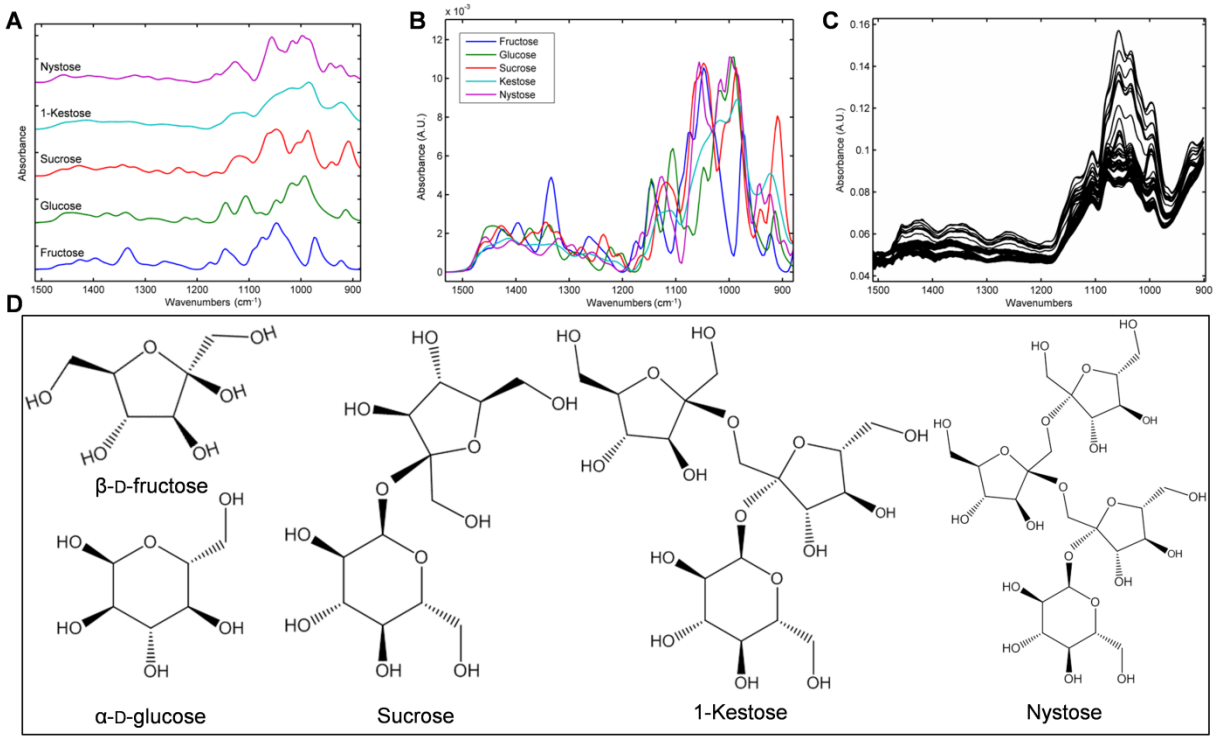

Fig. 4.1 Pure component powder FT-MIR ATR spectra and chemical structures for fructose, glucose, sucrose, 1-kestose and nystose. Spectra were collected from 3996 to 550 cm^{-1} with the 1500–900 cm^{-1} region showing the highest degree of variability between the sugars. The ATR diamond exhibits broad bands in the 2700–1800 cm^{-1} regions due to carbon bond absorption while hydrogen and nitrogen content is responsible for bands at 3100 cm^{-1} and 1500–1000 cm^{-1}, respectively (Thongnopkun and Ekgasit 2005). A shows stacked spectra in the 1500–900 cm^{-1} region for all sugars while B shows the same spectra overlaid, highlighting the extent of band overlap between the different sugars. C shows the same region for aqueous samples containing mixtures of the 5 sugars. D presents the chemical structures for the 5 sugars (Bolton et al. 2008)

Fig. 4.1A shows the unique banding pattern for the sugars while Fig. 4.1B shows overlaid pure component spectra where band overlap was evident. The 1500–1200 cm^{-1} bands result from C-O-H in plane bending. The region between 1200 and 900 cm^{-1} contained overlapping bands resulting from C-O stretching vibrations and included C-O-H, C-C-H and O-C-H bending of the anomeric configuration of carbohydrates that occurred between 950–750 cm^{-1} (Li-Chan et al. 2010). Spectra are shown up to 900 cm^{-1} as that was the cut-off for modelling purposes. The spectra of the liquid samples (Fig. 4.1C) can be considered composites of different concentrations of sugar mixtures in the yeast supernatant/buffer matrix. Due to the high degree of band overlap, multivariate methods are necessary for predicting specific sugar concentrations. Further complexities resulted from the aqueous nature of the samples used in

this study. The intensity of the peak at 3550–3200 cm^{-1} resulting from -OH bond stretching in sugars and water has been shown to be modified by their intermolecular interactions (Max and Chapados 2001). Furthermore, two aqueous hydrates each for glucose, fructose and sucrose are known with slightly altered spectra in the C-O stretch region for each of the hydrates. The aqueous hydrate abundances are concentration dependent (Max and Chapados 2007; Max and Chapados 2001). Fig. 4.1D shows the chemical structures of glucose, fructose, sucrose, 1-kestose and nystose. The sugars shared a high degree of structural similarity with ether, alkane, and -OH bonds present in all sugars. None of the sugars possessed a unique bond, rather different numbers and arrangements of the same types of bonds. Peak assignment to specific bond positions in the sugar molecule was not possible (and beyond the scope of this work) – only the type of bond could be assigned. The relevance of the sugar chemistries to the PLS models will be further discussed in section 4.3.3.

4.3.2 O-PLS analysis

It was of interest to detect the sources of predictive and uncorrelated variability in the datasets and for this purpose O-PLS was used. As opposed to spectral preprocessing methods that change the spectra, O-PLS does not alter the spectra, and therefore this method allows interpretation and insight into the influence of experimental and sampling issues. This was particularly necessary as several different Alpha-P FT-MIR ATR instruments were used over the course of the study (owing to their availability from the supplier). By way of example the O-PLS modelling of the glucose dataset will be discussed. A five component O-PLS model was calculated based on cross-validation. The model had one predictive component and 4 orthogonal components. Of the total variation in the X-block, 2.4% held information pertaining to the prediction of Y while 95.7% was orthogonal to Y. For the remaining sugars a maximum of 35.7% of X-block variation was used for the prediction of Y-block variation. The loadings plots for the predictive and orthogonal components indicate where spectral variation is correlated to analyte prediction or not. Positive or negative values for the predictive loadings indicate a correlation (positive or negative) between that wavenumber and analyte concentration. Values close to zero indicate little correlation between those wavenumbers and analyte concentration. Positive or negative values for the orthogonal component loadings indicate systematic variation unrelated to Y for that wavenumber. The direction of this variation is associated with the direction in the scores. If the orthogonal score for the first component is positive for a specific sample, then the variables that have a positive loading are generally high and vice versa. The scores and loadings for the predictive and first orthogonal component for the glucose O-PLS model are shown in Fig. 4.2A and B, respectively. Examination of the scores plot indicated that sample distribution was not random and grouping was evident. The separation of samples in the vertical direction was inconsistent with glucose concentration (data not shown). Rather, the

grouping appeared to be due to unknown factors that were batch related – samples were numbered according to batches.

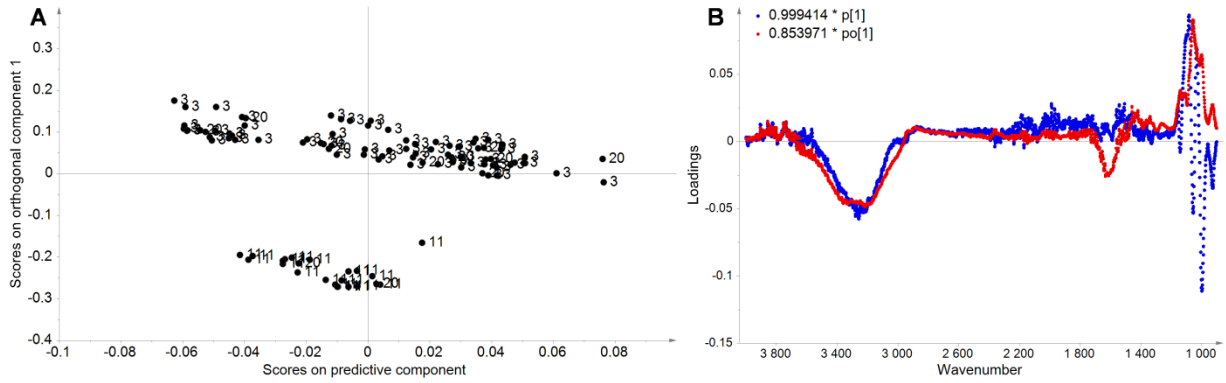

Fig. 4.2 **A** Scores plot of the predictive component versus orthogonal component 1 in O-PLS modelling of glucose. Sample batches are numbered. B shows the loadings for the predictive component (p1) in blue and the first orthogonal component (po1) in red

Samples in batch 11 grouped separately with some overlap between samples from batches 3 and 20. Similar batch groupings were observed in the O-PLS modelling of the other sugars. The loadings plot (Fig. 4.2B) for the predictive component (blue) indicated that wavenumbers in the 3500–3000 cm^{-1} region and below 1200 cm^{-1} were important for the prediction of Y. The orthogonal component (red) loadings were negative in the 3600–3000 cm^{-1} and 1600 cm^{-1} regions and positive below 1200 cm^{-1}. The extensive degree of spectral band overlap of the sugars (due to their structural similarity) and water discussed in section 3.1 could account for much of the X-block variation being uncorrelated to Y. Loadings were high in both the predictive and orthogonal components in the region below 1200 cm^{-1} which was where most of the differences between the spectra of pure components were identified. Similarly for the region of –OH stretching around 3500–3000 cm^{-1} where bands overlap for water and sugars. The trough in the orthogonal component in the 1600 cm^{-1} region corresponded to H-O-H bending of water. Carbohydrates do not absorb in this region but in aqueous solutions the concentration of water decreases with increased solute concentration, with a resulting diminished intensity of the H-O-H band (Max and Chapados 2007). Systematic variation in this region could be expected as a function of total solute concentration and hence the variation was orthogonal to the prediction of a specific sugar concentration. A peak or trough in this region was also observed in the orthogonal components for the other sugars. Based on the orthogonal component scores, much of the orthogonal variation in the 3500–3000 cm^{-1} and 1600 cm^{-1} regions stemmed from batch 11 samples.

Splitting the variation in the X-block into the predictive and orthogonal components aided in the understanding of the sources of variation and helped to clarify why so much of the spectral information was not correlated to Y-block variation. As much of the orthogonal variation could

be attributed to the similarities between the sugars and the batch effect, it can be appreciated that altering experimental design would not likely result in higher amounts of Y-block correlated X-variation. The Y-orthogonal variation was inherent to the experiments.

4.3.3 PLS calibrations

PLS regression models were generated for glucose, fructose, sucrose, 1-kestose and nystose. Reference models were built using raw, full spectra without any preprocessing (except mean centering and autoscaling of X and Y matrices, respectively) and served to evaluate the influence of preprocessing on the validation statistics/RMSEP of subsequent models. The comprehensive set of results is given in the supplementary Table S3-5. Data are shown for models where preprocessing, variable selection and combinations of the two treatments were applied. Model selections with regards to preprocessing, variable selection and number of components were made prior to test set validation – test sets were only used for evaluation of the chosen models. Preprocessing methods reported were not the same for all the sugars as certain methods did not yield acceptable results and were omitted. The calibration and validation statistics for the best models for each sugar are presented in Table 4.2 and the respective modelling will be discussed separately in the sections below. As the predictive power of O-PLS is the same as that for PLS modelling, only the PLS models were optimised using variable selection (iPLS) and subjected to independent test set validation.

Table 4.2 The optimised model results of PLS modelling of FT-MIR spectra for predicting concentrations of glucose, fructose, sucrose, 1-kestose and nystose

	Model				Calibration statistics			Validation statistics		Percentage improvement
								70% data split Average ± SD	50% data split	
	Wavenumbers used (cm^{-1})	**Number of samples**	**Preprocessing/ Variable selection**	**Number of LVs**	**R^2**	**RMSEC (g/l)**	**RMSECV (g/l)**	**RMSEP (g/l)**	**RMSEP (g/l)**	**%**
Glucose	[e]	112	Fingerprint + reverse iPLS	6	0.876	4.3	4.7	4.3 ± 0.4	5.0	20.7
Fructose	[h]	151	OSC + forward iPLS	5	0.973	2.1	2.2	2.0 ± 0.1	2.5	48.2
Sucrose	[l]	180	1st derivative + smoothing (SG) + forward iPLS	5	0.968	10.0	10.8	9.5 ± 0.8	10.9	30.1
1-Kestose	[o]	159	reverse iPLS	7	0.869	5.6	6.7	6.4 ± 0.4	7.5	23.5
Nystose	[v]	129	MSC + reverse iPLS	5	0.937	2.1	2.4	2.3 ± 0.4	3.0	21.7

LV Latent variable
SD Standard deviation (n = 3)
R^2 Coefficient of determination
RMSEC Root mean square error of calibration
RMSECV Root mean square error of cross validation
RMSEP Root mean square error of prediction
RPD Residual predictive deviation (standard deviation of validation samples/RMSEP)
iPLS Interval partial least squares (interval size of 50)
OSC Orthogonal signal correction (3 components, 10 iterations, 90% tolerance)
1st derivative (2nd order polynomial, 15 point window)
SG Savitzky Golay (2 order polynomial, 15 point window)
MSC Multiplicative scatter correction (Mean)

4.3.3.1 Glucose

The glucose model returning the lowest RMSECV was generated by the combination of *a priori* selection of the fingerprint-containing region and reverse iPLS (Table 4.2). Application of reverse iPLS involved the successive removal of intervals during local model analysis. Neither forward nor reverse iPLS on the full spectrum improved the model (Table S3), therefore *a priori* selection of the fingerprint-containing region was tested. It was evident from pure component spectra (Fig. 4.1) that this region (1500–900 cm^{-1}) included unique spectral banding patterns for all the sugars. This selection improved the average RMSEP of the model relative to reference by 9.9%. In combination with reverse iPLS the RMSEP was decreased from 5.5 to 4.3 g/l, which represented an improvement of 20.7% over the reference model RMSEP. Taken as a percentage of the average reference value for glucose samples, the improved RMSEP represents a 5% relative error. In comparison, using the data of Xie *et al.* (2009), a 16% relative error was calculated for glucose concentration in bayberry juice. Their study reported a similar calibration approach using real fruit juice samples, which allowed for a broad comparison to this study. The concentrations of glucose were low (<10 g/l) in their study so larger errors could be expected. Fig. 4.3 gives a regression overview of the model. The loadings plot (Fig. 4.3D) together with the SSQ (sum of squares) table (data not shown) indicated that the first LV captured 76.4% of the X-block variation and predicted 18.5% of the Y-block variation. The second LV used a further 19.9% of X to explain 60.9% of Y variation. The model used 6 LVs in total (Fig. 4.3A) to model the variation in the X- and Y-blocks. Addition of more LVs would have resulted in overfitting of the data. The captured variance statistics indicated that the first PLS factor modelled a combination of glucose correlated and uncorrelated information. The Hotellings T^2 versus Q residuals plot (Fig. 4.3C) identified samples that were unique (high T^2 values) and to a lesser degree some with relatively high Q residuals. This likely accounted for low proportion of Y-block variance captured in the first LV. The importance of wavenumbers below 1215 cm^{-1} for modelling glucose was highlighted in the loadings plot (Fig. 4.3D). These wavenumbers were in accordance with characteristic bands in the fingerprint region specific to the chemistry of glucose. Table 4.3 relates the results of the reverse iPLS modelling to the functional group assignments and the molecular vibrations of glucose. The alkane and ether bonds contributed to the spectral signature of glucose and the relevant wavenumber intervals were included in the PLS model. Max and Chapados (Max and Chapados 2007)reported similar spectral band assignments to glucose chemistry. They highlighted the C-O and C-C stretch vibrations in the 1000 cm^{-1} region.

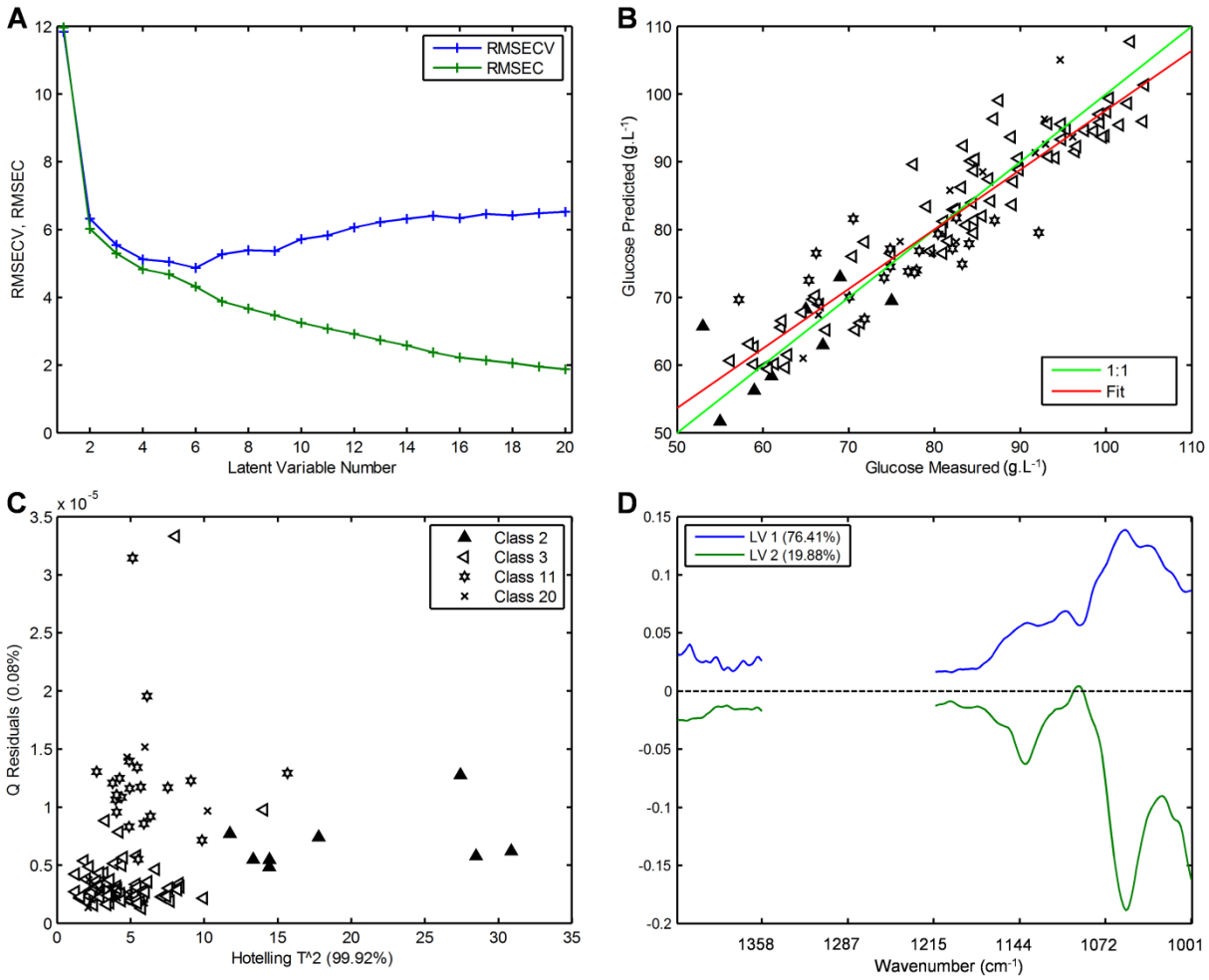

Fig. 4.3 Regression overview of the PLS calibration model for the determination of glucose in β-fructofuranosidase assay samples. A root mean square error of calibration (RMSEC) and cross validation (RMSECV) versus the number of latent variables (LVs), B predicted versus measured glucose values, C Hotelling T^2 versus Q residuals and D loadings plot for LVs 1 and 2

Table 4.3 Results of the interval selections from sugar calibration models and the band assignments relating to the chemistry of the respective sugars

Sugar	Intervals selected in best model (cm^{-1})	Functional group assignment*	Wavenumbers (cm^{-1})	Molecular vibrations **	Wavenumbers (cm^{-1})
Glucose	1428.9–1358.8	Alkane	~1585–1375	C-H bending	1470–1350
	1214.3–1001.2		~1225–900	C-C-H bending	1470–1200
		Ether	~1180–1100	C-O stretching	1200–900
Fructose	2930.8–2860.7	Alkane	~3130–2840	C-H stretching	3000–2800
	2787.7–2717.7				
	1500.4–1430.3	Alkane	~1585–1375	C-H bending	1470–1350
	1142.8–929.7		~1225–900	C-C-H bending	1470–1200
		Ether	~1180–1100	C-O stretching	1200–900
Sucrose	3789.0–3718.9	O-H free	~3700–3530	O-H stretching (free)	3650–3580
		O-H intermolecular	~3590–3200	O-H stretching (H-bonded)	3550–3200
	3145.3–3075.3	Alkane	~3140–2770	C-H stretching	3000–2800
	1858.0–1787.9	Not related to carbohydrates			
	1357.4–1144.2	Alkane	~1585–1375	C-H bending	1470–1350
	1071.3–1001.2		~1225–900	C-C-H bending	1470–1200
	928.3–901.1	Ether	~1180–1100	C-O stretching	1200–900
				C-O-H, C-C-H and O-C-H bending	950–750
1-Kestose	3996.4–3933.5	Not related to carbohydrates			
	3646.0–3361.3	O-H free	~3700–3530	O-H stretching (free)	3650–3580
	3288.4–3075.3	O-H intermolecular	~3590–3200	O-H stretching (H-bonded)	3550–3200
	3002.3–2503.1	Alkane	~3140–2770	C-H stretching	3000–2800
	2430.2–2288.5	Not related to carbohydrates			
	2215.6–2074.0				
	2001.0–1859.4				
	1786.5–1716.4				
	1500.4–1358.8	Alkane	~1585–1375	C-H bending	1470–1350
	1285.9–1215.8		~1225–900	C-C-H bending	1470–1200
	1071.3–901.1	Ether	~1180–1100	C-O stretching	1200–900
				C-O-H, C-C-H and O-C-H bending	950–750
Nystose	3996.4–3933.5	Not related to carbohydrates			
	3646.0–3575.9	O-H free	~3700–3530	O-H stretching (free)	3650–3580
	3502.9–3361.3	O-H intermolecular	~3590–3200	O-H stretching (H-bonded)	3550–3200
	3073.8–3003.7	Alkane	~3140–2770	C-H stretching	3000–2800
	2787.7–2717.7				
	2644.7–2217.0	Not related to carbohydrates			
	2144.1–2074.0				
	1858.0–1716.4				
	1571.9–1430.3	Alkane	~1585–1375	C-H bending	1470–1350
	1357.4–1287.3		~1225–900	C-C-H bending	1470–1200
	1214.3–901.1	Ether	~1180–1100	C-O stretching	1200–900
				C-O-H, C-C-H and O-C-H bending	950–750

*Software assigned
**Li-Chan et al. 2010

4.3.3.2 Fructose

The preprocessing combination of orthogonal signal correction (OSC) and forward iPLS returned the model with the largest reduction in RMSECV and its predictive ability was 48.2% improved over the reference model (Table 4.2). At 2 g/l the relative error was calculated to be 12% compared to 9% for the fructose data of Xie *et al.* (2009). The fructose concentrations were comparable to this study. As a supervised preprocessing method where the Y-block data is used as reference to identify uncorrelated variation in the X-block, OSC has been reported to overfit data (Trygg and Wold 2002). Examination of Fig. 4.4A indicated that the selected 5 LV model did not appear to overfit the data – the difference between RMSEC and RMSECV was small. The sample distribution was positively skewed, however log transformation of the Y block did not improve modelling statistics (data not shown). Most samples were generated from enzyme assays and the activities of the majority of enzymes were such that they did not produce much fructose. To expand the working concentration range, samples were generated by spiking with exogenous fructose (in random combination with the other four sugars). The first LV used 48.8% of X to explain 79.2% of Y. These statistics were indicative of a model where uncorrelated variation in the X-block was minimised. The loadings plot for the first two LVs highlighted the importance of wavenumbers from approximately 2900–2700 cm^{-1}, 1550–1450 cm^{-1} and 1200–900 cm^{-1}(Fig. 4.4D). These wavenumbers corresponded well with fructose chemistry as C–O and C–C stretch vibrations produce intense bands in the 1000 cm^{-1} region. C–C–H and C–O–H deformation modes produce low-intensity bands near 1400 cm^{-1} and C–H stretching produces bands in the 2700–3000 cm^{-1} (Max and Chapados 2007). Table 4.3 gives the precise wavenumber intervals where the contributions of the alkane and ether bonds to fructose chemistry were highlighted. In contrast to the one alkane group in the glucose molecule, fructose possesses two which likely explains the altered spectral bands in the 3000–2700 cm^{-1} and 1500–900 cm^{-1} regions. Inclusion of the 2930–2860 cm^{-1} region contributed positively to fructose model performance.

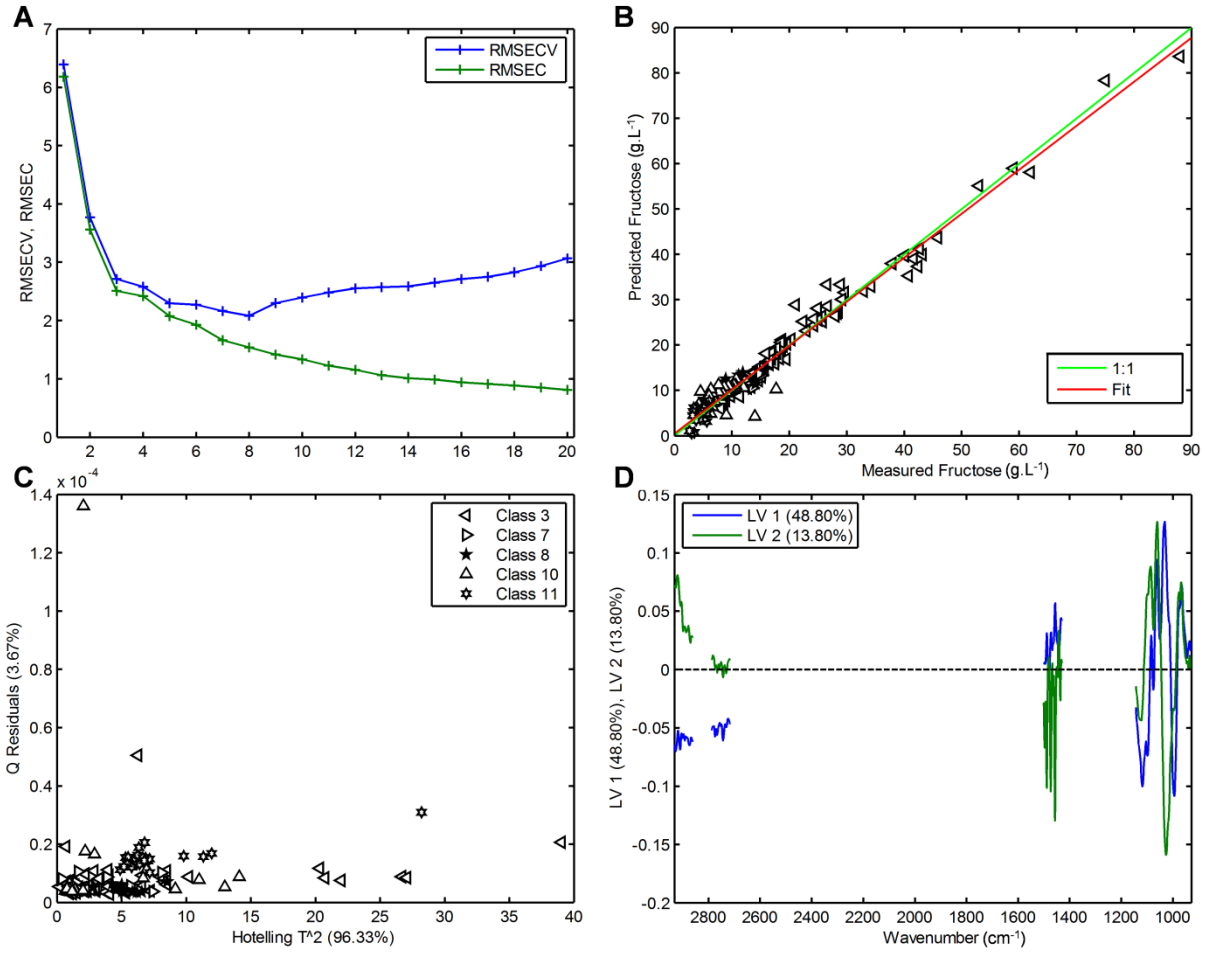

Fig. 4.4 Regression overview of the PLS calibration model for the determination of fructose in β-fructofuranosidase assay samples. A root mean square error of calibration (RMSEC) and cross validation (RMSECV) versus the number of latent variables (LVs), B predicted versus measured fructose values, C Hotelling T^2 versus Q residuals and D loadings plot for LVs 1 and 2

4.3.3.3 Sucrose

A 30.1% improvement in average RMSEP was obtained from the preprocessing combination of first derivative and smoothing with forward iPLS interval selection in the development of PLS models for sucrose prediction (Table 4.2). The reduction in RMSEP from 13.6 g/l to 9.5 g/l represented a significant improvement to the published sucrose model used in the screening of a β-fructofuranosidase variant library (Trollope et al. 2014). The 9.5 g/l average RMSEP was calculated to be 12% relative error in comparison to 13% for the sucrose data of Xie *et al.* (Xie et al. 2009). Similar tendencies pertaining to sample distribution and grouping of batches persisted, as discussed in section 3.2. The bimodal sample distribution (data not shown) could be attributed to enzyme properties used to generate samples. Highly active enzymes converted most of the sucrose and inactive enzymes did not act on the substrate, which produced samples with low and high sucrose concentrations, respectively. Fig. 4.5 gives the regression overview of the sucrose model.

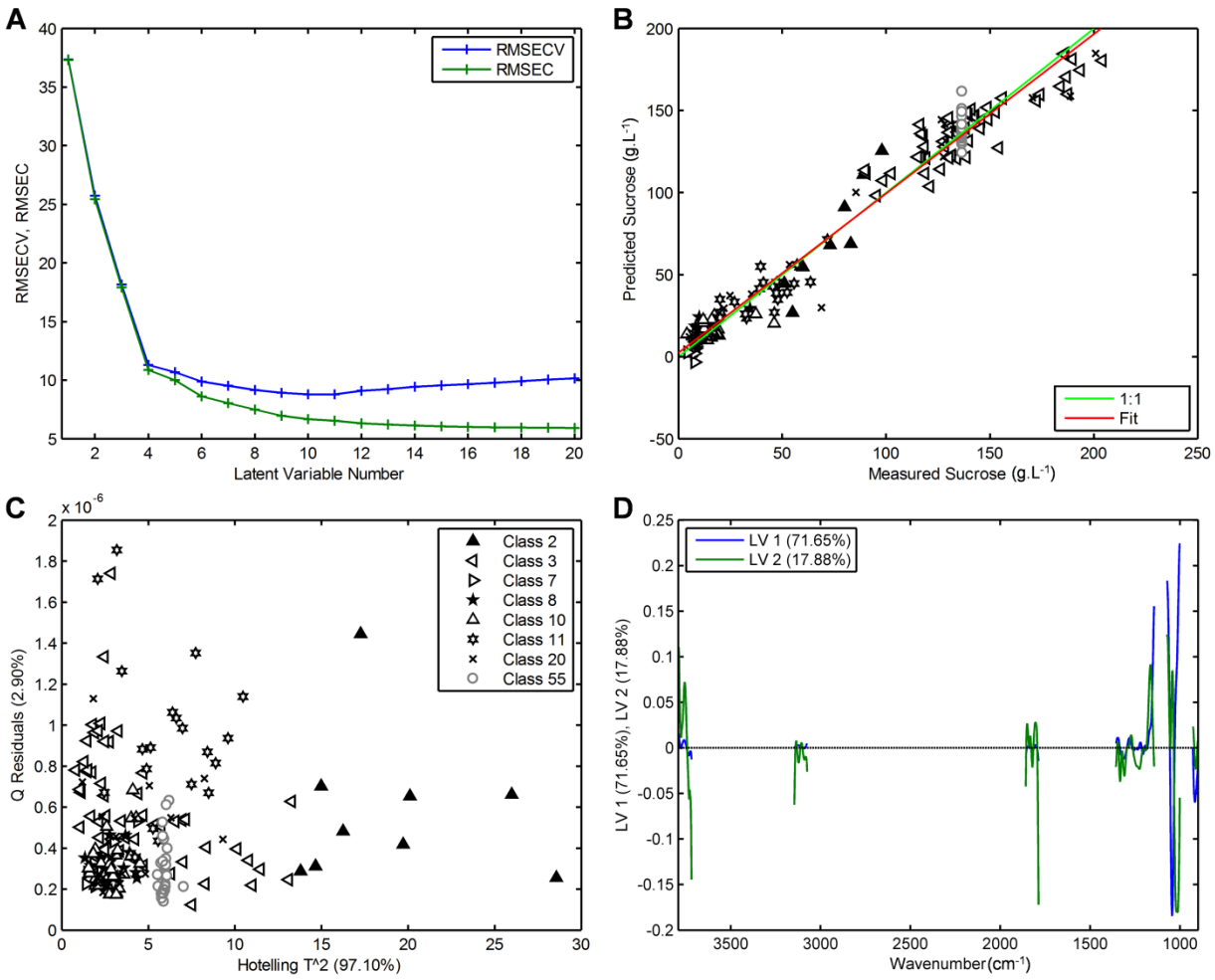

Fig. 4.5 Regression overview of the PLS calibration model for the determination of sucrose in β-fructofuranosidase assay samples. A root mean square error of calibration (RMSEC) and cross validation (RMSECV) versus the number of latent variables (LVs), B predicted versus measured sucrose values, C Hotelling T^2 versus Q residuals and D loadings plot for LVs 1 and 2

For the selected five LV model the loadings of the first two LVs, using 89.5% of X-block variation to describe 82.3% of Y-block variation, emphasised the importance of wavenumbers in a number of intervals (Fig. 4.5D). Table 4.3 gives the precise intervals used in sucrose modelling. The 1170–1145 cm^{-1} and 1070–1000 cm^{-1} intervals corresponded to the intense bands in the fingerprint region produced by C–O and C–C stretch vibrations (Max and Chapados 2001). As a disaccharide sucrose has similar types of bonds to its monomers however, with an additional ether bond (glycosidic bond) resulting from the condensation reaction between glucose and fructose. Table 4.3 shows that similar intervals were selected for sucrose modelling to those identified in glucose and fructose modelling. The sucrose model also included regions from 3789–3718 cm^{-1}, 3145–3075 cm^{-1} and 1858–1787 cm^{-1}. The first interval may correspond to free –OH bonds while the second interval may coincide with the intermolecular O-H or alkane band assignments as the limits of the bands are not absolute. The 1858–1787 cm^{-1} interval was unrelated to carbohydrate chemistry. This interval did not carry much weight in the model as indicated by the loadings plot - it did not feature in LV1.

4.3.3.4 1-Kestose

Numerous combinations of preprocessing and variable selection were tested but reverse iPLS on the full spectra (without preprocessing) provided the 1-kestose model with the second lowest RMSECV and the best predictive ability. It reflected a 23.5% improvement over raw, full spectra models (Table 4.2). Being the first report on the quantification of FOS using FT-MIR ATR spectroscopy, comparison of the 1-kestose and nystose model performances to published data was not possible. The first two LVs utilised 56.0% and 33.9% of X-block variation, respectively and together predicted 54.8% of Y-block variation. Fig. 4.6 gives the regression overview of 1-kestose modelling.

