# Expression of novel amylases in *Saccharomyces cerevisiae* for the efficient conversion of raw starch to bioethanol

Rosemary Anne Cripwell

Dissertation presented for the degree of Doctor of Philosophy in the Faculty of Science at Stellenbosch University




Promoter: Prof. W.H. van Zyl

Co-Supervisor: Dr S.H. Rose

March 2017

# Declaration

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

This dissertation includes 3 unpublished publications. The development and writing of the papers were the principal responsibility of myself 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: March 2017

**Declaration by the candidate:**

With regard to Chapters 1 to 6, the nature and scope of my contribution were as follows:

| **Nature of contribution** | **Extent of contribution** |
|---|---|
| Design and planning of experimental work, data capturing and analysis and preparation of the manuscript draft | 90% |

The following co-authors have contributed to Chapters 2 to 5:

| **Name** | **e-mail address** | **Nature of contribution** | **Extent of contribution** |
|---|---|---|---|
| Dr S.H. Rose | shrose@sun.ac.za | Editing and intellectual input | 10% |
| Prof. W.H. van Zyl | whvz@sun.ac.za | Editing and intellectual input | 10% |

Date: 13 February 2017

**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 2 to 5,

2. no other authors contributed to Chapters 2 to 5, 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 2 to 5 of this dissertation.

| **Signature** | **Institutional affiliation** | **Date** |
|---|---|---|
| **S.H. Rose** | Stellenbosch University | 13 February 2017 |
| **W.H. van Zyl** | Stellenbosch University | 13 February 2017 |

# Summary

Starchy biomass is an ideal, abundant substrate for bioethanol production. The cost effective conversion of starch requires a fermenting yeast that is able to produce starch hydrolysing enzymes and ferment glucose to ethanol in one step called consolidated bioprocessing (CBP). Despite the advantages, CBP yeasts have not yet been employed for the industrial processing of raw starch during bioethanol production.

Molecular biology has enabled the optimised expression of synthetically produced genes in *Saccharomyces cerevisiae*. The *Aspergillus tubingensis* raw starch hydrolysing α-amylase (*amyA*) and glucoamylase (*glaA*) encoding genes were codon optimised using different strategies and expressed in *S. cerevisiae* Y294. However, compared to the native coding sequences for the *amyA* and *glaA* genes, adapted synonymous codon usage resulted in a decrease in extracellular enzyme activity of 72% (30 nkat.ml$^{-1}$) and 69% (4 nkat.ml$^{-1}$), respectively.

Additional fungal amylase encoding genes (native and codon optimised) were expressed in *S. cerevisiae* Y294 and then screened for starch hydrolysis. Subsequently, *S. cerevisiae* Y294 laboratory strains were constructed to co-express the best α-amylase and glucoamylase gene variants and evaluated for raw starch fermentation. During raw starch fermentations, the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain displayed the highest carbon conversion (based on the percentage starch converted on a mol carbon basis) of 85%, compared to 54% displayed by the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain. Therefore, the native α-amylase (*temA_Nat*) and codon optimised glucoamylase (*temG_Opt*) genes, both originating from *Talaromyces emersonii*, presented the best amylase combination and were selected for further evaluation.

Amylolytic *S. cerevisiae* Ethanol Red™ and M2n industrial strains were constructed using the *amdS* marker (encoding for acetamidase). Strains co-expressing the *temA_Nat* and *temG_Opt* genes were selected for growth on acetamide as the sole nitrogen source. Amylolytic *S. cerevisiae* strains (Ethanol Red T12 and M2n T1) were compared in a CBP process (20% raw corn starch) at 30°C and 37°C. The maximum ethanol concentration produced at 30°C by the *S. cerevisiae* Ethanol Red T12 and M2n T1 strains was 86.5 g.l$^{-1}$ and 99.4 g.l$^{-1}$, respectively.

Fermentations were supplemented with different dosages of STARGEN 002™, an exogenous GSHE (granular starch hydrolysing enzyme) cocktail, to compare the amylolytic yeast strains to an industrial simultaneous saccharification and fermentation (SSF) process. Fermentation results for the *S. cerevisiae* Ethanol Red T12 strain with 10% of the recommended STARGEN™ dosage compared well with the SSF using *S. cerevisiae* Ethanol Red™

containing the full recommended STARGEN™ dosage, both having carbon conversions of 50% after 48 hours and 93% after 192 hours. This study also highlights the application of novel industrial amylolytic yeasts in combination with STARGEN™ for decreased fermentations times.

At 37°C, the amylolytic *S. cerevisiae* Ethanol Red T12 strain performed better than the *S. cerevisiae* M2n T1 strain, demonstrating its potential as a drop-in CBP yeast for existing bioethanol plants that use cold hydrolysis processes. The study also provided a novel enzyme combination (TemA_Nat and TemG_Opt) that efficiently hydrolyses raw corn starch. Finally, new light was shed on the importance of synonymous codon usage and the expression of native genes versus their codon optimised variants.

# Opsomming

Styselagtige biomassa is 'n ideale, volop substraat vir bio-etanol produksie. Die koste-effektiewe omskakeling van stysel vereis 'n fermenterende gis wat styselafbrekende ensieme produseer en glukose na etanol in een stap omskakel, bekend as gekonsolideerde bioprosessering (GBP). Ten spyte van die voordele, word GBP-giste nog nie vir die industriële verwerking van rou stysel na bio-etanol gebruik nie.

Molekulêre biologie het die optimale uitdrukking van sinteties-vervaardigde gene in *Saccharomyces cerevisiae* moontlik gemaak. Die kodonvolgorde van *Aspergillus tubingensis* gene wat vir die rou stysel hidroliserende α-amilase (*amyA*) en glukoamilase (*glaA*) kodeer, is met verskillende strategieë geoptimiseer en in *S. cerevisiae* uitgedruk. In vergelyking met die inheemse volgorde van die *amyA* en *glaA* gene, het aangepaste sinonieme kodongebruik egter onderskeidelik tot 'n afname van 72% (30 $nkat.ml^{-1}$) en 69% (4 $nkat.ml^{-1}$) in ekstrasellulêre ensiemaktiwiteit gelei.

Addisionele fungi amilase-koderende gene (inheems en kodon-geoptimiseerd) is in *S. cerevisiae* Y294 uitgedruk en dan vir rou stysel hidrolise getoets. Die *S. cerevisiae* Y294 laboratoriumstamme wat die beste α-amilase en glukoamilase geenvariante gesamentlik uitdruk, is vervolgens geskep en vir rou stysel fermentasie geëvalueer. Tydens rou stysel fermentasies, het die *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] gistam die hoogste rou stysel omskakeling getoon met 'n koolstof omskakeling van 85%, in vergelyking met 54% deur die *S. cerevisiae* Y294[AmyA-GlaA] verwysingstam. Die inheemse α-amilase (*temA_Nat*) en kodon-geoptimiseerde glukoamilase (*temG_Opt*) gene, beide van *Talaromyces emersonii* afkomstig, het die beste amilase kombinasie gelewer en is derhalwe vir verdere evaluering gekies.

Amilolitiese *S. cerevisiae* Ethanol Red™ en M2n industriële stamme is ontwikkel met behulp van die *amds* merker (kodeer vir asetamidase). Stamme wat die *temA_Nat* en *temG_Opt* gene gesamentlik uitdruk, is op asetamied as enigste stikstofbron geselekeer. Amilolitiese *S. cerevisiae* stamme (Ethanol Red T12 en M2n T1) is in 'n GBP-proses (20% rou mieliestysel) by 30°C en 37°C vergelyk. Die maksimum etanolkonsentrasie deur die *S. cerevisiae* Ethanol Red T12 en M2n T1 stamme gelewer by 30°C, was onderskeidelik 86.5 $g.l^{-1}$ en 99.4 $g.l^{-1}$. Fermentasies is met verskillende ladings van STARGEN 002™, 'n eksogene styselkorrel hidrolitiese ensiem-mengsel, aangevul ten einde die amilolitiese gisrasse in ‘n industriële gelyktydige versuikering en fermentasie (GVF) proses te vergelyk. Fermentasie resultate vir die *S. cerevisiae* Ethanol Red T12 stam met 10% van die aanbevole STARGEN™-lading het goed vergelyk met die *S. cerevisiae* Ethanol Red™ GVF met die volle aanbevole STARGEN™-lading. All twee het koolstof omskakelings van 50% na 48 uur en 93% na

192 ure. Hierdie studie beklemtoon ook die toepassing van unieke industriële amilolitiese giste in kombinasie met STARGEN™ vir verbeterde versuikering en fermentasie van rou mieliestysel.

Die stysel-afbrekende *S. cerevisiae* Ethanol Red T12 gisras het by 37°C beter as die *S. cerevisiae* M2n T1 ras gedoen, wat sy potensiaal uitlig as 'n GBP-gis vir toevoeging tot bestaande bio-etanol fabrieke wat koue hidrolise-prosesse gebruik. Die studie het ook 'n unieke ensiemkombinasie (TemA_Nat en TemG_Opt) gelewer wat rou mieliestysel doeltreffend hidroliseer. Laastens is nuwe lig gewerp op die belang van sinonieme kodongebruik en die uitdrukking van inheemse gene teenoor kodon-geoptimiseerde variante.

# List of abbreviations

| | |
|---|---|
| **ANOVA** | Analysis of variance |
| **BSA** | Bovine serum albumin |
| **CBP** | Consolidated bioprocessing |
| **CBM** | Carbohydrate binding domain |
| **CAI** | Codon adaption index |
| **CBI** | Codon bias index |
| **CUB** | Codon usage bias |
| **DDGS** | Distiller's dried grains with solubles |
| **DNS** | Dinitrosalicylic acid |
| **DCW** | Dry cell weight |
| **DSF** | Differential scanning fluorimetry |
| **EPA** | Environment protection authority |
| **EC** | Enzyme commission |
| **ER** | Endoplasmic reticulum |
| **FDA** | Food and drug administration |
| **GH** | Glycoside hydrolase |
| **GRAS** | Generally regarded as safe |
| **GSDE** | Granular starch degrading enzymes |
| **HPLC** | High performance liquid chromatography |
| **LAB** | Lactic acid bacteria |
| **MGT** | Mascoma grain technology |
| **mRNA** | Messenger RNA |
| **NADH** | Nicotinamide adenine dinucleotide |
| **ORF** | Open reading frame |
| **PDI** | Protein disulphide isomerase |
| **PCR** | Polymerase chain reaction |
| **SDS-PAGE** | Sodium dodecyl sulphate -polyacrylamide gel electrophoresis |
| **RNA** | Ribonucleic acid |
| **RSDE** | Raw starch degrading enzymes |
| **SEM** | Scanning electron microscope |
| **SBD** | Starch binding domain |
| **SHF** | Separate hydrolysis and fermentation |
| **SS** | Secretion signal |
| **SSF** | Simultaneous saccharification and fermentation |
| **SC** | Synthetic complete |
| **tRNA** | Transfer RNA |
| **YML** | Yeast mediated ligation |

# Acknowledgements

I thank God for his blessings, protection and the wonderful opportunity to conduct my studies in the Stellenbosch winelands.

My sincere thanks and appreciation is extended to the following people and institutions:

**Prof. W.H. van Zyl**, Department of Microbiology, Stellenbosch University, my supervisor for this study, for the opportunity to peruse my PhD studies, his enthusiasm and thought-provoking discussions during his study;

**Dr S.H. Rose**, Department of Microbiology, Stellenbosch University, my co-supervisor, for her wise words, support and proof reading of my chapters;

**Prof. M. Viljoen-Bloom**, for her helpful discussions and assistance with translating my summary section;

**My co-workers in the Bloom lab**, for their support, scientific input and tea-time discussions, especially **Lalie Kossatz** (my lunch buddy) for her help and friendship;

**The Staff of the Department of Microbiology**, for all their assistance and administrative help;

**Mrs Lisa Warburg**, for her assistance with HPLC analysis;

**Central Analytical Facilities (CAF)**, Stellenbosch University, especially Dr Stanton Hector for his assistance with real-time PCR;

**The National Research Foundation (NRF)** for financial support;

**My parents**, Sue and Roy Dobson, for their support (emotional and financial) and always believing in me;

**My brothers**, Richard, Timothy and Arthur, for their love, support and jokes about synthetic genes;

**My husband,** Jamie Cripwell, (to whom this thesis is dedicated) for his unending love, continued support and encouragement during our student years.

**To Jamie**

# Preface

This thesis is presented as a compilation of six chapters. Chapter 3 is written in the style of a journal to which the manuscript was submitted for review. Chapters 1, 2, 4 and 5 have been prepared in the same style. Extracts of this dissertation have been used for a provisional patent application (**Addendum A**).

**Chapter 1**: General introduction and project aims

**Chapter 2**: Literature review

**Chapter 3**: Expression and comparison of codon optimised *Aspergillus tubingensis* amylase variants in *Saccharomyces cerevisiae*

*Under review for publication*

**Chapter 4**: Novel raw starch amylase combinations for the construction of a CBP yeast

*In preparation for publication*

**Chapter 5**: Construction of an amylolytic CBP Ethanol Red™ strain

*In preparation for publication*

**Chapter 6**: General discussion and conclusions

# Table of contents

# List of figures

**Fig. 3.1** Schematic representation of the vector constructs used in this study. (**a**) The amy*A_Opt* (and derivatives), *amyA_CBI*, *glaA_Opt* (and *glaA_Opt→252*) and *glaA_CBI* genes were expressed from the yBBH1 multicopy episomal vector, whereas (**b**) the *PDI1* from *S. cerevisiae* was constitutively expressed from a single chromosomal integrative copy using the enolase 1 (*ENO1*) promoter and terminator sequences for transcriptional control.

**Fig. 3.2** Illustration of the relative frequency of codon usage for the (**a**) *amyA* and (**b**) *glaA* gene variants indicating regions I, II and III that were replaced to construct *amyA_Opt(→300)*, *amyA_Opt(→426)*, *amyA_Opt(→300 + 426)* and *glaA_Opt(→252)*, respectively. Schematic representation (**c** and **d**) of *amyA* and *glaA* gene constructs.

**Fig. 3.3** The effect of cultivation temperature on recombinant *S. cerevisiae* strains. Blue lines represent strains cultivated at 20°C and red lines strains cultivated at 30°C. The recombinant (**a**) α-amylase and (**b**) glucoamylase activities, as well as DCW (**c** and **d**) of the recombinant *S. cerevisiae* Y294 strains. Error bars represent the standard deviation of the three independent biological replicates. Supernatant from (**e**) (lanes 2,5) *S. cerevisiae* Y294[AmyA], (lane 3,6) Y294[AmyA_Opt] and (lane 4,7) Y294[AmyA_CBI], as well as (**f**) (lanes 2,5) Y294[GlaA], (lanes 3,6) Y294[GlaA_Opt] and (lane 4,7) Y294[GlaA_CBI] strains were subjected to SDS-PAGE with the protein size markers in lane M and *S. cerevisiae* Y294[BBH1] reference strain in lane 1. The degylcosylated (+) and (-) untreated (**g**) α-amylases and (**h**) glucoamylases were visualised by SDS-PAGE. The protein sizes are depicted on the left hand side.

**Fig. 3.4** DSF analysis of purified proteins. The melting curves of (**a**) AmyA, Amy_Opt and AmyA_CBI and (**b**) GlaA, GlaA_Opt and GlaA_CBI proteins. The relative fluorescence curves (**c** and **d**) for α-amylases and glucoamylases. The standard deviations for each sample set were calculated for the 5 repeats. SDS-PAGE was used to visualise (**e** and **f**) the purified α-amylases and glucoamylases. The protein size marker is in lane M and the sizes depicted on the left hand side.

**Fig. 3.5** Effect of *PDI1* co-expression and *S. cerevisiae* NI-C-D4 as the host strain on (**a** and **b**) extracellular α-amylase and glucoamylase activity and DCW (**c** and **d**) for α-amylase and glucoamylase *S. cerevisiae* strains. Error bars represent the standard deviation of three independent biological repeats. The supernatant of *S. cerevisiae* (**e**) α-amylase and (**f**) glucoamylase strains was subjected to SDS-PAGE. The protein size marker is in lane M and the sizes depicted on the left hand side.

**Fig. 3.6** Effect of synonymous codon substitutions on (**a**) extracellular α-amylase activity, (**b**) extracellular glucoamylase activity, (**c** and **d**) supernatant from *S. cerevisiae* Y294 strains was subjected to SDS-PAGE followed by silver staining. The protein size marker is in lane M and the sizes depicted on the left hand side.

**Fig. 4.1** Schematic representation of the final vector constructs used in this study. Amylase encoding genes were amplified using PCR and respectively cloned onto the yBBH1 and yBBH4 vectors (**a**, **b** and **c**). The $ENO1_P$-α-amylases-$ENO1_T$ cassettes were cloned onto the yBBH1-glucoamylase plasmids (**d**), to enable co-expression of the genes. *Bam*HI and *Bgl*II restriction enzyme sites were used for YML.

**Fig. 4.2** Extracellular α-amylase activity displayed by the *S. cerevisiae* Y294 strains expressing the (**a**) *ateA*, *amyA*, (**b**) *apuA* and (**c**) *temA* gene derivatives, respectively. The *S. cerevisiae* Y294[AmyA] strain was used for benchmark α-amylase production. Values represent the mean of three repeats and error bars represent the standard deviation. Supernatant from the *S. cerevisiae* Y294 strains (after 72 hours) was subjected to SDS-PAGE followed by silver staining. The arrows indicate the presence of the recombinant (**d**) AmyA, AteA, (**e**) ApuA and (**f**) TemA protein species, respectively. The *S. cerevisiae* Y294[BBH1] strain was used as the reference strain and the protein size marker (M) is depicted on the left hand side.

**Fig. 4.3** Extracellular glucoamylase activity displayed by the *S. cerevisiae* Y294 strains expressing the (**a**) *ateG, glaA* and (**b**) *temG* gene derivatives, respectively. The *S. cerevisiae* Y294[GlaA] strain was used for benchmark glucoamylase production. Values represent the mean of three repeats and error bars represent the standard deviation. Supernatant from the *S. cerevisiae* Y294 strains (after 72 hours) was subjected to SDS-PAGE followed by silver staining. The arrows indicate the presence of the recombinant **(c**) AteG and (**d**) TemG protein species, respectively. The *S. cerevisiae* Y294[BBH1] strain was used as the reference strain and the protein size marker (M) is depicted on the left hand side.

**Fig. 4.4** The amylolytic *S. cerevisiae* Y294 strains were evaluated on 200 $g.l^{-1}$ raw corn starch and 5 $g.l^{-1}$ glucose as sole carbohydrate source. The (**a** and **b**) ethanol and (**c** and **d**) glucose production was monitored overtime. Results from the best performing strains (left panel) and suboptimal strains (right panel) came from the same fermentation. Values represent the mean of three repeats and error bars represent the standard deviation.

**Fig. 4.5** The performance of *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] in a 2 litre bioreactor. (**a**) Ethanol concentrations at 26°C (-●-) and 30°C (-■-) and residual glucose concentrations at 26°C (-○-) and at 30°C (-□-) and (**b**) carbon conversion (based on the percentage starch converted on a mol carbon basis) at 26°C (-●-) and 30°C (-■-), respectively, with 2×$SC^{-URA}$ broth supplemented with 5 $g.l^{-1}$ glucose and 200 $g.l^{-1}$ raw corn starch. Values represent the mean of triplicate repeats and error bars represent the standard deviation.

**Fig. 5.1** Schematic representation of the final vector and gene cassettes used in this study. The $TEF_P$-*amdS*-$TEF_T$ cassette was cloned onto yBBH1 (**a**) to generate the yBBH1-amdSYM expression vector. The *ENO1 temA_Nat* and *temG_Opt* gene cassettes (**b**) were amplified using PCR and contained flanking regions homologous to the δ-integration sites.

**Fig. 5.2** Comparison of industrial transformants after integration of *temA* and *temG* gene cassettes. Ethanol produced (**a**) and carbon conversion (based on the percentage starch converted on a mol carbon basis) (**b**) displayed by *S. cerevisiae* Ethanol Red (-□-) and M2n (-○-) parental strains and *S. cerevisiae* Ethanol Red T1 (-▲-), T12 (-■-), M2n T1 (-◆-) and Mn2 T2 (-●-) amylolytic transformants at a fermentation temperature of 30°C on 200 g.l$^{-1}$ raw corn starch. SC-Ac (**c**) and SC-Acr (**d**) plate assays confirmed the ability of recombinant *S. cerevisiae* Ethanol Red T12 and M2n T1 strains to utilise acetamide and acrylamide, respectively, whereas the parental *S. cerevisiae* Ethanol Red and M2n strains indicated no growth.

**Fig. 5.3** Comparison between the laboratory *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain (-●-) and the industrial amylolytic *S. cerevisiae* Ethanol Red T12 strain at 30°C (-■-) and 37°C (-▲-). The production of ethanol (**a**), glucose (**b**), maltose (**c**) and glycerol (**d**) were compared using 2×SC$^{-URA}$ fermentation media that contained 5 g.l$^{-1}$ glucose and 200 g.l$^{-1}$ raw corn starch. Data are the mean of 3 repeats showing standard deviation.

**Fig. 5.4** Different fermentation broth conditions during fermentation at 37°C on 200 g.l$^{-1}$ raw corn starch. *S. cerevisiae* Ethanol Red T12 in YP (-◆-), YP citrate-acid buffer pH 5 (-■-), SC citrate-acid buffer pH 5 (-●-) and SC citrate-acid buffer pH 5 with 10 g.l$^{-1}$ extra $(NH_4)_2SO_4$ (-▲-). Ethanol (**a**), glucose (**b**), glycerol concentrations (**c**) and carbon conversion (based on the percentage starch converted on a mol carbon basis) (**d**) were compared. Data are the mean of 3 repeats showing standard deviation.

**Fig. 5.5** Ethanol concentrations produced by *S. cerevisiae* Ethanol Red strains during fermentation with 200 g.l$^{-1}$ corn starch at 30°C (**a**) and at 37°C (**b**), carbon conversion (based on the percentage starch converted on a mol carbon basis) at 30°C (**c**) and at 37°C (**d**). Untransformed *S. cerevisiae* Ethanol Red + 28 µl STARGEN (-▲-), Ethanol Red T12 (-■-), Ethanol Red T12 + 2.8 µl STARGEN (-◆-), Ethanol Red T12 + 4.6 µl STARGEN (-●-) and Ethanol Red T12 + 14 µl STARGEN (-▬-). Data are the mean of 3 repeats showing standard deviation.

**Fig. 5.6** Ethanol concentrations produced by *S. cerevisiae* M2n strains during fermentation with 200 g.l$^{-1}$ at 30°C (**a**) and 37°C (**b**), carbon conversion (based on the percentage starch converted on a mol carbon basis) at 30°C (**c**) and at 37°C (**d**). The untransformed *S. cerevisiae* M2n strain + 28 µl STARGEN (-▲-), M2n T1 (-■-), M2n T1 + 2.8 µl STARGEN (-◆-) and M2n T1 + 4.6 µl STARGEN (-●-). Data are the mean of 3 repeats showing standard deviation.

# List of tables

# Chapter 1

General introduction and project aims

## 1.1 Introduction

Energy availability plays a central role in the socio-economic development of countries. Combined with drastic changes in the crude oil prices over the last decade, this energy availability has had a significant effect on the global economy. Furthermore, there is an increased demand for transportation fuels throughout the world (Hahn-Hägerdal et al. 2006). Subsequently, the depletion of fossil fuel reserves has resulted in researchers focusing their efforts on the production of renewable and alternative energy resources (Connor and Atsumi 2010).

In order to decrease production costs and meet the mandate for renewable fuel blending, the International Energy Agency (IEA 2010) has promoted the use of cheap substrates for fuel production (energy-crops, food processing residues, as well as agricultural and forestry waste). Currently, the global bioethanol production is mainly produced from sugar and starchy feedstocks (Bai et al. 2008). Starch-based feedstocks, including grains (corn or wheat) and tubers (potatoes and cassava), are ideal substrates for biofuel production because they are renewable and available in large quantities. Since the majority of bioethanol produced is currently derived from starch, there is also an incentive to develop more cost-effective conversion processes for ethanol production from raw starch.

The conventional starch to ethanol process is a mature technology that requires a number of operational steps, as well as a high-energy input to cook the starch substrate (Goyal et al. 2005). However, an alternative process that eliminates the high cooking temperatures (i.e. cold hydrolysis) would be more energy-efficient for the conversion of raw starch to ethanol and reduce the ethanol production costs (Xiao et al. 2014). Several strategies can be followed to reduce production costs: (i) use a less expensive feedstock or agriculture waste product (Dellomonaco et al. 2010), (ii) use amylolytic strains to reduce the exogenous enzyme requirements and (iii) follow the concept of consolidated bioprocessing (CBP).

CBP is a single-step process whereby microorganisms are able to hydrolyse the biomass to monomeric sugars, while simultaneously converting the fermentable sugars to ethanol (Favaro et al. 2015). The integration of several fields of study is required, including synthetic biology, yeast engineering and fermentation technology, to develop an efficient amylolytic CBP yeast. With the assistance of molecular biology, metabolic and enzyme engineering strategies are being used to achieve renewable energy goals.

Starch based industries rely heavily on amylases, with microorganisms being the primary sources of these enzymes. The favourable properties that enzymes have as biocatalysts make them desirable for many different industrial uses; they have a high degree of specificity for their substrates and accelerate the rate of chemical reactions. Although amylases are

extensively used for industrial applications (Pandey et al. 2000), only a small percentage of bacterial and fungal strains meet the criteria for commercial production (e.g. strains of *Bacillus* sp., *Aspergillus* sp. and *Rhizopus* sp.) (Cereia et al. 2006; Zeng et al. 2011). Furthermore, a considerable amount of these amylases is required to convert raw starch to ethanol. Therefore, there is a need for novel recombinant (amylolytic) microorganisms that can effectively produce starch hydrolysing enzymes that can be used for amylase production in a continuous process.

Recombinant cell factories is a well used strategy for producing large quantities of pure enzymes. *Saccharomyces cerevisiae* is a frequently used host, since it has GRAS (generally regarded as safe) status and a well developed gene expression system. It is also the favoured industrial ethanol producer (Lynd et al. 2002). Although numerous studies have engineered *S. cerevisiae* to hydrolyse raw starch, ethanol production has not yet reached levels required for industrial application (Görgens et al. 2015) and major efforts are being made to increase the enzymatic saccharification of raw starch. The search for improved amylases, as well as the optimisation of known raw starch hydrolysing enzymes, is thus important for the development of amylolytic CBP yeast.

A large number of raw starch degrading amylases remain that have not been investigated for expression in *S. cerevisiae*. Sun et al. (2010) reviewed microorganisms that produce raw starch hydrolysing enzymes and concluded that low enzyme yield is the main limitation to their industrial application. Although many amylases have been considered for expression in alternative hosts, as listed by Robertson et al. (2006) and Reddy et al. (2009), experimental data to support their use for raw starch CBP is lacking.

Advances in gene expression tools used by molecular biologists have facilitated the engineering and codon optimisation of DNA molecules for improved expression in foreign hosts (Lux et al. 2012), with specific interest in genes encoding for industrially relevant enzymes. The design and *de novo* synthesis of genes and pathways is an emerging field of synthetic biology that has promoted the expression of foreign genes in heterologous hosts, such as *S. cerevisiae*, for improved protein yields. The re-designing of native genes at the nucleotide level will assist in understanding the relationship between gene sequence and gene expression (Wu et al. 2007). The decreased cost of synthetically produced DNA has enabled researchers to rapidly obtain foreign genes that can be codon optimised for expression in a particular host. This approach formed the basis of the study presented here, whereby novel fungal amylases were expressed in *S. cerevisiae* in order to construct improved industrial amylolytic yeast strains for the one-step conversion of raw corn starch to ethanol.

## 1.2 Aims and objectives of the study

The **first aim** of this study was to screen for and optimise novel amylase encoding genes for the hydrolysis of raw corn starch.

**Objectives**

i. Literature search and NCBI (National Center for Biotechnology Information) BLAST analysis (https://blast.ncbi.nlm.nih.gov) to identify novel amylases containing a starch binding domain (SBD).
ii. Clone and express amylase encoding genes in *S. cerevisiae* Y294 to identify the best α-amylase and glucoamylase enzymes in terms of extracellular amylase activity on soluble starch (using liquid assays).
iii. Compare the expression levels of codon optimised amylases versus their native counterparts.
iv. Investigate the effect of different secretion signals to ensure that the best conditions for heterologous protein secretion are selected.

The **second aim** was to develop amylolytic CBP yeast by co-expressing novel α-amylase and glucoamylase encoding gene combinations.

**Objectives**

i. Engineer *S. cerevisiae* Y294 to simultaneously express α-amylase and glucoamylase gene combinations.
ii. Perform fermentations with the different amylolytic *S. cerevisiae* Y294 strains, in order to select for the amylase enzyme combination that best converted raw starch (based on the percentage starch converted on a mol carbon basis).

The **third aim** was to produce ethanol through CBP using industrial amylolytic *S. cerevisiae* yeast strains.

**Objectives**

i. Engineer the industrial *S. cerevisiae* Ethanol Red™ and M2n strains to co-express the best α-amylase and glucoamylase combination for the one-step conversion of raw corn starch to ethanol.
ii. Compare ethanol concentrations produced by the amylolytic industrial strains to a simultaneous saccharification and fermentation (SSF) control, using untransformed *S. cerevisiae* Ethanol Red™ and M2n strains with STARGEN 002™ addition.

## 1.3 References

Bai FW, Anderson WA, Moo-Young M (2008) Ethanol fermentation technologies from sugar and starch feedstocks. Biotechnol Adv 26:89-105. doi: 10.1016/j.biotechadv.2007.09.002

Cereia M, Guimarães LHS, Peixoto-Nogueira SC, Jorge JA, Terenzi HF, Greene LJ, Polizeli MDLTM (2006) Glucoamylase isoform (GAII) purified from a thermophilic fungus *Scytalidium thermophilum* 15.8 with biotechnological potential. Afr J Biotechnol 5:1239-1245. ISSN 1684–5315

Connor MR, Atsumi, S (2010) Synthetic biology guides biofuel production. J Biomed Biotechnol 2010:541698. doi: 10.1155/2010/541698

Dellomonaco C, Fava F, Gonzalez R (2010) The path to next generation biofuels: successes and challenges in the era of synthetic biology. Microb Cell Fact 9:3. doi: 10.1186/1475-2859-9-3

Favaro L, Viktor M, Rose S, Viljoen-Bloom M, van Zyl W, Basaglia M, Cagnin L, Casella S (2015) Consolidated bioprocessing of starchy substrates into ethanol by industrial *Saccharomyces cerevisiae* strains secreting fungal amylases. Biotechnol Bioeng 112:1751-1760. doi: 10.1002/bit.25591

Görgens JF, Bressler DC, van Rensburg E (2015) Engineering *Saccharomyces cerevisiae* for direct conversion of raw, uncooked or granular starch to ethanol. Crit Rev Biotechnol 35:396-391. doi: 10.3109/07388551.2014.888048

Goyal N, Gupta JK, Soni SK (2005) A novel raw starch digesting thermostable α-amylase from Bacillus sp. I-3 and its use in the direct hydrolysis of raw potato starch. Enzyme Microb Technol 37:723–734. doi: 10.1016/j.enzmictec.2005.04.017

Hahn-Hägerdal B, Galbe M, Gorwa-Grauslund MF, Lidén G, Zacchi G (2006) Bio-ethanol - the fuel of tomorrow from the residues of today. Trends Biotechnol 24:549–56. doi: 10.1016/j.tibtech.2006.10.004

IEA (International Energy Agency) (2012) Sustainable production of second - generation biofuels potential and perspectives in major economies and developing countries. OECD/IEA, Paris. Available at: https:// www.iea.org/publications/freepublications/ publication/second_generation_biofuels.pdf. [Accessed 7 October 2016]

Lux MW, Bramlett BW, Ball DA, Peccoud J (2012) Genetic design automation: engineering fantasy or scientific renewal? Trends Biotechnol 30:120–126. doi: 10.1016/j.tibtech.2011.09.001

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506-577. doi: 10.1128/MMBR.66.3.506

Pandey A, Nigam P, Soccol CR, Soccol VT, Singh D, Mohan R (2000) Advances in microbial amylases. Biotechnol Appl Biochem 31:135–52. PMID: 10744959

Reddy LVA, Reddy OVS, Basappa SC (2009) Potentiality of yeasts in the direct conversion of starchy materials to ethanol and its relevance in the new millennium. In Satyanarayana T, Kunze G (Eds.). Yeast Biotechnology: Diversity and Applications. Netherlands: Springer, pp. 515–549

Robertson GH, Wong DW, Lee CC, Wagschal K, Smith MR, Orts WJ (2006) Native or raw starch digestion: a key step in energy efficient biorefining of grain. J Agric Food Chem 54:353-365. doi: 10.1021/jf051883m

Sørensen A, Lübeck M, Lübeck P, Ahring B (2013) Fungal beta-glucosidases: a bottleneck in industrial use of lignocellulosic materials. Biomolecules 3:612–631. doi: 10.3390/biom3030612

Sun H, Zhao P, Ge X, Xia Y, Hao Z, Liu J, Peng M (2010) Recent advances in microbial raw starch degrading enzymes. Appl Biochem Biotechnol 160:988–1003. doi: 10.1007/s12010-009-8579-y

Wu G, Zheng Y, Qureshi I, Zin HT, Beck T, Bulka B, Freeland SJ (2007) SGDB: a database of synthetic genes re-designed for optimizing protein over-expression. Nucleic Acids Res 35:D76-D79. doi: 10.1093/nar/gkl648

Xiao Z, Wu M, Grosse S, Beauchemin M, Lévesque M, Lau PCK (2014) Genome mining for new α-amylase and glucoamylase encoding sequences and high level expression of a glucoamylase from *Talaromyces stipitatus* for potential raw starch hydrolysis. Appl Biochem Biotechnol 172:73–86. doi: 10.1007/s12010-013-0460-3

Zeng Q, Wei C, Jin J, Wu C, Huang B (2011) Cloning of the gene encoding acid-stable alpha-amylase from *Aspergillus niger* and its expression in *Pichia pastoris*. Afr J Food Sci 5:668-675. ISSN: 1996-0794

# Chapter 2

Literature review

## 2.1 Starch

Starch is an abundant storage polysaccharide found in the leaves, flowers, seeds, stems and roots of plants. It is produced by green plants from glucose during photosynthesis in the chloroplasts and amyloplasts of leaves (Smith 2001; Tester et al. 2004). The annual starch production is estimated at 717 million metric tons, with United States, China and Brazil producing approximately 79% of this starch (Ranum et al. 2014). The largest percentage of starch is derived from corn (maize) (Singh et al. 2010), with tapioca, potato and wheat being the other primary industrial sources (de Souza and de Oliveira e Magalhães 2010). In 2012, the sale of starches (and derivatives) was estimated at $51.2 billion and is expected to reach $77.4 billion by 2018 (Santana and Meireles 2014).

Starch has numerous end use applications and this has led to a high global consumption rate, which is projected to reach 133.5 million tons by 2018 (Global Industry Analysts 2012). Starch does not require intense purification making it an economically attractive natural polymer for application in the food and beverage industry, with the corn starch market being divided into native starch, modified starch and sweeteners. Besides its use as a food source, starch has several non-food applications in the pharmaceutical, textile and biofuel industries (Santana and Meireles 2014). Furthermore, starch crops and residual starchy biomass are attractive feedstocks for bioethanol production (Streb and Zeeman 2012) because they represent a renewable and sustainable resource.

### 2.1.1 Molecular structure of starch

Starch typically consists of a mixture of amylose and amylopectin molecules (Fig. 2.1) and the relative amounts of these polymers determines the diverse chemical properties of starch granules (van der Maarel et al. 2002). Amylose (accounts for 10 - 30% of the granule mass) is a linear polymer consisting of up to 6000 glucose units linked by α-1,4 glycosydic bonds (de Souza and de Oliveira e Magalhães 2010; Olsen 2008). Amylose adopts a single helical structure (Fig. 2.2) that is randomly positioned within the amorphous lamella and is associated with the functional properties of starch (Streb and Zeeman 2012). The glycosidic bonds are stable at high pH, but degrades at low pH (van der Maarel et al. 2002). Amylopectin is a branched polymer (Fig. 2.2) that consists of short linear chains of 10 - 60 glucose units joined by α-1,4 bonds and α-1,6 linked side chains of 15 - 45 glucose units.

Starch is biosynthesised as granules containing growth rings that correspond to concentric semi-crystalline 120 - 400 nm thick shells, separated by amorphous regions (Fig. 2.3) (Tawil et al. 2011). It has a simple chemical structure and the architecture of native starch granules is influenced by the botanical source (origin, species and strain) from which the starch is obtained. Genes encoding for starch biosynthetic enzymes, as well as environmental

factors, influence the granule size distribution, crystallinity, organisation of the molecules within the granule and the chemical nature of the starch polymers (Fredriksson et al. 1998; Goldstein et al. 2016).

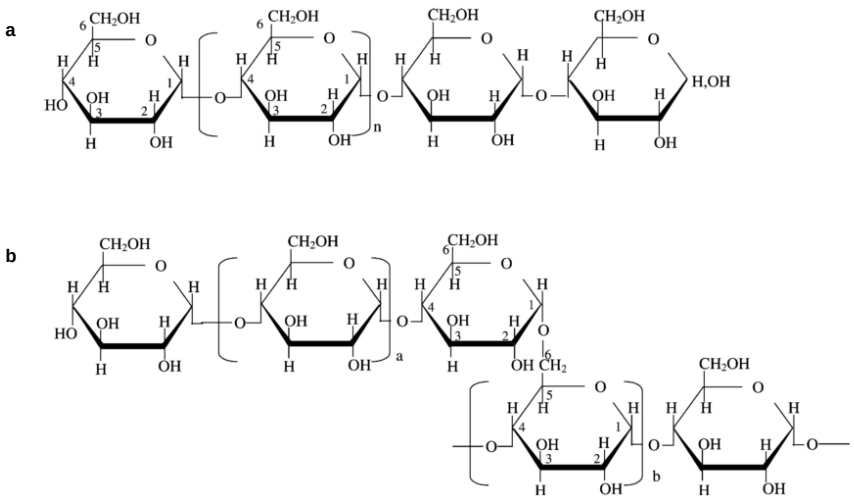


**Fig. 2.1** The structures of (**a**) amylose and (**b**) amylopectin adapted from Tester et al. (2004).

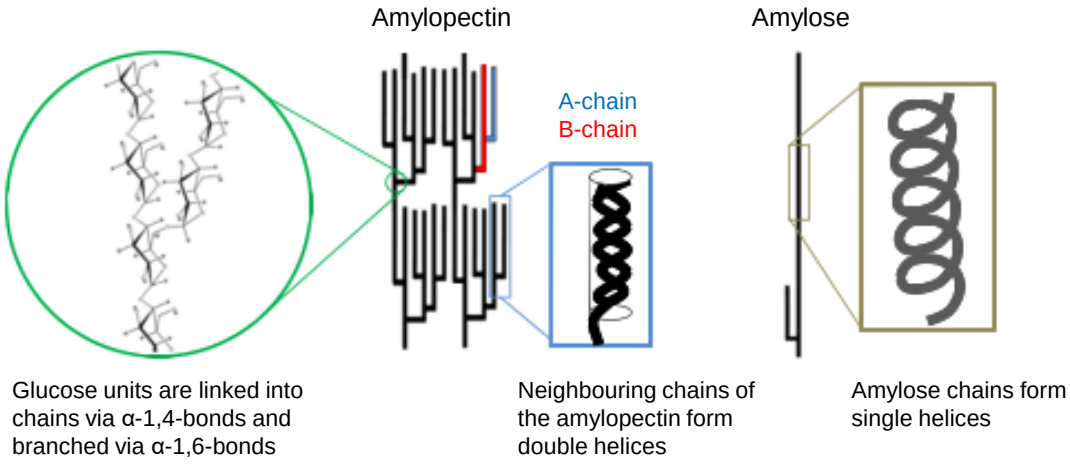


**Fig. 2.2** Schematic representation of starch structures. Amylopectin is characterised by left-handed double helices with six glucose units per turn that form between A-chains or longer B-chains. Amylose is linear and forms single helical structures. Adapted from Streb and Zeeman (2012).

Native starch can be classified into 3 groups (A-type, B-type and C-type) based on its crystalline polymorphs that result from differences in amylopectin packaging (Fig. 2.3) (Gallant et al. 1997; Streb and Zeeman 2012). These differences can be detected using X-ray scattering or solid state $^{13}$C-NMR spectroscopy (Buléon et al. 1998). The A-type generally occurs in cereal starches (wheat and corn) and is formed in warm and dry conditions. The

B-type crystallites are found in tubers, bananas and high-amylose starches and are formed in cold and wet conditions (Liu 2005). A-type crystallites have a chain length of about 23 - 29 glucose units, while the more open hydrated B-type hexagonal crystallites contain longer unbroken chain lengths of about 30 - 44 glucose molecules. The C-type structure contains a combination of A-type and B-type crystallites and is present in peas and beans.

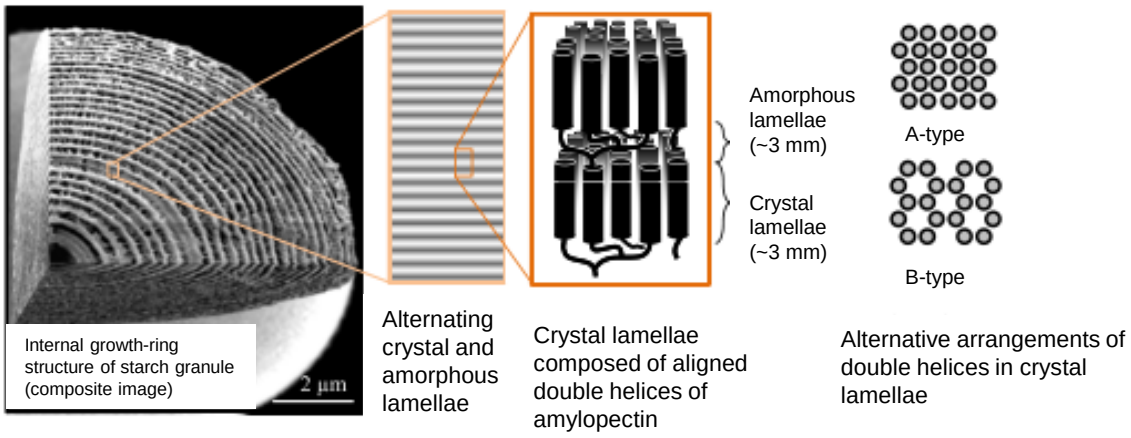


**Fig. 2.3** Amylopectin helices are arranged into ordered crystalline lamellae of two types, A-type (tightly packed) and B-type (open hexagonal pattern with a central, water-filled space). Crystalline lamellae alternate with amorphous lamellae and make up the growth rings (visible with light and electron microscopy). Adapted from Streb and Zeeman (2012).

The crystal structure of starch (determined by X-ray diffraction) is a major factor that affects resistance to hydrolysis (Sajilata et al. 2006). Starch crystallinity can vary between 15 - 45% depending on the origin of starch and its hydration level (Tawil et al. 2011). Starch digestibility is also influenced by the arrangement in A-type or B-type crystallites. Generally, A-type crystallites are reported to show a higher susceptibility to enzymatic hydrolysis than the B-type crystallites (Lehmann and Robin 2007). However, since crystallinity accounts for less than 50% of the starch granule, it is not the principle mode of organisation (Gallant et al. 1997). Consequently, the level of helical order is often more significant to the structure of starch, compared to the extent of crystalline order. Starch granules are comprised of alternating semi-crystalline and crystalline shells (Fig. 2.3) and a large part of the amylopectin that forms the semi-crystalline shells is in the double helical form, rather than the crystalline form.

### 2.1.2 Different types of starch

Besides classification by crystal structure, starch can be divided into native or modified starch. Native starches, also referred to as raw starches, are obtained directly from the plant biomass without additional processing steps. However, the use of native starch is limited by its thermal resistance and shear resistance; both of these factors are influenced by the granule rigidity,

lipid content and amylose:amylopectin ratio (Abbas et al. 2010; Singh et al. 2010). Modified starch, on the other hand, is native starch that has been physically or chemically changed (esterification, etherification, phosphorylation or pre-gelatinisation) to acquire desirable characteristics for specific industries; it is also referred to as soluble starch (Singh et al. 2010). Modified starch is used to improve cooking characteristics, reduce gelatinisation, increase the transparency and improve the texture of pastes. In the food industry, modified starch is used as a stabiliser, emulsifier, thickening agent, clouding agent, suspending agent and for freeze-thaw stability (Abbas et al. 2010). Modified starch also plays an important role in the paper, textile, plastic and biofuel industries. The use of starch requires the disruption of starch granules through acid/alkaline pretreatment, hydrothermal treatments or enzymatic hydrolysis. The latter is the most favoured process, since it is energy efficient and environmentally friendly.

The large demand for new starch resources has led to increased research efforts investigating structure, properties and possible applications of starch (Korus et al. 2004). Cereals (e.g. corn, wheat, rice, oats and barley) contain around 60 - 80% starch, legumes (e.g. chickpeas, beans and peas) 25 - 50% starch and tubers (e.g. potato and cassava) 60 - 90% starch (Santana and Meireles 2014). Several types of starches are known as “waxy” starches as a result of the shiny and wax-like appearance of the endosperm tissue from which they are derived. Waxy starch is almost comprised exclusively of amylopectin, with a minimal amount of amylose (< 15%). Due to their high crystallinity, these starches require a higher energy input for gelatinisation compared to normal starch (15 - 30% amylose) (Alcázar-Alay and Meireles 2015).

Starch quality is influenced by the lipid, protein and phosphorous content (Slattery et al. 2000; Santana and Meireles 2014). Cereal starches have a very low phosphate content and moderate viscosity. The amylose in these starches is complexed with lipids that form a weak crystalline structure and reinforce the granule structure (van der Maarel et al. 2002). On the other hand, potato starch has less crystallisation and a distinctly higher concentration of covalently bound phosphate, compared to cereal starches (O’Neill and Field 2015). The higher phosphate content is correlated with increased starch viscosity and it is likely that this will reduce the enzymatic hydrolysis of both modified and raw potato starch (Noda et al. 2008). The phosphate content reduces the hydrolytic ability of the amylases, since it hinders the attachment of amylases to the amylose chain containing phosphorylated glucosyl residues. Subsequently, phosphoryl-oligosaccharides are released during potato starch hydrolysis.

### 2.1.3 Amylases

Amylases are glycoside hydrolases hydrolysing O- and S-glycosyl compounds and thus have an Enzyme Commission number (EC number) of 3.2.1. The documented use of amylases dates back to the 9th century AD when malt was used to produce sweetener from arrowroot starch (Muralikrishna and Nirmala 2005) and the first starch hydrolysing enzyme was discovered by Kirchhoff in 1811 (Naidu and Saranraj 2013). The commercial use of fungal and bacterial amylases developed in the late 19th and early 20th century, respectively and by the 1930s these enzyme were being used in a number of different industrial sectors.

Amylases play a key role in present day biotechnology with applications ranging from food production, fermentation, biopharmaceutical applications, medicinal and clinical chemistry, as well as in the textile and paper industries (Pandey et al. 2000; Lee et al. 2015). Specifically, barley amylases are often used in the brewing industry, while fungal amylases are associated with commercial enzyme cocktails for raw starch hydrolysis (Gohel and Duan 2012) and the preparation of oriental foods (Sivaramakrishnan et al. 2006). During enzyme production, factors that influence the implementation of these enzymes include the yield, stability and production costs (Naidu and Saranraj 2013; Das et al. 2011). Emerging genetic engineering tools have facilitated the use of recombinant amylases and further promoted their use in industrial applications (Muralikrishna and Nirmala 2005). Recent technological advances in molecular biology allow for the manipulation of organisms with the aim of obtaining enzymes with desired or enhanced characteristics (Abdel-Fattah et al. 2013).

#### *2.1.3.1 α-Amylases*

The α-amylases (EC 3.2.1.1) are grouped in the glycoside hydrolase family 13 (GH 13) (Mehta and Satyanarayana 2013). This group comprises the largest family of glycoside hydrolases, with the majority of enzymes acting on starch, glycogen and related polysaccharides. An α-amylase consists of a single polypeptide chain that is folded into three domains (A, B and C) (Fig. 2.4). Domain A is the catalytic domain, domain B has an irregular structure and domain C is believed to stabilise the catalytic site of the enzyme by protecting the hydrophobic patch (Singh and Kayastha 2014). Most of the enzymes have an active site cleft found between domains A and B where a triad of catalytic residues (Asp, Glu and Asp) perform catalysis.

The α-amylase enzymes play a dominant role in carbohydrate metabolism and have entirely replaced the use of chemical hydrolysis in the starch-processing industry (Gupta et al. 2003; de Souza and de Oliveira Magalhaes 2010). α-Amylases were the first enzymes produced for use on a commercial scale and the annual sale in the global market was estimated at around $11 million (Sivaramakrishnan et al. 2006). These enzymes differ widely in their action

patterns and specificity. They originate from a variety of different hosts, e.g. humans, animals, plants, bacteria, yeast and fungi. However, microbial amylases from fungal and bacterial sources are more pH and temperature stable and are thus preferred for industrial applications. The production of α-amylases from *Bacillus licheniformis* and *Aspergillus sp*. comprises around 300 tonnes of pure enzyme protein per year (Sivaramakrishnan et al. 2006). The molecular weights of α-amylases vary from 10 - 210 kDa, but the size of microbial α-amylases is usually between 50 - 60 kDa (Gupta et al. 2003).

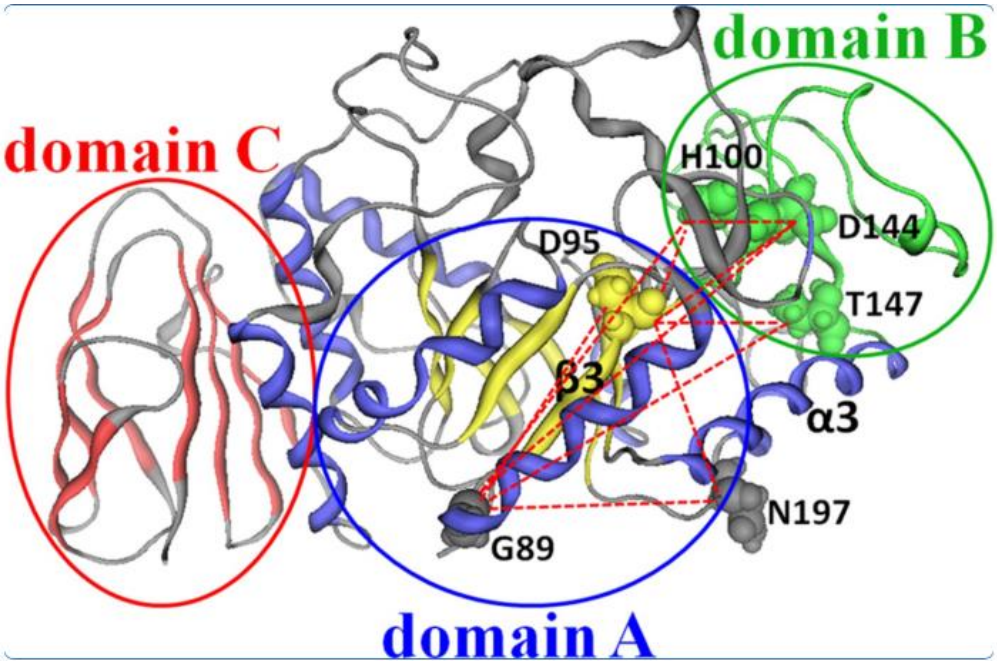


**Fig. 2.4** The three domains (A, B and C) of the α-amylase from *Bacillus subtilis* CN7 (Amy7C) (Wang et al. 2012).

Most α-amylases display maximum activity in the pH range between 4.5 – 7.0 (van Zyl et al. 2012) and thermostable acidic α-amylases are preferred for the industrial hydrolysis of starch, since starch slurry has a pH around 4.5 (Sharma and Satyanarayana 2013). Furthermore, from an economical and technical perspective, thermostable α-amylases are beneficial since they allow for higher operational temperatures. On the other hand, amylase candidates that display optimum activities closer to the temperature for recombinant yeast cultivation will improve the fermentation rate at lower temperatures (i.e. 30 - 37°C) by increasing the rate of starch hydrolysis (Carrasco et al. 2016). Alternatively, alkaline amylases are desirable for the detergent and food industries (Das et al. 2004), thus prompting searches for microbial strains expressing α-amylases with these properties.

#### *2.1.3.2 Glucoamylases*

Glucoamylases (glucan α-1,4-glucosidase, EC 3.2.1.3) are grouped in the GH 15 family and are exo-acting enzymes catalysing the hydrolysis of α-1,4- and α-1,6-glucosidic linkages, which release the inverted β-D-glucose from the non-reducing ends of starch (Chen et al. 2012). These enzymes are mainly used for the production of glucose syrup, high fructose corn syrup and bioethanol. These saccharifying enzymes have the ability to degrade large oligosaccharides containing up to 90% α-l,6 linkages. The pH and temperature optima of glucoamylases are generally in the range of 4.5 - 5.0 and 46 - 60°C, respectively, and these enzyme are relatively stable at higher temperatures (van Zyl et al. 2012). However, a few thermophilic strains produce glucoamylase with an optimum temperature of 70°C, such as *Rasamsonia emersonii*, *Halvina lanuginosa* and *Aspergillus niger* IMDCC No. 120 (James and Lee 1997). The molecular masses for glucoamylases can vary from 55 kDa (raw starch degrading glucoamylase from *Saccharomycopsis fibuligera*) to about 300 kDa (glucoamylase from *S. cerevisiae* var. *diastaticus*) (Hostinová and Gašperík 2010).

Glucoamylases are produced by a wide range of microorganisms. However, enzymes for commercial applications are produced by filamentous fungi because they are capable of secreting large quantities of the enzyme extracellularly. The industrial production of glucoamylases has focussed on *A. niger*, *Aspergillus awamori* and *Rhizopus oryzae* due to the stability of their enzymes (Lin et al. 2007). The most recent and comprehensive data on different glucoamylase producing strains has been reviewed by Kumar and Satyanarayana (2009).

#### *2.1.3.3 Production of industrial amylases*

Low-value agricultural residues have gained much interest over the last few decades, since they can be used an inexpensive raw material for enzyme production (Pandey et al. 2000). Industrially important enzymes have traditionally been produced using submerged fermentation, since it offers the benefits of controlling different parameters e.g. pH, temperature, aeration and oxygen transfer, as well as moisture. Solid state fermentation systems, however, are a promising alternative since they resemble the natural habitat of microorganisms (de Souza and de Oliveira e Magalhães 2010; Sundarram et al. 2014).

The bulk enzyme production requires minimal downstream processing and is often used as crude preparations, whereas enzyme applications in pharmaceutical and clinical sectors require high purity amylases (Pandey et al. 2000). Thus, the development of purification techniques will greatly enhance the use of these enzymes and enable additional applications in the medical sector. Enzyme are industrial catalysts and they need to tolerate the relatively harsh conditions that are often associated with industrial processes. Therefore, the pH profile,

pH stability and thermostability of enzymes are important factors to be considered in the development of fermentation processes (de Souza and de Oliveira e Magalhães 2010).

#### *2.1.3.4 Enzyme synergy*

Synergy occurs when the observed action of two or more enzymes (acting together in solution) is greater than the sum of their individual action (Wood and Garcia-Campayo 1990) and it is often reported as a percentage enhancement of activity (Gottschalk et al. 2010). Enzyme synergy is an important optimisation factor to consider when more than one enzyme is required for hydrolysis i.e. the saccharification step in a raw starch to ethanol process. A number of studies have reported a synergistic relationship between α-amylases and glucoamylases with varying α-amylase to glucoamylase activity ratios (Görgens et al. 2015). A report by Wong et al. (2007) suggested that the ratio between α-amylase and glucoamylase activities can be in the range of 3:1 to 1:3, on condition that the total enzyme activity remains constant and there is sufficient amounts of both enzymes. The review by Görgens et al. (2015) reported that efficient raw starch hydrolysis can be accomplished even though the ratio between glucoamylase and α-amylase activities varies; provided the α-amylase is in excess and there is sufficient activity levels for both enzymes.