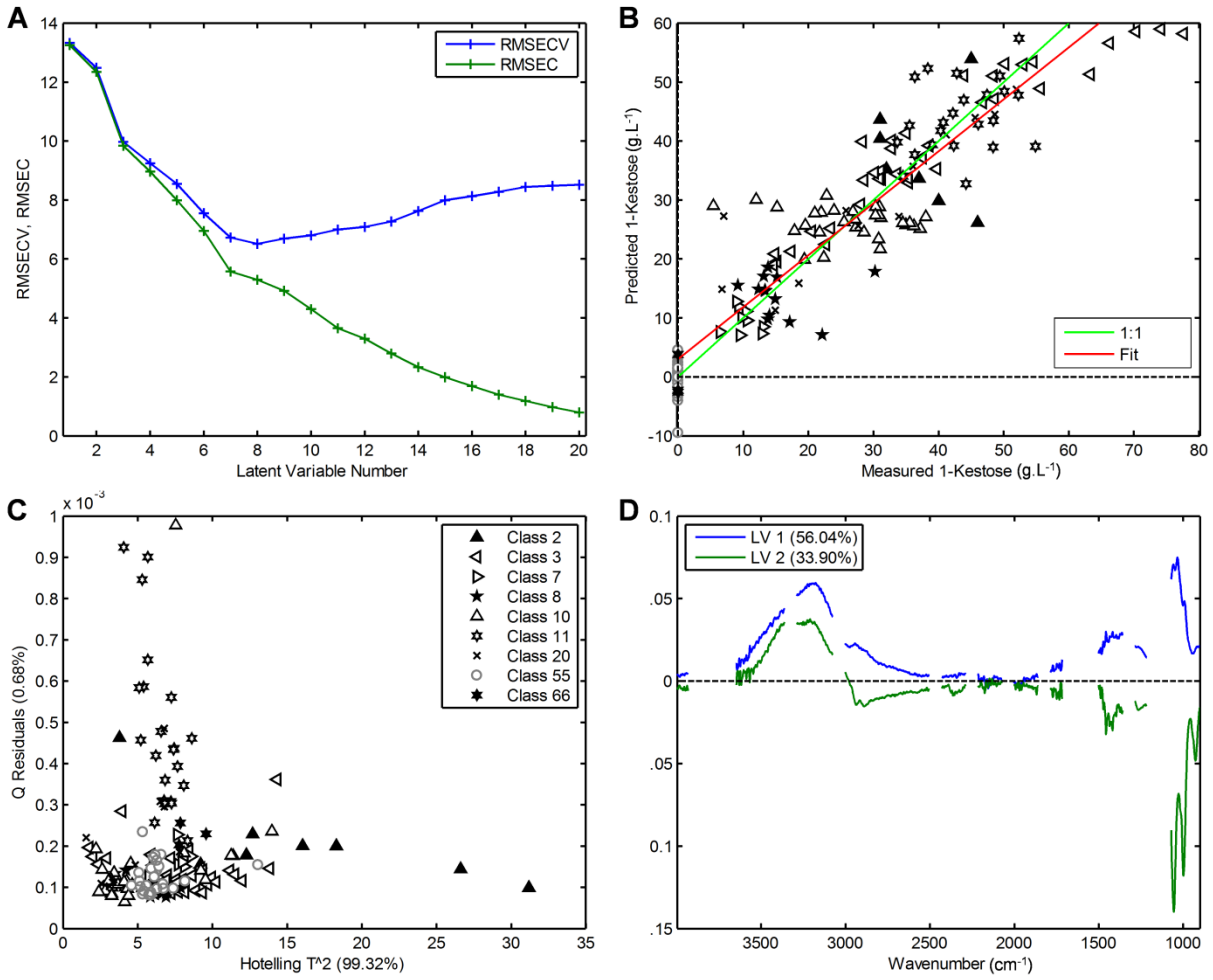

Fig. 4.6 Regression overview of the PLS calibration model for the determination of 1-kestose in β-fructofuranosidase assay samples. A root mean square error of calibration (RMSEC) and cross validation (RMSECV) versus the number of latent variables (LVs), B predicted versus measured 1-kestose values, C Hotelling T^2 versus Q residuals and D loadings plot for LVs 1 and 2.

As with the other sugars there were samples with high Hotelling T^2 scores, indicating unique samples which increased the complexity of modelling (Fig. 4.6C). Investigation of the loadings plot (Fig. 4.6D) indicated that these components modelled a combination of 1-kestose orthogonal and correlated variation – intervals between 2700 and 1500 cm^{-1} do not hold information pertaining to carbohydrates. The third LV used 6.59% of X variation to model a

further 16.49% of Y variation which appeared to hold more information pertaining to Y (data not shown). Similarly to the other sugars, the importance of the 3000–2700 cm^{-1} and 1000–900 cm^{-1} regions was evident. Table 4.3 gives a detailed breakdown of intervals used in the PLS model and the related chemistry. All of the intervals corresponding to O-H, alkane and ethers were used in the modelling. Although the same intervals were used, the differing number and positioning of the bonds in the 1-kestose molecule produced unique absorption bands (the case for all the sugars) and so their unique contributions to the models. Some intervals were included in the 1-kestose model that did not appear to be related to carbohydrate chemistry. Although not ideal, this can be expected in light of the investigations into orthogonal variation in the X block.

A common control used when working with genetically modified organisms is one where host organisms are transformed with a vector that that does not contain the gene of interest. It serves to confirm that the vector and transformation event did not produce the phenotype that is to be attributed to the gene of interest. In the modelling context, samples generated from control organisms were used in 1-kestose modelling as background/zero samples. The control yeast was unable to produce β-fructofuranosidase and hence samples generated from these supernatants did not contain any 1-kestose. Sample preparation was similar to other samples where the enzyme was present. To investigate their influence on the models, the zero samples (from different batches) were used only in the calibration set or omitted entirely. In both cases the average RMSEP increased. It was evident that the samples were important for modelling and thus were included in the final model. The purpose of their inclusion was not to predict such low concentrations due to the recognised sensitivity limitations of FT-MIR spectroscopy (Fig. 4.6C).

4.3.3.5 Nystose

A combination of MSC preprocessing and reverse iPLS provided the PLS model with one of the lowest RMSECVs and the lowest average RMSEP, representing a reduction of 21.7% in prediction error over the reference model (Table 4.2). Fig. 4.7 gives the regression overview of nystose modelling. The first LV used 87.9% of X-block variation to model only 10.3% of Y-block variation. This indicated a large amount of nystose uncorrelated variation modelled by this LV, despite the preprocessing applied. The spectra appeared to contain some noise (data not shown) but smoothing in combination with MSC and reverse iPLS did not improve model performance (Table S7).

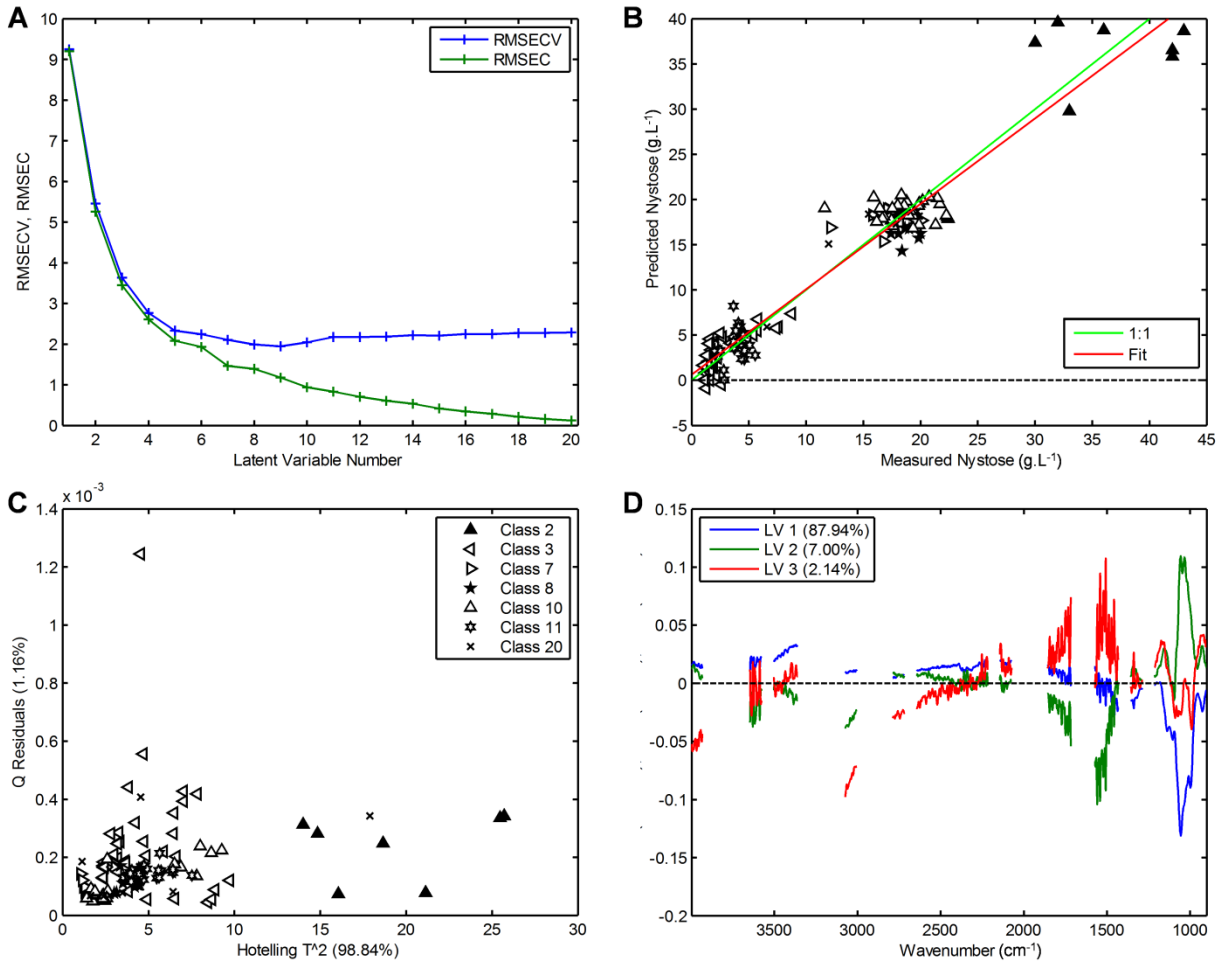

Fig. 4.7 Regression overview of the PLS calibration model for the determination of nystose in β-fructofuranosidase assay samples. A root mean square error of calibration (RMSEC) and cross validation (RMSECV) versus the number of latent variables (LVs), B predicted versus measured nystose values, C Hotelling T^2 versus Q residuals and D loadings plot for LVs 1, 2 and 3.

The loadings of wavenumbers below 1250 cm^{-1} for LV1 were high (Fig. 4.7D) which indicated that the band overlap from the other sugars was responsible for the large amount of Y-block uncorrelated variation modelled by this LV. Fig. 4.7B shows the measured vs predicted nystose with clearly uneven sample distribution. The three groupings represent inactive and active enzyme samples at lower concentrations with spiked samples having the highest nystose concentrations. As with fructose modelling, log transformation of the Y-block did not improve the models (data not shown). LVs 2 and 3 used 9.1% of X-block variation to account for 77.1% of Y-block variation. The loadings of these factors emphasised the importance of wavenumbers in the intervals 3073–3003 cm^{-1}, 1571–1430 cm^{-1} and 1214–901 cm^{-1} (Fig. 4.7D). These regions again corresponded to carbohydrate C–H stretching, C–C–H and C–O–H deformation modes and C–O and C–C stretch vibrations, respectively (Max and Chapados 2007; Max and Chapados 2001). Table 4.3 gives a detailed evaluation of the intervals selected for modelling in context of nystose chemistry. As for the other sugars the O-H, alkane and ether bonds featured as well as intervals not relating to carbohydrate chemistry. Removal of the intervals not relating to carbohydrate chemistry did not improve model performance (data not shown).

4.3.3.6 Investigation of the batch effect common to biotechnology experiments

An aspect receiving scant attention in the literature is the importance of sampling the variation represented by different sample batches used to develop calibration models. By way of example the influence of batch-wise cross validation on the 1-kestose model performance was examined. This sugar is heavily weighted in the direct library screening method presented in this work and factors shown to influence this model would by implication impact on the method as a whole. Different batches of samples represented separate yeast cultivation and assay events. The batch grouping of samples has been previously referred to in sections 3.3.2 and 3.3.3. The RMSECV for the reverse iPLS model cross validated using validation samples drawn from all batches (venetian blinds method with 10 data splits) was 6.7 g/l (Table 4.2Table 4.2). Using a random subsets cross validation method (10 data splits) returned a RMSECV of 7.0 g/l. Applying individual batches as cross validation sets (eight sets) resulted in an increased RMSECV of 13.4 g/l. Care was taken to select batches as cross validation sets that spanned similar concentration ranges to the batches used as calibration sets. Batch-wise cross validation for the other sugar models followed a similar trend of larger RMSECV values (data not shown). This result revealed the importance to calibration of sampling the variation represented by different sample batches. In conjunction with calibrating across the required concentration range to be predicted, this constitutes a sound approach to generating robust calibration models.

4.3.3.7 Additional multivariate tools for the identification of potential enzyme hits

Besides quantifying sugars, very useful information can be obtained from statistics generated during PLS modelling. Fig. 4.8 shows a panel of results for 1-kestose predictions from two independent sample batches not used in the calibrations. The top plot of Hotelling's T^2 versus Q residuals highlighted 10 samples with higher Hotelling's T^2 scores than the average samples. The four with the highest scores were highlighted (dark grey diamonds). Examination of the predicted 1-kestose plot (middle) indicated that these samples contained the highest levels of 1-kestose. The scores plot showed the grouping of the two sample batches and the highlighted samples appeared to be outliers, lying with the batch denoted by the squares. These plots clearly identified samples that differed from the average sample and hence provided a means to identify potential hits. When engineering enzymes, the outliers are of specific interest as the altered properties may well be the desired enzyme activity that is being screened for.

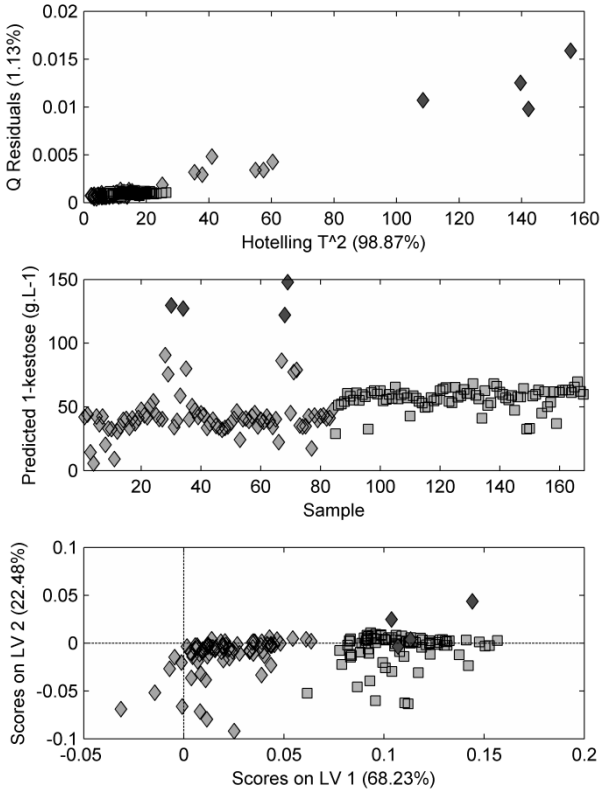

Fig. 4.8 Hotelling's T2 versus Q residuals (top), 1-kestose predictions (middle) and scores plot of first two latents variables (bottom) for new samples analysed by the optimised 1-kestose calibration model. Blocks and triangles denote different batches.

4.4 CONCLUSIONS

This study provided insight into the complexities of developing PLS calibration models for the prediction of structurally similar analyte concentrations in complex biological samples. Often, in real world scenarios all the theoretical requirements for generating ideal calibration models are not met. Instances of datasets with skewed sample distribution, analyte uncorrelated X-block variation and batch effects were discussed. Options of preprocessing and interval selection methods that can be applied to alleviate these modelling challenges were given. Interval selection served to improve modelling in all the datasets. The most likely explanation for this was that spectral intervals with high amounts of Y orthogonal variation (as investigated by O-PLS) were omitted from the models. This information could be used to optimise instrumental settings to acquire only useful data and minimise scanning times. It was found that matrix effects relating to different batches of samples affected modelling and to accurately predict future unknown samples, representatives of new batches should be included in the calibration model. Models presented were not the ultimate models and improving their predictive capacity remains an ongoing process.

PLS regression models were generated for the 5 sugars pertaining to the evaluation of β-fructofuranosidase activity. The reactions catalysed by this enzyme are complex as it

possesses a combination of hydrolytic and transferase activities where the products of the transferase activity also serve as substrates for hydrolytic activity. Hence, the ability to quantify all reaction products enables the comprehensive evaluation of enzyme activity. This capacity improves the quality of the assay that can be applied to screen large libraries of enzyme variants as it enables a direct screen for the sugars of interest, 1-kestose and nystose. The probability of identifying real enzyme hits, as opposed to false positives, is theoretically improved and would greatly facilitate the engineering of these proteins for desired attributes.

A tool is thus presented which meets many of the requirements for a high throughput screen – rapid, cost effective and direct quantification of multiple enzyme reaction substrates and products. Of topical relevance and hence a further advantage of the method, is the absence of environmentally hazardous waste generated during sample analysis. Automation of the screening process is possible with currently available equipment. In a high throughput format, this method would enable the use of random mutagenesis strategies for directed evolution of glycosyltransferase enzymes by facilitating the screening of large libraries and thus bypassing the requirement for a comprehensive understanding of enzyme structure-function relationships.

4.5 ACKNOWLEDGEMENTS

The authors would like to acknowledge Jamie Gallant for technical assistance with scanning FT-MIR ATR samples. Financial support for this work was provided by the Technology Innovation Agency (TIA), project PB110/08. The funders had no involvement in conducting the research or publication thereof. Kim Trollope was supported by a grant from TIA and Stellenbosch University. She also holds a L'Oreal-UNESCO For Women in Science Regional Fellowship in Sub-Saharan Africa.

4.6 LITERATURE CITED

De Abreu MA, Alvaro-Benito M, Plou FJ, Fernández-Lobato M, Alcalde M (2011) Screening β-fructofuranosidases mutant libraries to enhance the transglycosylation rates of β-(2→6) fructooligosaccharides. Comb Chem High Throughput Screen 14:730–738

Beine R, Moraru R, Nimtz M, Na'amnieh S, Pawlowski A, Buchholz K, Seibel J (2008) Synthesis of novel fructooligosaccharides by substrate and enzyme engineering. J Biotechnol 138:33–41

Blanch M, Goñi O, Sanchez-Ballesta MT, Escribano MI, Merodio C (2012) Characterisation and functionality of fructo-oligosaccharides affecting water status of strawberry fruit (*Fragraria vesca* cv. Mara de Bois) during postharvest storage. Food Chem 134:912–919

Bolton E, Wang Y, Thiessen P, Bryant S (2008) PubChem: Integrated Platform of Small Molecules and Biological Activities. In: Wheeler R, Spellmeyer D (eds) Annu. Rep. Comput. Chem. Elsevier, Oxford, pp 217–240

Brereton RG (2003) Chemometrics: Data analysis for the laboratory and chemical plant. John Wiley & Sons, Ltd., Chichester, UK

Buddington RK, Williams CH, Chen SC, Witherly SA (1996) Dietary supplement of neosugar alters the fecal flora and decreases activities of some reductive enzymes in human subjects. Am J Clin Nutr 63:709–716

Bureau S, Ruiz D, Reich M, Gouble B, Bertrand D, Audergon J-M, Renard CMGC (2009) Application of ATR-FTIR for a rapid and simultaneous determination of sugars and organic acids in apricot fruit. Food Chem 115:1133–1140

Cadet F, Pin FW, Rouch C, Robert C, Baret P (1995) Enzyme kinetics by mid-infrared spectroscopy: beta-fructosidase study by a one-step assay. Biochim Biophys Acta 1246:142–150

Corradini C, Lantano C, Cavazza A (2013) Innovative analytical tools to characterize prebiotic carbohydrates of functional food interest. Anal Bioanal Chem 405:4591–4605

Cozzolino D, Curtin C (2012) The use of attenuated total reflectance as tool to monitor the time course of fermentation in wild ferments. Food Control 26:241–246

Dominguez A, Nobre C, Rodrigues LR, Peres AM, Torres D, Rocha I, Lima N, Teixeira J (2012) New improved method for fructooligosaccharides production by *Aureobasidium pullulans*. Carbohydr Polym 89:1174–1179

Duarte IF, Barros A, Delgadillo I, Almeida C, Gil AM (2002) Application of FTIR spectroscopy for the quantification of sugars in mango juice as a function of ripening. J Agric Food Chem 50:3104–3111

Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. Anal Chim Acta 185:1–17

Gibson GR, Roberfroid MB (1995) Dietary Modulation of the Human Colonic Microbiota: Introducing the Concept of Prebiotics. J Nutr 125:1401–1412

Hidaka H, Hirayama M, Sumi N (1988) A fructooligosaccharide-producing enzyme from *Aspergillus niger* ATCC 20611. Agric Biol Chem 52:1181–1187

Jolliffe IT (2002) Principal Component Analysis, 2nd ed. Springer-Verlag, New York

Kuchner O, Arnold FH (1997) Directed evolution of enzyme catalysts. Trends Biotechnol 15:523–530

L'homme C, Peschet JL, Puigserver A, Biagini A (2001) Evaluation of fructans in various fresh and stewed fruits by high-performance anion-exchange chromatography with pulsed amperometric detection. J Chromatogr A 920:291–297

Li-Chan E, Chalmers JM, Griffiths P (2010) Applications of vibrational spectroscopy in food science: Volumes I and II: Analysis of food, drink and related materials. John Wiley & Sons, Chichester, UK

Maugeri F, Hernalsteens S (2007) Screening of yeast strains for transfructosylating activity. J Mol Catal B Enzym 49:43–49

Max J-J, Chapados C (2007) Glucose and fructose hydrates in aqueous solution by IR spectroscopy. J Phys Chem A 111:2679–2689

Max J-J, Chapados C (2001) Sucrose Hydrates in Aqueous Solution by IR Spectroscopy. J Phys Chem A 105:10681–10688

Naes T, Isaksson T, Fearn T, Davies T (2002) A User-friendly guide to multivariate calibration and classification. NIR Publications, Chichester, UK

Nemukula A, Mutanda T, Wilhelmi BS, Whiteley CG (2009) Response surface methodology: Synthesis of short chain fructooligosaccharides with a fructosyltransferase from *Aspergillus aculeatus*. Bioresour Technol 100:2040–2045

Norgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB (2000) Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. Appl Spectrosc 54:413–419

Otto M (1999) Chemometrics: statistics and computer application in analytical chemistry. Wiley-VCH, Chichester, UK

Riley MR, Arnold M a, Murhammer DW, Walls EL, DelaCruz N (1998) Adaptive calibration scheme for quantification of nutrients and byproducts in insect cell bioreactors by near-infrared spectroscopy. Biotechnol Prog 14:527–533

Rinnan Å, Berg F Van Den, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal Chem 28:1201–1222

Sangeetha PT, Ramesh MN, Prapulla SG (2005a) Recent trends in the microbial production, analysis and application of Fructooligosaccharides. Trends Food Sci Technol 16:442–457

Sangeetha PT, Ramesh MN, Prapulla SG (2005b) Maximization of fructooligosaccharide production by two stage continuous process and its scale up. J Food Eng 68:57–64

Schindler R, Le Thanh H, Lendl B, Kellner R (1998) Determination of enzyme kinetics and chemometric evaluation of reaction products by FTIR spectroscopy on the example of β-fructofuranosidase. Vib Spectrosc 16:127–135

Sharma S, Goodarzi M, Wynants L, Ramon H, Saeys W (2013) Efficient use of pure component and interferent spectra in multivariate calibration. Anal Chim Acta 778:15–23

Singh R, Singh R (2010) Production of fructooligosaccharides from inulin by endoinulinases and their prebiotic potential. Food Technol Biotechnol 48:435–450

Stenlund H, Gorzsás A, Persson P, Sundberg B, Trygg J (2008) Orthogonal projections to latent structures discriminant analysis modeling on in situ FT-IR spectral imaging of liver tissue for identifying sources of variability. Anal Chem 80:6898–6906

Stenlund H, Johansson E, Gottfries J, Trygg J (2009) Unlocking Interpretation in Near Infrared Multivariate Calibrations by Orthogonal Partial Least Squares. Anal Chem 81:203–209

Tapp HS, Kemsley EK (2009) Notes on the practical utility of OPLS. TrAC Trends Anal Chem 28:1322–1327

Thongnopkun P, Ekgasit S (2005) FTIR Spectra of faceted diamonds and diamond simulants. Diam Relat Mater 14:1592–1599

Trollope KM, Nieuwoudt HH, Görgens JF, Volschenk H (2014) Screening a random mutagenesis library of a fungal β-fructofuranosidase using FT-MIR ATR spectroscopy and multivariate analysis. Appl Microbiol Biotechnol 98:4063–4073

Trygg J, Wold S (2002) Orthogonal projections to latent structures (O-PLS). J Chemom 16:119–128

Vega R, Zúniga-Hansen ME (2011) Enzymatic synthesis of fructooligosaccharides with high 1-kestose concentrations using response surface methodology. Bioresour Technol 102:10180–10186

Wang J, Kliks MM, Jun S, Jackson M, Li QX (2010) Rapid analysis of glucose, fructose, sucrose, and maltose in honeys from different geographic regions using Fourier transform infrared spectroscopy and multivariate analysis. J Food Sci 75:C208–214

Waterhouse AL, Chatterton NJ (1993) Glossary of fructan terms. In: Suzuki M, Chatterton N (eds) Sci. Technol. Fruct. CRC Press, Boca Raton, FL, pp 2–7

Xie L, Ye X, Liu D, Ying Y (2009) Quantification of glucose, fructose and sucrose in bayberry juice by NIR and PLS. Food Chem 114:1135–1140

Yoshikawa J, Amachi S, Shinoyama H, Fujii T (2008) Production of fructooligosaccharides by crude enzyme preparations of beta-fructofuranosidase from *Aureobasidium pullulans*. Biotechnol Lett 30:535–539

4.7 APPENDIX: SUPPLEMENTARY INFORMATION

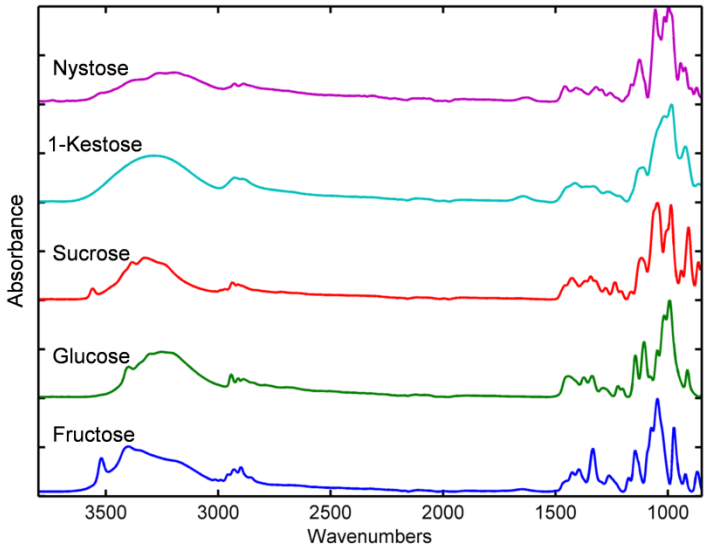

Fig. S1 Pure component powder FT-MIR ATR spectra for fructose, glucose, sucrose, 1-kestose and nystose. Spectra were collected from 3996 to 550 cm^{-1} with the 1500–900 cm^{-1} region showing the highest degree of variability between the sugars. The ATR diamond exhibits broad bands in the 2700–1800 cm^{-1} regions due to carbon bond absorption while hydrogen and nitrogen content is responsible for bands at 3100 cm^{-1} and 1500–1000 cm^{-1}, respectively (Thongnopkun and Ekgasit 2005). Stacked spectra are shown in the 3996–900 cm^{-1} region for all sugars.

Table S3 Glucose PLS models – combinations of preprocessing and interval selection methods

Glucose	Model			Calibration statistics			Validation statistics		Percentage improvement
							70% data split Average ± SD	50% data split	
	Wavenumbers used (cm^{-1})	Number of samples	Preprocessing/ Variable selection	Number of PLS factors	R^2	RMSEC ($g.L^{-1}$)	RMSECV ($g.L^{-1}$)	RMSEP ($g.L^{-1}$)	RMSEP ($g.L^{-1}$)	%
	3996.461 – 901.135	112	None	5	0.854	4.6	5.1	5.5 ± 0.2	6.19	0.0
	3996.461 – 901.135		MSC	5	0.861	4.5	5.0	5.0 ± 0.3	5.1	9.1
	3996.461 – 901.135		1st derivative	4	0.854	4.5	5.1	5.3 ± 0.9	5.8	2.3
	3996.461 – 901.135		1st derivative + smoothing (SG)	4	0.852	4.5	5.1	5.3 ± 1.5	5.2	3.1
	[a]		forward iPLS	5	0.840	4.7	5.3	6.2 ± 1.0	6.9	-13.2
	[b]		reverse iPLS	6	0.829	4.8	5.5	7.3 ± 1.1	6.6	-34.4
	[c]		MSC + iPLS forward	4	0.835	4.7	5.4	5.2 ± 0.6	5.6	3.7
	1600.587 – 901.135		Fingerprint	6	0.851	4.5	5.1	4.9 ± 1.0	5.6	9.9
	1600.587 – 901.135		MSC + fingerprint	6	0.868	4.3	4.8	4.9 ± 0.2	5.0	9.7
	[d]		Fingerprint + forward iPLS	6	0.865	4.4	4.9	4.7 ± 0.2	5.4	14.2
	[e]		Fingerprint + reverse iPLS	6	0.876	4.3	4.7	4.3 ± 0.4	5.0	20.8
	3996.461 – 901.135		OPLS	1+4	0.877	4.8	5.0	5.3 ± 0.4	5.0	2.7

Abbreviations:

SD Standard deviation (n = 3)
R^2 Coefficient of determination
RMSEC Root mean square error of calibration
RMSECV Root mean square error of cross validation
RMSEP Root mean square error of prediction
MSC Multiplicative scatter correction (Mean)
1st derivative (2nd order polynomial, 15 point window)
SG Savitzky Golay (2 order polynomial, 15 point window)
iPLS Interval partial least squares (interval size of 50)
OPLS Orthogonal partial least squares

iPLS wavenumber intervals:

[a] 3431.464:3361.376, 3288.427:3146.82, 2215.674:2145.559, 2001.091:1931.003, 1285.905:1144.298, 1071.349:1001.261, 928.312:901.135
[b] 3932.094:3504.43, 3431.464:2950.859, 2501.722:2431.634, 2358.685:2145.559, 2001.091:1931.003, 1858.05:1716.447, 1500.461:1358.854, 1214.386:1144.298, 1071.349:1001.261, 928.312:901.135
[c] 3431.464:3361.376, 3288.427:3218.339, 2215.647:2145.559, 2001.091:1931.003, 1858.054:1787.966, 1285.905:1144.298, 1071.349:1001.261, 928.312:901.135
[d] 1500.461:1358.854, 1214.386:1001.261
[e] 1428.942:1358.854, 1214.386:1001.261

Table S4 Fructose PLS models – combinations of preprocessing and interval selection methods

Fructose	Model				Calibration statistics			Validation statistics		Percentage improvement
								70% data split Average ± SD	50% data split	
	Wavenumbers used (cm^{-1})	Number of samples	Preprocessing/ Variable selection	Number of PLS factors	R^2	RMSEC (g/l)	RMSECV (g/l)	RMSEP (g/l)	RMSEP (g/l)	%
	3996.461 – 901.135	151	None	6	0.940	3.0	3.3	4.0 ± 0.1	3.7	0.0
	3996.461 – 901.135		MSC	6	0.951	2.7	3.0	3.4 ± 0.6	3.1	12.8
	3996.461 – 901.135		1st derivative	4	0.943	3.0	3.2	3.6 ± 0.1	3.8	9.1
	3996.461 – 901.135		1st derivative + smoothing (SG)	4	0.939	3.0	3.4	3.7 ± 0.6	3.9	5.7
	3996.461 – 901.135		OSC	4	0.942	2.7	3.3	2.8 ± 0.5	3.3	29.0
	[f]		forward iPLS	5	0.961	2.5	2.7	2.9 ± 0.2	3.2	26.8
	[g]		reverse iPLS	6	0.940	3.1	3.3	3.2 ± 0.4	4.3	17.5
	[h]		OSC + forward iPLS	5	0.973	2.1	2.5	2.0 ± 0.1	2.5	48.2
	1600.587 – 901.135		Fingerprint	5	0.963	2.6	2.7	2.2 ± 0.1	2.7	43.5
	1600.587 – 901.135		Fingerprint + OSC	4	0.969	2.3	2.5	2.9 ± 0.2	2.5	25.2
	3996.461 – 901.135		OPLS	1	0.950	3.1	3.3	3.3 ± 0.1	4.0	16.5

Abbreviations:

SD Standard deviation (n = 3)
R^2 Coefficient of determination
RMSEC Root mean square error of calibration
RMSECV Root mean square error of cross validation
RMSEP Root mean square error of prediction
MSC Multiplicative scatter correction (Mean)
1st derivative (2nd order polynomial, 15 point window)
SG Savitzky Golay (2 order polynomial, 15 point window)
OSC Orthogonal signal correction (3 components, 10 iterations, 90% tolerance)
iPLS Interval partial least squares
OPLS Orthogonal partial least squares

iPLS wavenumber intervals:

[f] 2859.315:2789.227, 2716.278:2646.189, 1500.461:1430.373, 1142.868:929.742
[g] 3996.461:3504.413, 3431.464:3289.857, 3216.908:2646.189, 2501.722:2217.078, 2144.129:2074.04, 2001.091:1573.41, 1428.942:901.135
[h] 2930.834:2860.745, 2787.796:2717.708, 1500.461:1430.373, 1142.868:929.742

Table S5 Sucrose PLS models – combinations of preprocessing and interval selection methods

Sucrose	Model				Calibration statistics			Validation statistics		Percentage improvement
								70% data split Average ± SD	50% data split	
	Wavenumbers used (cm^{-1})	Number of samples	Preprocessing/ Variable selection	Number of PLS factors	R^2	RMSEC (g/l)	RMSECV (g/l)	RMSEP (g/l)	RMSEP (g/l)	%
	3996.461 – 901.135	180	None	6	0.944	13.8	14.4	13.6 ± 0.9	15.3	0.0
	3996.461 – 901.135		MSC	6	0.943	13.8	14.6	15.6 ± 2.4	16.2	-14.8
	3996.461 – 901.135		1st derivative	5	0.962	10.4	11.8	11.4 ± 0.9	12.4	15.9
	3996.461 – 901.135		1st derivative + smoothing (SG)	5	0.963	10.5	11.8	12.0 ± 1.1	11.7	11.4
	3996.461 – 901.135		OSC	3	0.939	13.7	15.1	14.2 ± 1.8	15.4	-4.5
	[i]		forward iPLS	7	0.974	9.1	9.7	10.0 ± 0.3	10.67	26.0
	[j]		reverse iPLS	6	0.945	13.4	14.2	14.4 ± 0.3	15.28	-6.4
	[k]		1st derivative + forward iPLS	5	0.969	9.7	10.6	10.8 ± 0.7	9.2	20.8
	[l]		1st derivative + smoothing (SG) + forward iPLS	5	0.968	10.0	10.8	9.5 ± 0.8	10.9	30.1
	1600.587 – 901.135		Fingerprint only	7	0.976	8.8	9.5	9.9 ± 0.7	10.7	27.1
	[m]		forward iPLS + fingerprint	5	0.967	10.5	11.0	11.2 ± 0.6	12.0	17.6
	3996.461 – 901.135		OPLS	1	0.933	15.9	16.0	14.3 ± 3.5	14.3	-5.6

Abbreviations:

SD Standard deviation (n = 3)
R^2 Coefficient of determination
RMSEC Root mean square error of calibration
RMSECV Root mean square error of cross validation
RMSEP Root mean square error of prediction
MSC Multiplicative scatter correction (Mean)
1st derivative (2nd order polynomial, 15 point window)
SG Savitzky Golay (2 order polynomial, 15 point window)
OSC Orthogonal signal correction (3 components, 10 iterations, 90% tolerance)
iPLS Interval partial least squares (interval size of 50)
OPLS Orthogonal partial least squares

iPLS wavenumber intervals:

[i] 3502.983:3432.894, 3288.427:3218.339, 1786.535:1716.447, 1500.461:1430.373, 1214.386:1144.298, 1071.349:929.742
[j] 3996.461:3862.006, 3789.057:3361.376, 3288.427:3218.339, 3073.871:2860.745, 2787.796:2717.708, 2644.759:2574.671, 2430.203:2145.559, 2001.091:1287.335, 1214.386:929.742
[k] 3789.057:3718.969, 3145.389:3075.301, 2859.315:2789.227, 1929.573:1859.484, 1357.424:1144.298, 1071.349:1001.261, 928.312:901.135
[l] 3789.057:3718.969, 3145.389:3075.301, 1858.054:1787.966, 1357.424:1144.298, 1071.349:1001.261, 928.312:901.135
[m] 1500.461:1430.373, 1214.386:1144.298, 1071.349:929.742

Table S6 1-Kestose PLS models – combinations of preprocessing and interval selection methods

1-Kestose	Model			Calibration statistics		Validation statistics			Percentage improvement
						70% data split Average ± SD		50% data split	
	Wavenumbers used (cm^{-1})	Number of samples	Preprocessing/ Variable selection	Number of PLS factors	R^2	RMSEC (g/l)	RMSECV ($g.L^{-1}$)	RMSEP (g/l)	RMSEP (g/l)	%
	3996.461 – 901.135	159	None	6	0.819	7.2	7.8	8.4 ± 0.8	7.8	0.0
	3996.461 – 901.135		MSC	6	0.812	6.8	7.9	8.0 ± 0.1	7.5	5.3
	3996.461 – 901.135		MSC + Smoothing (SG)	7	0.839	6.2	7.4	7.5 ± 0.2	8.6	10.7
	3996.461 – 901.135		1st derivative	5	0.803	7.0	8.0	8.1 ± 1.3	8.5	3.1
	3996.461 – 901.135		1st derivative + smoothing (SG)	5	0.799	7.2	8.1	7.6 ± 0.1	8.3	10.1
	3996.461 – 901.135		OSC	4	0.843	6.5	7.3	7.6 ± 0.1	9.3	9.3
	[n]		forward iPLS	5	0.864	6.1	6.8	6.9 ± 0.5	8.5	18.2
	[o]		reverse iPLS	7	0.869	5.6	6.7	6.4 ± 0.4	7.5	23.5
	[p]		MSC + Smoothing (SG) + reverse iPLS	7	0.874	5.7	6.5	7.7 ± 0.8	7.8	8.4
	[q]		1st derivative + smoothing (SG) + reverse iPLS	6	0.867	5.6	6.7	8.2 ± 1.6	8.0	2.0
	[n]		reverse iPLS 0 in calibration only	7	0.869	5.6	6.7	8.0 ± 0.4	10.2	5.0
	[n]		reverse iPLS 0 excluded	5	0.689	7.5	8.4	8.2 ± 1.2	9.1	2.0
	1600.587 – 901.135		Fingerprint	7	0.811	6.8	8.0	8.7 ± 0.6	7.6	-3.0

1-Kestose	Model				Calibration statistics		Validation statistics			Percentage improvement
							70% data split Average ± SD		50% data split	
	Wavenumbers used (cm^{-1})	Number of samples	Preprocessing/ Variable selection	Number of PLS factors	R^2	RMSEC (g/l)	RMSECV ($g.L^{-1}$)	RMSEP (g/l)	RMSEP (g/l)	%
	[r]		Fingerprint + iPLS reverse	6	0.793	7.7	8.4	8.2 ± 0.2	8.4	2.4
	3996.461 – 901.135		OPLS	1	0.911	5.6	6.7	7.4 ± 0.4	9.1	11.8

Abbreviations:

SD Standard deviation (n = 3)
R^2 Coefficient of determination
RMSEC Root mean square error of calibration
RMSECV Root mean square error of cross validation
RMSEP Root mean square error of prediction
MSC Multiplicative scatter correction (Mean)
SG Savitzky Golay (2 order polynomial, 15 point window)
1st derivative (2nd order polynomial, 15 point window)
OSC Orthogonal signal correction (3 components, 10 iterations, 90% tolerance)
iPLS Interval partial least squares (interval size of 50)
OPLS Orthogonal partial least squares

iPLS wavenumber intervals:

[n] 3646.02:3575.932, 2215.647:2074.04, 1929.573:1859.484, 1786.535:1716.447, 1428.942:1358.854, 1071.349:1001.261
[o] 3996.461:3933.525, 3646.02:3361.376, 3288.427:3075.301, 3002.352:2503.152, 2430.203:2288.596, 2215.647:2074.04, 2001.091:1859.484, 1786.535:1716.447, 1500.461:1358.854, 1285.905:1215.817, 1071.349:901.135
[p] 3996.461:3862.006, 3717.539:3432.894, 3288.427:3146.82, 2930.834:2789.227, 2644.759:2503.152, 2215.647:2145.559, 1929.573:1501.891, 1428.942:1358.854, 1071.349:1001.261, 928.312:901.135
[q] 3996.461:3718.969, 3359.945:3289.857, 3216.908:3146.82, 3002.352:2932.264, 2787.796:2717.708, 2644.759:2431.634, 2358.685:2288.596, 2001.091:1931.003, 1643.498:1573.41, 1214.386:929.742
[r] 1571.98:1287.335, 1214.386:901.135

Table S7 Nystose PLS models – combinations of preprocessing and interval selection methods

Nystose	Model				Calibration statistics			Validation statistics		Percentage improvement
								70% data split Average ± SD	50% data split	
	Wavenumbers used (cm^{-1})	Number of samples	Preprocessing/ Variable selection	Number of PLS factors	R^2	RMSEC ($g.L^{-1}$)	RMSECV ($g.L^{-1}$)	RMSEP ($g.L^{-1}$)	RMSEP ($g.L^{-1}$)	%
	3996.461 – 901.135	129	None	4	0.911	2.6	2.9	2.9 ± 0.4	2.8	0.0
	3996.461 – 901.135		MSC	6	0.939	2.0	2.4	3.2 ± 0.3	3.6	-11.3
	3996.461 – 901.135		1st derivative	4	0.899	2.5	3.1	4.5 ± 0.2	3.3	-55.1
	3996.461 – 901.135		1st derivative + smoothing	4	0.902	2.5	3.0	3.7 ± 0.6	4.1	-25.1
	3996.461 – 901.135		MSC + smoothing	6	0.935	2.1	2.5	2.7 ± 0.2	2.7	7.6
	[s]		iPLS forward	4	0.940	2.2	2.4	2.6 ± 0.4	2.2	11.6
	[t]		iPLS reverse	6	0.945	2.0	2.3	2.7 ± 0.3	2.7	7.8
	[u]		MSC + forward iPLS	6	0.912	2.4	2.9	3.6 ± 0.2	3.9	-23.8
	[v]		MSC + reverse iPLS	5	0.937	2.1	2.4	2.3 ± 0.4	3.0	21.7
	[w]		MSC + smoothing + reverse iPLS	5	0.923	2.2	2.4	2.3 ± 0.2	2.4	20.2
	1600.587 – 901.135		fingerprint	6	0.941	2.1	2.4	2.7 ± 0.1	2.7	7.1
	[x]		fingerprint + reverse iPLS	5	0.925	2.5	2.7	2.6 ± 0.2	3.3	11.3
	[y]		fingerprint + MSC + reverse iPLS	4	0.925	2.5	2.7	2.6 ± 0.2	2.5	10.3
	3996.461 – 901.135		OPLS	1	0.957	2.1	2.4	3.0 ± 0.1	3.2	-2.7

Abbreviations:

SD Standard deviation (n = 3)
R^2 Coefficient of determination
RMSEC Root mean square error of calibration
RMSECV Root mean square error of cross validation
RMSEP Root mean square error of prediction
MSC Multiplicative scatter correction (Mean)
1st derivative (2nd order polynomial, 15 point window)
SG Savitzky Golay (2 order polynomial, 15 point window)
iPLS Interval partial least squares (interval size of 50)
OPLS Orthogonal partial least squares

iPLS wavenumber intervals:

[s] 2859.315:2717.708, 2644.759:2574.671, 2430.203:2360.115, 1500.461:1430.373, 1214.386:1001.261
[t] 3996.461:3862.006, 3789.057:2717.708, 2144.129:2002.522, 1929.573:1644.929, 1571.98:1358.854, 1285.905:1001.261
[u] 3717.539:3647.45, 3073.871:3003.783, 1643.498:1501.891, 1357.424:1287.335, 1214.386:1072.78, 999.831:929.742
[v] 3996.461:3933.525, 3646.02:3575.932, 3502.983:3361.376, 3073.871:3003.783, 2787.796:2717.708, 2644.759:2217.078, 2144.129:2074.04, 1858.054:1716.447, 1571.98:1430.373, 1357.424:1287.335, 1214.386:901.135
[w] 3996.461:933.525, 3789.057:3718.969, 3431.464:3361.376, 3288.427:2789.227, 2716.278:2431.634, 2144.129:2074.04, 1786.535:1716.447, 1571.98:1430.373, 1357.424:901.135
[x] 1600.587:1430.373, 1357.424:901.135
[y] 1600.587:1573.41, 1500.461:1001.261, 928.312:901.135

CHAPTER 5

RESEARCH RESULTS III

Directed evolution of loop regions in a fungal β-fructofuranosidase for improved fructooligosaccharide production

5 RESEARCH RESULTS III:

Directed evolution of loop regions in a fungal β-fructofuranosidase for improved fructooligosaccharide production

K.M. Trollope[1], J.F. Görgens[2], H. Volschenk[1]

[1]Department of Microbiology, Stellenbosch University, Private Bag X1, Stellenbosch, 7602, South Africa
[2]Department of Process Engineering, Private Bag X1, Stellenbosch University, Stellenbosch, 7602, South Africa

ABSTRACT

The *Aspergillus japonicus* β-fructofuranosidase is used in the industrially-important biotransformation of sucrose to fructooligosaccharides. Operating at high substrate loading and temperatures between 50-60 °C, the enzyme activity is negatively influenced by glucose product inhibition and thermal instability. To address these limitations, the solvent exposed loop regions of the β-fructofuranosidase were engineered using a combined crystal structure- and evolutionary-guided approach. This semi-rational approach yielded a functionally enriched first round library of 36 single amino acid substitution variants with 58% retaining activity. Of the active variants 71% displayed improved activity relative to the parent. The substitutions yielding the five most improved first round variants were subsequently exhaustively combined and evaluated. A four amino acid substitution variant was identified as the most improved and reduced the time to completion of a hitherto efficient industrial-like reaction by 26%. Purification of the five most improved variant enzymes and characterisation by isothermal denaturation assays indicated that all variants displayed improved thermostability relative to the parent, with the most thermostable variant displaying a 5.7 °C increased melting temperature. In addition, the variants displayed uniquely altered concentration-dependent substrate and product binding as determined by differential scanning fluorimetry. The altered catalytic activity was evidenced by increased specific activities of all five variants with the most improved variant doubling that of the parent. Variant homology modelling and calculated solvent accessible surface area changes for residues affected by the introduced substitutions were used to rationalise the effects of amino acid changes lacking direct interaction with substrates. This work again demonstrates that alterations to non-catalytic residues distant to the active site pocket can mediate long range effects that significantly improve enzyme activity.

5.1 INTRODUCTION

β-fructofuranosidases are family 32 glycoside hydrolase (GH32) enzymes that act on sucrose and related β-D-fructofuranosides (Cantarel et al. 2009). Also known as invertases (EC 3.2.1.26), they hydrolyse sucrose to produce invert sugar – an equimolar mixture of dextrorotatory D-glucose and levorotatory D-fructose (O'Sullivan and Tompson 1890). Available crystal structures for GH32 β-fructofuranosidases reveal that the enzymes display a bimodular arrangement of a N-terminal catalytic domain containing a five-bladed β-propeller fold linked to a C-terminal β-sandwich domain (Alberto et al. 2004; Alvaro-Benito et al. 2010b; Chuankhayan et al. 2010; Sainz-Polo et al. 2013). β-fructofuranosidases hydrolyse β-glycosidic bonds by a double displacement catalytic mechanism that retains the configuration of the fructose anomeric carbon (Koshland and Stein 1954). Multiple sequence alignments (MSAs) identified a highly conserved aspartate close to the N terminus that serves as the catalytic nucleophile and a glutamate residue that acts as a general acid/base catalyst (Pons et al. 2004). It is well known that some β-fructofuranosidases possess fructosyltransferase activity whereby the sugar moiety is transferred from the enzyme-fructosyl intermediate to a substrate other than water (Chuankhayan et al. 2010; Edelman 1956). This reaction forms the basis of fructooligosaccharide (FOS) synthesis from sucrose. Enzymes from *Aspergillus* spp. (Cruz et al. 1998; Hidaka et al. 1988; Zuccaro et al. 2008) and *Aureobasidium pullulans* (Yoshikawa et al. 2006) exhibit good propensities for the synthesis of inulin type FOS, with β-(2→1) linkages between fructose units. Synthesis of FOS (GF_n) from sucrose (GF) occurs via a disproportionation reaction with the reaction generalised as $GF_n + GF_n \rightarrow GF_{n-1} + GF_{n+1}$ (Jung et al. 1989; Nishizawa et al. 2001). In a batch reaction the initial products are glucose and 1-kestose (GF_2) and as the reaction progresses nystose (GF_3) and β-fructofuranosylnystose (GF_4) levels increase. Reaction conditions influence the dominance of hydrolytic or transferase reactions with high substrate concentrations favouring the latter (Hidaka et al. 1988).

The global demand for FOS is growing due to human health benefits associated with their consumption. FOS are prebiotics that selectively stimulate the growth of bifidobacteria, thereby promoting colonic health (Hidaka et al. 1986; Roberfroid 2007). Further claims as to the effect of FOS consumption relate to mineral absorption, lipid metabolism and the control of type II diabetes and have been extensively reviewed (Roberfroid 2007; Sangeetha et al. 2005; Singh and Singh 2010). Further to their health benefits, FOS are used in the food industry as low calorie sweeteners. They are also added to food products to improve their organoleptic properties and their inclusion allows producers to label their products as 'functional foods' – a claim that resonates with health conscious consumers (Roberfroid 2007; Sangeetha et al. 2005). Industrial biotransformation of sucrose to FOS is conducted in batch mode using the

β-fructofuranosidase from *A. niger* ATCC 20611. The enzyme is added to a 50–60% sucrose solution and the reaction proceeds at 50 – 60 °C for up to 20 h. These severe industrial conditions certainly impose limitations on activity. The fructosyltransferase activity of the enzyme has been shown to be non-competitively inhibited by the glucose product, limiting complete sucrose conversion (Nishizawa et al. 2001).

To improve competitiveness of enzymatic biotransformation processes, enzymes are usually engineered for enhanced performance under industrial conditions. Engineering transferases is mostly limited to rational approaches owing mainly to the absence of suitable high throughput screening methods, although progress in developing screens has been reported (Aharoni et al. 2006; de Abreu et al. 2011; Trollope et al. 2013). Understanding of the protein sequence motifs or structural features that impart high fructosyltransferase activity to fungal β-fructofuranosidases is incomplete and does not allow for the reliable designation of mutations to improve activity. Improvements to fructosyltransferase activity of fungal β-fructofuranosidases have been achieved by altering amino acids in the active site pocket (Alvaro-Benito et al. 2010a; Lafraya et al. 2011) and in the non-catalytic β-sandwich domain (de Abreu et al. 2011). Application of a semi-rational approach to engineer a β-fructofuranosidase would be feasible as it could combine the advantages of rational and random protein engineering strategies. Notably, it would meet the screening constraints imposed by HPLC quantification of FOS (Corradini et al. 2013; L'homme et al. 2001; Sangeetha et al. 2005).

Semi-rational protein engineering strategies employ MSAs of homologous proteins to guide the selection of mutations. MSAs provide a view of extant sequence variation and hence amino acids with proven fitness in the structures of homologous enzymes. MSAs thus provide means for selecting positions and types of amino acid substitutions with the objective of generating small, high quality libraries with increased functional information (Cole and Gaucher 2011). These approaches assume that amino acids residing in highly conserved regions are critical to protein function and alteration of these amino acids would prove deleterious (Miller and Kumar 2001). Subsequently, substitutions are most often targeted to regions of higher sequence diversity with different approaches utilising diverse criteria for the selection of these substitutions. Approaches that utilise protein family sequence diversity for mutation selection include the consensus sequence method, ancestral sequence resurrection, ancestral mutation method and reconstructing evolutionary adaptive paths as reviewed by Cole and Gaucher (2011). Furthermore, knowledge of proteins' phenotypic attributes can complement sequence-based mutation selection methods. For example, comparison of enzyme sequences from thermophilic and mesophilic organisms allowed the identification of residues from MSAs for substitution in order to identify stabilising residues (Haney et al. 1999; Perl et al. 2000).

The availability of crystal structures considerably facilitates the design of enzyme engineering strategies. Information on secondary, tertiary and quaternary structure is valuable to rationally reduce the number of substitutions identified from MSAs. Mutations in loops are generally better tolerated than substitutions that occur in secondary structures which may interfere with the specific geometries associated with helices and sheets (Chothia and Finkelstein 1990; Wintrode and Arnold 2000). Engineering loops has been shown to be an effective strategy for enhancing the activities of a lipase in DMSO and the thermostability of a neutral protease (Yedavalli & Rao 2013; Hardy et al. 1994). Furthermore, crystal structures provide opportunity to obtain accurate solvent accessible surface areas (SASAs) of the folded state from atomic coordinates (Kabsch and Sander 1983; Lee and Richards 1971; Shrake and Rupley 1973). In so doing, a quantitative representation of residue exposure can be determined. Sequence-based methods for the prediction of SASAs have also been developed in recent years (Ahmad et al. 2003a; Ahmad et al. 2003b; Joo et al. 2012). As loops often occur on protein surfaces the residues generally have high SASA values. That SASAs and evolutionary rates of sites are highly correlated (Franzosa and Xia 2009; Ramsey et al. 2011) further supports the notion of mutation tolerance in loop regions. Residue SASAs are thus useful criteria with which to select mutation positions.

Here the semi-rational directed evolution of the *A. niger* β-fructofuranosidase using a dual strategy is described. Only loop regions were selected for engineering using a crystal structure-guided approach, with amino acid substitutions selected based on scoring positions in homologous protein sequence alignments using sequence entropies and solvent accessibilities. The 36 variant library expressed in *Saccharomyces cerevisiae* was screened for variants that produced higher levels of FOS than the parent (wild type) enzyme. Enzymes were further screened for those producing more FOS than the parent under glucose inhibiting conditions in an attempt to identify variants relieved from product inhibition. Hits from the first round library were exhaustively combined and combination variants were screened for improved activity. A combination variant with four amino acid substitutions was identified that displayed a combination of improved thermostability and catalytic activity. It allowed for a 26% reduction in reaction time to produce FOS of a desired composition under industrial production conditions.

5.2 MATERIALS AND METHODS

5.2.1 Microbial strains and media

S. cerevisiae Y02321 [BY4741;*Mat a; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YIL162w(SUC2)::kanMX4*] obtained from the European *S. cerevisiae* Archive for Functional Analysis (EUROSCARF) served as host for the variant libraries (Brachmann et al. 1998). The

S. cerevisiae NI-C-D4 [*Mat*α; *trp1; ura3; pep4*] oversecretion phenotype strain was used when heterologous proteins were to be purified (Wang et al. 2001). *Escherichia coli* DH5α [*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*] (New England Biolabs, Midrand, South Africa) was used for cloning and amplification of plasmids. *E. coli* cells were grown at 37 °C in Luria Bertani broth supplemented with 100 µg/ml ampicillin or 50 µg/ml kanamycin, as appropriate.

5.2.2 DNA manipulations

All DNA manipulations were performed according to standard methods (Sambrook et al. 1989). Restriction enzymes and T4 DNA ligase were used according the specifications of the supplier (ThermoScientific, Waltham, Massachusetts, USA). CLC Main Workbench version 6.8.1 (Qaigen) was used for sequence analyses.

5.2.3 Gene synthesis and mutagenesis

The wild type (parent) gene was provided by DNA 2.0 (Menlo Park, CA, USA) as a synthetic construct combining the *Trichoderma reesei* endoxylanase 2 (*xln2*) secretion signal (Saarelainen et al. 1993) and the open reading frame of the *fopA* β-fructofuranosidase (GenBank accession number AB046383). The native *fopA* secretion signal was excluded. In the first round, 36 variants of the *fopA* β-fructofuranosidase gene were synthesised harbouring single amino acid substitutions. The substitutions were distributed throughout the protein sequence but were limited to loop regions as determined from the published crystal structure (3LF7) of Chuankhayan et al. (2010). The published crystal structures were determined for the *A. japonicus* β-fructofuranosidase. There is 99% homology on the DNA (AB046383, GU356596.1) and protein (BAB67771.1, ADK46938.1) levels of the β-fructofuranosidases of *A. niger* and *A. japonicus*, respectively. *Aspergillus niger* ATCC 20611 was reclassified by the curators of the ATCC culture collection (http://www.lgcpromochem-atcc. com) as *A. japonicus* and it was therefore assumed that the sequences and structures deposited in the databases are for the same gene/enzyme.

Positions for amino acid substitutions were selected by the strategy provided by DNA2.0. An alignment of homologous sequences to the 3LF7 structure was used from the homology-derived structures of proteins HSSP database (Sander and Schneider 1994). Positions within the MSA corresponding to secondary structural elements of the 3LF7 chain A were excluded. The solvent accessibility computed for each sequence position was normalised according to values obtained for Ala-X-Ala tripeptides (Shrake and Rupley 1973; Zielenkiewicz and Saenger 1992).

This provided a relative solvent accessibility at each position normalized by the side chain type. The sequence entropy for each position in the multiple sequence alignment was also obtained from the HSSP file for the Research Collaboratory for Structural Bioinformatics RCSB Protein Data Bank entry 3LF7. Sequence positions with relative solvent accessibility (RSA) greater than 50% and sequence entropy (SE) greater than 1.0 provided a list of positions in the structure-based multiple sequence alignment. From these alignment data, the most commonly observed substitutions were selected for inclusion in a 36 variant first round library. The mutations are listed in Table 5.1. A further 18 genes were synthesised containing exhaustive combinations of 5 single mutations that were determined to improve enzyme performance during first round screening (Table 5.2).

5.2.4 DNA cloning and yeast library generation

Cloning vectors (pJ227) containing the synthesised gene variants were digested with *Eco*RI and X*ho*I restriction enzymes (Fermentas) and directly ligated with the pJC1 yeast expression vector (Crous et al. 1995) digested with the same restriction enzymes. As there were no DNA clean-up steps, a mixture of plasmid backbones was used in the transformation of *E. coli*. Selection of *E. coli* transformants on LB agar plates supplemented with 100 μg/ml ampicillin ensured isolation of clones with the gene variant–pJC1 combination (due to the *BLA* marker on the pJC1 plasmid) and not re-circularised gene variant–pJ227, as the marker on the pJ227 cloning vector conferred resistance to kanamycin. The primer pair 5'GTTTAGTAGAACCTCGTGAAACTTA 3' and 5'ACTTAAAATACGCTGAACCCGAACAT3' was used to screen clones by polymerase chain reaction to ensure the presence of the 2000 base pair insert. Positive clones were further confirmed by restriction digest analysis. Yeast was transformed by the lithium acetate method described by Hill et al. (1991). The method was adapted to 96-well format by proportionally scaling down reagents.

5.2.5 Yeast cultivation and media

S. cerevisiae was cultivated at 30 °C in YPD (1 % yeast extract, 2 % peptone and 2 % glucose) or in synthetic medium, SC without uracil [2% carbon source, 0.67% yeast nitrogen base without amino acids (with ammonium sulphate; Difco Laboratories, Detroit, Michigan, USA) and 0.13 % amino acid dropout pool (Sherman et al. 1979)]. Glucose and galactose served as carbon source in solid and liquid SC^{-ura} media, respectively. Solid media contained 2% agar (Difco Laboratories).

Three yeast transformants per variant were manually transferred to individual wells of 2 ml round bottomed 96-deep-well microplates (Merck, Modderfontein, South Africa) containing 1.25 ml SC^{-ura} media. Mixing was facilitated by a single 2 mm glass bead (Merck, Modderfontein, South Africa) added to each well. Microplates were sealed with sterile, breathable AeraSealTM film (Excel Scientific Inc., Victorville, California, USA) and shaken at 200 rpm for 4 days. Fifty microliters of each culture were transferred to a fresh microplate and cultivated for a further 4 days. Master plates were generated on solid media using a 96-well replicator (Applikon Biotechnology, Delft, Netherlands). Culture supernatants were used in assays as source of enzyme after cell removal by centrifugation at 3000 rpm.

5.2.6 Library screening

Enzyme activity assays were performed in 96-well format by reacting 50 µl of culture supernatant with 50 µl of substrate at 55 °C for 2 h. The working concentration of substrate was 200 g/l sucrose (Fluka, Sigma-Aldrich, St. Louis, Missouri, USA) in 50 mM citrate phosphate buffer, pH 5.5. As determined previously (Trollope et al. 2014), 54 g/l glucose was added to the substrate solution to test for variants insensitive to product inhibition. Glucose, fructose, sucrose, 1-kestose, nystose and β-fructofuranosylnystose in assay samples were quantified using high performance liquid chromatography while Fourier transform mid-infrared attenuated total reflectance (FT-MIR ATR) spectroscopy quantified only sucrose. The details of the methods were described previously (Trollope et al. 2014). The rationale for supplying galactose as carbon source in liquid cultures was the separation of glucose and galactose by HPLC and hence any glucose present in assay samples was attributed to enzyme activity. Inhibition was calculated as the difference between uninhibited variant activity and inhibited activity divided by uninhibited activity. Data were normalised to the parental activity. The parent and all variants were cultivated and assayed with triplicate repeats. The cultivation and assay procedure was validated previously (Trollope et al. 2014).

5.2.7 Protein Purification

The top five performing variants were purified using immobilised metal affinity chromatography (IMAC). The proteins were N-terminal His-tagged by sub-cloning *Bgl*II-*Xho*I gene fragments into the same sites in a modified pJC1 yeast expression vector. The vector was modified by cloning a synthetic fragment (Geneart, Regensburg, Germany) encoding the *xyn2* secretion signal, six histidine residues and a factor Xa protease cleavage site into the *Eco*RI and *Bgl*II sites of the multiple cloning site. The synthetic DNA sequence read: gaattcGCGGCCGCATGGTTTCTTTCACATCCTTGTTGGCTGGTGTTGCTGCTATTTCCGGTG

TTTTGGCTGCTCCAGCTGCTGAAGTTGAATCCGTTGCTGTTGAGAAGAGACATCACCATCA CCATCACGGATCCGGCTCTGGATCTGGTATCGAGGGAagatct. Tagged gene variants were sequenced to verify integrity of the clones. Plasmids were transformed to *S. cerevisiae* NI-C-D4. Transformants were cultivated for 72 h in 50 ml double strength SC^{-ura} buffered with succinic acid at pH 6 [(2% glucose, 1.34% yeast nitrogen base without amino acids (with ammonium sulphate; Difco Laboratories, Detroit, Michigan, USA) and 0.26% amino acid dropout pool (Sherman et al. 1979)]. Antifoam 204 (Sigma-Aldrich, St. Louis, Missouri, USA) was added after 48 h of cultivation to a concentration of 0.025% (v/v). Following cell removal by centrifugation (3000 rpm), supernatants were concentrated 50 times by ultrafiltration using Amicon ultra-15 centrifugal filters with 10 kDA MWCO (Millipore, Molsheim, France). IMAC protein purification was performed under native conditions using Ni-NTA spin columns supplied by Qaigen (Venlo, Netherlands). For the removal of imidazole, buffer exchange with 10 mM Bis-Tris, pH 6.0 was performed using the aforementioned ultrafiltration devices. Protein concentration was determined using the bicinchoninic acid assay (Pierce Chemical Company, Rockford, Illinois, USA) with bovine serum albumin as standard.

5.2.8 Enzyme assays

A unit of enzyme was defined as the amount of protein that produced 1 mmol 1-kestose per minute from 10% (w/v) sucrose at 40 °C in 50 mM citrate phosphate buffer (pH 5.5). The definition approximates that of Hirayama et al. (1989).

5.2.9 Protein Electrophoresis

Samples were analysed by SDS-PAGE on an 8% resolving gel. Loading dye consisted of 60 mM Tris-HCl (pH 6.8), 25% glycerol, 2% SDS, 14 mM β-mercaptoethanol and bromophenol blue and gels were run in Tris-glycine buffer (25 mM Tris-HCl, 250 mM glycine, 0.1% SDS). Protein bands were visualized with a silver-stain (Gallagher and Sasse 2008).

5.2.10 Isothermal denaturation (ITD) and differential scanning fluorimetry (DSF)

Protein thermal denaturation assays were performed by the method described by Niesen et al. (2007). A 5x working concentration of SYPRO orange (Sigma-Aldrich, St. Louis, Missouri, USA) and 65 ng of protein were used in each reaction. Samples were incubated in a StepOnePlus Real-Time PCR machine (Applied Biosystems). For IDT samples were incubated at 55 °C for 10 h while DSF samples were incubated with the temperature increasing by 1 °C per minute from 25-95 °C. For experiments testing the influence of ligands on protein unfolding, sugars

were dissolved in 0.2 M citrate phosphate buffer pH 5.5 to give a final working concentration of 0.05 M. Sugar solutions were added to proteins stored on ice. Control reactions were performed for all experiments in which the enzymes and/or sugars were omitted. Multicomponent data were exported from the StepOne software to Excel 2010 and ROX dye filter data were used. Data points beyond the maximum fluorescence +4 were discarded. First derivatives were calculated in Statistica version 12 (StatSoft Inc.). In cases where 2 peaks were obtained after application of the derivative, the temperature of the second peak was used. One-way analysis of variance (ANOVA) was conducted to test for differences between treatments applied to enzymes.

5.2.11 Homology modelling

Homology models for the single amino acid variants at the four positions yielding the most improved variants and the five best combination variants were generated by the SWISS-MODEL web server (Arnold et al. 2006). The crystal structure 3LF7 (Chuankhayan et al. 2010) served as template for the automated modelling mode. Template and target sequences shared 99% identity.

5.2.12 Solvent accessible surface areas (SASAs)

SASA data for the four substitution positions that yielded the most improved first round variants as well as for the combination variants themselves were generated from their respective homology models. SASAs were computed for the folded (from the crystal structure) and unfolded states (sequence specific theoretical calculations) by the ProtSA web server (Estrada et al. 2009). Differences between the folded and mean unfolded ensembles were determined. To examine the influence of the amino acid substitutions on SASAs, differences between the parent and variants were further calculated.

5.2.13 FOS synthesis

FOS synthesis was performed with the parent enzyme and the variant displaying the highest specific activity. It was accomplished by reacting 10 U of parent enzyme per gram of sucrose and dosing the same amount of protein for the variant. The reaction was performed at 62 °C with shaking at 120 rpm. Working concentrations were 600 g/l sucrose dissolved in 50 mM citrate phosphate buffer, pH 5.5. Samples were taken hourly for 12 h and analysed by HPLC after appropriate dilution.

5.3 RESULTS AND DISCUSSION

5.3.1 Round I: Single amino acid substitution library screening

Table 5.1 lists the amino acid substitutions selected to generate the first round of variants. Although the substitutions were limited to the loop regions, they were otherwise distributed across the entire protein sequence length. Substitutions were made at 13 positions with multiple substitutions in 10 of the positions. Of the 36 variants, 15 were deemed inactive as evidenced by little or no sucrose conversion as quantified by HPLC and FT-MIR spectroscopy (data not shown). Only 9% of substitutions made in the loops of the β-sandwich domain resulted in inactive variants as opposed to 56% of substitutions in loops in the β-propeller domain. These data highlighted the sensitivity of the catalytic β-propeller domain to structural alterations. Amino acids D185 and Y261 were both located in the β-propeller domain and enzyme activity was abolished by all of the selected substitutions. Upon closer inspection of the crystal structure 3LF7 (Chuankhayan et al. 2010), these residues were positioned directly adjacent to residues participating in β-strand formation and substituted amino acids may have interfered with the secondary structure folding. With a total of 58% active variants, the first round library appeared to be functionally enriched when compared to previously reported semi-rationally designed libraries which ranged from 10–54% active variants (Chen et al. 2012). Of the active variants, 71% displayed improved activity relative to the parent in terms of nystose production. Fig. 5.1 shows the nystose and relative inhibition data for the 21 active variants. The larger the relative inhibition value the less sensitive the enzyme to glucose inhibition.

Table 5.1 Single amino acid substitutions made to the parent enzyme to generate the first round library of β-fructofuranosidase variants

Amino acid substitution	Position on crystal structure (3LF7A)	Domain	Relative solvent accessibility (%)	Sequence Entropy	Enzyme activity*
P28A	28	β-propeller	64	1.0	Active
G109S	109	β-propeller	79	1.5	Inactive
F140Y	140	β-propeller	50	2.0	Active
F140S	140	β-propeller	50	2.0	Inactive
F140T	140	β-propeller	50	2.0	Inactive
F140R	140	β-propeller	50	2.0	Inactive
A178P	178	β-propeller	55	1.7	Active
A178S	178	β-propeller	55	1.7	Active
A178Y	178	β-propeller	55	1.7	Inactive
D185N	185	β-propeller	65	1.2	Inactive
D185H	185	β-propeller	65	1.2	Inactive
D185Q	185	β-propeller	65	1.2	Inactive
Y261S	261	β-propeller	67	2.0	Inactive
Y261N	261	β-propeller	67	2.0	Inactive
Y261T	261	β-propeller	67	2.0	Inactive
G321N	321	β-propeller	56	2.1	Active
G321D	321	β-propeller	56	2.1	Active
G321E	321	β-propeller	56	2.1	Active
G321Y	321	β-propeller	56	2.1	Active
E389N	389	β-propeller	100	1.0	Active
E389A	389	β-propeller	100	1.0	Inactive
E389K	389	β-propeller	100	1.0	Inactive
D454Q	454	β-propeller	78	2.1	Active
D454G	454	β-propeller	78	2.1	Active
D454T	454	β-propeller	78	2.1	Inactive
E485P	485	β-sandwich	96	2.2	Active
E485Q	485	β-sandwich	96	2.2	Active
E485N	485	β-sandwich	96	2.2	Active
E485S	485	β-sandwich	96	2.2	Active
Q490S	490	β-sandwich	69	2.4	Active
Q490K	490	β-sandwich	69	2.4	Active
Q490N	490	β-sandwich	69	2.4	Active
T569P	569	β-sandwich	56	1.7	Active
T569N	569	β-sandwich	56	1.7	Active
T569A	569	β-sandwich	56	1.7	Active
N648D	648	β-sandwich	65	1.3	Inactive

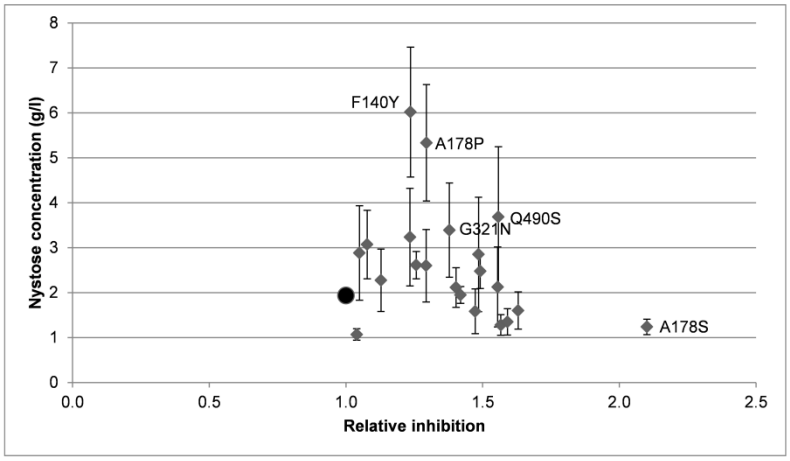

Fig. 5.1 Screening data for the active first round variants harbouring single amino acid substitutions. The parent enzyme is indicated by the filled circle. Error bars denote standard error (n = 3). Nystose data were generated by HPLC from assays performed under glucose inhibiting conditions. Sucrose consumption under normal and glucose inhibiting conditions was quantified by Fourier transform mid-infrared spectroscopy. Relative inhibition was expressed as the difference in sucrose consumption between the two conditions divided by uninhibited activity. Data were normalised to the parental activity. The most improved variants in terms of nystose production and/or relative inhibition are labelled

Nystose data, as quantified by HPLC, was chosen as better indicators of fructosyltransferase activities of the FopA variants under assay conditions. Time course data for FOS production have previously indicated that 1-kestose was the initial product with levels increasing rapidly. With reaction progression 1-kestose levels tapered off with concomitant increases in nystose and later GF4 levels (Nishizawa et al. 2001). As nystose levels consistently increased within a broader timeframe, they were taken as more reliable measures of enzyme activity for screening purposes in this study. Thus, for short reaction times (1-2 h) enzymes that produced more nystose than the parent were considered superior. Inhibition data were calculated from sucrose conversion data generated in the presence and absence of glucose inhibitor. Given its rapid and simple implementation, FT-MIR spectroscopy was used to quantify sucrose. Fig. 5.1 shows that variant A178S appeared to tolerate glucose better than other variants but at the expense of its fructosyltransferase activity. First round screening data were used to design combination variants of the top 5 performing round I variants. The top 4 nystose producing variants F140Y, A178P, Q490S and G321N and the single variant showing relief from glucose inhibition (A178S) were combined exhaustively to produce another 18 variants which included combinations of 2, 3 and 4 amino acid substitutions.