Substantially more amylase activity, specifically α-amylase activity, is needed for raw starch digestion than is the case with soluble starch digestion. The rate limiting step in starch hydrolysis is considered to be the conversion of raw starch oligosaccharides. Therefore, α-amylase has a far more significant role compared to that of glucoamylase because it provides numerous non-reducing ends that are available as substrates for glucoamylase activity (Wong et al. 2007; Yamada et al. 2010). This is contrary to the conventional cooked starch hydrolysis process, where a higher dosage of glucoamylase is required. The minimum enzyme dosages for efficient hydrolysis of raw corn starch were reported as 5 U.g starch$^{-1}$ for glucoamylase and 10 U.g starch$^{-1}$ for α-amylase (Görgens et al. 2015).

### 2.1.4 Starch binding domain

Carbohydrate-binding modules (CBMs) are non-catalytic ancillary domains, which function independently of the catalytic domain and are often present in glycoside hydrolases. CBMs that have an affinity for insoluble raw starch are generally referred to as starch binding domains (SBDs) (Peng et al. 2014). The SBD plays a fundamental role in granular starch hydrolysis by performing several simultaneous functions (Fig. 2.5). It binds to the starch molecules and thereby increases the concentration of substrate at the catalytic site. A strong correlation between raw starch hydrolysis and the adsorption of amylases to raw starch granules has been described for bacterial, yeast and fungal α-amylases and glucoamylases (Mitsuiki et al. 2005). The SBD may also disturb the structure of the starch by disrupting

polysaccharide chain interactions on the granule's surface, which subsequently enhances the amylolytic rate (Santiago et al. 2005; Barchiesi et al. 2015). It is usually composed of about one hundred amino acid residues and is present in a number of amylolytic enzymes of the glycoside hydrolase families (Fig. 2.5a). It is comprised of several β-strand segments forming an open-sided, distorted β-barrel structure (Fig. 2.5b) and it is connected to the catalytic domain by a glycosylated linker region (Juge et al. 2002; Barchiesi et al. 2015).

Sequence-based classification divides SBDs into the following ten CBM families: 20, 21, 25, 26, 34, 41, 45, 48, 53 and 58 (Peng et al. 2014). The CBM20 family has been shown to function in granular/raw starch binding and is the most generalised and well-studied family of SBDs. There are two starch-binding sites (SBS) in the CBM20s (Fig. 2.5b) that are positioned on the exterior of the active site area (Cockburn et al. 2014) and they are thought to have different functions. It has been reported that site 1 is probably used as the initial starch recognition site, whereas site 2 is associated with the specific recognition of certain regions of the starch polymer (Sorimachi et al. 1997).

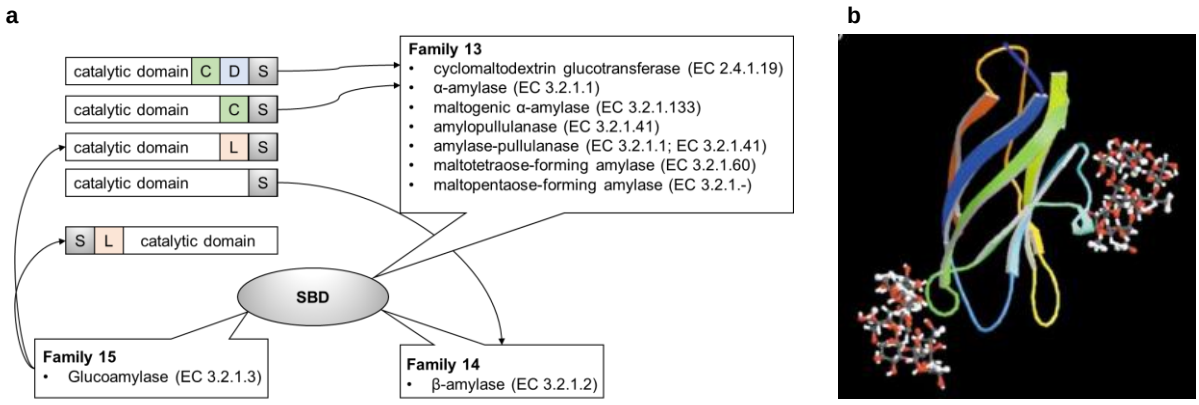


**Fig. 2.5** Starch binding domain (SBD) (**a**) occurrence within the glycoside hydrolase (GH) families 13, 14, and 15. L: long *O*-glycosylated linker region, C: C domain, D: D domain, S: SBD (adapted from Juge et al. 2002) and (**b**) *A. niger* glucoamylase SBD showing eight-stranded Greek key topology and two substrate binding sites (Rodríguez-Sanoja et al. 2005).

The CBM20s are grouped based on their amino acid sequences, substrate binding specificities and position in protein (middle, N- or C-terminal) (Barchiesi et al. 2015) (Fig. 2.5a). Microbial amylases that are involved in raw starch metabolism often contain a SBD at the C-terminus of the protein (Latorre-García et al. 2005), except for the α-amylase from *Thermoactinomyces vulgaricus* and the glucoamylase from *R. oryzae* that contain an N-terminal SBD (Santiago et al. 2005). A dramatic decline in raw starch hydrolysis has been shown when this domain is removed from the amylolytic proteins (Peng et al. 2014). Yet, the

α-amylase of *Saccharomycopsis fibuligera* lacks a distinct SBD, but has the ability to degrade raw starch (Janeček et al. 2014).

A common approach to engineering enzymes for starch hydrolysis is the addition of a SBD, which should theoretically result in increased affinity for starch (Dalmia and Nikolov 1991). The *A. niger* glucoamylase is produced in two main forms: glucoamylase-I (GA-I) and glucoamylase-II (GA-II). In a study by Dalmia and Nikolov (1991) the results showed that the presence of the SBD in the GA-I molecule lead to a 100-fold increase in affinity for the starch surface, compared to the GA-II molecule (which lacks a SBD). Furthermore, SBDs can also serve as affinity tags thereby facilitating protein purification (Latorre-García et al. 2005).

### 2.1.5 Starch processing

Starch firsts needs to be hydrolysed into fermentable sugars (glucose) before it can be converted to ethanol. This makes the use of starchy feedstocks a more complicated process compared to the fermentation of simple sugars because more processing steps are required. In order to achieve complete degradation of starch, two main groups of amylolytic enzymes are required: α-amylases for liquefying starch and glucoamylases for the saccharification step (Białas et al. 2014). The α-amylase supply shorter oligosaccharides by an endo-wise random disruption of the large molecules (Fujii et al. 1988). The endo-catalytic events increase the number of substrate sites (non-reducing ends) for the exo-acting enzymes; this action is synergistic and leads to enhanced conversion rates (Robertson et al. 2006).

Conventional starch processing comprises two hydrolysis stages followed by fermentation (Fig. 2.6). First, the starch slurry is gelatinised in a jet cooker at 100 - 105°C, for 5 minutes. This is followed by two liquefaction steps using thermostable α-amylase at high temperatures (> 95°C, for a maximum of 3 hours). The second hydrolysis stage, saccharification, involves cooling of the slurry to 60°C and the addition of a glucoamylase to release fermentable sugars (Robertson et al. 2006; Mehta and Satyanarayana 2014). This conventional process has several disadvantages, such as the high energy input required to reach the temperatures needed for gelatinisation, as well as pH adjustment steps (illustrated in Fig. 2.6). Commercial α-amylases have a pH optimum around 6 - 6.5 and therefore the pH of the hydrolysis environment (fermentation broth) needs to be lowered (pH 4 - 5) to provide optimal hydrolysis conditions before glucoamylase addition (Liakopoulou-Kyriakides et al. 2001).

The direct hydrolysis of raw starch at sub-gelatinisation temperatures has the potential to substantially reduce operation costs. However, raw starch degrading enzymes (RSDEs) are required because the substrate composition is more compact compared to that of gelatinised starch. The RSDEs result in the formation of pores, as the granules are broken apart, and these pores facilitate the diffusion of the amylases. Many industrial and biological processes

employ enzymatic hydrolysis e.g. starch metabolism in plants, digestion by mammals, fermentation processes, glucose syrup manufacturing and bioethanol production (Tawil et al. 2011). However, although fermentable sugars can be produced in a more energy efficient process, the use of commercial enzymes is expensive. Therefore, the processing of low-cost starchy biomass for ethanol production using the cold hydrolysis process is not yet economically viable.

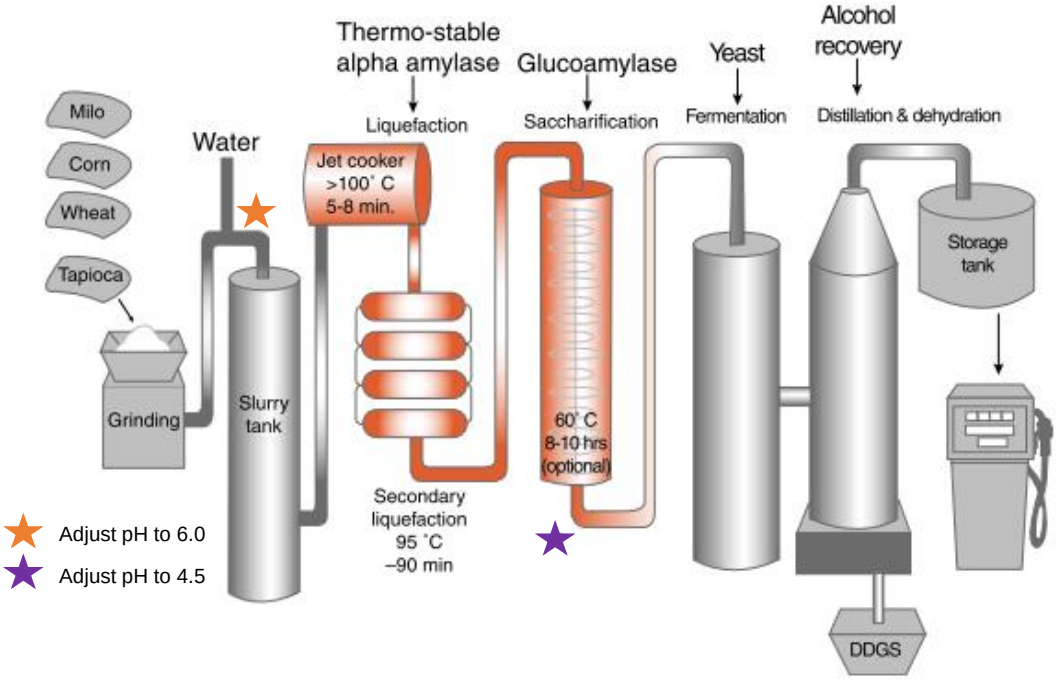


**Fig. 2.6** Conventional process for the conversion of starch to ethanol: liquefaction, saccharification and fermentation, indicating two pH adjustment steps and DDGS (distiller's dried grains with solubles). Adapted from Schubert (2007).

One of the main goals in obtaining cost-effective starch conversion is the expression of starch hydrolysing genes by a fermenting yeast. The consolidated bioprocessing (CBP) approach involves the use of a single organism that is able to achieve liquefaction, hydrolysis and fermentation of starch in a single fermentation vessel (van Zyl et al. 2012). The engineering of yeast to utilise alternative substrates for the production of value added products and chemicals can then be investigated once efficient conversion of raw starch has been achieved.

### *2.1.5.1 Gelatinisation of starch*

Liquefaction comprises two steps: gelatinisation and dextrinisation. Gelatinisation is the permanent alteration of the starch granule from ordered, semi-crystalline granules to an amorphous state and occurs in the presence of water (Ai and Jane 2015). During heating, water enters the amorphous space of the starch molecules and this results in swelling, which causes the hydrogen bonds to break (Aiyer 2005). Helical structures unwind and the crystal structure dissolves into an amorphous form, as the amylose leaches out of the granule.

Subsequently, the slurry viscosity increases to form a gel/paste. Since the intermolecular bonds of the starch molecules are broken, the reactivity of cooked starch towards amylolytic enzymes is significantly enhanced (Szymanowska-Powałowska et al. 2012). The process of starch recrystallisation is called retrogradation. It occurs after a starch gel changes from a dissolved and dissociated state to an associated state (van der Maarel et al. 2002).

Starch gelatinisation temperature is determined by the distribution and chain length of amylopectin (Jane et al. 1999). The gelatinisation temperature is the temperature at which the phase transition of starch granules occurs, from an ordered to a disordered state (Ubwa et al. 2012). The gelatinisation onset temperature ($T_o$), peak temperature ($T_p$) and conclusion temperature ($T_c$) are used to compare thermal properties of starch (Ao and Jane 2007). Conditions for starch gelatinisation are determined by the plant type, degree of cross-linking of the amylopectin, as well as the salt, sugar and lipid content. B-type starches exhibit a lower gelatinisation temperature ($T_0$ 58.2 °C) compared to most A-type starches, even though their chain length is slightly longer (Ai and Jane 2015).

Cereals with a high amylose content are more resistant to gelatinisation compared to those with normal to high amylopectin content (Gómez et al. 2016). Gelatinisation temperatures for starches are in the following ranges: white corn, 72 - 78°C; yellow corn, 66 - 72°C; white sorghum, 74 - 82°C; brown sorghum, 74 - 82°C; potato, 55 - 66°C; wheat, 52 - 66°C; triticale, 55 - 70°C and rice, 66 - 82°C. Identifying starch species with lower gelatinisation temperature will help reduce the cost of food processing. It will also assist in determining which starches are suitable for use in the pharmaceutical industry e.g. as a binder and in drug coating (Ubwa et al. 2012).

#### *2.1.5.2 Raw starch hydrolysis*

At low temperatures (<50°C) starch is in its solid phase i.e. not gelatinised and insoluble in aqueous media (Cinelli et al. 2015). Raw starch granules have a densely compacted polycrystalline structure and a larger particle size in comparison to gelatinised starch, this increases the difficulty of enzymatic hydrolysis (Lin et al. 2011). The first step in raw starch digestion is the adsorption of amylases to the granule's surface and it is possible that enzyme-substrate interactions include imperfect docking or binding events that do not result in hydrolysis (Robertson et al. 2006). Several factors including the physical architecture (degree of branching), molecular structure of the solid substrate and the molecular configuration of the enzyme may result in inefficient catalysis. Additional factors that influence the hydrolysis of starch in its heterogeneous phase include: enzyme diffusion in the medium, substrate accessibility, recognition leading to adsorption, formation of the enzyme-substrate complex and catalytic action (Gallant et al. 1997; Tawil et al. 2011).

It was previously believed that the hydrolysis of starch started at the surface of the granule. However, recent studies have shown that native cereal starches, namely corn and sorghum, contain peripheral pores and channels, which facilitate the penetration of α-amylase. This allows for hydrolysis to begin inside the granule and work progressively toward the surface. Dissimilarly, potato and other B-type starches (tuber starches) are digested by exo-corrosion, whereby enzyme attack starts on granule's surface. These differences explain why cereal starches are more easily digested compared to tuber starches. In addition, the particle size of the starch and the surface area affect the hydrolysis process (Kim et al. 2008). Higher enzymatic susceptibility is noted for starch with a smaller granule size, whereas a large, smooth surface (together with specific supramolecular properties) explains the resistance of potato granules to enzymatic attack (Lehmann and Robin 2007).

RSDEs are ubiquitous in nature and can be produced by recombinant hosts on a large scale (de Souza and de Oliveira e Magalhães 2010; Cinelli et al. 2015). The hydrolysis of raw starch is significantly slower than that of soluble starch because it takes longer for the double helical structures to unwind. Therefore, increased enzyme loadings are needed for raw starch hydrolysis compared to soluble starch hydrolysis. When the four most economical native starches are ordered in relation to their susceptibility to hydrolysis then (in diminishing order) corn ≥ wheat > cassava > potato (Szymanowska-Powałowska et al. 2012).

The efficient and cost effective conversion of raw starch into ethanol is the major hurdle in the production of biofuel by simultaneous saccharification and fermentation (SSF). Presently, low protein yield (in the enzyme production process) and high cost are the two main factors that impede the application of RSDEs. Therefore, several methods are being explored in order to make enzymatic hydrolysis economically attractive, such as the overexpression of genes encoding for RSDEs in recombinant strains, the use of cheaper raw materials, optimised media and culture conditions, as well as efficient downstream processes for enzyme recovery (Sun et al. 2010).

There are several advantages to producing bioethanol from raw starch using SSF (cold hydrolysis) as opposed to the conventional "cooked starch" process. Omitting liquefaction and starch gelatinisation makes the SSF method an attractive strategy for bioethanol production and the use of RSDEs has the potential to reduce the total energy cost for ethanol production by approximately 30% (Vu et al. 2010). However, it is important that the recombinant enzymes used for hydrolysis perform well at fermentation temperatures (around 30 - 37°C) (Cinelli et al. 2015). Currently, there is no industrial process that relies solely on recombinant yeasts for direct ethanol production from raw starch (van Zyl et al. 2012).

The use of granular/raw starch hydrolysing enzymes can eliminate the use of high temperatures used for starch gelatinisation and improve the efficiency of starch conversion into ethanol (Gohel and Duan 2012). STARGEN 001™ is a commercial amylase preparation designed for hydrolysing raw starch from cereals such as corn, while STARGEN 002™ is designed for hydrolysing rye, wheat, triticale and barley (Genencor International, California, USA). These enzyme cocktails contain an α-amylase from *Aspergillus kawachi* and glucoamylase from *A. niger* (Huang et al. 2015) and are used in combination with a fermenting yeast. However, recent advances in the production of RSDEs (as a result of engineering recombinant expression hosts) and the commercial availability of RSDE cocktails has allowed for the development of low temperature processes for starch SSF (Görgens et al. 2015). The next step will be to decrease the reliance on commercial amylases by using a CBP yeast that produces the recombinant enzymes required for the one-step conversion of starch to ethanol.

### 2.1.6 Industrial starch to ethanol processes

Over the last decade, the development of recombinant yeast for CBP has increased significantly. Not only for lignocellulosic ethanol production, but also for ethanol production from starch. Although CBP technologies have not yet been fully integrated into industrial processes, there are several facilities that use semi-CBP yeast to produce ethanol. Mascoma Corporation have developed a recombinant *S. cerevisiae* strain that expresses a glucoamylase gene (TransFerm®, supplied by Lallemand, Inc) and the strain has also been further engineered to have reduced glycerol production (TransFerm® Yield+) (http://www.lallemandbds.com). Lowering glycerol concentrations under anaerobic conditions is a key factor for improved ethanol production (Basso et al. 2008). Glycerol is used as an electron acceptor to balance yeast metabolism and if less glycerol is produced then there is more carbon available that can be used for ethanol production. The TransFerm® ethanol technology allows for a drop-in MGT™ (Mascoma grain technology) yeast to produce ethanol from liquefied grains, thus eliminating the need for exogenous glucoamylase addition. Optimal processing conditions for this technology include a temperature range between 30 - 35°C for fermentation.

TransFerm® and TransFerm® Yield+ yeasts are commercially available and represent the first yeast strains to be engineered for starch saccharification and fermentation that have gone through formal FDA (food and drug administration) and EPA (environment protection authority) review (www.mascoma.com). This technology is currently employed in approximately 20% of operational corn ethanol facilities in the United States. It has combined the conventional saccharification and fermentation steps of the starch to ethanol process, thereby simplifying the process and lowering enzyme costs. However, the MGT™ yeast only

produces a glucoamylase enzyme and therefore the heating step is still required for starch liquefaction. Consequently, it remains an energy intensive process.

POET, on the other hand, uses BPX™ technology involving the SSF of raw starch. Together with Novozymes (a leading commercial enzyme provider) they have developed a patent protected blend of α-amylases and glucoamylases. BPX™ technology (promoted to commercial scale in 2004) eliminates the use of a jet cooker to liquefy starch (www.poet.com). This technology represents an industrial "cold-process" and has been implemented in nearly all of POET's ethanol production facilities (24 out of 27), with an annual ethanol production of $5.8\times10^9$ litres (Rasmussen et al. 2015). Furthermore, since high temperatures to cook to starch are avoided, the DDGS (distiller's dried grains with solubles), which are a byproduct of the dry grind process, have increased nutritional value and desirable characteristics e.g. colour and increased amino acid digestibility. Since POET does not use antibiotics to control contamination in their starch to ethanol process, their certified DDGS can be fed to laying hens and other livestock (Smith 2011). The quality of DDGS is important because in addition to animal feed, they are value-added products for several industrial applications (Muthaiyan et al. 2011).

Commercial enzymes are costly, therefore it would be of great value to the bioethanol economy if a "drop-in" CBP yeast could be used that expresses both an α-amylase and glucoamylase. Effective fermentations to produce ethanol using a CBP yeast will require tolerance to high concentrations of both glucose and ethanol (Alper et al. 2006). Ethanol producing strains need to produce high yields of ethanol, reaching approximately 90 - 93% of the theoretical maximum of 0.51 grams ethanol per gram glucose (Bai et al. 2008). Glycerol concentrations depend on fermentation conditions and typically reach levels of about 1.0% ($w.v^{-1}$). Furthermore, an essential factor that contributes to downstream purification processes and costs is the final ethanol titre, since increasing yeast's ethanol productively lowers capital costs. Therefore, the ethanol industry is continually searching for novel yeast strains that could meet these standards. Even a small improvement in the ethanol titre, yield or productivity will have a significant effect on the overall production costs (Bai et al. 2008).

#### *2.1.6.1 Microbial contamination*

The high temperatures used in the conventional process for starch liquefaction (90 − 105°C) not only gelatinises the starch but also help in reducing bacterial contamination (Rasmussen et al. 2015). Proliferation of unwanted lactic acid bacteria (LAB) results in ethanol yield losses, which can have significant economic consequences (Muthaiyan et al. 2011). These bacteria compete with the ethanologenic yeast for nutrients such a glucose (Skinner and Leathers 2004) and severe LAB contamination can lead to ethanol plant

shutdowns for cleaning; this negatively affects ethanol productivity (Rasmussen et al. 2015). In order to reduce microbial contamination ethanol distilleries have traditionally used antibiotics such as penicillin, erythromycin, tylosin and virginiamycin (Olmstead 2012). However, this has led to antibiotic resistant bacterial strains and the FDA have prohibited or limited sales of DDGS that are contaminated with antibiotic residues.

In the last decade, many ethanol plants have changed their approach towards controlling microbial contamination. The Institute for Agriculture and Trade Policy's (IATP) reported that more than 40% of ethanol plants in the USA now use some form of antibiotic-free antimicrobial (IATP 2009). Alternative strategies to control microbial contamination during bioethanol production include, ozonation of uncooked corn mash (Rasmussen et al. 2015) and the use of chemical treatments (e.g. acid), natural compounds (e.g. chitosan and bacteriocins), as well as plant-derived compounds (e.g. hops) (Muthaiyan et al. 2011). Furthermore, the use of bacteriophage endolysins represents a novel approach for reducing the occurrence of *Lactobacillus fermentum* during fermentation (Khatibi et al. 2014; Liu et al. 2015). In addition to antimicrobials, several processing strategies are a routine practice to help control microbial contamination, such as lowering the pH of the mash to 4.0 or less and using large yeast inocula (≥ 2% $v.v^{-1}$) (Narendranath and Power 2005).

#### *2.1.6.2 Low value byproducts*

In 2014, approximately 56 billion litres of corn ethanol was produced in the USA. Resulting in large amounts of fibre-protein rich byproducts (e.g. 40 million metric tons of distiller's grains) (Xiang and Runge 2016). The two major processes in the industrial production of ethanol from corn are (i) wet mill and (ii) dry grind, with the latter being the most predominant. The resulting product from the wet milling step in corn starch processing is called corn bran. This recalcitrant cereal byproduct consists of the outmost layers of the corn kernel and is mainly comprised of saccharides namely, the monosaccharides arabinose and xylose (pentose sugars) (Agger et al. 2011).

Utilisation of these byproducts will add value to the overall starch to ethanol process. Arabinoxylan contributes to the rigidity of cell wall structure and is often associated with the hardness of the endosperm (Gamlath et al. 2008). Corn is used in large quantities for bioethanol production and it is estimated that of the 5% ($w.w^{-1}$) of processed substrate is corn bran. Therefore, the hydrolysis of corn bran using additional enzymes (such as xylanases) has the potential to increase the sustainability of the biofuel industry and make ethanol production from starchy biomass more attractive.

## 2.2 Engineering strains for starch hydrolysis

Novel microorganisms are constantly being evaluated for amylase production. Although amylases have been used for decades, only a few fungal and bacterial strains have the ability to produce these enzymes at levels that meet the criteria for commercial production (de Souza and de Oliveira e Magalhães 2010). Therefore, an attractive alternative is to heterologously express amylases in an appropriate host. The heterologous expression of genes plays a fundamental role in functional genomics initiatives. Numerous examples of eukaryotic and prokaryotic systems are used for the expression of foreign genes and the subsequent protein production. Thus, heterologous expression is in theory a straightforward practise (Wu et al. 2004).

Several factors need to be considered when constructing a robust recombinant host organism. In order to avoid bottlenecks in gene expression these factors include: transcription, translation, protein folding, glycosylation, cell viability, secretion signals and synonymous codon usage (Damasceno et al. 2012; Xiao et al. 2014). Proteins that are larger than 100 kDa are generally expressed in eukaryotic expression systems, while prokaryotic systems are used for proteins smaller than 30 kDa (Botte et al. 2016). Mammalian cells, fungi or the baculovirus system are routinely used for the production of recombinant proteins that require glycosylation.

The *Escherichia coli* overexpression system is the most convenient and frequently used cell-free system for the production of recombinant proteins; it is inexpensive and has a fast turnover (Yamaguchi and Miyazaki 2014; Demain and Vaishnav 2009). However, it is not suitable for expression of large multi-domain starch proteins. Overall, 39% of recombinant proteins are made by *E. coli*, 35% by Chinese hamster ovary cells (CHO cells), 15% by yeasts, 10% by other mammalian systems and 1% by other bacteria/other systems (Demain and Vaishnav 2009).

*Pichia pastoris* is also widely used for the production of recombinant proteins, since it can produce disulfide bonds and is capable of protein glycosylation (Daly and Hearn 2005). In comparison to *S. cerevisiae,* glycosylation is less extensive in *P. pastoris* due to shorter chain lengths of *N*-linked high-mannose oligosaccharides, usually up to 20 residues compared to 50 – 150 residues in *S. cerevisiae*. However, *Pichia* in unable to produce chaperone proteins that are essential for the proper folding of many proteins, which is a major disadvantages of using this host for heterologous gene expression (Demain and Vaishnav 2009).

### 2.2.1 Heterologous expression in *S. cerevisiae*

Initially the majority of synthetic biology tools were developed and tailored for gene expression in bacteria, especially for *E. coli*. However, over the last few decades, there has been an increase in yeast synthetic biological tools aimed at engineering yeast strains for industrial application. Furthermore, the production of fuels and chemicals from renewable biomass is being assisted largely by the advancements in molecular biology and recombinant DNA technologies, with the main focus on the model organism *S. cerevisiae* (Tsai et al. 2015).

*S. cerevisiae* is the most utilised organism for industrial bioethanol production due to its good fermentation capacity, high ethanol productivity, ethanol tolerance and GRAS (generally regarded as safe) status (Görgens et al. 2015; Shi et al. 2016). Subsequently, it has been extensively used for protein production. Recombinant protein production in yeast is advantageous for a number of reasons, namely proteins are secreted into the culture medium thus avoiding toxicity from intracellularly accumulated material and protein purification methods are simplified (Damasceno et al. 2012). The protein product is relatively pure, since cell harvest and disruption steps can be avoided (Madhavan and Sukumaran 2014).

An in-depth understanding of heterologous gene expression in *S. cerevisiae* has been generated over the last few decades with a wide range of genetic tools available for strain development and improvement. In addition, the genetic competence to accept foreign DNA has facilitated targeted chromosomal manipulations. However, with regards to the starch bioprocessing industry, the main disadvantage is *Saccharomyces'* lack of native starch hydrolysing enzymes (Eksteen et al. 2003), with the exception of the amylolytic *Saccharomyces diastaticus* strain (Adam et al. 2004). Therefore, it is necessary to use recombinant techniques, such as stable chromosomal integration, to engineer robust industrial *S. cerevisiae* yeast strains to produce extracellular amylases (Favaro et al. 2013).

The main challenge with regards to engineering a robust amylolytic yeast is the ability to hydrolyse raw starch at high concentrations, while simultaneously fermenting the hydrolysed sugars to ethanol (Favaro et al. 2015). Compared to laboratory strains, industrial strains have valuable characteristics including higher ethanol tolerance and a higher tolerance to acids and sugars (Gírio et al. 2010; Pereira et al. 2010). A well-used transformation platform for the heterologous expression of foreign genes in *S. cerevisiae* uses the high copy 2-micron plasmids, which can be maintained at around 10 - 50 copies per cell. However, industrial *S. cerevisiae* strains cannot be genetically engineered using these plasmids because gene stability requires the use of selectable markers and thus the need to maintain selection pressure (Shi et al. 2016).

Haploid yeast strains can be easily engineered because they exhibit good mating ability, they can readily take up exogenous DNA and they contain convenient selectable (auxotrophic) markers. The use of haploid laboratory strains allows for single genes to be disrupted and the resulting transformants can be easily selected. Thus, laboratory *S. cerevisiae* strains have been mutated and selected for easy handling in the laboratory (e.g. no flocculation). However, they are unsuited to industrial application because of their lower genetic stability and lack of robustness (Demeke et al. 2013).

Alternatively, gene integration into targeted DNA sequences on the yeast's chromosomes (δ-sequences of the Ty retrotransposon and ribosomal DNA) allows for multiple gene integration. This has assisted high expression levels in *S. cerevisiae* (Favaro et al. 2015). However, the use of δ-sites generates transformants that have different expression efficiencies. The exact positions of the integrated cassettes are unknown and their numbers could vary substantially among transformants (Kavšček et al. 2015).

Diploidisation is as a promising strategy for improved fermentation ability of industrial *S. cerevisiae* strains. The reason being that polyploidy yeast strains (including diploid strains) are usually more robust, have higher specific growth rates, biomass yields and tolerances to various stresses, compared to haploid strains (Yamada et al. 2011). These strains lack auxotrophic mutations (Hashimoto et al. 2005) making them suited to industrial applications. However, a number of technical challenges are associated with the genetic manipulation of industrial strains. They are genetically more complex compared to laboratory strains and often present aneuploidy (abnormal number of chromosomes) and/or polyploidy (more than two paired (homologous) sets of chromosomes), have poor sporulation efficiencies and unstable mating types (Steensels et al. 2014). Single gene disruption is more difficult to achieve in diploid or polyploid strains (Le Borgne 2012) because multiple copies of the same gene are present and all the copies on different chromosomes have to be inactivated. Furthermore, the ability to eliminate or recycle selection markers used during industrial strain construction is important due to the limited number of dominant selection markers available.

Traditional yeast chromosomal editing techniques for recombinant industrial strain construction were time consuming and are often plagued by low transformation efficiencies. However, the use of molecular biology cloning tools has helped overcome some of the difficulties associated with gene integration and yeast transformation. Recently the EasyClone vector set has been developed for integration of foreign DNA into *S. cerevisiae's* genome (Jensen et al. 2014) and the Cre–loxP system is frequently used for removing marker cassettes during the successive manipulation of multiply deletant strains (Stovicek et al. 2015). The CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 system is another engineering approach that allows for effective genome

editing and it makes use of marker-flanked integration cassettes. However, potential drawbacks of this technique include the requirement for a specific guide RNA (gRNA) plasmid for each target locus (Kavšček et al. 2015) and off-target effects, which cause unwanted mutations (Gupta and Musunuru 2014).

### 2.2.2 CBP yeast

CBP is a widely recognised concept that combines biological processes in a single reaction vessel and relies on the engineering of yeast strains (Olson et al. 2012); numerous examples of CBP yeast have been discussed for lignocellulose hydrolysis (Lynd et al. 2002). Furthermore, the engineering of CBP *S. cerevisiae* strains for raw starch hydrolysis and fermentation has made substantial progress over the last two decades. However, there are fewer studies reporting the use of amylolytic CBP *S. cerevisiae* strains for raw starch hydrolysis compared to cellulolytic CBP yeast. Although examples of genetically engineered strains for the production of ethanol from renewable feedstocks has been described (Favaro et al. 2015), there is limited experimental data showing the use and performance of these strains on an industrial scale. The main reason being that the hydrolytic ability of current recombinant amylolytic yeast strains has not yet reached the levels required for efficient raw starch hydrolysis.

Recent processes based on RSDEs are focusing on enhancing the amount of carbon conversion of the starch-based feedstock to glucose monomers, so that there is more glucose available for fermentation to ethanol (with minimal commercial enzyme supplementation). Molecular biology offers practical optimisation approaches (e.g. codon optimisation of nucleotide sequences) for improving the expression of genes in a foreign host, in order to enhance the production of amylases. Amylolytic strains for efficient CBP of natural starchy biomass were first described by Favaro et al. (2015); codon-optimised variants of the *Thermomyces lanuginosus* glucoamylase (TLG1) and *S. fibuligera* α-amylase (SFA1) genes were δ-integrated into two *S. cerevisiae* strains (M2n and MEL2) with promising industrial traits. However, current strains are unable to produce ethanol concentrations that can compare to those achieved from the conventional process (Görgens et al. 2015), due to ineffective starch hydrolysis to glucose. Therefore, there is still a need to improve raw starch conversion by recombinant amylolytic yeast.

Yeast cells can suffer from a variety of different stresses during a fermentation such as contamination, high temperatures (35 - 37°C), ethanol accumulation (> 15% ($v.v^{-1}$), acetic acid (> 0.05 ($w.v^{-1}$)) and lactic acid (> 0.8 ($w.v^{-1}$)) accumulation. The inhibitory effect of high ethanol concentrations can be intensified by the presence of other fermentation byproducts or high temperature fluctuations (Bai et al. 2008). One of the main disadvantages of CBP is that

recombinant enzymes for starch hydrolysis have an optimum temperature around 50 - 60°C, while fermenting microbes require temperatures between 20 - 35°C for ethanol production (Lin et al. 2012). High temperature fermentations are desirable for the ethanol industry because they reduce the costs associated with cooling (Abdel-Banat et al. 2010), as well as allow for shorter fermentation times and reduced operational costs (Banat et al. 1998; Qiu et al. 2015). Higher fermentation temperatures are closer to the enzymes' optimum temperature, thus allowing for an increased rate in substrate hydrolysis, which subsequently leads to increased ethanol productivity. Therefore, in order to facilitate maximum hydrolysis, it is important to have a CBP yeast that is capable of growth and fermentation at higher temperatures. Thermotolerant industrial *S. cerevisiae* strains are thus highly sought after for the CBP of raw starch to ethanol.

The recent developments in synthetic biology, metabolic engineering and protein engineering support the goal of generating ethanol with a single organism and will hopefully help in alleviating some of the main challenges that are experienced during recombinant protein production. The development of a suitable CBP yeast for raw starch hydrolysis requires the consideration of all the bottlenecks in the gene expression process such as transcription, translation, protein folding, secretion and cell viability.

## 2.3 Synthetic biology

A review by Cameron et al. (2014) defined synthetic biology as "the use of molecular biology tools and techniques to forward-engineer cellular behaviour" and the article highlighted the main accomplishments that have become milestones over the last decade, as well as the obstacles and applications. Recent progress in synthetic biology has provided the tools to develop biological engineering as an application of biology, in the same manner that chemical engineering is an application of chemistry (McArthur and Fong 2010). It encompasses the design and construction of biological parts and systems, as well as the re-designing of existing systems for improved functions (Fig. 2.7). Two main approaches associated with synthetic biology are the manufacturing of synthetic molecules/biological parts to mimic naturally occurring molecules and the use of natural molecules to construct systems or circuits that act unnaturally (Steensels et al. 2014).

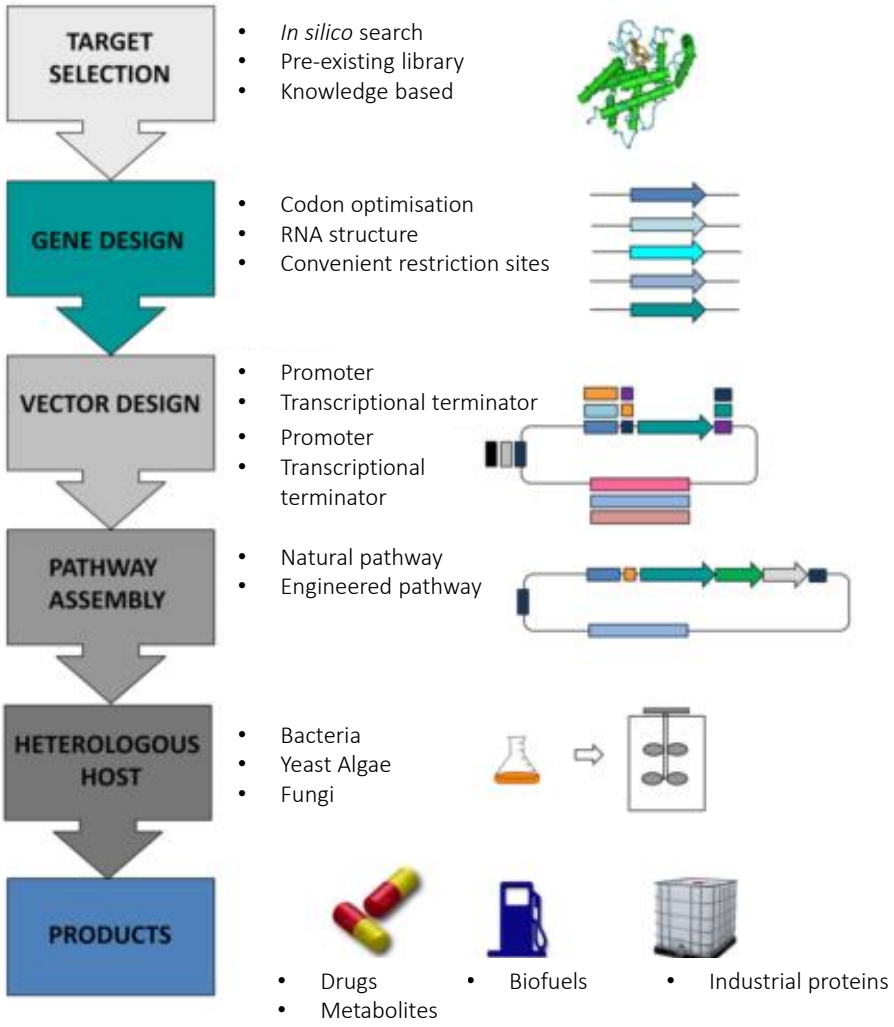


**Fig. 2.7** Schematic representation of an approach used by molecular biologists for the design of novel bio-based parts and devices, as well as the engineering of heterologous expression hosts. Adapted from Elena et al. (2014).

Over the last decade DNA sequence information has increased consistently as a result of genome and metagenome sequencing projects, which has boosted the rational design and *de novo* synthesis of DNA. Scientists are now able to manipulate cellular mechanisms and study the fundamental aspects behind gene regulation, with the individual genetic elements often referred to in literature as 'BioBricks' (Venton 2014). A BioBrick is a DNA sequence that must have certain characteristics such as being able to send and receive standard biochemical signals, as well as being able to be restricted and ligated into a linear sequence of other BioBricks (Tucker and Zilinskas 2006). Several types of computational tools have been developed using *E. coli* and *S. cerevisiae* strains as the two model organisms (Cameron et al. 2014), to facilitate the design of artificial genes. Thus, synthetic biology has become a flexible tool for many different applications. However, what may apply to the expression of one gene does not always hold true for the expression of another gene and information encoded in the genetic code is essential for efficient gene expression.

### 2.3.1 Synthetic DNA

One of the most common synthetic biology tools used in research is the synthesis of DNA. The use of artificial DNA fragments is integrated into a wide range of applications, namely metabolic engineering, gene circuit design and genome synthesis (Ma et al. 2012). Pre-existing template DNA is no longer a prerequisite and complete genes may be synthesised *de novo. De novo* gene synthesis is rapidly becoming the most popular and efficient method for obtaining functional genetic constructs. The ability to generate double-stranded DNA synthetically has increased from less than 100 bp to over 1 Mb in recent years (Ma et al. 2012), thus making gene amplification through conventional PCR methods redundant (Wu et al. 2007). Gene synthesis can be combined with tailored design options such as codon optimisation and the production of RNA interference (RNAi) resistant genes. It has also benefited the field of medical research for vaccine production (by eliminating the need to obtain genes from pathogenic organisms), microbiome engineering and cell therapy (Ruder et al. 2011).

The use of synthetic DNA technology offers a powerful tool that is changing the way scientists approach gene expression. Gene synthesis has become a key tool in recombinant DNA technology and provides a practical way in which numerous nucleotide changes can be made to an original gene sequence. Synthetic genes have an additional advantage over molecular cloning, which is flexibility of the gene design (Wu et al. 2007). Since most of the desired protein sequences are not available as physical DNA (Welch et al. 2009), gene synthesis offers a unique niche in the exploitation of foreign genes for heterologous gene expression. Researchers are able to design and have more control over expression systems because they can have gene sequences optimised for specific expression hosts (Gustafsson et al. 2004). Synthetic genes can be synthesised based on amino acid sequences and allow for protein-coding sequences to be re-designed (using synonymous codons) for a number of different purposes. Synthetic biology offers numerous advantages related to non-native DNA sequences, namely:

- manipulate codon usage bias for expression in a particular host (Lanza et al. 2014);
- change GC content to increase messenger RNA (mRNA) stability (Gustafsson et al. 2004);
- modify transcriptional and translational control regions (Young and Alper 2010);
- remove/add post translation modification sites (e.g. glycosylation sites) (Wang 2015);
- change restriction site patterns (Gustafsson et al. 2012);
- modify ribosome binding sites and mRNA degradation sites (McArthur and Fong 2010) and
- adjust translational rates to allow for proper protein folding (Young and Alper 2010).

Synthetic DNA is also changing the way scientists address protein engineering challenges. The ability to synthesise entire genes, novel genetic pathways and even whole genomes has applications that can be applied to many scientific and industrial fields (Hughes et al. 2011; Villalobos et al. 2006). In order to benefit fully from the advantage of synthetic genes, reliable criteria for designing genes is required. The aim is to reach a point whereby high gene expression levels are guaranteed and genes can be cloned directly into biosynthetic pathways, or biological circuits, without having to endure series of elimination by trial and error (Welch et al. 2009b).

### 2.3.2 Applications in synthetic biology

Although the majority of molecular tools (vectors, genetic controllers and gene expression protocols) have been developed and optimised for bacteria, especially for *E. coli*, the availability of synthetic biological techniques for yeast is increasing. These tools can be utilised in many different research fields, including enhancing the ability of yeast to produce fuels and chemicals from renewable biomass (Tsai et al. 2015). The different synthetic biology approaches for engineering *S. cerevisiae* include: (i) DNA assembly techniques, (ii) genome editing techniques and (iii) the construction of genes, pathways and synthetic genomes. The availably of BioBricks is often a bottleneck in the advancement of synthetic biology, but one of the ultimate challenges is to completely map and design a functional cell that is optimised for a specific industrial purpose (Steensels et al. 2014).

The availability of synthetic DNA, specifically synthetic genes, is allowing for the rapid advancement of molecular science and the development of biologically engineered solutions. This will help address global problems related to environment, energy and health concerns. However, the application of genetically engineered yeast for producing biofuels has three main challenges to overcome, namely substrate range, the development of new synthesis pathways for producing advanced biofuels, as well as decreased product inhibition (Tsai et al. 2015).

### 2.3.3 Synonymous codons and codon bias

The genetic code consists of 64 codons (sequences of 3 nucleotides) that are used to encode for 20 amino acids, as well as the start and stop codons. In theory, the redundancy of the genetic code enables the same protein sequence to be encoded for by alternative nucleotide sequences and the term ‘synonymous codons’ is used to refer to codons that encode for the same amino acid (Lynn et al. 2002). The occurrence of synonymous codons within any genome is not uniform and codon usage frequencies are unequal for most synonymous codons within naturally occurring genomes (Wu et al. 2007). This bias is believed to be a passive reflection of the mutational/evolutionary biases at work in a genome (Sharp et al. 1993; Hockenberry et al. 2014).

Studies investigating synonymous codon usage have been pursued for over 30 years and since then inquiries have shown that bias in codon usage is associated with a variety of factors, including genomic base composition, mutational bias and selection for, or against particular sequence motifs (used as control elements) (Hockenberry et al. 2014). Codon bias is also widespread across diverse taxa (Sharp et al. 2005) and although substantial computer generated data has been analysed, the reasons for the selection of translational optimal codons remains unclear. Possible suggestions include: maximised speed of elongation, minimised costs of proofreading, or maximised accuracy of translation (Stoletzki and Eyre-Walker 2007).

A significant impact on gene expression levels and protein folding can occur when the codon usage is changed (Elena et al. 2014). Subsequently, the expression of many foreign genes has increased by employing different strategies that redesign DNA coding sequences (Hu et al. 2013). However, this approach is not always successful (Nørholm et al. 2012), since codon usage varies between genes from the same organism (Hockenberry et al. 2014). From a translational perspective, optimal codons were considered to be those that were best recognised by the most abundant tRNAs (transfer RNAs) and there is a correlation between the frequency of these codons in a gene and gene expression levels. Thus, according to this theory, gene expression levels are thought to be associated with the strength of codon usage (Henry and Sharp 2007). This implies that natural selection for a particular group of “optimal” codons is not random (Stoletzki and Eyre-Walker 2007; Plotkin and Kudla 2011). However, proving which features of a gene sequences have been selected for by natural selection is extremely difficult (Wu et al. 2007).

Host-specific codon usage bias (CUB) results from the varying distribution of preferred codons across all organisms (Lanza et al. 2014) and this becomes a major factor when expressing foreign genes in a recombinant host. Codon usage is determined by the combined actions of mutation, drift and selection (Hockenberry et al. 2014). Since CUB is associated with translational efficiency, a common approach for codon optimisation strategies is to substitute rare codons with more frequently occurring ones, in so doing matching the CUB of the host organism (Qian et al. 2012; Lanza et al. 2014).

In recent years, however, more in-depth studies have been conducted that suggest that this concept of globally “optimal/abundant” or “suboptimal/rare” codons is misguided. A specific codon may be considered as “optimal” with regards to translational efficiency, while on the other hand it could be considered to be suboptimal with regards to mRNA secondary structure. Saunders and Deane (2010) reported that synonymous codons vary in their tendency for protein secondary structures and some codons are associated with buried/ non-aggregation-prone sites, while others favour exposed/aggregation-prone sites (Lee et al. 2010). These

contrasting situations have led to a notion that codon usage is position dependent and therefore the role of individual codons needs to be analysed in relation to their location within a gene (Hockenberry et al. 2014).

The codons surrounding the 5' terminal of proteins have been a particular focus, with some studies highlighting the negative effect of rare codons particularly at 5' termini (Hu et al. 2013). However, rare codons may influence co-translational protein folding. Clarke and Clark (2010) revealed an enrichment of rare codons at both the 5' and 3' end of ORFs from *E. coli* and other prokaryotes and concluded that clusters of rare codon are intentionally located and enriched at *E. coli* gene termini. This may suggest that rare codons aid in a number of functions, including protein biogenesis, folding, secretion and interactions with chaperone proteins. The "ramp" effect is also a commonly referred to a condition that is associated with gene expression. It is characterised by the increased occurrence of functionally relevant rare codons immediately after the initiator codon (~30 – 50 codons after the AUG) and is also associated with the involvement of chaperone proteins at the end of the ribosome exit tunnel (Ragionieri et al. 2015).

Recent advances in sequencing and synthetic biology have provided more information pertaining to the role and importance of codon bias. Significant patterns have been uncovered, suggesting new hypotheses for protein synthesis that involve codon adaption, ribosome availability and translation rates (Plotkin and Kudla 2011). However, there are too many gaps in the overall knowledge on the role that synonymous codons play. The limitations to the use of CUB in estimating gene expression levels needs to be explored on a fundamental level in order to effectively optimise protein production strategies. It appears that the rate of protein synthesis is (mostly) controlled by codon usage in the mRNA (Jha and Komar 2011). However, since the mechanism of translation differs between eukaryotes and prokaryotes, their codon usage cannot be compared with respect to their role in protein folding (Clarke and Clark 2010).

#### *2.3.3.1 Codon optimisation*

Codon optimisation is defined as the "*in silico* design of an optimal coding sequence for a given protein using a distinct arrangement of alternative codons" (Liss et al. 2012) and is a practical aspect of molecular biology. The redundancy of the genetic code allows for numerous possibilities of DNA sequences that can encode for the same protein. Foreign proteins are often produced at low levels because wild-type foreign genes have not evolved for optimum expression in alternative expression hosts (Chung and Lee 2012). The GC content and codon usage of genes are the two main sequence features recognised to influence gene expression (Barahimipour et al. 2015). Furthermore, studies have shown that a well optimised gene can

improve gene expression by varying amounts. In *P. pastoris,* increases of up to 10-fold improvement in protein production have been reported (Hu et al. 2013). However, the fundamental molecular mechanisms behind these factors still remains unclear.

Many hypotheses have been made relating to the use of rare versus abundant codons leading to numerous algorithms for optimised gene sequence design. Yet, predicting the ideal gene sequence remains unclear, since the host organism is an additional parameter that needs to be taken into consideration. Gene synthesis vendors use different algorithms to optimise DNA sequences. However, there is currently no one method that guarantees improved gene expression. Furthermore, there is a lack of experimental data to support the different types of design principles used to design the optimal gene sequence (Welch et al. 2009b). Subsequently, the resulting codon optimisation and expression levels may differ depending on which company was used for gene synthesis. For example, the gene synthesis company DNA2.0 (California, USA) uses an algorithm that codon optimises a gene sequence so that the codon bias distribution for the target gene reflects the codons that confer expression during artificial induction (Welch et al. 2009b). An increased understanding of synonymous codon usage and protein folding is required in order to efficiently tackle the development of heterologous gene (native versus codon optimised) expression in industrially important hosts.

Many studies have shown that synthetically designed genes can enhance gene expression (Jung and McDonald 2011; Elena et al. 2014; Ragionieri et al. 2015). Gustafsson et al. (2004) compared gene expression levels from natural gene sequences with their codon optimised counterparts in identical systems. The position of the codons, the surrounding codons, as well as their role within the mRNA can all impact on synonymous codon choices. Codon optimisation was used successfully to increase the expression of the keratinase (*kerA*) gene from the *B. licheniformis* in *P. pastoris* and activity for the two optimised gene variants increased from 195 $U.ml^{-1}$ to 324 $U.ml^{-1}$ and 293 $U.ml^{-1}$, respectively (Hu et al. 2013). However, the lack of understanding regarding the factors involved in a rational design strategy (for a protein-coding gene) hampers progress in this field (Wu et al. 2007).

Codon optimisation techniques alter the codon usage pattern, which may result in increased expression levels, however it has also been shown to compromise the final tertiary confirmation of recombinant proteins (Yadava and Ockenhouse 2003). Therefore, choosing a reliable method to best design synthetic genes is not a straightforward process. Furthermore, there is a tendency in the scientific community to only publish effective optimisation experiments. This makes it difficult to draw conclusions regarding the use of different algorithms and strategies for the successful optimisation of gene sequences.

#### *2.3.3.2 Codon adaption index*

One of the popular assumptions surrounding CUB is that rare codons are suboptimal and therefore their usage in some studies has been reduced or avoided in coding sequences (Hockenberry et al. 2014). The codon adaptation index (CAI; Sharp and Li 1987) is a measure of usage of preferred codons (codon bias of a gene towards common codons). This strategy uses a subset of highly expressed codons and avoids rarely used codons (for the host of interest as the reference set), to maximise the success of gene expression (Fox and Erill 2010). The CAI optimisation strategy is an alternative to the algorithms designed to codon optimise genes. CAI calculations require the classification of highly expressed genes from a given host organism (Grote et al. 2005). It is an approach that was used in the early days of gene design and is designated as “one amino acid-one codon” because it uses the most abundant codon of the host to encode all occurrences of a given amino acid in the optimised sequence. The CAI is the prevailing empirical measure of expressivity (Grote et al. 2005), with an average CAI for highly expressed genes being 0.97 and 0.21 for poorly expressed genes (Thanaraj and Argos 1996). Therefore, if highly expressed genes are rich in preferred codons, then these codons should work well for heterologous gene expression.

Many studies have documented the successful expression of genes that were designed by maximising CAI and it is often regarded as the “gold standard” among CBIs (codon bias index) (Fox and Erill 2010). However, there is a lack of understanding with regards to the complete picture and genetic relationships involved. One of the main drawbacks to this method is that a gene with this design approach will have strongly transcribed mRNA and contain a high concentration of a subset of codons. Recombinant proteins may be produced at levels as high as 60% of total cell mass. This can potentially result in an imbalance in the tRNA pool and a subsequent depletion of certain tRNAs, causing reduced growth (Villalobos et al. 2006).

#### *2.3.3.3 Codon usage affects speed of translation*

Bioinformatics studies have been used to indicate possible instances of ribosome pausing during the translation of many membrane proteins (Nørholm et al. 2012). It is believed that rare/low abundance codons in the mRNA are a genetically inherent mechanism that the host cells uses to slow down the rate of protein synthesis at specific sites by creating pauses during mRNA translation (Yadava and Ockenhouse 2003; Rosano and Ceccarelli 2009); this allows ribosomes to bind more efficiently to the mRNA and helps to avoid ribosomal bottlenecks (Ragionieri et al. 2015).

It is possible that variation of translation kinetics by synonymous codon usage may also help nascent chains fold correctly, since a slower translation rate allows for individual folding events to take place (Morgunov and Babu 2014; Pechmann and Frydman 2013). The principle behind

this theory is that additional time is needed for the corresponding rare species of tRNA to be delivered to the ribosome, thereby slowing down translation (Fig. 2.8). It has been proposed that translation pauses are an important part of protein synthesis that allow for proper formation of secondary structures, conversion of α-helices and independent folding of domains (Makhoul and Trifonov 2002; Jha and Komar 2011). If the nascent peptide chain does not develop its native-like structure then protein function is affected (Baker 2000).

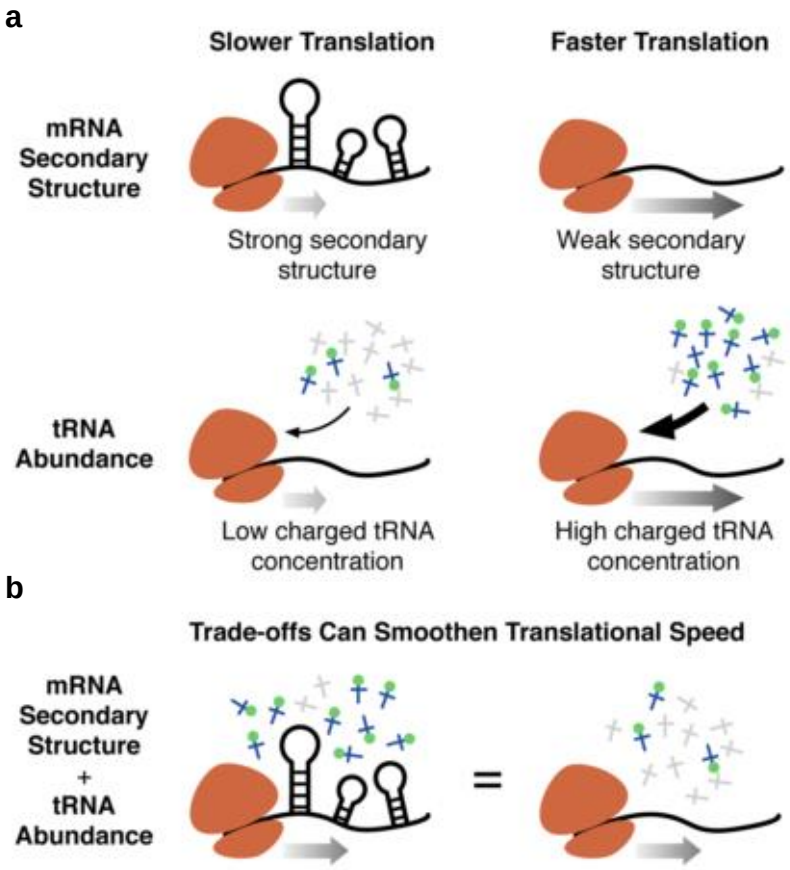


**Fig. 2.8** The trade-off between secondary structure and tRNA-concentration affects translation rate. (**a**) mRNA secondary structure and an abundance of cognate amino acid tRNA affect the elongation speed. (**b**) A trade-off of the negative effect of one factor with the positive effect of the other results in a more fluent elongation rate (Gorochowski et al. 2015).