5.3.2 Round II: Combination variants screening

Table 5.2 lists the combination variants grouped by number of substitutions and their activity data arranged per group in descending order for nystose. The parent (wild type) data is given in bold font. Sixteen of the 18 variants showed improved activity relative to the parent. The 4

combination variant F140Y-A178P-G321N-Q490S proved to be the most improved variant with nystose levels 65% higher than the parent – 53.8 g/l versus 32.6 g/l. In round I variant A178P produced one of the highest levels nystose while A178S showed the highest tolerance for glucose. In round II only F140Y-A178S-G321N displayed poorer fructosyltransferase activity than the parent. Furthermore, all the variants with the A178S substitution performed worse than the variants with the A178P substitution. Although combining A178S with other substitutions did recover activity relative to the parent, it was generally assumed to be a deleterious substitution and did not prove permissive in a combinatorial context. No trend emerged for relative inhibition of combination variants and the A178S substitution (data not shown). Single substitutions ranked in decreasing order of nystose production A178P, G321N, Q490S, F140Y and A178S. A178P and G321N always proved to be good substitutions be it in isolation, in combination with each other and with either Q490S and/or F140Y. The contributions of Q490S and F140Y to enzyme activity were combination dependent. Together in a combination of 2 they were the poorest performers of the combination mutants. When combined with A178P, F140Y improved enzyme activity to a greater degree than Q490S – 37.6 g/l nystose produced as opposed to 34.7 g/l. In a combination of 2 with G321N, both F140Y and Q490S resulted in the same amount of nystose produced at approximately 44 g/l. However, the 1-kestose values differed with F140Y-G321N levels at 79.3 g/l while G321N-Q490S levels were 72.2 g/l. It is probable that G321N-Q490S was a more efficient enzyme as decreasing 1-kestose levels with similar nystose levels indicate GF4 production. No significant differences in the levels of sucrose, glucose and fructose supported this deduction. For a 3 combination variant the opposite for Q490S was true – A178P-G321N-Q490S produced more nystose than F140Y- A178P-G321N.

Table 5.2 Activity data for combination variants. Data are arranged in order of decreasing nystose per number of substitutions. Values in brackets are standard error (n = 3). The five most improved variants are coded

Number of substitutions	F140Y	A178P	A178S	G321N	Q490S	Glucose (g/l)	Fructose (g/l)	Sucrose (g/l)	1-Kestose (g/l)	Nystose (g/l)
4 (V1)	**x**	**x**		**x**	**x**	74.6 (± 3.4)	14.5 (± 0.9)	177.7 (± 1.3)	67.2 (± 0.5)	53.8 (± 2.0)
4	**x**		**x**	**x**	**x**	50.3 (± 19.6)	14.6 (± 6.1)	186.1 (± 3.0)	71.9 (± 6.4)	33.3 (± 1.5)
3 (V4)		**x**		**x**	**x**	68.1 (± 1.6)	11.2 (± 0.6)	177.8 (± 0.6)	63.1 (± 0.9)	50.8 (± 1.5)
3 (V5)	**x**	**x**			**x**	69.0 (± 3.9)	12.0 (± 1.2)	180.2 (± 3.0)	61.0 (± 3.7)	47.4 (± 2.5)
3 (V3)	**x**	**x**		**x**		67.0 (± 0.5)	14.1 (± 2.0)	176.5 (± 2.2)	65.3 (± 1.4)	47.0 (± 0.7)
3	**x**			**x**	**x**	64.8 (± 0.9)	10.2 (± 1.6)	173.4 (± 2.9)	75.0 (± 1.0)	42.4 (± 2.3)
3	**x**		**x**		**x**	60.6 (± 0.6)	12.0 (± 0.4)	168.4 (± 0.8)	80.7 (± 0.2)	34.2 (± 0.6)
3			**x**	**x**	**x**	58.7 (± 0.6)	22.3 (± 0.5)	167.1 (± 1.5)	80.6 (± 0.6)	34.0 (± 0.4)
3	**x**		**x**	**x**		47.3 (± 1.0)	18.4 (± 1.5)	119.9 (± 1.1)	86.5 (± 1.1)	15.6 (± 0.7)
2 (V8)		**x**		**x**		67.1 (± 1.4)	15.3 (± 0.8)	174.1 (± 0.9)	69.2 (± 2.5)	47.4 (± 0.6)
2				**x**	**x**	57.7 (± 7.5)	11.5 (± 0.6)	176.5 (± 3.2)	72.2 (± 1.1)	44.7 (± 1.6)
2	**x**			**x**		66.3 (± 1.1)	12.9 (± 1.9)	173.1 (± 0.5)	79.3 (± 1.1)	44.3 (± 1.2)
2	**x**	**x**				67.3 (± 5.8)	14.0 (± 0.8)	172.2 (± 0.2)	81.4 (± 1.3)	42.0 (± 0.9)
2		**x**			**x**	21.0 (± 9.4)	12.2 (± 1.8)	178.7 (± 3.5)	59.0 (± 8.0)	37.6 (± 5.2)
2	**x**				**x**	54.2 (± 9.2)	16.0 (± 5.9)	173.2 (± 3.3)	76.8 (± 2.4)	34.7 (± 4.1)
2			**x**	**x**		59.3 (± 1.3)	17.0 (± 1.2)	165.0 (± 0.7)	85.2 (± 1.8)	34.4 (± 0.6)
2	**x**		**x**			57.8 (± 1.7)	15.0 (± 1.4)	170.1 (± 4.4)	76.1 (± 7.1)	34.2 (± 4.4)
2			**x**		**x**	61.1 (± 1.8)	13.1 (± 0.3)	167.5 (± 2.7)	87.5 (± 2.4)	33.6 (± 0.9)
1		**x**				43.5 (± 21.3)	10.8 (± 0.4)	180.8 (± 5.7)	77.2 (± 2.6)	40.1 (± 2.1)
1				**x**		49.0 (± 11.5)	13.2 (± 0.6)	175.3 (± 6.2)	79.1 (± 1.8)	37.6 (± 1.4)
1					**x**	48.5 (± 13.0)	16.3 (± 3.1)	174.9 (± 3.0)	83.7 (± 3.0)	33.8 (± 3.0)
1	**x**					59.2 (± 1.0)	15.3 (± 1.2)	165.1 (± 2.9)	82.9 (± 1.3)	32.9 (± 1.9)
1			**x**			46.7 (± 6.4)	18.7 (± 4.0)	164.2 (± 4.7)	68.0 (± 19.7)	22.8 (± 2.3)
Parent	**-**	**-**	**-**	**-**	**-**	**44.1 (± 14.9)**	**17.7 (± 3.7)**	**166.1 (± 2.2)**	**69.1 (± 12.8)**	**32.6 (± 1.2)**

5.3.3 Improved variant characterisation

This section applies to the parent and the five most improved combination variants. For simplicity the combination variants were named V1 (F140Y-A178P-G321N-Q490S), V3 (F140Y-A178P-G321N), V4 (A178P-G321N-Q490S), V5 (F140Y-A178P-Q490S) and V8 (A178P-G321N).

5.3.3.1 Protein purification and electrophoresis

Enzymes were purified using IMAC. Fig. 5.2 shows the crude yeast supernatants and purified enzymes on a silver stained SDS-PAGE gel.

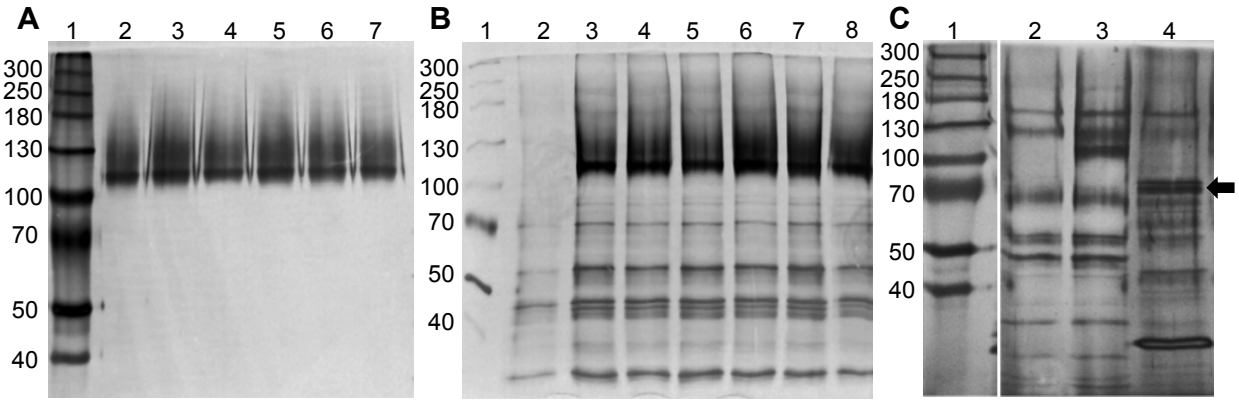

Fig. 5.2 SDS-PAGE gels of purified (A) and crude (B, C) parent and five most improved combination variant β-fructofuranosidases. All gels were 8% and silver stained. A shows the IMAC purified enzymes. Lanes 2 to 7 show the parent and variant enzymes. B shows crude supernatants of *S. cerevisiae* NI-C-D4[fopA]. Lane 2 is a reference as yeast was transformed with empty vector pJC1. Lanes 3 to 8 show the parent and the five most improved variants. C shows the result of PNGase treatment of the crude *S. cerevisiae* NI-C-D4[fopA] supernatant. Lane 2 reference, lane 3 untreated and lane 4 treated sample. The Spectra multicolor high range protein ladder (Thermo Scientific) was loaded in lane 1 of all the gels and served as molecular weight marker

The high degree of enzyme purity following IMAC is evident in gel A. The intense bands between 100 and 130 kDa correspond to the parent and variant β-fructofuranosidases – the band was absent in the reference in which supernantant from *S. cerevisiae* transformed with plasmid backbone, pJC1, was loaded (gel B, lane 2). The same reference was used in gels B and C. The proteins were larger than the expected 69 kDa as deduced from the protein sequence but the shift was attributed to N-glycosylation of the proteins by *S. cerevisiae*. Following the removal of the glycosylation with PNGase F, bands of the expected 69 kDa were obtained (gel C, lane 4). All variants had 8 putative N-glycosylation sites which was unchanged from the glycosylation pattern of the parent (Blom et al. 2004). Chuankhayan et al. (2010) proposed that altered N-glycosylation patterns from heterologous hosts may influence enzyme stability. As all enzymes were expressed in the same host and retained the same putative N-

glycosylation pattern, altered variant stability (hence, altered enzyme activity) that could arise from altered N-glycosylation patterns due to differences in primary protein structures was ruled out.

5.3.3.2 Specific activity

Specific activities for the purified enzymes were determined. Results are shown in Fig. 5.3. The specific activity of the parent enzyme was 1652 U/mg protein which was lower than the 2650 U/mg protein reported by Nishizawa et al. (2001). Heterologous enzyme production and experimental disparities could account for this. Differences in activity between the native enzyme and the heterologous *fopA* gene product were also reported by Yanai et al. (2001). All the variants had significantly ($p = 0.0023$ for V8) higher specific activities than the parent, with V1 as much as double.

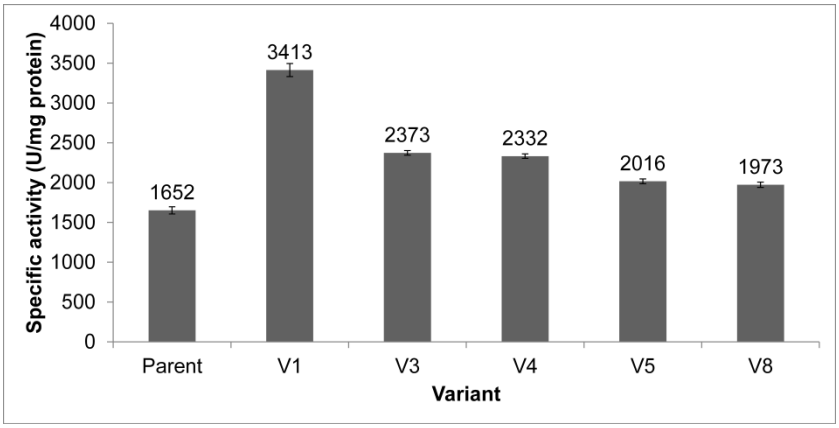

Fig. 5.3 Specific activity data for the purified parent and five most improved combination variants. Values above each bar indicate the average 1-kestose units per milligram purified enzyme. Error bars denote standard error (n = 3)

5.3.3.3 Isothermal denaturation (ITD)

To investigate the stability of the improved variants ITD, a method that quantifies protein stability by measuring protein unfolding caused by heat denaturation in the presence of SYPRO orange dye was employed (Epps et al. 2001). Typically this method is used to quantify stability and ligand affinity for a given protein. The method was employed in a modified sense in that the proteins were the variable factor. Purified proteins were incubated for 10 h at 55 °C to examine the influence of protracted exposure to temperatures routinely used for industrial FOS synthesis reactions. Fig. 5.4 shows that after 2 h the fluorescence measured for the parent enzyme increased dramatically whereas the variants showed more measured increases. The increased fluorescence can be interpreted as thermal denaturation of the enzymes. Upon protein

unfolding, binding of the SYPRO orange dye to newly exposed hydrophobic amino acid residues results in increased fluorescence (Lo et al. 2004). It thus appeared that the variant enzymes were more thermostable than the parent. Slight differences in stability were observed between the variants which could be attributed to the differing combinations of amino acid substitutions.

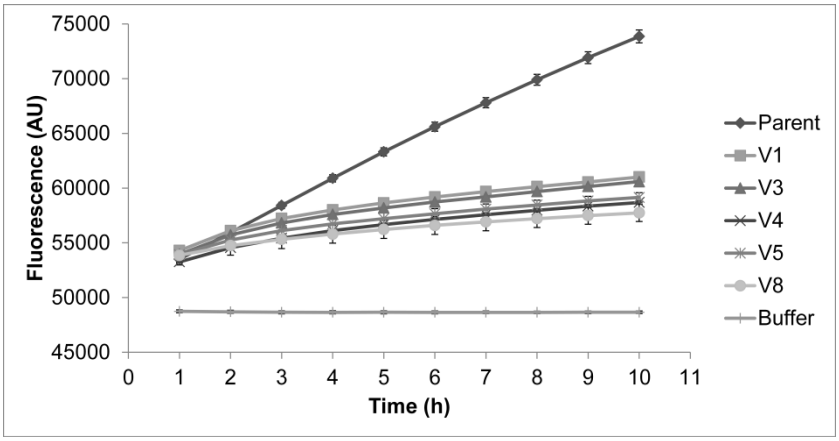

Fig. 5.4 Isothermal denaturation data for the parent and five most improved combination variants. Purified proteins were incubated with SYPRO orange at 55°C and fluorescence was monitored for 10 h. Increased fluorescence indicates thermal denaturation of the protein as SYPRO orange binds to newly exposed hydrophobic amino acids. Error bars denote standard error (n = 5)

5.3.3.4 Differential Scanning Fluorimetry (DSF)

5.3.3.4.1 Melting temperatures (T_m) and pH optima

To further investigate the thermostability of the variants DSF was employed (Niesen et al. 2007). The principle is similar to ITD however instead of maintaining a set temperature, it is increased by 1 °C per minute. The T_m of a protein is the temperature at which half of the protein molecules are unfolded and reflects the transition midpoint of the fluorescence versus temperature curve (Niesen et al. 2007; Privalov 1979; Schellman 1997). Factors influencing protein stability include buffers, salts and detergents and also specific interactions with ligands. Factors that promote stability delay the thermally induced unfolding and result in increased T_m. DSF experiments were conducted for the parent FopA and 5 variants at pH values ranging from 4 to 7. Fluorescence intensity curves for the parent displayed the typical two-state unfolding transition (folded to unfolded with no stable intermediates) (Niesen et al. 2007; Zucker et al. 2010) at all pH levels tested – profiles were sigmoidal and similar to Fig. 5.5A.

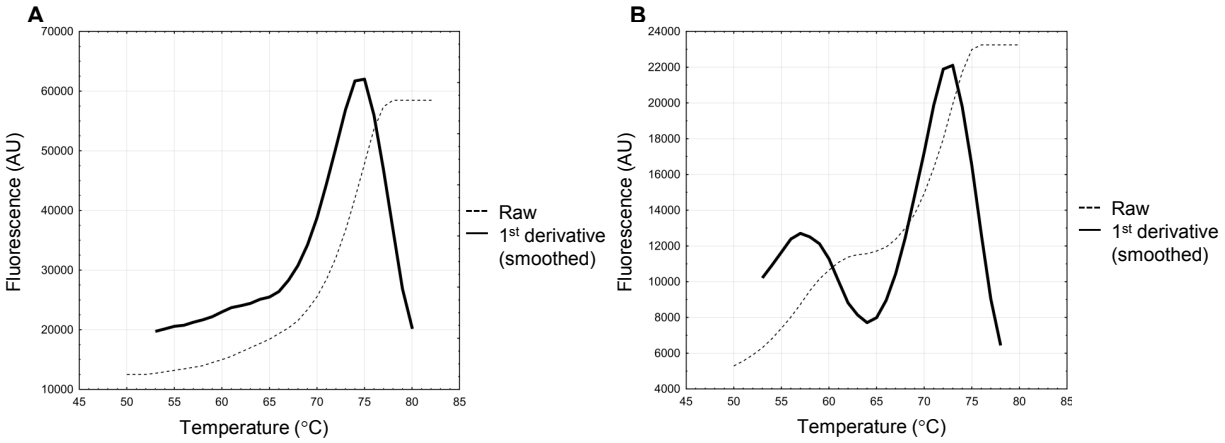

Fig. 5.5 Fluorescence intensity data generated by differential scanning fluorimetry for purified V1 at pH 4 (A) and 7 (B). Due to the multiple-state transitions during thermal denaturation of V1 at pH7 (B), the first derivative was applied to obtain peak maxima. In cases where two peaks were obtained, the second peak maximum was taken as the melting temperature of the enzyme

However, all the variants' profiles were pH dependent as shown in Fig. 5.5. At low pH the unfolding proceeded via the typical two-state transition (Fig. 5.5A) while at higher pH unfolding occurred via multiple-state transitions (Fig. 5.5B). This was attributed to a mutation(s) that stabilised a portion of the protein and thus more energy was required for complete unfolding. The maximum fluorescence at pH 4 was in excess of 60000 AU whereas at pH 7 it was 22000 AU. Hirayama *et al*. (1989) reported the optimum pH for the wild type enzyme in terms of activity and stability to be between 5 and 6 and 6.5, respectively. At pH values below 4.5 the stability and activity of the enzyme were severely compromised. Therefore, the high fluorescence and two-state transition of the variant proteins at low pH was attributed to the unfavourable conditions imposed on the enzyme at pH 4. The ionization states of amino acids were likely altered and hence contributed to extensive protein denaturation. T_m values were calculated from the first derivatives applied to the fluorescence data (Fig. 5.5) (Niesen et al. 2007). For samples presenting multiple transitions the maximum of the second peak was used to determine the T_m. The results at each pH for the enzymes are shown in Fig. 5.6. Similar to the ITD results, it was clear that the variants were more thermostable than the parent. The pH optima profiles did not significantly change due to amino acid substitution – the optima for all enzymes were at pH 5.0. At optimum pH, V8 displayed a 5.7 °C increased T_m relative to the parent (73.7 °C) making it the most thermostable variant. V1, V3 and V4 all displayed improvements of 5.3 °C while V5 was improved by 4.3 °C. The differences between the three T_m values were all statistically significant ($p<0.05$). Considering the combinations of substitutions in the variants, A178P and G321N appeared to be responsible for the improved thermostability. All variants harboured the A178P substitution and they were all more stable than the parent. As V1 (F140Y-A178P-G321N-Q490S), V3 (F140Y-A178P-G321N) and V4 (A178P-G321N-Q490S) all displayed the same improvement, it appeared that F140Y and

Q490S were marginally destabilising substitutions in terms of thermostability – inclusion of either one or both of them resulted in a decreased T_m relative to V8. As V8 (A178P-G321N) was the most improved and V5 (F140Y-A178P-Q490S) the least, it was concluded that G321N contributed positively to stability.

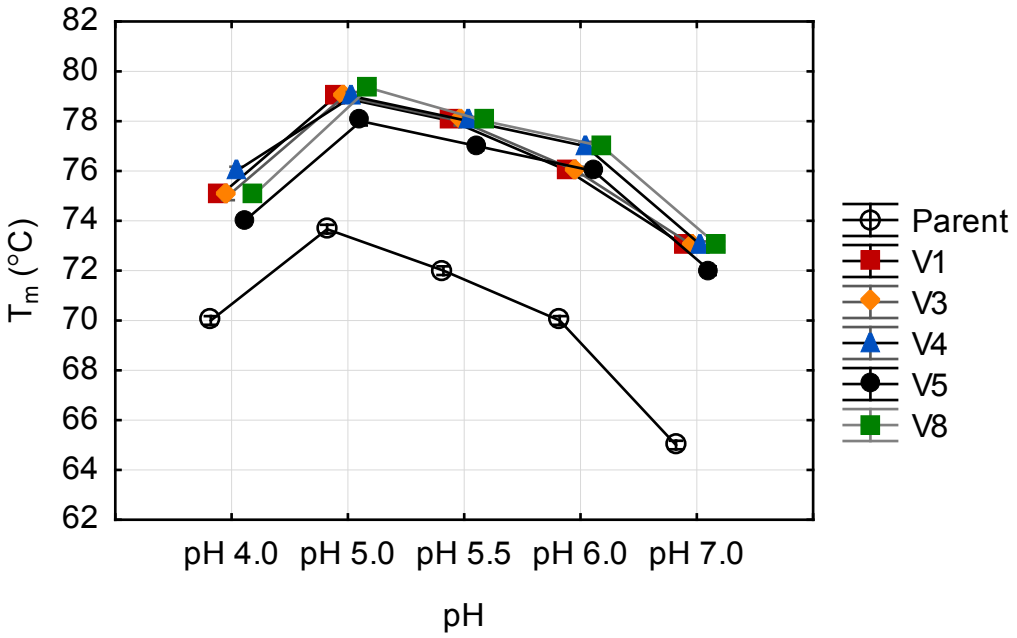

Fig. 5.6 Differential scanning fluorimetry-derived melting temperatures for purified parent and five most improved combination β-fructofuranosidase variants at pH values ranging from 4-7. Error bars denote 0.95 confidence intervals (n = 3)

5.3.3.4.2 Ligand affinities

Although V8 was the most thermostable enzyme it did not display the highest specific activity. DSF was used to perform comparative investigations into substrate interactions between the parent and variant proteins. More effective ligand binding to enzymes results in delayed thermal denaturation and higher T_m values. Results for ΔT_m (T_m at given substrate concentration minus T_m at 0 substrate concentration) at different substrate and product concentrations are shown in Fig. 5.7. At 0.2 M sucrose increased T_m values were observed for all enzymes (Fig. 5.7A). Further increased sucrose concentrations resulted in diminished substrate binding for all enzymes as evidenced by lower T_m values. At 0.5 M sucrose the parent, V5 and V8 displayed the same T_m as in the absence of substrate whereas variants V1, V3 and V4 displayed increased T_m values – differences were small, ΔT_m of 1 °C or less, but significant ($p<0.05$ for V1, V3 and V4). It was apparent that the substrate binding at 0.5 M sucrose was altered for enzymes with higher specific activities, namely V1, V3 and V4 (Fig. 5.3).

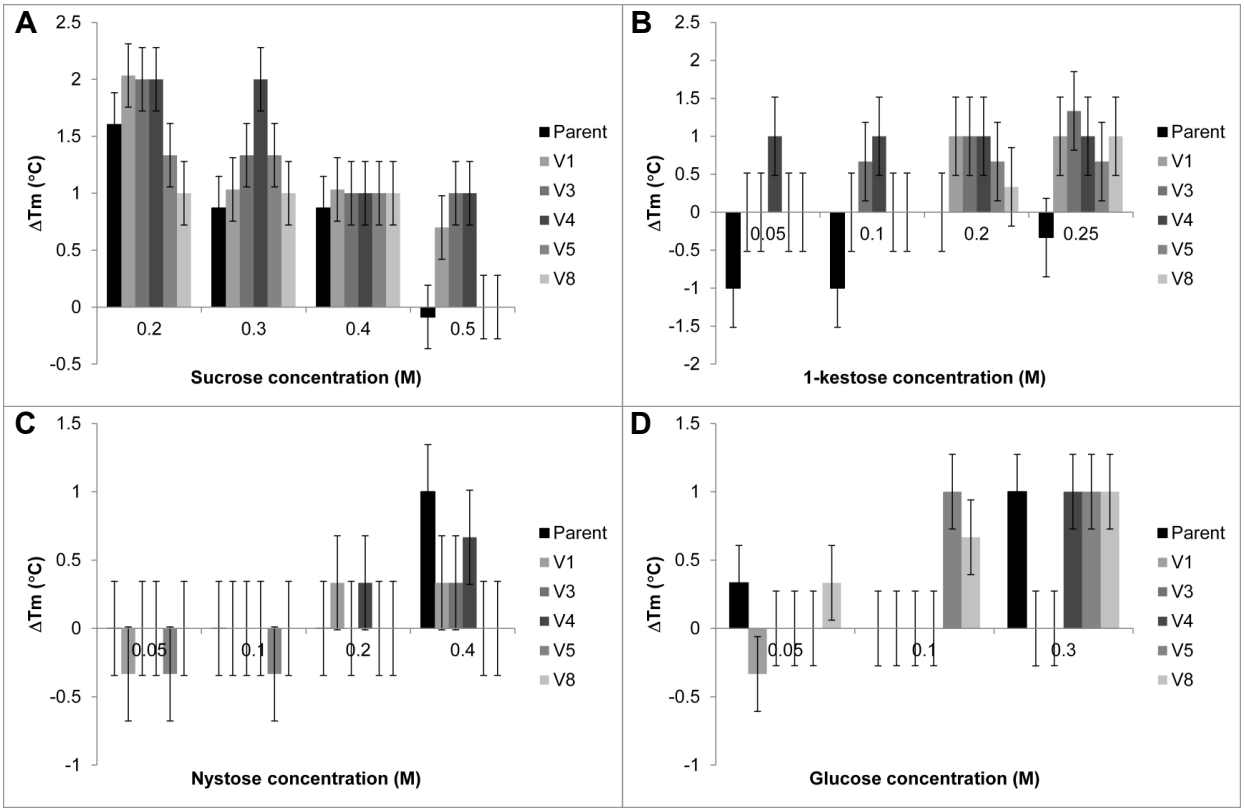

Fig. 5.7 ΔT_m values for the parent and 5 most improved combination variants in the presence of substrates sucrose (A), 1-kestose (B) and nystose (C). Differences between the given concentrations and the zero substrate T_ms are shown. Error bars denote 0.95 confidence intervals (n = 3)

In the case of 1-kestose all variants bound the substrate with higher affinity than the parent (Fig. 5.7B). Increased 1-kestose concentrations increased the stability of all variants except V4 which did not display an altered T_m at all concentrations tested. Enzyme stabilisation as a result of nystose binding was not evident at concentrations up to 0.2 M – no significant changes in T_m were observed (Fig. 5.7C). However, at 0.4 M the parent and V4 were stabilised significantly by nystose binding. The other variants did not show significant improvements in T_m as a result of nystose binding. Finally, binding of glucose (product) to enzymes was also examined and results indicated that the two most improved variants V1 and V3 appeared to have diminished affinities for glucose – T_m values were not affected at glucose concentrations up to 0.3 M while the parent, V3, V4 and V5 were stabilised by 1 °C. In summary, the most improved variant, V1 displayed increased affinities for sucrose and 1-kestose and decreased affinities for nystose and glucose at the highest substrate concentrations tested. The T_m of the parent was unaffected by sucrose or 1-kestose while it was increased by nystose and glucose, again at the highest concentrations tested. Taken in combination, these data indicated that subtle changes to the range of substrate and product affinities as well as improved thermostability were responsible for improved enzyme activities. It is plausible that the amino acid substitutions altered the active site such that the enzymes were relieved to an extent from product inhibition, a condition that is well documented in the literature for β-fructofuranosidases. Further kinetic

characterisation is required to fully understand the impact of substitutions on the enzymes' activities.

5.3.4 Characterising amino acid substitutions using solvent accessible surface areas

5.3.4.1 Homology model quality

Homology models were generated for the single amino acid substitution variants at positions 140, 178, 321 and 490, namely the positions yielding the most improved first round variants and the five most improved combination variants. The wild type *A. japonicus* crystal structure (3LF7) served as template which had in excess of 99% amino acid identity with the targets. The structure was solved at a resolution of 2 Å and had R and R-free values of 0.213 and 0.24, respectively (Chuankhayan et al. 2010). The quality scores for the modelling are given in Table 5.3.

Table 5.3 Quality scores for homology models of β-fructofuranosidase variants at the four positions yielding the most improved variants. The crystal structure 3LF7 served as template

Homology model	QMEAN4*	GMQE**
F140Y	-0.66	1
F140T	-0.59	1
F140S	-0.58	1
F140R	-0.61	1
A178Y	-0.65	1
A178S	-0.65	1
A178P	-0.55	1
G321D	-0.64	1
G321E	-0.65	1
G321N	-0.65	1
G321Y	-0.65	1
Q490K	-0.61	1
Q490N	-0.61	1
Q490S	-0.61	1

* Qualitative Model Energy Analysis
** Global Model Quality Estimation

The qualitative model energy analysis (QMEAN4) score provides a local and global estimation of model quality while global model quality estimation (GMQE) includes properties of the target-template alignment. The QMEAN4 score reflects a weighted linear combination of four structural descriptors including torsion angle potential over three amino acid residues, two pairwise distance-dependent potentials of Cβ interactions and predicted secondary structures and a residue solvation potential (Benkert et al. 2008). The GMQE combines alignment

properties and the QMEAN4 score. The scale of GMQE is between zero and one with higher numbers indicating higher quality models. Table 5.3 indicates that all models had GMQE scores of 1 which reflected the high degree of homology between the target and template sequences. The homology models were in essence identical to the crystal structure template thus the B-factors were examined to obtain an idea of the accuracy of the loop regions. Loop regions are generally flexible regions and have relatively high B-factors. Fig. 5.8 shows the B factors for 3LF7 with the positions where substitutions were made marked by the arrows. It is apparent that for the marked positions and their immediate vicinities the B-factors were average. Thus it was concluded that the loops in the crystal structure were modelled with a relatively high degree of accuracy. Similar deductions were made for the homology models.

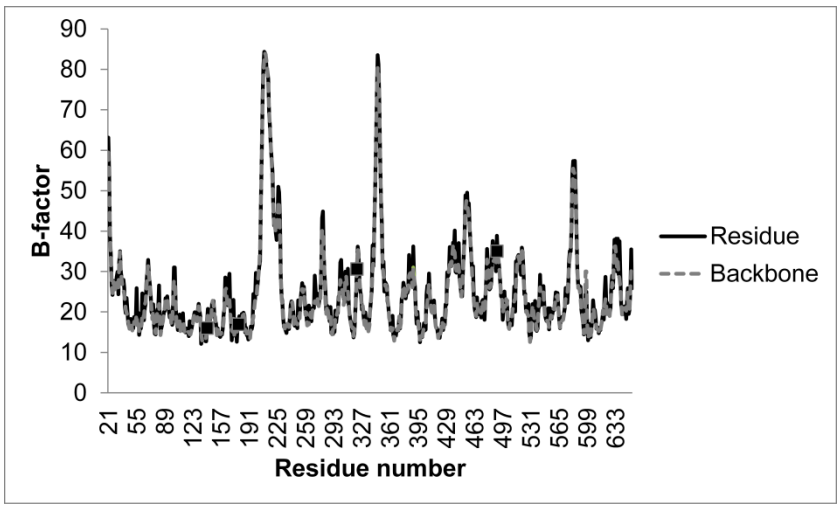

Fig. 5.8 Experimental B-factors for the X-ray crystal structure (3LF7) of the *Aspergillus japonicus* β-fructofuranosidase (Berman 2000). Positions where amino acid substitutions were made to produce variant enzymes are indicated by filled squares on the line and labelled arrows above the line

5.3.5 Solvent accessible surface areas (SASAs)

The hydrophobic effect is a major factor contributing to protein folding (Baldwin 2007) and the stability of the folded state (Pace et al. 1996). The contribution of the hydrophobic effect to the free energy of folding can be determined from the change in SASA upon folding (Baldwin 2007; Wesson and Eisenberg 1992). Differences between the folded and mean unfolded ensembles were determined for the parent and the variants. To examine the influence of the amino acid substitutions on SASAs, differences between the parent and variants were further calculated. Results for these differences are shown in Figs. 9-12. SASAs of the residues were altered in the regions surrounding the substitutions with the remainder of the protein not exhibiting large changes. Upon protein folding the majority of residues became less exposed. A larger SASA value (i.e. more positive) for the difference between the parent and variant indicated that residues were less exposed upon folding in the variant. Smaller SASA values (i.e. more

negative) indicated that the residue was more exposed in the folded variant than the parent. Data for the combination variants were similar to the single amino acid substitution variants at the relevant positions and thus only the results for the single substitution variants are shown and discussed.

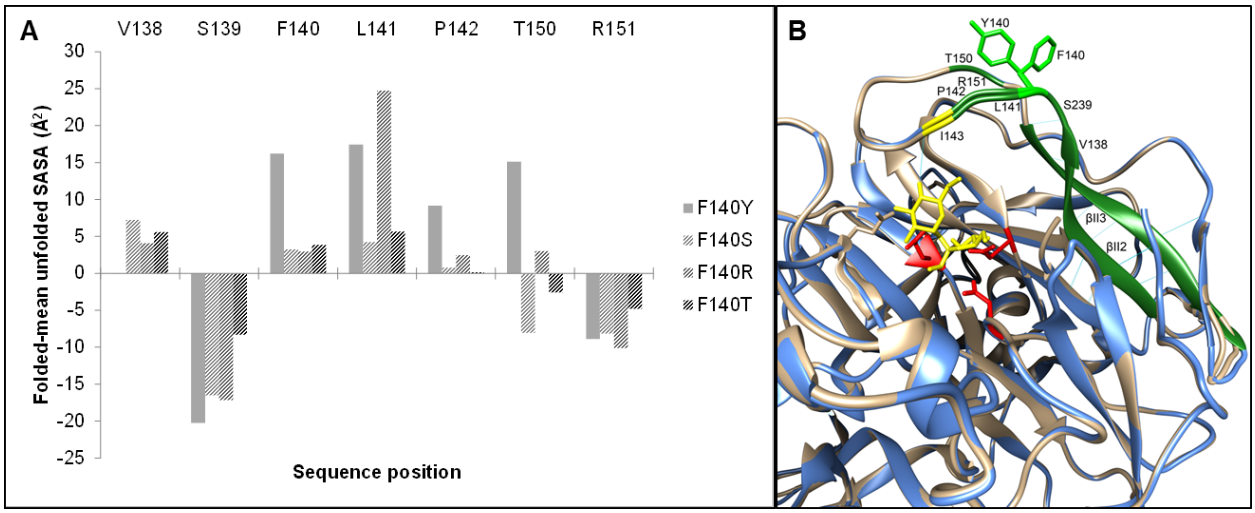

Fig. 5.9 Solvent accessible surface area (SASA) data (difference between folded and unfolded state) for single amino acid substitution variants at position 140 (A) and the 3D structure of superimposed models of the parent (beige) and F140Y (blue) (B). In A differences between the parent SASAs and the variant SASAs are shown. Variants are arranged in order of decreasing enzyme activity. Inactive variants are denoted by pattern filled bars. In B active site residues are coloured red, sucrose substrate and residues with direct interactions yellow, H-bonds are shown as turquoise lines. Residues with altered SASAs are labelled and coloured dark green. Substituted residue side chains are visible and coloured bright green. β-strands are also coloured dark green and labelled

At position 140 four amino acid substitutions were made, namely F140Y, F140S, F140R and F140T. Screening data indicated that only F140Y remained active (Table 5.1). In addition to F140, surrounding residue SASAs affected by substitution were V138, S139, L141, P142, T150 and R151 (Fig. 5.9A). Besides V138 which formed part of a β-strand, all the other residues were loop residues (Fig. 5.9B). Altering the SASA of V138 proved to be detrimental to activity – it was unchanged in the active variant F140Y while for the inactive F140 variants, V138 became less exposed. It is likely that decreased exposure of V138 interfered with the hydrogen bonding between strands 1–3 in the β-sheet of blade 2 in the β-propeller domain (numbering and nomenclature of structures as per Chuankhayan et al. 2010). Decreased exposure of residues Y140, L141, P142 and T150 and increased exposure of S139 and R151 also appeared to contribute to improved variant activity. I143 is directly involved in hydrogen bonding with substrates and although it did not display a much altered SASA in the variant, decreased exposure of its neighbours may have impacted on positioning of substrates and hence I143 interaction with substrates. Furthermore, L141, P142 and I143 were positioned in the loop between the β-strands forming part of the entrance to the active site pocket (blade II of the β-propeller). In terms of amino acid properties, a F to Y substitution did not constitute an

inordinate change in amino acid structure. Y also possesses an aromatic amino acid side chain with an additional –OH group which renders the residue less hydrophobic (Betts and Russell 2003). This decreased hydrophobicity may have been advantageous when considering the exposed surface position the residue occupied.