Gorochowski et al. (2015) demonstrated that there is a relationship between codon choice and mRNA secondary structure. Synonymous variations can alter mRNA secondary structure, which can have a knock on effect influencing the efficiency of translation and rate of mRNA degradation (Hunt et al. 2014) (Fig. 2.8). A reduced translational efficiency may result in altered folding patterns and subsequent function of the translated polypeptides. Thus, protein synthesis is directly linked to co-translational folding and mRNA sequences are evolutionary conserved with a strong association to protein folding (Pechmann and Frydman 2013).

## 2.4 Post-translational modifications and secretion

The production of functional proteins is closely related to the cellular machinery of the host microorganism (Macauley-Patrick et al. 2005). Yeasts are widely used as expression hosts and are able to perform a number of different post-translational modifications with respect to secreted proteins. These are similar to those that are performed by higher eukaryotes, such as correct folding, disulphide bond formation, *N*- and *O*-linked glycosylation, as well as the proteolytic processing of signal sequences (Çelik and Çalık 2012). Various optimisation strategies have addressed rate-limiting factors that occur between mRNA translation and protein secretion. A few of the post-translational modification routes are discussed below.

### 2.4.1 Secretion signals

In the 1970s the discovery of signal sequences was a major breakthrough in cell biology (Hegde and Bernstein 2006) and successful protein secretion is the fundamental step in efficient protein production. Most secretory proteins and many membrane proteins contain a cleavable N-terminal signal sequence. This secretion signal or signal peptide consists of 15 - 30 amino acids and directs the pro-peptide to the endoplasmic reticulum (ER) and through the secretory pathway for both eukaryotic and prokaryotic peptides (Nielsen et al. 1997). This pro-peptide is cleaved during translocation by dibasic endo-peptidases, such as Kex2 (Yang et al. 2013) or furin, in the Golgi apparatus (Daly and Hearn 2005) and the mature peptides are packaged into secretory vesicles and transported to the cell's surface. Alternatively, extracellular signal peptidases cleave secretion signals shortly after translocation (Hiller et al 2004), in order to release the mature peptide.

Proteins produced without their native pro-region can result in a slower exit from the ER and in some cases misfolded products are formed (Daly and Hearn 2005). Therefore, engineering the addition of the Kex2 protease recognition site (Lys–Arg), when designing recombinant heterologous secretion signals, may improve protein secretion (Njokweni et al. 2012). Endo-proteolysis by the Kex2 protease has been identified as a possible rate limiting step in protein secretion, where expression levels are high. Subsequently, the overexpression of the *KEX2* gene has been shown to relieve this rate limiting step (Shuster 1991).

Secretion signals are considered to be relatively heterogeneous, since they can function interchangeably between different species. There is little homology in the sequence of these peptides, but all contain a basic amino acid in the N-terminal region that is followed by a hydrophobic core region (Futatsumori-Sugai and Tsumoto 2010). Computer software can now be used to analyse the hydrophobicity of a protein sequence and allows for automated identification of signal peptides and prediction of the cleavage sites in amino acid sequences from different organisms (Hiller et al. 2004).

A review by Hegde and Bernstein (2006) examined the diversity among secretion signals and gave evidence to conclude that secretion signals encode for more information than was originally thought. These peptides may have a number of post cleavage functions e.g. antigen presentation and act as possible intermediates in other biological functions, which remain to be explored. Furthermore, efficient protein secretion is not solely directed by a signal sequence, but is also influenced by the structural nature of the protein (Cereghino et al. 2002). The increase in available DNA sequence data and the industrial need for efficient recombinant systems has stimulated a strong interest in the role of secretion signals and scientists are still exploring the diversity in secretion signal functionality (Hegde and Bernstein 2006). In some cases the native secretion signal may not be effective for heterologous gene expression and replacement with an alternative signal sequence is recommended. Currently there is no single secretion signal that functions efficiently in various hosts (Tan et al. 2002).

A number of secretion signals have been described for use in expressing recombinant peptides or proteins, e.g. a novel secretion signal (SS) was cloned by Tan et al. (2002) that is capable of directing the secretion of recombinant proteins from both prokaryotes and eukaryotes; their strategy offers a flexible approach to improving gene expression. Furthermore, fungal secretion signals have been repeatedly used for recombinant protein production in yeasts, e.g. secretion signals from *R. oryzae's* α-amylase and *A. awamori's* glucoamylase. In addition, many yeast expression systems use the *S. cerevisiae* α-mating type secretion signal to direct secretion (Madhavan and Sukumaran 2014). On the other hand, *S. cerevisiae* and *P. pastoris* strains are known to have low specificity for the recognition of signal sequences. Therefore, when expressing recombinant proteins the native signal sequence may be effective for gene expression and successful protein secretion. Yet, expression levels vary substantially when changing the secretion signal and there is currently no definitive method to predetermine whether a native signal will result in adequate secretion of biologically active proteins (Daly and Hearn 2005). Ideally, selected secretion signals should be compatible to a variety of proteins and be effective in a variety of hosts.

### 2.4.2 Glycosylation

Post-translational glycosylation is an abundant modification that is species-, tissue- and cell-type-specific (Demain and Vaishnav 2009) and occurs in almost all secreted eukaryotic proteins. It is a fundamental process for protein function and cell physiology and involves the attachment of a sugar moiety to a protein either during or after translation. *N*-linked and *O*-linked glycosylation are both involved in the confirmation and activity of proteins, protection from proteolytic degradation and secretion (Fig. 2.9). *N*-glycan moieties (Fig. 2.9a) also play a significant role in protein folding and the secretion of proteins from the ER and the Golgi apparatus. The inherent structural diversity of glycans allows glycosylation to be effective in

generating diversity (Mitra et al. 2006). *O*-GalNAcylation, is a common *O*-glycosylation that occurs on secreted proteins and there are eight types of core *O*-GalNAc structures (Fig. 2.9b) (Roth et al. 2012).

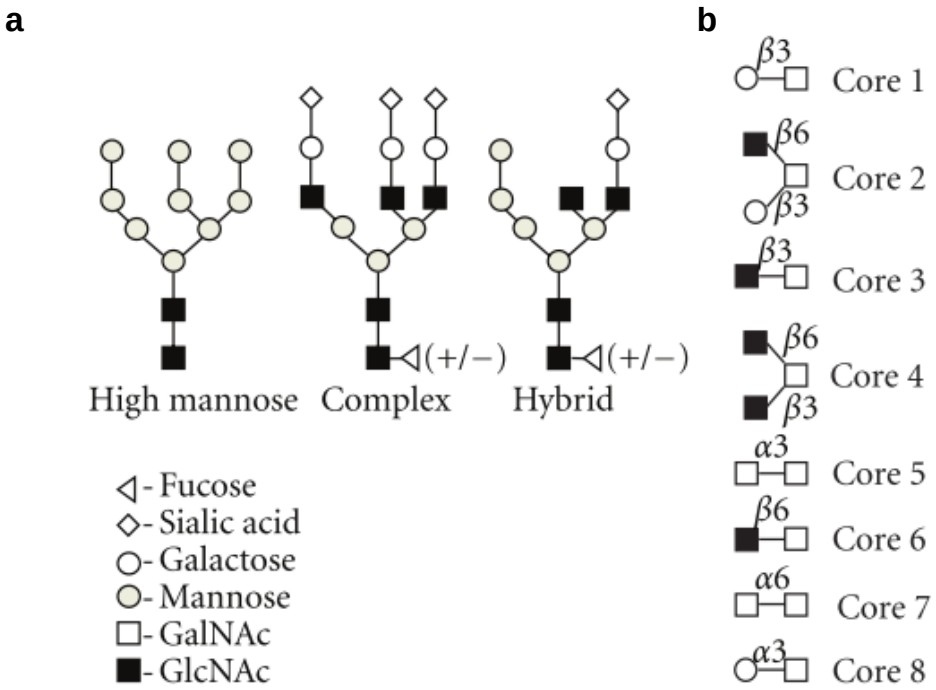


**Fig. 2.9** The basic core structures of (**a**) *N*-glycans and (**b**) *O*-glycans (Roth et al. 2012).

The degree of protein glycosylation depends on the expression host; *S. cerevisiae* is known to have the most extensive glycosylation system, while glycosylation by *Yarrowia lipolytica* involves significantly less residues. Subsequently, glycosylation by *S. cerevisiae* can lead to hyper-glycosylation, which reduces the secretion rate (Çelik and Çalık 2012). In most situations, protein glycosylation begins while the protein is being synthesised, thus suggesting that this type of modification may have an important role in protein folding and oligomerisation. In addition to mediating interactions between protein subunits, glycans can also act like chaperones by assisting in protein folding. Misfolding and aggregation of proteins often occurs when glycosylation is suppressed or inhibited and this leads to non-functional proteins (Mitra et al. 2006).

### 2.4.3 Protein folding

In order for proteins to function properly they need to achieve the correct conformation, which is an intricate and complex process. Protein folding and the subsequent secretion of proteins involves many interacting participants, namely tRNA availability and the amount of protein being expressed. This step is followed by the folding of the newly translated polypeptides into three-dimensional conformations. Protein chains can adopt a number of different

conformations, each relying on the co-operation of non-covalent interactions (Hartl et al. 2011) and the number of conformations increases exponentially with the length of the peptide (Baker 2000). Protein folding generally begins with the formation of secondary structures (α-helices and β-sheets), followed by the rapid generation of disulfide bonds in the ER (Daly and Hearn 2005). Therefore, it is important to understand the mechanism of protein folding and the factors may affect the process (Jha and Komar 2011).

The resulting tertiary structure corresponds to a distinctive biologically-active state, which has been selected for by evolution. However, complications can arise inside the cell which may lead to misfolding or protein aggregation and loss of function occurs (Mokry et al. 2015). The native structure is only acquired once all the interactions are formed between the different domains (Dobson 2004). In eukaryotic expression systems, many of the synthesised proteins are secreted to an extracellular environment. These proteins contain a pro-region, which is essential for proper folding and in some cases oligomerisation (Daly and Hearn 2005).

According to the Anfinsen's principle, the amino acid sequence itself is sufficient to direct native thermodynamically driven protein folding (Hunt et al. 2014). The original studies of Anfinsen investigated the spontaneous refolding of small denatured proteins *in vitro* and this was believed to demonstrate that the native three-dimensional structure of a protein was directed by its amino acid sequence (Anfinsen 1973). However, advances in protein studies have disputed this original principle. Studies investigating synonymous mutations have shown that changes in the codon usage can alter co-translational protein folding and can result in conformational differences (Hunt et al. 2014), thus affecting the tertiary structure.

Pechmann and Frydman (2013) investigated patterns of conserved optimal and non-optimal codons, which associate with the secondary structure of the translated polypeptides. They concluded that there are evolutionarily conserved signatures in the mRNA sequences that are associated with folding patterns of the encoded polypeptides. Following the translation of mRNA into polypeptide chains, the nascent chain moves from the ribosome tunnel into the ER, a process accompanied by interactions with export targeting particles (Jha and Komar 2011). Single domain proteins complete their folding post-translationally (after chain termination and release from the ribosome), while proteins consisting of several domains may fold co-translationally as the domains emerge sequentially from the ribosome (Kim et al. 2013).

Aggregation of proteins during the folding stage results in low protein yields. Thus, a number of different strategies are employed to avoid this aggregation and enhance the refolding of proteins such as small molecule additives. Some molecules (co-solutes) improve structure formation or collapse, while others increase flexibility or solubility of the proteins (Tsumoto et al. 2003). Co-solutes can be divided into two groups, (i) folding enhancers and

(ii) aggregation suppressors, which include polyols, salts (ammonium sulphate and magnesium chloride) and amino acids (glycine and alanine). Studies on creating the "perfect folding environment" are still ongoing, particularly with regards to the production of industrial proteins (Gasser et al. 2008).

Correct *in vivo* protein folding is one of the key factors that needs to be considered when choosing a suitable expression host for heterologous gene expression and it is important to note that not all native recombinant proteins can be effectively secreted in yeast (Damasceno et al. 2012). Increasing the rate of one step in the pathway can result in a bottleneck further down the line in the expression system. Stress situations and stress reactions can severely influence the efficiency of an expression system. The following is a list of potential bottlenecks associated with protein production (Mattanovich et al. 2004; Damasceno et al. 2012): (i) efficient transcription by using strong promoters; (ii) codon usage; (iii) translocation determined by the secretion signal peptide; (iv) processing and folding in the ER and Golgi; (v) secretion out of the cell and (vi) protein turnover by proteolysis.

### 2.4.4 Cultivation temperature

Temperature affects a cell's metabolism, as well as the production of proteins that assist in protein folding (Hsp70 family, ER-membrane proteins, etc.). Studies involving decreased cultivation temperatures (from 30°C to 20 – 25°C) have improved heterologous production of proteins in both *E. coli* and *P. pastoris*. (Li et al. 2001; Sørensen and Mortensen 2005; Li et al. 2007). One theory is that a lower growth temperature results in lower specific growth rates, thus slowing the rate of folding and enabling recombinant proteins to reach their native conformation. This leads to improvements in protein yield and may improve the solubility of a number of difficult proteins (Sørensen and Mortensen 2005).

A lower temperature also helps in decreasing the formation of inclusion bodies (Yamaguchi and Miyazaki 2014) and reduces proteolytic degradation of recombinant proteins (in the culture medium), since the amount of proteases released from the dead cells is reduced (Li et al. 2001; Sørensen and Mortensen 2005; Li et al. 2007). Therefore, a lower growth temperature can increase the stability and potential for correct protein folding. High temperatures can negatively impact protein folding by promoting aggregation reactions and favouring conformational stress. The production of recombinant proteins at lower temperatures presents an interesting model for studying the dynamics of protein folding and misfolding and may lead to improvements in the quality of downstream products (Gasser et al. 2008). However, a disadvantage of using lower cultivation temperatures is that cultivation/fermentation times are increased.

### 2.4.5 Molecular chaperones

Molecular chaperones, a ubiquitous class of folding modulators, are multi domain proteins that are found in the cytosol of prokaryotic and eukaryotic cells (Baneyx and Mujacic 2004). Chaperones have evolved to assist the folding of native proteins (Fig. 2.10) (Saibil 2015) and form a diverse group of unrelated proteins (Liberek et al. 2008). They form a network of pathways that are associated with substrate polypeptides from the moment of initial synthesis on ribosomes to the final stages of folding. They also function to protect subunits from heat shock during the assembly of complexes (heat shock proteins - Hsps), as well as to prevent protein aggregation or mediate targeted unfolding/disassembly (Young et al. 2004; Saibil 2015).

Chaperones are constitutively expressed under normal growth condition and play an important role in the conformational quality control of the proteome, by participating in various interactions with non-native polypeptides (Baneyx and Mujacic 2004). From a mechanical point of view, the action of molecular chaperones depends on the differential exposure of structured hydrophobic domains to the solvent to bind nonpolar segments that would normally be hidden within the core of the peptide. Many proteins require the assistance of chaperones in order to fold, thus these proteins are key components in the flow of generic information: DNA↔RNA↔polypeptide↔folded protein (Benjamin and Mcmillan 1998).

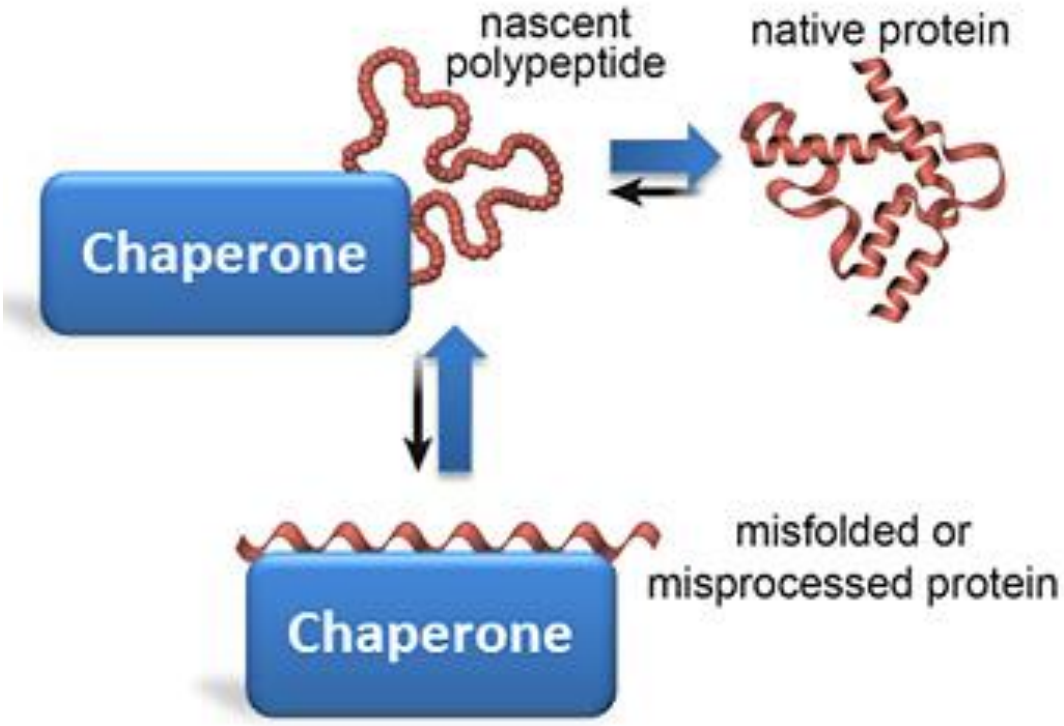


**Fig. 2.10** The role of chaperones in assisting native protein folding (Sanders 2014).

The three general types of chaperones are classified based on the manner in which they interact with the peptides (Tiroli-Cepeda and Ramos 2011). Foldases (e.g., chaperonin, Hsp70/DnaK and Hsp60/GroEL) facilitate the net refolding/unfolding of the peptides in an ATP-dependent manner and assist in helping the polypeptides reach their native confirmation. Holding chaperones (e.g., IbpB) are associated with partially folded proteins, they bind client proteins in an ATP-independent manner, while waiting for folding chaperones to become available. These chaperones not only secure the protein for further action by other types of chaperones, but also protect proteins from misfolding and aggregation (Tiroli-Cepeda and Ramos 2011). Typical examples of holders are small Hsps (sHsps) and Hsp40s. The least understood type of chaperones are those associated with protein disaggregation. An example of a disaggregating chaperone is ClpB, which is involved in the solubilisation of aggregated proteins during stressful conditions (Baneyx and Mujacic 2004).

Chaperones are associated with a wide range of different substrates. Their expression levels increase during stress indicating their importance with regards to a cell's health. In addition to their classification based on function, chaperones are also grouped based on their sequence homology and approximate molecular weight (MW). Stress-related Hsps were originally named according to their molecular weight (Hsp40s, Hsp60s, Hsp70s, Hsp90s, Hsp100s) and are upregulated upon heat shock or other stresses (Kim et al. 2013). They are also the most conserved proteins present in both prokaryotes and eukaryotes and allow the cell to survive potentially lethal conditions (Schmitt et al. 2006). More recently their function as molecular chaperones has been given more attention (Feder and Hofmann 1999; Saibil 2015), specifically the role of Hsps in the pathology of human diseases (Horváth et al. 2008).

Chaperone proteins are also involved in several aspects of proteome maintenance, including assistance in macromolecular complex assembly, transportation and degradation of protein, as well as the refolding of proteins denatured due to stress conditions (Kim et al. 2013). Further investigation into the chaperone systems will provide a greater understanding of the complex role that these proteins play in protein folding and the mechanisms by which protein misfolding and aggregation cause disease. From a biotechnology perspective, increased knowledge on chaperone regulation will help in improving the expression of recombinant genes and protein secretion in various host systems.

Over the last few decades, numerous studies have focused on protein folding during *in vitro* experiments. Results have revealed fascinating insights into the mechanisms by which chaperone systems function (Kim et al. 2013). However, these analyses are from *in vitro* experiments and don't apply directly to *in vivo* protein folding. The influence of the cellular environment on protein folding and stability, as well as the effect of translation on the folding process is still being explored. In order to gain a deeper understanding of the role that

chaperones play, a broad systems biology approach is needed that combines ribosome profiling, quantitative proteomics and computational modelling (Kim et al. 2013).

### *2.4.5.1 Protein disulphide isomerase*

Protein disulphide isomerase (PDI) is a multifunctional soluble protein (55 kDa), which is found in all eukaryotic organisms. It is induced by stress and has been discovered in many other cellular locations, including the ER, the cell surface, cytosol, mitochondria and extracellular matrix (Parakh and Atkin 2015). PDI has general chaperone activity and disulphide interchange activity (Fig. 2.11). As a chaperone, it assists unfolded or incorrectly folded proteins by providing an environment in which the protein can attain its native state (Ferrari and Söling 1999). It is also responsible for the isomerisation, formation and rearrangement of protein disulphide bonds. PDI is an essential protein in *S. cerevisiae* and many yeast engineering studies include the overexpression of *PDI* for improved protein folding (Çelik and Çalık 2012).

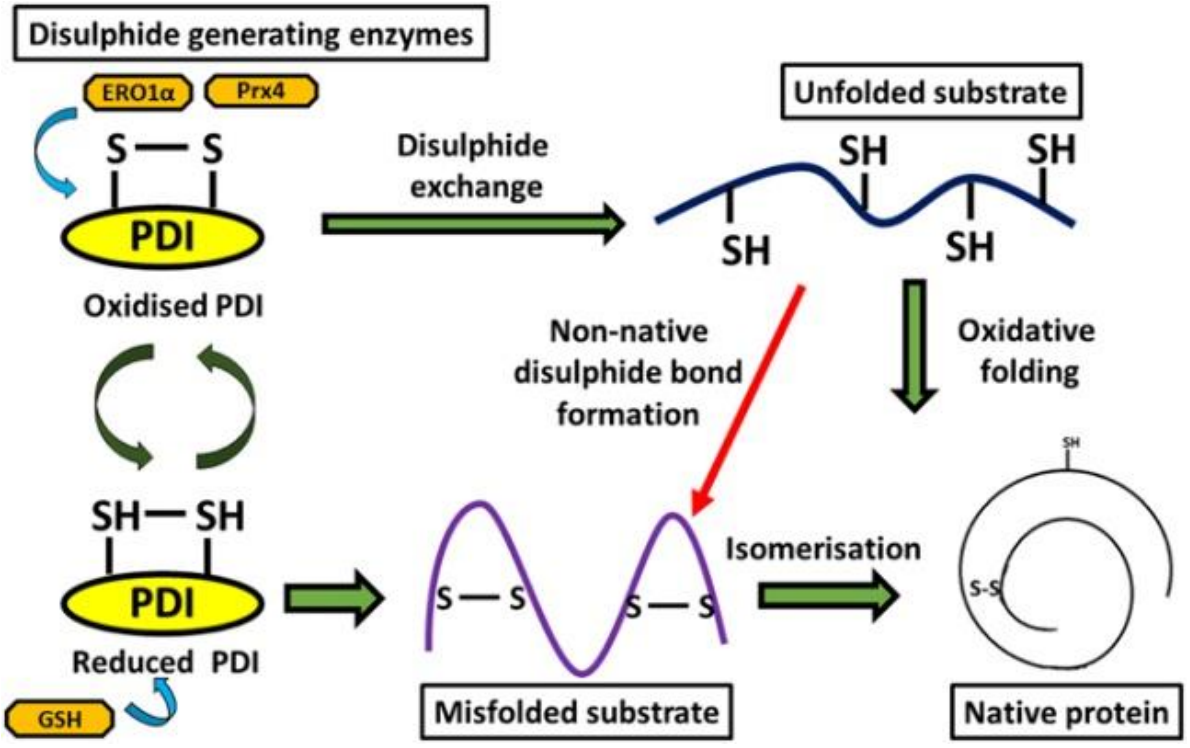


**Fig. 2.11** The role of protein disulphide isomerase (PDI) in assisting protein folding; showing disulphide bond formation and redox reactions. Oxidative folding of PDI assists disulphide bond formation in native protein substrates, while reduced PDI enables isomerisation of non-native bonds in protein substrates (Parakh and Atkin 2015).

#### *2.4.5.2 Polyols*

Polyols, such as glycerol and sugars, act as chemical chaperones to supress protein aggregation. Polyol co-solvents are useful in the treatment of protein misfolding diseases. These compounds are associated with stabilising the natively folded conformation of proteins (in relation to the denatured state) (Mishra et al. 2007), but the mechanisms of action remain unknown. A study by Mishra et al. (2007) suggested that the effect of polyols, in relation to the folding of large proteins, is due to preferential hydration which favours structure formation in the folding intermediates. Polyols are preferentially excluded from the area surrounding the protein's surface and are able to maintain solvophobic interactions. This results in the formation of hydrogen bonds that provide a key role in supporting the native conformation of the protein and thus have a stabilising effect (Amadi et al. 2014).

## 2.5 This study

The economic production of biofuels from starch-based substrates requires an integrated approach for the molecular engineering of yeast. Specific attention directed at predicting optimal pathways, gene assembly and high-throughput screening methods is needed for improved enzyme production. The use of synthetic genes has increased over the last decade (Welch et al. 2009) and this has allowed molecular biologists to screen for novel enzymes with enhanced properties, without first having to acquire the genomic DNA. It has also allowed for the ability to clone, express and compare native and codon optimised genes when constructing a recombinant host.

The most economical way to produce bioethanol from a starchy feedstock is to use a single organism that is able to degrade the raw starch without a heat pretreatment step and then ferment the resulting sugars to ethanol in a one-step process. *S. cerevisiae* is known for its high fermentative capacity, high ethanol yield and its high ethanol tolerance. However, it is incapable of hydrolysing starch or dextrins. Therefore, strains of *S. cerevisiae* have the potential to be used in a one-step CBP environment, provided that they can be engineered to produce amylolytic enzymes (α-amylase and glucoamylase) with superior hydrolytic activity. These amylases are industrially important enzymes that are widely used in the biofuel industry for the hydrolysis of starch. However, there a few RSDEs that have been cloned and expressed in *S. cerevisiae* for the hydrolysis of raw starch to ethanol, which provides a unique opportunity for the novel research prosed in this study.

In **Chapter 3**, a fundamental approach was taken to investigate the expression of two amylase genes that had been redesigned on a nucleotide level, using synonymous codons. The use of optimised genes for the construction of an amylolytic CBP yeast is a relatively new field of

study however, and so there is a lack of experimental data on codon bias and the effect that it has on gene expression and protein production (when using *S. cerevisiae* as host). There have been contradictory findings regarding synonymous codon usage and the importance of mRNA secondary structure with respect to protein folding (Wu et al. 2007). Thus two different approaches for the optimisation of the *A. tubingensis* amylase genes were used in this chapter, i.e. codon optimisation using a gene design algorithm (DNA2.0, California, USA) and the CBI approach. The expression of the native and codon adapted amylase variants was subsequently compared in *S. cerevisiae* Y294 using enzyme assays and protein characterisation. The recombinant *S. cerevisiae* strains expressing the native α-amylase and glucoamylase from *A. tubingensis* (Viktor et al. 2013) were used as benchmark strains.

In addition to this gene optimisation approach, additional raw starch amylases were also screened. All fungal amylases contained the SBD associated with raw starch hydrolysis. **Chapter 4** of this study focused on expressing novel raw starch amylase combinations with high catalytic activity in *S. cerevisiae* for the efficient conversion of raw corn starch to glucose. During the screening process the best amylase combination was selected for its superior raw starch hydrolytic ability.

Lastly, in **Chapter 5**, industrial *S. cerevisiae* strains expressing the best combination of α-amylase and glucoamylase genes were constructed using a novel gene integration method. These strains were subsequently evaluated in fermentations using high loadings of raw corn starch (200 $g.l^{-1}$). Different fermentation temperatures were also investigated in order to determine the amylolytic strains' thermotolerance. Industrial amylolytic *S. cerevisiae* strains that can effectively convert raw corn starch to ethanol in one step represent a potential "drop-in" CBP solution to curb enzyme costs in industry.

## 2.6 References

Abbas KA, Khalil SK, Meor Hussin AS (2010) Modified starches and their usages in selected food products: A review study. J Agric Sci 2:90–100. doi: 10.5539/jas.v2n2p90

Abdel-Banat BMA, Hoshida H, Ano A, Nonklang S, Akada R (2010) High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast? Appl Microbiol Biotechnol 85:861-867. doi: 10.1007/s00253-009-2248-5

Abdel-Fattah YR, Soliman NA, El-Toukhy NM, El-Gendi H, Ahmed RS (2013) Production, purification, and characterization of thermostable α-amylase produced by *Bacillus licheniformis*, isolate AI20. J Chmistry 2013:673173. doi: 10.1155/2013/673173

Adam AC, Latorre-García L, Polaina J (2004) Structural analysis of glucoamylase encoded by the *STA1* gene of *Saccharomyces cerevisiae* (var. *diastaticus*). Yeast 21:379-388. doi: 10.1002/yea.1102

Agger J, Johansen KS, Meyer AS (2011) pH catalyzed pretreatment of corn bran for enhanced enzymatic arabinoxylan degradation. N Biotechnol 28:125-135. doi: 10.1016/j.nbt.2010.09.012

Ai Y, Jane J (2015) Gelatinization and rheological properties of starch. Starch - Stärke. 67:213-224. doi: 10.1002/star.201400201

Aiyer PV (2005) Amylases and their applications. African J Biotechnol 4:1525-1529. ISSN: 1684-5315

Alcázar-Alay SC, Meireles MAA (2015) Physicochemical properties, modifications and applications of starches from different botanical sources. Food Sci Technol 35:215-236. doi: 10.1590/1678-457X.6749

Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314:1565-1568. doi: 10.1126/science.1131969

Amadi OC, Okolo BN, Mateo C, Guisan JM, Pessela BC (2014) Enhancing thermal stability of immobilized raw starch digesting amylase (RSDA) using different additives. J Biocatal Biotransformation 2:2 doi: 10.4172/2324-9099.1000109

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223-230. doi: 10.1126/science.181.4096.223

Ao Z, Jane J (2007) Characterization and modelling of the A- and B-granule starches of wheat, triticale, and barley. Carbohydr Polym 67:46-55. doi: 10.1016/j.carbpol.2006.04.013

Bai FW, Anderson WA, Moo-Young M (2008) Ethanol fermentation technologies from sugar and starch feedstocks. Biotechnol Adv 26:89-105. doi: 10.1016/j.biotechadv.2007.09.002

Baker D (2000) A surprising simplicity to protein folding. Nature 405:39-42. doi: 10.1038/35011000

Banat IM, Nigam P, Singh D, Marchant R, McHale AP (1998) Ethanol production at elevated temperatures and alcohol concentrations: Part I – Yeasts in general. World J Microbiol Biotechnol 14:809-821. doi: 10.1023/A:1008802704374

Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in *Escherichia coli*. Nat Biotechnol 22:1399-1408. doi: 10.1038/nbt1029

Barahimipour R, Strenkert D, Neupert J, Schroda M, Merchant SS, Bock R (2015) Dissecting the contributions of GC content and codon usage to gene expression in the model alga *Chlamydomonas reinhardtii*. Plant J 84:704-717. doi: 10.1111/tpj.13033

Barchiesi J, Hedin N, Gomez-Casati DF, Ballicora MA, Busi MV (2015) Functional demonstrations of starch binding domains present in *Ostreococcus tauri* starch synthases isoforms. BMC Res Notes 8:613. doi: 10.1186/s13104-015-1598-6

Basso LC, de Amorim HV, de Oliveira AJ, Lopes ML (2008) Yeast selection for fuel ethanol production in Brazil. FEMS Yeast Res 8:1155-1163. doi: 10.1111/j.1567-1364.2008.00428.x

Benjamin IJ, Mcmillan DR (1998) Stress (heat shock) proteins: molecular chaperones in cardiovascular biology and disease. Circ Res 83:117-132. doi: 10.1161/01.RES.83.2.117

Białas W, Czerniak A, Szymanowska-Powałowska D (2014) Kinetic modeling of simultaneous saccharification and fermentation of corn starch for ethanol production. Acta Biochim Pol 61:153-162. PMID: 24649484

Botte M, Deniaud A, Schaffitzel C (2016) Chapter 6: Cell-free synthesis of macromolecular complexes. In. Vega MC. ed. Advanced Technologies for Protein Complex Production and Characterization. Switzerland: Springer, pp. 79-95

Buléon A, Colonna P, Planchot V, Ball S (1998) Starch granules: structure and biosynthesis. Int J Biol Macromolec 23:85–112. doi.org/10.1016/S0141-8130(98)00040-3

Cameron DE, Bashor CJ, Collins JJ (2014) A brief history of synthetic biology. Nat Rev Microbiol 12:381-390. doi: 10.1038/nrmicro3239

Carrasco M, Villarreal P, Barahona S, Alcaíno J, Cifuentes V, Baeza M (2016) Screening and characterization of amylase and cellulase activities in psychrotolerant yeasts. BMC Microbiol 16:21. doi: 10.1186/s12866-016-0640-8

Çelik E, Çalık P (2012) Production of recombinant proteins by yeast cells. Biotechnol Adv 30:1108-1118. doi: 10.1016/j.biotechadv.2011.09.011

Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM (2002) Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr Opin Biotechnol 13:329-332. doi: 10.1016/S0958166902003300

Chen W, Xie T, Shao Y, Chen F (2012) Phylogenomic relationships between amylolytic enzymes from 85 strains of fungi. PLoS One. 7:e49679. doi:10.1371/journal.pone.0049679.

Chung BK, Lee DY (2012) Computational codon optimization of synthetic gene for protein expression. BMC Syst Biol 6:134. doi: 10.1186/1752-0509-6-134

Cinelli BA, Castilho LR, Freire DMG, Castro AM (2015) A brief review on the emerging technology of ethanol production by cold hydrolysis of raw starch. Fuel 150:721-729. doi: 10.1016/j.fuel.2015.02.063

Clarke TF 4th, Clark PL (2010) Increased incidence of rare codon clusters at 5' and 3' gene termini: implications for function. BMC Genomics 11:118. doi: 10.1186/1471-2164-11-118

Cockburn D, Wilkens C, Ruzanski C, Andersen S, Willum Nielsen J, Smith AM, Field RA, Willemoes M, Abou Hachem M, Svensson B (2014) Analysis of surface binding sites (SBSs) in carbohydrate active enzymes with focus on glycoside hydrolase families 13 and 77 — a mini-review. Biologia (Bratisl) 69:705-712. doi: 10.2478/s11756-014-0373-9

Dalmia BK, Nikolov ZL (1991) Characterization of glucoamylase adsorption to raw starch. Enzyme Microb Technol 13:982-990. doi: 10.1016/0141-0229(91)90121-P

Daly R, Hearn MTW (2005) Expression of heterologous proteins in *Pichia pastoris:* a useful experimental tool in protein engineering and production. J Mol Recognit 18:119-138. doi: 10.1002/jmr.687

Damasceno LM, Huang C, Batt CA (2012) Protein secretion in *Pichia pastoris* and advances in protein production. Appl Microbiol Biotechnol 93:31-39. doi:10.1007/s00253-011-3654-z

Das K, Doley R, Mukherjee AK (2004) Purification and biochemical characterization of a thermostable, alkaliphilic, extracellular α-amylase from *Bacillus subtilis* DM-03, a strain isolated from the traditional fermented food of India. Biotechnol Appl Biochem 40:291-298. doi: 10.1042/BA20040034

Das S, Singh S, Sharma V, Soni ML (2011) Biotechnological applications of industrially important amylase enzyme. Int J Pharma Bio Sci 2:486-496. ISSN: 0975-6299

de Souza PM, de Oliveira e Magalhães P (2010) Application of microbial α-amylase in industry - A review. Brazillian J Microbiol 41:850-861. doi: 10.1590/S1517-83822010000400004

Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. Biotech Adv 27:297–306. doi: 10.1016/j.biotechadv.2009.01.008

Demeke MM, Dumortier F, Li Y, Broeckx T, Foulquié-Moreno MR, Thevelein JM (2013) Combining inhibitor tolerance and D-xylose fermentation in industrial *Saccharomyces cerevisiae* for efficient lignocellulose-based bioethanol production. Biotechnol Biofuels 6:120. doi: 10.1186/1754-6834-6-120

Dobson CM (2004) Principles of protein folding, misfolding and aggregation. Semin Cell Dev Biol 15:3-16. doi: 10.1016/j.semcdb.2003.12.008

Eksteen JM, van Rensburg P, Cordero Otero RR, Pretorius IS (2003) Starch fermentation by recombinant *Saccharomyces cerevisiae* strains expressing the α-amylase and glucoamylase genes from *Lipomyces kononenkoae* and *Saccharomycopsis fibuligera*. Biotechnol Bioeng 84:639–646. doi: 10.1002/bit.10797

Elena C, Ravasi P, Castelli ME, Peirú S, Menzella HG (2014) Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. Front Microbiol 5:21. doi: 10.3389/fmicb.2014.00021

Favaro L, Basaglia M, Casella S (2012) Processing wheat bran into ethanol using mild treatments and highly fermentative yeasts. Biomass Bioenergy 46:605-617. doi: 10.1016/j.biombioe.2012.07.001

Favaro L, Jooste T, Basaglia M, Rose SH, Saayman M, Görgens JF, Casella S, van Zyl WH (2013) Designing industrial yeasts for consolidated bioprocessing of starchy biomass to ethanol. Bioengineered 4:97–102. doi: 10.4161/bioe.22268

Favaro L, Viktor M, Rose S, Viljoen-Bloom M, van Zyl W, Basaglia M, Cagnin L, Casella S (2015) Consolidated bioprocessing of starchy substrates into ethanol by industrial *Saccharomyces cerevisiae* strains secreting fungal amylases. Biotechnol Bioeng 112:1751-1760. doi: 10.1002/bit.25591

Feder ME, Hofmann GE (1999) Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. Annu Rev Physiol 61:243-282. doi: 10.1146/annurev.physiol.61.1.243

Ferrari DM, Söling HD (1999) The protein disulphide-isomerase family: unravelling a string of folds. Biochem J 339:1-10. doi: 10.1042/0264-6021:3390001

Fox JM, Erill I (2010) Relative codon adaptation: a generic codon bias index for prediction of gene expression. DNA Res 17:185-196. doi: 10.1093/dnares/dsq012

Fredriksson H, Silverio J, Andersson R, Eliasson A-C, Åman P (1998) The influence of amylose and amylopectin characteristics on gelatinization and retrogradation properties of different starches. Carbohydr Polym 35:119-134. doi: 10.1016/S0144-8617(97)00247-6

Fujii M, Homma T, Taniguchi M (1988) Synergism of alpha-amylase and glucoamylase on the hydrolysis of native starch granules. Biotechnol Bioeng 32:910-915. doi: 10.1002/bit.260320710

Futatsumori-Sugai M, Tsumoto K (2010) Signal peptide design for improving recombinant protein secretion in the baculovirus expression vector system. Biochem Biophys Res Commun 391:931-935. doi: 10.1016/j.bbrc.2009.11.167.

Gallant DJ, Bouchet B, Baldwin PM (1997) Microscopy of starch: evidence of a new level of granule organization Carbohydr Polym 32:177-191. doi: 10.1016/S0144-8617(97)00008-8

Gamlath J, Aldred GP, Panozzo JF (2008) Barley (1-3; 1-4)-β-glucan and arabinoxylan content are related to kernel hardness and water uptake. J Cereal Sci 47:365-371. doi: 10.1016/j.jcs.2007.05.017

Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodríguez-Carmona E, Baumann K, Giuliani M, Parrilli E, Branduardi P, Lang C, Porro D, Ferrer P, Tutino ML, Mattanovich D, Villaverde A (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact 7:11. doi: 10.1186/1475-2859-7-11

Gírio FM, Fonseca C, Carvalheiro F, Duarte LC, Marques S, Bogel-Łukasik R (2010) Hemicelluloses for fuel ethanol: A review. Bioresour Technol 101:4775-4800. doi: 10.1016/j.biortech.2010.01.088

Global Industry Analysts, Inc. (GIA) (2012) Available at: http://www.prweb.com/releases/native_starch_dry_starch/modified_liquid_starch/prweb9945359.htm [Accessed 1 October 2016]

Gohel V, Duan G (2012) No-cook process for ethanol production using Indian broken rice and pearl millet. Int J Microbiol 2012:680232. doi: 10.1155/2012/680232

Goldstein A, Annor G, Putaux JL, Hebelstrup KH, Blennow A, Bertoft E (2016) Impact of full range of amylose contents on the architecture of starch granules. Int J Biol Macromol 89:305-318. doi:10.1016/j.ijbiomac.2016.04.053

Görgens JF, Bressler DC, van Rensburg E (2015) Engineering *Saccharomyces cerevisiae* for direct conversion of raw, uncooked or granular starch to ethanol. Crit Rev Biotechnol 35:396-391. doi: 10.3109/07388551.2014.888048

Gorochowski TE, Ignatova Z, Bovenberg RAL, Roubos JA (2015) Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. Nucleic Acids Res 43:1-11. doi: 10.1093/nar/gkv199

Gottschalk LMF, Oliveira RA, da Silva EP (2010) Cellulases, xylanases, β-glucosidase and ferulic acid esterase produced by *Trichoderma* and *Aspergillus* act synergistically in the hydrolysis of sugarcane bagasse Biochem Eng J 51:72–78. doi: 10.1016/j.bej.2010.05.003

Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 33:526-531. doi: 10.1093/nar/gki376

Gupta R, Gigras P, Mohapatra H, Goswami VK, Chauhan B (2003) Microbial α-amylases: a biotechnological perspective. Process Biochem 38:1599-1616. doi: 10.1016/S0032-9592(03)00053-0

Gupta RM, Musunuru K (2014) Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. J Clin Invest 124:4154–4161. doi: 10.1172/JCI72992

Gustafsson C, Govindarajan S, Minshull J, Park M (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22:346-353. doi: 10.1016/j.tibtech.2004.04.006.

Gustafsson C, Minshull J, Govindarajan S, Ness J, Villalobos A, Welch M (2012) Engineering genes for predictable protein expression. Protein Expr Purif 83:37-46. doi: 10.1016/j.pep.2012.02.013

Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. Nature 475:324-332. doi: 10.1038/nature10317

Hashimoto S, Ogura M, Aritomi K, Hoshida H, Nishizawa Y, Akada R (2005) Isolation of auxotrophic mutants of diploid industrial yeast strains after UV mutagenesis. Appl Environ Microbiol 71:312-319. doi: 10.1128/AEM.71.1.312

Hegde RS, Bernstein HD (2006) The surprising complexity of signal sequences. Trends Biochem Sci 31:563-571. doi: 10.1016/j.tibs.2006.08.004

Henry I, Sharp PM (2007) Predicting gene expression level from codon usage bias. Mol Biol Evol 24:10-12. doi: 10.1093/molbev/msl148

Hiller K, Grote A, Scheer M, Münch R, Jahn D (2004) PrediSi: prediction of signal peptides and their cleavage positions. Nucleic Acids Res 32:375-379. doi: 10.1093/nar/gkh378

Hockenberry AJ, Sirer MI, Amaral LAN, Jewett MC (2014) Quantifying position-dependent codon usage bias. Mol Biol Evol 31:1880-1893. doi: 10.1093/molbev/msu126

Horváth I, Multhoff G, Sonnleitner A, Vígh L (2008) Membrane-associated stress proteins: More than simply chaperones. Biochim Biophys Acta 1778:1653-1664. doi: 10.1016/j.bbamem.2008.02.012

Hostinová E, Gašperík J (2010) Yeast glucoamylases: molecular-genetic and structural characterization. Biologia (Bratisl) 65:559-568. doi: 10.2478/s11756-010-0077-8

Hu H, Gao J, He J, Yu B, Zheng P, Huang Z, Mao X, Yu J, Han G, Chen D (2013) Codon optimization significantly improves the expression level of a keratinase gene in *Pichia pastoris*. PLoS One 8:e58393. doi: 10.1371/journal.pone.0058393

Huang H, Qureshi N, Chen MH, Liu W, Singh V (2015) Ethanol production from food waste at high solids content with vacuum recovery technology. J Agric Food Chem 63:2760–2766. doi: 10.1021/jf5054029

Hughes RA, Miklos AE, Ellington AD (2011) Gene synthesis: methods and applications. Methods Enzymol 498:277–309 doi: 10.1016/B978-0-12-385120-8.00012-7

Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C (2014) Exposing synonymous mutations. Trends Genet 30:308-321. doi: 10.1016/j.tig.2014.04.006

IATP (2009) Institute for Agriculture and Trade Policy, 2009 annual report. Available at: http://www.iatp.org/files/2009AnnualReport_2-2-11.pdf [Accessed 1 October 2015]

James JA, Lee BH (1997) Glucoamylases: Microbial sources, industrial applications and molecular biology - A review. J Food Biochem 21:1-52. doi: 10.1111/j.1745-4514.1997.tb00223.x

Jane J, Chen YY, Lee LF, McPherson AE, Wong KS, Radosavljevic M, Kasemsuwan T (1999) Effects of amylopectin branch chain-length and amylose content on the gelatinization and pasting properties of starch. Cereal Chem 76:629-637. doi: 10.1094/CCHEM.1999.76.5.629

Janeček Š, Svensson B, MacGregor EA (2014) α-Amylase: an enzyme specificity found in various families of glycoside hydrolases. Cell Mol Life Sci 71:1149-1170. doi: 10.1007/s00018-013-1388-z

Jensen NB, Strucko T, Kildegaard KR, David F, Maury J, Mortensen UH, Forster J, Nielsen J, Borodina I (2014) EasyClone: method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*. FEMS Yeast Res 14:238-248. doi: 10.1111/1567-1364.12118

Jha S, Komar AA (2011) Birth, life and death of nascent polypeptide chains. Biotechnol J 6:623-640. doi: 10.1002/biot.201000327

Juge N, Le Gal-Coëffet M, Furniss CSM, Gunning AP, Kramhøft B, Morris VJ, Svensson B (2002) The starch binding domain of glucoamylase from *Aspergillus niger* : overview of its structure, function, and role in raw-starch hydrolysis. Biol Bratislava 11:239-245. ISSN: 0006-3088

Jung SK, McDonald K (2011) Visual gene developer: a fully programmable bioinformatics software for synthetic gene optimization. BMC Bioinformatics 12:340. doi: 10.1186/1471-2105-12-340

Kavšček M, Stražar M, Curk T, Natter K, Petrovič U (2015) Yeast as a cell factory: current state and perspectives. Microb Cell Fact 14:94. doi: 10.1186/s12934-015-0281-x

Khatibi P, Roach DR, Donovan DM, Hughes SR, Bischoff, KM (2014) *Saccharomyces cerevisiae* expressing bacteriophage endolysins reduce *Lactobacillus* contamination during fermentation. Biotechnol Biofuels 7:104. doi: 10.1186/1754-6834-7-104

Kim JC, Kong BW, Kim MJ, Lee SH (2008) Amylolytic hydrolysis of native starch granules affected by granule surface area. J Food Sci 73:621-624. doi: 10.1111/j.1750-3841.2008.00944.x

Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Ulrich Hartl F (2013) Molecular chaperone functions in protein folding and proteostasis. Annu Rev Biochem 82:323-355. doi: 10.1146/annurev-biochem-060208-092442

Kumar P, Satyanarayana T (2009) Microbial glucoamylases: characteristics and applications. Crit Rev Biotechnol 29: 225–255. doi: 10.1080/07388550903136076

Korus J, Juszczak L, Witczak M, Achremowicz B (2004) Influence of selected hydrocolloids on triticale starch rheological properties. Int J Food Sci Technol 39:641-652. doi: 10.1111/j.1365-2621.2004.00823.x

Lanza AM, Curran KA, Rey LG, Alper HS (2014) A condition-specific codon optimization approach for improved heterologous gene expression in *Saccharomyces cerevisiae*. BMC Syst Biol 8:33. doi: 10.1186/1752-0509-8-33

Latorre-García L, Adam AC, Manzanares P, Polaina J (2005) Improving the amylolytic activity of *Saccharomyces cerevisiae* glucoamylase by the addition of a starch binding domain. J Biotechnol 118:167-176. doi: 10.1016/j.jbiotec.2005.03.019

Le Borgne S (2012) Genetic engineering of industrial strains of *Saccharomyces cerevisiae*. Methods Mol Biol 824:451–465. doi: 10.1007/978-1-61779-433-9_24

Lee Y, Zhou T, Tartaglia GG, Vendruscolo M, Wilke CO (2010) Translationally optimal codons associate with aggregation-prone sites in proteins. Proteomics 10:4163-4171. doi:10.1002/pmic.201000229

Lee Y, Park D, Choi Y (2015) Characterization of maltotriose production by hydrolyzing of soluble starch with α-amylase from *Microbulbifer thermotolerans* DAU221. Appl Microbiol Biotechnol 99:3901-3911. doi: 10.1007/s00253-014-6186-5

Lehmann U, Robin F (2007) Slowly digestible starch - its structure and health implications: a review. Trends Food Sci Technol 18:346-355. doi: 10.1016/j.tifs.2007.02.009

Li P, Anumanthan A, Gao X, Ilangovan K, Suzara VV, Düzgüneş N, Renugopalakrishnan V (2007) Expression of recombinant proteins in *Pichia pastoris*. Appl Biochem Biotechnol 142:105-124. doi: 10.1007/s12010-007-0003-x

Li Z, Xiong F, Lin Q, D'Anjou M, Daugulis AJ, Yang DS, Hew CL (2001) Low-temperature increases the yield of biologically active herring antifreeze protein in *Pichia pastoris*. Protein Expr Purif 21:438-445. doi: 10.1006/prep.2001.1395

Liakopoulou-Kyriakides M, Karakatsanis A, Stamatoudis M, Psomas S (2001) Synergistic hydrolysis of crude corn starch by α-amylases and glucoamylases of various origins. Cereal Chem 78:603–607. doi: 10.1094/CCHEM.2001.78.5.603

Liberek K, Lewandowska A, Zietkiewicz S, Ziętkiewicz S (2008) Chaperones in control of protein disaggregation. EMBO J 27:328-335. doi: 10.1038/sj.emboj.7601970

Lin SC, Liu WT, Liu SH, Chou WI, Hsiung BK, Lin IP, Sheu CC, Dah-Tsyr Chang M (2007) Role of the linker region in the expression of *Rhizopus oryzae* glucoamylase. BMC Biochem 8:9. doi: 10.1186/1471-2091-8-9

Lin H, Xian L, Zhang Q, Luo X, Xu, Q, Yang Q, Duan C, Liu J, Tang J, Feng J (2011) Production of raw cassava starch-degrading enzyme by *Penicillium* and its use in conversion of raw cassava flour to ethanol. J Ind Microbiol Biotechnol 38:733-742. doi: 10.1007/s10295-010-0910-7

Lin Y, Zhang W, Li C, Sakakibara K, Tanaka S, Kong H (2012) Factors affecting ethanol fermentation using *Saccharomyces cerevisiae* BY4742. Biomass Bioenergy 47:395-401. doi: 10.1016/j.biombioe.2012.09.019

Liss M, Daubert D, Brunner K, Kliche K, Hammes U, Leiherer A, Wagner R (2012) Embedding permanent watermarks in synthetic genes. PLoS One 7:e42465. doi: 10.1371/journal.pone.0042465

Liu Q (2005) Chapter 7: Understanding starches and their role in foods. In Cui SW ed. Food Carbohydrates: Chemistry, Physical Properties, and Applications. CRC Press Taylor and Francis Group, pp. 310-355

Liu M, Bischoff KM, Gill JJ, Mire-Criscione MD, Berry JD, Young R, Summer EJ (2015) Bacteriophage application restores ethanol fermentation characteristics disrupted by *Lactobacillus fermentum*. Biotechnol Biofuels 8:132. doi: 10.1186/s13068-015-0325-9

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506-577. doi: 10.1128/MMBR.66.3.506

Lynn DJ, Singer GA, Hickey DA (2002) Synonymous codon usage is subject to selection in thermophilic bacteria. Nucleic Acids Res 30:4272-4277. doi: 10.1093/nar/gkf546

Gómez LM, Posada SL, Olivera M (2016) Starch in ruminant diets: a review. Rev Colomb Ciencias Pecu 29:77-90. doi: 10.17533/udea.rccp.v29n2a01

Ma S, Tang N, Tian J (2012) DNA synthesis, assembly and applications in synthetic biology. Curr Opin Chem Biol 16:260-267. doi: 10.1016/j.cbpa.2012.05.001

Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22:249-270. doi: 10.1002/yea.1208

Madhavan A, Sukumaran RK (2014) Promoter and signal sequence from filamentous fungus can drive recombinant protein production in the yeast *Kluyveromyces lactis*. Bioresour Technol 165:302-308. doi: 10.1016/j.biortech.2014.03.002

Makhoul CH, Trifonov EN (2002) Distribution of rare triplets along mRNA and their relation to protein folding. J Biomol Struct Dyn 20:413-420. doi: 10.1080/07391102.2002.10506859.