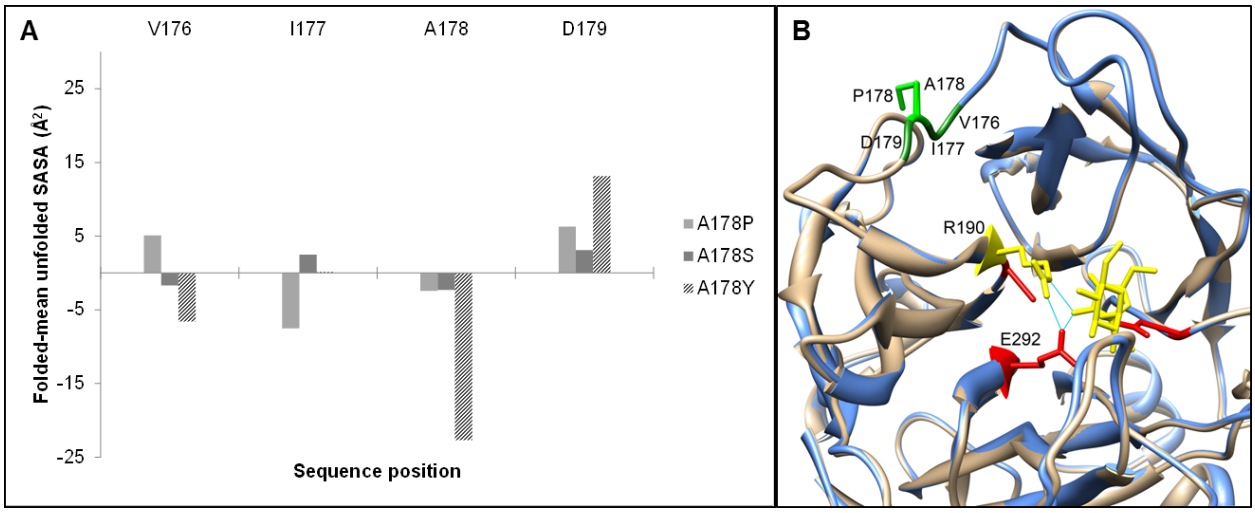

Fig. 5.10 Solvent accessible surface area (SASA) data (difference between folded and unfolded state) for single amino acid substitution variants at position 178 (A) and the 3D structure of superimposed models of the parent (beige) and A189P (blue) (B). In A differences between the parent SASAs and the variant SASAs are shown. Variants are arranged in order of decreasing enzyme activity. Inactive variants are denoted by pattern filled bars. In B active site residues are coloured red, sucrose substrate and residues with direct interactions yellow, H-bonds are shown as turquoise lines. Residues with altered SASAs are labelled and coloured dark green. Substituted residue side chains are visible and coloured bright green

At position 178 substitutions were A178P, A178S and A178Y. A178P was the most active variant followed by A178S. A178Y was inactive (Table 5.1). The SASAs of V176, I177, A178 and D179 were affected, with the active variants showing only minor changes (Fig. 5.10A). The A178Y substitution resulted in a greatly increased exposure of the Y residue and to a lesser extent V176. D179 was considerably less exposed. Substituting Y for A is generally disfavoured in all proteins (Betts and Russell 2003) – replacing a small, neutral amino acid with a large, aromatic and somewhat more hydrophilic residue had a detrimental effect on enzyme activity. On the other hand, the A178P substitution did not greatly influence solvent exposure at position 178 but decreased the exposure of V176 and increased exposure of I177. The opposite effects resulted in decreased enzyme activities. According to Betts and Russel (2003) an A to P substitution is generally disfavoured but in extracellular proteins it is a neutral substitution. P substitutions restrict the number of available main chain conformations due to the constraints on their allowed psi and phi angles. This reduces the entropy of the unfolded state (Stites and Pranata 1995) and was confirmed by the ProtSA analyses. The SASA of A178 was equal to the SASA of the average unfolded A in a set of 19 proteins used in their analyses while for P178 it was 25 – 40% lower than the average unfolded A (data not shown). Entropy

(as well as heat capacity and enthalpy) is strongly correlated to SASA (Baldwin 2007; Wesson and Eisenberg 1992). As protein stability is governed by the difference in Gibbs free energy between the folded and unfolded states a reduction in the entropy of the unfolded state contributes to thermodynamic stability (Wallach 1993). These findings are in agreement with previous studies where P substitutions in loops increased both loop rigidity and hence protein stability (Hardy et al. 1993; Watanabe et al. 1994; Wu and Arnold 2013).

A178P is located in the loop leading to the β-strand which holds R190 and D191 (Fig. 5.10B). R190 forms a salt bridge with the E292 acid/base catalyst and D191 is the transition-state stabiliser in the catalytic triad. R190 also forms hydrogen bonds with substrates (Chuankhayan et al. 2010). It is conceivable that backbone changes due to the A178P substitution could influence the active site configuration and also catalytic activity. Similar findings were made by Schallmey et al. (2013) for loop proline substitutions and altered active site geometries of a halohydrin dehalogenase.

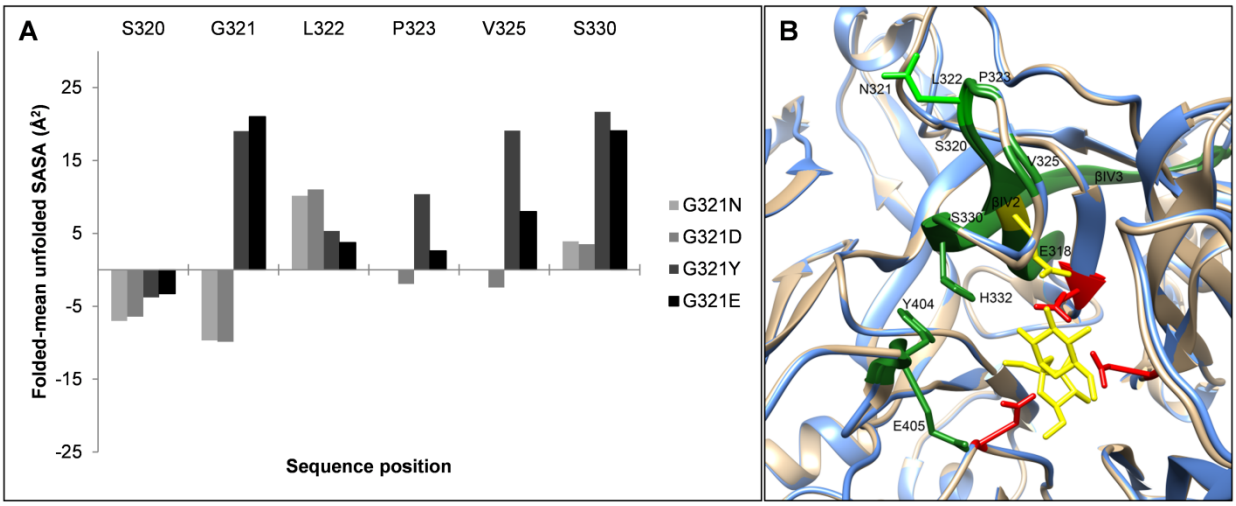

Fig. 5.11 Solvent accessible surface area (SASA) data (difference between folded and unfolded state) for single amino acid substitution variants at position 321 (A) and the 3D structure of superimposed models of the parent (beige) and G321N (blue) (B). In A differences between the parent SASAs and the variant SASAs are shown. Variants are arranged in order of decreasing enzyme activity. In B active site residues are coloured red and sucrose substrate and residues with direct interactions yellow. Residues with altered SASAs are labelled and coloured dark green. Substituted residue side chains are visible and coloured bright green. β-strands are also coloured dark green and labelled

All variants generated at position 321 were active (Table 5.1) – the substitutions were G321N, G321D, G321Y and G321E. They are listed in order of decreasing activity, however there were not major differences between the variants. Substitutions at this position affected the loop connecting the second and third strands of blade IV of the β-propeller. Together with loops in blades I, II and V these loops encompass the entrance to the active-site pocket. Furthermore, strands 2 and 3 in blade IV harbour residues E318 and H332, respectively which form part of the negatively charged active-site pocket. H332 is involved in a stacking interaction with Y404

which stabilises the folding of the enzyme and positions Y404 and E405 such that they stabilise the substrate at the +2 subsite. The latter is integral to the formation of inulin-type FOS (Chuankhayan et al. 2010). Alterations to the conformation of the loop harbouring position 321 may thus impact on the positioning of residues regulating entrance of substrates to the active site pocket, the geometry of the active site pocket as well as substrate binding. In terms of SASAs the four substitutions either had similar effects – all slightly increased the exposure of S320 and all decreased the exposure of L322 and S330 – or opposing effects on surrounding residues (Fig. 5.11A). At position 321, N and D substitutions increased exposure of the residue while Y and E substitutions greatly decreased exposure. It appeared that increased exposure at this position correlated with slightly higher enzyme activity. Findings by Zhang et al. (2009) related local solvent accessibility to residue flexibility and showed that solvent exposure of adjacent residues promotes flexibility of a central residue, while burial of adjacent residues inhibits the flexibility of a central residue, irrespective of the exposure state of the central residue. Considering the altered solvent exposures of residues in the vicinity of position 321, the flexibility of L322 was limited by Y and E substitutions at position 321 while N and D substitutions promoted flexibility. It is likely that altered flexibility of the loop impacted on the positioning of neighbouring residues Q327, V328 and S329 which lined the entrance to the active pocket (Fig. 5.11B). Substituted residues varied in size and charge and it thus appeared that position 321 was tolerant to a range of amino acid type substitutions. This was most likely due to the lack of direct noncovalent interactions of the loop with neighbouring secondary structures.

Finally, three substitutions were made at position 490 namely, Q490S, Q490K and Q490N. Again, all three variants were active (Table 5.1Table 5.1). Q490S was marginally more active than Q490K (3.7 g/l vs 3.2 g/l nystose) but notably improved over Q490N (2.1 g/l nystose). In contrast to the other three positions which were located in the β-propeller, position 490 was situated in the β-sandwich domain. By virtue of its distance from the active site, this is a residue that would not have been rationally identified for substitution. The function of the domain remains largely unknown. In the *Saccharomyces cerevisiae* invertase a role for the β-sandwich domain in dimer interaction and hence an octameric biological assembly is shown but this assembly appears to be unique to this GH32 enzyme (Sainz-Polo et al. 2013). In the *S. occidentalis* β-fructofuranosidase, its dimeric nature is mediated by interactions between residues from both subunits, some in the β-sandwich domain. Dimer formation is important for substrate specificity and thus the function of the β-sandwich domain was assigned (Alvaro-Benito et al. 2010b). FopA is a monomer so this likely does not apply to this enzyme (Chuankhayan et al. 2010). The SASAs of residues either side of position 490 were not greatly

impacted by the substitutions. However, the Q490K substitution resulted in a largely decreased solvent exposure of the residue (Fig. 5.12A).

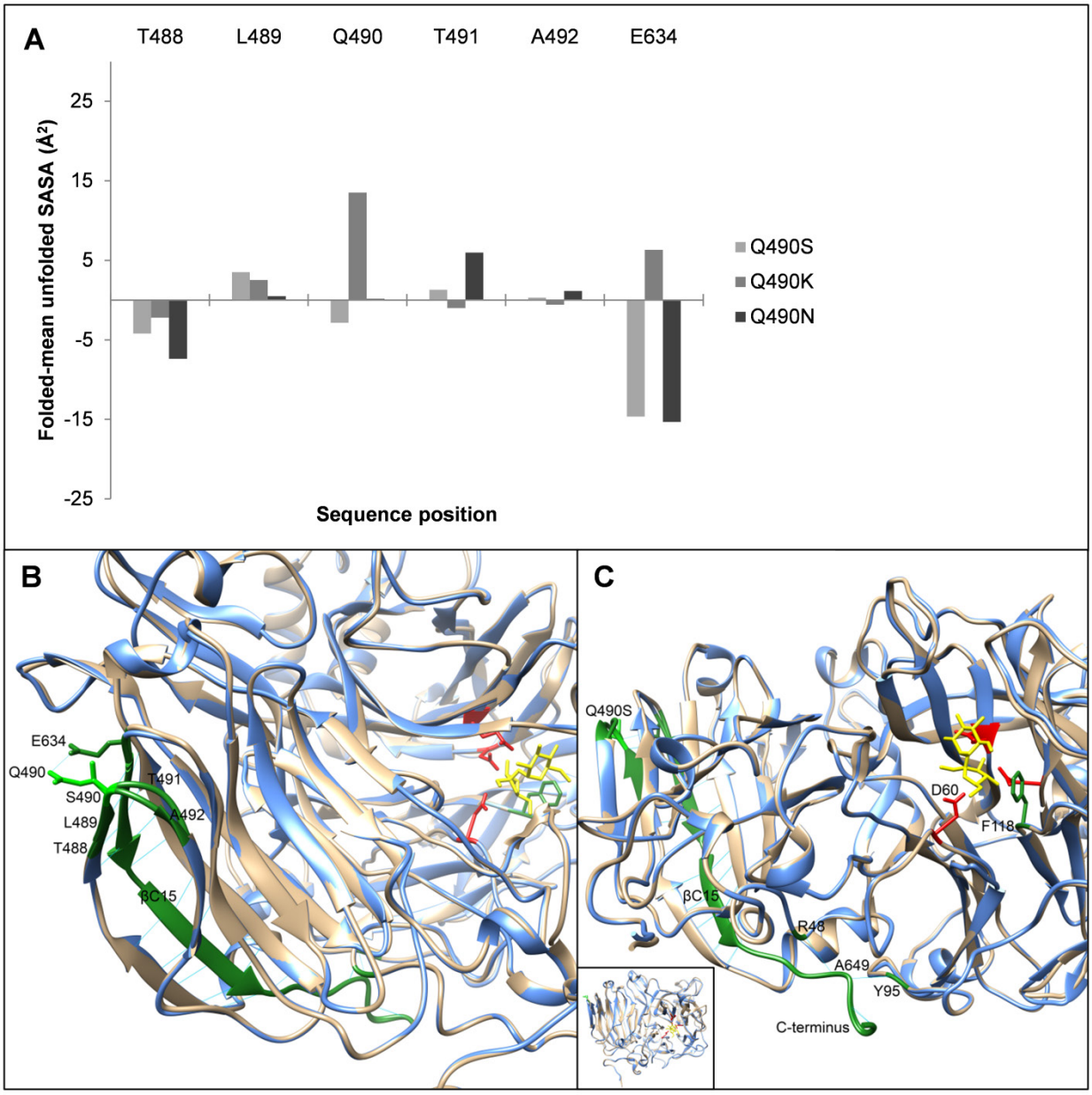

Fig. 5.12 Solvent accessible surface area (SASA) data (difference between folded and unfolded state) for single amino acid substitution variants at position 490 (A) and the 3D structure of superimposed models of the parent (beige) and Q490S (blue) (B and C). In A differences between the parent SASAs and the variant SASAs are shown. Variants are arranged in order of decreasing enzyme activity. In B and C active site residues are coloured red, sucrose substrate and residues with direct interactions yellow, H-bonds are shown as turquoise lines. Residues with altered SASAs are labelled and coloured dark green. Substituted residue side chains are visible and coloured bright green. β-strands are also coloured dark green and labelled. B shows a side view with clear visibility of position 490 and surrounding residues. C shows a front view with clear visibility of the C-terminus and active site pocket. The insert shows the whole enzyme structure highlighting the distance of the substitution (green) from the active site

As indicated in the figure, the SASAs of E634 were impacted by substitutions at position 490. In variant Q490S an increased residue exposure at position 490 was paired with a corresponding

increased exposure of E634. For Q490K the opposite was true – decreased exposure of K490 was paired with decreased exposure of E634. Either effect has a positive influence on enzyme activity. The exposure of N490 remained unchanged while E634 was greatly increased. This 'unpaired' change in exposure of the residues was met by a less improved variant activity albeit still slightly superior to the parental activity. All substitutions retained the polar nature of the residue at position 490. In the native state E643 was situated opposite Q490 in close enough proximity to allow side chain H-bonding between OE1 and NE2 of the respective residues (Fig. 5.12B). This together with other noncovalent interactions positioned β-strand C15 such that a network of H-bonds can form that stabilise the anti-parallel β-sheet comprised of β-strands C1, C2, C15, C4, C11, C12 and C13. In Q490S the side chain H-bond with E643 was no longer present. Furthermore, β-strand C15 preceded the 10 residue C-terminus which was stabilised by H-bonds between pairs A649/Y95 and A649/R48 (Fig. 5.12C). Y95 was located upstream in the loop harbouring F118 which provided hydrophobic interactions with substrates. R48 is upstream in the loop harbouring D60, one of the catalytic residues. In a rather convoluted manner the importance of Q490 to enzyme stability and activity can be appreciated and hence the effects of altered of residue exposure. In order to gain complete clarity on the effects of the amino acid substitutions, crystal structures of the variants will be required. The SASA analyses contribute information on residue exposure which impacts on protein flexibility and adds another step towards understanding how amino acid networks in proteins contribute to function.

5.3.6 FOS synthesis

As an ultimate test of variant performance, equal amounts of purified parent FopA enzyme and V1 were used to produce FOS under conditions similar to those used in industry. In commercial FOS products the ratio of GF_2:GF_3:GF_4 is approximately 42:47:10 and depending on the product, compromise up to 95% of the dry mass after chromatographic separation of glucose, fructose and sucrose (Bujacz et al. 2011; Molis et al. 1996; Nishizawa et al. 2001; Saulnier et al. 2007). Fig. 5.13 shows HPLC data for all the relevant sugars. Regarding the FOS it is evident that V1 has a higher catalytic activity relative to the parent as the levels of nystose and GF_4 were higher. Higher 1-kestose levels for the parent highlighted the difference in activities as this is the initial FOS species produced from sucrose. It in turn serves as substrate for the formation of nystose and GF_4. As equal amounts of protein were dosed, it can be deduced that V1 has a higher turnover number (k_{cat}) than the parent. The fructose data for both enzymes were virtually identical which indicated that the hydrolytic activity was unchanged in the variant. V1 consumed marginally more sucrose than the parent but it was not reflected in the amount of total FOS produced. As the differences in molecular weights between sucrose and FOS are large,

experimental error could account for the discrepancy (a given amount of sucrose translates into a small amount of FOS which could be missed given a maximum of 10% error of quantification).

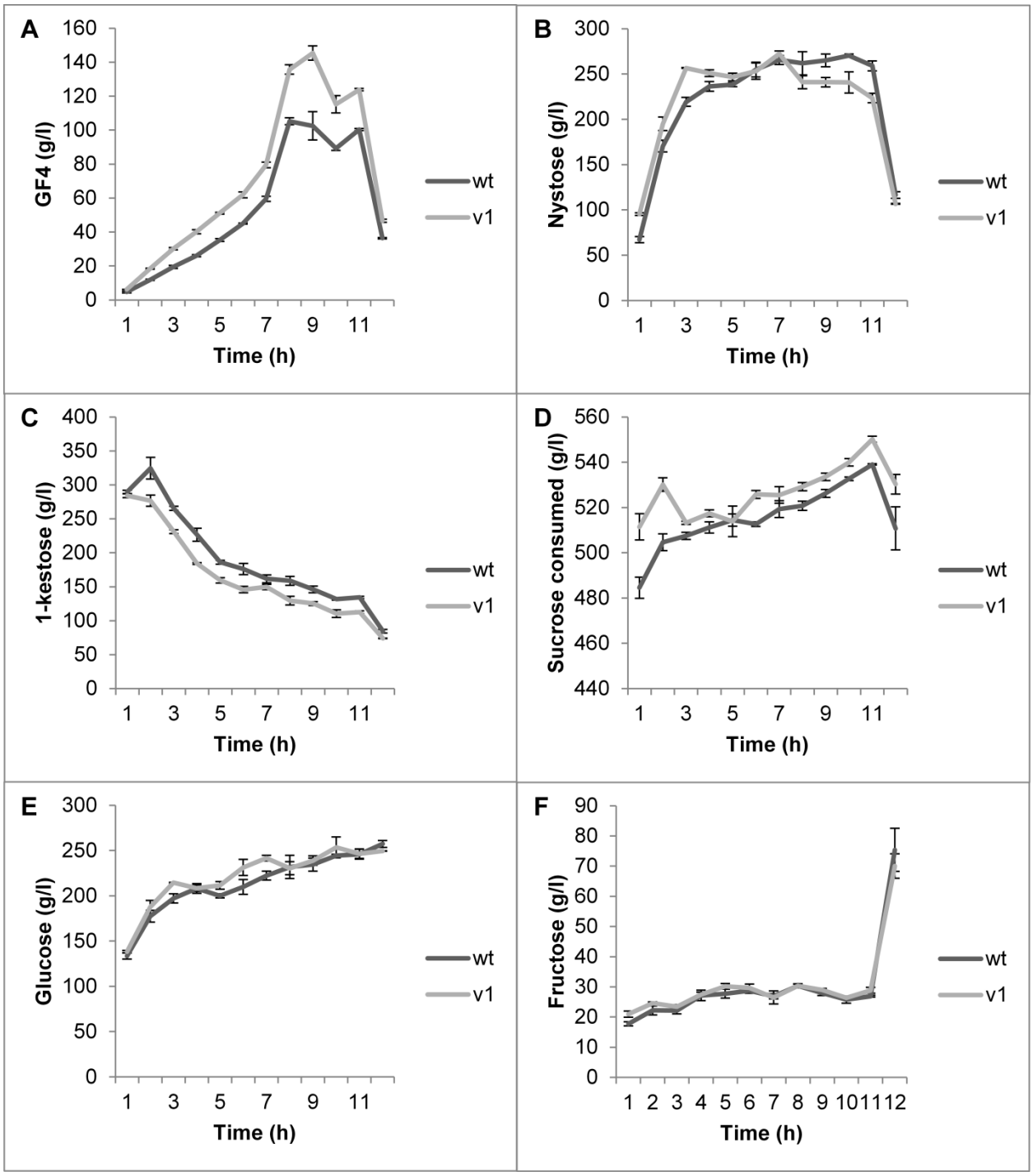

Fig. 5.13 Time course FOS synthesis by the purified parent and most improved variant (V1) enzymes. The enzyme dosage was 10 KU per gram sucrose with a starting concentration of 600 g/l sucrose. The reaction was conducted at 62 °C, pH 5.5 with shaking at 120 rpm. Error bars denote standard error (n = 3)

Similarly for glucose, which is an indicator of global enzyme activity, clear differences at all the time points were not evident. Although the reaction was followed for 12 h, the endpoint for a typical industrial reaction was regarded as the time when total FOS composition comprised 10%

GF_4. V1 reached 10% GF_4 in 4.6 h while it took the parent 6.2 h to produce FOS of similar composition. This difference represented a 26% reduction in time required to complete the reaction. Otherwise stated, enzyme dosage could be reduced to achieve the same result in 6.2 h. Fox and Clay (2009) derived the average velocity based on time-integrated enzyme behaviour over the reaction course and termed the expression the catalytic effectiveness of an enzyme. They noted that this parameter 'expresses the ability of an enzyme to meet the only parameter that is ultimately relevant … *viz*. time to completion'. Together with Eisenthal et al. (2007), they deemed k_{cat}/K_M an inappropriate metric to compare the activity of two enzymes acting on the same substrate. Thus, future kinetic characterisation of the variant could focus on obtaining a value for the enzyme's catalytic effectiveness in order to quantify the differences between the enzymes under industrially applicable conditions. Yields of FOS on sucrose added and consumed were 93% and 79%, respectively for the parent and 89% and 76% for V1. Yields are given for the time points reflecting 10% GF_4. Given the experimental error these yields can be regarded as equivalent. Interestingly, the hydrolytic activity of both enzymes dominated after 11 h as reflected by the sharp increase in fructose and corresponding decreases in GF_4, nystose, sucrose and to a lesser extent, 1-kestose levels.

5.4 CONCLUSIONS

Improving enzymes' catalytic performance by protein engineering still remains a somewhat hit-or-miss exercise despite the extensive research efforts in the field. Random mutagenesis is a popular and effective engineering technique which requires little prior knowledge about enzyme function or structure. However, the screening burden is very high. In scenarios where screening is the limiting factor rational approaches are a necessity. They however, require detailed knowledge about the protein. Structural predictions from sequence are already a reality, to a certain degree, but high resolution crystal structures for the protein of interest still remain first prize for rational design. Furthermore, many reports identify mutations that improved enzyme activity that could not have been rationally identified, hence the value of semi-rational approaches that employ aspects of both aforementioned strategies.

This study detailed the engineering of a β-fructofuranosidase using a crystal structure-guided approach to engineer only loop regions with high solvent exposure. The strategy avoided mutations in defined secondary structural elements which generally do not tolerate structural changes. In addition to structure, the approach also utilised an evolutionary-guided approach to select amino acid substitutions with proven fitness in the structures of homologous enzymes. This strategic combination likely resulted in the enriched functionality of the library – 58% of the first round library was active. Supplementary to identifying optimal amino acid substitutions, the

information obtained from the enriched functionality allows design of future mutations in regions that have demonstrated tolerance to mutation under the selected criteria. For example, the β-sandwich domain displayed high tolerance to mutation as evidenced by few inactive substitution variants. In a similar strategy focused on engineering loops, however employing site saturation mutagenesis at 90 loop residues in a lipase, Yedavalli and Rao (2013) obtained 10% of active mutants that were improved relative to the parent. In this study 71% of active mutants were improved over the parent and thus highlights the value of the strategy to improve the probability of identifying beneficial substitutions.

All possible combinations of five beneficial mutations were tested to obtain a four amino acid substitution variant that was the most improved variant, with a specific activity that was double that of the parent. Tested under conditions approximating its industrial application, the variant displayed an improved catalytic effectiveness by reducing the time to completion of the reaction by 26%. The industrial relevance of the catalyst was thus demonstrated. The stepwise combination of substitutions ensured exploitation of the full potential of the superior amino acids identified in the first round of screening. Data showed that the effect of combining the top four first round substitutions was cumulative and delivered the best variant compared to the three or two combination, or even single substitution variant.

ITD and DSF were used to gain an understanding of the effects of the substitutions on protein stability. A178P and G321N were identified as positive contributors to thermostability. Proline substitutions in loops have been linked to improved thermostability due to backbone modifications that increase loop rigidity (Goihberg et al. 2007; Schallmey et al. 2013) and data obtained in this work supports these findings. DSF data also showed that substrate and inhibitor binding altered the stability of the variants which supported the biochemical assay data and gave insight into the mechanisms of altered catalytic activity. Extrapolation from the DSF thermostability data would suggest that V1 can be applied under reaction conditions at least 5°C higher than currently employed. Optimisation of the required enzyme loading and optimal temperature for FOS synthesis is required, however it is probable that time to completion can be further reduced as it is accepted that reactions kinetics are enhanced at elevated temperatures.

Although catalytic pocket residues seem the more obvious choice for amino acid substitutions aimed at improving specific activity and substrate/product interactions, it has been demonstrated that changes in distant loops are able to mediate altered enzyme activity. SASA data from crystal structures are valuable to understand the impact of the amino acid substitutions on the particular residue in question as well as on residues in close proximity. Not surprisingly, a generalised pattern for altered SASA and enzyme activity did not emerge as

impacts of a substitution are expected to be sequence and three dimensional structure context-dependent. With time, improved understanding of the impact of mutations may allow for the prediction of their impact. It will most likely be in conjunction with other structural properties such as residue flexibility and packing densities (Yeh et al. 2013; Yeh et al. 2014; Zhang et al. 2009). In conclusion, the analyses in this work have shown that substitutions in solvent exposed loops mediate long range interactions which putatively alter protein flexibility and active site geometry, which in turn modify enzyme activity.

5.5 ACKNOWLEDGEMENTS

Niël van Wyk, Steven Pelly, Erick Strauss, Javier Sancho and Martin Kidd are thanked for their valuable insights. Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). This work forms part of the Technology Innovation Agency (TIA) funded project PB110/08. Kim Trollope was supported by a grant from TIA and Stellenbosch University. She also holds a L'Oreal-UNESCO For Women in Science Regional Fellowship in Sub-Saharan Africa.

5.6 LITERATURE CITED

Aharoni A, Thieme K, Chiu CPC, Buchini S, Lairson LL, Chen H, Strynadka NCJ, Wakarchuk WW, Withers SG (2006) High-throughput screening methodology for the directed evolution of glycosyltransferases. Nat Methods 3:609–614

Ahmad S, Gromiha MM, Sarai A (2003a) RVP-net: online prediction of real valued accessible surface area of proteins from single sequences. Bioinforma 19 :1849–1851

Ahmad S, Gromiha MM, Sarai A (2003b) Real value prediction of solvent accessibility from amino acid sequence. Proteins 50:629–635

Alberto F, Bignon C, Sulzenbacher G, Henrissat B, Czjzek M (2004) The three-dimensional structure of invertase (beta-fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. J Biol Chem 279:18903–18910

Alvaro-Benito M, de Abreu M, Portillo F, Sanz-Aparicio J, Fernández-Lobato M (2010a) New insights into the fructosyltransferase activity of *Schwanniomyces occidentalis* ß-fructofuranosidase, emerging from nonconventional codon usage and directed mutation. Appl Environ Microbiol 76:7491–7499

Alvaro-Benito M, Polo A, González B, Fernández-Lobato M, Sanz-Aparicio J (2010b) 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 285:13930–13941

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinforma 22 :195–201

Baldwin RL (2007) Energetics of protein folding. J Mol Biol 371:283–301

Benkert P, Tosatto SCE, Schomburg D (2008) QMEAN: A comprehensive scoring function for model quality assessment. Proteins 71:261–277

Berman HM (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242

Betts MJ, Russell RB (2003) Amino acid properties and consequences of subsitutions. In: Barnes MR, Gray IC (eds) Bioinformatics for geneticists.Wiley, New York

Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 4:1633–1649

Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast 14:115–32

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

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 37:D233–238

Chen MMY, Snow CD, Vizcarra CL, Mayo SL, Arnold FH (2012) Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. Protein Eng Des Sel 25:171–178

Chothia C, Finkelstein a V (1990) The classification and origins of protein folding patterns. Annu Rev Biochem 59:1007–1039

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

Cole MF, Gaucher EA (2011) Utilizing natural diversity to evolve protein function: applications towards thermostability. Curr Opin Chem Biol 15:399–406

Corradini C, Lantano C, Cavazza A (2013) Innovative analytical tools to characterize prebiotic carbohydrates of functional food interest. Anal Bioanal Chem 405:4591–4605

Crous JM, Pretorius IS, van Zyl WH (1995) Cloning and expression of an *Aspergillus kawachii* endo-1,4-beta-xylanase gene in *Saccharomyces cerevisiae*. Curr Genet 28:467–473

Cruz R, Cruz VD, Belini MZ, Belote JG, Vieira CR (1998) Production of fructooligosaccharides by the mycelia of *Aspergillus japonicus* immobilized in calcium alginate. Bioresour Technol 65:139–143

De Abreu MA, Alvaro-Benito M, Plou FJ, Fernández-Lobato M, Alcalde M (2011) Screening β-fructofuranosidases mutant libraries to enhance the transglycosylation rates of β-(2→6) fructooligosaccharides. Comb Chem High Throughput Screen 14:730–738

Edelman J (1956) The formation of oligosaccharides by enzymic transglycosylation. In: Nord FF (ed) Adv. Enzymol., Vol XVII. Interscience Publishers, Inc., New York, p 189

Eisenthal R, Danson MJ, Hough DW (2007) Catalytic efficiency and kcat/KM: a useful comparator? Trends Biotechnol 25:247–249

Epps DE, Sarver RW, Rogers JM, Herberg JT, Tomich PK (2001) The ligand affinity of proteins measured by isothermal denaturation kinetics. Anal Biochem 292:40–50

Estrada J, Bernadó P, Blackledge M, Sancho J (2009) ProtSA: a web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble. BMC Bioinformatics 10:104-112

Fox RJ, Clay MD (2009) Catalytic effectiveness, a measure of enzyme proficiency for industrial applications. Trends Biotechnol 27:137–140

Franzosa E a, Xia Y (2009) Structural determinants of protein evolution are context-sensitive at the residue level. Mol Biol Evol 26:2387–2395

Gallagher SR, Sasse J (2012) Staining Proteins in Gels. Curr Protoc Essent Lab Tech. 6:7.4:7.4.1–7.4.14

Goihberg E, Dym O, Tel-Or S (2007) A single proline substitution is critical for the thermostabilization of *Clostridium beijerinckii* alcohol dehydrogenase. Proteins: Struct., Funct., Bioinf. 204:196–204

Haney PJ, Badger JH, Buldak GL, Reich CI, Woese CR, Olsen GJ (1999) Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. Proc Natl Acad Sci U S A 96:3578–3583

Hardy F, Vriend G, Veltman OR, van der Vinne B, Venema G, Eijsink VGH (1993) Stabilization of *Bacillus stearothermophilus* neutral protease by introduction of prolines. FEBS Lett 317:89–92

Hidaka H, Eida T, Takizawa T, Tokunahga T, Tashiro Y (1986) Effects of fructo-oligosaccharides on intestinal flora and human health. Bifidobact Microflora 5:37–50

Hidaka H, Hirayama M, Sumi N (1988) A fructooligosaccharide-producing enzyme from *Aspergillus niger* ATCC 20611. Agric Biol Chem 52:1181–1187

Hill J, Donald K, Griffiths D (1991) DMSO-enhanced whole cell yeast transformation. Nucleic Acids Res 19:5791

Hirayama M, Sumi N, Hidaka H (1989) Purification and properties of a Fructoooligosaccharide-producing β-fructofuranosidase from *Aspergillus niger* ATCC 20611. Agric Biol Chem 53:667–673

Joo K, Lee SJ, Lee J (2012) Sann: Solvent accessibility prediction of proteins by nearest neighbor method. Proteins Struct Funct Bioinforma 80:1791–1797

Jung KH, Yun JW, Kang KR, Lim JY, Lee JH (1989) Mathematical model for enzymatic production of fructo-oligosaccharides from sucrose. Enzyme Microb Technol 11:491–494

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637

Koshland DE, Stein SS (1954) Correlation of bond breaking with enzyme specificity; cleavage point of invertase. J Biol Chem 208:139–148

L'homme C, Peschet JL, Puigserver A, Biagini A (2001) Evaluation of fructans in various fresh and stewed fruits by high-performance anion-exchange chromatography with pulsed amperometric detection. J Chromatogr A 920:291–297

Lafraya Á, Sanz-Aparicio J, Polaina J, Marín-Navarro J (2011) Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase. Appl Environ Microbiol 77:6148–6157

Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. J Mol Biol 55:379–400

Lo M-C, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M, Ellestad G (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem 332:153–159

Miller MP, Kumar S (2001) Understanding human disease mutations through the use of interspecific genetic variation. Hum Mol Genet 10 :2319–2328

Molis C, Flourie B, Ouarne F, Gailing MF, Lartigue S, Guibert A, Bornet F, Galmiche JP (1996) Digestion, excretion, and energy value of fructooligosaccharides in healthy humans. Am J Clin Nutr 64:324–328

Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2:2212–2221

Nishizawa K, Nakajima M, Nabetani H (2001) Kinetic Study on Transfructosylation by β-Fructofuranosidase from *Aspergillus niger* ATCC 20611 and Availability of a Membrane Reactor for Fructooligosaccharide Production. Food Sci Technol Res 7:39–44

O'Sullivan C, Tompson FW (1890) LX. - Invertase: a contribution to the history of an enzyme or unorganised ferment. J Chem Soc Trans 57:834

Pace CN, Shirley BA, McNutt M, Gajiwala K (1996) Forces contributing to the conformational stability of proteins. FASEB J 10:75–83

Perl D, Mueller U, Heinemann U, Schmid FX (2000) Two exposed amino acid residues confer thermostability on a cold shock protein. Nat Struct Biol 7:380–383

Pons T, Naumoff DG, Martínez-Fleites C, Hernández L (2004) Three acidic residues are at the active site of a beta-propeller architecture in glycoside hydrolase families 32, 43, 62, and 68. Proteins 54:424–432

Privalov PL (1979) Stability of proteins: small globular proteins. Adv Protein Chem 33:167–241

Ramsey DC, Scherrer MP, Zhou T, Wilke CO (2011) The relationship between relative solvent accessibility and evolutionary rate in protein evolution. Genetics 188:479–488

Roberfroid MB (2007) Inulin-type fructans: functional food ingredients. J Nutr 137:2493S–2502S

Saarelainen R, Paloheimo M, Fagerström R, Suominen P, Nevalainen K (1993) Cloning, sequencing and enhanced expression of the *Trichoderma reesei* endoxylanase II (pI 9) gene xln2. Mol Gen Genet 241:497–503

Sainz-Polo MA, Ramírez-Escudero M, Lafraya A, González B, Marín-Navarro J, Polaina J, Sanz-Aparicio J (2013) Three-dimensional structure of *Saccharomyces invertase*: role of a non-catalytic domain in oligomerization and substrate specificity. J Biol Chem 288:9755–9766

Sambrook J, Fritsch E, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sander C, Schneider R (1994) The HSSP database of protein structure-sequence alignments. Nucleic Acids Res 22:3597–3599

Sangeetha PT, Ramesh MN, Prapulla SG (2005) Recent trends in the microbial production, analysis and application of fructooligosaccharides. Trends Food Sci Technol 16:442–457

Saulnier DM a, Molenaar D, de Vos WM, Gibson GR, Kolida S (2007) Identification of prebiotic fructooligosaccharide metabolism in *Lactobacillus plantarum* WCFS1 through microarrays. Appl Environ Microbiol 73:1753–1765

Schallmey M, Floor RJ, Hauer B, Breuer M, Jekel P a, Wijma HJ, Dijkstra BW, Janssen DB (2013) Biocatalytic and structural properties of a highly engineered halohydrin dehalogenase. Chembiochem 14:870–881

Schellman JA (1997) Temperature, stability, and the hydrophobic interaction. Biophys J 73:2960–2964

Sherman F, Fink G, Lawrence C (1979) Methods in Yeast Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Shrake A, Rupley JA (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. J Mol Biol 79:351–371

Singh R, Singh R (2010) Production of fructooligosaccharides from inulin by endoinulinases and their prebiotic potential. Food Technol Biotechnol 48:435–450

Stites WE, Pranata J (1995) Empirical evaluation of the influence of side chains on the conformational entropy of the polypeptide backbone. Proteins 22:132–140

Trollope KM, Nieuwoudt HH, Görgens JF, Volschenk H (2014) Screening a random mutagenesis library of a fungal β-fructofuranosidase using FT-MIR ATR spectroscopy and multivariate analysis. Appl Microbiol Biotechnol 98:4063–4073

Wallach J (1993) Protein stability and stabilization through protein engineering. Biochem Educ 21:111

Wang B-D, Chen D-C, Kuo T-T (2001) Characterization of a *Saccharomyces cerevisiae* mutant with oversecretion phenotype. Appl Microbiol Biotechnol 55:712–720

Watanabe K, Masuda T, Ohashi H, Mihara H, Suzuki Y (1994) multiple proline substitutions cumulatively thermostabilize *Bacillus cereus* ATCC7064 oligo-1,6-glucosidase. Eur J Biochem 226:277–283

Wesson L, Eisenberg D (1992) Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci 1:227–235

Wintrode PL, Arnold FH (2000) Temperature adaptation of enzymes: lessons from laboratory evolution. Adv Protein Chem 55:161–225

Wu I, Arnold FH (2013) Engineered thermostable fungal Cel6A and Cel7A cellobiohydrolases hydrolyze cellulose efficiently at elevated temperatures. Biotechnol Bioeng 110:1874–1883

Yanai K, Nakane A, Kawate A, Hirayama M (2001) Molecular cloning and characterization of the fructooligosaccharide-producing beta-fructofuranosidase gene from *Aspergillus niger* ATCC 20611. Biosci Biotechnol Biochem 65:766–773

Yedavalli P, Rao NM (2013) Engineering the loops in a lipase for stability in DMSO. Protein Eng Des Sel 26:317–324

Yoshikawa J, Amachi S, Shinoyama H, Fujii T (2006) Multiple beta-fructofuranosidases by *Aureobasidium pullulans* DSM2404 and their roles in fructooligosaccharide production. FEMS Microbiol Lett 265:159–163

Zhang H, Zhang T, Chen K, Shen S, Ruan J, Kurgan L (2009) On the relation between residue flexibility and local solvent accessibility in proteins. Proteins 76:617–636

Zielenkiewicz P, Saenger W (1992) Residue solvent accessibilities in the unfolded polypeptide chain. Biophys J 63:1483–1486

Zuccaro A, Götze S, Kneip S, Dersch P, Seibel J (2008) Tailor-made fructooligosaccharides by a combination of substrate and genetic engineering. Chembiochem 9:143–149

Zucker FH, Stewart C, dela Rosa J, Kim J, Zhang L, Xiao L, Ross J, Napuli AJ, Mueller N, Castaneda LJ, Nakazawa Hewitt SR, Arakaki TL, Larson ET, Subramanian E, Verlinde CLMJ, Fan E, Buckner FS, Van Voorhis WC, Merritt EA, Hol WGJ (2010) Prediction of protein crystallization outcome using a hybrid method. J Struct Biol 171:64–73

CHAPTER 6

RESEARCH RESULTS IV

Sequence-based functional sub-classification of sucrolytic enzymes in glycosyl hydrolase family 32 for the identification of high-level fructooligosaccharide synthesis

6 RESEARCH RESULTS IV:

Sequence-based functional sub-classification of sucrolytic enzymes in glycosyl hydrolase family 32 for the identification of high-level fructooligosaccharide synthesis

K.M. Trollope[1], N. van Wyk[1], M.A. Kotjomela[1], J.F. Görgens[2], H. Volschenk[1]

[1] Department of Microbiology, Stellenbosch University, Private Bag X1, Stellenbosch, 7602, South Africa
[2] Department of Process Engineering, Private Bag X1, Stellenbosch University, Stellenbosch, 7602, South Africa

ABSTRACT

Sucrolytic enzymes hydrolyse sucrose or generate fructose polymers with low degrees of polymerisation (DP) known as fructooligosaccharides (FOS). FOS synthesis capacity differs between studied sucrolytic enzymes. As the consumption of FOS is beneficial to animal and human health, the increased demand has focused attention on novel enzymes able to efficiently convert a commodity sugar to a valuable nutraceutical. Classification of FOS synthesising enzymes based on protein sequence alone would facilitate *in-silico* identification of novel catalysts as large amounts of sequence data lie untapped. This work presents the development of conserved motif sequence logos which enable the rapid identification of enzymes displaying high-level FOS synthesis. New light is shed on the sequence context of active site residues in three previously identified motifs in glycosyl hydrolase family 32 β-fructofuranosidases. Using the proposed WMN**D**PNG, W**E**C and the GQIG**D**PC, F**E**T motifs, the FOS synthesis abilities of the disputed *Aspergillus niger* extracellular β-fructofuranosidase and a putative *Aureobasidium pullulans* β-fructofuranosidase were predicted and experimentally verified. FOS data confirmed the successful prediction of their FOS synthesis capacity. Structural comparison of enzymes displaying low and high-level FOS synthesis identified steric hindrance between nystose (DP4) and a loop region unique to low-level FOS synthesising enzymes. This loop is proposed to limit the synthesis of higher DP FOS, a phenomenon observed among enzymes displaying low-level FOS synthesis. A structural determinant distant to active site residues was thus also an identifier of FOS synthesis capacity. The tool presented may also be useful to improve the understanding of the structure-function relationships of β-fructofuranosidases by facilitating the identification of mutations in groups of enzymes that have been functionally sub-classified.

6.1 INTRODUCTION

β-fructofuranosidases (also known as invertases) catalyse the hydrolysis of terminal non-reducing β-D-fructofuranoside residues in β-D-fructofuranosides such as sucrose. Based on their substrate specificity they are classified according to the nomenclature system of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) as Enzyme commission (EC) number 3.2.1.26 (Myrbäck 1960; Neumann and Lampen 1967). Although valuable in its own right, this classification does not reflect the structural features of enzymes and does not take into account divergent or convergent evolutionary events (Henrissat 1991). In the sequence-based, and hence structural, Carbohydrate-Active Enzymes (CAZy) classification system, β-fructofuranosidases appear in glycosyl hydrolase (GH) family 32. Family GH32 includes a spectrum of enzymes, in addition to β-fructofuranosidase, displaying activities on a broad range of fructose-containing substrates as reflected by their respective EC classifications: endo-inulinase (EC 3.2.1.7), β-2,6-fructan 6-levanbiohydrolase (EC 3.2.1.64), 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 3.2.1.154), 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.-), cycloinulo-oligosaccharide fructanotransferase (EC 2.4.1.-) (Lombard et al. 2014). To date over 2500 enzymes from archaea, bacteria, eukaryota and a single viral candidate have been classified in GH32. Only 242 have been functionally characterised which reflects the impact of next-generation DNA sequencing technologies (NGS). Owing to their similar tertiary structure, catalytic residues and mechanisms of action GH32 is grouped together with family GH68 into clan GH-J (Henrissat and Bairoch 1996). Members of this clan possess a 5-fold β-propeller domain (Alberto et al. 2004); an aspartate residue acting as catalytic nucleophile and a glutamate residue serving as a catalytic proton donor (Pons et al. 2004); and all retain the configuration of the anomeric carbon in the substrate (Koshland and Stein 1954). EC classified activities in family GH68, which only contains archaeal and bacterial representatives, include levansucrase (EC 2.4.1.10), β-fructofuranosidase (EC 3.2.1.26), and inulosucrase (EC 2.4.1.9). Although they share similar enzyme activities, the main difference between family GH32 and GH68 enzymes is that enzymes belonging to the latter do not possess the C-terminal β-sandwich domain (Sainz-Polo et al. 2013). Families GH32 and GH68 together with families GH43 (α-L-arabinases) and GH62 (β-xylosidases) constitute the β-fructosidase superfamily (Naumoff 2001).

Given the vast amounts of genome sequence data being generated by NGS, accurate genome sequence annotation is key to extracting meaningful information from such data. GH32 members can be predicted from sequence is based on conserved domains which have specific signatures. The National Centre for Biotechnology Information (NCBI) makes use of the conserved domain database (CCD) for the functional annotation of protein sequences. For this purpose domain and protein family alignment models from Pfam, SMART, COG, TIGRFAM and the NCBI Protein Clusters database are imported (Marchler-Bauer et al. 2011). The European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI) uses algorithms from InterPro which is an international consortium of databases which include PROSITE, HAMAP, Pfam, PRINTS, ProDom, SMART, TIGRFAM, PIRSF, SUPERFAMILY, CATH-Gene3D and PANTHER (Hunter et al. 2012). Functional annotation of GH32 enzymes is based on the identification of the N-terminal β-propeller domain and the C-terminal β-sandwich domains (Alberto et al. 2004) with specific emphasis on the active site residues, situated in the N-terminal domain. A vast number of arrangements of these domains in combination with each other and other domains, often of unknown function (DUF), are possible (Finn et al. 2014). Both NCBI and EMBL-EBI use the Pfam domain signatures PF0251 and PF08244 for the GH32 N- and C- terminal domains, respectively. These domain signatures are represented by profile hidden Markov models (HMM) for protein families built from an aligned set of curator-defined family-representative sequences (Finn et al. 2014). Profile HMMs are probabilistic models used for the statistical inference of homology (Eddy 1998; Krogh et al. 1994). A high-quality seed alignment provides the basis for the position-specific amino-acid frequencies, gap and length parameters in the profile HMM.

At sub-domain level, conserved motifs have been identified in GH32 proteins with Gunasekaran *et al.* (1990) being the first to report well conserved A–F motifs in the N-terminal region. These motifs contain identical or similar amino acids and were proposed to be important for expression of enzyme activity. An additional B1 region was included by Pons et al. (1998) which included a predicted α-helix around which residues specific to levanase, yeast, and plant invertases were located. Region G was identified in the C-terminal region of *A. niger* inulinases (Ohta et al. 1998) and *B. subtilis* levanase (Martin et al. 1987). This motif was proposed to be involved in the binding of high molecular weight fructans, as it was absent in enzymes hydrolysing sucrose and raffinose. Motifs A, D and E were shown to harbour active site residues (Pons et al. 2004; Reddy and Maley 1996). Comparison of the Pfam HMM logo (Fig. 6.3) for the GH32 N-terminal domain (EMBL-EBI, 2013) to the A, D and E motifs of relatively large GH32 multiple sequence alignments (MSAs) (Alméciga-Díaz et al. 2011; Pons et al. 2004; Yuan et al. 2006) highlighted the importance of these regions in the functional annotation of GH32 sequences. Results from early studies with relatively small sequence datasets thus are reflected, to some extent, in more

recent, comprehensive bioinformatic evaluations The consensus sequences for the A, D and E motifs from the HMM logo are WMN**D**PNG, FR**D**P and MW**E**CPDF, with the boldface letters representing the active site residues.

Identification of conserved domains functionally annotates sequences to the family level with further analyses required to identify substrate specificities. The NC-IUBMB Enzyme Commission (EC) nomenclature system (Dixon and Webb 1958; McDonald and Tipton 2014) elegantly complements the CAZy classification system in that regard. Experimental evidence is the basis for the NC-IUBMB classification which in turn informs the identification of domain signatures and smaller sequence motifs. GH32 enzymes are well characterised experimentally due to their function in sucrose metabolism which in itself has received much research attention. Of particular interest is the production of fructooligosaccharides (FOS) which are important functional food ingredients (Roberfroid 2007). FOS can be synthesised by the enzymatic transfer of successive fructose units to sucrose, the initial acceptor molecule. Enzymes of fungal origin are the preferred FOS producers as only a single enzyme is required, as opposed to two enzymes as found in higher plants (Gadegaard et al. 2008; Lammens et al. 2009; Tamura et al. 2009). Also, the production yield of FOS by plant enzymes is low in comparison to fungal FOS producing enzymes (Yun 1996). Alternatively, FOS can be derived from the enzymatic hydrolysis of inulin, a fructose-rich plant storage polysaccharide. Both endo- and exo-acting inulinases are included in family GH32. Commercial production of inulin-type FOS from sucrose is performed either by whole cell biotransformation or enzymes originating from *Aspergillus* spp. and *Aureobasidium pullulans* (Hidaka et al. 1988; Hirayama et al. 1989; Maiorano et al. 2008; Yun 1996). Inulin-type FOS are characterised by β(2→1) linkages between the sugar monomers (Waterhouse and Chatterton 1993). Inconsistencies with regards to the nomenclature of FOS synthesising enzymes exist. The terms *fructosyltransferase* (EC 2.4.1.9), *invertase* or *β-fructofuranosidase* (EC 3.2.1.26) are used because in addition to sucrose hydrolysis, some β-fructofuranosidases exhibit transferase activity at high substrate concentrations which results in the formation of FOS (Edelman 1956; Hirayama et al. 1989; Yun 1996).

Phylogenetic analyses of family GH32 enzymes (Alméciga-Díaz et al. 2011) and specifically fungal candidates have been conducted (Parrent et al. 2009; Yuan et al. 2006). Alméciga-Díaz et al. (2011) identified seven clades for GH32 *fructosyltransferase* enzymes with five representing plant enzymes and one clade each for fungal and bacterial sequences. The fungal clade comprised two branches. They proposed that fructosyltransferases from fungi and bacteria likely evolved from dicotyledonous fructosyltransferases. The analyses of Parrent et al. (2009) identified 9 lineages of GH32 fungal genes. Their analyses were broader than those of

Alméciga-Díaz et al. (2011) in the sense that they included sequences for inulinases as well as *invertases*. In the case of invertases they identified intra- and extracellular invertases, with the extracellular invertases split in two groups. Yuan et al. (2006) examined the inulin-modifying enzymes of *A. niger* in the context of functionally characterised GH32 enzymes from yeast and fungi. They arrived at a similar phylogenetic separation of the sequences with clades representing exo-inulinases, endo-inulinases and invertases. The latter enzyme class was represented in two branches harbouring yeast and filamentous fungal invertases, respectively.

NGS technologies are increasingly enabling the generation of whole genome sequence data on a routine basis. Genomes are sequenced and annotated to investigate a myriad of biological questions and the spin-off of these undertakings is the public availability of vast amounts of unmined data with regard to other research questions. In the post-genomic era bioprospecting has progressed from conventional culture-based methods, with novel sequences easily obtained from public sequence repositories. The ability to identify enzymes with a preferred activity directly from a predicted protein sequence would greatly facilitate *in silico* bioprospecting for novel catalysts. Sequence annotation to the CAZy family or the fourth EC digit level (substrate specificity) (Nagao et al. 2014) is already a reality for new sequences displaying homology to previously classified sequences. This level of annotation however, is still unable to specify the efficiency of the enzyme.

In the context of the health and commercial importance of FOS, using a data integration approach a means was developed to functionally subclassify sucrolytic enzymes (EC 3.2.1.26 or 2.4.1.9), based on protein sequence alone, as exhibiting either low- or high-level FOS synthesis activity. Otherwise stated as enzymes having either predominantly hydrolytic or good fructosyltransferase activity. This classification method was developed by the reappraisal of previously identified clades and conserved sequence motifs for GH32 enzymes in light of published experimental data. Using this method, numerous uncharacterised enzymes were identified from public databases that can serve as departure point for testing novel oligosaccharide synthesising biocatalysts. The validity of these findings were tested experimentally on two enzymes predicted by the model to belong to each of the identified functional sub-groups, viz. low- and high-level FOS synthesis. Experimental data provided clarity on the activity of the *A. niger* extracellular β-fructofuranosidase. Structural comparison of enzymes belonging to each of the sub-groups was also performed.

6.2 Materials and methods

6.2.1 GH32 sequences

Protein sequences for GH32 enzymes were obtained from the Protein database on the NCBI server (http://www.ncbi.nlm.nih.gov/). Sequences for functionally characterised fungal invertases/β-fructofuranosidases/fructosyltransferases were obtained using the published accession numbers (Table 6.1). Protein sequences of all deposited fungal GH32 enzymes were retrieved from the Protein database (accessed in June 2014) using *GH32* as the search term. The species *fungi* and Source databases *RefSeq* filters were applied and 566 sequences were returned. Sequences shorter than 200 amino acids were removed as generally GH32 β-fructofuranosidases range from 500 to 700 amino acids. Furthermore, sequences displaying a redundancy greater than 99% were removed in Jalview (Waterhouse et al. 2009). The remaining 258 sequences served as the dataset for analysis of the fungal GH32 enzymes. Their accession numbers are listed in supplementary Table S8. Another dataset was obtained from the fungal phylogenomic complete genome database (FunyCG) on the FUNYBASE server (http://genome.jouy.inra.fr/funybase/) (Marthey et al. 2008) using the search term *fructofuranosidase*. The curated homology derived secondary structure of proteins (HSSP) databank (Sander and Schneider 1994) dataset for sequences displaying homology to *Aspergillus japonicus* fructosyltransferase crystal 3LF7 (Chuankhayan et al. 2010) was obtained from the server (http://mrs.cmbi.ru.nl/m6/) in June 2014.

6.2.2 Sequence alignments and phylogenetic trees

MSAs were performed using the MUSCLE algorithm (Edgar 2004) in CLC Main Workbench version 7.0.3 (Qiagen, Venlo, Netherlands). The MSAs were manually adjusted to correct gap placement. Phylogenetic trees were constructed using the Neighbour-joining method (Saitou and Nei 1987) with 1000 bootstraps and distances corrected by the Jukes-Cantor model (Jukes and Cantor 1969). In the analyses involving experimentally characterised enzymes, two bacterial invertase GH68 sequences were used as an outgroup – *Streptococcus mutans*, GenBank accession number AAA88584.1 and *Bacillus megaterium*, GenBank accession number ADF38395.1.

Table 6.1 GH32 enzymes experimentally characterised with respect to fructooligosaccharide synthesis

Enzyme	NCBI Accession number	Organism	Substrate concentration	Time	Maximum percentage FOS in reaction mixture	Reference
SucA β –fructofuranosidase	CAK45778.1* ABB59678.1	*Aspergillus niger* strain AB1.13	10%	1 h	56% FOS ($GF_{2,3}$)	Yuan et al., 2006; Zuccaro, Götze, Kneip, Dersch, & Seibel, 2008
Suc1 β-fructofuranosidase precursor	S33920	*Aspergillus niger* B60	20%	2 h	61%	Boddy et al., 1993 This work
Sequence 1	AAQ31829.1	*Scopulariopsis brevicaulis* IFO4843	Not reported in patent – assumed to be similar to *P. roquefortii*	Not reported in patent – assumed to be similar to *P. roquefortii*	Not reported in patent – assumed to be similar to *P. roquefortii*	Yanai, Nakane, & Toshiaki, 2003
Sequence 3	AAQ31828.1	*Penicillium roquefortii* IAM7254	60 wt % sucrose	23 h	39.4% ($GF_{2,3}$)	Yanai, Nakane, & Toshiaki, 2003
F1 β-fructofuranosidase	BAE63792.1 XP_001824925 AP007167	*Aspergillus oryzae* KB	20% sucrose	24 h	58.7% FOS ($GF_{2,3,4}$)	Kurakake et al., 2008; Machida et al., 2005
F2 β-fructofuranosidase	BAE62112.1 XP_001823245 AP007164–1378	*Aspergillus oryzae* KB	20% sucrose	24 h	8.6%	Kurakake et al., 2008; Machida et al., 2005
fopA β-fructofuranosidase	AB046383.1	*Aspergillus japonicus***	50% sucrose	24 h	~60%	Hidaka, Hirayama, & Sumi, 1988; Hirayama, Sumi, & Hidaka, 1989; Yanai, Nakane, Kawate, & Hirayama, 2001
sucB invertase	ABB59679	*Aspergillus niger* N402	34.2%	12 h	Not reported for TLC - traces of GF_2	Goosen et al., 2007
SFT fructosyltransferase	CAB89083.1	*Aspergillus sydowii* IAM 2544	Not reported	1 week	Not reported – FOS up to DP 10	Heyer & Wendenburg, 2001
Suc2 invertase	NP_012104.1	*Saccharomyces*	60%	24-48 h	2.5% FOS (GF_2)	Lafraya, Sanz-

Enzyme	NCBI Accession number	Organism	Substrate concentration	Time	Maximum percentage FOS in reaction mixture	Reference
		cerevisiae S288C				Aparicio, Polaina, & Marín-Navarro, 2011; Taussig & Carlson, 1983
β-fructofuranosidase	ADN34605	*Schwanniomyces occidentalis* ATCC 26077	34.2%	6 h	6% FOS (GF_2)	Alvaro-Benito, de Abreu, Portillo, Sanz-Aparicio, & Fernández-Lobato, 2010
β-fructofuranosidase	ACL79833.1	*Xanthophyllomyces dendrorhous* ATCC MYA-131	41%	50 h	16%	Linde et al., 2009
β-fructofuranosidase	AGV22100.1	*Ceratocystis monoliformis* CMW 10134	1.7%	1 h	Not reported. Hydrolysis activity dominated.	van Wyk, Trollope, Steenkamp, Wingfield, & Volschenk, 2013
β-fructofuranosidase	JGI *Aureobasidium pullulans* var. subglaciale EXF-2481 v1.0 scaffold_16:248726-250621 Protein ID 31127	*Aureobasidium pullulans* CBS 12338	20%	3 h	11.5%	This work Nordberg et al., 2014
Sucrose:Sucrose 1-Fructosyltransferase	AJ000493.1	*Aspergillus foetidus* NRRL 337	34.2%	1 h	Not reported. Hydrolysis activity dominated.	Rehm, Willmitzer, & Heyer, 1998

* Accession number for the 1 exon protein however, the 2 exon DNA was cloned and expressed. The protein is assumed to be comprised of 2 exons after *A. niger* mRNA processing.
** Previously classified as *Aspergillus niger* ATCC 20611

6.2.3 Sequence alignment logos

The seed alignment for the GH32 N-terminal domain (PF00251) was obtained from EMBL-EBI (http://pfam.xfam.org/family/PF00251#tabview=tab3, 24/07/2014). A subset of 32 sequences displaying the GQIG**D**PC and F**E**T motifs was extracted from the 258 fungal GH32 sequences obtained from the NCBI server. Skyline server was used to generate alignment logos (Wheeler et al. 2014). Logos were generated such that positive scoring letters i.e. above background frequency were shown and the letter height in each stack represented estimated letter probability. Alignment processing was selected to use observed counts – 100% identical sequences were removed prior to logo building.

6.2.4 Microbial strains, media and culturing conditions

All microbial strains used in this study are listed with their relevant genotypes in Table 6.2. *Escherichia coli* was used for cloning and amplification of plasmids. *A. pullulans* and *A. niger* strains were used for the isolation of their β-fructofuranosidase encoding genes. *A. niger* was kindly provided by Petra Dersch (Technical University of Braunschweig, Germany). The invertase-deficient yeasts *Saccharomyces cerevisiae* and *Komagataella (Pichia) pastoris* DSMZ 70382 (Sreekrishna et al. 1987) served as hosts for the expression of the β-fructofuranosidases. *E. coli* cells were grown at 37 °C in Luria Bertani broth (BioLab, Midrand, South Africa) supplemented with 100 µg/ml ampicillin or 50 µg/ml kanamycin, as appropriate (Sambrook et al. 1989). *A. pullulans* was cultivated in medium containing 50 g/l glucose, 5 g/l yeast extract, 5 g/l NH_4SO_4, 1 g/l NaCl, 5 g/l KH_2PO_4, 0.2 g/l $MgSO_4$ (Cheng et al. 2011) at 30 °C with shaking at 200 rpm for 72 h. *S. cerevisiae* was cultivated aerobically, shaking at 200 rpm, at 30 °C in rich YPD medium containing 20 g/l glucose, 10 g/l yeast extract and 20 g/l peptone and or in synthetic minimal medium containing 20 g/l glucose, 6.7 g/l yeast nitrogen base without amino acids (Difco Laboratories, Detroit, USA), 1.3 g/l amino acid pool without uracil (Ausubel et al. 2001). For enhanced enzyme production *S. cerevisiae* was cultivated for 72 h in double strength minimal medium containing 20 g/l glucose, 13.4 g/l yeast nitrogen base without amino acids, 2.6 g/l amino acid pool without uracil. The medium was buffered with 20 g/l succinic acid with the pH adjusted to 6.0 using 10 N NaOH. *K. pastoris* was cultivated aerobically, with shaking at 200 rpm, at 30 °C in rich YPD medium supplemented with 50 ug/ml zeocin, as appropriate. At the enzyme production phase *K. pastoris* was cultivated for 48 h. Solid media contained 20 g/l agar (BioLab, Midrand, South Africa).

Table 6.2 Microbial strains and plasmids used in this study

Strain or plasmid	Relevant genotype or construct	Reference or source
Bacterial strains		
Escherichia coli DH5α	fhuA2Δ(argF-lacZ)U169 *phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs, Midrand, South Africa
Fungi		
Aureobasidium pullulans var. subglaciale CBS 12338	Wild isolate	(Zalar et al. 2008)
Aspergillus niger SKAn1015	*prt-13 sucA pyrG+*	(Zuccaro et al. 2008)
Yeast strains		
Saccharomyces cerevisiae EUROSCARF Y02321	BY4741 *Mat a his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 YIL162w(SUC2)::kanMX4*	(Brachmann et al. 1998)
Komagataella pastoris DSMZ 70382 (*Pichia pastoris*)	Type strain	(Yamada et al. 1995)
Yeast transformants		
K. pastoris pGAPZ*fopA*	*Sh ble* GAP_p*-xln2-fopA-*$AOX1_T$	This study
K. pastoris pGAPZ*ΔfopA598*	*Sh ble* GAP_p*-xln2-*Δ*fopA598-*$AOX1_T$	This study
K. pastoris pGAPZ*suc1*	*Sh ble* GAP_p*-suc1-*$AOX1_T$	This study
K. pastoris pGAPZ*sucA*	*Sh ble* GAP_p*-sucA-*$AOX1_T$	This study
S. cerevisiae pJC1*fopA*	*Mat a his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 YIL162w(SUC2)::kanMX4 [*$PGK1_P$*-xln2-fopA-*$PGK1_T$*]*	This study
S. cerevisiae pJC1*ApINV*	*Mat a his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 YIL162w(SUC2)::kanMX4 [*$PGK1_P$*-ApINV-*$PGK1_T$*]*	This study
Plasmids		
pJC1	*bla URA3* $PGK1_P$*-*$PGK1_T$	(Crous et al. 1995)
pGAPZαB	*Sh ble* GAP_p*-mfα-*$AOX1_T$	Invitrogen, Life Technologies, Thermo Fisher Scientific, Waltham, USA
pJ901*fopA*	*kanMX* $AOX1_P$*-xln2-fopA-*$AOX1_T$	This study; DNA 2.0 (Menlo Park, USA)
pGAPZ*fopA*	*Sh ble* GAP_p*-xln2-fopA-*$AOX1_T$	This study
pJ227*ΔfopA598*	*kanMX-xln2-*Δ*fopA598*	This study; DNA 2.0 (Menlo Park, USA)
pGAPZ*ΔfopA598*	*Sh ble* GAP_p*-xln2-*Δ*fopA598-*$AOX1_T$	This study
pGAPZ*suc1*	*Sh ble* GAP_p*-suc1-*$AOX1_T$	This study
pGAPZ*sucA*	*Sh ble* GAP_p*-sucA-*$AOX1_T$	This study
pJC1*fopA*	*bla URA3* $PGK1_P$*-xln2-fopA-*$PGK1_T$	This study
pJet*ApINV*	*bla ApINV*	This study, Fermentas, Thermo Fisher Scientific, Waltham, USA
pJC1*ApINV*	*bla URA3* $PGK1_P$*-ApINV-*$PGK1_T$	This study

mfα – *S. cerevisiae* mating factor alpha secretion signal sequence (first 89 amino acids)
xln2 – *Trichoderma reesei* endoxylanase II secretion signal sequence (33 amino acids)

6.2.5 DNA manipulations and β-fructofuranosidase cloning

All DNA manipulations were performed according to standard methods (Sambrook et al. 1989). Restriction enzymes and T4 DNA ligase were used according the specifications of the supplier (Fermentas, ThermoScientific, Waltham, USA). The plasmids used and generated in this study are listed in Table 6.2. All genes cloned by PCR were sequenced to verify their integrity (Sanger et al. 1977). The *fopA* gene (AB046383) was cloned from *A. niger* ATCC 20611 by Yanai et al. (2001) but the organism was later reclassified as *A. japonicus* by the curators of the ATCC culture collection (http://www.lgcpromochem-atcc.com). An unpublished sequence deposited at GenBank (GU356596) for the *A. japonicus* CB05 fructosyltransferase displays 99% sequence identity to *fopA* on the DNA and protein levels. To prevent the perpetuation of the aforementioned inaccuracy, the *fopA* gene product will be referred to in this study as the *A. japonicus* β-fructofuranosidase. The corresponding β-fructofuranosidase gene (GenBank accession number AB046383) was codon optimised and synthesised by DNA 2.0 (Menlo Park, USA). The native secretion signal was replaced by the *Trichoderma reesei* endoxylanase II (*xln2*) secretion signal (Saarelainen et al. 1993). The construct was subcloned as a 2,033 bp *Eco*RI-*Xho*I fragment from the pJ901 IP-Free© *P. pastoris* expression vector (DNA 2.0, Menlo Park, USA) to the *Eco*RI-*Xho*I digested pJC1 *E. coli–S. cerevisiae* shuttle vector (Crous et al. 1995). For expression in *K. pastoris* the same fragment was subcloned into the same sites of a modified pGAPZαB plasmid (Life Technologies) in which the mating factor alpha sequence was removed. A truncated version of the *fopA* gene fused to the *xln2* secretion signal was synthesised by DNA 2.0 and shipped in the pJ227 cloning vector. The truncated version was designated Δ*fopA598* and encoded a protein 38 amino acids shorter on the C-terminal end than the full length protein. Position 598 was represents the 598th amino acid downstream from the mature protein start i.e. the numbering excludes amino acids encoded by the native secretion signal. Similarly, *Δfop*A598 was cloned as an *Eco*RI-*Xho*I fragment into the modified pGAPZαB.

Two versions of the *A. niger* β-fructofuranosidase were cloned from genomic DNA isolated from a spore suspension. A 1770 bp sequence was amplified by PCR from genomic DNA using the Suc1-L and Suc1-R primer pair (Table 6.3). This coding sequence was described by Boddy et al. (1993) and comprises one exon. It was named *suc1*. Re-examination of the genome sequence contig An08c0280 containing *suc1* (GenBank accession AM270183.1) by Yuan et al. (2006) revealed a 55 base pair intron at the 3' end of the open reading frame. This was also confirmed by REFSEQ annotation XM_001393172. A construct joining the 2 exons (i.e. removing the intron) was synthesised by PCR using the primer pair Suc1-L and Suc1_wo_intron-R (Table 6.3) and the cloned s*uc1* as template. The Suc1_wo_intron-R primer encoded the second exon and was attached to the first exon (*suc1*) during PCR. In line with the

nomenclature used in the published literature this sequence was named s*ucA*. Both versions were cloned into the modified pGAPZ at the *Eco*RI-*Xho*I sites.

A putative β-fructofuranosidase from *A. pullulans* was cloned from genomic DNA by PCR with the primers ApINV-L and ApINV-R. The jgi|Aurpu_var_sub1|31127|fgenesh1_pg.16_#_95 JGI accession number was used. The gene was cloned as an *Eco*RI-*Xho*I fragment into the same sites in pJC1.

All the expression plasmids generated in this study placed the β-fructofuranosidase genes under the control of constitutive promoters – the *S. cerevisiae* plasmids contained the phosphoglycerate kinase (*PGK1*) promoter while the *K. pastoris* plasmids contained the glyceraldehyde-3-phosphate dehydrogenase (*GAP*) promoter. Furthermore, all the plasmids contained secretion signals to facilitate extracellular protein production.

6.2.6 Microbial transformations

Bacterial transformation was performed according to standard methods (Sambrook et al. 1989). A lithium acetate-DMSO procedure (Hill et al. 1991) was used for the transformation of *S. cerevisiae* while the condensed competent cell preparation method and transformation by electroporation (Lin-Cereghino et al. 2005) was used to generate *K. pastoris* transformants. The transformants were verified by PCR with gene specific primers or vector specific primers (Table 6.3). The pGAP forward and 3'AOX1 primer pair was used to screen *K. pastoris* pGAPZ*fopA* and pGAPZΔ*fopA598* transformants while the pMBRE4-1'F and pMBRE4-3'R pair was used to screen *S cerevisiae* pJC1*fopA* transformants.

6.2.7 Protein Electrophoresis and zymograms

Yeast supernatants were subjected to 8% SDS-PAGE. Loading dye consisted of 60 mM Tris-HCl (pH 6.8), 25% glycerol, 2% SDS, 14 mM β-mercaptoethanol and bromophenol blue and gels were run in Tris-glycine buffer (25 mM Tris-HCl, 250 mM glycine, 0.1% SDS). After electrophoresis, protein bands were visualized with a silver-stain (Gallagher and Sasse 2012). For zymogram analyses, following electrophoresis, gels were washed in 0.05 M citrate-phosphate buffer pH 6.0 for 1 h to remove SDS. Gels were then incubated at 60 °C in the presence of 0.05 M sucrose for 15 minutes. After the removal of the sucrose solution, gels were placed in a boiling 100 mM NaOH solution containing 0.2% triphenyltetrazolium chloride (Sigma-Aldrich, St. Louis, USA) (Chaira et al. 2011).

Table 6.3 A list of the primers used in this study

Primer Name	Restriction Enzyme site*	Sequence 5' to 3'
Suc1-L	*Pac*I *Eco*RI	actgTTAATTAAGAATTCatgaagcttcaaacggcttccg
Suc1-R	*Asc*I *Xho*I	tcGGCGCGCCTCGAGttaagactgacgatccggccaag
Suc1_wo_intron-R	*Xho*I	acgtCTCGAGttaagactgacgatccggccaagcatcatacaa gccctcggaaacggtcacgttggagaacgcaacaccacccacgc cattctggaagaaactgatgttcgtggaattggcgtaccaagaacga acccaagtactcaac
SucA-R	*Pac*I *Xho*I	cgTTAATTAACTCGAGttaagactgacgatccggccaag
ApINV-L	*Pac*I	atgcTTAATTAAatgaagagctcccttgctc
ApINV-R	*Asc*I	tgacGGCGCGCCttagttaccagtccaag
pGAP forward	-	gtccctatttcaatcaattgaa
3´ AOX1	-	gcaaatggcattctgacatcc
pMBRE4-1'F	-	gtttagtagaacctcgtgaaactta
pMBRE4-3'R	-	acttaaaatacgctgaacccgaacat

* Restriction enzyme recognition sequences are given in uppercase.

6.2.8 Enzyme assays

As an initial indicator of enzyme functionality, yeast transformants were streaked on solid, minimal media containing 20 g/l sucrose to evaluate the ability of yeast to use sucrose as sole carbon source. Assays to determine International units (IU) and FOS production were performed as described by van Wyk et al. (2013). Briefly, IU of enzyme was defined as μmol of product released per minute following incubation of enzyme with 0.05 M sucrose in 0.05 M citrate phosphate buffer. The enzymes produced by *K. pastoris* were incubated at pH 5.0 at 50 °C for five minutes while the enzymes produced by *S. cerevisiae* were incubated at pH 6.0 at 57.5 °C. Reducing sugars were determined using the 3,5-dinitrosalicyclic acid (DNS) method (Miller 1959). Assays to evaluate FOS production by the recombinant β-fructofuranosidases were conducted with 0.58 M sucrose in 0.05 M citrate-phosphate buffer for 2 h. Reactions were dosed with one IU of enzyme. Enzymes produced by *K. pastoris* were incubated at pH 5.0 at 55 °C while enzymes produced by *S. cerevisiae* were incubated at pH 6.0 at 57 °C. Following appropriate dilution, reaction products were quantified with high performance liquid chromatography as described by van Wyk et al. (2013).