Mattanovich D, Gasser B, Hohenblum H, Sauer M (2004) Stress in recombinant protein producing yeasts. J Biotechnol 113:121-135. doi: 10.1016/j.jbiotec.2004.04.035

McArthur GH, Fong SS (2010) Toward engineering synthetic microbial metabolism. J Biomed Biotechnol 2010:459760. doi: 10.1155/2010/459760

Mehta D, Satyanarayana T (2013) Biochemical and molecular characterization of recombinant acidic and thermostable raw-starch hydrolysing a-amylase from an extreme thermophile *Geobacillus thermoleovorans*. J Mol Catal B Enzym 85-86:229-238. doi: 10.1016/j.molcatb.2012.08.017

Mishra R, Bhat R, Seckler R (2007) Chemical chaperone-mediated protein folding: stabilization of P22 tailspike folding intermediates by glycerol. Biol Chem 388:797-804. doi: 10.1515/BC.2007.096

Mitra N, Sinha S, Ramya TNC, Surolia A (2006) N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. Trends Biochem Sci 31:156-163. doi: 10.1016/j.tibs.2006.01.003

Mitsuiki S, Mukae K, Sakai M, Goto M, Hayashida S, Furukawa K (2005) Comparative characterization of raw starch hydrolyzing α-amylases from various *Bacillus* strains. Enzyme Microb Technol 37:410-416. doi: 10.1016/j.enzmictec.2005.02.022

Mokry DZ, Abrahão J, Ramos CH (2015) Disaggregases, molecular chaperones that resolubilize protein aggregates. An Acad Bra Cienc 87:1273-1292. doi: 10.1590/0001-3765201520140671

Morgunov AS, Babu MM (2014) Optimizing membrane-protein biogenesis through nonoptimal-codon usage. Nat Struct Mol Biol 21:1023-1025. doi: 10.1038/nsmb.2926

Muralikrishna G, Nirmala M (2005) Cereal α-amylases—an overview. Carbohydr Polym 60:163-173. doi: 10.1016/j.carbpol.2004.12.002

Muthaiyan A, Limayem A, Ricke SC (2011) Antimicrobial strategies for limiting bacterial contaminants in fuel bioethanol fermentations. Prog Energy Combust Sci 37:351–370. doi: 10.1016/j.pecs.2010.06.005

Naidu MA, Saranraj P (2013) Bacterial amylase: A review. Int J Pharmacceutical Biol Arch 4:274-287. ISSN 0976-3333

Narendranath NV, Power R (2005) Relationship between pH and medium dissolved solids in terms of growth and metabolism of lactobacilli and *Saccharomyces cerevisiae* during ethanol production. Appl Environ Microbiol 71:2239-2243. doi: 10.1128/AEM.71.5.2239-2243.2005

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1-6. PMID: 10065837

Njokweni A, Rose SH, van Zyl WH (2012) Fungal β-glucosidase expression in *Saccharomyces cerevisiae*. J Mol Microbiol Biotechnol 39:1445–1452. doi: 10.1007/s10295-012-1150-9

Noda T, Takigawa S, Matsuura-Endo C, Suzuki T, Hashimoto N, Kottearachchi NS, Yamauchi H, Zaidul IS (2008) Factors affecting the digestibility of raw and gelatinized potato starches. Food Chem 110:465-470. doi: 10.1016/j.foodchem.2008.02.027

Nørholm MH, Light S, Virkki MT, Elofsson A, von Heijne G, Daley DO (2012) Manipulating the genetic code for membrane protein production: What have we learnt so far? Biochim Biophys Acta 1818:1091-1096. doi: 10.1016/j.bbamem.2011.08.018

O'Neill EC, Field RA (2015) Underpinning starch biology with in vitro studies on carbohydrate-active enzymes and biosynthetic glycomaterials. Front Bioeng Biotechnol 3:136. doi: 10.3389/fbioe.2015.00136

Olmstead J (2012) Bugs in the system - How the FDA fails to regulate antibiotics in ethanol production. Available at: http://www.iatp.org/files/2012_05_02_AntibioticsInEthanol_JO_0.pdf [Accessed 1 October 2015]

Olsen HS (2008) Enzymes in food processing. In Rehm HJ, Reed G. (Eds.). In Biotechnology Set. Wiley-VCH Verlag GmbH: Weinheim, Germany, pp. 663-736

Olson DG, McBride JE, Shaw AJ, Lynd LR (2012) Recent progress in consolidated bioprocessing. Curr Opin Biotechnol 23:396-405. doi: 10.1016/j.copbio.2011.11.026

Pandey A, Nigam P, Soccol CR, Soccol VT, Singh D, Mohan R (2000) Advances in microbial amylases. Biotechnol Appl Biochem 31:135-152. PMID: 10744959

Parakh S, Atkin JD (2015) Novel roles for protein disulphide isomerase in disease states: a double edged sword? Front Cell Dev Biol 3:30. doi: 10.3389/fcell.2015.00030

Pechmann S, Frydman J (2013) Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat Struct Mol Biol 20:237-243. doi: 10.1038/nsmb.2466

Peng H, Zheng Y, Chen M, Wang Y, Xiao Y, Gao Y (2014) A starch-binding domain identified in α-amylase (AmyP) represents a new family of carbohydrate-binding modules that contribute to enzymatic hydrolysis of soluble starch. FEBS Lett 588:1161-1167. doi: 10.1016/j.febslet.2014.02.050

Pereira FB, Guimarães PMR, Teixeira JA, Domingues L (2010) Selection of *Saccharomyces cerevisiae* strains for efficient very high gravity bio-ethanol fermentation processes. Biotechnol Lett 32:1655-1661. doi: 10.1007/s10529-010-0330-9

Plotkin JB, Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet 12:32-42. doi: 10.1038/nrg2899

Qian W, Yang JR, Pearson NM, Maclean C, Zhang J (2012) Balanced codon usage optimizes eukaryotic translational efficiency. PLoS Genet 8:e1002603. doi: 10.1371/journal.pgen.1002603

Qiu Z, Deng Z, Tan H, Zhou S, Cao L (2015) Engineering the robustness of *Saccharomyces cerevisiae* by introducing bifunctional glutathione synthase gene. J Ind Microbiol Biotechnol 42:537-542. doi: 10.1007/s10295-014-1573-6

Ragionieri L, Vitorino R, Frommlet J, Oliveira JL, Gaspar P, Ribas de Pouplana L, Santos MA, Moura GR (2015) Improving the accuracy of recombinant protein production through integration of bioinformatics, statistical and mass spectrometry methodologies. FEBS J 282:769-787. doi: 10.1111/febs.13181

Ranum P, Peña-Rosas JP, Garcia-Casal MN (2014) Global maize production, utilization, and consumption. Ann N Y Acad Sci 1312:105-12. doi: 10.1111/nyas.12396

Rasmussen ML, Koziel JA, Jane JL, Pometto AL (2015) Reducing bacterial contamination in fuel ethanol fermentations by ozone treatment of uncooked corn mash. J Agric Food Chem 63:5239-5248. doi: 10.1021/acs.jafc.5b00563

Robertson GH, Wong DW, Lee CC, Wagschal K, Smith MR, Orts WJ (2006) Native or raw starch digestion: a key step in energy efficient biorefining of grain. J Agric Food Chem 54:353-365. doi: 10.1021/jf051883m

Rodríguez-Sanoja R, Oviedo N, Sánchez S (2005) Microbial starch-binding domain. Curr Opin Microbiol 8:260-267. doi: 10.1016/j.mib.2005.04.013

Rosano GL, Ceccarelli EA (2009) Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. Microb Cell Fact 8:41. doi: 10.1186/1475-2859-8-41

Roth Z, Yehezkel G, Khalaila I (2012) Identification and quantification of protein glycosylation. Int J Carbohydr Chem 2012:640923. doi: 10.1155/2012/640923

Ruder W, Lu T, Collins JJ (2011) Synthetic biology moving into the clinic. Science 333:1248-1252. doi: 10.1126/science.1206843

Saibil H (2015) Chaperone machines for protein folding, unfolding and disaggregation. Nat Rev Mol Cell Biol 14:630-642. doi: 10.1038/nrm3658

Sajilata MG, Singhal RS, Kulkarni PR (2006). Resistant starch - A review. Compr Rev Food Sci Food Saf 5:1-17. doi:10.1111/j.1541-4337.2006.tb00076.x.

Santana ÁL, Meireles MA (2014) New starches are the trend for industry applications: A review. Food Public Heal 4:229-241. doi: 10.5923/j.fph.20140405.04

Santiago M, Linares L, Sánchez S, Rodríguez-Sanoja R (2005) Functional characteristics of the starch-binding domain of *Lactobacillus amylovorus* α-amylase. Biologia (Bratisl) 60:111-114

Saunders R, Deane CM (2010) Synonymous codon usage influences the local protein structure observed. Nucleic Acids Res 38:6719–6728. doi: 10.1093/nar/gkq495

Sanders R (2014) New front in war on Alzheimer's, other protein-folding diseases. Available at: http://news.berkeley.edu/2014/10/16/new-front-in-war-on-alzheimers-other-protein-folding-diseases [Accessed 1 August 2016]

Schmitt E, Gehrmann M, Brunet M, Multhoff G, Garrido C (2006) Intracellular and extracellular functions of heat shock proteins: repercussions in cancer therapy. J Leukoc Biol 81:15-27. doi: 10.1189/jlb.0306167

Schubert C (2007) Can biofuels finally take center stage? Nat Biotechnol 24:777-784. doi: 10.1038/nbt0706-777

Sharma A, Satyanarayana T (2013) Microbial acid-stable α-amylases: Characteristics, genetic engineering and applications. Process Biochem 48:201–211. doi: 10.1016/j.procbio.2012.12.018

Sharp PM, Li WH (1987) The codon adaptation index a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15:1281-1295. PMID: 3547335

Sharp PM, Stenico M, Peden JF, Lloyd AT (1993) Codon usage: mutational bias, translational selection, or both? Biochem Soc Trans 4:835-841. PMID: 8132077

Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE (2005) Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Res 33:1141-1153. doi: 10.1093/nar/gki242

Shi S, Liang Y, Zhang MM, Ang EL, Zhao H (2016) A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in *Saccharomyces cerevisiae*. Metab Eng 33:19-27. doi: 10.1016/j.ymben.2015.10.011

Shuster JR (1991) Gene expression in yeast: protein secretion. Curr Opin Biotechnol 2:685-690. PMID: 1367718

Singh AV, Nath LK, Singh A (2010) Pharmaceutical, food and non-food applications of modified starches: a critical review. Electron J Environ Agric Food Chem 9:1214-1221. ISSN: 1579-4377

Singh K, Kayastha AM (2014) α-Amylase from wheat (*Triticum aestivum*) seeds: Its purification, biochemical attributes and active site studies. Food Chem 162:1-9. doi: 10.1016/j.foodchem.2014.04.043

Sivaramakrishnan S, Gangadharan D, Nampoothiri K, Soccol CR, Pandey A (2006) α-Amylases from microbial sources - an overview on recent developments. Food Technol Biotechnol 44:173-184. ISSN: 1330-9862

Skinner KA, Leathers TD (2004) Bacterial contaminants of fuel ethanol production. J Ind Microbiol Biotechnol 31:401-408. doi: 10.1007/s10295-004-0159-0

Slattery CJ, Kavakli IH, Okita TW (2000) Engineering starch for increased quantity and quality. Trends Plant Sci 5:291-298. doi: 10.1016/S1360-1385(00)01657-5

Smith AM (2001) The biosynthesis of starch granules. Biomacromolecules 2:335-341. doi: 10.1021/bm000133c

Smith R (2011) Antibiotic-free DDGS launched. Available at: http://feedstuffsfoodlink.com/story-antibioticfree-ddgs-launched-71-66394 [Accessed 2 October 2015]

Sørensen HP, Mortensen KK (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microb Cell Fact 4:1. doi: 10.1186/1475-2859-4-1

Sorimachi K, Le Gal-Coëffet MF, Williamson G, Archer DB, Williamson MP (1997) Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to β-cyclodextrin. Structure 5:647-661. doi: 10.1016/S0969-2126(97)00220-7

Steensels J, Snoek T, Meersman E, Nicolino MP, Voordeckers K, Verstrepen KJ (2014) Improving industrial yeast strains: exploiting natural and artificial diversity. FEMS Microbiol Rev 38:947-995. doi: 10.1111/1574-6976.12073

Stoletzki N, Eyre-Walker A (2007) Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. Mol Biol Evol 24:374-381. doi: 10.1093/molbev/msl166

Stovicek V, Borja GM, Forster J, Borodina I (2015) EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial *Saccharomyces cerevisiae* strains. J Ind Microbiol Biotechnol 42:1519-1531. doi: 10.1007/s10295-015-1684-8

Streb S, Zeeman S (2012) Starch metabolism in Arabidopsis. Arabidopsis Book 9:e0160. doi: 10.1199/tab.0160

Sun H, Zhao P, Ge X, Xia Y, Hao Z, Liu J, Peng M (2010) Recent advances in microbial raw starch degrading enzymes. Appl Biochem Biotechnol 160:988-1003. doi: 10.1007/s12010-009-8579-y

Sundarram A, Pandurangappa T, Murthy K (2014) α-Amylase production and applications: a review. J Appl Environ Microbiol 2:166-175. doi: 10.12691/jaem-2-4-10

Szymanowska-Powałowska D, Lewandowicz G, Błaszczak W, Szwengiel A (2012) Structural changes of corn starch during fuel ethanol production from corn flour. BioTechnologia 93:333-341. doi: 10.5114/bta.2012.46587

Tan NS, Ho B, Ding JL (2002) Engineering a novel secretion signal for cross-host recombinant protein expression. Protein Eng 15:337-345. doi: 10.1093/protein/15.4.337

Tawil G, Viksø-Nielsen A, Rolland-Sabaté A, Colonna P, Buléon A (2011) In depth study of a new highly efficient raw starch hydrolyzing α-amylase from *Rhizomucor* sp. Biomacromolecules 12:34-42. doi: 10.1021/bm100913z

Tester RF, Karkalas J, Qi X (2004) Starch—composition, fine structure and architecture. J Cereal Sci 39:151-165. doi: 10.1016/j.jcs.2003.12.001

Thanaraj TA, Argos P (1996) Ribosome-mediated translational pause and protein domain organization. Protein Sci 5:1594-1612. doi: 10.1002/pro.5560050814

Tsai CS, Kwak S, Turner TL, Jin Y-S (2015) Yeast synthetic biology toolbox and applications for biofuel production. FEMS Yeast Res 15:1-15. doi: 10.1111/1567-1364.12206

Tsumoto K, Ejima D, Kumagai I, Arakawa T (2003) Practical considerations in refolding proteins from inclusion bodies. Protein Expr Purif 28:1-8. doi: 10.1016/S1046-5928(02)00641-1

Tiroli-Cepeda A, Ramos C (2011) An overview of the role of molecular chaperones in protein homeostasis. Protein Pept Lett 18:101-109. PMID: 21121892

Tucker JB, Zilinskas RA (2006) The promise and perils of synthetic biology. New Atlantis 12:25-45. PMID: 16832953

Ubwa ST, Abah J, Asemave K, Shambe T (2012) Studies in the gelatinization temperature of some cereal starches. Int J Chem 4:22-28. doi: 10.5539/ijc.v4n6p22.

van der Maarel MJ, van der Veen B, Uitdehaag JC, Leemhuis H, Dijkhuizen L (2002) Properties and applications of starch-converting enzymes of the alpha-amylase family. J Biotechnol 94:137-155. doi: 10.1016/S0168-1656(01)00407-2

van Zyl WH, Bloom M, Viktor MJ (2012) Engineering yeasts for raw starch conversion. Appl Microbiol Biotechnol 95:1377-1388. doi: 10.1007/s00253-012-4248-0

Vu HV, Pham TA, Kim K (2010) Improvement of a fungal strain by repeated and sequential mutagenesis and optimization of solid-state fermentation for the hyper-production of raw-starch-digesting enzyme. J Microbiol Biotechnol 20:718-726. doi: 10.4014/jmb.0908.08016

Venton D (2014) Core Concept: Synthetic biology—change, accelerated: Proc Natl Acad Sci USA 111:16978-16979. doi: 10.1073/pnas.1419688111

Viktor MJ, Rose SH, van Zyl WH, Viljoen-Bloom M (2013) Raw starch conversion by *Saccharomyces cerevisiae* expressing *Aspergillus tubingensis* amylases. Biotechnol Biofuels 6:167. doi: 10.1186/1754-6834-6-167

Villalobos A, Ness JE, Gustafsson C, Minshull J, Govindarajan S (2006) Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics 7:285. doi: 10.1186/1471-2105-7-285

Wang C, Huang R, He B, Du Q (2012) Improving the thermostability of alpha-amylase by combinatorial coevolving-site saturation mutagenesis. BMC Bioinformatics 13:263. doi: 10.1186/1471-2105-13-263

Wang T (2015) Engineering yeast for cellulosic ethanol production. Austin J Chem Eng 2:1018. ISSN: 2381-8905

Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, Gustafsson C (2009) Design parameters to control synthetic gene expression in *Escherichia coli*. PLoS One 4:e7002. doi: 10.1371/journal.pone.0007002

Welch M, Villalobos A, Gustafsson C, Minshull J (2009b) You're one in a googol: optimizing genes for protein expression. J R Soc Interface. doi: 10.1098/rsif.2008.0520.focus

Wong DW, Robertson GH, Lee CC, Wagschal K (2007) Synergistic action of recombinant alpha-amylase and glucoamylase on the hydrolysis of starch granules. Protein J 26:159-164. doi: 10.1007/s10930-006-9057-9

Wood T, Garcia-Campayo V (1990) Enzymology of cellulose degradation. Biodegradation 1:147-161. doi: 10.1007/BF00058833

Wu G, Dress L, Freeland SJ (2007) Optimal encoding rules for synthetic genes: the need for a community effort. Mol Syst Biol 3:134. doi: 10.1038/msb4100176

Wu X, Jörnvall H, Berndt KD, Oppermann U (2004) Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance. Biochem Biophys Res Commun 313:89-96. doi: 10.1016/j.bbrc.2003.11.091

Xiang Z, Runge T (2016) Emulsifying properties of succinylated arabinoxylan-protein gum produced from corn ethanol residuals. Food Hydrocoll 52:423-430. doi: 10.1016/j.foodhyd.2015.07.018

Xiao S, Shiloach J, Betenbaugh MJ (2014) Engineering cells to improve protein expression. Curr Opin Struct Biol 26:32-38. doi: 10.1016/j.sbi.2014.03.005

Yadava A, Ockenhouse CF (2003) Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. Infect Immun 71:4961-4969. doi: 10.1128/IAI.71.9.4961-4969.2003

Yamada R, Tanaka T, Ogino C, Fukuda H, Kondo A (2010) Novel strategy for yeast construction using δ-integration and cell fusion to efficiently produce ethanol from raw starch. Appl Microbiol Biotechnol 85:1491-1498. doi: 10.1007/s00253-009-2198-y

Yamada R, Taniguchi N, Tanaka T, Ogino C, Fukuda H, Kondo A (2011) Direct ethanol production from cellulosic materials using a diploid strain of *Saccharomyces cerevisiae* with optimized cellulase expression. Biotechnol Biofuels 4:8. doi: 10.1186/1754-6834-4-8

Yamaguchi H, Miyazaki M (2014) Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. Biomolecules 4:235-251. doi: 10.3390/biom4010235

Yang S, Kuang Y, Li H, Liu Y, Hui X, Li P, Jiang Z, Zhou Y, Wang Y, Xu A, Li S, Liu P, Wu D (2013) Enhanced production of recombinant secretory proteins in *Pichia pastoris* by optimizing Kex2 P1' site. PLoS One. 8:e75347. doi: 10.1371/journal.pone.0075347.

Young E, Alper H (2010) Synthetic biology: tools to design, build, and optimize cellular processes. J Biomed Biotechnol 2010:13078. doi: 10.1155/2010/130781

Young JC, Agashe VR, Siegers K, Hartl FU (2004) Pathways of chaperone-mediated protein folding in the cytosol. Nat Rev Mol Cell Biol 5:781-791. doi: 10.1038/nrm1492

# Chapter 3

Expression and comparison of codon optimised *Aspergillus tubingensis* amylase variants in *Saccharomyces cerevisiae*

# Expression and comparison of codon optimised *Aspergillus tubingensis* amylase variants in *Saccharomyces cerevisiae*

Rosemary A. Cripwell, Shaunita H. Rose, Willem H. van Zyl

Department of Microbiology, Stellenbosch University, Private Bag X1, 7602 Matieland, South Africa.

## 3.1 Abstract

Starch can be used as feedstock for bioethanol production, once an effective raw starch utilising *Saccharomyces cerevisiae* strain has been constructed. The *Aspergillus tubingensis* α-amylase (*amyA*) and glucoamylase (*glaA*) genes were expressed in the laboratory *S. cerevisiae* Y294 strain and the effect of codon optimisation and codon bias strategies were evaluated. Codon optimisation to favour the *S. cerevisiae* codon bias resulted in a decrease in extracellular enzyme activity of 72% (30 nkat.ml$^{-1}$) and 68% (4 nkat.ml$^{-1}$) compared to the expression of the native *amyA* and *glaA* genes, respectively, after 96 hours of cultivation. Lower levels of protein were secreted and dissimilar *N*-glycosylation patterns observed. A lower cultivation temperature and co-expression with the *PDI1* gene increased extracellular activity levels of the codon optimised α-amylase and glucoamylase, respectively. Despite the identical amino acid sequence of the GlaA, GlaA_Opt and GlaA_CBI proteins, differential scanning fluorimetry revealed changes in the glucoamylase proteins' melting temperatures (> 3°C). Shifts in the fluorescence curves were also observed suggesting changes in glucoamylase tertiary structure. Overall, these results suggest that the native genes have translational information in their mRNA sequence that is lost during codon optimisation, leading to aberrant folding. Specific synonymous codon changes to the coding regions of *amyA* and *glaA* genes confirms that codon optimisation is not always the best strategy for recombinant protein production and that there is crucial translational information present within the coding sequence that controls protein production levels, influences protein secretion and protein folding.

**Keywords:** • α-amylase • glucoamylase • codon optimisation • protein secretion • protein folding

## 3.2 Introduction

Amylases account for 25 - 33% of the international enzyme market and are used for many industrial processes (de Oliveira et al. 2015), including the hydrolysis of starch for bioethanol production. The enzyme industry is growing rapidly and relies on new technologies for the identification and characterisation of novel enzymes with improved activities (Li et al. 2012). A cost effective way to produce bioethanol directly from uncooked starch would be by expressing heterologous amylase genes in *Saccharomyces cerevisiae* in a process called consolidated bioprocessing (CBP). Current trends include the use of synthetic biology tools for the investigation and improvement of heterologous gene expression, instead of conventional cloning techniques (Wu et al. 2007).

The flow of genetic information from DNA via mRNA to protein is affected by numerous factors that ultimately dictate efficient heterologous gene expression and protein folding (or function). The promoter selected and GC content of the gene affect the transcription of the genetic material, whereas the codon usage and complexity of the mRNA secondary structure may hinder the efficiency of translation and significantly influence the rate of protein synthesis (Zhou et al. 2015). Therefore, the use of *de novo* gene design is a powerful technique to investigate how the optimisation of DNA sequences can improve gene expression and protein production.

In the past, the central dogma of molecular biology proposed that synonymous codon changes, that do not alter the protein sequence, will not affect the resulting protein and therefore have no functional consequence (Plotkin and Kudla 2011; Hunt et al. 2014). Although the general perception surrounding “optimal” or “suboptimal” codons is often misguided (Hockenberry et al. 2014), it is now recognised that this phenomenon, termed codon-usage bias, plays a significant role in gene expression and cellular function. Recent studies have shown that codon changes can significantly affect a variety of processes, ranging from RNA processing to recombinant protein production by affecting translational speed and accuracy, regulation of co-translational folding and protein secretion (Plotkin and Kudla 2011; Shah et al. 2015).

Historically there have been two approaches to the design of synthetic genes, firstly the “one amino acid – one codon” (Sharp and Li 1987) approach and secondly codon randomisation, which is based on the frequency distribution of codons (i.e. codon optimisation) (Elena et al. 2014). In recent studies it is the latter approach that is

commonly employed for gene synthesis (Welch et al. 2009). In general, the more rare codons a gene contains the less likely it is that the protein will be produced at reasonable levels (Gustafsson et al. 2004). Therefore, improved gene expression can be obtained through synonymous codon usage where the codon bias for the target gene reflects the codon usage of the expression host (Gustafsson et al. 2004). However, the presence of rare codons may not necessarily be a hindrance, but could be a result of species evolution/natural selection (Wu et al. 2007).

The expression of codon optimised genes in *S. cerevisiae* for heterologous protein production remains an empirical process leaving much scope to explore additional factors that affect the expression of optimised genes. This study aimed to determine how codon bias of optimised amylase genes affects extracellular enzyme production and activity in *S. cerevisiae*, using synthetic variants of the *Aspergillus tubingensis* α-amylase (*amyA*) and glucoamylase (*glaA*) genes. The objective was to obtain more information regarding the relationship between adapted synonymous codon usage and extracellular enzyme levels, which would be valuable for the development of a high-level expression system.

## 3.3 Materials and methods

### 3.3.1 Media and cultivation conditions

Unless stated otherwise, all chemicals were of analytical grade and were obtained from Merck (Darmstadt, Germany). The *Escherichia coli* transformants were cultivated at 37°C in Terrific Broth (12 $g.l^{-1}$ tryptone, 24 $g.l^{-1}$ yeast extract, 4 $ml.l^{-1}$ glycerol, 0.1 M potassium phosphate buffer) containing 100 $\mu g.ml^{-1}$ ampicillin for selective pressure (Sambrook et al. 1989). The *S. cerevisiae* Y294 strain was maintained on YPD agar plates (10 $g.l^{-1}$ yeast extract, 20 $g.l^{-1}$ peptone, 20 $g.l^{-1}$ glucose and 20 $g.l^{-1}$ agar) and transformants were selected and maintained on $SC^{-URA}$ agar plates containing 6.7 $g.l^{-1}$ yeast nitrogen base without amino acids (BD-Diagnostic Systems, Maryland, USA), 20 $g.l^{-1}$ glucose, 1.5 $g.l^{-1}$ yeast synthetic drop-out medium supplements (Sigma-Aldrich, Steinheim, Germany) and 20 $g.l^{-1}$ agar. All *S. cerevisiae* strains were aerobically cultivated on a rotary shaker (200 rpm) at 30°C, in 125 ml Erlenmeyer flasks containing 20 ml double strength $SC^{-URA}$ medium ($2\times SC^{-URA}$) containing 13.4 $g.l^{-1}$ yeast nitrogen base without amino acids (BD-Diagnostic Systems, Maryland, USA), 20 $g.l^{-1}$ glucose and 3 $g.l^{-1}$ yeast synthetic drop-out medium

supplements. Unless otherwise stated, all cultures were inoculated to a concentration of $1\times10^6$ cells.ml$^{-1}$.

### 3.3.2 Strains and plasmids

The genotypes of the bacterial and yeast strains, as well as the plasmids used in this study are summarised in Table 3.1.

**Table 3.1**. Strains and plasmids used in this study

| Strains and plasmids | Genotype | Reference/Source |
|---|---|---|
| *E. coli DH5α* | *supE44 ΔlacU169 (ϕ80lacZ*ΔM15*) hsdR17 recA1 endA1 gyrA96 thi-1 relA1* | Sambrook et al. (1989) |
| ***S. cerevisiae* strains** | | |
| Y294 | α *leu2*-3,112 *ura3*-52 *his3 trp1-289* | ATCC 201160 |
| NI-C-D4 | α *trp1 ura3 pep4* | Wang et al. (2001) |
| Y294[BBH1] | *URA3 ENO1*$_P$*-ENO1*$_T$ | Viktor et al. (2013) |
| Y294[AmyA] | *URA3 ENO1*$_P$*-amyA-ENO1*$_T$ | Viktor et al. (2013) |
| Y294[GlaA] | *URA3 ENO1*$_P$*-glaA-ENO1*$_T$ | Viktor et al. (2013) |
| Y294[BBH1-AmyA_Opt] | *URA3 ENO1*$_P$*-amyA_Opt-ENO1*$_T$ | This study |
| Y294[BBH1AmyA_CBI] | *URA3 ENO1*$_P$*-amyA_CBI-ENO1*$_T$ | This study |
| Y294[BBH1-AmyA_Opt(→300)] | *URA3 ENO1*$_P$*-amyA_Opt(→300)-ENO1*$_T$ | This study |
| Y294[BBH1-AmyA_Opt(→426)] | *URA3 ENO1*$_P$*-amyA_Opt(→426)-ENO1*$_T$ | This study |
| Y294[BBH1-AmyA_Opt(→300 + 426)] | *URA3 ENO1*$_P$*-amyA_Opt(→300 + 426)-ENO1*$_T$ | This study |
| Y294[BBH1-GlaA_Opt] | *URA3 ENO1*$_P$*-glaA_Opt-ENO1*$_T$ | This study |
| Y294[BBH1-GlaA_CBI] | *URA3 ENO1*$_P$*-glaA_CBI-ENO1*$_T$ | This study |
| Y294[BBH1-GlaA_Opt(→252) | *URA3 ENO1*$_P$*-glaA_Opt(→262)-ENO1*$_T$ | This study |
| NI-C-D4[AmyA_Opt] | *URA3 ENO1*$_P$*-amyA_Opt-ENO1*$_T$ | This study |
| NI-C-D4[GlaA_Opt] | *URA3 ENO1*$_P$*-glaA_Opt-ENO1*$_T$ | This study |
| Y294[BBH1-AmyA_Opt & PDI1] | *URA3 ENO1*$_P$*-amyA_Opt-ENO1*$_T$ *; TRP1 ENO1*$_P$*-PDI1-ENO1*$_T$ | This study |
| Y294[BBH1-GlaA_Opt & PDI1] | *URA3 ENO1*$_P$*-glaA_Opt-ENO1*$_T$ ; *TRP1 ENO1*$_P$*-PDI1-ENO1*$_T$ | This study |
| **Plasmids** | | |
| yBBH1 | *URA3 ENO1*$_P$ *-ENO1*$_T$ | Njokweni et al. (2012) |
| yBBH1-AmyA | *bla URA3 ENO1*$_P$*-amyA-ENO1*$_T$ | Viktor et al. (2013) |
| yBBH1-GlaA | *bla URA3 ENO1*$_P$*-glaA-ENO1*$_T$ | Viktor et al. (2013) |
| yBBH1-AmyA_Opt | *bla URA3 ENO1*$_P$*-amyA_Opt-ENO1*$_T$ | This study |
| yBBH1-AmyA_CBI | *bla URA3 ENO1*$_P$*-amyA_CBI-ENO1*$_T$ | This study |
| yBBH1-GlaA_Opt | *bla URA3 ENO1*$_P$*-glaA_Opt-ENO1*$_T$ | This study |
| yBBH1-GlaA_CBI | *bla URA3 ENO1*$_P$*-glaA_CBI-ENO1*$_T$ | This study |
| yBBH1-AmyA_Opt(→426) | *bla URA3 ENO1*$_P$*-amyA_Opt(→426)-ENO1*$_T$ | This study |
| yBBH1-AmyA_Opt(→300) | *bla URA3 ENO1*$_P$*-amyA_Opt(→300)-ENO1*$_T$ | This study |
| yBBH1-AmyA_Opt(→300 + 426) | *bla URA3 ENO1*$_P$*-amyA_Opt(→*300 + 426*)-ENO1*$_T$ | This study |
| yBBH1-GlaA_Opt(→252) | *bla URA3 ENO1*$_P$*-glaA_Opt(→252)-ENO1*$_T$ | This study |
| YIplac204* | *bla TRP1* | ATCC 87591 |
| YIplacPdi1* | *bla TRP1 ENO1*$_P$*-PDI1-ENO1*$_T$ | This laboratory |

*Single copy integration vector

### 3.3.3 DNA manipulations

Standard protocols were followed for all DNA manipulations and *E. coli* transformations (Sambrook et al. 1989). The enzymes used for restriction digests and ligations were purchased from Inqaba Biotec (Pretoria, South Africa) and used as recommended by the supplier. Digested DNA was eluted from 0.8% agarose gels using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, California, USA). Plasmid DNA was isolated from *S. cerevisiae* Y294 strains using the High Pure Plasmid Isolation kit (Roche, Mannheim, Germany). Sequence verification of the final vector constructs was performed by the dideoxy chain termination method, with an ABI PRISM™ 3100 Genetic Analyser (CAF, Stellenbosch University, South Africa).

### 3.3.4 Amylase genes and GenBank accession numbers

The *amyA_Opt* and *gla_Opt* genes were codon optimised for expression in *S. cerevisiae* (DNA2.0, California, USA) and the DNA sequences deposited (Accession numbers KX959309 and KX959311, respectively). The *amyA_CBI* and *glaG_CBI* genes were designed to have a codon bias index (CBI; Carbone et al. 2003) of 1.0 using JCAT software (JAVA Codon Adaption Tool – http://www.jcat), they were synthesised by GenScript (Piscataway, USA) and the DNA sequences deposited (Accession numbers KX959310 and KX959312, respectively). The native *amyA* and *glaA* genes (Viktor et al. 2013) were also expressed in *S. cerevisiae* for comparative purposes. Synthetic copies of the *amyA* and *glaA* genes were obtained that encoded for proteins with identical amino acid sequence to that of the native *amyA* (Accession number JF809672) and *glaA* (Accession number AY528665) genes.

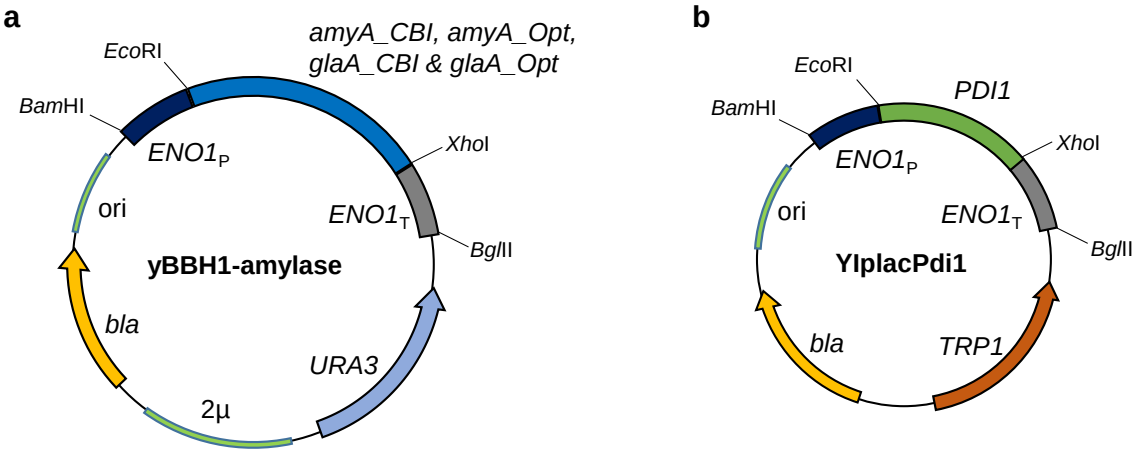


**Fig. 3.1** Schematic representation of the vector constructs used in this study. (**a**) The amy*A_Opt* (and derivatives), *amyA_CBI*, *glaA_Opt* (and *glaA_Opt→252*) and *glaA_CBI* genes were expressed from the yBBH1 multicopy episomal vector, whereas (**b**) the *PDI1* from *S. cerevisiae* was constitutively expressed from a single chromosomal integrative copy using the enolase 1 (*ENO1*) promoter and terminator sequences for transcriptional control.

### 3.3.5 Yeast strain construction

Synthetic *amyA_Opt, amyA_CBI, glaA_Opt* and *glaA_CBI* genes were subcloned individually onto the episomal yBBH1 plasmid (Fig. 3.1a, Table 3.1). The yBBH1-AmyA_Opt and yBBH1-GlaA_Opt vectors were linearised with *Hin*dIII and *Xba*I, respectively. Three regions were identified for codon exchange between native and optimised amylase genes (Fig. 3.2a and 3.2b).

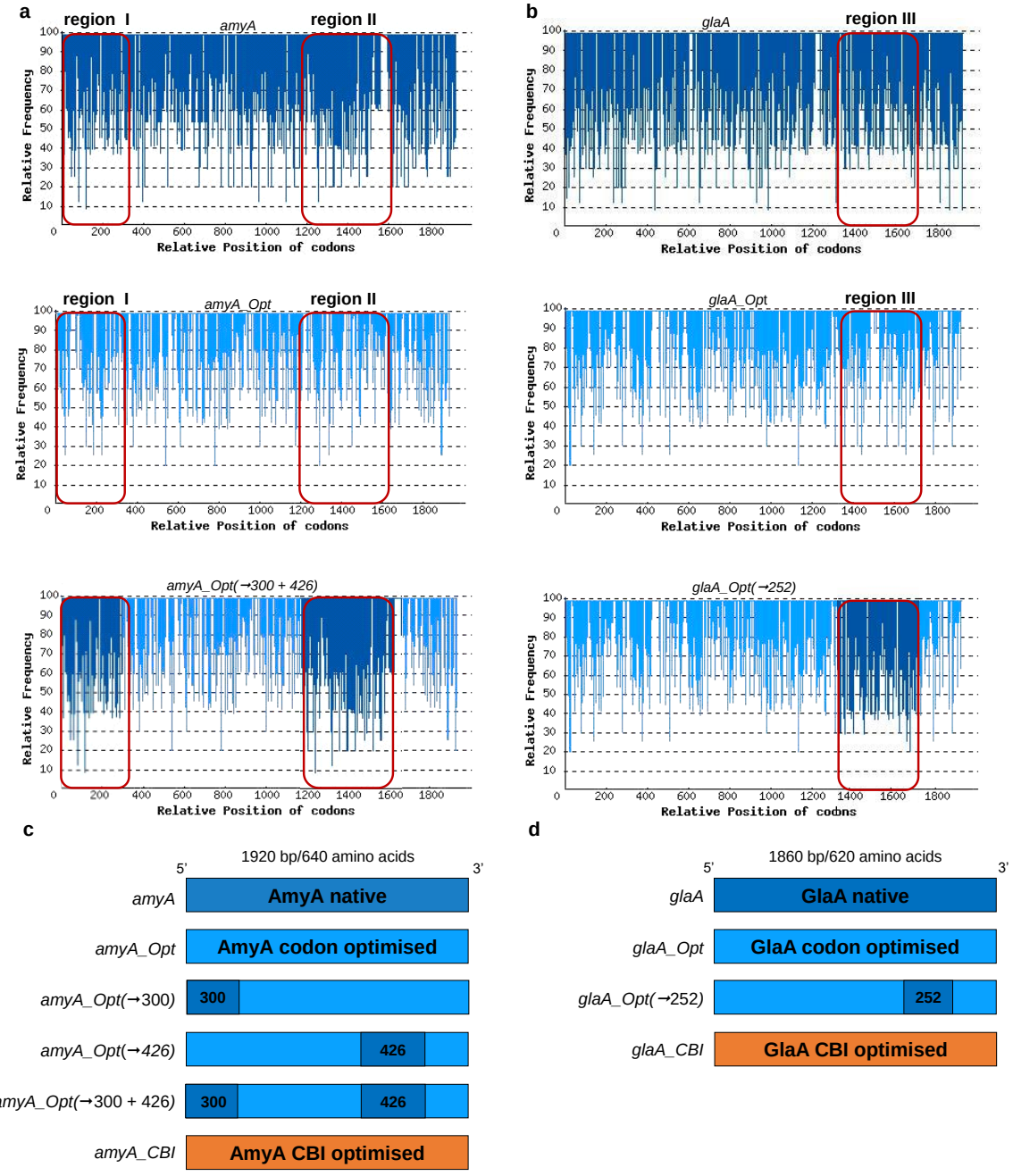


**Fig. 3.2** Illustration of the relative frequency of codon usage for the (**a**) *amyA* and (**b**) *glaA* gene variants indicating regions I, II and III that were replaced to construct *amyA_Opt(→300)*, *amyA_Opt(→426)*, *amyA_Opt(→300 + 426)* and *glaA_Opt(→252)*, respectively. Schematic representation (**c** and **d**) of *amyA* and *glaA* gene constructs.

Region II (base pairs 1188 to 1614) of *amyA* and region III (base pairs 1359 to 1611) of *glaA* were cloned onto the linearised vectors by means of yeast mediated ligation (Cho et al. 1999) to obtain yBBH1-AmyA_Opt(→426) and yBBH-GlaA_Opt(→252), respectively (Fig. 3.2c and Fig. 3.2d). Subsequently, the first 300 bp of the native *amyA* gene (Region I - Fig. 3.2a) was restricted from yBBH1-AmyA and used to replace the corresponding sequences on yBBH1-AmyA_Opt and yBBH1-AmyA_Opt(→426) to generate the yBBH1-AmyA_Opt(→300) and yBBH1-AmyA_Opt(→300 + 426) vectors, respectively (Fig. 3.2c and 3.2d).

The *S. cerevisiae's PDI1* gene was cloned into the YIplac204 vector under the control of the *ENO1* promoter and terminator sequences to generate YIplacPdi1 (Fig. 3.1b). The YIplacPdi1 was transformed to the *S. cerevisiae* Y294[AmyA_Opt] and Y294[GlaA_Opt] strains and transformants selected for growth on $SC^{-URA-TRP}$ plates. Plasmid DNA was transformed to electro-competent *S. cerevisiae* cells (Cho et al. 1999) using a Bio-Rad system (GenePluserXcell TM, Bio-Rad, California, USA) at 1.4 kV, 200 Ω and 25 µF using 0.2 cm electroporation cuvettes. The presence of integrated *PDI1* was confirmed using colony PCR with the ENOCASS-L: 5'-gtgcggtatttcacaccgcataggagatcgatcccaattaatgtgagttacctcactc-3' and ENOCASS-R: 5'-cgggcctcttcgctattacgccagagcttagatct-3' primers. The biomass of the yeast strains were calculated as dry cell weight (DCW) according Den Haan et al. (2007).

### 3.3.6 Characterisation of recombinant strains and enzymes

#### *3.3.6.1 Amylase assays*

For quantitative assays, yeast transformants were cultured in 20 ml $2\times SC^{-URA}$ medium in 125 ml Erlenmeyer flasks with agitation at 200 rpm and sampling at 24 hour intervals. The supernatant was harvested and extracellular enzymatic activity levels were assessed colourimetrically (xMark™ Microplate Spectrophotometre, Bio-Rad, California, USA) using the reducing sugar assay with glucose as standard (Miller 1959). The α-amylase and glucoamylase activities were determined at pH 5 and at 37°C according to the method described by Viktor et al. (2013). Enzymatic activities were expressed as nano-katals per ml ($nkat.ml^{-1}$), with nkat defined as the enzyme activity needed to produce 1 nmol of glucose per second under the described assay conditions.

### 3.3.6.2 *Protein analysis*

Protein samples were separated using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) as described by Laemmli (1970). Proteins were visualised using the silver staining method (O'Connell and Stults 1997). The broad-range Page Ruler Prestained SM0671 Protein Ladder (Fermentas, Shenzhen, China) was used as a molecular mass marker. The protein content of crude extracellular extracts was determined using the Bio-Rad protein reagent (Bio-Rad, California, USA) as directed by the manufacturer. Protein concentrations were determined colourimetrically at 750 nm (xMark$^{TM}$ Microplate Spectrophotometre, Bio-Rad, California, USA), using bovine serum albumin (BSA) as standard.

### 3.3.6.3 *Protein deglycosylation*

The supernatant of *S. cerevisiae* Y294[AmyA], Y294[AmyA_Opt], Y294[AmyA_CBI], Y294[GlaA], Y294[GlaA_Opt] and Y294[GlaA_CBI] strains was harvested by centrifugation of the culture at 4000 × g for two minutes. Protein samples were deglycosylated using the N-glycosidase F kit (PNGase F, New England Biolabs, Massachusetts, USA) as described by the manufacturer. Untreated AmyA and GlaA proteins were prepared in the same manner, but lacked the PNGase F enzyme.

### 3.3.6.4 *Preparation of partially purified enzymes*

Yeast strains were cultivated in 2 litre Erlenmeyer flasks containing 500 ml 2×SC$^{-URA}$ and 0.2 M succinate buffer (pH 6), at 30°C with orbital shaking at 200 rpm. Cultures were centrifuged after 96 hours of cultivation and filtered through 0.45 µm filters (Millipore, Massachusetts, USA) to obtain the supernatant. The Minitan (Millipore Massachusetts, USA) system with a 10 kDa cut-off membrane was used to concentrate the supernatant samples to ~200 ml. The Amicon ultrafiltration system (Millipore) equipped with a 30 kDa cut-off membrane was used to concentrate samples to 50 ml. All purification steps were carried out at 4°C.

### 3.3.6.5 *Differential scanning fluorimetry*

The StepOne real time polymerase chain reaction (PCR) instrument (Applied Biosystems) was used for differential scanning fluorimetry (DSF) analysis. SYPRO® Orange Protein Gel Stain (Sigma-Aldrich, Steinheim, Germany) was used according to the protocol outlined by Niesen et al. (2007) using 0.5 mg.ml$^{-1}$ purified protein in 0.05 M citrate buffer (pH 5.0). Fluorescence readings were monitored between

25°C - 95°C (increasing the temperature in increments of 1°C.min$^{-1}$). Melting curves were exported into GraphPad Prism® 6.01 software and a Boltzmann curve was fitted to determine melting temperature values. The temperature at which 50% of the protein is unfolded ($T_m$) can be estimated from the inflexion point of the curves.

### 3.3.7 Statistical analysis

Data was analysed using the Student's t-test and by a single factor ANOVA (analysis of variance).

## 3.4 Results

Bioethanol production from uncooked starch has been demonstrated by the co-expression of the *A. tubingensis amyA* and *glaA* genes in *S. cerevisiae,* resulting in the secretion of α-amylases and glucoamylases, respectively (Viktor et al. 2013). In this study, the synthetic *amyA_Opt, amyA_CBI*, *glaA_Opt* and *glaA_CBI* genes (codon optimised variants of the *amyA* and *glaA* genes) were used to investigate the effect of adapted synonymous codon usage on enzyme production and activity. The α-amylases produced by the *S. cerevisiae* Y294[AmyA], Y294[AmyA_Opt] and Y294[AmyA_CBI] strains will be referred to as AmyA, AmyA_Opt and AmyA_CBI, respectively, although they have the same amino acid sequence. Similarly, the glucoamylases produced by the *S. cerevisiae* Y294[GlaA], Y294[GlaA_Opt] and Y294[GlaA_CBI] strains will be referred to as GlaA, GlaA_Opt and GlaA_CBI, respectively.

### 3.4.1 Effect of cultivation temperature on extracellular enzyme activity

Duplicate *S. cerevisiae* Y294 cultivations were performed at 20°C and 30°C, the latter being routinely used as cultivation temperature (Fig. 3.3) (Wasungu and Simard 1982). Cultivation at a lower temperature resulted in an increase in extracellular amylase activity observed for all strains, towards the end of the cultivation period. Significantly lower extracellular activity levels were observed for the *S. cerevisiae* Y294 strains expressing the optimised amylase genes (Fig. 3.3a and 3.3b) compared to the *S. cerevisiae* Y294 strains expressing the native *amyA* and *glaA* genes from *A. tubingensis*. The activities observed for the *S. cerevisiae* Y294[AmyA_Opt] and Y294[AmyA_CBI] strains were comparable at both 20°C and 30°C with no significant difference in α-amylase activity detected at 72 and 96 hours. However, although the

extracellular α-amylase activities of AmyA, AmyA_Opt and AmyA_CBI at 20°C were higher than at 30°C, the relative percentage increase in extracellular activity for all three α-amylase at 20°C was statistically the same ($p$ = 0.8990).

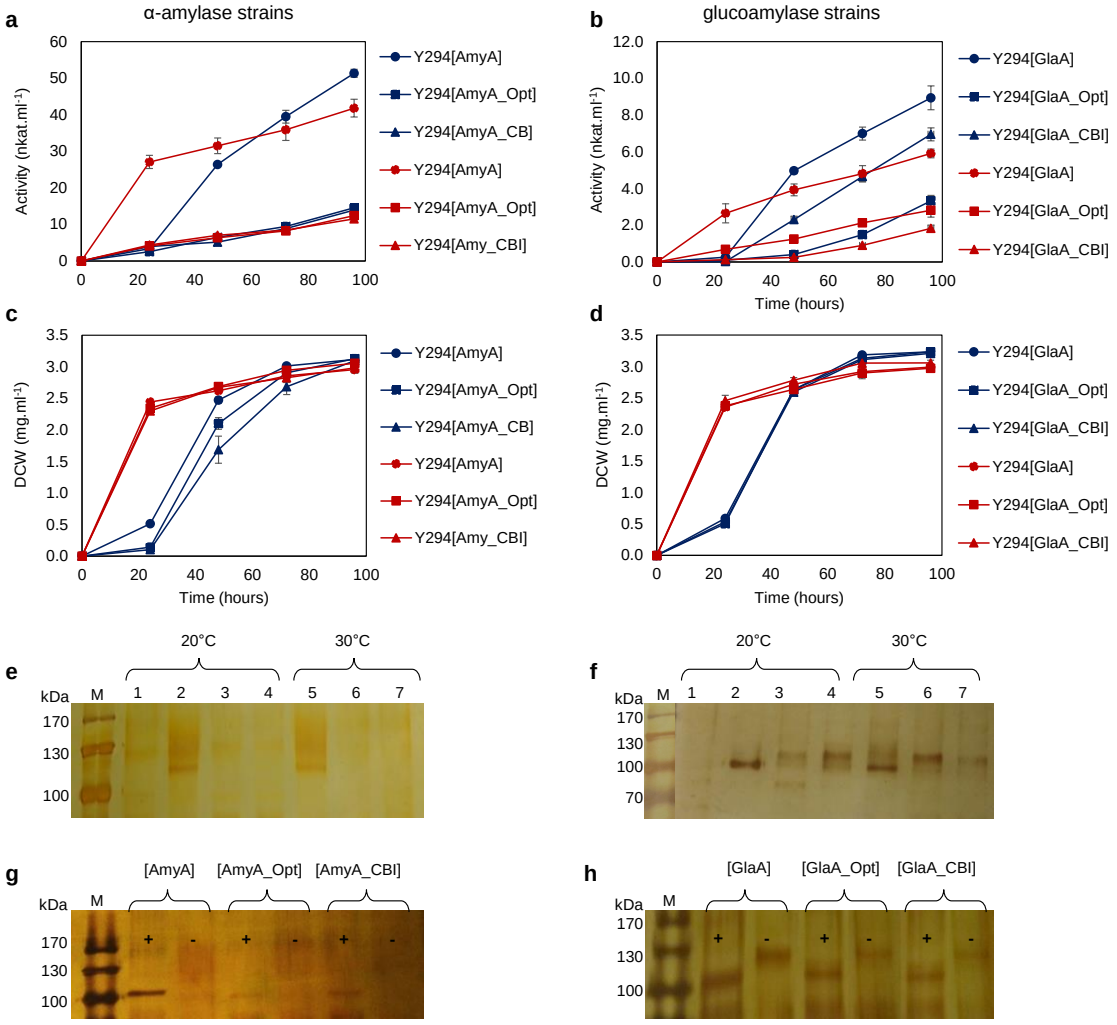


**Fig. 3.3** The effect of cultivation temperature on recombinant *S. cerevisiae* strains. Blue lines represent strains cultivated at 20°C and red lines strains cultivated at 30°C. The recombinant (**a**) α-amylase and (**b**) glucoamylase activities, as well as DCW (**c** and **d**) of the recombinant *S. cerevisiae* Y294 strains. Error bars represent the standard deviation of the three independent biological replicates. Supernatant from (**e**) (lanes 2,5) *S. cerevisiae* Y294[AmyA], (lane 3,6) Y294[AmyA_Opt] and (lane 4,7) Y294[AmyA_CBI], as well as (**f**) (lanes 2,5) Y294[GlaA], (lanes 3,6) Y294[GlaA_Opt] and (lane 4,7) Y294[GlaA_CBI] strains were subjected to SDS-PAGE with the protein size markers in lane M and *S. cerevisiae* Y294[BBH1] reference strain in lane 1. The degylcosylated (+) and (-) untreated (**g**) α-amylases and (**h**) glucoamylases were visualised by SDS-PAGE. The protein sizes are depicted on the left hand side.

With cultivation at 30°C, extracellular glucoamylase activity produced by the *S. cerevisiae* Y294[GlaA_CBI] and Y294[GlaA_Opt] strains was 2-fold lower compared to the *S. cerevisiae* Y294[GlaA] strain (Fig. 3.3b). A decrease in cultivation temperature resulted in a significant increase in extracellular glucoamylase activity for

both the *S. cerevisiae* Y294[GlaA_CBI] and Y294[GlaA] strains. A 3.6-fold increase in GlaA_CBI activity was observed after 96 hours of cultivation at 20°C (Fig. 3.3b) compared to 30°C, whereas the activity of the native GlaA only increased by 1.4-fold. The cultivation temperature had no significant effect on GlaA_Opt activity after 96 hours. Overall, extracellular amylase activity (Fig. 3.3a and 3.3.b) and productivity levels (data not shown) for all strains increased steadily after 48 hours of cultivation.

As expected, all strains reached stationary phase faster with cultivation at 30°C, which was closer to the preferred cultivation temperature for this laboratory *S. cerevisiae* Y294 strain (Fig. 3.3c and 3.3d). The strains expressing glucoamylase produced similar growth patterns at 20°C and 30°C. However, the *S. cerevisiae* Y294[AmyA_CBI] strain produced significantly less biomass (over the first 72 hours) compared to the benchmark *S. cerevisiae* Y294[AmyA] strain when cultivated at 20°C, which was less noticeable at 30°C (Fig. 3.3c). Once all strains reached stationary phase, the biomass production was similar (approximately 3 $mg.ml^{-1}$ DCW produced) (Fig. 3.3c and 3.3d).

SDS-PAGE analysis showed correlation between activity levels and recombinant protein concentrations (Fig. 3.3a, b, e and f). The α-amylases and glucoamylases were respectively produced as different heterogeneously glycosylated species (Fig. 3.3e and 3.3f). However, after deglycosylation the α-amylase (Fig. 3.3g) and glucoamylase (Fig. 3.3f) recombinant protein species were approximately the same size, around 100 kDa and 110 kDa, respectively.

### 3.4.2 DSF analysis of enzymes

The supernatant from the *S. cerevisiae* Y294[AmyA], Y294[AmyA_Opt], Y294[AmyA_CBI], Y294[GlaA], Y294[GlaA_Opt] and Y294[GlaA_CBI] strains was concentrated and partially purified for further analysis. The purified α-amylases had similar specific activities, while the specific activities for the glucoamylases differed significantly ($p$ = 0.0140) (Table 3.2).

**Table 3.2**. Amylase activity before and after protein purification

| Protein | Activity before purification ($nkat.ml^{-1}$) | Specific activity after purification ($nkat.mg^{-1}$protein) | Protein | Activity before purification ($nkat.ml^{-1}$) | Specific activity after purification ($nkat.mg^{-1}$protein) |
|---|---|---|---|---|---|
| AmyA | 50.0 | 395 ± 11 | GlaA | 2.4 | 86 ± 4 |
| AmyA_Opt | 17.8 | 407 ± 11 | GlaA_Opt | 1.3 | 54 ± 11 |
| AmyA_CBI | 17.6 | 369 ± 18 | GlaA_CBI | 1.0 | 27 ± 2 |

The purified α-amylases and glucoamylases were analysed using DSF. Low initial background and good transition was observed for all proteins. The fluorescence curves for the α-amylases were similar (Fig. 3.4a and 3.4c) and peaked at the same temperature (68 - 69°C), while the curves for the glucoamylases (Fig. 3.4b and 3.4d) were shifted and peaked at different temperatures (68, 69 and 72°C). The melting temperature ($T_m$) values for the codon optimised AmyA_Opt and AmyA_CBI, as well as the GlaA_Opt and GlaA_CBI proteins were higher than the native AmyA and GlaA proteins, respectively, indicating a difference in the stability of the proteins. The α-amylases presumably achieved the correct confirmation upon folding, whereas the GlaA_Opt and GlaA_CBI displayed different folding pattern (Fig. 3.4b).

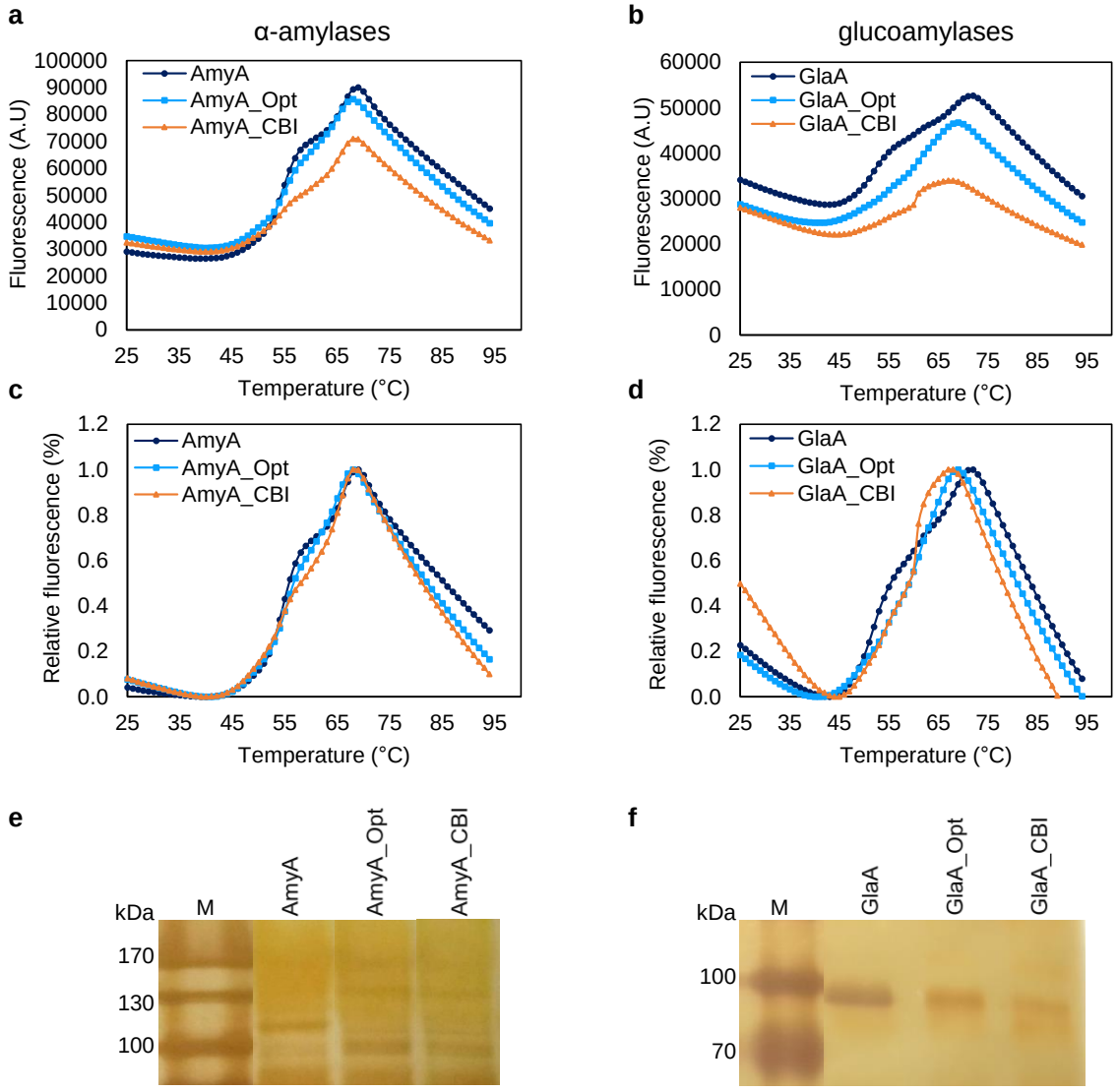


**Fig. 3.4** DSF analysis of purified proteins. The melting curves of (**a**) AmyA, Amy_Opt and AmyA_CBI and (**b**) GlaA, GlaA_Opt and GlaA_CBI proteins. The relative fluorescence curves (**c** and **d**) for α-amylases and glucoamylases. The standard deviations for each sample set were calculated for the 5 repeats. SDS-PAGE was used to visualise (**e** and **f**) the purified α-amylases and glucoamylases. The protein size marker is in lane M and the sizes depicted on the left hand side.

The SDS-PAGE analysis of the purified proteins (Fig. 3.4e and 3.4f) agrees with the trends in specific activity (Table 3.2). Although crude protein concentrations were similar for the glucoamylase proteins before purification, partially purified GlaA had greater specific activity compared to the GlaA_Opt and GlaA_CBI proteins, 1.6-fold and 3.2-fold respectively (Table 3.2). Even after purification, SDS-PAGE analysis indicated a more distinct protein species for the recombinant GlaA protein (approximately 90 kDa) compared to the GlaA_Opt and GlaA_CBI proteins (Fig. 3.4f). GlaA has a calculated molecular weight of 68 kDa. On the other hand, the partially purified protein species for the α-amylase proteins (Fig. 3.4e) had the same intensity, but different degrees of glycosylation. Recombinant AmyA was approximately 110 kDa, while the AmyA_Opt and AmyA_CBI proteins were approximately 100 kDa. AmyA has a calculated molecular weight of 69.6 kDa.

### 3.4.3 Effects of chaperone co-expression and host strain

Different strategies were followed to elucidate the problem with synthetic gene expression. Co-expression of *amyA_Opt* and *glaA_Opt* with the *S. cerevisiae PDI1* gene resulted in an increase in extracellular levels of activity (Fig. 3.5a and 3.5b). At 72 hours, the *S. cerevisiae* Y294[AmyA_Opt & PDI1] strain indicated an increase of 1.32 nkat.ml$^{-1}$ in α-amylase activity ($p = 0.0281$) compared to the *S. cerevisiae* Y294[AmyA_Opt] strain. Similarly, the glucoamylase activity improved 2-fold at 72 hours (Fig. 3.5b). The overexpression of *PDI1* also resulted in an increase in AmyA_Opt and GlaA_Opt crude protein concentrations (Fig. 3.5e and 3.5f).

The *S. cerevisiae* NI-C-D4 and Y294 strains were compared as host strains for the expression of the *amyA_Opt* and *glaA_Opt* genes. The *S. cerevisiae* NI-C-D4[AmyA_Opt] strain grew to a similar cell density, but displayed a higher volumetric activity than the *S. cerevisiae* Y294[AmyA_Opt] strain (Fig. 3.5a and 3.5c). SDS-PAGE analysis did not indicate any significant difference in size or concentration of the recombinant α-amylases secreted by the different strains (Fig. 3.5e). In contrast, the *S. cerevisiae* NI-C-D4[GlaA_Opt] strain grew to a higher biomass concentration, which contributed to the higher levels of activity and increased extracellular glucoamylase concentration (Fig. 3.5b, d and f). The *S. cerevisiae* NI-C-D4[GlaA_Opt] strain also displayed a 3.8-fold higher productivity (2.10 nkat.mg DCW$^{-1}$) (productivity data not shown).

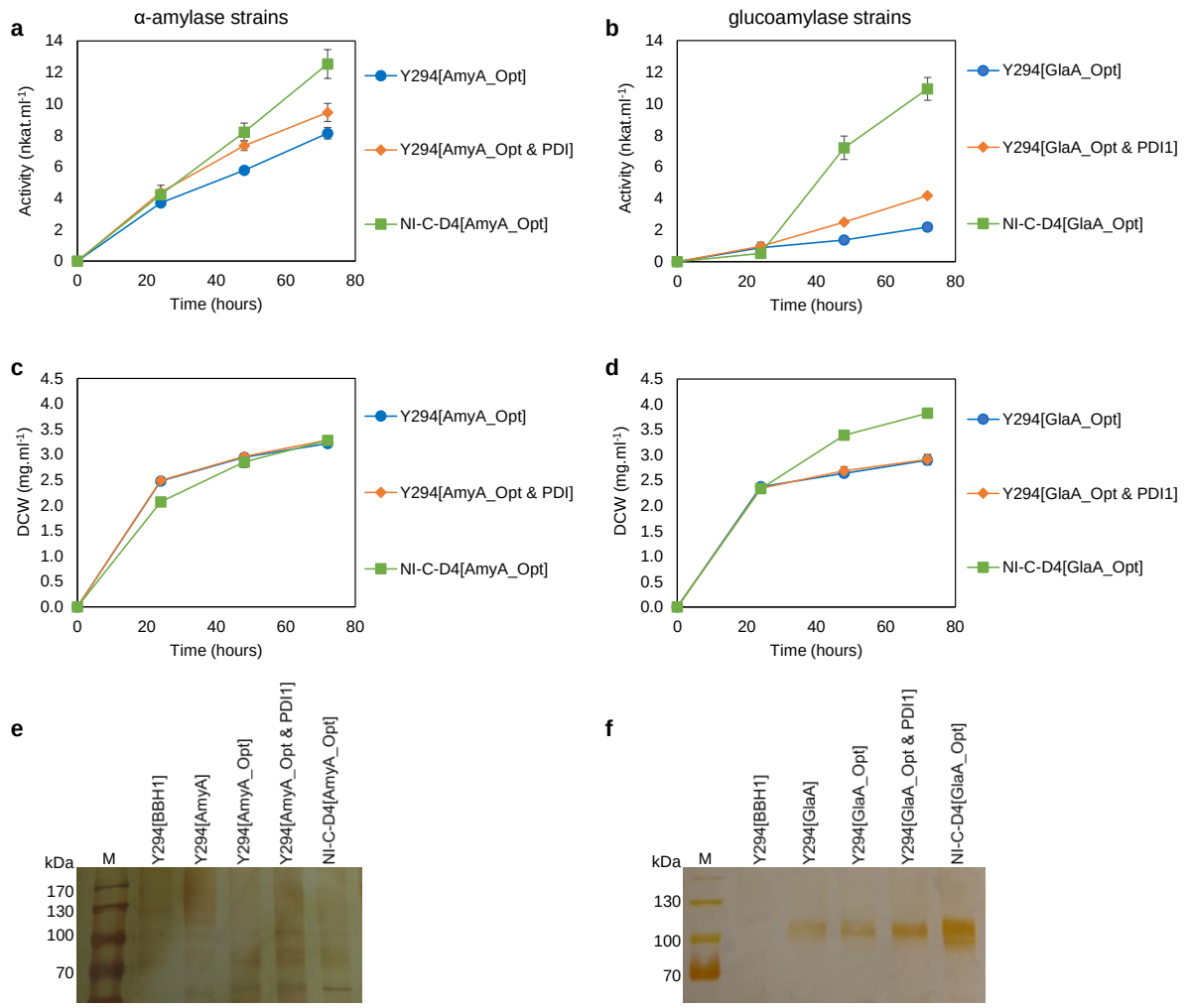


**Fig. 3.5** Effect of *PDI1* co-expression and *S. cerevisiae* NI-C-D4 as the host strain on (**a** and **b**) extracellular α-amylase and glucoamylase activity and DCW (**c** and **d**) for α-amylase and glucoamylase *S. cerevisiae* strains. Error bars represent the standard deviation of three independent biological repeats. The supernatant of *S. cerevisiae* (**e**) α-amylase and (**f**) glucoamylase strains was subjected to SDS-PAGE. The protein size marker is in lane M and the sizes depicted on the left hand side.

### 3.4.4 Effects of codon usage

Two regions on the *amyA_Opt* sequence and one region on *glaA_Opt* sequence were replaced with the corresponding nucleotides from the native *amyA* and *glaA* gene sequences, respectively (Fig. 3.2) and expressed in the *S. cerevisiae* Y294 strain. Replacement with the native *amyA* region I (first 300 base pairs), which included the native secretion signal, resulted in a 29% decrease in the level of α-amylase activity by the *S. cerevisiae* Y294[AmyA_Opt(→300)] strain (Fig. 3.6a). Furthermore, replacement with *amyA* region II (Fig. 3.2a) resulted in no significant difference in the extracellular activity displayed by the *S. cerevisiae* Y292[AmyA_Opt(→426)] strain (10.03 nkat.ml$^{-1}$) at 72 hours, compared to the *S. cerevisiae* Y294[Amy_Opt] strain

(10.27 nkat.ml$^{-1}$) (Fig. 3.6a). However, replacement of both regions (I and II) had a positive effect on the extracellular activity produced by the *S. cerevisiae* Y294[AmyA_Opt(→300+426)] strain and a significant increase in activity (1.92 nkat.ml$^{-1}$) after 72 hours ($p$ = 0.0102). Similar growth profiles (data not shown) were obtained for all strains.

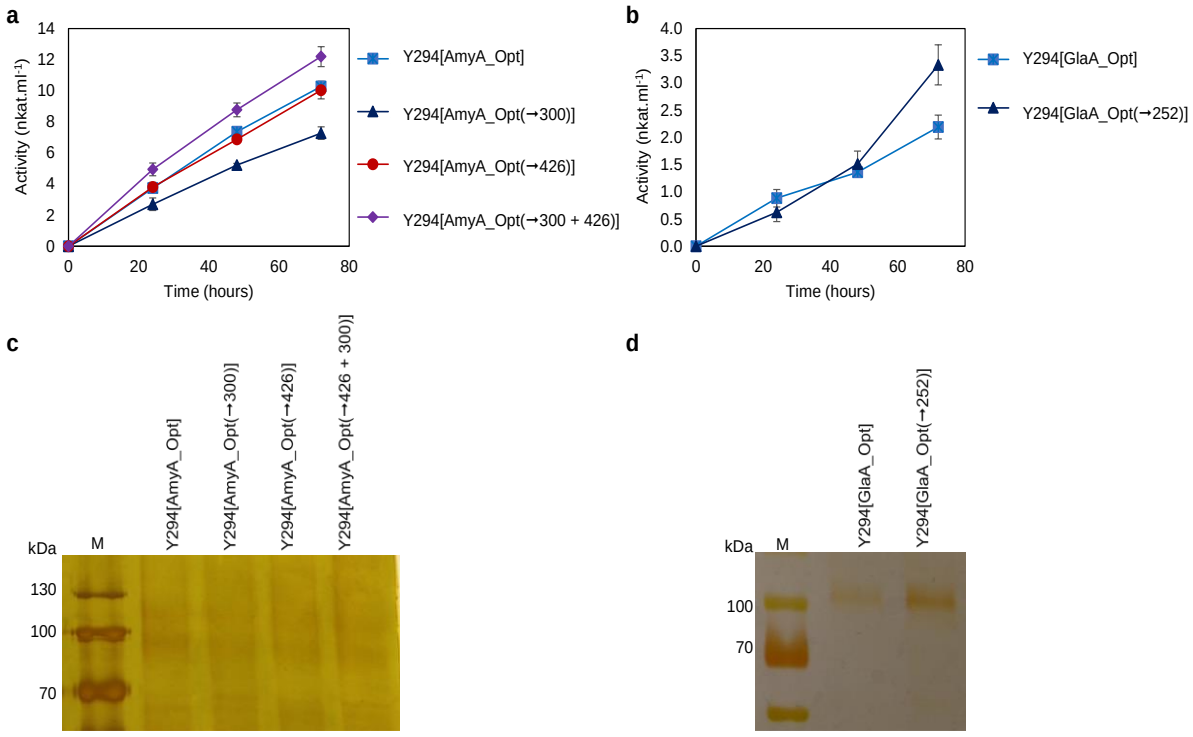


**Fig. 3.6** Effect of synonymous codon substitutions on (**a**) extracellular α-amylase activity, (**b**) extracellular glucoamylase activity, (**c** and **d**) supernatant from *S. cerevisiae* Y294 strains was subjected to SDS-PAGE followed by silver staining. The protein size marker is in lane M and the sizes depicted on the left hand side.

Replacement of part of the codon optimised *glaA_Opt* sequence with the 252-bp fragment from the native *glaA* sequence (Region III) resulted in the construction of *glaA_Opt(*→252) (Fig. 3.2b). A considerable difference in the extracellular recombinant protein concentration was observed for the *S. cerevisiae* Y294[GlaA_Opt(→252)] strain, with SDS-PAGE analysis confirming a more distinct protein species of approximately 110 kDa (Fig. 3.6d). This correlated with a significant increase in extracellular glucoamylase activity of 1.14 nkat.ml$^{-1}$ (after 72 hours) ($p$ = 0.0099), compared to extracellular activity produced by the *S. cerevisiae* Y294[GlaA_Opt] strain (Fig. 3.6b).

## 3.5 Discussion

The native *A. tubingensis* α-amylase (AmyA) and glucoamylase (GlaA) used in this study represented the benchmark α-amylase and glucoamylase, respectively. These enzymes have been shown previously to be effective in the single-step conversion of raw starch to ethanol (Viktor et al. 2013). The co-expression of the native *amyA* and *glaA* genes in an industrial *S. cerevisiae* strain represented significant progress towards developing an amylolytic CBP yeast that could be used in a cold hydrolysis process, without the need for exogenous enzyme addition (Viktor et al. 2013). Therefore, codon optimisation strategies were employed in order to try increase gene expression levels and protein secretion.

The *amyA_Opt* and *glaA_Opt* genes were optimised using a gene design algorithm of a synthetic DNA supplier. The codon optimised *amyA_Opt* and *glaA_Opt* genes had a resulting CBI of 0.27 and 0.28, respectively (Carbone et al. 2003), which were similar to the CBI values of the native *amyA* and *glaA* genes, 0.27 and 0.25, respectively. However, the respective expression of *amyA_Opt* and *glaA_Opt* in *S. cerevisiae* (Fig. 3.3) did not result in improved extracellular activity levels. This highlighted that the CBI strategy and codon optimisation algorithms cannot be relied upon to ensure optimal gene expression and there is clearly additional information “hidden” in the nucleotide sequence necessary for protein production and folding. Since the presence of rare codons can result in low levels of expression (Gustafsson et al. 2004), the *amyA_CBI* and *glaA_CBI* genes were designed based on the “one amino acid - one codon” principle with most abundant codons of the host being used to encode all occurrences of a given amino acid in the optimised sequence. A major drawback to this optimisation approach is that highly transcribed mRNA can result in an imbalance in the tRNA pool (Elena et al. 2014) leading to possible tRNA depletion.

The synthetic and native genes encoded for the same amino acid sequence, yet the strains expressing the native gene sequences produced more extracellular activity at both cultivation temperatures (Fig. 3.3a and 3.3b). The lower cultivation temperature had the biggest impact on the resulting GlaA_CBI extracellular activity (after 96 hours of cultivation). The *glaA_CBI* gene contained no rare codons and therefore lacked potential translational pause sites. This could have lead to improper folding of the protein, changes in glycosylation and an altered functional state (Mitra et al. 2006). Glucoamylases are generally active as dimers (Jørgensen et al. 2008), therefore

small changes in monomer structure or glycosylation patterns could affect dimer formation (McKinnon et al. 2010).

A lower cultivation temperature has been hypothesised to increase the amount of correctly folded protein, since decreased translation rates allow the nascent peptide chains more time to fold correctly (Rosano and Ceccarelli 2009; Hockenberry et al. 2014). Thus it is likely that GlaA_CBI protein structure was improved by a slower rate of folding, leading to the 3.6-fold increase in extracellular activity when the *S. cerevisiae* Y294[GlaA_CBI] strain was cultivated at 20°C (Fig. 3.3b). During the exponential phase, the *S. cerevisiae* Y294[AmyA_CBI] strain grew slower than the other strains at 20°C (Fig. 3.3a), which indicated that heterologous protein secretion might be affecting the metabolic stress of the cell (Ilmén et al. 2011). The lack of translational pause sites may also have influenced proper protein folding leading to futile recycling of the enzyme and decreased protein secretion. All amylases were secreted into the extracellular medium, but the AmyA_Opt, AmyA_CBI, GlaA_Opt and GlaA_CBI were present at lower protein levels compared to the AmyA and GlaA proteins, when crude proteins were separated by SDS-PAGE (Fig. 3e and f). These heterogeneous glycosylation patterns are associated with protein folding, which was affected by the codon optimisation strategies (Fig. 3.3g and 3.3h).

The partially purified AmyA, AmyA_Opt and AmyA_CBI proteins displayed similar specific activities (Table 3.2) despite the differences in sizes (Fig. 3.4e). Therefore, the lower extracellular volumetric activity displayed by the *S. cerevisiae* Y294[AmyA_Opt] and Y294[AmyA_CBI] strains (Fig. 3.3a) is not due to the differences in glycosylation patterns or protein structure (Fig. 3.4a), but is perhaps attributed to decreased recombinant enzyme secretion (Fig. 3.3e). On the other hand, the partially purified glucoamylases showed significantly different specific activities (Table 3.2), but their sizes were similar (similar levels of glycosylation) when separated by SDS-PAGE (Fig. 3.4f). The extracellular enzyme activities at 20°C (Fig. 3.3a and 3.3.b) combined with the increase in protein concentrations (Fig. 3.3e and 3.3f) suggested that at a lower cultivation temperature the proteins were able to fold better thereby reaching their native conformation.

DSF is a high-throughput screening method that can offer insights into protein structure/stability, when studying the unfolding of proteins (Niesen et al. 2007; Vollrath et al. 2014). DSF was subsequently used to confirm that proteins unfold

differently in spite of them having identical amino acid sequences. These results helped validate the differences observed when comparing the activities of the recombinant strains (Fig. 3.3b). The shift in fluorescence curves for the GlaA_CBI and GlaA_Opt proteins suggested that these proteins had an altered tertiary structure compared to that of the native glucoamylase and may explain the catalytic differences observed during the enzymatic assays (Fig. 3.3b). A difference in initial fluorescence was observed for the denaturation profiles of the glucoamylase proteins (Fig. 3.4b), suggesting potential differences in hydrophobicity. Some amino acids are more likely to be buried, while others preferred to be exposed to a solvent (Saunders and Deane 2010), thus leading to structural changes. Therefore, it is speculated that the decreased levels of volumetric activity of the *S. cerevisiae* Y294[GlaA_Opt] and Y294[GlaA_CBI] strains (Fig. 3.3b) was attributed to the lower levels of secreted protein (Fig. 3.3f), as well as possible differences in structures of the GlaA_Opt and GlaA_CBI enzymes (Fig. 3.4b).

In eukaryotic cells, protein disulfide isomerase (PDI1), found in the endoplasmic reticulum (ER), catalyses disulfide bond exchange and assists in the protein folding of newly synthesised proteins (Davis et al. 2000; Parakh and Atkin 2015). It also functions as a chaperone and protects native protein chains from misfolding (Hartl and Hayer-Hartl 2002). Increasing PDI1 activity in bacterial, yeast and insect cell expression systems has been reported to increase secretion and activity of heterologous proteins that contain disulfide bridges (Davis et al. 2000; Smith et al. 2004). The AmyA and GlaA proteins contain 5 and 4 disulphide bonds, respectively (http://scratch.proteomics.ics.uci.edu/). Co-expression of *PDI1* resulted in a significant increase in volumetric activity and enzyme concentration of the AmyA_Opt and GlaA_Opt proteins (Fig. 3.5a and 3.5b). This suggested that overexpression of *PDI1* assisted in protein folding, thereby contributing to increased specific activity (of the crude protein) and streamlining the secretion process. These results supported the hypothesis that the recombinant AmyA_Opt and GlaA_Opt proteins may not fold correctly, when expressed in *S. cerevisiae* Y294.

It is well documented that *S. cerevisiae* hyperglycosylates foreign proteins (Demain and Vaishnav 2009), which can prevent proper dimer assembly (Nørskov-Lauritsen et al. 2015). The *S. cerevisiae* NI-C-D4 strain, on the other hand, is known for its oversecreting nature and low-glycosylation of proteins

(van Wyk et al. 2010). It is also characterised by high enzyme production, which is partly due to its protease deficiency (Wang et al. 2001). It is possible that the protein folding ability of the *S. cerevisiae* NI-C-D4[AmyA_Opt] and NI-C-D4[GlaA_Opt] strains was superior to that of the *S. cerevisiae* Y294[AmyA_Opt] and Y294[GlaA_Opt] strains, resulting in enhanced extracellular levels of amylase activity (Fig. 3.5). These improvements might be due to the increased expression levels of the native chaperone genes, differences in glycosylation patterns or an enhanced secretion capacity.

Synonymous codon bias can affect multiple levels of cellular biology (Hunt et al. 2014) and synonymous codon substitutions at the 5' end of mRNA can have a significant impact on mRNA secondary structure and stability (Wu et al. 2007). The first 30 - 50 codons are responsible for slowing down the initial translation rate to enable efficient binding of the ribosomes to the mRNA (Angov 2011). Therefore, the first 300 nucleotides of the *amyA_Opt* gene were replaced with nucleotides from the native *amyA* gene (Fig. 3.2a – region I). However, this resulted in a significant decrease in extracellular activity of 29% after 72 hours, for the *S. cerevisiae* Y294[AmyA_Opt(→300)] strain (Fig. 3.6a). It is speculated that specific nucleotide changes resulted in an increase in the translational speed, which could have resulted in aggregation of the AmyA_Opt(→300) peptides and decreased protein secretion.

Rare codons in mRNA are associated with translational pausing, which may positively assist in co-translational folding. Rare codons at the start of helices leads to a decline in translation speed observed with the transition from coil to helix and coil to strand (Saunders and Deane 2010). Thus, the DNA and RNA sequences contain vital messages for translational pauses that are imbedded within the protein coding region (Hartl 2011; Makhoul and Trifonov 2002). The algorithm proposed by Carbone et al. (2003) revealed distinctive codon usage patterns before the sequence encoding for the carbohydrate binding module (CBM) of the *amyA* and *glaA* genes (Fig. 3.2a and 3.b), whereas the CBI-optimised and codon optimised sequences lacked regions of either rare or highly expressed codons in the non-defined linker areas before the CBM20.

Region II in the *amyA* gene contained two areas of highly expressed codons encoding for DYITYKNDPIT and ATTSSSSSAAATTS, respectively. As expected, the replacement of region II did not have a significant impact on the extracellular α-amylase activity or enzyme concentration produced by the *S. cerevisiae*

Y294[AmyA_Opt(→426)] strain (Fig. 3.6a and 3.6c), since the highly expressed codons would only increase the translational speed before the CBM and not the entire mRNA. However, the combined replacement of Regions I and II resulted in an 19% increase in extracellular activity (with an accompanied increase in enzyme concentration), after 72 hours of cultivation of the *S. cerevisiae* Y294[AmyA_Opt(→300 + 426)] strain (Fig. 3.6a). These two areas most probably complemented each other by having a positive effect on translation or protein folding.

Region III of the *glaA* gene contained an area of favourable codons (encoding for SSVPG) and an area of rare codons (encoding for GSGGV) before the CBM domain. The increased activity observed by substituting codons from *glaA_Opt* with the corresponding synonymous codons from *glaA* to form *glaA_Opt*(→252) (Fig. 3.6b) suggested that the native codons may have partially restored some of the features that were lost during codon optimisation. The 252 bps from *glaA* are associated with a glycosylated linker region (Sauer et al. 2000). Therefore, it is possible that the native codons helped improve the structure of the linker region, which contributed to the stabilisation of the functional C-terminal starch binding domain (Lin et al. 2007) or resulted in improved translational accuracy (Zhang and Ignatova 2009). However, functional cooperation between amylase domains is not well understood and requires further investigation.

The analysis of how codon bias affects gene expression is a prerequisite for improving gene design algorithms for different expression hosts (Welch et al. 2009). This study highlighted that designer proteins are not always as active as naturally occurring proteins and it appears that underlying translational information is lost when synonymous codons are used. Codon optimisation and codon bias strategies do not always result in improved protein production; fundamental differences in extracellular enzyme activity, protein secretion, glycosylation patterns and DSF analysis were observed is this study.

Combined with the data from other studies, there is a clear pattern that insufficient protein folding limits heterologous expression and that manipulation of certain ER processes may help overcome this bottleneck. The protein folding conditions were improved by changing the cultivation temperature and overexpressing the *PDI1* gene, subsequently leading to enhanced protein secretion and higher levels of extracellular activity for the *S. cerevisiae* Y294 strains expressing the *amyA_Opt* and *glaA_Opt*

genes. Therefore, new tools are needed that aim to predict the functional impact of synonymous codon usage in heterologous gene expression using *S. cerevisiae*.

### 3.6 Compliance with ethical standards

Research did not involve any human participants or animals.

### 3.7 Conflict of interests

The authors declare that they have no competing interests.

### 3.8 Acknowledgements

The authors would like to thank the National Research Foundation (NRF) for financial support to the Chair of Energy Research: (grant number UID 86423 awarded to WHvZ).

### 3.9 References

Angov E (2011) Codon usage: nature's roadmap to expression and folding of proteins. Biotechnol J 6:650–659. doi: 10.1002/biot.201000332

Carbone A, Zinovyev A, Képès F (2003) Codon adaptation index as a measure of dominating codon bias. Bioinformatics 19:2005–2015. doi: 10.1093/bioinformatics/btg272

Cho KM, Yoo YJ, Kang HS (1999) δ-Integration of endo/exoglucanase and β-glucosidase genes into the yeast chromosomes for direct conversion of cellulose to ethanol. Enzyme Microb Technol 25:23–30. doi: doi.org/10.1016/S0141-0229(99)00011-3

Davis R, Schooley K, Rasmussen B, Thomas J, Reddy P (2000) Effect of PDI overexpression on recombinant protein secretion in CHO cells. Biotechnol Progr 16:736–743. doi: 10.1021/bp000107q

de Oliveira APA, Silvestre AM, Alves-Prado HF, Rodrigues A, da Paz MF, Fonseca, GG, Leite RSR (2015) Bioprospecting of yeasts for amylase production in solid state fermentation and evaluation of the catalytic properties of enzymatic extracts. Afr J Biotechnol 14:1215-1223. doi: 10.5897/AJB2014.14062

Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. Biotech Adv 27:297–306. doi: 10.1016/j.biotechadv.2009.01.008

Den Haan R, Rose SH, Lynd LR, van Zyl WH (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. Metab Eng 9:87-94. doi: 10.1016/j.ymben.2006.08.005

Elena C, Ravasi P, Castelli ME, Peirú S, Menzella HG (2014) Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. Front Microbiol 5:21. doi: 10.3389/fmicb.2014.00021

Gustafsson C, Govindarajan S, Minshull J, Park M (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22:346–353. doi: 10.1016/j.tibtech.2004.04.006

Hartl FU, Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295:1852–1858. doi: 10.1126/science.1068408

Hartl FU 2011 Chaperone-assisted protein folding: the path to discovery from a personal perspective. Nat Med 17:1206-1210. doi: 10.1038/nm.2467

Hockenberry AJ, Sirer MI, Amaral LA, Jewett MC (2014) Quantifying position-dependent codon usage bias. Mol Biol Evol 31:1880–1893. doi: 10.1093/molbev/msu126

Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C (2014) Exposing synonymous mutations. Trends Genet 30:308–321. doi: 10.1016/j.tig.2014.04.006

Ilmén M, den Haan R, Brevnova E, McBride J, Wiswall E, Froehlich A, Koivula A, Voutilainen SP, Siika-aho M, la Grange DC, Thorngren N, Ahlgren S, Mellon M, Deleault K, Rajgarhia V, van Zyl WH, Penttilä M (2011) High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*. Biotechnol Biofuels 4:30. doi: 10.1186/1754-6834-4-30

Jørgensen AD, Nøhr J, Kastrup JS, Gajhede M, Sigurskjold BW, Sauer J, Svergun DI, Svensson B, Vestergaard B (2008) Small angle x-ray studies reveal that *Aspergillus niger* glucoamylase has a defined extended conformation and can form dimers in solution. J Biol Chem 283:14772–14780. doi: 10.1074/jbc.M801709200

Laemmli UK (1970) Clearage of structural proteins during the assembly of the head of becteriophage T4. Nature 227:680–685. doi: 10.1038/227680a0

Li S, Yang X, Yang S, Zhu M, Wang X (2012) Technology prospecting on enzymes: application, marketing and engineering. Comput Struct Biotechnol J. 2:e201209017. doi: 10.5936/csbj.201209017

Lin, SC, Liu WT, Liu SH, Chou WI, Hsiung BK, Lin IP, Sheu CC, Chang M (2007) Role of the linker region in the expression of *Rhizopus oryzae* glucoamylase. BMC Biochem 8:9. doi: 10.1186/1471-2091-8-9

Makhoul CH, Trifonov EN (2002) Distribution of rare triplets along mRNA and their relation to protein folding. J Biomol Struct Dyn 20:413–420. doi: 10.1080/07391102.2002.10506859

McKinnon TA, Goode EC, Birdsey GM, Nowak AA, Chan AC, Lane DA, Laffan MA (2010) Specific *N*-linked glycosylation sites modulate synthesis and secretion of von Willebrand factor. Blood 116:640–648. doi: 10.1182/blood-2010-02-267450

Miller GL (1959) Use of dinitrosalicyclic acid reagent for determination of reducing sugars. Anal Chem 31:426–428. doi: 10.1021/ac60147a030

Mitra N, Sinha S, Ramya TN, Surolia A (2006) *N*-linked oligosaccharides as outfitters for glycoprotein folding, form and function. Trends Biochem Sci 31:156–163. doi: 10.1016/j.tibs.2006.01.003

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. doi: 10.1038/nprot.2007.321

Njokweni A, Rose SH, van Zyl WH (2012) Fungal β-glucosidase expression in *Saccharomyces cerevisiae*. J Mol Microbiol Biotechnol 39:1445–1452. doi: 10.1007/s10295-012-1150-9

Nørskov-Lauritsen L, Jørgensen S, Bräuner-Osborne H (2015) *N*-glycosylation and disulfide bonding affects GPRC6A receptor expression, function, and dimerization. FEBS Lett 589:588–597. doi: 10.1016/j.febslet.2015.01.019

O'Connell K, Stults JT (1997) Identification of mouse liver proteins on two-dimensional electrophoresis gels by matrix-assisted laser desorption/ionization mass spectrometry of in situ enzymatic digests. Electrophoresis 18:349–359. doi: 10.1186/1754-6834-6-167

Parakh S, Atkin JD (2015) Novel roles for protein disulphide isomerase in disease states: a double edged sword? Front Cell Dev Biol 3: 30. doi: 10.3389/fcell.2015.00030

Plotkin JB, Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. Nature Rev Genet 12:32–42. doi: 10.1038/nrg2899

Rosano GL. Ceccarelli EA (2009) Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. Microb Cell Fact 8:41. doi: 10.1186/1475-2859-8-41

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press

Sauer J, Sigurskjold BW, Christensen U, Frandsen TP, Mirgorodskaya E, Harrison M, Roepstorff P, Svensson B (2000) Glucoamylase: structure/function relationships, and protein engineering. Biochim Biophys Acta 1543:275–293. doi: 10.1016/S0167-4838(00)00232-6

Saunders R, Deane CM (2010) Synonymous codon usage influences the local protein structure observed. Nucleic Acids Res 38:6719–6728. doi: 10.1093/nar/gkq495

Shah K, Cheng Y, Hahn B, Bridges R, Bradbury NA, Mueller DM (2015) Synonymous codon usage affects the expression of wild type and F508del CFTR. J Mol Biol 427:1464–1479. doi: 10.1016/j.jmb.2015.02.003

Sharp PM, Li WH (1987) The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15:1281–1295. doi: 10.1093/nar/15.3.1281

Smith JD, Tang BC, Robinson AS (2004) Protein disulfide isomerase, but not binding protein, overexpression enhances secretion of a non-disulfide-bonded protein in yeast. Biotechnol Bioeng 85:340–350. doi: 10.1002/bit.10853

Van Wyk N, den Haan R, van Zyl WH (2010) Heterologous production of NpCel6A from *Neocallimastix patriciarum* in *Saccharomyces cerevisiae*. Enzyme Microb Technol 46:378-383. doi: 10.1016/j.enzmictec.2009.11.005

Viktor MJ, Rose SH, van Zyl WH, Viljoen-Bloom M (2013) Raw starch conversion by *Saccharomyces cerevisiae* expressing *Aspergillus tubingensis* amylases. Biotechnol Biofuels 6:167. doi: 10.1186/1754-6834-6-167

Vollrath F, Hawkins N, Porter D, Holland C, Boulet-Audet M (2014) Differential scanning fluorimetry provides high throughput data on silk protein transitions. Sci Rep 4:5625. doi: 10.1038/srep05625

Wang D, Chen C, Kuo T (2001) Characterization of a *Saccharomyces cerevisiae* mutant with oversecretion phenotype. Appl Microbiol Biotechnol 55:712–720. doi: 10.1007/s002530100594

Wasungu KM, Simard RA (1982) Growth characteristics of bakers' yeast in ethanol. Biotechnol Bioeng 24:1125-1134. doi: 10.1002/bit.260240509

Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, Gustafsson C (2009) Design parameters to control synthetic gene expression in *Escherichia coli*. PloS ONE. 4: e7002. doi: 10.1371/journal.pone.0007002

Wu G, Dress L, Freeland SJ (2007) Optimal encoding rules for synthetic genes: the need for a community effort. Mol Syst Biol 3: art.134. doi: 10.1038/msb4100176

Zhang G, Ignatova Z (2009) Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. PLoS ONE 4: e5036. doi: 10.1371/journal.pone.0005036

Zhou WJ, Yang JK, Mao L, Miao LH (2015) Codon optimization, promoter and expression system selection that achieved high-level production of *Yarrowia lipolytica* lipase in *Pichia pastoris*. Enzyme Microb Technol 71:66–72. doi: 10.1016/j.enzmictec.2014.10.007

# Chapter 4

Novel raw starch amylase combinations for the construction of a CBP yeast

# Novel raw starch amylase combinations for the construction of a CBP yeast

Rosemary A. Cripwell, Shaunita H. Rose, Willem H. van Zyl

Department of Microbiology, Stellenbosch University, Private Bag X1, 7602 Matieland, South Africa.

## 4.1 Abstract

Starchy substrates are abundant feedstocks for bioethanol production. Cost-effective utilisation of starch requires consolidated bioprocessing (CBP) whereby a single amylolytic yeast can produce the enzymes required for starch hydrolysis, while simultaneously converting the resultant glucose to ethanol. Novel fungal α-amylase and glucoamylase encoding gene combinations were expressed in the *Saccharomyces cerevisiae* Y294 laboratory strain and evaluated for their starch hydrolysing ability. The recombinant *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain expressing the codon optimised glucoamylase and native α-amylase from *Talaromyces emersonii* produced 51.7 $g.l^{-1}$ ethanol from 200 $g.l^{-1}$ raw corn starch after 120 hours of fermentation, compared to 33.1 $g.l^{-1}$ produced by the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain. The *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain displayed an 85% carbon conversion (based on the percentage starch converted on a mol carbon basis) after 192 hours, compared to the 54% by the benchmark strain. Thus the TemG_Opt and TemA_Nat enzyme combination was efficient in the hydrolysis of raw corn starch. This demonstrated the superior hydrolytic effect of the *T. emersonii* amylases and that the novel enzyme combination shows potential for use in industrial fermentation processes. The single-step conversion of raw corn starch represents significant progress towards the implementation of an amylolytic CBP yeast and the elimination of the heat pretreatment of starch.

**Keywords:** • native genes • codon optimised • amylases • CBP • raw corn starch

## 4.2 Introduction

Cost effective, renewable and sustainable energy is a global concern, which has increased investigations into alternative fuel sources. Starch is an industrially important substrate for the production of biofuel, proteins and chemicals

(Li et al. 2011; Naguleswaran et al. 2013). It is produced by plants as an energy store and consists of α-1,4 linked glucose units with α-1,6 branching points. The amylose and amylopectin polymers are densely packed in starch granules forming a semi-crystalline structure with inter- and intra-molecular bonds. A combination of α-amylase and glucoamylase enzymes is required for the complete hydrolysis of starch (Białas et al. 2014). Amylases are widely distributed in nature (Jeang et al. 2002) and the vast availability of starch utilising microbes makes them the preferred source of amylolytic enzymes (Rana et al. 2013).

Starch granules are insoluble in cold water and are often resistant to enzymatic hydrolysis (Uthumporn et al. 2010). Therefore, the conventional process for starch conversion involves two hydrolysis steps. During liquefaction, thermostable α-amylase are added and starch is heated to about 100°C until it is gelatinised. This is followed by the addition of glucoamylase for saccharification, which involves the conversion of the liquefied starch into glucose as final product (Presečki et al. 2013). This traditional starch conversion process carries high operational costs (about 10 - 20% of the fuel value of ethanol produced) resulting from the large amount of energy required for the heating process (Cinelli et al. 2015).

Consolidated bioprocessing (CBP) is still in the early stages of development, but it offers a promising approach for cost-effective biofuel production from starchy biomass (Salehi Jouzani and Taherzadeh 2015). CBP systems use a single organism that is able to perform the liquefaction and hydrolysis of starch, as well as ferment the resulting sugars to ethanol at low temperatures i.e. cold hydrolysis (Wong et al. 2007). The cold process requires amylases that have the ability to digest raw starch efficiently at yeast fermentation conditions (i.e. pH and temperature), in order to eliminate the heating requirement (Jeang et al. 2002; Białas et al. 2014). However, the main disadvantages to cold starch hydrolysis are high enzyme loadings and the cost of commercial enzymes e.g. STARGEN™ (Dupont-Danisco, Itasca, USA).

Few raw starch hydrolysing amylases have been reported to date (Mamo and Gessesse 1999; Robertson et al. 2006; Celińska et al. 2015). They differ from conventional amylases in their affinity and interaction with the microcrystalline structures of starch granules. A starch binding domain (SBD) is a key characteristic of these enzymes and enables them to bind effectively to the surface of raw starch granules. The yeast *S. cerevisiae* is an efficient ethanol producer, but it lacks the ability

to degrade starch (Yamada et al. 2011). Raw starch amylase encoding genes from *Lipomyces kononenkoae* and *Saccharomycopsis fibuligera* (Eksteen et al. 2003; Knox et al. 2004), *Rhizopus arrhizus* (Yang et al. 2011) and *Aspergillus tubingensis* (Viktor et al. 2013) are among the few that have been expressed in *S. cerevisiae*. However, current amylase production by yeast does not yet support efficient conversion of raw starch to ethanol in a single step required for CBP on a commercial scale (Görgens et al. 2015).

A CBP yeast for raw starch industrial application without the addition of commercial amylases remains elusive. In this study, α-amylases and glucoamylases from *Aspergillus terreus*, *Aureobasidium pullulans*, *Chaetomium thermophilum*, *Humicola grisea*, *Neosartorya fischeri*, *Rhizomucor pusillus*, *Talaromyces emersonii*, *Talaromyces stipitatus* and *Thermomyces lanuginosus* were screened for activity on starch and compared to the *S. cerevisiae* Y294[AmyA] and Y294[GlaA] benchmark strains, respectively (Viktor et al. 2013). Thereafter, several different amylolytic *S. cerevisiae* Y294 strains were constructed and compared to the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain (Viktor et al. 2013) and evaluated for their ability to hydrolyse raw corn starch and ferment the resulting glucose to ethanol at a high substrate loading (200 $g.l^{-1}$ raw corn starch).

## 4.3 Materials and methods

### 4.3.1 Media and cultivation conditions

All chemicals were of analytical grade and were obtained from Merck (Darmstadt, Germany), unless otherwise stated. *Escherichia coli* DH5α (Takara Bio Inc.) was used for vector propagation. The *E. coli* transformants were selected for on Luria Bertani agar (Sigma-Aldrich, Steinheim, Germany), containing 100 $\mu g.ml^{-1}$ ampicillin and cultivated at 37°C in Terrific Broth (12 $g.l^{-1}$ tryptone, 24 $g.l^{-1}$ yeast extract, 4 $ml.l^{-1}$ glycerol, 0.1 M potassium phosphate buffer) containing 100 $\mu g.ml^{-1}$ ampicillin for selective pressure (Sambrook et al. 1989).

The *S. cerevisiae* Y294 strain was maintained on YPD plates (10 $g.l^{-1}$ yeast extract, 20 $g.l^{-1}$ peptone and 20 $g.l^{-1}$ glucose and 15 $g.l^{-1}$ agar ) and amylolytic transformants were selected and maintained on $SC^{-URA}$ plates (containing 6.7 $g.l^{-1}$ yeast nitrogen base without amino acids (BD-Diagnostic Systems, Maryland, USA), 20 $g.l^{-1}$ glucose, 1.5 $g.l^{-1}$ yeast synthetic drop-out medium supplements (Sigma-Aldrich, Germany),

2% corn starch (Sigma-Aldrich, Germany) and 15 $g.l^{-1}$ agar). The *S. cerevisiae* strains were aerobically cultivated on a rotary shaker (200 rpm) at 30°C, in 125 ml Erlenmeyer flasks containing 20 ml double strength $SC^{-URA}$ medium ($2\times SC^{-URA}$ containing 13.4 $g.l^{-1}$ yeast nitrogen base without amino acids (BD-Diagnostic Systems), 20 $g.l^{-1}$ glucose and 3 $g.l^{-1}$ yeast synthetic drop-out medium supplements (Sigma-Aldrich, Germany). All cultures were inoculated to a concentration of $1\times10^6$ $cells.ml^{-1}$.

### 4.3.2 Strains and plasmids

The genotypes of the bacterial and fungal strains, as well as the plasmids used in this study are summarised in Table 4.1.

### 4.3.3 DNA manipulations

Standard protocols were followed for all DNA manipulations and *E. coli* transformations (Sambrook et al. 1989). All genes were synthesised by GenScript (Piscataway, New Jersey, USA), based on the nucleotide accession numbers listed under **section 4.3.4**. The internal *Eco*RI, *Xho*I, *Bam*HI and *Bgl*II restriction sites were avoided, but the amino acid sequence remained unaffected. The polymerase chain reaction (PCR) was performed using a Perkin Elmer Gene Amp® PCR System 2400 and TaKaRa Ex Taq™ (Takara Bio Inc, Japan) as per the manufacturer's recommendations. The amylase genes were amplified using primers (Inqaba Biotec, South Africa) (Table 4.2) designed for yeast mediated ligation (YML) and visualised on a 0.8% agarose gel. DNA was eluted from agarose gels with the Zymoclean™ Gel Recovery Kit (Zymo Research, California, USA).

The amylase genes were subcloned individually onto the yBBH1 or yBBH4 plasmid (Fig. 4.1a, b and c) in order to construct the expression vectors listed in Table 4.1. The yBBH4 vector (Fig. 4.1c) contained the sequence encoding for the *XYNSEC* secretion signal of the *Trichoderma reesei xyn2* (Den Haan et al. 2007) for directing the secretion of the amylases. The *ENO1*$_P$-α-amylase-*ENO1*$_T$ cassettes were amplified from the yBBH1-α-amylase vectors using YML cassette primers: ENOCASS-L: gtgcggtatttcacaccgcataggagatcgatcccaattaatgtgagttacctcactc and ENOCASS-R: cgggcctcttcgctattacgccagagcttagatct and cloned on the *Bgl*II site of yBBH1-glucoamylase or yBBH4-glucoamylase vectors (Fig. 4.1c and 4.1d). Sequence verification of the final vector constructs was performed by the dideoxy chain termination method, with an ABI PRISM™ 3100 Genetic Analyser (CAF, Stellenbosch University).

**Table 4.1**. Strains and plasmids used in this study

| Strains and plasmids | Genotype | Reference |
|---|---|---|
| *E. coli DH5α* | *supE44 ΔlacU169 (ϕ80lacZ*ΔM15) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1* | Sambrook et al. (1989) |
| ***S. cerevisiae* strains** | | |
| Y294 | *α leu2*-3,112 *ura3*-52 *his3 trp1-289* | ATCC 201160 |
| Y294[BBH1] | *URA3 ENO1*$_P$-*ENO1*$_T$ | Viktor et al. (2013) |
| Y294[AmyA][1] | *URA3 ENO1*$_P$-*amyA-ENO1*$_T$ | Viktor et al. (2013) |
| Y294[GlaA][1] | *URA3 ENO1*$_P$-*glaA-ENO1*$_T$ | Viktor et al. (2013) |
| Y294[AmyA-GlaA][1] | *URA3 ENO1*$_P$-*amyA-ENO1*$_T$; *ENO1*$_P$-*glaA-ENO1*$_T$ | Viktor et al. (2013) |
| Y294[ApuA_Nat][1] | *URA3 ENO1*$_P$-*apuA_Nat-ENO1*$_T$ | This study |
| Y294[ApuA_Opt-NatSS][1] | *URA3 ENO1*$_P$-*NatSS-apuA_Opt-ENO1*$_T$ | This study |
| Y294[ApuA_Nat-XYNSEC] | *URA3 ENO1*$_P$-*XYNSEC-apuA_Nat-ENO1*$_T$ | This laboratory |
| Y294[ApuA_Opt-XYNSEC] | *URA3 ENO1*$_P$-Opt*XYNSEC-apuA_Opt-ENO1*$_T$ | This study |
| Y294[ApuA_Opt-OptXYNSEC] | *URA3 ENO1*$_P$-Opt*XYNSEC-apuA_Opt-ENO1*$_T$ | This study |
| Y294[AteA_Nat] [1] | *URA3 ENO1*$_P$-*ateA_Nat-ENO1*$_T$ | This study |
| Y294[AteA_Nat-XYNSEC] | *URA3 ENO1*$_P$-*XYNSEC-ateA_Nat-ENO1*$_T$ | This laboratory |
| Y294[TemA_Nat][1] | *URA3 ENO1*$_P$-*temA_Nat-ENO1*$_T$ | This study |
| Y294[TemA_Opt] | *URA3 ENO1*$_P$-*temA_Opt-ENO1*$_T$ | This study |
| Y294[TemA_Opt-XYNSEC] | *URA3 ENO1*$_P$-*XYNSEC-temA_Opt-ENO1*$_T$ | This study |
| Y294[TemA_Nat- XYNSEC] | *URA3 ENO1*$_P$-*XYNSEC-temA_Nat-ENO1*$_T$ | This study |
| Y294[TemA_Opt-NatSS][1] | *URA3 ENO1*$_P$-*NatSS-temA_Opt-ENO1*$_T$ | This study |
| Y294[AteG_Nat][1] | *URA3 ENO1*$_P$-*ateG_Nat-ENO1*$_T$ | This study |
| Y294[AteG_Nat-XYNSEC] | *URA3 ENO1*$_P$-*XYNSEC-ateG_Nat-ENO1*$_T$ | This study |
| Y294[AteG_Opt-XYNSEC] | *URA3 ENO1*$_P$-*XYNSEC-ateG_Opt-ENO1*$_T$ | This study |
| Y294[AteG_Opt-NatSS][1] | *URA3 ENO1*$_P$-*NatSS-ateG_opt-ENO1*$_T$ | This study |
| Y294[TemG_Nat][1] | *URA3 ENO1*$_P$-*temG_Nat-ENO1*$_T$ | This study |
| Y294[TemG_Opt] | *URA3 ENO1*$_P$-*temG_Opt-ENO1*$_T$ | This study |
| Y294[TemG_Opt-XYNSEC][1] | *URA3 ENO1*$_P$-*XYNSEC-temG_Opt-ENO1*$_T$ | This study |
| Y294[TemG_Nat-XYNSEC][1] | *URA3 ENO1*$_P$-*XYNSEC-temG_Nat-ENO1*$_T$ | This study |
| Y294[TemG_Opt-NatSS][1] | *URA3 ENO1*$_P$-*NatSS-temG_Opt-ENO1*$_T$ | This study |
| Y294[TemG_Opt-AmyA] | *URA3 ENO1*$_P$-*temG_Opt-ENO1*$_T$; *ENO1*$_P$-*amyA-ENO1*$_T$ | This study |
| Y294[TemG_Opt-TemA_Nat] | *URA3 ENO1*$_P$-*temG_Opt-ENO1*$_T$; *ENO1*$_P$-*temA_Nat-ENO1*$_T$ | This study |
| Y294[TemG_Opt-TemA_Opt] | *URA3 ENO1*$_P$-*temG_Opt-ENO1*$_T$; *ENO1*$_P$-*temA_Opt-ENO1*$_T$ | This study |
| Y294[TemG_Opt-AteA_Nat] | *URA3 ENO1*$_P$-*temG_Opt-ENO1*$_T$; *ENO1*$_P$-*ateA_Nat-ENO1*$_T$ | This study |

**Table 4.1**. Strains and plasmids used in this study continued

| Strains and plasmids | Genotype | Reference |
|---|---|---|
| Y294[TemG_Opt-ApuA_Nat] | *URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-apuA_Nat-ENO1*$_T$ | This study |
| Y294[GlaA_Nat-TemA_Nat] | *URA3 ENO1*$_P$*-glaA-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| Y294[TemG_Nat-AmyA] | *URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-amyA-ENO1*$_T$ | This study |
| Y294[TemG_Nat-AteA_Nat] | *URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-ateA_Nat-ENO1*$_T$ | This study |
| Y294[TemG_Nat-ApuA_Nat] | *URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-apuA_Nat-ENO1*$_T$ | This study |
| Y294[AteG_Nat-XYNSEC-AmyA] | *URA3 ENO1*$_P$*-ateG_Nat-ENO1*$_T$; *ENO1*$_P$*-amyA-ENO1*$_T$ | This study |
| **Plasmids** | | |
| yBBH1 | *bla URA3 ENO1*$_P$*-ENO1*$_T$ | Njokweni et al. (2012) |
| yBBH4 | *bla URA3 ENO1*$_P$*-XYNSEC-ENO1*$_T$ | Njokweni et al. (2012) |
| yBBH1-AmyA | *bla URA3 ENO1*$_P$*-amyA-ENO1*$_T$ | Viktor et al. (2013) |
| yBBH1-GlaA | *bla URA3 ENO1*$_P$*-glaA-ENO1*$_T$ | Viktor et al. (2013) |
| yBBH1-AteA_Nat | *bla URA3 ENO1*$_P$*-ateA_Nat-ENO1*$_T$ | This study |
| yBBH1-ApuA_Nat | *bla URA3 ENO1*$_P$*-apuA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemA_Nat | *bla URA3 ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemA_Opt | *bla URA3 ENO1*$_P$*-temA_Opt-ENO1*$_T$ | This study |
| yBBH1-AteG_Nat-XYNSEC | *bla URA3 ENO1*$_P$*-ateG_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Nat | *bla URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$ | This study |
| yBBH1-TemG_Nat-ApuA_Nat | *bla URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-apuA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Nat-AmyA | *bla URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-amyA-ENO1*$_T$ | This study |
| yBBH1-TemG_Nat-AteA_Nat | *bla URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-ateA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt-ApuA_Nat | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-apuA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt-AmyA | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-amyA-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt-AteA_Nat | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-ateA-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt-TemA_Nat | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt-TemA_Opt | *bla URA3 ENO1*$_P$*-temG_Opt -ENO1*$_T$; *ENO1*$_P$*-temA_Opt-ENO1*$_T$ | This study |
| yBBH1-GlaA-TemA_Nat | *bla URA3 ENO1*$_P$*-glaA-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| yBBH4-AteG_Nat-XYNSEC-AmyA | *bla URA3 ENO1*$_P$*-XYNSEC-ateG_Nat-ENO1*$_T$; *ENO1*$_P$*-amyA-ENO1*$_T$ | This study |

[1] native secretion signal

_Nat: native coding sequence; _Opt: codon optimised coding sequences (GenScript, USA); -NatSS: native secretion signal; -XYNSEC: native secretion signal from *Trichoderma reesei Xyn2* gene; -OptXYNSEC: codon optimised-XYNSEC secretion signal

### 4.3.4 Amylase genes and Genbank accession numbers

The following amylases were cloned and expressed in *S. cerevisiae* Y294. The native glucoamylases from *A. pullulans (*Accession no. HM246718), *A. terreus* (Accession no. XP_001213553), *H. grisea (*Accession no. M89475), *T. emersonii* (Accession no. AJ304803) and *T. lanuginosus (*Accession no. EF545003), as well as the native α-amylases from *A. pullulans* (Accession no. AEH03024), *A. terreus* (Accession no. XM_001209405), *N. fischeri* (Accession no. XP_001265628), *R. pusillus* (Accession no. AGJ52081) and *T. emersonii* (Accession no. XM_013469492). Coding sequences for the glucoamylases from *C. thermophilum* (Accession no. ABD96025), *T. stipitatus* (Accession no. XP_002484948), *A. terreus* and *T. emersonii,* as well for α-amylases from *A. pullulans* and *T. emersonii* were codon optimised for expression in *S. cerevisiae* (GenScript, Piscataway, New Jersey, USA). *T. emersonii* has recently been classified as *Rasamsonia emersonii* (Houbraken et al. 2012), but will be referred to by its original name in this study.