6.2.9 Three dimensional structure alignments

The crystal structure models for *S. cerevisiae* Suc2 invertase (4EQV) and *A. japonicus* fructosyltransferase complexed with nystose (3LEM) were obtained from the RCSB Protein Data Bank (http://www.rcsb.org/pdb/home/home.do). Structural alignments were performed in USCF Chimera version 1.7, with solvent-excluded molecular surfaces created by the MSMS package (Pettersen et al. 2004; Sanner et al. 1996). The alignment for superimposition was generated using the best-aligning pair of chains between reference and match structures using the Needleman-Wunsch alignment algorithm and the BLOSUM-62 matrix. Secondary structure scores were weighted to 30%. Homology modelling of the ApINV was performed using *S. cerevisiae* invertase (4EQV) as template. Models were generated by the SWISS-MODEL web server (Arnold et al. 2006). Template and target sequences shared 33.7% identity.

6.3 RESULTS

6.3.1 Phylogenetic analyses and identification of conserved sequence motifs

Table 6.1 summarises the search results for experimentally characterised fungal GH32 enzymes for which the corresponding protein sequences have been reported. Only enzymes acting on sucrose were included i.e. β-fructofuranosidases and fructosyltransferases, while inulinases were excluded. A phylogenetic tree was constructed from the MSA of these protein sequences revealed distinct groupings of the sequences (Fig. 6.1).

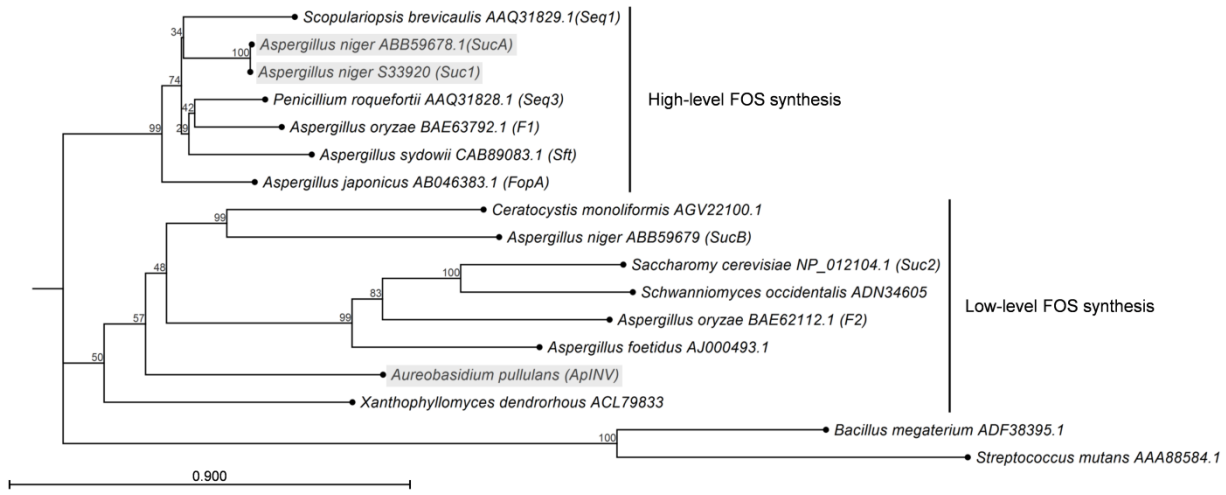

Fig. 6.1 Neighbour-joining phylogeny of experimentally characterised fungal GH32 enzymes displaying sucrolytic activity. GenBank protein accession numbers are given with trivial names indicated in brackets, where available. Bootstrap support values based on 1000 replicates are given as percentages above individual branches. Enzymes highlighted in grey were experimentally evaluated in this study

As expected, the bacterial GH32 and GH68 sequences formed an outgroup clade with the fungal sequences grouping within two distinct clades. Enzymes that have been shown to produce reaction mixtures containing in excess of ca. 20% FOS of total sugars grouped together, while the second group was comprised of enzymes producing limited amounts of FOS. The cut-off of 20% was an arbitrary assignment from the reported FOS levels – it served to create the distinction between low and high amounts of produced FOS.

Examination of the MSA used to generate the tree indicated that the sequences for the two groups differed in the active site residue-containing A and E motifs. The sequences of the enzymes exhibiting high-level FOS synthesis contained the GQIG**D**PC and F**E**T motifs while the enzymes exhibiting low-level FOS synthesis displayed the WMN**D**PNG and **E**CP motifs characteristic of the GH32 N-terminal domain HMM (active site residues indicated in bold). To exclude the possibility that the observed grouping in Fig. 6.1 was an artefact of sequence sampling size, phylogenetic analyses of all 258 fungal GH32 enzymes from the NCBI repository were performed. Known enzymes exhibiting high-level FOS synthesis were located and the sequences of immediate neighbours in the tree were investigated for the GQIG**D**P and F**E**T motifs. It emerged that all sequences displaying the GQIG**D**P and F**E**T motifs (or highly similar sequences) were grouped together as indicated in Fig. 6.2. These 32 protein sequences were extracted and realigned.

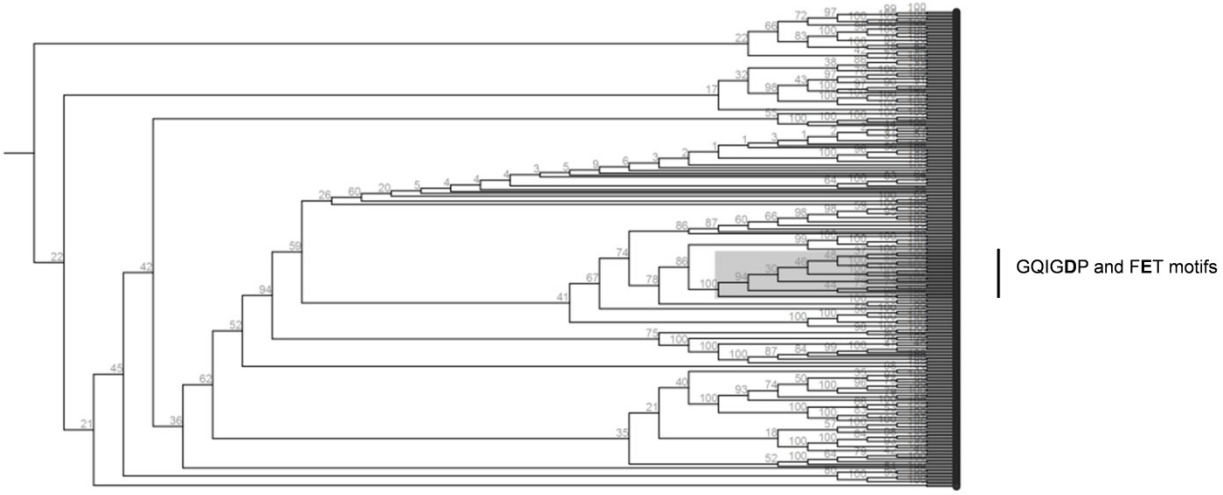

Fig. 6.2 Neighbour-joining phylogeny of fungal GH32 enzymes. Sequences displaying GQIGDP and FET motifs in the highly conserved active site residue-containing A and E motifs are highlighted in the grey block. Bootstrap support values based on 1000 replicates are given as percentages above individual branches

Alignment logos for the active site residue-containing motifs A, D and E were generated and are shown in Fig. 6.3. There were clear differences between the GH32 HMM A and E motifs and the group displaying putative high-level FOS synthesis. Motif D was relatively similar between the two groups. Putative signature motifs were thus identified for enzymes displaying high-level FOS synthesis. Other databases were further searched for β-fructofuranosidases with the

identified signature motifs and a list of novel enzymes putatively displaying high-level FOS synthesis is given in Table S8. This list serves as a rational departure point for bioprospecting for novel FOS synthesising enzymes.

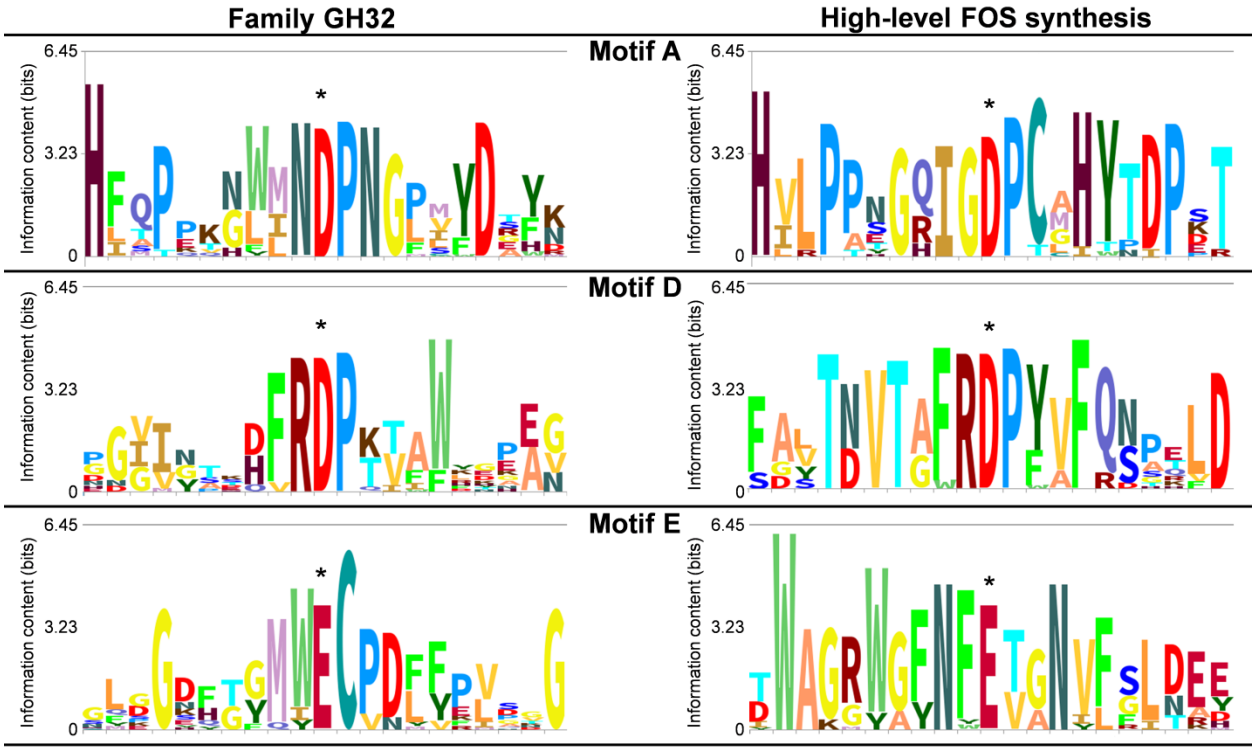

Fig. 6.3 Sequence logos for the conserved motifs harbouring active site residues in the profile hidden Markov model of family GH32 enzymes and a group displaying putative high-level FOS synthesis. Alignments used to build the logos were obtained from the EMBL-EBI webpage for Pfam domain 00251 and a subset of fungal GH32 sequences obtained from the NCBI server that were aligned by the MUSCLE algorithm, respectively. The Skylign server generated the sequence logos such that stack height represented the information content of the alignment position. Positive scoring letters i.e. above background frequency were shown and the letter height in each stack represented estimated letter probability. The catalytic residues are denoted by the asterisks

6.3.2 Heterologous production of two fungal β-fructofuranosidases

To test the hypothesis that the phylogenetic grouping of fungal sucrolytic GH32 enzymes was related to FOS synthesis activity, two β-fructofuranosidases whose protein sequences were proposed to represent low- and high-level FOS synthesis activities were functionally characterised. The prediction was based on the presence of either the WMNDPNG and ECP or the GQIGDPC and FET sequence motifs in motifs A and E, respectively. The β-fructofuranosidase from *A. niger* (Suc1 and SucA) was predicted to be a high-level FOS producer while the β-fructofuranosidase from *A. pullulans* (ApINV) was predicted to be a low-level FOS producer. Their positions in the phylogenetic tree of functionally characterised enzymes are highlighted in grey (Fig. 6.1). The activities of these β-fructofuranosidases were

compared to the *A. japonicus* FopA, the enzyme used for commercial production of FOS and used in this study as a benchmark for high-level FOS synthesis (Nishizawa et al. 2001).

6.3.2.1 A. niger *β-fructofuranosidases*

Two versions of the *A. niger* β-fructofuranosidase were heterologously expressed in *K. pastoris*. The reasoning for this follows. The original report by Boddy et al. (1993) on the characterisation of the purified enzyme and its gene sequence indicated that the *suc1* gene contained a single exon. Furthermore, they did not investigate FOS synthesis by the purified enzyme. Yuan et al. (2006) predicted the presence of an intron one base pair upstream of the termination codon. They verified the presence of the intron by reverse transcriptase PCR and hence predicted an additional 39 extra amino acids on the C-terminal end of Suc1. The presence of the intron was also previously proposed by Yanai et al. (2001) but neither groups examined the enzyme activity experimentally. Based on the report by Yanai et al. (2001), Yuan et al. (2006) speculated that SucA (comprised of two exons) lacked fructosyltransferase activity and that fopA displayed much higher fructosyltransferase activity than Suc1 from *A. niger* strain B60. Zuccaro et al. (2008) produced oligosaccharides using Suc1, although it was deduced from their cloning strategy that they were using the two exon enzyme (SucA). Furthermore, they claimed that Suc1 possessed superior FOS synthesis activity to FopA even though they did not directly compare the activities of the two enzymes. To obtain clarity on the activity of the *A. niger* extracellular β-fructofuranosidase, which the bioinformatic analysis predicted to have high-level FOS synthesis abilities, both Suc1 and SucA were produced in *K. pastoris*. To facilitate a direct comparison FopA was also produced in *K. pastoris*. Finally, a truncated version of FopA (ΔFopA598) was produced such that it resembled Suc1 to allow for investigation of the role of the C-terminal region (encoded by the second exon in *sucA*) in enzyme activity. The full length *fopA* coding sequence did not contain an intron (Yanai et al. 2001).

Fig. 6.4 shows the results of the SDS-PAGE (A) and zymogram analysis (B) of the culture supernatants of *K. pastoris* expressing the aforementioned enzymes. On the SDS-PAGE gel photograph bands were visible in the region of 150 kDa that corresponded to the enzymes. A band at that size was absent for the yeast transformed with the empty pGAPZ vector. The bands were larger than the theoretical predicted weights for the mature proteins of 66 kDa (SucA) and 69 kDa (FopA), however de-*N*-glycosylation with PNGase F yielded the expected band sizes (data not shown). The bands for Suc1 and ΔFopA598 were slightly lower than SucA and FopA, owing to the C-terminal truncations. Hydrolytic activity was evident on the zymogram between 130 and 170 kDa for Suc1, SucA and FopA. The ΔFopA598 enzyme was deemed

inactive as no band was visible on the zymogram. In addition, the *K. pastoris* pGAPZ*ΔFopA598* transformants did not display growth on solid media with sucrose as sole carbon source.

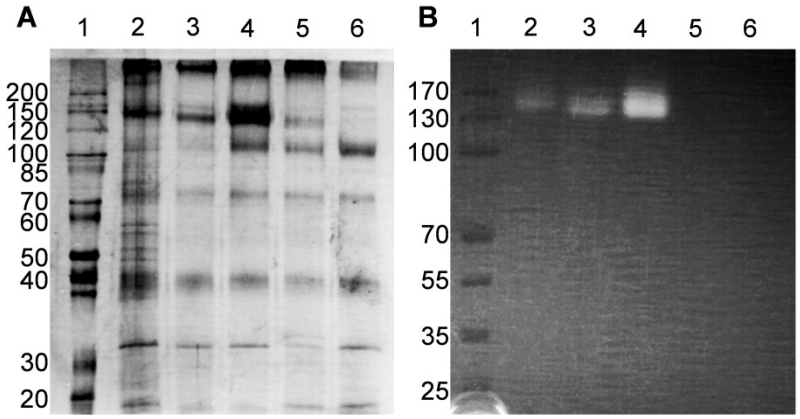

Fig. 6.4 SDS-PAGE and zymogram gels of Suc1, SucA, FopA and ΔFopA598. **A** Silver-stained 8% polyacrylamide gel of culture supernatants of *K. pastoris* expressing the four β-fructofuranosidases. Lane 2 shows Suc1, Lane 3 SucA, Lane 4 FopA, Lane 5 ΔFopA598 and Lane 6 contains the control strain transformed with the empty pGAPZ vector. **B** Zymogram gel of the four β-fructofuranosidases loaded in the same order as for SDS-PAGE. In both gels, Lanes 1 contained the protein molecular weight markers. The PageRuler unstained protein ladder and the PageRuler prestained protein ladder (Thermo Scientific) were used in A and B, respectively

6.3.2.2 A. pullulans *β-fructofuranosidase*

A putative β-fructofuranosidase was cloned from *A. pullulans* and expressed in *S. cerevisiae*. As in the section above, *fopA* was also expressed in *S. cerevisiae* to facilitate a direct comparison of enzyme activities. The sequence of the cloned *ApINV* was 99.9% identical to jgi|Aurpu_var_sub1|31127|fgenesh1_pg.16_#_95 on the DNA level. Translation of the cloned gene sequence revealed two amino acid substitutions, T61I and T170A relative to the JGI database protein sequence.

The results of the SDS-PAGE and zymogram analyses are shown in Fig. 6.5. Bands corresponding to the FopA and ApINV enzymes were visible between 100 and 130 kDa. The band was absent in the control strain supernatant. Similarly to the β-fructofuranosidases expressed in *K. pastoris*, the band sizes were larger than the expected 69 and 66.8 kDa for FopA and ApINV, respectively. De-*N*-glycosylation yielded bands of the expected sizes (data not shown). Both enzymes displayed hydrolytic activity on sucrose as evidenced by bands between 100 and 130 kDa visible on the zymogram gel.

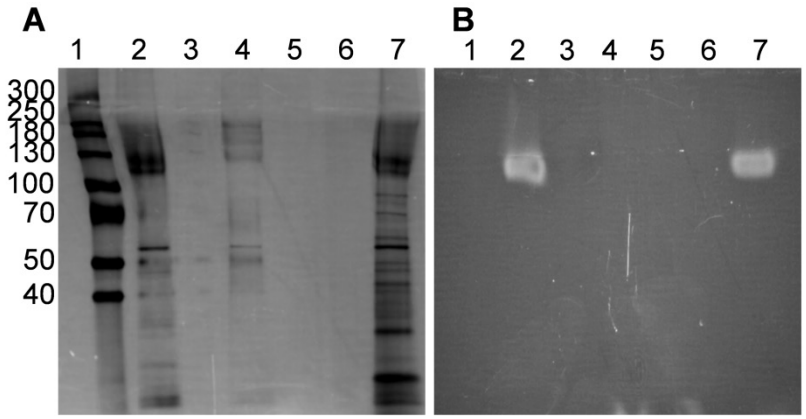

Fig. 6.5 SDS-PAGE and zymogram gels of ApINV and FopA. A Silver-stained 8% polyacrylamide gel of culture supernatants of *S. cerevisiae* expressing the two β-fructofuranosidases. Lane 1 contained the Spectra Multicolor High Range Protein Ladder (Thermo Scientific), Lane 2 shows FopA, Lane 4 contains the control strain transformed with the empty pJC1 vector and Lane 7 shows ApINV. B Zymogram gel of the two β-fructofuranosidases loaded in the same order as for SDS-PAGE

6.3.3 FOS synthesis by the β-fructofuranosidases

Following cultivation of the yeasts, supernatants were evaluated for global enzyme activity using a DNS assay. Supernatants were appropriately diluted to react 1 IU of enzyme with 0.58 M of buffered sucrose for 2 h. FOS were quantified using HPLC and the results are shown in Fig. 6.6.

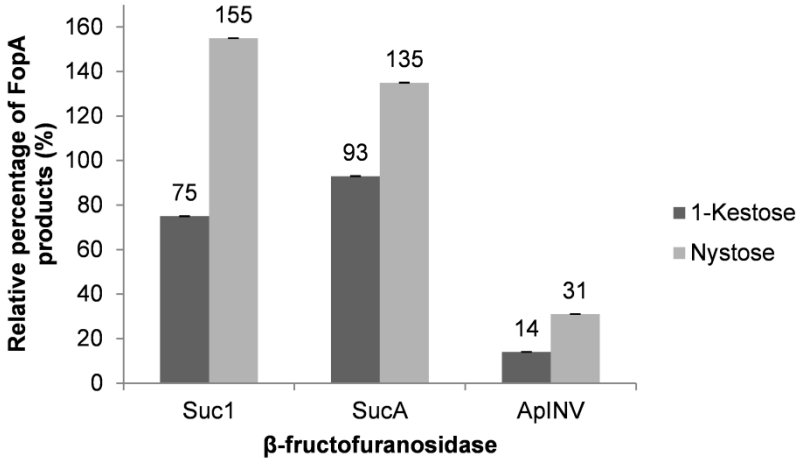

Fig. 6.6 Fructooligosaccharide production by *Aspergillus niger* Suc1 and SucA and *Aureobasidium pullulans* ApINV β-fructofuranosidases relative to *Aspergillus japonicus* FopA. One International unit of enzyme was reacted with 0.58 M buffered sucrose for 2 h. Fructooligosaccharides were quantified using high performance liquid chromatography. Error bars denote percentage standard error (n = 3)

FOS data clearly indicated that Suc1 and SucA possessed high-level FOS synthesis activities while ApINV displayed low-level FOS synthesis. ApINV only produced 14 and 31% of the 1-kestose and nystose produced by FopA, respectively. Although Suc1 produced 75% of the 1-

kestose produced by FopA while SucA produced 93%, Suc1 was deemed to have superior activity owing to the higher amounts of nystose produced. Time course studies of β-fructofuranosidases have shown that as the reaction progresses, levels of FOS with higher degrees of polymerisation increase up to a point after which they decline as FOS are degraded again (Nishizawa et al. 2001; our laboratory). In this study equal global units of enzyme were dosed and reactions were conducted for a fixed time, thus enzyme fructosyltransferase activity was the variable factor. Reactions were conducted at the optimal temperature of 55 °C reported for SucA (Zuccaro et al. 2008). In order to produce higher levels of nystose the enzyme must possess a higher transferase activity which is evidenced as increased substrate turnover. 1-Kestose serves as substrate for nystose production, the levels of which were correspondingly lower for Suc1. HPLC analysis of β-fructofuranosylnystose (GF_4) indicated that Suc1 produced 5.2 mM as opposed to 3.9 mM by SucA and 2.9 mM by FopA. All means were significantly different ($p<0.05$). On the other hand glucose levels, which represent global enzyme activity, did not differ significantly (data not shown).

6.3.4 Investigation of the active pocket geometries of enzymes displaying low-level and high-level FOS synthesis

The three dimensional crystal structure models for *S. cerevisiae* invertase (Suc2) (Sainz-Polo et al. 2013), *A. japonicus* fructosyltransferase complexed with nystose (Chuankhayan et al. 2010) and the ApINV homology model were superimposed to compare the catalytic pockets of enzymes exhibiting low-level and high-level FOS synthesis. The evaluation of superimpositions across 238 fully populated columns in the final alignment gave an overall root-mean-square deviation (RMSD) of 1.273 Å. Fig. 6.7 shows the superimposed models. The overall fit appeared good, despite the low degrees of sequence similarity and supports the view that protein fold is more highly conserved than sequence. It has been shown that the secreted form of Suc2 produces limited amounts of 6-kestose and no or undetectable amounts of nystose (Lafraya et al. 2011). The FOS data indicates similar low levels of 1-kestose and nystose for ApINV relative to FopA (Fig. 6.6). Inspection of Fig. 6.7 may provide an explanation for these observations. Nystose (GF_3) is shown with the terminal fructose moiety bound in the -1 subsite and the terminal glucose moiety in the +3 subsite. A clash is evident between loop residues of Suc2 and ApINV and the glucose moiety while this is not the case for *A. japonicus* fructosyltransferase. Thus a spatial incompatibility exists for the two enzymes displaying low-level FOS synthesis and FOS ligands with higher degrees of polymerisation. The obstructing loop includes 11 amino acid residues from N42 to L52 in 4EQV.

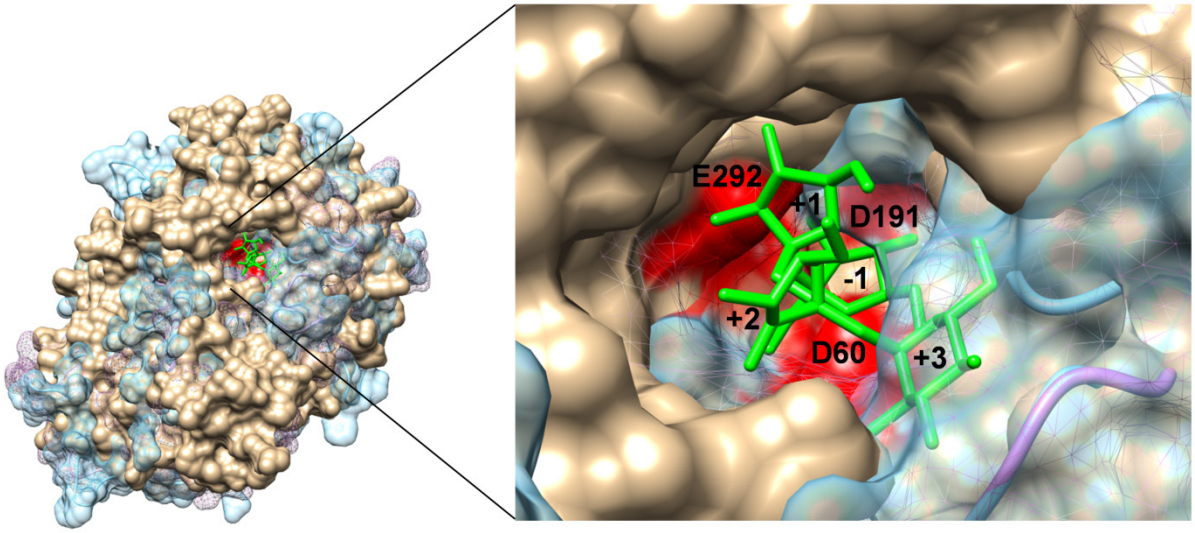

Fig. 6.7 Superimposed three-dimensional structures showing the solvent excluded molecular surfaces for *S. cerevisiae* invertase (plum mesh, 4EQV chain A), *A. japonicus* fructosyltransferase (tan) complexed with nystose (GF_3) shown in green (3LEM) and ApINV (blue, homology model). The catalytic triad residues are numbered according to 3LEM and are shown in red. The spatial clash between the glucose moiety in the +3 subsite and the loop residues of 4EQV and ApINV is highlighted in the zoomed-in catalytic pocket.

Investigation of the loop region in the MSA of experimentally characterised fungal GH32 enzymes displaying sucrolytic activity clearly indicated that the loop was longer in enzymes displaying low-level FOS synthesis activity. A gap was present in all of the sequences of enzymes displaying high-level FOS synthesis (Fig. 6.8).

```
                                                                  160                                 180
                                                                   |                                   |
Streptococcus mutans AAA88584.1      EADITTPKNT  IDEYGLTEQA  RKIATEAGIN  LSSLTQKQVE
Bacillus megaterium ADF38395.1       ----------  --SYGISHIT  RDNMVKIP--  -QQQNSDQFK
Aspergillus foetidus AJ000493.1      --------NG  TYHLFFQYNP  GGIEWGN-IS  WGHATSEDL-
Aspergillus oryzae BAE62112.1        ------SKRG  VYHLYYQYNP  TATVAGN-QH  WGHATSPDL-
Schwanniomyces occidentalis ADN34605 ------KTAK  LWHLYFQYNP  NATAWGQPLY  WGHATSNDL-
Saccharomyces cerevisiae NP_012104.1 (Suc2) ------EKDA  KWHLYFQYNP  NDTVWGTPLF  WGHATSDDL-
Aureobasidium pullulans (ApINV)      ------PTED  LYHAFYQWHP  EHINWGN-IS  WGHATSKDM-
Aspergillus niger ABB59679 (SucB)    ------PATG  VYHLSFQWNP  KGNDWGN-IS  WGHAVSNDL-
Xanthophyllomyces dendrorhous ACL79833 ------RADG  SIHAGYQSHP  KHIQWGN-IS  QGAAYSSDF-
Ceratocystis monoliformis AGV22100.1 ------TANK  KYHLGMQWTP  DGIAWNN-IT  WGAAVSDDL-
Aspergillus sydowii CAB89083.1 (Sft) ------PATG  LFHVGFLHNG  TGI-------  -SSVYTDDL-
Aspergillus japonicus AB046383.1 (FopA) ------PSTG  LFHVGFLHDG  DGI-------  -AGATTANL-
Scopulariopsis brevicaulis AAQ31829.1 (Seq1) ------PETG  IFHVGWLYNG  NGA-------  -SGATTEDL-
Aspergillus oryzae BAE63792.1 (F1)   ------PDTG  LFHVGFLYNG  SGI-------  -AGATTDDM-
Penicillium roquefortii AAQ31828.1 (Seq3) ------PKTG  LFHVGWLY--  SGI-------  -SGATTDDL-
Aspergillus niger S33920 (Suc1)      ------PSTG  LFHVGFLHDG  SGI-------  -SSATTDDL-
Aspergillus niger ABB59678.1 (SucA)  ------PSTG  LFHVGFLHDG  SGI-------  -SSATTDDL-
```

Fig. 6.8 A section of the multiple sequence alignment of experimentally characterised fungal GH32 enzymes displaying sucrolytic activity. Numbering is for the *S. mutans* sequence. GenBank protein accession numbers are given with trivial names indicated in brackets, where available. The loop highlighted by the block is longer in enzymes displaying low-level FOS synthesis than in those producing high levels of fructooligosaccharides and is proposed to cause spatial incompatibilities in the catalytic pocket that hinder the formation of fructooligosaccharides, especially those with higher degrees of polymerisation

Regarding the three dimensional structure in the regions of the conserved motifs A, D and E, motif A displays the most obvious disparities between to two types of β-fructofuranosidases. The N and G residues directly adjacent to the catalytic nucleophile in the WMN**D**PNG and GQIG**D**PC are the most different, with N possessing a bulky side chain while G has no side chain. The positioning of the A, D and E motifs in the superimposed structures was very similar (data not shown) and could be anticipated as the catalytic residues are highly conserved and would require precise spatial positioning in order to fulfil their functions. Additional H-bonds are formed between the substrate and the N residue while the space created by the G residue adjacent to the active site residue may reduce steric hindrance caused by the N side chain and promote transferase activity.

6.4 Discussion

An understanding of the structure-function relationship in enzymes is extremely valuable and enables strategic design to tailor catalysts for specific applications or to identify novel starting templates. Owing to the advances in DNA sequencing technologies and computational biology, large amounts of sequence data are available that can be analysed to identify regions in enzymes where variation has been tolerated by evolution. The assumption is made that regions displaying high levels of sequence diversity are responsible for functional differences between related enzymes. Furthermore, the larger the sequence dataset, the higher the confidence in the extrapolations drawn from characterised enzymes to unknown homologs. However, underrepresented groups are often lost in the generalisations. For example, in family GH32 the majority of sequences contain the WMN**D**PNG, W**E**C motifs, hence the signature HMM logos. The β-fructofuranosidases displaying high-level FOS synthesis and the GQIG**D**PC, F**E**T motifs are very much in the minority. Many of the experiments investigating the functionalities of GH32 enzymes to date were conducted prior to the availability of large homologous sequence datasets. Hence, some deductions from (semi-)rationally designed mutations exist which are difficult to apply to the extended sequence space. This study attempted to extrapolate experimental findings from a small number of studies (sequences) to a large sequence space, identify trends from the expanded view and again apply the findings in a focussed manner. Sequence motif identifiers were proposed for GH32 enzymes displaying high level FOS synthesis from the analysis of available fungal GH32 sequences in the context of extrapolations drawn from a handful of studies. Using the proposed WMN**D**PNG, W**E**C and the GQIG**D**PC, F**E**T motifs, the function of two novel sequences were correctly predicted and experimentally verified. The sequence motif for enzymes displaying high-level FOS synthesis was clearly different to the HMM for GH32 enzymes (Fig. 6.3) which may be attributed to the underrepresentation of these enzymes in the sequenced fungal genomes. Approximately 32 of

258 GH32 fungal sequences displayed the relevant motifs. This underrepresentation phenomenon may be linked to different fungal nutritional modes in relation to plants and an extended ancestral state reconstruction similar to the study by Parrent et al. (2009) may provide some insights into the role of high-level FOS synthesis in fungal nutritional strategies.

Interestingly, the data showed that the single exon Suc1 displayed enhanced FOS synthesis capabilities over its two exon counterpart, SucA (Fig. 6.6). The presence of the intron directly before the termination codon has been previously noted by Yanai et al. (2003) in the genes of the *P. roqueforti* and *S. brevicaulis*. The two exon versions of these enzymes were shown to have superior activity to enzymes comprised of only a single exon. Thus it was hypothesised that the second exon played a role in the determination of fructosyltransferase activity but the data obtained in this work refuted this. Furthermore, the data also showed that the region encoded by the second exon in the aforementioned enzymes was crucial to the functioning of FopA. The role of the second exon remained unclear but it may contribute to enzyme stability over extended incubation periods and in so doing, improve enzyme activity. Verification of this will require further investigation. Further bioinformatic analysis of the prevalence of the intron in the coding sequences of β-fructofuranosidases is also required. This investigation has, however clarified a protracted dialogue on the activity of the single- and two exon versions of the *A. niger* extracellular β-fructofuranosidase and is comparative activity to FopA.

The R**D**P sequence motif harbouring the transition state stabilising catalytic residue was similar for both low- and high-level FOS synthesising enzymes. Strikingly, the catalytic residue as well as its flanking residues were completely conserved across the GH32 family, which was not the case for the A and E motifs. This observation may imply that the RDP region is neither involved in determining transferase activity nor product specificity, which are highly variable attributes among GH32 β-fructofuranosidases. Mutagenesis studies on *S. cerevisiae* Suc2 improved synthesis of 6-kestose 10-fold by substituting residues in the WMN**D**PNG motif, namely W19Y and N21S (Lafraya et al. 2011). Synthesis of 1-kestose was unchanged which implies that the WMN**D**PNG motif is more likely involved with hydrolase/transferase activity rather than product specificity. In a comparison with plant enzymes, Chuankhayan et al. (2010) deduced that the G adjacent to the catalytic nucleophile in the equivalent GQIG**D**PC motif was responsible for product specificity. As plant and yeast/fungal GH32 enzymes form distinct phylogenetic clades (Alméciga-Díaz et al. 2011), it may be more applicable to extrapolate experimental findings to closer related enzymes. Together with the observation that characterised enzymes displaying high-level FOS synthesis produce inulin type (β(2→1)) FOS, the findings of Lafraya et al. (2011) were taken to support the hypothesis that the WMN**D**PNG/GQIG**D**PC motif is responsible for determining hydrolase/transferase activity.

A clear understanding on the function of the E motif was elusive. Different combinations of the A and E motifs were seen. For example, ApINV possessed the WMN**D**PNG motif in combination with a F**E**V motif although the consensus combination in the GH32 HMM is WMN**D**PNG combined with the W**E**C motif. Similarly, *A. niger* SucB possessed a WMN**D**PNG and W**E**V combination. Both enzymes combined hydrolase and transferase motifs. Again these observations supported the notion that the WMNDPNG motif is characteristic of enzymes displaying low-level FOS, however the F**E**T motif is not unique to the enzymes displaying high-level FOS synthesis. That the A motif is in all probability responsible for hydrolase/transferase activity and that the D motif was not involved in product specificity (6-kestose vs 1-kestose), a logical deduction was that the E motif may be connected to product specificity. Investigation of the proposed reaction mechanism of a fructosyltransferase (Chuankhayan et al. 2010) informed the view of the involvement of the E motif in product specificity, as the general acid/base catalytic residue (W**E**C or F**E**T) is directly involved in the bond formation between fructose moieties. The involvement of the catalytic nucleophile and surrounding residues in determining hydrolase/transferase activity left the general acid/base and surrounding residues as the obvious candidates. No clear correlation existed between the presence of the respective consensus E motifs and product specificity although in a few cases the correlation was made. Substitution of the P residue in the vicinity of the acid/base catalytic residue (Y**E**CP) for a V improved the specificity of 6-kestose production by the *S. occidentalis* β-fructofuranosidase which supports the hypothesis made in this study (Alvaro-Benito et al. 2010). Omissions on 6-kestose data in the literature may be one explanation for the lack of correlation as analytical standards are not easily obtained for 6-kestose. Furthermore, 6-kestose may not be the major reaction product and hence be below detection limits. There is evidence that GH32 enzymes produce mixtures of β(2→1) and β(2→6) linked FOS (Hirayama et al. 1989; Linde et al. 2009; Rehm et al. 1998; Sainz-Polo et al. 2013) and this may account for the variability observed in the E motif.