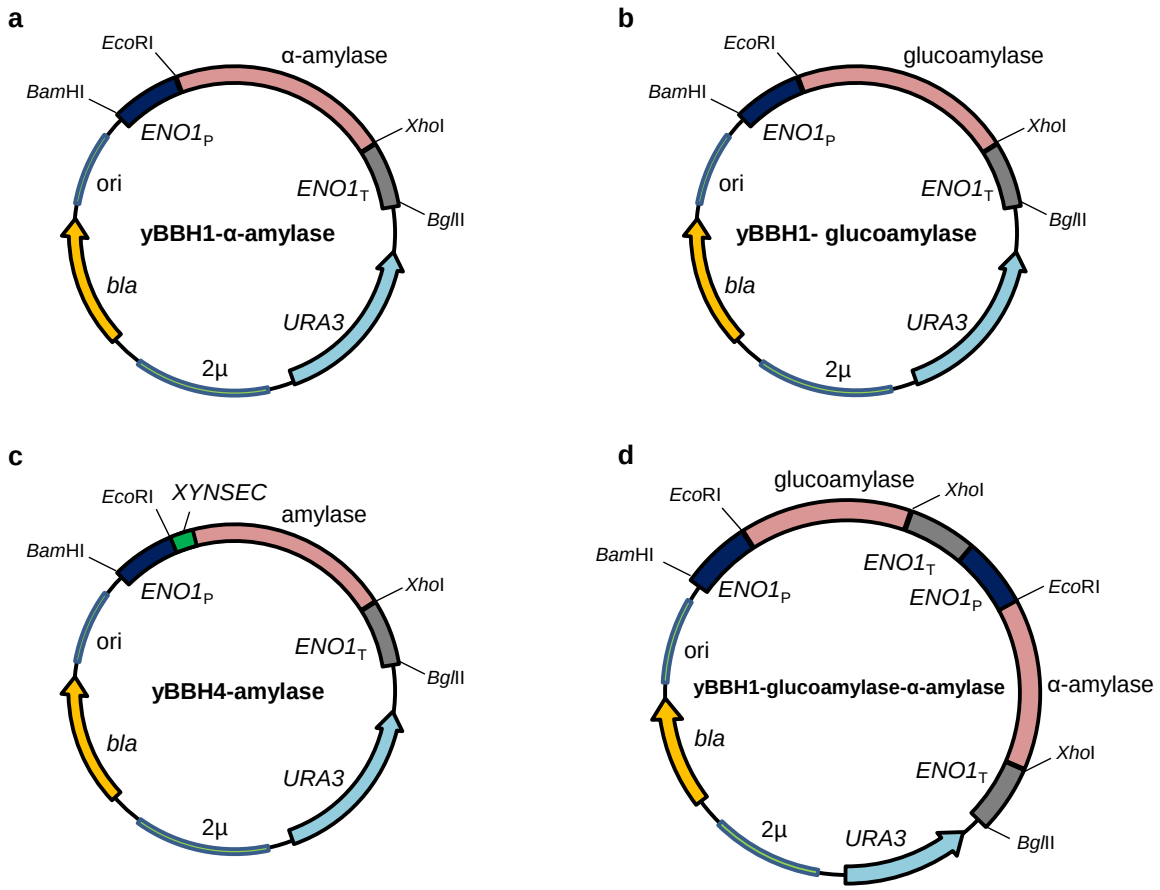


**Fig. 4.1** Schematic representation of the final vector constructs used in this study. Amylase encoding genes were amplified using PCR and respectively cloned onto the yBBH1 and yBBH4 vectors (**a**, **b** and **c**). The *ENO1*$_P$-α-amylases-*ENO1*$_T$ cassettes were cloned onto the yBBH1-glucoamylase plasmids (**d**), to enable co-expression of the genes. *Bam*HI and *Bgl*II restriction enzyme sites were used for YML.

**Table 4.2**. PCR oligo-primers used in this study with the relevant restriction sites underlined (*Eco*RI = gaattc; *Nru*I = tcgcga; *Xho*I = ctcgag)

| Gene name (host organism) | Sequence (5'-3') | Signal peptide[1] |
|---|---|---|
| *apuA* (*A. pullulans)* | ApuA_Nat-L: tgcttatcaacacacaaacactaaatcaaagaattcatggcagccaactacgtttctcgattgttg<br>ApuA_N-R:gactagaaggcttaatcaaaagctctcgagtcacccctgccaagtattgctgaccgatgc<br>ApuA_Opt-NatSS-L: tctctacttgaccgggttggtgcagtgtttgactccagctcaatggagaagtcaatctat<br>ApuA_Opt-R: ggactagaaggcttaatcaaaagctctcgagctaaccttgccatgtattggagactgagg<br>ApuA_optXynSec-L: gaacccgtggctgtggagaagcgctcgcgattgactccagctcaatggagaagtc<br>ApuA_Opt-R: ggactagaaggcttaatcaaaagctctcgagctaaccttgccatgtattggagactgagg | 22 |
| *ateA* (*A. terreus)* | AteA_Nat-L: tgcttatcaacacacaaacactaaatcaaagaattcatgaagtggacctcctcgctcctcctcta<br>AteA_Nat-R: gactagaaggcttaatcaaaagctctcgagtcacctccaagtatcagcaactgtcaccgt | 20 |
| *temA* (*T. emersonii)* | TemA_Nat-L: tgcttatcaacacacaaacactaaatcaaagaattcatgacgcctttcgtcctcacggcc<br>TemA_Nat-R ggactagaaggcttaatcaaaagctctcgagctatctccatgtgtcgacaatcgtctccg<br>TemA_Opt-NatOptSS-L: tgcttatcaacacacaaacactaaatcaaagaattcatgacccctttgttttgacagcc<br>TemA_Opt-R: ggactagaaggcttaatcaaaagctctcgagctatctccaagtgtcaacaatagtttcag<br>TemA_Nat-xynsecSS-L: gaacccgtggctgtggagaagcgctcgcgattgaccccggccgaatggcgcaaacaat<br>TemA_Opt-xynsecSS-L gaacccgtggctgtggagaagcgctcgcgattgacaccagccgaatggagaaagcaatc<br>TemA_Opt-NatSS-L: tcttgctggggaatgccgtgttggccttgacaccagccgaatggagaaagc | 19 |
| *ateG* (*A. terreus)* | AteG_Nat-L: tgcttatcaacacacaaacactaaatcaaagaattcatgacgcgcattctcaccctcgcccttcat<br>AteG_Nat-R: ggactagaaggcttaatcaaaagctctcgagctagcgccaagtggtgttcaccaccgcggt<br>AteG_Opt-NatSS-L: gggctggctcttgtccaaagtgttgttggggcaccacaattggctcctagagcaactaca<br>AteG_Opt-R: tggactagaaggcttaatcaaaagctctcgagctatctccaggttgtgttgacaacggcg<br>AteG_Nat-xynSS-L: gaacccgtggctgtggagaagcgctcgcgagctccccaattggcccccagagcgacaacc | 20 |
| *temG* (*T. emersonii)* | TemG_Nat-L: tgcttatcaacacacaaacactaaatcaaagaattcatggcgtccctcgttgctggcgctctctgc<br>TemG_Nat-R: ggactagaaggcttaatcaaaagctctcgagtcactgccaactatcgtcaagaatggcggt<br>TemG_Nat-xynsecSS-L:gaacccgtggctgtggagaagcgctcgcgacgagcgcccgttgcagcgcgagccaccggt<br>TemG_Opt-xynsecSS-L: gaacccgtggctgtggagaagcgctcgcgaagagccccagtcgcagccagagcaacaggt<br>TemG_Opt-R: gactagaaggcttaatcaaaagctctcgagtcattgccaagagtcgtccaagattgcggt<br>TemG_Opt-NatOptSS-L: ttatcaacacacaaacactaaatcaaagaattcatggcctccttagtcgcaggtgcctta<br>TemG_Opt-NatSS-L: atcctgggcctgacgcctgctgcatttgcaagagccccagtcgcagccagagcaacaggt | 20 |

[1]The length (amino acids) of putative signal peptides was analysed using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP).

### 4.3.5 Yeast transformations

The *S. cerevisiae* Y294 strain was grown overnight in 5 ml YPD broth and prepared according to Cho et al. (1999). After electroporation, 1 ml of YPDS was immediately added to the cuvette. Cultures were incubated at 30°C for 1 hour prior to plating out onto $SC^{-URA}$ plates containing 2% starch. Plates were incubated at 30°C for 2 - 3 days and then transferred to 4°C for 24 hours to allow the starch to precipitate.

### 4.3.6 Activity assays

For quantitative assays, yeast transformants were cultured in 20 ml 2×$SC^{-URA}$ medium in 125 ml Erlenmeyer flasks with agitation at 200 rpm and sampling at 24 hour intervals. The supernatant was harvested and extracellular enzymatic activity levels were assessed colourimetrically (xMark™ Microplate Spectrophotometre, Bio-Rad, San Francisco, USA) using the reducing sugar assay with glucose as standard (Miller 1959). The α-amylase activities were determined after a 5 minute incubation with 0.2% soluble corn starch in 0.05 M citrate-acid buffer (pH 5) at 37°C.

Glucoamylase activity was determined by incubating 50 µl supernatant with 450 µl of 0.2% soluble corn starch in 0.05 M citrate-acid buffer (pH 5) at 37°C for 15 minutes. The glucose concentration was determined using the D-Glucose Assay Kit (Megazyme, Wicklow, Ireland) with absorbance measured at 510 nm (xMark™ Microplate Spectrophotometre, Bio-Rad, San Francisco, USA). Enzymatic activities were expressed as nano-katals per ml ($nkat.ml^{-1}$), with nkat defined as the enzyme activity needed to produce 1 nmol of glucose per second under the described assay conditions.

### 4.3.7 Protein analysis

The recombinant *S. cerevisiae* Y294 strains were cultivated in 125 ml Erlenmeyer flasks containing 20 ml 2×$SC^{-URA}$ medium for 3 days. Twenty microliters of supernatant was added to protein loading buffer and the samples boiled for 3 minutes to denature the proteins. The recombinant proteins were separated using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Tris-glycine buffer (Sambrook et al. 1989). Electrophoresis was carried out at 100 V for ± 90 minutes at ambient temperature and protein species were visualised using the silver staining method (O'Connell and Stults 1997). The broad-range Page Ruler Prestained SM0671 Protein Ladder (Fermentas, China) was used as a molecular mass marker.

### 4.3.8 Fermentation studies

Precultures were cultured in 60 ml 2×SC$^{-URA}$ media in 250 ml Erlenmeyer flasks and incubated at 30°C with agitation of 200 rpm. Fermentations were performed with 2×SC$^{-URA}$ media containing 200 g.l$^{-1}$ raw corn starch and 5 g.l$^{-1}$ glucose and inoculated with a 10% (v.v$^{-1}$) inoculum. Ampicillin (100 µg.ml$^{-1}$) and streptomycin (50 µg.ml$^{-1}$) were added to inhibit bacterial contamination. Agitation and incubation were performed on a magnetic multi-stirrer at 30°C, with daily sampling through a syringe needle pierced through the rubber stopper.

For bioreactor experiments, precultures were cultivated in 120 ml 2×SC$^{-URA}$ media in 500 ml Erlenmeyer flasks at 30°C with agitation at 200 rpm. Bioreactor fermentations were performed in a 2 litre MultiGen Bioreactor (New Brunswick Scientific Corporation, New Jersey, USA) containing 2×SC$^{-URA}$ media supplemented with 200 g.l$^{-1}$ raw corn starch and 5 g.l$^{-1}$ glucose as carbohydrate source. A 10% (v.v$^{-1}$) inoculum was used in a total working volume of 1 litre. Fermentations were carried out at 26°C and 30°C with stirring at 300 rpm and daily sampling through a designated sampling port. All fermentation experiments were performed in triplicate.

#### *4.3.8.1 High performance liquid chromatography analysis*

Ethanol, glucose, maltose, glycerol and acetic acid concentrations were quantified with high performance liquid chromatography (HPLC) using a Surveyor Plus liquid chromatograph (Thermo Scientific) consisting of a liquid chromatography pump, autosampler and refractive index (RI) detector. The compounds were separated on a Rezex RHM Monosaccharide 7.8×300 mm column (00H0132-K0, Phenomenex) at 60°C with 5 mM $H_2SO_4$ as mobile phase at a flow rate of 0.6 ml.min$^{-1}$.

#### *4.3.8.2 Analytical methods and calculations*

The theoretical $CO_2$ concentrations were calculated according to Favaro et al. (2015). The glucose equivalent is defined as the mass of glucose resulting from the complete hydrolysis of starch, i.e. 1.11 grams of glucose per gram of starch. The available carbon (mol C in 100% hydrolysed substrate) was calculated based on the available glucose equivalents and the carbon conversion is defined as the percentage starch converted on a mol carbon basis. This carbon conversion was calculated from ethanol, glucose, maltose, glycerol, acetic acid and $CO_2$ concentrations. The ethanol yield (% of the theoretical yield) was calculated as the amount of ethanol produced per gram of consumed glucose. The ethanol productivity was calculated based on ethanol concentration produced per hour ($g.l^{-1}.h^{-1}$).

### 4.3.9 Statistical analysis

Data was analysed using the Student's t-test.

## 4.4 Results

### 4.4.1 Functional expression of recombinant amylases

The *S. cerevisiae* Y294 strain was used as host for the heterologous gene expression of recombinant amylases. Recombinant strains were constructed to express either an α-amylase or glucoamylase encoding gene (Table 4.1) and evaluated for their ability to hydrolyse corn starch using the *S. cerevisiae* Y294[AmyA] and Y294[GlaA] strains, respectively, as benchmarks strains (Viktor et al. 2013). All of the recombinant *S. cerevisiae* Y294 strains evaluated in this study were able to hydrolyse soluble starch (demonstrated by zones of hydrolysis during plate assays – data not shown). However, several amylase candidates showed significantly lower levels of extracellular activity ($nkat.ml^{-1}$), when compared to the benchmark *S. cerevisiae* Y294 strains expressing the *amyA* and *glaA* genes (data no shown). Thus, the following genes were omitted from further evaluation: native glucoamylases from *A. pullulans, H. grisea* and *T. lanuginosus,* as well as the codon optimised α-amylases from *N. fischeri, R. pusillus* and codon optimised glucoamylases from *C. thermophilum* and *T. stipitatus.* The different gene variants for the *ateA*, *apuA*, *temA*, *ateG* and *temG* genes contained different DNA sequences, but encoded for the same amino acid sequence (for the mature protein).

#### *4.4.1.1 α-Amylases*

The *ateA_Nat* gene was efficiently expressed by the *S. cerevisiae* Y294[AteA_Nat] strain, but the extracellular levels of activity were consistently lower than that of the *S. cerevisiae* Y294[AmyA] benchmark strain (Fig. 4.2a). Replacing the native secretion signal with the native *XYNSEC* did not result in significant differences in either extracellular activity or the amount of AteA crude protein produced by recombinant *S. cerevisiae* Y294[AteA_Nat-XYNSEC] (Fig. 4.2a and 4.2d). The extracellular protein levels of AmyA and AteA were similar (Fig. 4.2d).

The *S. cerevisiae* Y294[ApuA_Nat] and Y294[TemA_Nat] strains displayed more extracellular α-amylase activity on soluble starch (Fig. 4.2b and 4.2c) than the *S. cerevisiae* Y294[AmyA] benchmark strain. Codon optimisation of the *apuA_Nat* and *temA_Nat* genes resulted in less extracellular activity due to a decrease in extracellular enzyme concentration (Fig. 4.2e and 4.2f). Changing the secretion signal also resulted in a decrease in extracellular enzyme concentration with a negative impact on extracellular activity (Fig. 4.2c and 4.2d).

SDS-PAGE analysis of the supernatant indicated that most of these α-amylases were glycosylated. The ApuA and AteA protein species (calculated molecular weights of 65.3 kDa and 64.1 kDa, respectfully) (Fig 4.2b and d) were the least glycosylated α-amylases with a putative recombinant size of around 70 kDa, while TemA (calculated molecular weight of 66.29 kDa) had a higher degree of glycosylation (Fig. 4.2f) and a putative size of around 90 kDa. The large heterogeneous smear between 110 and 150 kDa for the AmyA protein is consistent with that of a previous report (Viktor et al. 2013).

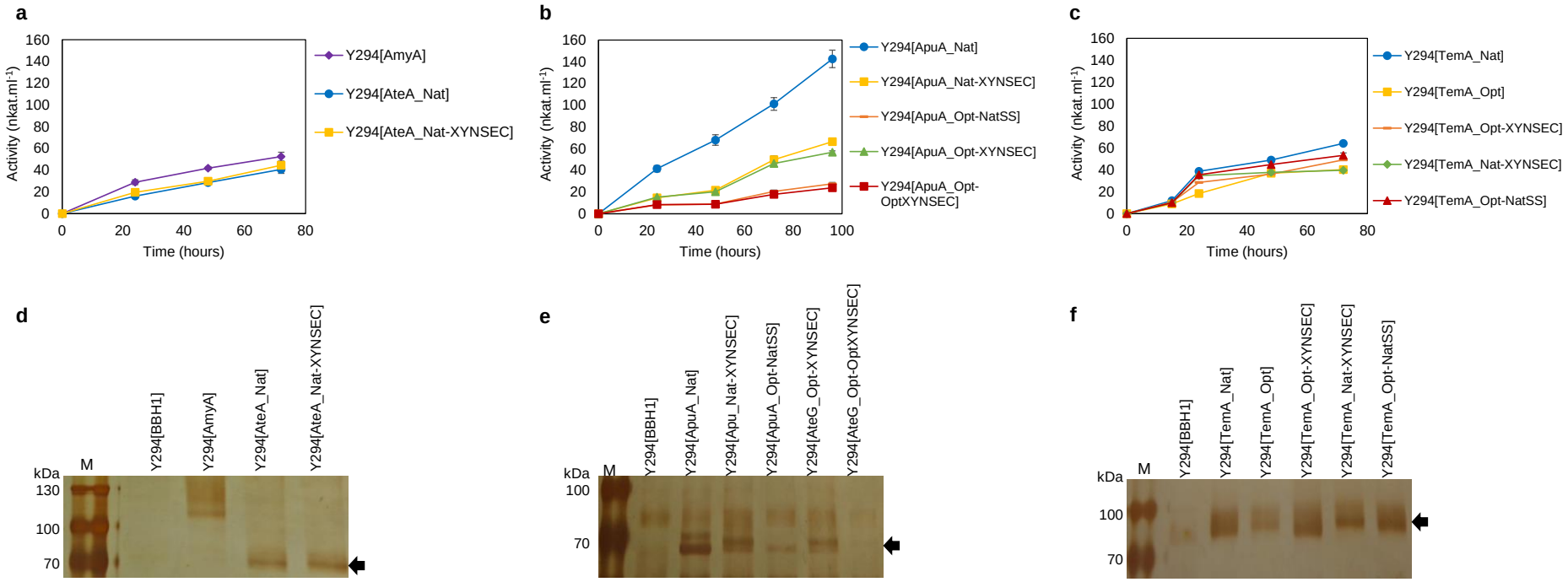


**Fig. 4.2** Extracellular α-amylase activity displayed by the *S. cerevisiae* Y294 strains expressing the (**a**) *ateA*, *amyA*, (**b**) *apuA* and (**c**) *temA* gene derivatives, respectively. The *S. cerevisiae* Y294[AmyA] strain was used for benchmark α-amylase production. Values represent the mean of three repeats and error bars represent the standard deviation. Supernatant from the *S. cerevisiae* Y294 strains (after 72 hours) was subjected to SDS-PAGE followed by silver staining. The arrows indicate the presence of the recombinant (**d**) AmyA, AteA, (**e**) ApuA and (**f**) TemA protein species, respectively. The *S. cerevisiae* Y294[BBH1] strain was used as the reference strain and the protein size marker (M) is depicted on the left hand side.

### *4.4.1.2 Glucoamylases*

The replacement of the *ateG_Nat* secretion signal with the *XYNSEC* sequence improved extracellular glucoamylase activity, albeit less than the activity displayed by the *S. cerevisiae* Y294[GlaA] benchmark strain (Fig. 4.3a). The *S. cerevisiae* Y294[AteG_Opt-XYNSEC] and Y294[AteG_Nat-XYNSEC] strains produced similar levels of activity, which exceeded the activity by the strains containing the native *ateG* secretion signal. The *S. cerevisiae* Y294[AteG_Opt-NatSS] strain secreted no visible protein (Fig. 4.3c) confirming that the native *ateG* secretion signal negatively affected protein secretion. Codon optimisation did not have a visible effect on the extracellular amount of AteG protein produced, despite the increase in extracellular activity (Fig. 4.3a and 4.3c).

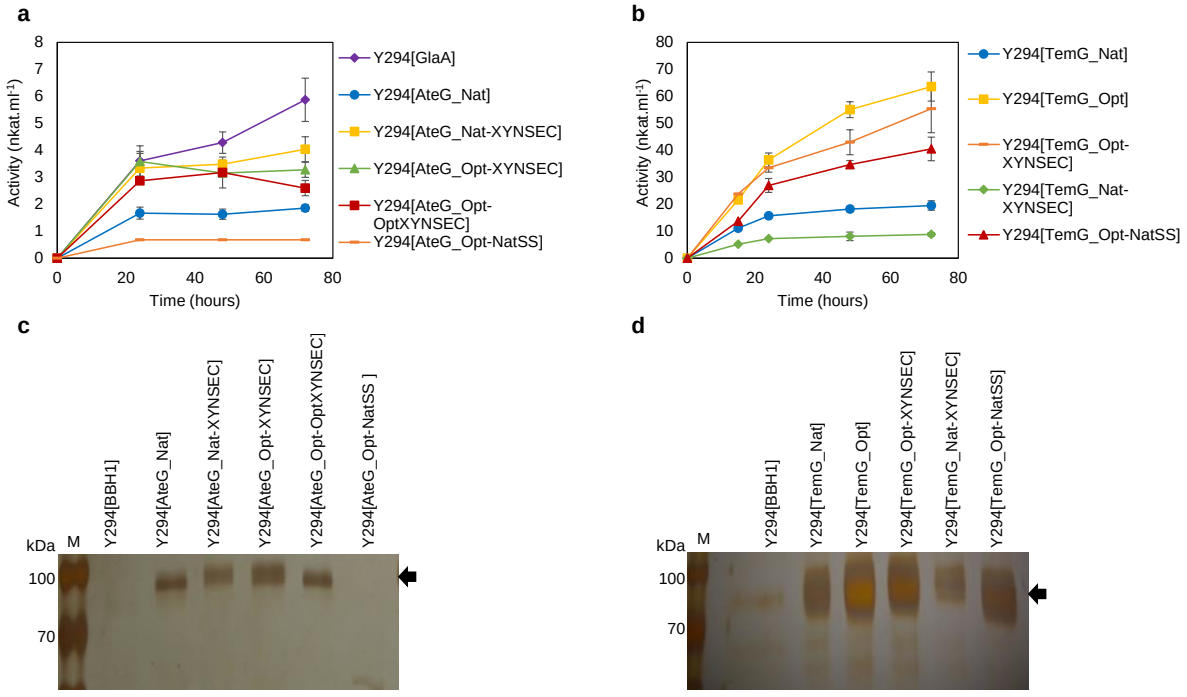


**Fig. 4.3** Extracellular glucoamylase activity displayed by the *S. cerevisiae* Y294 strains expressing the (**a**) *ateG, glaA* and (**b**) *temG* gene derivatives, respectively. The *S. cerevisiae* Y294[GlaA] strain was used for benchmark glucoamylase production. Values represent the mean of three repeats and error bars represent the standard deviation. Supernatant from the *S. cerevisiae* Y294 strains (after 72 hours) was subjected to SDS-PAGE followed by silver staining. The arrows indicate the presence of the recombinant (**c**) AteG and (**d**) TemG protein species, respectively. The *S. cerevisiae* Y294[BBH1] strain was used as the reference strain and the protein size marker (M) is depicted on the left hand side.

A significant increase in extracellular glucoamylase activity was observed when the *temG* gene was codon optimised (Fig. 4.3b). At 72 hours, extracellular activity for the *S. cerevisiae* Y294[TemG_Opt] strain was > 3-fold higher than the *S. cerevisiae*

Y294[TemG_Nat] strain and > 10-fold higher than the Y294[GlaA] benchmark strain. Changing secretion signals for the expression of the *temG* indicated that the optimised *temG* secretion signal contributed to enhanced protein secretion and extracellular activity (Fig. 4.3b and 4.3d), whereas replacement with the XYNSEC had a negative impact on protein production.

SDS-PAGE analysis of the supernatant indicated that these glucoamylases were glycosylated. The AteG protein species (calculated molecular weight of 65.73 kDa) (Fig. 4.3c) had a putative size of around 95 kDa, while the TemG protein (calculated molecular weight of 63.57 kDa) was less glycosylated with a putative size of around 85 kDa (Fig. 4.3d). Moreover, the intensity of the recombinant protein species visualised using SDS-PAGE showed correlation with the extracellular enzyme activity levels for all amylases.

### 4.4.2 Raw corn starch fermentations

The amylase encoding gene variants that resulted in the highest levels of extracellular activity when expressed in *S. cerevisiae* Y294 (*apuA_Nat*, *ateA_Nat*, *temA_Nat, temA_Opt*, *ateG_Nat-XYNSEC*, *temG_Nat* and *temG_Opt*), together with the benchmark (*amyA* and *glaA*) genes, were then used to construct amylolytic strains that produced a different α-amylase and glucoamylase combination (Table 4.1). The recombinant yeast strains were evaluated (using 100 ml serum bottle fermentations) for their ability to hydrolyse raw starch and ferment glucose at a high substrate loading under oxygen-limited conditions (Fig. 4.4).

At 192 hours, the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain produced the highest ethanol concentration (62.2 g. $l^{-1}$), which was 59.7% of the theoretical ethanol yield (Fig. 4.4a). After 120 hours, the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain had produced 51.7 g. $l^{-1}$ ethanol, which represented a 1.6-fold improvement on the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain ($p$ = 0.0013). Ethanol levels of 38.6 g.$l^{-1}$ and 39.4 g.$l^{-1}$produced by the *S. cerevisiae* Y294[TemG_Opt-ApuA_Nat] and Y294[TemG_Opt-AteA_Nat] strains, respectively, were also higher than the benchmark strain (at 120 hours). The *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain accumulated 46.3 g.$l^{-1}$ residual glucose after 192 hours of fermentation, while insignificant glucose concentrations (< 5 g.$l^{-1}$) were detected in the other fermentations (Fig. 4.4c).

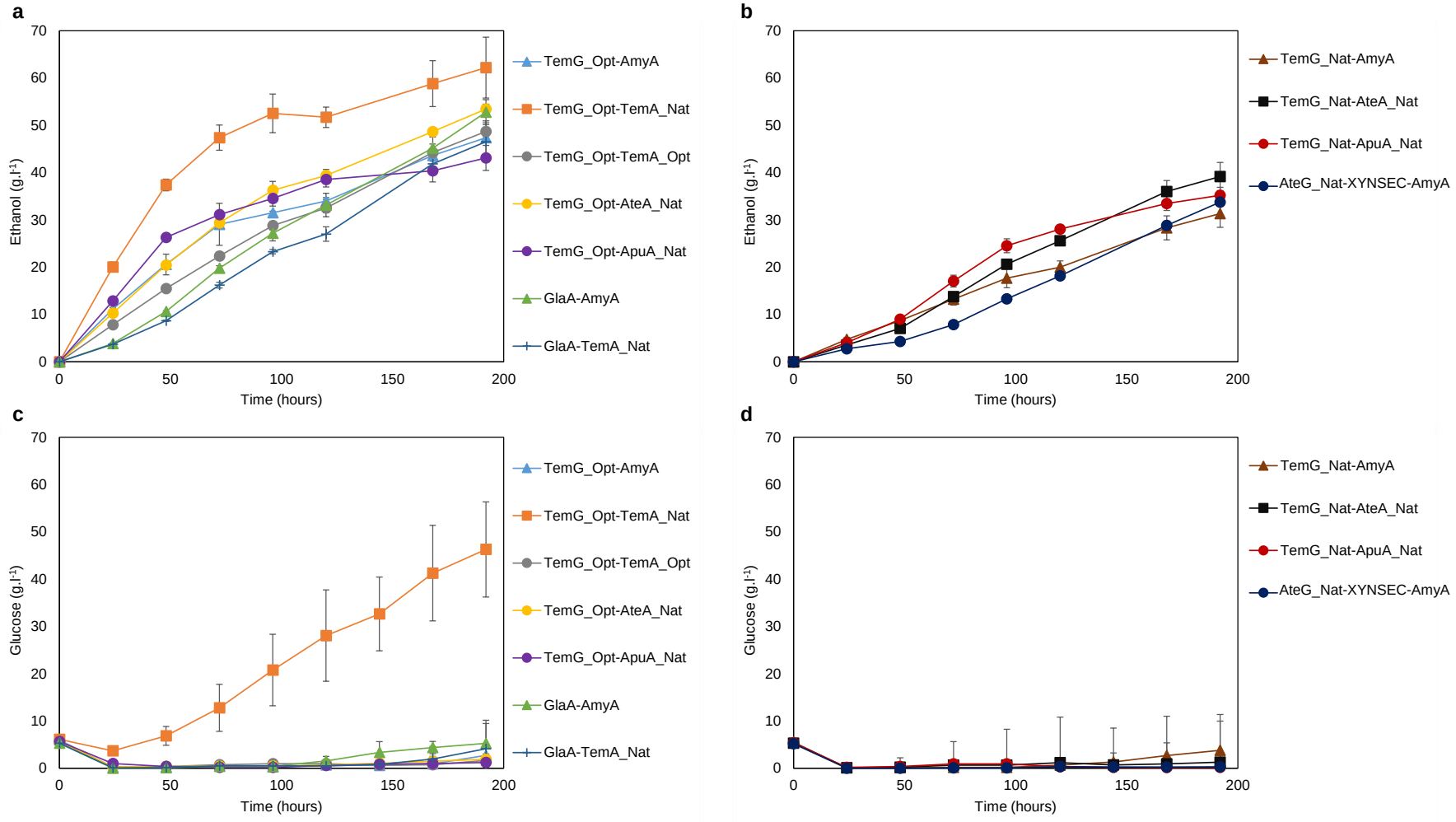


**Fig. 4.4** The amylolytic *S. cerevisiae* Y294 strains were evaluated on 200 g.l$^{-1}$ raw corn starch and 5 g.l$^{-1}$ glucose as sole carbohydrate source. The (**a** and **b**) ethanol and (**c** and **d**) glucose production was monitored overtime. Results from the best performing strains (left panel) and suboptimal strains (right panel) came from the same fermentation. Values represent the mean of three repeats and error bars represent the standard deviation.

The *S. cerevisiae* Y294 strains expressing the TemG_Nat-AmyA, TemG_Nat-ApuA_Nat and AteG_Nat-XYNSEC-AmyA amylase combinations produced less ethanol compared to the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain (Fig. 4.4a and 4.4b), with little to no residual glucose detected (Fig. 4.4d). Overall, results depicted in Fig. 4.4c indicated that the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain was superior to the other strains and the secreted recombinant enzyme combination was effective in hydrolysing raw corn starch at fermentation temperatures. At 192 hours, the carbon conversion obtained by the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain was 57% higher than that of the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain, whereas the *S. cerevisiae* Y294[TemG_Opt-AteA_Nat] strain produced comparable results to that of the benchmark strain (Table 4.3).

**Table 4.3**. Products formed by the *S. cerevisiae* Y294 strains after 192 hours of fermentation at 30°C in 2×SC$^{-URA}$ broth with glucose (5 g.l$^{-1}$) and raw corn starch (200 g.l$^{-1}$)

| ***S. cerevisiae* Y294 strains** | **[TemG_Opt-AmyA]** | **[TemG_Opt-TemA_Nat]** | **[TemG_Opt-TemA_Opt]** | **[TemG_Opt-AteA_Nat]** | **[TemG_Opt-ApuA_Nat]** | **[GlaA-AmyA]** | **[GlaA-TemA_Nat]** |
|---|---|---|---|---|---|---|---|
| **Substrate (g.l$^{-1}$)** | | | | | | | |
| Raw starch (dry weight) | 185 | 185 | 185 | 185 | 185 | 185 | 185 |
| Glucose equivalent | 208.5 | 208.5 | 208.5 | 208.5 | 208.5 | 208.5 | 208.5 |
| **Products (g.l$^{-1}$)** | | | | | | | |
| Glucose | 2.72 | 46.30 | 1.67 | 1.94 | 1.21 | 5.30 | 4.12 |
| Glycerol | 4.76 | 6.64 | 2.40 | 3.43 | 2.45 | 2.46 | 2.26 |
| Maltose | 1.09 | 1.03 | 1.07 | 1.14 | 0.95 | 1.02 | 1.17 |
| Acetic acid | 1.91 | 1.66 | 0.60 | 0.85 | 0.61 | 0.61 | 0.56 |
| Ethanol | 47.40 | 62.20 | 48.71 | 53.46 | 43.12 | 52.78 | 46.56 |
| $CO_2$[1] | 45.33 | 59.50 | 46.59 | 51.13 | 41.25 | 50.48 | 44.53 |
| **Total** | 103.21 | 177.33 | 101.04 | 111.95 | 89.60 | 112.65 | 99.20 |
| Carbon conversion (%) | 49.50 | 85.05 | 48.46 | 53.69 | 42.97 | 54.03 | 47.58 |
| Ethanol yield[2] (% of theoretical yield) | 45.46 | 59.67 | 46.72 | 51.28 | 41.36 | 50.63 | 44.66 |
| Ethanol productivity[3] | 0.247 | 0.324 | 0.254 | 0.278 | 0.225 | 0.275 | 0.242 |

[1]$CO_2$ concentrations were deduced from the ethanol produced
[2]Ethanol yield (% of the theoretical yield) was calculated as the amount of ethanol produced per gram of consumed glucose (at a specific time point)
[3]Ethanol productivity was calculated based on ethanol concentrations produced per hour (g.l$^{-1}$.h$^{-1}$)

The *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain was evaluated in a 2 litre bioreactor (1 litre working volume) under two fermentation temperatures (26°C and 30°C) (Fig. 4.5). After 192 hours, the final ethanol concentration (83.8 g.l$^{-1}$) was significantly higher at a fermentation temperature of 26°C (Fig. 4.5a), however the carbon conversion percentages were similar (79 - 81%). After 192 hours, a decrease in fermentation temperature resulted in 1.8-fold improvement in the ethanol concentration and no residual glucose was detected at a fermentation temperature of 26°C (Fig. 4.5a). The carbon conversion displayed by the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain (at 30°C) was similar for both fermentation types (100 ml bottle fermentations and bioreactor), 85% and 81% respectively, after 192 hours (Table 4.3 and Fig. 4.5b).

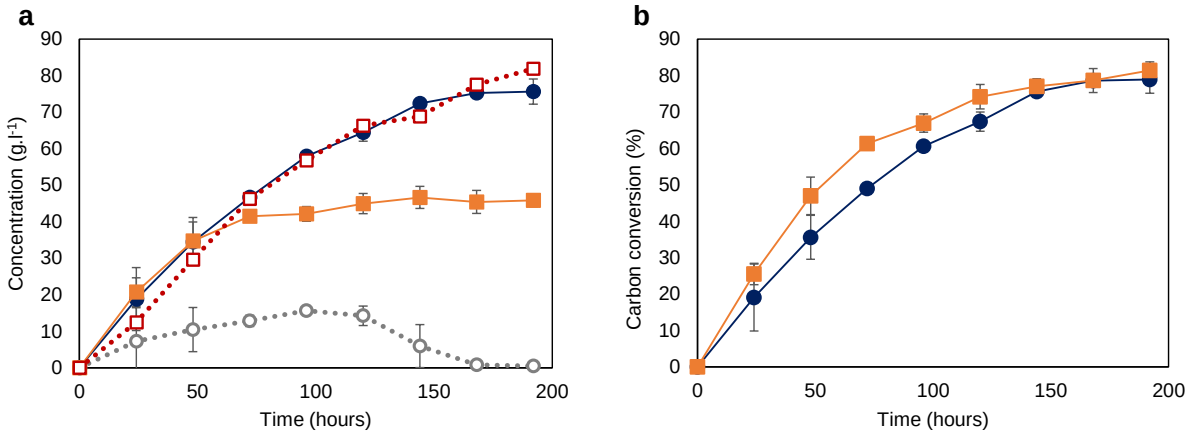


**Fig. 4.5** The performance of *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] in a 2 litre bioreactor. (**a**) Ethanol concentrations at 26°C (-●-) and 30°C (-■-) and residual glucose concentrations at 26°C (-○-) and at 30°C (-□-) and (**b**) carbon conversion at 26°C (-●-) and 30°C (-■-), respectively, with 2×SC$^{-URA}$ broth supplemented with 5 g.l$^{-1}$ glucose and 200 g.l$^{-1}$ raw corn starch. Values represent the mean of triplicate repeats and error bars represent the standard deviation.

## 4.5 Discussion

A selection of amylases from various fungi have been investigated independently by several research groups with raw starch hydrolysing enzymes being favoured for starch conversion to ethanol (Robertson et al. 2006; Viktor et al. 2013; Favaro et al. 2015; Celińska et al. 2015). Approximately 10% of all amylases contain SBD (Sun et al. 2010), which are classically associated with the adsorption of these enzymes to raw starch granules thereby enhancing the amylolytic rate and the subsequent hydrolysis (Santiago et al. 2005; Mitsuiki et al. 2005). Thus, for this study, the presence of a SBD was a prerequisite when selecting amylases for expression in

*S. cerevisiae*. The chosen amylase genes were heterologously expressed in order to choose the enzymes with the highest extracellular enzyme activity on corn starch, as well as to investigate the effect of adapted synonymous codon usage on gene expression (Table 4.1).

Previous studies have shown that high levels of gene expression can be correlated to the codon adaptation index (CAI) (Carbone et al. 2003). A CAI value of 1.0 is considered to be ideal, while GenScript recommends that a CAI of > 0.8 is rated as good for expression in the desired expression organism. Analysis of the genes' CAI values using GenScript's OptimumGeneTM (http://www.genscript.com/cgi-bin/tools/rare_codon_analysis) indicated that all CAI values increased when the genes were optimised. GenScript's algorithm for gene optimisation aims to improve gene expression and therefore the synthetic amylase genes in this study were codon optimised for expression in *S. cerevisiae*. However, results indicated that increased gene expression and protein secretion was not guaranteed by codon optimisation strategies (Fig. 4.2 and 4.3).

The strains expressing the *apuA_Nat* and *temA_Nat* genes were superior to the strains expressing their codon optimised counterparts *apuA_Opt-NatSS*/*apuA_Opt-OptXYNSEC* and *temA_Opt*, respectively (Fig. 4.2b and 4.2c), while optimisation of the *temG* coding sequence resulted in a significant increase in TemG_Opt protein secreted by the *S. cerevisiae* Y294[TemG_Opt] strain (Fig. 4.3d). Increased recombinant protein secretion correlated with enhanced levels of extracellular activity, which suggested similar specific activities (Fig. 4.2e and 4.2f) and SDS-PAGE analysis indicated that codon optimisation did not affect amylase protein size (Fig. 4.2 and 4.3). Based on the deduced amino acid sequences, the predicted molecular weights of the unglycosylated amylases were around 64 - 70 kDa, which is in agreement with previous reports on similar amylases (Gupta et al. 2003).

The *temA_Nat* α-amylase gene had a CAI of 0.61 compared to the *temA_Opt* gene with a CAI of 0.91. However, after 72 hours the *S. cerevisiae* Y294[TemA_Nat] strain produced 59% more extracellular α-amylase activity than the *S. cerevisiae* Y294[TemA_Opt] strain. The *temG_Nat* glucoamylase gene had a CAI of 0.58 compared to *temG_Opt* gene, which had a CAI of 0.91. The extracellular glucoamylase activity for the *S. cerevisiae* Y294[TemG_Nat] and Y294[TemG_Opt] strains represented a > 3-fold and > 10-fold fold improvement, respectively, compared

to the *S. cerevisiae* Y294[GlaA] benchmark strain. Therefore, even for genes originating from the same species (in this case *T. emersonii*), significant differences in protein secretion and extracellular enzyme activities were observed between native and codon optimised genes. Thus, CAI values alone cannot be relied upon for improving gene expression in *S. cerevisiae*.

Secretion signals are used to direct the propeptide to the endoplasmic reticulum (ER) and then through the secretory pathway (Futatsumori-Sugai and Tsumoto 2010). Therefore, signal peptides represented an important factor to consider when improving the concentration of secreted protein. The XYNSEC secretion signal from the *Trichoderma reesei* β-xylanase 2 gene has been used successfully for the secretion of a number of proteins (van Wyk et al. 2010; van Rensburg et al. 2012; Favaro et al. 2013) and was used in this study for comparative purposes. All of the native enzymes selected for this study were successfully secreted using their native secretion peptides. Yet, the replacement of the native *ateG* signal peptide encoding sequence with the *XYNSEC* sequence resulted in enhanced extracellular activity (Fig. 4.3a). However, in general the XYNSEC secretion signal was less effective than the proteins' native secretion signals. Furthermore, results from Fig. 4.3d suggested that the *temG_Opt* secretion signal may be a good candidate for improving protein secretion in *S. cerevisiae,* since it resulted in better protein production compared to that produced by the *S. cerevisiae* Y294[TemG_Opt-NatSS] strain.

Following the identification of successful amylase candidates, novel gene combinations were expressed in *S. cerevisiae* Y294 in order to obtain an amylolytic yeast suitable for raw starch CBP. It has been previously reported that starch fermentation by genetically engineered strains is limited by the glucoamylase activity (Inlow et al. 1988), but in a more recent review the limiting factor in raw starch hydrolysis was attributed to α-amylase activity (Görgens et al. 2015). The type of starchy biomass (used as substrate) is likely to affect the ratio of amylases, but if a recombinant amylolytic yeast is able to produce highly active enzymes, an exact ratio should not be a limiting factor.

During cultivation on 200 $g.l^{-1}$ raw corn starch, simultaneous expression of the α-amylase and glucoamylase combinations in *S. cerevisiae* resulted in varying ethanol yields (Fig. 4.4a and 4.4.b). After 72 hours, the carbon conversion displayed by the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain was 2.7-fold higher than the

*S. cerevisiae* Y294[AmyA-GlaA] benchmark strain. The *S. cerevisiae* Y294[TemG_Opt-ApuA_Nat] and Y294[TemG_Opt-AteA_Nat] strains also outperformed the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain (Fig. 4.4a) in the early stages of fermentation (> 2.4-fold higher ethanol concentrations after 48 hours). Substantially higher ethanol concentrations were obtained, compared to the modified amylolytic yeast strain constructed by Yamakawa et al. (2012), which produced 46.5 $g.l^{-1}$ ethanol from 200 $g.l^{-1}$ of raw corn starch. Furthermore, these results showed considerable improvements when compared to amylolytic CBP systems listed in a recent review by Salehi Jouzani and Taherzadeh (2015). The carbon conversion displayed by the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain on raw corn starch (Table 4.3) represented the highest reported for amylolytic *S. cerevisiae* Y294 strains in fermentations with high substrate loading and low inoculums.

Overall, the recombinant *S. cerevisiae* strains with higher levels of glucoamylase, i.e. those expressing the *temG_Opt* glucoamylase, hydrolysed starch better than the *S. cerevisiae* Y294 strains with the *temG_Nat* glucoamylase and faster than the *S. cerevisiae* strains expressing the *glaA* glucoamylase. However, the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain displayed a significantly higher carbon conversion (~1.6-2 fold) compared to any of the other recombinant *S. cerevisiae* Y294 strains expressing the *temG_Opt* glucoamylase (Table 4.3). This suggested that there was a unique synergistic effect between the *T. emersonii* TemG_Opt and TemA_Nat enzymes that outperformed the other TemG_Opt-α-amylase combinations.

A synergistic effect was also observed for the *A. tubingensis* enzyme combination. At 192 hours, the 54% carbon converison displayed by the *S. cerevisiae* Y294[GlaA-AmyA] strain was 9% higher than the carbon conversion displayed by the *S. cerevisiae* Y294[TemG_Opt-AmyA] strain (49%) (Table 4.3), even though TemG_Opt was superior to GlaA in terms of extracellular glucoamylase activity (Fig. 4.3). This highlighted the importance of comparing different enzyme combinations in the chosen expression host. Even though extracellular amylase activities differed on soluble starch (Fig. 4.2 and 4.3), enzymes originating from the same host may have a superior synergistic hydrolytic effect as a result of their modes of action and affinity for raw starch. Presečki et al. (2013) developed a mathematical model to explain the synergism between a glucoamylase and two α-amylases (in different combinations) and showed that the type and combinations of amylases

affected enzyme synergy. Furthermore, whether an α-amylase is classified as “liquefying” or “saccharifying” may also attribute to the synergist relationship (Liakopoulou-Kyriakides et al. 2001).

The AmyA α-amylase displayed a greater extracellular activity on soluble starch, compared to the AteA_Nat α-amylase (Fig. 4.2a). However, during fermentation studies enzyme combinations containing the AteA_Nat α-amylase facilitated a faster rate of raw starch conversion compared to the enzyme combinations with AmyA (Fig. 4.4). The AteA_Nat α-amylase also contributed to higher ethanol productivity (compared to the AmyA α-amylase) when combined with the TemG_Opt and TemG_Nat glucoamylases, respectively (Fig. 4.4 and Table 4.3). This suggested that the AteA_Nat enzyme may have performed better on raw starch compared to the AmyA enzyme, or it had a superior synergistic effect with the TemG glucoamylases.

Dissimilarly, the extracellular activity produced by the *S. cerevisiae* Y294[ApuA_Nat] strain (expressing the native α-amylase from *A. pullulans*) was 2.7-fold higher than that of the *S. cerevisiae* Y294[AmyA] benchmark strain (Fig. 4.2a and 4.2b), but overall the carbon conversion by the amylolytic *S. cerevisiae* Y294[TemG_Opt-ApuA_Nat] strain was 13% lower than the *S. cerevisiae* Y294[TemG_Opt-AmyA] strain (Table 4.3). Therefore, AmyA may either have had improved raw starch converting ability, or a better synergistic relationship with TemG_Opt, compared to ApuA_Nat (Fig. 4.4a).

Chi et al. (2009) demonstrated that the glucoamylase from *A. pullulans* hydrolysed potato starch granules (type-B crystallinity) better than raw corn starch granules (type-A crystallinity), although type-B starch structures are usually more resistant to enzyme hydrolysis (Man et al. 2013). Corn starch has a higher amylose content and smaller granule diameter compared to potato starch (Hii et al. 2012) and the combination of these properties are known to influence the rate and extent of starch hydrolysis (Naguleswaran et al. 2013). Results from this study (Fig. 4.2 and 4.4) highlighted a prime example where starch structure affected the action of different amylolytic enzymes.

Although *S. cerevisiae* is known for its ethanol tolerance, the Y294 strains were inhibited by fermentation conditions at a cultivation temperature of 30°C and thus ethanol concentrations did not exceed 63 $g.l^{-1}$ (Fig. 4.5 and 4.6). The poor fermentative

performance by the *S. cerevisiae* Y294 laboratory strain at 30°C was not as a result of inadequate recombinant protein secretion or low enzymatic activity, since glucose concentrations increased rapidly throughout the fermentation with the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain (Fig. 4.4b) and the final carbon conversion after 192 hours was 85% (Table 4.3).

Raw starch fermentation by the recombinant *S. cerevisiae* strains is often disadvantaged by long cultivations times required for sufficient enzyme secretion. However, it was clear from the fermentation results for the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain (Fig. 4.5) that volumetric productivity of ethanol and starch conversion rates were high. Furthermore, a fermentation temperature of 26°C relieved physiological stress on the yeast cells allowing for improved glucose conversion. After 192 hours, the carbon conversion displayed by the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain were the similar (81 - 85%) for the 100 ml serum bottles and bioreactor fermentations (Table 4.3 and Fig. 4.5b). Thus suggesting that the lower temperature was the main factor to favour glucose fermentation and that the extracellular enzyme activity was not significantly affected by a lower temperature (since the final carbon conversion remained the same at both fermentation temperatures). Therefore, decreasing the fermentation temperature confirmed that it was possible to increase the conversion of glucose to ethanol and improve the theoretical ethanol yield.

Schmidt et al. (2006) provided several definitions for ethanol tolerance, one of which was the effect of ethanol concentrations on the ability of a cell to metabolise sugar. Biochemical and physiological responses occur when yeast are exposed to accumulating ethanol titres (Schmidt et al. 2006) and as a result the *S. cerevisiae* Y294 strains were likely to experience compromised membrane structure and protein function at 30°C. The presence of ethanol changes the composition of the phospholipid bilayer making it permeable to small molecules. Since many cellular functions rely on membrane integrity, high ethanol concentrations can have a number of adverse effects on the yeast cells. In this study, the negative effects of ethanol accumulation were avoided by lowering the fermentation temperature to 26°C.

Few genes have been cloned and sequenced from thermophilic fungi. Glucoamylases from *T. lanuginosus* (Thorsen et al. 2006), *T. emersonii* (Nielsen et al. 2002), *H. grisea* var. thermoidea (Allison et al. 1992) and *C. thermophilum* (Chen et al. 2007) have

been expressed in fungal host strains. However, the *T. emersonii* amylases have not been expressed in *S. cerevisiae*. *T. emersonii* is an acidicothermophilic fungus that is industrially important and well recognised for its production of glycoside hydrolases (GHs) with special enzymatic properties, especially cellulases (Amore and Faraco 2012; Wang et al. 2014). However, few studies have investigated its starch hydrolysing enzymes.

*T. emersonii* has the potential to contend with industrially important fungal enzyme producers i.e. species of *Aspergillus* (*A. awamori* and *A. oryzae*) with regards to the production of amylases. In 2002, the first *T. emersonii* amylase was cloned and heterologously expressed in *A. niger* (Nielsen et al. 2002). Upon purification, the glucoamylase indicated improved half-life and high specific activity towards maltose, isomaltose and maltoheptaose. Furthermore, when compared with other fungal amylases, the *T. emersonii* amylases had a high thermal stability (Bunni et al. 1989) and have been used in the baking industry (Waters et al. 2010); thus demonstrating their potential for starch processing. This study confirmed their efficiency in raw starch hydrolysis for ethanol production. Furthermore, the thermostability and economical production of *T. emersonii* amylases are advantageous characteristics, which would be desirable to the biofuel industry.

## 4.6 Conclusion

Currently, industry lacks the implementation of an amylolytic CBP yeast that expresses both an α-amylase and glucoamylase. This study focused on the selection of highly active amylases with the ability to convert raw starch to glucose and led to the identification and evaluation of novel amylase combinations for the hydrolysis of raw corn starch. The recombinant *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain was superior in its ability to convert 85% of the available carbon in 200 $g.l^{-1}$ raw corn starch within 192 hours. Thus, this unique TemG_Opt-TemA_Nat enzyme combination represented a promising candidate for the industrial conversion of uncooked starch. Further investigations are required that focus on improving the volumetric productivity and yeast thermotolerance, in order to decrease fermentations times and ensure that all available glucose is fermented at standard fermentation temperatures.

## 4.7 Acknowledgments

The authors would like to thank Mrs Lisa Warburg for assistance with HPLC analysis and the National Research Foundation of South Africa for financial support.

## 4.8 References

Allison DS, Rey MW, Berka RM, Armstrong G, Dunn-Coleman NS (1992) Transformation of the thermophilic fungus *Humicola grisea* var. thermoidea and over- production of *Humicola* glucoamylase. Curr Genet 21:225–229. doi: 10.1007/BF00336845

Amore A, Faraco V (2012) Potential of fungi as category I Consolidated BioProcessing organisms for cellulosic ethanol production. Renew Sust Energ Rev 16:3286–3301. doi.10.1016/j.rser.2012.02.050

Białas W, Czerniak A, Szymanowska-Powałowska D (2014) Kinetic modeling of simultaneous saccharification and fermentation of corn starch for ethanol production. Acta Biochimica Polonica 61:153–162. PMID: 24649484

Bunni L, McHale L, McHale AP (1989) Production, isolation and partial characterization of an amylase system produced by *Talaromyces emersonii* CBS 814.70. Enzyme Microb Technol 11:370–375. doi.10.1016/0141-0229(89)90022-7

Carbone A, Zinovyev A, Képès F (2003) Codon adaptation index as a measure of dominating codon bias. Bioinformatics 19:2005–2015. doi: 10.1093/bioinformatics/btg272

Celińska E, Białas W, Borkowska M, Grajek W (2015) Cloning, expression, and purification of insect (*Sitophilus oryzae*) alpha-amylase, able to digest granular starch, in *Yarrowia lipolytica* host. Appl Microbiol Biotechnol 99:2727–2739. doi: 10.1007/s00253-014-6314-2

Chen J, Zhang Y, Zhao C, Li A, Zhou Q, Li D (2007) Cloning of a gene encoding thermostable glucoamylase from *Chaetomium thermophilum* and its expression in *Pichia pastoris*. J Appl Microbiol 103:2277–2284. doi: 10.1111/j.1365-2672.2007.03475.x

Chi Z, Wang F, Chi Z, Yue L, Liu G, Zhang T (2009) Bioproducts from *Aureobasidium pullulans*, a biotechnologically important yeast. Appl Microbiol Biotechnol 82:793–804. doi: 10.1007/s00253-009-1882-2

Cho KM, Yoo YJ, Kang HS (1999) δ-Integration of endo/exoglucanase and β-glucosidase genes into the yeast chromosomes for direct conversion of cellulose to ethanol. Enzyme Microb Technol 25:23–30. doi: 10.1016/S0141-0229(99)00011-3

Cinelli BA, Castilho LR, Freire DMG, Castro AM (2015) A brief review on the emerging technology of ethanol production by cold hydrolysis of raw starch. Fuel 150:721-729. doi: 10.1016/j.fuel.2015.02.063

Den Haan R, Rose SH, Lynd LR, van Zyl WH (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. Metab Eng 9:87—94. doi: 10.1016/j.ymben.2006.08.005

den Haan R, Kroukamp H, Mert M, Bloom M, Görgens JF, van Zyl WH (2013) Engineering *Saccharomyces cerevisiae* for next generation ethanol production. J Chem Technol Biotechnol 88:983–991. doi: 10.1002/jctb.4068

Eksteen JM, van Rensburg P, Cordero Otero RR, Pretorius IS (2003) Starch fermentation by recombinant *Saccharomyces cerevisiae* strains expressing the α-amylase and glucoamylase genes from *Lipomyces kononenkoae* and *Saccharomycopsis fibuligera*. Biotechnol Bioeng 84:639–646. doi: 10.1002/bit.10797

Favaro L, Jooste T, Basaglia M, Rose SH, Saayman M, Görgens JF, Casella S, van Zyl WH (2013) Designing industrial yeasts for consolidated bioprocessing of starchy biomass to ethanol. Bioengineered 4:97–102. doi: 10.4161/bioe.22268

Favaro L, Viktor M, Rose SH, Viljoen-Bloom M, van Zyl WH, Basaglia M, Cagnin L, Casella S (2015) Consolidated bioprocessing of starchy substrates into ethanol by industrial *Saccharomyces cerevisiae* strains secreting fungal amylases. Biotechnol Bioeng 112:1751–1760. doi: 10.1002/bit.25591

Futatsumori-Sugai M, Tsumoto K (2010) Signal peptide design for improving recombinant protein secretion in the baculovirus expression vector system. Biochem Biophys Res Commun 391:931–935. doi: 10.1016/j.bbrc.2009.11.167

Görgens JF, Bressler DC, van Rensburg E (2015) Engineering *Saccharomyces cerevisiae* for direct conversion of raw, uncooked or granular starch to ethanol. Crit Rev Biotechnol 35:369–391. doi: 10.3109/07388551.2014.888048

Gupta R, Gigras P, Mohapatra H, Goswami VK, Chauhan B (2003) Microbial α-amylases: a biotechnological perspective. Process Biochem 38:1599–1616. doi: 10.1016/S0032-9592(03)00053-0

Hii SL, Tan JS, Ling TC, Ariff AB (2012) Pullulanase: role in starch hydrolysis and potential industrial applications. Enzyme Res 2012:921362. doi: 10.1155/2012/921362

Houbraken J, Spierenburg H, Frisvad JC (2012) *Rasamsonia*, a new genus comprising thermotolerant and thermophilic *Talaromyces* and *Geosmithia* species. Antonie Van Leeuwenhoek 101:403–421. doi: 10.1007/s10482-011-9647-1

Inlow D, McRae J, Ben-Bassat A (1988) Fermentation of corn starch to ethanol with genetically engineered yeast. Biotechnol Bioeng 32:227–234. doi: 10.1002/bit.260320213

Jeang CL, Chen LS, Chen MY, Shiau RJ (2002) Cloning of a gene encoding raw-starch-digesting amylase from a *Cytophaga* sp. and its expression in *Escherichia coli*. Appl Environ Microbiol 68:3651-3654. doi: 10.1128/AEM.68.7.3651-3654.2002

Knox AM, Du Preez JC, Kilian SG (2004) Starch fermentation characteristics of *Saccharomyces cerevisiae* strains transformed with amylase genes from *Lipomyces kononenkoae* and *Saccharomycopsis fibuligera*. Enzyme Microb Technol 34:453–460. doi: 10.1016/j.enzmictec.2003.12.010

Li H, Sun W, Gao Y, Wu Y, Huang L, Huang EZ, Wang A, Yin X, Wang Q, Xie T, Zeng Z (2011) Cloning recombinant expression and characterization of a new glucoamylase gene from *Aureobasidium pullulans* NRRL 12974 and its potential application in raw potato starch degradation. Afr J Biotechnol 10:9122–9131. doi: 10.5897/AJB11.819

Liakopoulou-Kyriakides M, Karakatsanis A, Stamatoudis M, Psomas S (2001) Synergistic hydrolysis of crude corn starch by α-amylases and glucoamylases of various origins. Cereal Chem 78:603–607. doi: 10.1094/CCHEM.2001.78.5.603

Mamo G, Gessesse A (1999) Purification and characterization of two raw-starch-digesting thermostable α-amylases from a thermophilic *Bacillus*. Enzyme Microb Technol 25:433–438. doi: 10.1016/S0141-0229(99)00068-X

Man J, Yang Y, Huang J, Zhang C, Zhang F, Wang Y, Gu M, Liu Q, Wei C (2013) Morphology and structural properties of high-amylose rice starch residues hydrolysed by amyloglucosidase. Food Chem 138:2089–2098. doi: 10.1016/j.foodchem.2012.12.009

Miller GL (1959) Use of dinitrosalicyclic acid reagent for determination of reducing sugars. Anal Chem 31:426–428. doi: 10.1021/ac60147a030

Mitsuiki S, Mukae K, Sakai M, Goto M, Hayashida S, Furukawa K (2005) Comparative characterization of raw starch hydrolyzing α-amylases from various *Bacillus* strains. Enzyme Microb Technol 37:410–416. doi: 10.1016/j.enzmictec.2005.02.022

Naguleswaran S, Vasanthan T, Hoover R, Bressler D (2013) The susceptibility of large and small granules of waxy, normal and high-amylose genotypes of barley and corn starches toward amylolysis at sub-gelatinization temperatures. Food Res Int 51:771–782. doi: 10.1016/j.foodres.2013.01.057

Nielsen BR, Lehmbeck J, Frandsen TP (2002) Cloning, heterologous expression, and enzymatic characterization of a thermostable glucoamylase from *Talaromyces emersonii*. Protein Expr Purif 26:1–8. doi: 10.1016/S1046-5928(02)00505-3

Njokweni A, Rose SH, van Zyl WH (2012) Fungal β-glucosidase expression in *Saccharomyces cerevisiae*. J Mol Microbiol Biotechnol 39:1445–1452. doi: 10.1007/s10295-012-1150-9

O'Connell K, Stults JT (1997) Identification of mouse liver proteins on two-dimensional electrophoresis gels by matrix-assisted laser desorption/ionization mass spectrometry of in situ enzymatic digests. Electrophoresis 18:349–359. doi: 10.1186/1754-6834-6-167

Presečki AV, Blažević ZF, Vasić-Rački D (2013) Complete starch hydrolysis by the synergistic action of amylase and glucoamylase: impact of calcium ions. Bioprocess Biosyst Eng 36:1555–1562. doi: 10.1007/s00449-013-0926-2

Rana N, Walia A, Gaur A (2013) α-Amylases from microbial sources and its potential applications in various industries. Natl Acad Sci Lett 36:9–17. doi: 10.1007/s40009-012-0104-0

Robertson GH, Wong DW, Lee CC, Wagschal K, Smith MR, Orts WJ (2006) Native or raw starch digestion: a key step in energy efficient biorefining of grain. J Agric Food Chem 54:353–365. doi: 10.1021/jf051883m

Salehi Jouzani G, Taherzadeh MJ (2015) Advances in consolidated bioprocessing systems for bioethanol and butanol production from biomass: a comprehensive review. Biofuel Res J 2:152–195. doi: 10.18331/BRJ2015.2.1.4

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press

Santiago M, Linares L, Sánchez S, Rodríguez-Sanoja R (2005) Functional characteristics of the starch-binding domain of *Lactobacillus amylovorus* α-amylase. Biologia (Bratislava) 60:111–114.

Schmidt SA, Tran T, Chambers PJ, Herderich MJ, Pretorius IS (2006) Developing indicators of wine yeast performance: an overview of the impact of ethanol stress. Austral New Zealand Wine Industry J 21:24–30.

Sun H, Zhao P, Ge X, Xia Y, Hao Z, Liu J, Peng M (2010) Recent advances in microbial raw starch degrading enzymes. Appl Biochem and Biotechnol 160:988–1003. doi: 10.1007/s12010-009-8579-y

Thorsen TS, Johnsen AH, Josefsen K, Jensen B (2006) Identification and characterization of glucoamylase from the fungus *Thermomyces lanuginosus*. Biochim Biophys Acta 1764:671–676. doi: 10.1016/j.bbapap.2006.01.009

Uthumporn U, Zaidul ISM, Karim AA (2010) Hydrolysis of granular starch at sub-gelatinization temperature using a mixture of amylolytic enzymes. Food and Bioprod Process 88:47-54. doi: 10.1016/j.fbp.2009.10.001

van Rensburg E, den Haan R, Smith J, van Zyl WH, Görgens JF (2012) The metabolic burden of cellulase expression by recombinant *Saccharomyces cerevisiae* Y294 in aerobic batch culture. Appl Microbiol and Biotechnol 96:197–209. doi: 10.1007/s00253-012-4037-9

van Wyk N, den Haan R, van Zyl WH (2010) Heterologous production of NpCel6A from *Neocallimastix patriciarum* in *Saccharomyces cerevisiae*. Enzyme Microb Technol 46:378-383. doi: 10.1016/j.enzmictec.2009.11.005

Viktor MJ, Rose SH, van Zyl WH, Viljoen-Bloom M (2013) Raw starch conversion by *Saccharomyces cerevisiae* expressing *Aspergillus tubingensis* amylases. Biotechnol Biofuels 6:167. doi: 10.1186/1754-6834-6-167

Wang K, Luo H, Shi P, Huang H, Bai Y, Yao B (2014) A highly-active endo-1,3-1,4-beta-glucanase from thermophilic *Talaromyces emersonii* CBS394.64 with application potential in the brewing and feed industries. Process Biochem 49:1448-1456. doi: 10.1016/j.procbio.2014.06.003

Waters DM, Murray PG, Ryan LA, Arendt EK, Tuohy MG (2010) *Talaromyces emersonii* thermostable enzyme systems and their applications in wheat baking systems. J Agric Food Chem 58:7415–7422. doi: 10.1021/jf100737v

Wong DW, Robertson GH, Lee CC, Wagschal K (2007) Synergistic action of recombinant alpha-amylase and glucoamylase on the hydrolysis of starch granules. Protein J 26:159-164 doi: 10.1007/s10930-006-9057-9

Yamada R, Yamakawa S, Tanaka T, Ogino C, Fukuda H, Kondo A (2011) Direct and efficient ethanol production from high-yielding rice using a *Saccharomyces cerevisiae* strain that express amylases. Enzyme Microb Technol 48:393–396. doi: 10.1016/j.enzmictec.2011.01.002

Yamakawa S, Yamada R, Tanaka T, Ogino C, Kondo A (2012) Repeated fermentation from raw starch using *Saccharomyces cerevisiae* displaying both glucoamylase and α-amylase. Enzyme Microb Technol 50:343–347. doi: 10.1016/j.enzmictec.2012.03.005

Yang S, Jia N, Li M, Wang J (2011) Heterologous expression and efficient ethanol production of a *Rhizopus* glucoamylase gene in *Saccharomyces cerevisiae*. Mol Biol Rep 38:59–64. doi: 10.1007/s11033-010-0077-3

# Chapter 5

Construction of an amylolytic CBP Ethanol Red™ strain

# Construction of an amylolytic CBP Ethanol Red™ strain

Rosemary A. Cripwell, Shaunita H. Rose, Willem H. van Zyl

Department of Microbiology, Stellenbosch University, Private Bag X1, 7602 Matieland, South Africa.

## 5.1 Abstract

Consolidated bioprocessing (CBP) integrates enzyme production, saccharification and fermentation into a one-step process. This strategy represents a promising alternative for economic ethanol production from starchy biomass. Recombinant *Saccharomyces cerevisiae* Ethanol Red™ and M2n yeast strains were used to produce ethanol directly from raw corn starch. Two δ-integration gene cassettes were constructed to allow for the simultaneous multiple integration of the codon optimised *Talaromyces emersonii* glucoamylase gene (*temG_Opt*) and the native *T. emersonii* α-amylase gene (*temA_Nat*) into the yeasts' genomes. The *T. emersonii* amylases were both constitutively expressed under the control of the *ENO1* promoter, using the δ-integration DNA transformation system. During the fermentation of 200g.l$^{-1}$ raw corn starch, the amylolytic industrial strains were able to ferment raw starch to ethanol in a single step with high ethanol yields. After 192 hours at 30°C, the *S. cerevisiae* Ethanol Red T12 and M2n T1 strains (containing integrated *temA_Nat* and *temG_Opt* gene cassettes) produced 86.5 g.l$^{-1}$ and 99.4 g.l$^{-1}$ ethanol, respectively, corresponding to carbon conversions (percentage starch converted on a mol carbon basis) of 84% and 96%, respectively. The addition of STARGEN 002™ in combination with the recombinant amylolytic yeast improved the rate of ethanol production and allowed for a 90% reduction in the enzyme dosage, compared to the conventional simultaneous saccharification and fermentation (SSF) process with the untransformed host strains. The amylolytic industrial strains were also able to grow on acrylamide as the sole nitrogen source demonstrating their potential novel application in environmental acrylamide reduction.

**Keywords:** • CBP • raw corn starch • Ethanol Red • amylases • acrylamide

## 5.2 Introduction

Starch is a renewable substrate that is a readily available raw material in most regions of the world (Mobini-Dehkordi and Javan 2012). There are numerous types of starchy

biomass that represent attractive substrates for bioethanol production, namely corn (maize), wheat, oats, rice, potato and cassava (Nigam and Singh 1995). For decades, amylases from various microbial sources have been used in starch based industries, which has led to amylases being among the most important enzymes used for industrial applications (Pandey et al. 2000; Sivaramakrishnan et al. 2006; Rana et al. 2013). However, only a limited number of fungal and bacterial strains meet the criteria for commercial amylase production. Therefore, new microorganisms are continuously screened for amylase activity, especially for applications in the biofuel industry.

The conventional process for the conversion of starch to ethanol requires a heat intensive liquefaction step to gelatinise the starch and thermostable α-amylases, followed by saccharification with glucoamylases. The high temperatures required for the initial processes usually account for approximately 30 - 40% of the total energy required for ethanol production (Szymanowska-Powałowska et al. 2012). An alternative to this is a cold hydrolysis process at temperatures below the onset of starch gelatinisation (65°C for corn) (Robertson et al. 2006). Benefits to this process include reduced energy requirements and a higher nutritional content for the distiller's dried grains with solubles (DDGS) (Nkomba et al. 2016). DDGS are produced in large quantities during bioethanol production and represent a valuable ingredient for livestock feed (Brehmer et al. 2008).

Consolidated bioprocessing (CBP) using a single organism combines enzyme production, hydrolysis and fermentation into a one-step process for bioethanol production at low temperatures (van Zyl et al. 2012). This technology has developed rapidly over the last decade and is a promising approach for the economic production of biofuel from lignocellulosic and starchy feedstocks (Salehi Jouzani and Taherzadeh 2015). However, CBP has not yet been implemented on a commercial level (den Haan et al. 2015), with the main challenge being the availability of an ideal microorganism that can express suitable enzymes and have a high fermentation capacity. CBP would simplify operational processes (e.g. number of control steps and reaction vessels) and therefore reduce maintenance and production costs.

The comprehensive review on consolidated bioprocessing systems by Salehi Jouzani and Taherzadeh (2015) highlighted different CBP strategies, diversity in the substrate types, as well as the organisms involved in fermenting the sugars. Currently no

industrial process uses an amylolytic yeast in a CBP process, but the commercial production of granular starch hydrolysing enzyme (GSHE) cocktails has allowed for the development of simultaneous saccharification and fermentation (SSF) processes (at lower temperatures) for ethanol production from starchy substrates (Balcerek and Pielech-Przybylska 2013; Szymanowska-Powałowska et al. 2014; Nkomba et al. 2016).

It is estimated that the use of raw starch hydrolysing enzymes for ethanol production reduces energy costs by 10 - 20% (Robertson et al. 2006). Genencor's STARGEN 002™ cocktails (Dupont-Danisco, Itasca, USA) hydrolyse raw starch at low temperatures (48°C is recommended for SSF), while POET (Sioux Falls, South Dakota, USA) uses a patented blend of Novozymes enzymes (POET BPX technology) in an SSF process (Görgens et al. 2015). On the other hand, a bioengineered *Saccharomyces cerevisiae* strain that secretes a glucoamylase, TransFerm® from Lallemand and developed by Mascoma Corporation, is commercially available (http://www.ethanoltech.com/transferm). However, the TransFerm® yeast lacks the required α-amylase enzyme for starch liquefaction (den Haan et al. 2015) and is therefore only a semi-CBP yeast. The TransFerm® yeast strain represents a drop-in organism for the conventional (warm) process, since it only consolidates the saccharification and fermentation processes after starch liquefaction.

A market has developed for a drop-in CBP yeast that is able to express both a raw starch α-amylase and glucoamylase for complete starch hydrolysis. One of the main challenges remains the simultaneous production of these enzymes with high substrate affinities and specific activity (den Haan et al. 2013). In addition, fermentation requirements are ethanol concentrations in excess of 10 - 12% ($w.v^{-1}$) within 48 to 72 hours (Bothast and Schlicher 2005).

In this study, two industrial *S. cerevisiae* strains were selected, namely Ethanol Red™ and the M2n distillery yeast. Ethanol Red™ is one of the most widely used yeast strains for first generation bioethanol production (Stovicek et al. 2015). Gene integration and the acetamide selection method were used for the engineering of the industrial yeast strains. The use of the *amdS* gene as a dominant marker enabled the selection of recombinant prototrophic strains on acetamide (Solis-Escalante et al. 2013), which replaced the conventional selection method that required antibiotics. The industrial amylolytic strains were evaluated at high solids loadings and were able to

hydrolyse raw corn starch to glucose, with an ethanol yield close to the theoretical maximum. The amylolytic *S. cerevisiae* strains represented suitable drop-in candidates for the existing cold fermentation process.

## 5.3 Materials and methods

### 5.3.1 Media and cultivation conditions

All chemicals were of analytical grade and were obtained from Merck (Darmstadt, Germany), unless otherwise stated. *Escherichia coli* DH5α (Takara Bio Inc.) was used for vector propagation. The *E. coli* transformants were selected for on Luria Bertani agar (Sigma-Aldrich, Steinheim, Germany), containing 100 μg.ml$^{-1}$ ampicillin and cultivated at 37°C in Terrific Broth (12 g.l$^{-1}$ tryptone, 24 g.l$^{-1}$ yeast extract, 4 ml.l$^{-1}$ glycerol, 0.1 M potassium phosphate buffer) containing 100 μg.ml$^{-1}$ ampicillin for selective pressure (Sambrook et al. 1989).

The *S. cerevisiae* parental strains were maintained on YPD agar plates (10 g.l$^{-1}$ yeast extract, 20 g.l$^{-1}$ peptone, 20 g.l$^{-1}$ glucose and 20 g.l$^{-1}$ agar). The *S. cerevisiae* Y294 transformants were selected for and maintained on SC$^{-URA}$ agar plates (6.7 g.l$^{-1}$ yeast nitrogen base without amino acids (BD-Diagnostic Systems, Sparks, Maryland, USA), 20 g.l$^{-1}$ glucose and 1.5 g.l$^{-1}$ yeast synthetic drop-out medium supplements (Sigma-Aldrich, Germany) and 20 g.l$^{-1}$ agar). *S. cerevisiae* strains were aerobically cultivated on a rotary shaker (200 rpm) at 30°C, in 125 ml Erlenmeyer flasks containing 20 ml double strength SC$^{-URA}$ medium (2×SC$^{-URA}$ containing 13.4 g.l$^{-1}$ yeast nitrogen base without amino acids (BD-Diagnostic Systems), 20 g.l$^{-1}$ glucose and 3 g.l$^{-1}$ yeast synthetic drop-out medium supplements). Fermentation media for the *S. cerevisiae* Y294 strains comprised of 2×SC$^{-URA}$ containing 5 g.l$^{-1}$ glucose and 200 g.l$^{-1}$ raw corn starch, whereas the medium for the *S. cerevisiae* Ethanol Red and M2n strains was YP containing 5 g.l$^{-1}$ glucose and 200 g.l$^{-1}$ raw corn starch. Ampicillin (100 μg.ml$^{-1}$) and streptomycin (50 μg.ml$^{-1}$) were added to inhibit bacterial contamination. All cultures were inoculated to a concentration of 1×10$^{6}$ cells.ml$^{-1}$, unless otherwise stated.

SC media (yeast synthetic drop-out medium omitted) containing 2% starch was used to maintain industrial transformants. The *S. cerevisiae* Ethanol Red and M2n transformants were selected for on SC-Ac plats (SC plates with $(NH_4)_2SO_4$ replaced by 0.6 g.l$^{-1}$ acetamide and 6.6 g.l$^{-1}$ $K_2SO_4$) and transferred to SC-Acr plates

(SC-Ac with 0.71 g.l$^{-1}$ acrylamide replacing the acetamide). For plate assays, 2% soluble starch was added to SC-Ac and SC-Acr plates. SC-Fac plates (SC media containing 2.3 g.l$^{-1}$ fluoroacetamide) was used to remove the yBBH1-amdSYM episomal vector from the transformants. The pH in all the media was adjusted to 6.0.

### 5.3.2 Strains and plasmids

The genotypes of the bacterial and yeast strains, as well as the plasmids used in this study are summarised in Table 5.1.

**Table 5.1**. Strains and plasmids used in this study

| Strains and plasmids | Genotype | Reference/ Source |
|---|---|---|
| *E. coli DH5α* | *supE44 ΔlacU169 (ϕ80lacZ*ΔM15*) hsdR17 recA1 endA1 gyrA96 thi-1 relA1* | Sambrook et al. (1989) |
| ***S. cerevisiae* strains** | | |
| Y294 | *α leu2*-3,112 *ura3*-52 *his3 trp1-289* | ATCC 201160 |
| Y294[amdSYM] | *URA3 TEF*$_P$*-amdS-TEF*$_T$ | This study |
| Y294[TemG_Opt-TemA_Nat] | *URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Opt-ENO1*$_T$ | Chapter 4 |
| Ethanol Red[1] | *MATa/α* prototroph | Fermentis, Lesaffre, France |
| M2n | *MATa/α* prototroph | Favaro et al. (2015) |
| Ethanol Red T1[2] | δ-integration of *ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| Ethanol Red T12[2] | δ-integration of *ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| M2n T1[2] | δ-integration of *ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| M2n T2[2] | δ-integration of *ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| **Plasmids** | | |
| yBBH1 | *bla URA3 ENO1*$_P$*-ENO1*$_T$ | Njokweni et al. (2012) |
| yBBH1-TemA_Nat[3] | *bla URA3 ENO1*$_P$*-temA_Nat-ENO1*$_T$ | Chapter 4 |
| yBBH1-TemG_Opt[4] | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$ | Chapter 4 |
| yBBH1-TemG_Opt-TemA_Nat | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | Chapter 4 |
| pUG-amdSYM[5] | *bla TEF*$_P$*-amdS-TEF*$_T$ | Solis-Escalante et al. (2013) |
| yBBH1-amdSYM | *bla URA3 TEF*$_P$*-amdS-TEF*$_T$ | This study |

[1]Ethanol Red™ Version 1, referred to as Ethanol Red
[2]Amylolytic transformants (T) contain integrated copies of *ENO1*$_P$*-temA_Nat-ENO1*$_T$ and *ENO1*$_P$*-temG_Opt-ENO1*$_T$ gene cassettes, the number indicates the transformant number during the screening process
[3]Accession no. XM_013469492 for the native *Talaromyces emersonii α-amylase* (*temG_Nat*)
[4]Accession no. AJ304803 for the native *T. emersonii glucoamylase* (*temG_Opt* encodes for the codon optimised gene)
[5]Assession no. P30669 for pUG-amdSYM plasmid

### 5.3.3 DNA manipulations

Standard protocols were followed for all DNA manipulations and *E. coli* transformations (Sambrook et al. 1989). The enzymes used for restriction digests and ligations were purchased from Inqaba Biotec (Pretoria, South Africa) and used as recommended by the supplier. Digested DNA was eluted from 0.8% agarose gels using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, California, USA). The *temA_Nat* and *temG_Opt* gene cassettes (*ENO1* promoter and terminator) (Fig. 5.1b) were amplified through polymerase chain reaction (PCR) using Delta-ENO1 primers (Table 5.2) together with the yBBH1-TemA_Nat and yBBH1-TemG_Opt vectors (**Chapter 4**), respectively, as template.

#### *5.3.3.1 Plasmid construction*

The *TEF*$_P$-*amdS*-*TEF*$_T$ gene cassette was amplified from pUG-amdSYM through PCR using the amdSYMCas primers (Table 5.2) and cloned onto yBBH1 using yeast mediated ligation (YML) yielding plasmid yBBH1-amdSYM (Fig. 5.1a). The *Ashbya gossypii TEF* promoter regulated the expression of the acetamidase-encoding gene (*amdS*) for the selection of transformants on SC-Ac plates. The yBBH1-amdSYM plasmid was retrieved from the *S. cerevisiae* Y294[amdSYM] strain and transformed into *E. coli* DH5α in order to obtain a high concentration of plasmid DNA. Plasmid DNA was isolated using the High Pure Plasmid Isolation kit (Roche, Mannheim, Germany) and sequence verification was performed by the dideoxy chain termination method, with an ABI PRISM™ 3100 Genetic Analyser (CAF, Stellenbosch University, South Africa).

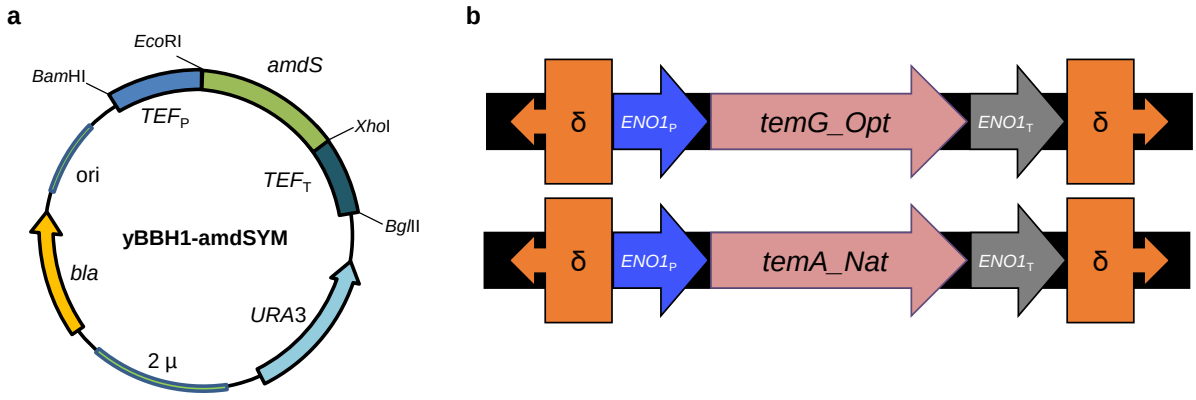


**Fig. 5.1** Schematic representation of the final vector and gene cassettes used in this study. The *TEF*$_P$-*amdS*-*TEF*$_T$ cassette was cloned onto yBBH1 (**a**) to generate the yBBH1-amdSYM expression vector. The *ENO1 temA_Nat* and *temG_Opt* gene cassettes (**b**) were amplified using PCR and contained flanking regions homologous to the δ-integration sites.

**Table 5.2**. PCR primers designed and used in this study with the relevant restriction sites underlined (*Eco*RI = gaattc; *Xho*I =ctcgag, *Bam*HI = ggatcc, *Bgl*II = agatct)

| Primer name | Sequence (5'-3') |
|---|---|
| amdSYMCas:L | ccgcgcgttggccgattcattaatccaggatccacatggaggcccagaataccctccttgac |
| amdSYMCas:R | gggcctcttcgctattacgccagagcttagatctcagtatagcgaccagcattcacatacttaa |
| Delta-ENO1p:L | tggaataaaaatccactatcgtctatcaactaatagttatattatcaatatattatcatatacggtgttaagatga tgacataagttatgagaagctgtcggatcccaattaatgtgagttacctcac |
| Delta-ENO1t:R | tgagatatatgtgggtaattagataattgttgggattccattgttgataaaggctataatattaggtatacagaat atactagaagttctcctcgaggatagatctcctatgcggtgtgaaataccgc |
| TemG_Opt:L | ttatcaacacacaaacactaaatcaaagaattcatggcctccttagtcgcaggtgcctta |
| TemG_Opt:R | gactagaaggcttaatcaaaagctctcgagtcattgccaagagtcgtccaagattgcggt |
| TemA_Nat:L | tgcttatcaacacacaaacactaaatcaaagaattcatgacgcctttcgtcctcacggcc |
| TemA_Nat:R | ggactagaaggcttaatcaaaagctctcgagctatctccatgtgtcgacaatcgtctccg |

### 5.3.4 Yeast transformations

Electro-competent *S. cerevisiae* Y294, Ethanol Red and M2n cells were prepared according to Cho et al. (1999) and transformed by means of electroporation using a Bio-Rad system (GenePluserXcell TM, Bio-Rad, Hercules, California, USA). For the transformation of industrial strains, amylase DNA (*temA_Nat* and *temG_Opt ENO1* linear DNA cassettes) were simultaneously transformed into the yeasts genomes using the yBBH1-amdSYM episomal vector which contained the selection marker (Fig. 5.1). After electroporation, 1 ml of YPDS was immediately added to the cuvettes. Cells were incubated at 30°C for 3 hours. Transformants were selected for by plating the transformation mix on to SC-Ac plates containing 2% starch (adapted from Solis-Escalante et al. 2013) and incubated at 30°C for 4 days. The integration of the linear expression cassette DNA into the yeast genome was confirmed by PCR using gene specific primers (Table 5.2).

### 5.3.5 Marker recycling

Plasmid curing was performed on the industrial recombinant strains as described by Solis-Escalante et al. (2013). The removal of the yBBH1-amdSYM containing the acetamide marker was achieved by growing cells overnight in 5 ml liquid YPD and transferring 20 µl to a 125 ml Erlenmeyer flask containing 10 mL SC-Fac. Marker-free single colonies were obtained by plating 100 µl of culture on SC-Fac solid media containing 2% starch and confirmed by colony PCR. The amylolytic strains' genomic DNA was isolated using the ZR fungal/bacterial DNA miniprep kit (Zymo Research, California, USA) and was then used as a template for real-time PCR.

### 5.3.6 Quantitative PCR

Oligo primers for real-time PCR were designed using IDT's PrimerQuest Tool (http://eu.idtdna.com/PrimerQuest/Home/Index). Special attention was given to primer length (18 – 22 bp), annealing temperature (58 – 62°C), base composition, 3′-end stability and amplicon size (75 – 100 bp). All primers were synthesised by Inqaba Biotech (Pretoria, South Africa) with reverse phase cartridge purification and are listed in Table 5.3. The performance of all primers was experimentally confirmed by conventional PCR to ensure that there was no formation of primer dimers and to confirm the amplification of a single region with the correct amplicon length.

**Table 5.3**. List of candidate reference genes and target genes including details of primers and amplicons for each gene

| Gene name | Amplicon length (bp) | Primers (5'-3') |
|---|---|---|
| *URA3* | 92 | L: cgtggatgatgtggtctctac<br>R:gttcaccctctaccttagcatc |
| *temA_Nat* | 100 | L: gcgatgtcactgagaggatcta<br>R: gaaatccagatggccgtgaa |
| *temG_Opt* | 95 | L: tacaggtggtttgggtgaac<br>R: ctctcaatgctggaccatctc |

Real-time PCR was carried out on a StepOne real time PCR instrument (Applied Biosystems) using white-walled PCR plates (96 wells). A × 2 KAPA HRM Fast Master Mix (containing a fast proof-reading polymerase, dNTPs, stabilisers and EvaGreen® dye) was used according to the manufacturer's instructions (KAPA Biosystems). Reactions were prepared in a total volume of 20 µl containing, 2.5 mM $MgCl_2$, 0.2 µM of each primer and 1 - 10 ng DNA. The cycle conditions were set as follows: initial template denaturation at 95°C for 30 seconds, followed by 45 cycles of denaturation at 95°C for 5 seconds and combined primer annealing/elongation at 60°C for 20 seconds and a final denaturation at 95°C for 1 minute to ensure all amplicons were fully melted. The yBBH1-TemG_Opt-TemA_Nat plasmid DNA was used to set up the standard curves (starting with $1\times10^7$ copies and making a 1:10 serial dilution) using primer pairs listed in Table 5.3. Genomic DNA concentrations were standardised to 10 ng for all samples. The PCR efficiency for each of the primer sets was calculated using StepOne software (Applied Biosystems). The number of copies of the *temG_Opt* and *temA_Nat* genes was calculated using the standard curve method described by Chen et al. (2012), using *URA3* as the reference gene.

### 5.3.7 Fermentations

The *S. cerevisiae* Y294 precultures were cultured in 100 ml 2×SC$^{-URA}$ medium in 500 ml Erlenmeyer flasks, whereas the *S. cerevisiae* Ethanol Red and M2n precultures were cultivated similarly in YPD medium. All precultures were incubated at 30°C with agitation at 200 rpms until stationary phase. The *S. cerevisiae* Y294 fermentations were performed in 2×SC$^{-URA}$ media, whereas the *S. cerevisiae* Ethanol Red and M2n fermentations were performed in YP starch media and inoculated with a 10% (v.v$^{-1}$) inoculum from the stationary preculture. Agitation and incubation were performed on a magnetic multi-stirrer platform (Velp Scientifica, Italy) at 30°C and 37°C, with daily sampling through a syringe needle pierced through the rubber stopper.

The exogenous GSHE cocktail used to supplement the fermentation process was STARGEN 002™ (now referred to as STARGEN) obtained from Dupont Industrial Biosciences (Palo Alto, California, USA) with an activity minimum of 570 GAU.gm$^{-1}$ (http://www.genencor.com) and used according to the manufactures instructions. STARGEN contained *Aspergillus kawachii* α-amylase expressed in *Trichoderma reesei* and a glucoamylase from *T. reesei* that work synergistically to hydrolyse granular starch to glucose (Huang et al. 2015).

#### *5.3.7.1 High performance liquid chromatography and analytical methods*

Ethanol, glucose, maltose, glycerol and acetic acid concentrations were quantified with high performance liquid chromatography (HPLC) using a Surveyor Plus liquid chromatograph (Thermo Scientific) consisting of a liquid chromatography pump, autosampler and refractive index (RI) detector. The compounds were separated on a Rezex RHM Monosaccharide 7.8 × 300 mm column (00H0132-K0, Phenomenex) at 60°C with 5 mM $H_2SO_4$ as mobile phase at a flow rate of 0.6 ml.min$^{-1}$. The theoretical $CO_2$ yields were calculated according to Favaro et al. (2015). The carbon conversion (percentage starch converted on a mol carbon basis) was calculated from ethanol, glucose, maltose, glycerol, acetic acid and $CO_2$ concentrations.

### 5.3.8 Statistical analysis

Data was analysed using the Student's t-test.

## 5.4 Results

The *T. emersonii temA_Nat* and *temG_Opt* genes encode for valuable amylase enzymes for use in the production of biofuel and are produced and secreted during cultivation on raw corn starch (**Chapter 4**). The linear *ENO1*$_P$-*temA_Nat-ENO1*$_T$ and *ENO1*$_P$-*temG_Opt-ENO1*$_T$ DNA gene cassettes (Fig. 5.1b), flanked by the δ sequence, were amplified and integrated into the δ-integration sites in the *S. cerevisiae* Ethanol Red and M2n industrial strains' genomes, in order to generate multi-copy integrants (Kim et al. 2011). The *amdS* gene was present on an episomal vector (Fig. 5.1a) to enable plasmid curing for easy recycling of the marker.

### 5.4.1 Industrial strain screening

The *S. cerevisiae* transformants were screened on SC plates containing 2% corn starch and transformants that produced zones of hydrolysis were selected for further testing (Fig. 5.2c). PCR was used to confirm the integration of both the *ENO1*$_P$-*temA_Nat-ENO1*$_T$ and *ENO1*$_P$-*temG_Opt-ENO1*$_T$ gene cassettes, respectively. The four strains showing the highest extracellular amylase activity were then evaluated under fermentative conditions (Fig. 5.2a and 5.2b). Significant differences in the carbon conversion were noted during the early stages of fermentation (Fig. 5.2b). However, after 192 hours, carbon conversion started to plateau, representing ~80% conversion of corn starch. The *S. cerevisiae* Ethanol Red T12 and M2n T1 strains hydrolysed starch and fermented the sugars quicker than the *S. cerevisiae* Ethanol Red T1 and M2n T2 strains (Fig. 5.2b and Table 5.4), therefore, they were selected for further evaluation under different fermentation conditions.

Plasmid curing of the amylolytic strains was performed by plating cultures onto SC-Fac plates containing 2% soluble corn starch. Quantitative PCR assays were performed using the genomic DNA from the cured amylolytic *S. cerevisiae* transformants, in order to determine the number of integrated copies of both *temA_Nat* and *temG_Opt* genes, respectively (Fig. 5.2d). The *S. cerevisiae* Ethanol Red T1, M2n T1 and M2n T2 strains contained single copies of *temA_Nat* and *temG_Opt* gene cassettes, whereas the *S. cerevisiae* Ethanol Red T12 contained 1 copy of *temA_Nat* and 2 copies of *temG_Opt*.

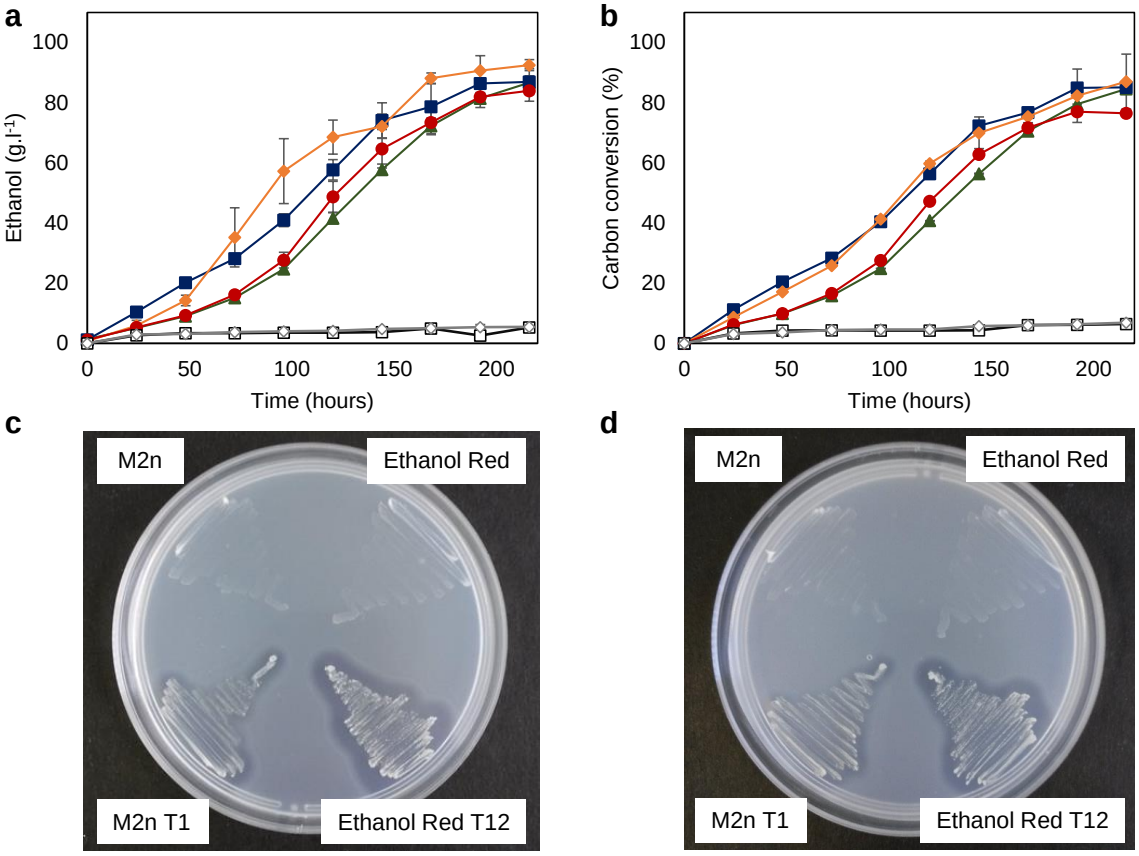


**Fig. 5.2** Comparison of industrial transformants after integration of *temA* and *temG* gene cassettes. Ethanol produced (**a**) and carbon conversion (percentage starch converted on a mol carbon basis) (**b**) displayed by *S. cerevisiae* Ethanol Red (-□-) and M2n (-○-) parental strains and *S. cerevisiae* Ethanol Red T1 (-▲-), T12 (-■-), M2n T1 (-◆-) and Mn2 T2 (-●-) transformants at a fermentation temperature of 30°C on 200 g.l$^{-1}$ raw corn starch. SC-Ac (**c**) and SC-Acr (**d**) plate assays confirmed the ability of recombinant *S. cerevisiae* Ethanol Red T12 and M2n T1 strains to utilise acetamide and acrylamide, respectively, whereas the parental *S. cerevisiae* Ethanol Red and M2n strains indicated no growth.

**Table 5.4**. Product formation by *S. cerevisiae* strains after 144 hours of fermentation at 30°C

| ***S. cerevisiae*** | **Ethanol Red T1** | **Ethanol Red T12** | **M2n T1** | **M2n T2** |
|---|---|---|---|---|
| **Substrate (g.l$^{-1}$)** | | | | |
| Raw starch weighed | 200 | 200 | 200 | 200 |
| Glucose weighed | 5 | 5 | 5 | 5 |
| Raw starch (dry weight) | 185 | 185 | 185 | 185 |
| Glucose equivalent | 208.5 | 208.5 | 208.5 | 208.5 |
| **Products (g.l$^{-1}$)** | | | | |
| Glucose | 0.82 | 0.67 | 0.60 | 0.72 |
| Glycerol | 2.39 | 3.40 | 1.92 | 2.29 |
| Acetic acid | 0.49 | 0.46 | 0.76 | 0.35 |
| Ethanol | 57.76 | 74.19 | 72.19 | 64.68 |
| Maltose | 0.99 | 1.09 | 1.01 | 1.08 |
| $CO_2$ | 55.25 | 70.94 | 69.05 | 61.87 |
| **Total** | 117.68 | 150.76 | 145.53 | 131.00 |
| Carbon conversion (%) | 56.44 | 72.31 | 69.80 | 62.83 |
| Ethanol yield[1] (% of theoretical yield) | 55.41 | 71.17 | 69.25 | 62.05 |
| Ethanol productivity[2] | 0.40 | 0.52 | 0.50 | 0.45 |

[1]Ethanol yield (% of the theoretical yield) was calculated as the amount of ethanol produced per gram of consumed glucose (at the specific time point)
[2]Ethanol productivity was calculated based on ethanol concentrations produced per hour (g.l$^{-1}$.h$^{-1}$)

The fermentation vigour of the amylolytic *S. cerevisiae* Ethanol Red T12 strain at 30°C and 37°C was compared to the laboratory *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain at 30°C (Fig. 5.3). The *S. cerevisiae* Ethanol Red T12 strain was able to ferment all the available glucose (Fig. 5.3b) at a fermentation temperature of 30°C and produced significantly less glycerol during the fermentation (Fig. 5.3d). This indicated a more efficient carbon conversion for ethanol (Bideaux et al. 2006). However, at a temperature of 37°C, ethanol levels did not increase significantly after 144 hours (Fig. 5.3a) and high levels of residual glucose were present (> 40 $g.l^{-1}$ after 264 hours). Maltose concentrations were similar at both fermentation temperatures (Fig. 5.3c).

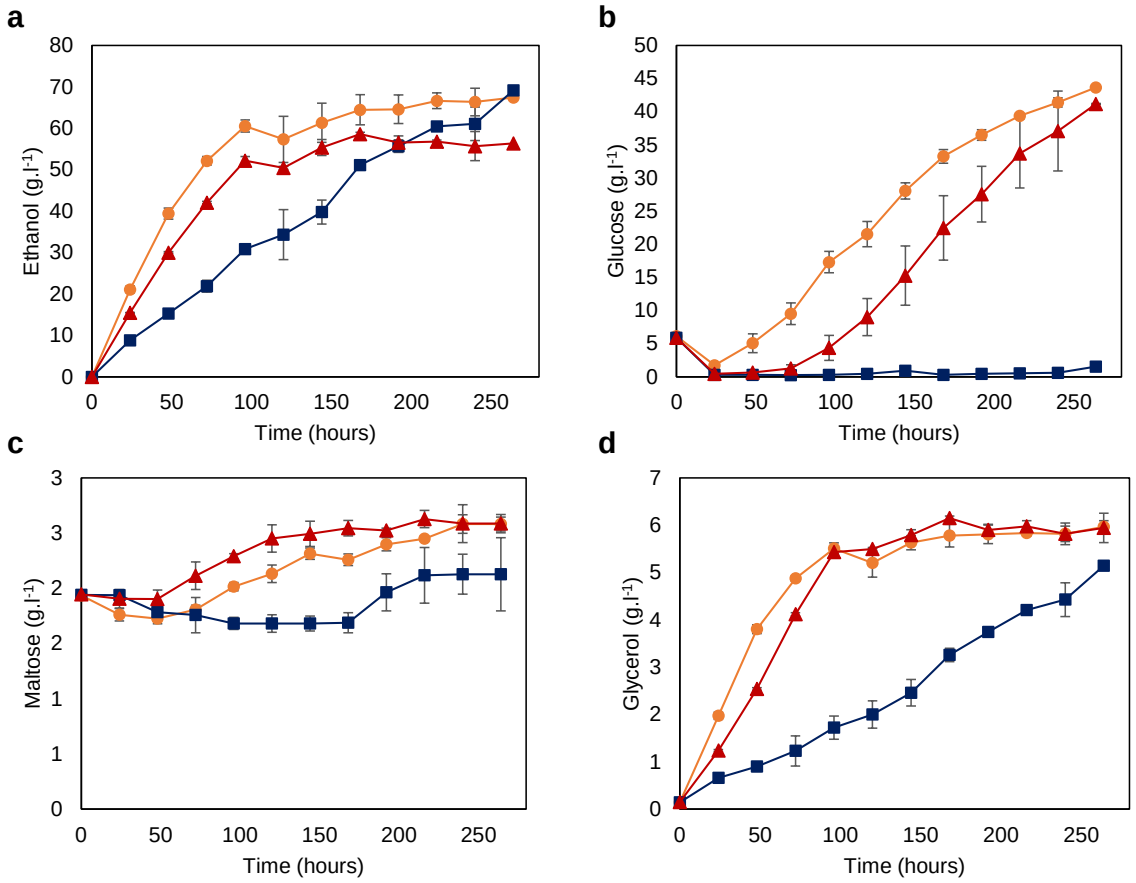


**Fig. 5.3** Comparison between the laboratory *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain (-●-) and the industrial amylolytic *S. cerevisiae* Ethanol Red T12 strain at 30°C (-■-) and 37°C (-▲-). The production of ethanol (**a**), glucose (**b**), maltose (**c**) and glycerol (**d**) were compared using 2×$SC^{-URA}$ fermentation media that contained 5 $g.l^{-1}$ glucose and 200 $g.l^{-1}$ raw corn starch. Data are the mean of 3 repeats showing standard deviation.

The evaluation of different media conditions (Fig. 5.4) was subsequently undertaken in order to determine whether buffered fermentation media (pH 5), the type of media (YP versus SC), or the addition of extra nitrogen (in the form of $(NH_4)_2SO_4$) could increase the efficiency of glucose fermentation by the *S. cerevisiae* Ethanol Red T12 strain at a fermentation temperature of 37°C.

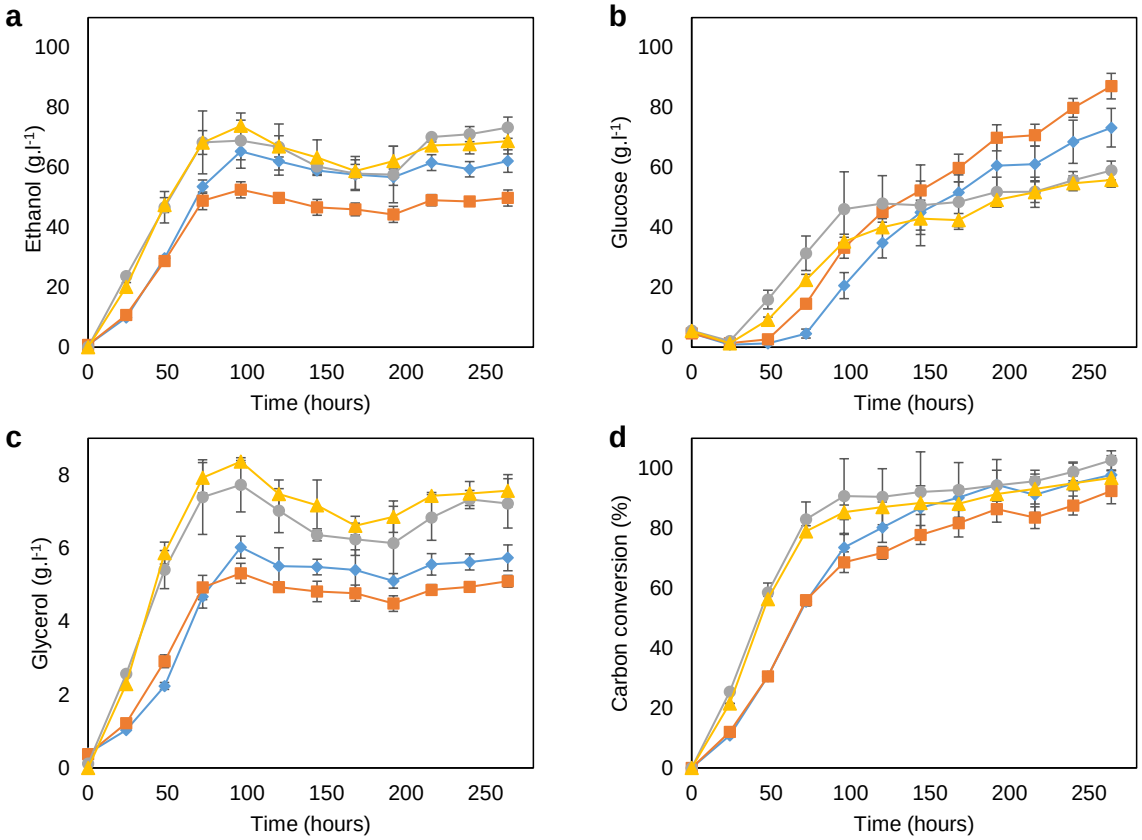


**Fig. 5.4** Different fermentation broth conditions during fermentation at 37°C on 200 g.l$^{-1}$ raw corn starch. *S. cerevisiae* Ethanol Red T12 in YP (-◆-), YP citrate-acid buffer pH 5 (-■-), SC citrate-acid buffer pH 5 (-●-) and SC citrate-acid buffer pH 5 with 10 g.l$^{-1}$ extra $(NH_4)_2SO_4$ (-▲-). Ethanol (**a**), glucose (**b**), glycerol concentrations (**c**) and carbon conversion (percentage starch converted on a mol carbon basis) (**d**) were compared. Data are the mean of 3 repeats showing standard deviation.

YP starch media (unbuffered) had a pH lower than 5 and this was more favourable for ethanol production, compared to the buffered YP broth (pH 5) (Fig. 5.4a). The addition of extra ammonium sulphate (10 g.l$^{-1}$) to the buffered SC fermentation broth did not increase the final ethanol concentrations or carbon conversion (Fig 5.4a and 5.4d), indicating sufficient nitrogen levels in the fermentation broth. Increased residual glucose concentrations were observed when YP media was used (Fig. 5.4b), while higher glycerol concentrations were noted when the fermentation was performed in SC media (Fig. 5.4c). SC media is less nutrient rich compared to YP media and more glycerol was produced to maintain the cytosolic redox balance under fermentation conditions (Bergman 2001). The higher glycerol concentrations contributed to the increased carbon conversion, especially during the first 120 hours of fermentation. Overall, the results in Fig. 5.4 showed that the media composition (SC vs YP and the pH) affected ethanol and glycerol production. However, changes in the type of media only affected the carbon conversion during the first 120 hours of fermentation. After

192 hours, the differences in carbon conversion were less apparent (between 92 – 100%).

### 5.4.2 Fermentations with STARGEN

The recommended STARGEN dosage was calculated as 1.4 µl.g$^{-1}$ starch, according to the manufacturer's specifications. The amylolytic S. *cerevisiae* Ethanol Red T12 and M2n T1 strains were compared to a simulated conventional SSF process (parental *S. cerevisiae* Ethanol Red/M2n strains + the full dosage of STARGEN) with a substrate loading of 200 g.l$^{-1}$ raw corn starch. Three different enzyme dosages were evaluated based on the percentage of the recommended enzyme loading: 2.8 µl (10%), 5.6 µl (20%) and 14 µl (50%) and compared to the SSF process, which had 28 µl STARGEN per 100 ml fermentation (representing 100% of the recommended dosage). The addition of exogenous enzymes significantly increased ethanol concentrations and enhanced the ethanol productivity (g.l$^{-1}$.h$^{-1}$) (Fig. 5.5 and 5.6), during the first 72 hours of fermentation.

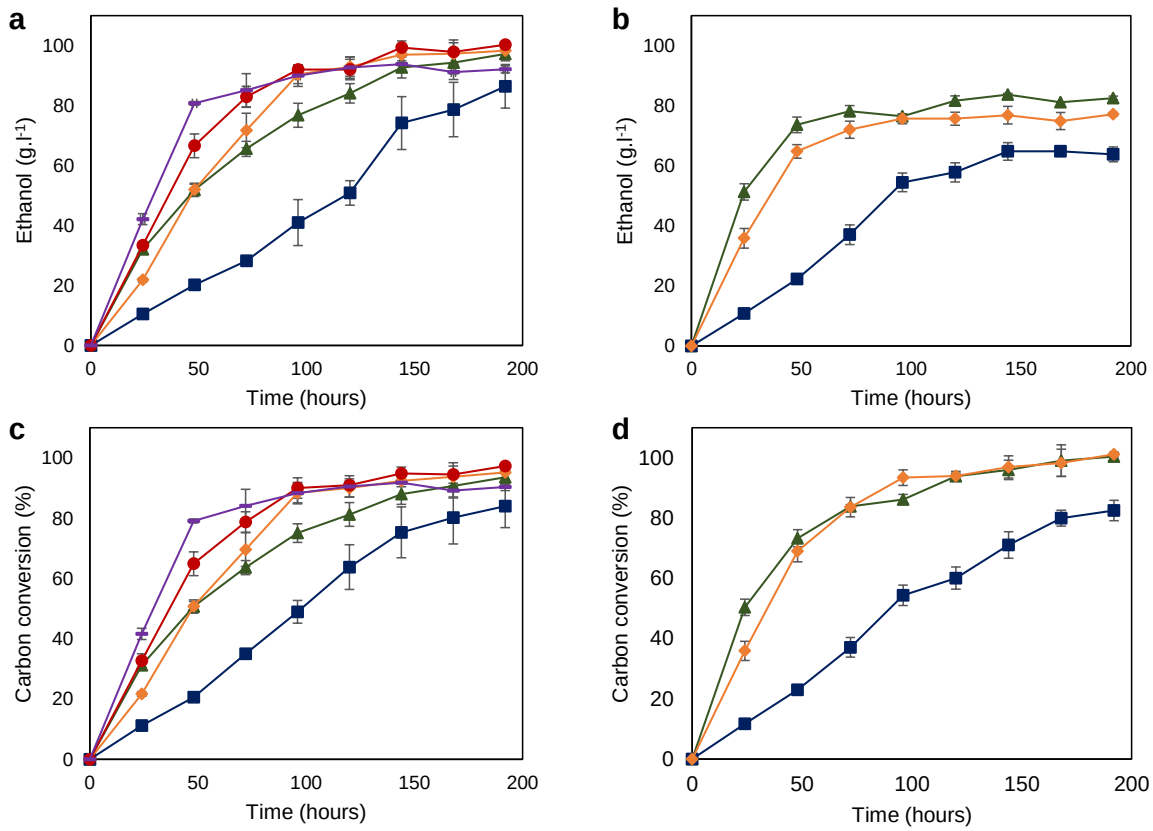


**Fig. 5.5** Ethanol concentrations produced by *S. cerevisiae* Ethanol Red strains during fermentation with 200 g.l$^{-1}$ corn starch at 30°C (**a**) and at 37°C (**b**), carbon conversion (percentage starch converted on a mol carbon basis) at 30°C (**c**) and at 37°C (**d**). Untransformed *S. cerevisiae* Ethanol Red + 28 µl STARGEN (-▲-), Ethanol Red T12 (-■-), Ethanol Red T12 + 2.8 µl STARGEN (-◆-), Ethanol Red T12 + 4.6 µl STARGEN (-●-) and Ethanol Red T12 + 14 µl STARGEN (-▬-). Data are the mean of 3 repeats showing standard deviation.

At a fermentation temperature of 30°C, the ethanol profiles for the *S. cerevisiae* Ethanol Red and M2n parental strains were similar for the respective condition (Fig. 5a and 6a, Table 5.5). By 48 hours, the *S. cerevisiae* Ethanol Red T12 strain supplemented with 2.8 µl STARGEN had produced the same amount of ethanol (52 $g.l^{-1}$) and showed the same carbon conversion (50%), compared to that of the control SSF process with untransformed *S. cerevisiae* Ethanol Red supplemented with 28 µl STARGEN (Fig. 5.5a and 5.5c). A similar trend was observed for the *S. cerevisiae* M2n T1 strain supplemented with 2.8 µl STARGEN, compared to the *S. cerevisiae* M2n parental strain (Fig. 5.6a and 5.6c).