There are numerous reports detailing single amino acid substitutions that alter hydrolase/transferase activity, substrate and product specificity and may involve amino acids distant to the active site. Often these mutations are context dependent in that interaction with other amino acid residues in the protein are required to create conditions that are permissive for the resultant phenotype. Mutation of the equivalent amino acid in another enzyme may not yield similar results due to a lack of context. Identification of the responsible contexts is highly complex but statistical approaches are emerging to compute networks of evolutionarily conserved protein networks (Süel et al. 2003; Baussand & Carbone 2009; Dib et al. 2014). These networks are responsible for allosteric communication in proteins by physically linking

distant functional sites in the folded protein. Identification of these networks in GH32 enzymes and evaluation of reported substitutions in context of the networks will likely significantly improve the understanding of the structure-function relationships in GH32 enzymes.

In this study naturally occurring sequences were used to generate conserved sequence motifs that include active site residues, which enable the identification of enzymes with low- and high-level FOS synthesising abilities. They inherently possess the required three dimensional architecture and residue interaction networks that permit their functioning. Besides the rapid identification of novel enzymes for FOS synthesis, groups of enzymes identified by fitting logo profiles may serve as good starting points to investigate mutations responsible for particular aspects of enzyme functioning. As more sequences and targeted experimental results become available, the refinement of sequence logos for the identification of specific functions will improve the successes of *in-silico* bioprospecting.

6.5 ACKNOWLEDGEMENTS

Michael Bester is thanked for his inputs. Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). Financial support for this work was provided by the Technology Innovation Agency (TIA), project PB110/08. Kim Trollope was supported by grants from TIA and Stellenbosch University. She also holds a L'Oreal-UNESCO For Women in Science Regional Fellowship in Sub-Saharan Africa.

6.6 LITERATURE CITED

Alberto F, Bignon C, Sulzenbacher G, Henrissat B, Czjzek M (2004) The three-dimensional structure of invertase (beta-fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. J Biol Chem 279:18903–18910

Alméciga-Díaz CJ, Gutierrez AM, Bahamon I, Rodríguez A, Rodríguez M a, Sánchez OF (2011) Computational analysis of the fructosyltransferase enzymes in plants, fungi and bacteria. Gene 484:26–34

Alvaro-Benito M, de Abreu M, Portillo F, Sanz-Aparicio J, Fernández-Lobato M (2010) New insights into the fructosyltransferase activity of *Schwanniomyces occidentalis* ß-fructofuranosidase, emerging from nonconventional codon usage and directed mutation. Appl Environ Microbiol 76:7491–7499

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinforma 22 :195–201

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (2001) Current Protocols in Molecular Biology

Baussand J, Carbone A (2009) A combinatorial approach to detect coevolved amino acid networks in protein families of variable divergence. PLoS Comput Biol 5:e1000488

Boddy LM, Bergès T, Barreau C, Vainstein MH, Dobson MJ, Ballance DJ, Peberdy JF (1993) Purification and characterisation of an *Aspergillus niger* invertase and its DNA sequence. Curr Genet 24:60–66

Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast 14:115–132

Chaira N, Smaali I, Besbes S, Mrabet A, Lachiheb B, Ferchichi A (2011) Production of fructose rich syrups using invertase from date palm fruits. J Food Biochem 35:1576–1582

Cheng K-C, Demirci A, Catchmark JM, Puri VM (2011) Effects of initial ammonium ion concentration on pullulan production by *Aureobasidium pullulans* and its modeling. J Food Eng 103:115–122

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

Crous JM, Pretorius IS, van Zyl WH (1995) Cloning and expression of an *Aspergillus kawachii* endo-1,4-beta-xylanase gene in *Saccharomyces cerevisiae*. Curr Genet 28:467–473

Dib L, Silvestro D, Salamin N (2014) Evolutionary footprint of coevolving positions in genes. Bioinformatics 30:1241–1249

Dixon M, Webb EC (1958) Enzymes. Longmans, Green & Co. and Academic Press, London New York

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763

Edelman J (1956) The formation of oligosaccharides by enzymic transglycosylation. In: Nord FF (ed) Adv. Enzymol., Vol XVII. Interscience Publishers, Inc., New York, p 189

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113

EMBL (2013) Family: *Glyco_hydro_32N* (PF00251). http://pfam.xfam.org/family/PF00251#tabview=tab4. Accessed June 2014

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014) Pfam: the protein families database. Nucleic Acids Res 42:D222–230

Gadegaard G, Didion T, Folling M, Storgaard M, Andersen CH, Nielsen KK (2008) Improved fructan accumulation in perennial ryegrass transformed with the onion fructosyltransferase genes *1-SST* and *6G-FFT*. J Plant Physiol 165:1214–1225

Gallagher SR, Sasse J (2012) Staining Proteins in Gels. Curr. Protoc. Essent. Lab. Tech. John Wiley & Sons, Inc., pp 6:7.4:7.4.1–7.4.14

Goosen C, Yuan X-L, van Munster JM, Ram AFJ, van der Maarel MJEC, Dijkhuizen L (2007) Molecular and biochemical characterization of a novel intracellular invertase from *Aspergillus niger* with transfructosylating activity. Eukaryot Cell 6:674–681

Gunasekaran P, Karunakaran T, Cami B, Mukundan AG, Preziosi L, Baratti J (1990) Cloning and sequencing of the *sacA* gene: characterization of a sucrase from *Zymomonas mobilis*. J Bacteriol 172:6727–6735

Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280 (Pt 2):309–316

Henrissat B, Bairoch A (1996) Updating the sequence-based classification of glycosyl hydrolases. Biochem J 316 (Pt 2):695–696

Heyer AG, Wendenburg R (2001) Gene cloning and functional characterization by heterologous expression of the fructosyltransferase of *Aspergillus sydowi* IAM 2544. Appl Env Microbiol 67:363–370

Hidaka H, Hirayama M, Sumi N (1988) A fructooligosaccharide-producing enzyme from *Aspergillus niger* ATCC 20611. Agric Biol Chem 52:1181–1187

Hill J, Donald K, Griffiths D (1991) DMSO-enhanced whole cell yeast transformation. Nucleic Acids Res 19:5791

Hirayama M, Sumi N, Hidaka H (1989) Purification and properties of a Fructoooligosaccharide-producing β-fructofuranosidase from *Aspergillus niger* ATCC 20611. Agric Biol Chem 53:667–673

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duquenne L, Finn RD, Fraser M, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, McMenamin C, Mi H, Mutowo-Muellenet P, Mulder N, Natale D, Orengo C, Pesseat S, Punta M, Quinn AF, Rivoire C, Sangrador-Vegas A, Selengut JD, Sigrist CJA, Scheremetjew M, Tate J, Thimmajanarthanan M, Thomas PD, Wu CH, Yeats C, Yong S-Y (2012) InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res 40:D306–312

Jukes TH, Cantor CR (1969) Mammalian Protein Metabolism. In: Munro HN (ed)Academic Press, New York, pp 21–32

Koshland DE, Stein SS (1954) Correlation of bond breaking with enzyme specificity; cleavage point of invertase. J Biol Chem 208:139–148

Krogh A, Brown M, Mian IS, Sjölander K, Haussler D (1994) Hidden Markov models in computational biology. Applications to protein modeling. J Mol Biol 235:1501–1531

Kurakake M, Ogawa K, Sugie M, Takemura A, Sugiura K, Komaki T (2008) Two types of β-fructofuranosidases from *Aspergillus oryzae* KB. J Agric Food Chem 56:591–596

Lafraya Á, Sanz-Aparicio J, Polaina J, Marín-Navarro J (2011) Fructo-oligosaccharide synthesis by mutant versions of *Saccharomyces cerevisiae* invertase. Appl Environ Microbiol 77:6148–6157

Lammens W, Le Roy K, Schroeven L, Van Laere A, Rabijns A, Van den Ende W (2009) Structural insights into glycoside hydrolase family 32 and 68 enzymes: functional implications. J Exp Bot 60:727–740

Lin-Cereghino J, Wong W, Giang W (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. Biotechniques 38:44–48

Linde D, Macias I, Fernández-Arrojo L, Plou FJ, Jiménez A, Fernández-Lobato M (2009) Molecular and biochemical characterization of a beta-fructofuranosidase from *Xanthophyllomyces dendrorhous*. Appl Environ Microbiol 75:1065–1073

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–495

Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, Kusumoto K-I, Arima T, Akita O, Kashiwagi Y, Abe K, Gomi K, Horiuchi H, Kitamoto K, Kobayashi T, Takeuchi M, Denning DW, Galagan JE, Nierman WC, Yu J, Archer DB, Bennett JW, Bhatnagar D, Cleveland TE, Fedorova ND, Gotoh O, Horikawa H, Hosoyama A, Ichinomiya M, Igarashi R, Iwashita K, Juvvadi PR, Kato M, Kato Y, Kin T, Kokubun A, Maeda H, Maeyama N, Maruyama J, Nagasaki H, Nakajima T, Oda K, Okada K, Paulsen I, Sakamoto K, Sawano T, Takahashi M, Takase K, Terabayashi Y, Wortman JR, Yamada O, Yamagata Y, Anazawa H, Hata Y, Koide Y, Komori T, Koyama Y, Minetoki T, Suharnan S, Tanaka A, Isono K, Kuhara S, Ogasawara N, Kikuchi H (2005) Genome sequencing and analysis of *Aspergillus oryzae*. Nature 438:1157–1161

Maiorano AE, Piccoli RM, da Silva ES, de Andrade Rodrigues MF (2008) Microbial production of fructosyltransferases for synthesis of pre-biotics. Biotechnol Lett 30:1867–1877

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko M V, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:D225–229

Marthey S, Aguileta G, Rodolphe F, Gendrault A, Giraud T, Fournier E, Lopez-Villavicencio M, Gautier A, Lebrun M-H, Chiapello H (2008) FUNYBASE: a FUNgal phYlogenomic dataBASE. BMC Bioinformatics 9:456

Martin I, Débarbouillé M, Ferrari E, Klier A, Rapoport G (1987) Characterization of the levanase gene of *Bacillus subtilis* which shows homology to yeast invertase. Mol Gen Genet MGG 208:177–184

McDonald AG, Tipton KF (2014) Fifty-five years of enzyme classification: advances and difficulties. FEBS J 281:583–592

Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31:426–428

Myrbäck K (1960) Invertases. In: Boyer PD, Lardy H, Myrbäck K (eds) Enzym., 2nd ed. Academic Press, New York, pp 379–396

Nagao C, Nagano N, Mizuguchi K (2014) Prediction of detailed enzyme functions and identification of specificity determining residues by random forests. PLoS One 9:e84623

Naumoff DG (2001) Beta-fructosidase superfamily: homology with some alpha-L-arabinases and beta-D-xylosidases. Proteins 42:66–76

Neumann NP, Lampen JO (1967) Purification and properties of yeast invertase. Biochemistry 6:468–475

Nishizawa K, Nakajima M, Nabetani H (2001) Kinetic study on transfructosylation by β-fructofuranosidase from *Aspergillus niger* ATCC 20611 and availability of a membrane reactor for fructooligosaccharide production. Food Sci Technol Res 7:39–44

Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, Smirnova T, Grigoriev I V, Dubchak I (2014) The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res 42:D26–31

Ohta K, Akimoto H, Matsuda S, Toshimitsu D, Nakamura T (1998) Molecular cloning and sequence analysis of two endoinulinase genes from *Aspergillus niger*. Biosci Biotechnol Biochem 62:1731–1738

Parrent JL, James TY, Vasaitis R, Taylor AF (2009) Friend or foe? Evolutionary history of glycoside hydrolase family 32 genes encoding for sucrolytic activity in fungi and its implications for plant-fungal symbioses. BMC Evol Biol 9:148–164

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612

Pons T, Naumoff DG, Martínez-Fleites C, Hernández L (2004) Three acidic residues are at the active site of a beta-propeller architecture in glycoside hydrolase families 32, 43, 62, and 68. Proteins 54:424–432

Pons T, Olmea O, Chinea G, Beldarraín a, Márquez G, Acosta N, Rodríguez L, Valencia A (1998) Structural model for family 32 of glycosyl-hydrolase enzymes. Proteins 33:383–395

Reddy A, Maley F (1996) Studies on identifying the catalytic role of Glu-204 in the active site of yeast invertase. J Biol Chem 271:13953–13957

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

Roberfroid MB (2007) Inulin-type fructans: functional food ingredients. J Nutr 137:2493S–2502S

Saarelainen R, Paloheimo M, Fagerström R, Suominen P, Nevalainen K (1993) Cloning, sequencing and enhanced expression of the *Trichoderma reesei* endoxylanase II (pI 9) gene *xln2*. Mol Gen Genet 241:497–503

Sainz-Polo MA, Ramírez-Escudero M, Lafraya A, González B, Marín-Navarro J, Polaina J, Sanz-Aparicio J (2013) Three-dimensional structure of *Saccharomyces invertase*: role of a non-catalytic domain in oligomerization and substrate specificity. J Biol Chem 288:9755–9766

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sambrook J, Fritsch E, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sander C, Schneider R (1994) The HSSP database of protein structure-sequence alignments. Nucleic Acids Res 22:3597–3599

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. Biopolymers 38:305–320

Sreekrishna K, Tschopp JF, Fuke M (1987) Invertase gene (*SUC2*) of *Saccharomyces cerevisiae* as a dominant marker for transformation of *Pichia pastoris*. Gene 59:115–125

Süel GM, Lockless SW, Wall M a, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol 10:59–69

Tamura K, Kawakami A, Sanada Y, Tase K, Komatsu T, Yoshida M (2009) Cloning and functional analysis of a fructosyltransferase cDNA for synthesis of highly polymerized levans in timothy (*Phleum pratense* L.). J Exp Bot 60:893–905

Taussig R, Carlson M (1983) Nucleotide sequence of the yeast *SUC2* gene for invertase. Nucleic Acids Res 11:1943–1954

Van Wyk N, Trollope KM, Steenkamp ET, Wingfield BD, Volschenk H (2013) Identification of the gene for β-fructofuranosidase from *Ceratocystis moniliformis* CMW 10134 and characterization of the enzyme expressed in *Saccharomyces cerevisiae*. BMC Biotechnol 13:100–111

Waterhouse AL, Chatterton NJ (1993) Glossary of fructan terms. In: Suzuki M, Chatterton N (eds) Sci. Technol. Fruct. CRC Press, Boca Raton, FL, pp 2–7

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191

Wheeler TJ, Clements J, Finn RD (2014) Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinformatics 15:7–16

Yamada Y, Matsuda M, Maeda K, Mikata K (1995) The phylogenetic relationships of methanol-assimilating yeasts based on the partial sequences of 18S and 26S ribosomal RNAs: the proposal of *Komagataella* gen. nov. (Saccharomycetaceae). Biosci Biotechnol Biochem 59:439–44

Yanai K, Nakane A, Kawate A, Hirayama M (2001) Molecular cloning and characterization of the fructooligosaccharide-producing β-fructofuranosidase gene from *Aspergillus niger* ATCC 20611. Biosci Biotechnol Biochem 65:766–773

Yanai K, Nakane A, Toshiaki K (2003) β-fructofuranosidase and gene thereof. US patent 6566111

Yuan X-L, Goosen C, Kools H, van der Maarel MJEC, van den Hondel C a MJJ, Dijkhuizen L, Ram AFJ (2006) Database mining and transcriptional analysis of genes encoding inulin-modifying enzymes of *Aspergillus niger*. Microbiology 152:3061–3073

Yun JW (1996) Fructooligosaccharides—Occurrence, preparation, and application. Enzyme Microb Technol 19:107–117

Zalar P, Gostincar C, de Hoog GS, Ursic V, Sudhadham M, Gunde-Cimerman N (2008) Redefinition of *Aureobasidium pullulans* and its varieties. Stud Mycol 61:21–38

Zuccaro A, Götze S, Kneip S, Dersch P, Seibel J (2008) Tailor-made fructooligosaccharides by a combination of substrate and genetic engineering. Chembiochem 9:143–149

6.7 APPENDIX: SUPPLEMENTARY INFORMATION

Table S8 List of protein sequences harbouring the conserved sequence motifs, GQIGDP and FET, which impart putative high-level FOS synthesis activity. Entries marked with an asterisk were used for the generation of sequence logos for conserved motifs. Unmarked entries were retrieved from the respective databases based on BLAST homology to characterised enzymes with high-level FOS synthesis abilities or *fructofuranosidase* keyword search. Only BLAST hits displaying the GQIGDP motif were included

Accession Number	Organism	Source database
*gi|630184394|ref|XP_007846416.1|	*Moniliophthora roreri* MCA 2997	NCBI
*gi|238592184|ref|XP_002392832.1|	*Moniliophthora perniciosa* FA553	NCBI
*gi|114190041|gb|EAU31741.1|	*Aspergillus terreus* NIH2624	NCBI
*gi|114192365|gb|EAU34065.1|	*Aspergillus terreus* NIH2624	NCBI
*gi|211586019|emb|CAP93760.1|	*Penicillium chrysogenum* Wisconsin 54-1255	NCBI
*gi|7649391|emb|CAB89083.1|	*Aspergillus sydowii*	NCBI
*gi|398404245|ref|XP_003853589.1|	*Zymoseptoria tritici* IPO323	NCBI
*gi|300953014|gb|ADK46938.1|	*Aspergillus japonicus*	NCBI
*gi|295982394|pdb|3LF7|	*Aspergillus japonicus*	NCBI
*gi|90660167|gb|ABD97344.1|	*Aspergillus japonicus*	NCBI
*gi|15620806|dbj|BAB67771.1|	*Aspergillus niger*	NCBI
*gi|187981491|gb|EDU48117.1|	*Pyrenophora tritici-repentis* Pt-1C-BFP	NCBI
*gi|317155113|ref|XP_003190558.1|	*Aspergillus oryzae* RIB40	NCBI
*gi|635513567|gb|KDE85371.1|	*Aspergillus oryzae* 100-8	NCBI
*gi|159023686|gb|ABW87267.1|	*Aspergillus oryzae*	NCBI
*gi|169780922|ref|XP_001824925.1|	*Aspergillus oryzae* RIB40	NCBI
*gi|425771688|gb|EKV10125.1|	*Penicillium digitatum* PHI26	NCBI
*gi|211584647|emb|CAP74172.1|	*P. chrysogenum* Wisconsin 54-1255	NCBI
*gi|313767042|gb|ADR80690.1|	*Aspergillus niger*	NCBI
*gi|2599506|gb|AAC08047.1|	*Aspergillus niger* IBT10sb	NCBI
*gi|566080619|gb|AHC54391.1|	*Aspergillus niger*	NCBI
*gi|81177521|gb|ABB59678.1|	*Aspergillus niger*	NCBI
*gi|134077738|emb|CAK45778.1|	*Aspergillus niger* CBS 513.88	NCBI
*gi|477508549|gb|ENH61842.1|	*Fusarium oxysporum f. sp. cubense* race 1	NCBI
*gi|46128075|ref|XP_388591.1|	*Fusarium graminearum* PH-1	NCBI
*gi|485547|pir|S33920|	*Aspergillus niger*	NCBI
*gi|629643712|ref|XP_007789440.1|	*Eutypa lata* UCR-EL1	NCBI
*gi|631245667|ref|XP_007917755.1|	*Togninia minima* UCRPA7	NCBI
gi|358371493|dbj|GAA88101.1	*Aspergillus kawachii* IFO 4380	NCBI
*tr|C7YM80|	*Nectria haematococca* ATCC MYA-4622	HSSP
*tr|E3RQG1|	*Pyrenophora teres f. teres*	HSSP
*tr|B8NTE9|	*Aspergillus flavus* ATCC 200026	HSSP
*Aurpu2p4_002860	*Aureobasidium pullulans* ATCC62921	Genozymes
FVET_02085	*Fusarium verticilloides* strain 7600	FunyBASE
AB05676.1	*Alternaria brassicicola* strain	FunyBASE

Accession Number	Organism	Source database
	ATCC 96836	
ref|XP_007835784.1|	*Pestalotiopsis fici* W106-1	NCBI
ref|XP_008026158.1|	*Setosphaeria turcica* Et28A	NCBI
gb|EMD86040.1|	*Bipolaris maydis* C5	NCBI
ref|XP_007691109.1|	*Bipolaris oryzae* ATCC 44560	NCBI
gb|EYE96443.1|	*Aspergillus ruber* CBS 135680	NCBI
gb|EUC31783.1|	*Bipolaris zeicola* 26-R-13	NCBI
ref|XP_007701777.1|	*Bipolaris sorokiniana* ND90Pr	NCBI
gb|EUN28642.1|	*Bipolaris victoriae* FI3	NCBI
gb|KEQ96897.1|	*Aureobasidium subglaciale* EXF-2481	NCBI
gb|EST08223.1|	*Pseudozyma brasiliensis* GHG001	NCBI
emb|CCF49382.1|	*Ustilago hordei*	NCBI
jgi|Aspac1|55546|fgenesh1_pm.23_#_70	*Aspergillus aculeatus* ATCC 16872	JGI 29/07/2014
jgi|Aspca3|202165|fgenesh_isotigs_kg.1_#_55_#_isotig06770	*Aspergillus carbonarius* ITEM 5010	JGI 29/07/2014

NCBI National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/).
HSSP Homology Derived Secondary Structure Of Proteins (http://mrs.cmbi.ru.nl/m6/)
Genozymes (http://www.fungalgenomics.ca/wiki/Fungal_Genomes)
FunyBASE (http://genome.jouy.inra.fr/funybase/FUNYCG/WEB/CGI-BIN/funycg_infos.cgi?idproject=70)
JGI Joint Genome Institute, United States Department of Energy (http://jgi.doe.gov/

CHAPTER 7

GENERAL DISCUSSION AND CONCLUSIONS

7 GENERAL DISCUSSION AND CONCLUSIONS:

7.1 GENERAL DISCUSSION AND CONCLUSIONS

Industrial production of fructooligosaccharides (FOS) relies on two different enzymatic processes depending on the available substrate and technologies. Inulin extracted from chicory can be hydrolysed by microbial endoinulinases to yield FOS. Alternatively, the biotransformation of sucrose to FOS is carried out by β-fructofuranosidases, particularly those possessing high fructosyltransferase activities. Reaction conditions for FOS synthesis from sucrose typically involve high sucrose concentrations of up to 700 g/l and fairly high reaction temperatures of 50-60°C (Nishizawa et al. 2001; Yun 1996). Glucose, a by-product of fructosyltransferase activity, is responsible for enzyme inhibition and limits the sucrose to FOS conversion yields to ca. 60% FOS (Jung et al. 1989; Nishizawa et al. 2001). The stability of the enzyme is also significantly affected by reaction temperatures (Yun and Song 1999). To address the limitations of the current enzymatic FOS synthesis bioprocess the modification of enzyme traits to improve its robustness is required.

The aim of this study was to engineer the *A. japonicus* β-fructofuranosidase (FopA) for improved synthesis of inulin-type FOS from sucrose. Superior activity included any property, such as increased thermostability, relief from product inhibition and/or higher specific activity. FopA was selected as it displays high fructosyltransferase activity and is well studied in terms of sequence, structure and function (Chuankhayan et al. 2010; Hidaka et al. 1988; Hirayama et al. 1989; Yanai et al. 2001). This prior knowledge afforded an opportunity to investigate the specific determinants of high fructosyltransferase activity, which to date remain unclear. The fact that FopA is used for the commercial production of inulin-type FOS for which the prebiotic effects have been verified by experimental data and human clinical trials, and hence can substantiate claims that FOS consumption results in improved human and animal health, was a further motivating factor for the selection of FopA. An improved enzyme variant would be of considerable industrial interest to reduce the cost of FOS manufacturing.

A dual approach was employed to engineer FopA as to increase the probability of obtaining an improved enzyme variant(s). A random mutagenesis and a semi-rational approach for the selection of mutations were tested. The random mutagenesis approach was selected to harness the potential of the randomness of mutations as precise structural knowledge of the enzyme regions involved in the phenotypic presentation of product inhibition, specific activity and thermal stability was unavailable. The semi-rational approach afforded the additional opportunity to reduce the number of variants to be screened yet theoretically increased the functional content of the library. By selecting evolutionary tolerated mutations and avoiding

regions displaying defined secondary structures, which are sensitive to changes that alter precise geometric conformations, the result was indeed a small, functionally enriched library.

A high throughput screen (HTS) is a prerequisite for successful application of a random mutagenesis approach to improve enzyme traits (Kazlauskas and Bornscheuer 2009; Romero and Arnold 2009). The selection and screen should reflect the desired result as this experimental stage is both time-consuming and expensive. Very few reports detail the directed evolution of β-fructofuranosidases, especially using a random mutagenesis approach, as a HTS for FOS quantification, to the best of my knowledge, is unavailable. The studies using random mutagenesis to engineer the *Schwanniomyces occidentalis* β-fructofuranosidase make use of an indirect colorimetric-based HTS that enzymatically quantifies glucose and fructose separately. A high glucose to fructose ratio was used to identify variants that theoretically produced higher levels of 6-kestose (de Abreu et al. 2013; de Abreu et al. 2011). The screen was followed by high performance liquid chromatography to confirm that the selected variants produce the desired product, 6-kestose. The first law of directed evolution states that 'you get what you screen for' (Schmidt-Dannert and Arnold 1999). The enzymatic screen of de Abreu et al. can potentially misidentify hits, as enzymes with very low activity, for example, may also give high glucose to fructose ratios, leading to the identification of false positives. A screen that directly quantifies FOS would be advantageous to increase confidence in initial screening data, which are critical to a directed evolution experiment.

The work done in this study on the simultaneous quantification of glucose, fructose, sucrose, 1-kestose and nystose using Fourier transform mid infrared attenuated total reflectance (FT-MIR ATR) spectroscopy combined with multivariate data analysis provides, for the first time, a rapid direct FOS product quantification method that is amenable to high throughput library screening. In addition, it is also a cost effective, environmentally friendly method that does not require large amounts of costly consumables nor does it generate toxic waste products during sample analysis. In these two aspects the method developed in this study is superior to the currently available methods for screening β-fructofuranosidase libraries and FOS quantification. The method is however, sensitive to sample matrix effects and requires fairly sophisticated data analysis techniques. For this reason, ongoing partial least squares calibration model development was proposed to address the matrix effects, which in practise only requires tweaking of the model by including additional unknown samples from new assay batches for which reference values have been determined. For both the enzymatic and spectroscopic methods to be applied in a high-throughput format, access to expensive, sophisticated liquid handling robotic equipment and well-plate adaptors is necessary. As all compounds have unique infrared spectra the assay method developed in this study can be readily adapted for the quantification of numerous other enzyme products, provided they are present at detectable

concentrations. In essence the method has intersected two independent disciplines – infrared spectroscopy combined with multivariate data analysis which specialises in discriminating between different samples using multiple variables; and protein engineering which relies on using a single variable, i.e. enzyme activity to identify unique samples (enzyme variants). By applying partial least squares regression for calibration, multiple, easily measured variables (wavenumbers) are transformed into a single variable (sugar concentration) which may be used to rapidly discriminate between multiple enzyme variants. Combining the two techniques provides a novel and powerful way to resolve the constraint on engineering a β-fructofuranosidase by random mutagenesis, or other methods that generate large libraries, due to the absence of a HTS for FOS products.

In the enzyme engineering field, there has been a move away from screening increasingly larger variant libraries to ensure sufficient sampling of mutational diversity to the construction of smaller, functionally enriched libraries using prior knowledge of enzyme structure and function. Screening limitations often necessitate the construction of semi-rational libraries and at the outset of this study, this scenario was presented. The availability of the crystal structure of the *A. japonicus* fructosyltransferase (Chuankhayan et al. 2010) facilitated the design of a small variant library targeting substitutions to solvent exposed loop regions. Multiple sequence alignments also provided guidance for the selection of the identity of the mutations and their position, targeting regions with high sequence entropy. The combination of four amino acid substitutions, identified after screening an initial library of single amino acid substitution variants, yielded a much improved variant relative to the wild type parent, FopA. The variant displayed an ca. doubled specific activity and a 26% reduction in reaction time to obtain a FOS composition (1-kestose:nystose:β-fructofuranosylnystose) approximating that of a commercial FOS preparation. As this is the first report on the directed evolution of an efficient β-fructofuranosidase displaying high level FOS synthesis, comparison of the results from this work to other studies is problematic. Nevertheless, results on the directed evolution of the *S. occidentalis* β-fructofuranosidase (originally a low level FOS producer) indicate that the transglycosylation activity could be increased ~5.5-fold at 16 h reaction time and 2.2-fold at 136 h, where FOS synthesis peaked (de Abreu et al. 2013). The specific activity determination in this work was performed using different reaction parameters, specifically with only 1 h of reaction time, to allow for comparison of the data generated in this study with published data on FopA. Theoretically, one would expect it more likely to obtain major improvements in a certain trait in an enzyme that does not display a high propensity for the specified trait compared to an enzyme that does. That is, one would expect to be able to greatly improve the FOS synthesis capacity of an enzyme displaying low level FOS synthesis abilities rather than in one already displaying high level FOS synthesis, as such an enzyme has naturally evolved to perform the specified function. As this work aimed to produce FOS of a specified composition, the improved

variant data is not on an equal footing to the results obtained in other studies that report the highest total FOS levels, as such data do not consider FOS composition. When higher degree of polymerisation (DP) FOS species are present in the mixture the total FOS level is lower, and therefore does not allow for direct comparison of enzyme performance. Two molecules of 1-kestose are converted to one molecule of nystose, with the release of a glucose molecule. The total mass included in the FOS concentration value is thus lower, as glucose is excluded.

Practically, the only valid comparison in this section of the work is that of the best variant to the parent. Only four of 11 amino acids that were shown to improve single amino acid substitution variant activities were combined in the most improved variant, with their effects being additive. This generates scope to combine the remaining seven substitutions and investigate to what extent the additional substitutions may improve variant activity. The main limitation of this approach is the cost of variant gene synthesis. It is however, the most feasible way to generate the library that has all combinations of the 11 substitutions.

As to which approach is superior, in the case of FopA which has significant amounts of associated prior knowledge, a semi-rational approach is favoured for directed evolution because it proved to be the approach that required the least amount of effort to obtain a variant that displayed improved synthesis of inulin-type FOS. It required only 54 variants in total to be screened. In further agreement with Kazlauskas and Bornscheuer (2009) the semi-rational approach was superior as it provided new knowledge that engineering loops is an effective way to increase the elusive specific activity of a β-fructofuranosidase and less surprisingly, its thermostability.

Next generation sequencing technologies have caused an explosion in the amount of available genomic sequence data and has created the need to characterise new sequences using computational methods, rather than traditional experimental characterisation. Traditional methods are simply no longer practically feasible as there is too much new sequence data. At the same time, the large number of available homologous sequences has created a unique opportunity to study how nature has evolved particular enzyme (as well as other protein) traits. By studying the sequences and structures of characterised enzymes in the context of their homologs it is possible to identify specific sub-domain regions that have been conserved in enzymes displaying a specific property. The hypothesis that certain critical amino acid residues in the context of a specific architecture are responsible for certain enzyme traits motivated the search, in this study, for conserved sequence motifs that would reflect this. The design of the semi-rational variant library and the evaluation of data collected from the directed evolution work precipitated such thinking, strengthening the proposal made in the project aim statement that the engineering of FopA would create new avenues to investigate the determinants of

fructosyltransferase activity which are responsible for the high level FOS synthesis. The approach followed to develop a protein sequence-based strategy to identify novel GH32 enzymes displaying high-level FOS synthesis deviates from the conventional mutagenesis-based approaches to identify specific amino acids responsible for a specified trait. It is a more inclusive way to study the structure-function relationship in proteins and complements conventional methodology. The tool developed in this section of work may be applied in the development of new protein engineering strategies. It facilitates the rapid identification of novel sequences of enzymes displaying high fructosyltransferase activity that can serve as starting material for directed evolution studies. It may also inform the selection of homologous sequences to be used in the design of mutations in a semi-rational engineering approach. On a more fundamental level the sequence-based functional classification of GH32 enzymes may benefit studies investigating the importance of FOS in fungal physiology. It may help to give clues as to why some fungi possess multiple GH32 enzymes if the enzymes can be functionally annotated more accurately. The method successfully developed in this study to functionally sub-classify β-fructofuranosidases may be applied to other enzyme families, if supporting experimental data is available.

The data presented in the body of the thesis have been adequate to meet the aim and specific objectives set out at the onset of this study. The best variant, engineered using a semi-rational rather than a random approach, showed improvements to its thermal stability, relief from glucose inhibition and specific activity without any obvious compromise to overall enzyme performance. Such enhancements will in all likelihood improve the bioprocess economics of FOS production by either reducing the required enzyme dosage or reaction times to obtain FOS of a specified composition. Identifying this variant signified meeting the first objective and overall aim of the project. The development of the FT-MIR ATR spectrometry and chemometrics-based sugar quantification method met the second objective of the study targeted at developing a rapid variant library screening method, the advantages of which are self-evident in a protein engineering context. The final objective was met following the successful development of a protein sequence-based strategy to functionally annotate FOS synthetic capacity of GH32 β-fructofuranosidases.

The finding that by engineering loop regions it is possible to alter enzyme activity supports other data that implicate residue network interactions within a protein to be crucial to enzyme function (Baussand and Carbone 2009; Süel et al. 2003). Evaluation of such networks within FopA in light of mutagenesis data obtained in this study may improve understanding of the protein functioning. The tools developed in this study will aid in the development of focused protein engineering strategies for β-fructofuranosidases which may increase the knowledge extracted from such endeavours and further contribute to the understanding of the structure-function of

β-fructofuranosidases. Improved understanding of function will facilitate reliable tailoring of specific enzyme traits and may ultimately find utility in a range of applications, such as endeavours to glycosylate alternate substrates such as proteins, lipids or small organic molecules.

7.2 LITERATURE CITED

Baussand J, Carbone A (2009) A combinatorial approach to detect coevolved amino acid networks in protein families of variable divergence. PLoS Comput Biol 5:e1000488

De Abreu M, Alvaro-Benito M, Sanz-Aparicio J, Plou FJ, Fernandez-Lobato M, Alcalde M (2013) Synthesis of 6-Kestose using an Efficient β-Fructofuranosidase Engineered by Directed Evolution. Adv Synth Catal 355:1698–1702

De Abreu MA, Alvaro-Benito M, Plou FJ, Fernández-Lobato M, Alcalde M (2011) Screening β-fructofuranosidases mutant libraries to enhance the transglycosylation rates of β-(2→6) fructooligosaccharides. Comb Chem High Throughput Screen 14:730–738

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

Hidaka H, Hirayama M, Sumi N (1988) A fructooligosaccharide-producing enzyme from *Aspergillus niger* ATCC 20611. Agric Biol Chem 52:1181–1187

Hirayama M, Sumi N, Hidaka H (1989) Purification and properties of a Fructoooligosaccharide-producing β-fructofuranosidase from *Aspergillus niger* ATCC 20611. Agric Biol Chem 53:667–673

Jung KH, Yun JW, Kang KR, Lim JY, Lee JH (1989) Mathematical model for enzymatic production of fructo-oligosaccharides from sucrose. Enzyme Microb Technol 11:491–494

Kazlauskas RJ, Bornscheuer UT (2009) Finding better protein engineering strategies. Nat Chem Biol 5:526–529

Nishizawa K, Nakajima M, Nabetani H (2001a) Kinetic study on transfructosylation by β-fructofuranosidase from *Aspergillus niger* ATCC 20611 and availability of a membrane reactor for fructooligosaccharide production. Food Sci Technol Res 7:39–44

Romero PA, Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol 10:866–876

Schmidt-Dannert C, Arnold FH (1999) Directed evolution of industrial enzymes. Trends Biotechnol 17:135–136

Süel GM, Lockless SW, Wall M a, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol 10:59–69

Yanai K, Nakane A, Kawate A, Hirayama M (2001) Molecular cloning and characterization of the fructooligosaccharide-producing β-fructofuranosidase gene from *Aspergillus niger* ATCC 20611. Biosci Biotechnol Biochem 65:766–773

Yun JW (1996) Fructooligosaccharides—Occurrence, preparation, and application. Enzyme Microb Technol 19:107–117

Yun JW, Song SK (1999) Enzymatic Production of Fructooligosaccharides from Sucrose. In: Bucke C (ed) Carbohydr. Biotechnol. Protoc. Humana Press, Totowa, N.J., pp 141–151