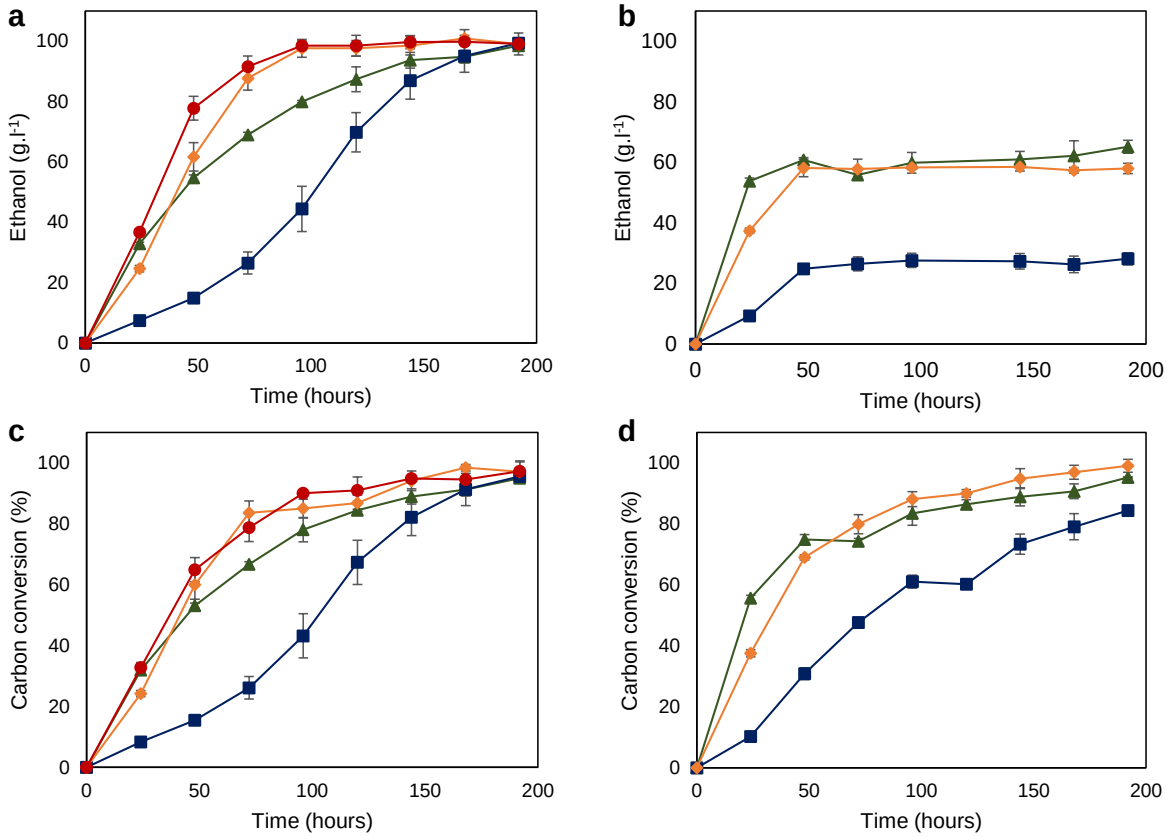


**Fig. 5.6** Ethanol concentrations produced by *S. cerevisiae* M2n strains during fermentation with 200 $g.l^{-1}$ at 30°C (**a**) and 37°C (**b**), carbon conversion (percentage starch converted on a mol carbon basis) at 30°C (**c**) and at 37°C (**d**). The untransformed *S. cerevisiae* M2n strain + 28 µl STARGEN (-▲-), M2n T1 (-■-), M2n T1 + 2.8 µl STARGEN (-◆-) and M2n T1 + 4.6 µl STARGEN (-●-). Data are the mean of 3 repeats showing standard deviation.

After 96 hours, ethanol produced by the *S. cerevisiae* Ethanol Red T12 strain supplemented with 2.8 µl STARGEN (90.4 $g.l^{-1}$) was similar to the amount of ethanol produced by the *S. cerevisiae* Ethanol Red T12 strain supplemented with 5.6 µl STARGEN (92.0 $g.l^{-1}$) (Fig. 5.5a). The carbon conversion displayed by these two strains was also similar (between 88 – 90%), at 96 hours (Fig. 5.5c). This represented a significant increase compared to the *S. cerevisiae* Ethanol Red control

strain, which produced 76.8 g.l$^{-1}$ ethanol and displayed a 75% carbon conversion after 96 hours. Therefore, the addition of 2.8 µl STARGEN (10% of the recommended dosage) was sufficient to produce results that were comparable to an SSF control. If the aim, however, is to decrease the fermentation time then higher dosages of STARGEN can be used in combination with the *S. cerevisiae* Ethanol Red T12 strain. An enzyme dosage representing 50% (14 µl of STARGEN) did not improve the final ethanol concentrations, however this dosage did result in a decreased fermentation time, with the maximum ethanol concentration being reached at 96 hours, instead of 192 hours. This was due to a higher ethanol productivity during the initial days of fermentation (Fig. 5.5a). Therefore, the use of the *S. cerevisiae* Ethanol Red T12 CBP strain can reduce the reliance on commercial enzyme used, as well as reduce the fermentation times.

Similar trends were observed for the *S. cerevisiae* M2n strains at a fermentation temperature of 30°C, compared to the *S. cerevisiae* Ethanol Red equivalent fermentations (Fig. 5.5a and 5.6a). However, the final ethanol concentration for the *S. cerevisiae* M2n T1 strain was significantly higher, > 10 g.l$^{-1}$ compared to *S. cerevisiae* Ethanol Red T12, after 192 hours ($p$ = 0.0392). At 30°C, the low residual levels of glucose and maltose in the fermentation broth (Table 5.5) indicated a rapid glucose uptake by all the strains.

**Table 5.5**. Product formation by *S. cerevisiae* Ethanol Red and M2n strains after 192 hours of fermentation at 30°C in YP media supplemented with different STARGEN dosages

| ***S. cerevisiae* strains** | **Ethanol Red** | **M2n** | **Ethanol Red T12** | **M2n T1** | **Ethanol Red T12** |
|---|---|---|---|---|---|
| STARGEN added (µl) | 28 | 28 | 2.8 | 2.8 | 5.6 |
| **Substrate (g.l$^{-1}$)** | | | | | |
| Raw starch weighed | 200 | 200 | 200 | 200 | 200 |
| Glucose weighed | 5 | 5 | 5 | 5 | 5 |
| Glucose equivalent | 208.5 | 208.5 | 208.5 | 208.5 | 208.5 |
| **Products (g.l$^{-1}$)** | | | | | |
| Glucose | 0.02 | 0.31 | 0.02 | 3.28 | 0.12 |
| Glycerol | 4.07 | 4.30 | 4.76 | 4.59 | 5.22 |
| Acetic acid | 0.00 | 0 | 0.90 | 0.31 | 0.96 |
| Ethanol | 97.23 | 98.49 | 98.37 | 99.08 | 100.32 |
| Maltose | 0.79 | 0.71 | 0.31 | 0.37 | 0.26 |
| $CO_2$ | 93.00 | 94.21 | 94.09 | 94.77 | 95.96 |
| **Total** | 195.11 | 198.02 | 198.44 | 202.40 | 202.85 |
| Carbon conversion (%) | 93.58 | 94.98 | 95.17 | 97.07 | 97.29 |
| Ethanol yield[1] (% of theoretical yield) | 93.26 | 94.48 | 94.36 | 95.04 | 96.23 |
| Ethanol productivity[2] | 0.51 | 0.51 | 0.51 | 0.52 | 0.52 |

[1]Ethanol yield (% of the theoretical yield) was calculated as the amount of ethanol produced per gram of consumed glucose (at a specific time point)

[2]Ethanol productivity was calculated based on ethanol concentrations produced per hour (g.l$^{-1}$.h$^{-1}$)

At 37°C, the *S. cerevisiae* Ethanol Red T12 strain had a higher ethanol tolerance and was able to ferment for longer (compared to the *S. cerevisiae* M2n T1 strain) producing a 2.3-fold increase in ethanol concentration at 192 hours (Fig. 5.5b and 5.6b). Although the recombinant *S. cerevisiae* M2n T1 strain produced more ethanol at 30°C, it was severely affected at a higher fermentation temperature (Fig.5.6a and 5.6b). At 37°C, the ethanol concentrations plateaued after 48 hours for all the *S. cerevisiae* M2n fermentations (Fig. 5.6b). The extent of carbon conversion displayed by the *S. cerevisiae* Ethanol Red T12 strain was similar (~83%) at the two fermentation temperatures (Fig. 5.5c and 5.5d), while the carbon conversion displayed by the *S. cerevisiae* M2n T1 strain was 13% higher at 30°C, compared to the carbon conversion at 37°C (Fig. 5.6c and 5.6d). Both the amylolytic *S. cerevisiae* Ethanol Red T12 and M2n T1 strains had lower ethanol productivity at 37°C, compared to at 30°C and residual glucose levels were > 40 $g.l^{-1}$ at 37°C (data not shown), which represented a large amount of unfermented glucose. Overall, results showed that temperature tolerance played a major role on the fermentation vigour of industrial *S. cerevisiae* Ethanol Red T12 and M2n T1 strains. The addition of STARGEN in combination with the amylolytic yeast strains reduced the fermentation time and increased the carbon conversion, compared to the control with untransformed strains and the recommended enzyme dosage.

## 5.5 Discussion

Starch-rich biomass is currently the main substrate for bioethanol production and can be efficiently hydrolysed by α-amylases and glucoamylases (Viktor et al. 2013). The *S. cerevisiae* Ethanol Red strain is a widely used industrial yeast, predominantly applied in first-generation bioethanol production from corn and wheat; it was the primary expression host used in this study. It is characterised by excellent fermentation capacity and yield, high robustness and stress tolerance (Demeke et al. 2013). The *S. cerevisiae* M2n strain is a South African distillery yeast and was used in this study for comparative purposes. The construction of a CBP yeast that can simultaneously express heterologous amylases and produce ethanol efficiently could yield more cost-effective ethanol production from starchy feedstocks.

Industrial strains can be used as the platform for heterologous amylase expression intended for first generation bioethanol production (Favaro et al. 2013). After the initial screening process, four recombinant strains expressing the *temG_Opt* and *temA_Nat*

gene cassettes (the *S. cerevisiae* Ethanol Red T1/T12 and *S. cerevisiae* M2n T1/T2 transformants) were selected for further evaluation (Fig. 5.2a). The *S. cerevisiae* M2n T1 strain performed better than the *S. cerevisiae* Ethanol Red T12 strain at 30°C and achieved a maximum ethanol concentration of 99.4 $g.l^{-1}$, which was 15% higher than the *S. cerevisiae* Ethanol Red T12 strain, after 192 hours (Fig. 5.5a and 5.6a). However, at 37°C is was clear that the *S. cerevisiae* Ethanol Red T12 strain had a greater fermentation vigour and was more tolerant to ethanol and higher temperatures, compared to the *S. cerevisiae* M2n T1 strain (Fig. 5.5b and 5.6b).

Results from this study showed significant improvements when compared to the industrial *S. cerevisiae* M2n[TLG1-SFA1] and MEL2[TLG1-SFA1] amylolytic strains (Favaro et al. 2015) that produced 64 $g.l^{-1}$ ethanol from 200 $g.l^{-1}$ raw corn starch, corresponding to 55% of the theoretical ethanol yield, as well as the *S. cerevisiae* Mnuα1[AmyA-GlaA] strain (Viktor et al. 2013) that produced 65.83 $g.l^{-1}$ ethanol (after 10 days) representing 57% of the theoretical ethanol yield. Ethanol yields (% of the theoretical) obtained from the recombinant industrial strains in this study were > 90% and thus represented a significant improvement on previously constructed amylolytic strains.

Final ethanol concentrations were also higher than those reported for the amylolytic haploid yeast strain, which produced 46.5 $g.l^{-1}$ of ethanol from 200 $g.l^{-1}$ of raw corn starch after 120 hours of fermentation (Yamakawa et al. 2012). The amylolytic yeast strains in this study were superior in their ethanol production, producing > 50 $g.l^{-1}$ and > 60 $g.l^{-1}$ ethanol for the *S. cerevisiae* Ethanol Red T12 and M2n T1 strains, respectively, after 120 hours (Fig. 5.5a and 5.6a). Furthermore, since the recombinant amylases were secreted into the fermentation broth they had increased physical contact with the starch granules, compared to recombinant yeast that displayed amylases on the cell's surface (Yamakawa et al. 2012). This eliminated potential bottlenecks and facilitated improved starch hydrolysis because the enzymes were able to penetrate into the granules and create pores more quickly.

During fermentation with the amylolytic *S. cerevisiae* Ethanol Red T12 and M2n T1 strains there was an initial "lag" phase in carbon conversion up until 48 hours (Fig. 5.5c and 5.6c). This was expected, since the strains first had to adjust to the fermentation conditions and produce amylases *de novo*. On the other hand, during the SSF process

with STARGEN (Fig. 5.5a and 5.6a) the enzymes were in abundance at the start of the fermentation and rapidly produced glucose upon addition. Therefore, although the *S. cerevisiae* Ethanol Red T12 and M2n T1 strains achieved a high carbon conversion (Fig. 5.2b), supplementation with STARGEN (Fig. 5.5 and 5.6) increased ethanol productivity at the start of the fermentation, especially at 30°C.

In the industrial cold hydrolysis set-up for bioethanol production, commercial amylase enzymes are only added at the beginning of the process and therefore their overall efficiency will decrease over time. However, the amylolytic CBP yeasts were able to continually replenish the recombinant enzymes in the fermentation broth and thus displayed an overall increase in carbon conversion, when the fermentation was supplemented with STARGEN (Fig. 5.5c/d and 5.6c/d). The cost of commercial enzyme addition has been estimated at 4.8 US cents per gallon, representing 8.3% of the total possessing costs in ethanol production from corn (Wong et al. 2010). The amylolytic *S. cerevisiae* Ethanol Red T12 and M2n T1 strains represented a novel alternative for lowering the enzyme dosage for raw starch hydrolysis, as well as being able to provide constant amylolytic activity for a continuous cold fermentations process. Furthermore, the use of amylolytic CBP yeast would allow for a simplified fermentation design, since pretreatment steps and costs can be bypassed (Salehi Jouzani and Taherzadeh, 2015).

There are a number of factors commonly associated with a stuck fermentation, including the yeast strain, nitrogen availability and glucose concentration (Henderson and Block, 2014). However, fermentation temperature is considered as one of the main bottlenecks with regards to ethanol production by SSF and CBP strategies. Fig. 5.4 showed the performance of the *S. cerevisiae* Ethanol Red T12 strain in different fermentation media and results confirmed that extra nitrogen (in the form of $(NH_4)_2SO_4$)) did not increase the fermentation of glucose to ethanol, at a temperature of 37°C. Furthermore, increasing the pH of the conventional YP fermentation medium (to pH 5) did not improve fermentation conditions. Therefore, a lower pH was more favourable for starch conversion when using the TemG_Opt and TemA_Nat enzymes from *T. emersonii*, which have a pH optimum around 4 - 4.5 (Nielsen et al. 2002).

The demand for higher temperature fermentations began in the 1980s (Abdel-Banat et al. 2010). High-temperature fermentations may assist in making the

simultaneous fermentation and ethanol extraction process more suitable for fuel ethanol production, decrease operational costs (especially in regions with hot climates where cooling of fermentation vessels is required), improve hydrolysis conditions and reduce the risk of contamination (Banat et al. 1998). Currently, the fermentation temperatures used in industry are between 30 - 34°C (Mukhtar et al. 2010). However, the internal temperature of a fermentation vessel exceeds these temperatures due to exothermic metabolic activities, as well as environmental temperatures in higher-temperature regions. Subsequently, the efficiency of ethanol production can be severly decreased if the yeast is unable to continue fermenting at temperatures higher than 34°C (in the presence of high ethanol concentrations).

The effect of temperature on fermentation products has been described by a number of different research groups (Favaro et al. 2013b; Woo et al. 2014). Although industrial strains of *S. cerevisiae* are known for their high ethanol tolerance and relatively high ethanol concentrations, many of these strains still lack the ability to continue fermenting glucose at temperatures that are higher than their normal growth temperature (~30-34°C depending on the strain) (Fig. 5.5b). Moreover, ethanol concentrations of approximately 10% ($wt.vol^{-1}$) will reduce the fermentative activity of yeast by approximately 50% (Henderson and Block 2014) and inhibit cell growth and viability. This leads to lower productivity and lower ethanol yields (Stanley et al. 2010). In order to improve yeasts' ethanol tolerance, the understanding of the cellular impact of ethanol toxicity needs to be explored.

Results for the comparison of ethanol production by recombinant *S. cerevisiae* Y294 and Ethanol Red T12 strains (Fig. 5.3) were in agreement with a study by Favaro et al. (2013b). They showed that at 30°C the laboratory *S. cerevisiae* Y294 strain had a lower fermentation vigour compared to the industrial strain at 30°C. The decreased ability to consume glucose could be explained by the *S. cerevisiae* Y294 strain displaying an optimum cultivation temperature around 25°C not 30°C. Similarly, the amylolytic *S. cerevisiae* Ethanol Red T12 strain had reduced fermentation vigour at 37°C, compared to at 30°C (Fig. 5.5b).

Reduced glycerol concentrations were observed when lower fermentation temperatures were used, indicating that better carbon conversion to ethanol occurred at a fermentation temperature of 30°C, compared to 37°C (Fig. 5.3d). Carbon source

utilisation was important for the optimisation of ethanol production (Navarrete et al. 2014) and results showed that the fermentation media influenced glycerol production (Fig. 5.4). The commercially available TransFerm™ Yield+ yeast (Mascoma and Lallemand Biofuels and Distilled Spirts) was engineered to produce significantly less glycerol during fermentations, so that more carbon can be utilised for ethanol production. In this study, the accumulating glycerol concentrations were below the conventional (10 $g.l^{-1}$) (Huang et al. 2015) and therefore would not have had a significant effect on the yeast cells.

In recent years more attention has been given to acrylamide, which is considered a potential carcinogen (Liu et al. 2013). The occurrence of acrylamide is widespread and is often produced by industrial processes and during the cooking of carbohydrate foods at high temperatures (Charoenpanich 2013). One of the main pathways for acrylamide formation in food is via the Maillard reaction (Zhang et al. 2007). Although, several microorganisms have been reported as acrylamide degraders e.g. *Bacillus*, *Pseudomonas*, *Rhodococcus* and *Aspergillus* (Hynes 1970; Charoenpanich 2013), few studies have investigated the use of these microorganisms in the food and beverage industries (Wakaizumi et al. 2009).

*S. cerevisiae* has GRAS (generally regarded as safe) status and therefore represents an ideal candidate for the reduction of acrylamide. The amylolytic strains constructed in this study were transformed using the *amdS* marker, which enabled the strains to use acrylamide in addition to acetamide as the sole nitrogen sources (Fig. 5.2). To our knowledge, this is the first report of an acrylamide utilising amylolytic *S. cerevisiae* strain. The use of recombinant *S. cerevisiae* strains would be a novel approach for acrylamide degradation. In addition to the biofuel industry, these recombinant strains (or the spent yeast) could be used for other applications in a number of different industries (e.g. food and beverage), for biowaste and wastewater treatment, as well as in bioaugmentation strategies aimed at acrylamide-contaminated soil and wastewater (Liu et al. 2013).

## 5.6 Conclusion

Few studies have engineered *S. cerevisiae* Ethanol Red for the expression of gene cassettes or adapted it for desired characteristics. Demeke et al. (2013b) developed a D-xylose fermenting strain, Wallace-Salinas and Gorwa-Grauslund (2013) developed a strain capable of fermenting spruce hydrolysate and Stovicek et al. (2015b) introduced a xylose consumption pathway. To our knowledge, this study represented the first to engineer *S. cerevisiae* Ethanol Red for the co-expression of both an α-amylase and glucoamylase gene for efficient raw corn starch conversion. It also represented the first study to investigate the effects of STARGEN supplementation in combination with an amylolytic CBP yeast (the commercial TransFerm® yeast only expresses a glucoamylase).

Improved carbon conversion of raw corn starch was achieved in this study, however a fermentation temperature of 30°C enabled higher ethanol concentrations, compared to fermentations at 37°C. The amylolytic *S. cerevisiae* Ethanol Red T12 and M2n T1 strains represented a drop-in CBP yeast substitute for the existing cold fermentation process that is currently used by major ethanol producers such as POET (Sioux Falls, South Dakota, USA). Although high-temperature fermentations are more practical for industrial ethanol production, results showed that ethanol tolerance/thermotolerance are still the main bottlenecks with regards to constructing CBP yeast for the industrial production of bioethanol. Therefore, future studies aimed at ethanol tolerance are required in order to engineer an amylolytic CBP yeast that can ferment at higher temperatures.

## 5.7 Acknowledgments

The authors would like to thank Mrs Lisa Warburg for her assistance with HPLC analysis and the National Research Foundation (South Africa) for financial support.

## 5.8 References

Abdel-Banat BMA, Hoshida H, Ano A, Nonklang S, Akada R (2010) High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast? Appl Microbiol Biotechnol 85:861-867. doi: 10.1007/s00253-009-2248-5

Balcerek M, Pielech-Przybylska K (2013) Effect of simultaneous saccharification and fermentation conditions of native triticale starch on the dynamics and efficiency of process and composition of the distillates obtained. J Chem Technol Biotechnol 88:615–622. doi: 10.1002/jctb.3873

Banat IM, Nigam P, Singh D, Marchant R, McHale AP (1998) Ethanol production at elevated temperatures and alcohol concentrations: Part I – Yeasts in general. World J Microbiol Biotechnol 14:809-821. doi: 10.1023/A:1008802704374

Bergman LW (2001) Growth and maintenance of yeast. Methods Mol Biol 177:9-14. doi:10.1385/1-59259-210-4:009.

Bideaux C, Alfenore S, Cameleyre X, Molina-Jouve C, Uribelarrea JL, Guillouet SE (2006) Minimization of glycerol production during the high-performance fed-batch ethanolic fermentation process in *Saccharomyces cerevisiae*, using a metabolic model as a prediction tool. Appl Environ Microbiol 72:2134–2140. doi: 10.1128/AEM.72.3.2134-2140.2006

Bothast RJ, Schlicher MA (2005) Biotechnological processes for conversion of corn into ethanol. Appl Microbiol Biotechnol 67:19–25. doi:10.1007/s00253-004-1819-8

Brehmer B, Bals B, Sanders J, Dale B (2008) Improving the corn-ethanol industry: Studying protein separation techniques to obtain higher value-added product options for distillers grains. Biotechnol Bioeng 101:49–61. doi: 10.1002/bit.21881

Charoenpanich J (2013) Removal of acrylamide by microorganisms. Applied Bioremediation - Active and Passive Approaches. doi: 10.5772/56150

Chen Y, Partow S, Scalcinati G, Siewers V, Nielsen J (2012) Enhancing the copy number of episomal plasmids in *Saccharomyces cerevisiae* for improved protein production. FEMS Yeast Res 12:598-607. doi:10.1111/j.1567-1364.2012.00809.x

Cho KM, Yoo YJ, Kang HS (1999) δ-Integration of endo/exoglucanase and β-glucosidase genes into the yeast chromosomes for direct conversion of cellulose to ethanol. Enzyme Microb Technol 25:23–30. doi: 10.1016/S0141-0229(99)00011-3

Demeke MM, Dumortier F, Li Y, Broeckx T, Foulquié-Moreno MR, Thevelein JM (2013) Combining inhibitor tolerance and D-xylose fermentation in industrial *Saccharomyces cerevisiae* for efficient lignocellulose-based bioethanol production. Biotechnol Biofuels 6:120. doi: 10.1186/1754-6834-6-120

Demeke MM, Dietz H, Li Y, Foulquié-Moreno MR, Mutturi S, Deprez S, Den Abt T, Bonini BM, Liden G, Dumortier F, Verplaetse A, Boles E, Thevelein JM (2013b) Development of a D-xylose fermenting and inhibitor tolerant industrial *Saccharomyces cerevisiae* strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. Biotechnol Biofuels 6:89. doi: 10.1186/1754-6834-6-89

den Haan R, Kroukamp H, Mert M, Bloom M, Görgens JF, van Zyl WH (2013) Engineering *Saccharomyces cerevisiae* for next generation ethanol production. J Chem Technol Biotechnol 88:983–991. doi: 10.1002/jctb.4068

den Haan R, van Rensburg E, Rose SH, Görgens J, van Zyl WH (2015) Progress and challenges in the engineering of non-cellulolytic microorganisms for consolidated bioprocessing. Curr Opin Biotechnol 33:32–38. doi: 10.1016/j.copbio.2014.10.003

Favaro L, Jooste T, Basaglia M, Rose SH, Saayman M, Görgens JF, Casella S, van Zyl WH (2013) Designing industrial yeasts for consolidated bioprocessing of starchy biomass to ethanol. Bioengineered 4:97–102. doi: 10.4161/bioe.22268

Favaro L, Basaglia M, Trento A, van Rensburg E, García-Aparicio M, van Zyl WH, Casella S (2013b) Exploring grape marc as trove for new thermotolerant and inhibitor-tolerant *Saccharomyces cerevisiae* strains for second-generation bioethanol production. Biotechnol Biofuels 6:168. doi: 10.1186/1754-6834-6-168

Favaro L, Viktor M, Rose S, Viljoen-Bloom M, van Zyl W, Basaglia M, Cagnin L, Casella S (2015) Consolidated bioprocessing of starchy substrates into ethanol by industrial *Saccharomyces cerevisiae* strains secreting fungal amylases. Biotechnol Bioeng 112:1751-1760. doi: 10.1002/bit.25591

Görgens JF, Bressler DC, van Rensburg E (2015) Engineering *Saccharomyces cerevisiae* for direct conversion of raw, uncooked or granular starch to ethanol. Crit Rev Biotechnol 35:396-391. doi: 10.3109/07388551.2014.888048

Henderson CM, Block DE (2014) Examining the role of membrane lipid composition in determining the ethanol tolerance of *Saccharomyces cerevisiae*. Appl Environ Microbiol 80:2966–2972. doi: 10.1128/AEM.04151-13

Huang H, Qureshi N, Chen MH, Liu W, Singh V (2015) Ethanol production from food waste at high solids content with vacuum recovery technology. J Agric Food Chem 63:2760–2766. doi: 10.1021/jf5054029

Hynes MJ (1970) Induction and repression of amidase enzymes in *Aspergillus nidulans*. J Bacteriol 103:482–487. PMCID: PMC248107

Kim HR, Im YK, Ko HM, Chin, JE, Kim IC, Lee HB, Bai S (2011) Raw starch fermentation to ethanol by an industrial distiller's yeast strain of *Saccharomyces cerevisiae* expressing glucoamylase and α-amylase genes. Biotechnol Lett 33:1643–1648. doi: 10.1007/s10529-011-0613-9

Liu ZH, Cao YM, Zhou, QW, Guo K, Ge F, Hou JY, Hu SY, Yuan S, Dai YJ (2013) Acrylamide biodegradation ability and plant growth-promoting properties of *Variovorax boronicumulans* CGMCC 4969. Biodegradation 24:855–864. doi: 10.1007/s10532-013-9633-6

Mobini-Dehkordi M, Javan FA (2012) Application of alpha-amylase in biotechnology. J Biol Today's World 1:15–20. doi: 10.15412/J.JBTW

Mukhtar K, Asgher M, Afghan S, Hussain K, Zia-Ul-Hussnain S (2010) Comparative study on two commercial strains of *Saccharomyces cerevisiae* for optimum ethanol production on industrial scale. J Biomed Biotechnol 210:419586. doi: 10.1155/2010/419586

Navarrete C, Nielsen J, Siewers V (2014) Enhanced ethanol production and reduced glycerol formation in *fps1*Δ mutants of *Saccharomyces cerevisiae* engineered for improved redox balancing. AMB Express 4:86. doi: 10.1186/s13568-014-0086-z

Nielsen BR, Lehmbeck J, Frandsen TP (2002) Cloning, heterologous expression, and enzymatic characterization of a thermostable glucoamylase from *Talaromyces emersonii*. Protein Expr Purif 26:1–8. doi: 10.1016/S1046-5928(02)00505-3

Nigam P, Singh D (1995) Enzyme and microbial systems involved in starch processing. Enzyme Microb Technol 17:770-778. doi:10.1016/0141-0229(94)00003-A

Njokweni A, Rose SH, van Zyl WH (2012) Fungal β-glucosidase expression in *Saccharomyces cerevisiae*. J Mol Microbiol Biotechnol 39:1445–1452. doi: 10.1007/s10295-012-1150-9

Nkomba EY, van Rensburg E, Chimphango AF, Görgens JF (2016) The influence of sorghum grain decortication on bioethanol production and quality of the distillers' dried grains with solubles using cold and conventional warm starch processing. Bioresour Technol 203:181–189. doi: 10.1016/j.biortech.2015.12.045

Pandey A, Nigam P, Soccol CR, Soccol VT, Singh D, Mohan R (2000) Advances in microbial amylases. Biotechnol Appl Biochem 31:135-152. PMID: 10744959.

Rana N, Walia A, Gaur A (2013) α-Amylases from microbial sources and its potential applications in various industries. Natl Acad Sci Lett 36:9–17. doi: 10.1007/s40009-012-0104-0

Robertson GH, Wong DW, Lee CC, Wagschal K, Smith MR, Orts WJ (2006) Native or raw starch digestion: a key step in energy efficient biorefining of grain. J Agric Food Chem 54:353-365. doi: 10.1021/jf051883m

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press

Salehi Jouzani G, Taherzadeh MJ (2015) Advances in consolidated bioprocessing systems for bioethanol and butanol production from biomass: a comprehensive review. Biofuel Res J 5:152–195. doi: 10.18331/BRJ2015.2.1.4

Sivaramakrishnan S, Gangadharan D, Nampoothiri K, Soccol CR, Pandey A (2006) α-Amylases from microbial sources - an overview on recent developments. Food Technol Biotechnol 44:173-184. ISSN: 1330-9862

Solis-Escalante D, Kuijpers NGA, Bongaerts N, Bolat I, Bosman L, Pronk JT, Daran JM, Daran-Lapujade P (2013) amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. FEMS Yeast Res 13: 126–139. doi: 10.1111/1567-1364.12024

Stanley D, Bandara A, Fraser S, Chambers PJ, Stanley GA (2010) The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*. J Appl Microbiol 109:13–24. doi: 10.1111/j.1365-2672.2009.04657.x

Stovicek V, Borodina I, Forster J (2015) CRISPR–Cas system enables fast and simple genome editing of industrial *Saccharomyces cerevisiae strains*. Metab Eng Commun 2:13–22. doi: 10.1016/j.meteno.2015.03.00

Stovicek V, Borja GM, Forster J, Borodina I (2015b) EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial *Saccharomyces cerevisiae* strains. J Ind Microbiol Biotechnol 42:1519-1531. doi: 10.1007/s10295-015-1684-8

Szymanowska-Powałowska D, Lewandowicz G, Kubiak P, Błaszczak W (2014) Stability of the process of simultaneous saccharification and fermentation of corn flour. The effect of structural changes of starch by stillage recycling and scaling up of the process. Fuel 119:328–334. doi.org/10.1016/j.fuel.2013.11.034

van Zyl WH, Bloom M, Viktor MJ (2012) Engineering yeasts for raw starch conversion. Appl Microbiol Biotechnol 95:1377-1388. doi: 10.1007/s00253-012-4248-0

Viktor MJ, Rose SH, van Zyl WH, Viljoen-Bloom M (2013) Raw starch conversion by *Saccharomyces cerevisiae* expressing *Aspergillus tubingensis* amylases. Biotechnol Biofuels 6:167. doi: 10.1186/1754-6834-6-167

Wakaizumi M, Yamamoto H, Fujimoto N, Ozeki K (2009) Acrylamide degradation by filamentous fungi used in food and beverage industries. J Biosci Bioeng 108:391-393. doi: 10.1016/j.jbiosc.2009.05.004

Wallace-Salinas V, Gorwa-Grauslund MF (2013) Adaptive evolution of an industrial strain of *Saccharomyces cerevisiae* for combined tolerance to inhibitors and temperature. Biotechnol Biofuels 6:151. doi: 10.1186/1754-6834-6-151

Wong D, Batt Throne SB, Robertson GH, Lee CC, Wagschal KC (2010) Chromosomal integration of both an alpha-amylase and a glucoamylase gene in *Saccharomyces cerevisiae* for starch conversion. Ind Biotechnol 6:112–119. doi: 10.1089/ind.2010.0010.

Woo JM, Yang KM, Kim SU, Blank LM, Park JB (2014) High temperature stimulates acetic acid accumulation and enhances the growth inhibition and ethanol production by *Saccharomyces cerevisiae* under fermenting conditions. Appl Microbiol Biotechnol 98:6085–6094. doi: 10.1007/s00253-014-5691-x

Yamakawa S, Yamada R, Tanaka T, Ogino C, Kondo A (2012) Repeated fermentation from raw starch using *Saccharomyces cerevisiae* displaying both glucoamylase and α-amylase. Enzyme Microb Techno 50:343–347. doi: 10.1016/j.enzmictec.2012.03.005

Zhang Y, Ren Y, Zhao H, Zhang Y (2007) Determination of acrylamide in Chinese traditional carbohydrate-rich foods using gas chromatography with micro-electron capture detector and isotope dilution liquid chromatography combined with electrospray ionization tandem mass spectrometry. Anal Chim Acta 584:322–332. doi: 10.1016/j.aca.2006.10.061

# Chapter 6

General discussion and conclusions

# General discussion and conclusions

## 6.1 Heterologous gene expression

Recent advances in synthetic biology and metabolic engineering have contributed to advanced biofuel production becoming more economically feasible (Dellomonaco et al. 2010). The production of fuels and chemicals from renewable sources falls under a sub-group of biotechnology known as white biotechnology (also referred to as industrial biotechnology) and involves the use of living systems for the production of value added products. The framework for this study combined these research fields and employed recombinant DNA strategies for improved amylase production by yeast. This enabled the efficient conversion of raw starch to ethanol, using industrial amylolytic *Saccharomyces cerevisiae* strains.

Molecular biology has provided the necessary tools needed to optimise gene sequences encoding for industrially important enzymes that are suitable for use in biofuel production (Connor and Atsumi 2010; Elena et al. 2014). Subsequently, an assortment of software programs have been developed to assist in the reverse translation of proteins in order to optimise coding sequences for gene expression, but there is still a lot to learn when it comes to the conceptual design of gene sequences (Wu et al. 2007). Numerous studies have explored codon optimisation strategies (Hoekema et al. 1987; Welch et al. 2009; Chung and Lee 2012; Lanza et al. 2014), since codon bias affects gene expression (Henry and Sharp 2007). Results from this study have emphasised the need for a systematic analysis of the relationship between synonymous codon usage and gene expression. Therefore, future research is needed in order to develop efficient platforms for enhanced gene expression in industrial *S. cerevisiae* strains and to refine design algorithms used for gene optimisation.

Differences in codon usage is one of the major factors affecting gene expression levels (Elena et al. 2014). However, there is an ongoing debate over what factors are the most important for improving heterologous gene expression. Moreover, the expression of fungal proteins in yeast (eukaryotic organism) is fundamentally different to the expression of human proteins in *Escherichia coli* (a prokaryote). Therefore, the guidelines that apply to the one scenario may not be applicable/suitable to another. Furthermore, fungal and other foreign genes do not have the same codon choice

patterns as yeast, but rather a codon bias typical to their origin that may negatively affect gene expression in a recombinant host (Hoekema et al. 1987).

Synthetic genes offer several advantages when compared to the cloning of genomic DNA (Presnell and Benner 1988). Firstly, it is easier to obtain protein sequences as opposed to the corresponding native DNA sequence and secondly, codon usage can be adapted (using synonymous codons) and "optimised" for a specific expression host. Generally, synthetic genes encode for the same protein as the gene of interest, but the synthetic nucleotide sequence may contain a number of modifications depending on the requirements for expression (e.g. manipulation of restriction sites, glycosylation sites and adaption of GC content) (Richardson et al. 2006). Yet, few studies have evaluated and compared codon optimised genes to that of their native counterparts (in terms of the resulting extracellular activity), thus resulting in a lack of experimentally supported design principles (Wu et al. 2007).

It is difficult to draw conclusions from studies that have expressed codon optimised genes because the examples in literature differ in many aspects (type of optimisation algorithm, expression host, etc.) and most often only successful optimisation experiments are published (Welch et al. 2009). Thus, there remains a need to understand how genomic codon usage bias has evolved to regulate the expression of genes (Angov 2011). The choice of promoters and secretion signal peptides are also important elements for optimising heterologous protein production by *S. cerevisiae*.

During **Chapter 4** of this study, different gene variants that encoded for the same mature protein were evaluated. Subsequently, we were able to compare the expression of several native fungal genes to their respective codon optimised counterparts. The genes that encoded for the best enzyme combination originated from *Talaromyces emersonii*; this combination included a native α-amylase (*temA_Nat*), whereas the glucoamylase was encoded for by the codon optimised gene variant (*temG_Opt*). Thus even though the genes originated from the same host organism, the specific gene variant that performed the best was not the same. Therefore, the rational approach taken in this study has contributed to the understanding of synonymous codon usage. It also highlighted the importance of comparing native and codon optimised gene expression in *S. cerevisiae*.

## 6.2 Amylase production in yeast

Given that only a small percentage of amylases are able to hydrolyse raw starch, there have been very few studies that have characterised and investigated these raw starch hydrolysing enzymes. Raw starch amylases generally have lower percentages of enzymatic saccharification, compared to soluble starch hydrolysing enzymes. This is as a result of several factors that may impeded hydrolysis, namely substrate interactions, starch composition, imperfect docking, as well as poor adsorption of the enzymes to the granule's surface. Although the exploration of amylases for raw starch hydrolysis at fermentation temperatures has been investigated, only a limited number of studies have had success in engineering yeast for the one-step conversion of starch using the CBP (consolidated bioprocessing) concept.

The technology for the *in vivo* homologous recombination of genes in yeast is well established, with various approaches aimed at integrating heterologous genes into the yeast genome (Shao et al. 2009; Favaro et al. 2013). The reiterated DNA sequences such as δ-sequences of the Ty retrotransposon and ribosomal DNA are commonly used target sites, which allow for multiple integrations of gene copies. Recently, research has focused on advancing the genetic engineering toolbox for the manipulation of industrial strains, by exploring the integration of genes into the yeast's genome in a single step (Stovicek et al. 2015).

Industrial yeast strains are robust and preferred over laboratory strains for their fermenting ability. However, numerous technical challenges arise when engineering industrial strains for commercial applications and transformation efficiencies are relatively low (Le Borgne 2012). Low copy numbers using δ-mediated DNA integration are common (Ekino et al. 2002) and heterologous gene expression using this integration method is affected by the ploidy of the strains. Therefore, in order to improve the transformation efficiency and increase the gene copy number, alternative integration methods need to be explored e.g. CRISPR (clustered regularly interspaced short palindromic repeats)–Cas9 based system for gene integration (Jessop-Fabre et al. 2016).

The industrial yeast strains (*S. cerevisiae* Ethanol Red™ and M2n) were more robust compared to the laboratory *S. cerevisiae* Y294 strain at 30°C. Yet, despite their increased thermotolerance, a fermentation temperature of 37°C was too high for optimal ethanol production. In addition to thermotolerance, there are several factors

that can decrease the efficiency of a fermentation process many of which create a stressful environment for the fermenting yeast. The effect of high temperature is intensified by ethanol concentrations that exceed 3% ($w.v^{-1}$) and this affects the yeast cell's membrane causing protein denaturation. Furthermore, when the cultivation temperature increases above the optimum growth temperature, the specific glucose uptake of *S. cerevisiae* is affected by changes to the physiology of the yeast cells (Woo et al. 2014).

The fermentation results from this study clearly highlighted the effect of temperature on the yeast's ability to ferment glucose. Ethanol tolerance is a complex phenotype (Snoek et al. 2015) and since there are several industrial and economic advantages to producing ethanol at higher temperatures (Mukhtar et al. 2010), it is important to understand the factors influencing ethanol tolerance and what physiology processes are being affected. An understanding of the fermentation kinetics at higher temperatures (> 30°C for industrial strains) and the subsequent effect on sugar transporters will help elucidate methods for improving the uptake of glucose when ethanol titres exceed 3% ($w.v^{-1}$).

The accumulation of glycerol during fermentations at 30°C and 37°C for recombinant *S. cerevisiae* Y294 and Ethanol Red strains, respectively, validated the effect of fermentation conditions on carbon utilisation. Glycerol is a major by-product during ethanol production by *S. cerevisiae*, signifying 4 - 5% of the carbon source consumption; other products include biomass, carbon dioxide and acetic acid production. The glycerol pathway is essential to reoxidise the NADH produced during the anaerobic cultivation of *S. cerevisiae*, resulting in the production of glycerol. Several studies have used metabolic engineering to decrease the production of glycerol, in order to improve the carbon flow and construct a yeast that produces more ethanol. Navarrete et al. (2014) investigated the combinatorial effect of reducing glycerol export and formation and successfully engineered *S. cerevisiae* with an increased ethanol yield of 4.6%.

Temperature is also an important parameter for the optimisation of aerobic heterologous amylase production in yeast systems. In agreement with this study (**Results from Chapter 3**), Dragosits et al. (2011) showed that a lower cultivation temperature positively influenced recombinant protein production in several prokaryotic and eukaryotic hosts. The influence of temperature on microbial growth is

a widely studied topic and many mathematical models have been developed to quantify and predict its effects (Salvadó et al. 2011). Yeast cultivation temperatures have a considerable influence on cell metabolism, as well as strain productivity. The increase in extracellular enzyme activity at a lower cultivation temperature can also be attributed to a number of factors, including enhanced protein folding pathways, increased cell viability, changes in gene regulation and increases in enzyme stability (half-life) (Li et al. 2001; Hong et al. 2002; Gasser et al. 2007).

A major hurdle in molecular biology is understanding the process by which proteins fold. Misfolded proteins or aggregation will prevent the protein from functioning normally and it is not a guarantee that proteins having the same amino acid sequence (encoded for by different nucleotide sequences) will fold to form the same tertiary structure. In this study, the codon optimisation strategies failed to improve the protein production by recombinant yeast strains expressing the codon adapted/optimised *amyA* and *glaA* genes from *Aspergillus tubingensis* (**Chapter 3**). The use of differential scanning fluorimetry, overexpression of chaperone *PDI1*, the exchange of specific rare/abundant codons and a lower cultivation temperature suggested that the mechanics behind protein folding and the subsequent secretion of proteins into the extracellular environment is a complex process that involves several interacting participants. The effect that codon translation rates have on co-translational folding is poorly understood, yet it remains an essential factor for producing proteins encoded for by codon optimised genes.

In this study, not all codon optimised genes performed worse than their native counterparts. The *S. cerevisiae* Y294[TemG_Opt] strain expressing the codon optimised glucoamylase from *Talaromyces emersonii (temG_Opt)* outperformed all other glucoamylase producing *S. cerevisiae* strains, including the *S. cerevisiae* Y294[TemG_Nat] strain expressing the native glucoamylase (*temG_Nat*) from *T. emersonii* (**Chapter 4**). Together with the native α-amylase (*temA_Nat*) from *T. emersonii*, these genes formed a novel amylase combination expressed by *S. cerevisiae* for raw starch hydrolysis. Furthermore, replacing the secretion signals did not increase the extracellular enzyme production. Thus the DNA coding sequences for the native secretion signals of the fungal amylases were preferred for directing the extracellular secretion of the amylases.

Cinelli et al. (2015) highlighted recent examples of ethanol production via granular starch hydrolysis, but the examples used wild-type/untransformed *S. cerevisiae* strains and commercial enzymes. A major limitation of fermentation studies, using amylolytic *S. cerevisiae* yeast, is the lack of data to compare CBP processes in combination with granular starch hydrolysing enzymes (GSHE) cocktails (Görgens et al. 2015). In this study, several different enzyme loadings of exogenous GSHEs were evaluated in combination with the amylolytic *S. cerevisiae* Ethanol Red T12 and M2n T1 strains. Results indicated that a low enzyme dosage of STARGEN 002™ (10% of the recommended loading) together with the amylolytic yeast was sufficient for ethanol concentrations to exceed 94% of the theoretical yield. (**Chapter 5**). Thus, underlining the novel approach to a one-step CBP process for starch processing.

## 6.3 Conclusions

- Adapted synonymous codon optimisation of the *amyA* and *glaA* genes from *A. tubingensis* resulted in decreased extracellular activity levels.
- The codon adaption index (CAI) is not a reliable strategy for the optimisation of fungal amylases for expression in *S. cerevisiae*.
- Overexpressing *PDI1* and lowering the cultivation temperature, respectively, improved the conditions for protein folding, leading to enhanced secretion and productivity for the strains expressing the *amyA_Opt* and *glaA_Opt* genes.
- Synonymous codon usage does not guarantee successful gene expression and insufficient protein folding limits heterologous expression.
- The expression of the *T. emersonii* codon optimised glucoamylase gene (*temG_Opt*) and native α-amylase gene (*temA_Nat*) facilitated the best conversion of raw corn starch during fermentation studies and represented a novel enzyme combination for raw starch hydrolysis.
- The recombinant *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain converted 85% of the available carbon in 200 $g.l^{-1}$ raw corn starch within 192 hours.
- The *temA_Nat* and *temG_Opt* linear gene cassettes were successfully integrated into the *S. cerevisiae* Ethanol Red and M2n strains' genomes.

- At 30°C, the *S. cerevisiae* M2n T1 strain displayed the highest ethanol concentration (99.4 g.l$^{-1}$ after 192 hours).
- At 37°C, the *S. cerevisiae* Ethanol Red T12 strain performed better than the *S. cerevisiae* M2n T1 strain in terms of ethanol production.
- Supplementation with STARGEN 002™ allowed for a 90% reduction in the exogenous enzyme dosage, with ethanol yields comparable to an SSF control. The SSF control consisted of the untransformed parental strains (*S. cerevisiae* Ethanol Red and M2n) with the recommended STARGEN 002™ loading.
- During fermentation studies, temperature was the main factor affecting the conversion of glucose to ethanol.
- The industrial amylolytic yeast strains were able to grow on acetamide and acrylamide as sole nitrogen sources demonstrating their potentially novel application in reducing acrylamide contamination.

## 6.4 Future research

### 6.4.1 Molecular biology approach

Recent studies in our laboratory (unpublished data) have shown that the choice of constitutive promoter can have a significant impact on the resulting protein production. Therefore, investigation of alternative promoters (e.g. the constitutively expressed *TEF* promoter from *Ashbya gossypii* or *S. cerevisiae*) needs to be considered for the expression of amylase genes. Promoters drive the expression of foreign genes and they have different regulatory mechanisms. An alternative promoter could enhance the expression of certain amylases and result in increased protein production. Therefore, a molecular approach to improve the expression of *temA_Nat* and *temG_Opt* would be to redesign linear gene cassettes so that the genes are expressed under the control of different promoters.

The amylases used in this study remained active long after the maximum ethanol concentration was reached. However, increased amylase expression at the beginning of the fermentation would decrease the initial lag in ethanol production and subsequently shorten the fermentation time. An alternative gene integration approach (e.g. CRISPR-Cas9 targeted genome editing) is suggested as an approach for

increasing gene copy numbers and improving transformation efficiencies. The use of the laboratory *S. cerevisiae* Y294 is an ideal host for routine screening, but it has a limited ethanol tolerance at 30°C (it can only tolerate ethanol concentrations ~60 $g.l^{-1}$). Thus, it is not an ideal host for evaluating amylase combinations that are highly effective with regard to enzymatic saccharification of raw corn starch. Therefore, it is fundamental to use industrial *S. cerevisiae* strains to evaluate different factors for improved raw starch hydrolysis.

### 6.4.2 Controlling microbial contamination

Lactic acid bacteria (LAB) contamination is a major problem in the biofuel industry, resulting in reduced ethanol titres. This weakens the economic viability of bioethanol production, especially when using starchy biomass in the cold hydrolysis process (as opposed to the conventional process using high temperatures to gelatinise and liquefy starch granules). Therefore, there is a need to combine an antimicrobial agent with the amylolytic CBP yeast.

Further research should focus on screening for novel, economical, broad spectrum antimicrobial products. Antimicrobial compounds from natural sources such as medicinal plants and herbs have not been sufficiently evaluated for their use in fermentations to control contamination. The added benefit to these compounds is that they would be environmentally friendly compared to the antibiotics traditionally used in biofuel production and their use will allow for distiller's dried grains with solubles (DDGS) to be used as animal feed.

### 6.4.3 Acrylamide reduction using *S. cerevisiae* Ethanol Red T12

This study demonstrated that the industrial amylolytic yeast were able to grow on acrylamide ($C_3H_5NO$) as the sole nitrogen source. Therefore, these strains should be further evaluated for their potential to reduce acrylamide (group 2A carcinogen) contamination. Boiling does not result in acrylamide formation, since the water temperature is around 100°C and thus below the 120°C threshold. However, during the cooking of starchy foods (at temperatures above 120°C) a chemical reaction occurs between asparagine and reducing sugars. Foods containing high levels of acrylamide include: breakfast cereals, potato chips and coffee. The average acrylamide consumption by children is twice as high as for adults, which is concerning

because it can have a negative impact on their health. Therefore, from a medical perspective it is important to reduce the health threat from acrylamide.

It is essential to develop cost effective, environmentally friendly methods to control acrylamide waste. Novozymes currently produces Acrylaway® (http://www.novozymes.com), which has been used to reduce acrylamide in a several food products. However, the use of this product to treat contaminated waste water would be a misuse of resources, because the waste has no economic value. Supernatant from the recombinant *S. cerevisiae* Ethanol Red T12 strain would, however, represent a more affordable option, whereby the cell-free extract containing acetamidase could be used to treat waste water before it is disposed of in the environment. This will lower the level of acrylamide contamination. Alternatively, acrylamide contaminated waste from the starch industry (e.g. potato waste) could be used as a feedstock for bioethanol production by the recombinant yeast expressing acetamidase, since carbon and nitrogen sources are products of acrylamide degradation.

### 6.4.4 Further evaluation of *S. cerevisiae* Ethanol Red T12

Starchy biomass is an attractive feedstock from which to produce bioethanol. However, corn is a staple food source in South Africa and thus its usage for ethanol production is prohibited. Therefore, there is scope to investigate alternative starchy substrates for bioethanol production e.g. triticale (an abundant disease resistant crop that can tolerate harsh growth conditions) and potato waste.

During this study, industrial amylolytic yeast were constructed that demonstrated suitable drop-in CBP organisms for the industrial production of bioethanol. However, starchy biomass contains several other components, such as proteins, cellulose, hemicellulose and lignin. Therefore, it is desirable to improve starch hydrolysis by expressing additional hydrolases. The amylolytic *S. cerevisiae* Ethanol Red T12 strain can be engineered further to express auxiliary enzymes that would assist in the hydrolysis of biomass, such as triticale, that contain starch and small amounts of other polysaccharides (e.g. cellulose and xylan). This would improve the utilising of the feedstock and increase the final ethanol titres. The expression of proteases would also facilitate in the hydrolysis of starchy (corn) biomass, by hydrolysing proteins (called oleosins) that provide structure to the corn kernels.

During bioethanol production, the morphology and structure of native starch granules affects its bioconversion efficiency to fermentable sugars. However, an additional advantage to using alternative starchy biomass such as grains (e.g. triticale and sorghum) is that they have their own native amylases (β-amylase and α-amylase). The native enzymes would supplement the recombinant amylolytic enzymes and enhance the rate of starch hydrolysis; this would subsequently increase ethanol productivity. Furthermore, studying the hydrolysis of different types of starch granules will also provide information on the relationship between bioconversion factors during the CBP of starch for bioethanol production, as well as other industrial applications.

Ethanol tolerance is one of the key characteristics for an ethanol-producing yeast, especially when high gravity fermentations are considered, e.g. raw starch dry weight loadings of 30 - 40%. Applications in fuel production require robust industrial *S. cerevisiae* strains that are able to tolerate high ethanol yields and higher fermentation temperatures (> 30°C). Therefore, there is a need to use strain improvement techniques in order to develop robust temperature tolerant strains that would be more suited to the industrial production of bioethanol from starchy materials. Furthermore, proteomic studies are necessary for understanding the molecular basis of thermotolerance, as well as to identify key proteins that play a role in the thermotolerance response. Although temperature tolerance is an important environmental stress, few studies have investigated the thermotolerance of industrial *S. cerevisiae* strains using a proteomic approach.

One of the major renewable energy goals is the establishment of an advanced biofuels economy (integrates biomass conversion processes and fuel production) that can use raw starch technology for the production of bioethanol. Raw starch technology can also be extended to the production of other higher value green chemicals. This would promote the establishment of biorefineries, as part of a future bio-based economy, in southern Africa and provide the opportunity for South Africa to become a technology leader in this field.

## 6.5 References

Angov E (2011) Codon usage: nature's roadmap to expression and folding of proteins. Biotechnol J 6:650-659. doi:10.1002/biot.201000332

Cinelli BA, Castilho LR, Freire DMG, Castro AM (2015) A brief review on the emerging technology of ethanol production by cold hydrolysis of raw starch. Fuel 150:721-729. doi: 10.1016/j.fuel.2015.02.063

Celińska E, Białas W, Borkowska M, Grajek W (2015) Cloning, expression, and purification of insect (*Sitophilus oryzae*) alpha-amylase, able to digest granular starch, in *Yarrowia lipolytica* host. Appl Microbiol Biotechnol 99:2727-2739. doi: 10.1007/s00253-014-6314-2

Chung BK, Lee DY (2012) Computational codon optimization of synthetic gene for protein expression. BMC Syst Biol 6:134. doi: 10.1186/1752-0509-6-134

Connor MR, Atsumi, S (2010) Synthetic biology guides biofuel production. J Biomed Biotechnol 2010:541698. doi: 10.1155/2010/541698

Dellomonaco C, Fava F, Gonzalez R (2010) The path to next generation biofuels: successes and challenges in the era of synthetic biology. Microb Cell Fact 9:3. doi: 10.1186/1475-2859-9-3

Dragosits M, Frascotti G, Bernard-Granger L, Vázquez F, Giuliani M, Baumann K, Rodríguez-Carmona E et al. (2011) Influence of growth temperature on the production of antibody Fab fragments in different microbes: a host comparative analysis. Biotechnol Prog 27:38-46 doi: 10.1002/btpr.524

Ekino K, Hayashi H, Moriyama M, Matsuda M, Goto M, Yoshino S, Furukawa K (2002) Engineering of polyploid *Saccharomyces cerevisiae* for secretion of large amounts of fungal glucoamylase. Appl Environ Microbiol 68:5693-5697. doi: 10.1128/AEM.68.11.5693-5697.2002

Elena C, Ravasi P, Castelli ME, Peirú S, Menzella HG (2014) Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. Front Microbiol 5:21. doi: 10.3389/fmicb.2014.00021

Favaro L, Jooste T, Basaglia M, Rose SH, Saayman M, Görgens JF, Casella S, van Zyl WH (2013) Designing industrial yeasts for consolidated bioprocessing of starchy biomass to ethanol. Bioengineered 4:97–102. doi: 10.4161/bioe.22268

Gasser B, Maurer M, Rautio J, Sauer M, Bhattacharyya A, Saloheimo M, Penttilä M, Mattanovich D (2007) Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. BMC Genomics 8:179. doi: 10.1186/1471-2164-8-179

Görgens JF, Bressler DC, van Rensburg E (2015) Engineering *Saccharomyces cerevisiae* for direct conversion of raw, uncooked or granular starch to ethanol. Crit Rev Biotechnol 35:396-391. doi: 10.3109/07388551.2014.888048

Henry I, Sharp PM (2007) Predicting gene expression level from codon usage bias. Mol Biol Evol 24:10-12. doi: 10.1093/molbev/msl148

Hoekema A, Kastelein RA, Vasser M, de Boer HA (1987) Codon replacement in the *PGK1* gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. Mol Cell Biol 7:2914-2924. PMID: 2823108

Hong F, Meinander NQ, Jönsson LJ (2002) Fermentation strategies for improved heterologous expression of laccase in *Pichia pastoris*. Biotechnol Bioeng 79:438-449. doi: 10.1002/bit.10297

Jessop-Fabre MM, Jakočiūnas T, Stovicek V, Dai Z, Jensen MK, Keasling JD Borodina I (2016) EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into *Saccharomyces cerevisiae* via CRISPR-Cas9. Biotechnol J 11:1110-1117. doi: 10.1002/biot.201600147.

Lanza AM, Curran KA, Rey LG, Alper HS (2014) A condition-specific codon optimization approach for improved heterologous gene expression in *Saccharomyces cerevisiae*. BMC Syst Biol 8:33. doi: 10.1186/1752-0509-8-33

Le Borgne S 2012. Genetic engineering of industrial strains of *Saccharomyces cerevisiae*. Methods Mol Biol 824:451–465. doi: 10.1007/978-1-61779-433-9_24

Li Z, Xiong F, Lin Q, d'Anjou M, Daugulis AJ, Yang DS, Hew CL (2001) Low-temperature increases the yield of biologically active herring antifreeze protein in *Pichia pastoris*. Protein Expr Purif 21:438-445. doi: 10.1006/prep.2001.1395

Mukhtar K, Asgher M, Afghan S, Hussain K, Zia-Ul-Hussnain S (2010) Comparative study on two commercial strains of *Saccharomyces cerevisiae* for optimum ethanol production on industrial scale. J Biomed Biotechnol 2010:419586. doi: 10.1155/2010/419586

Navarrete C, Nielsen J, Siewers V (2014) Enhanced ethanol production and reduced glycerol formation in *fps1*Δ mutants of *Saccharomyces cerevisiae* engineered for improved redox balancing. AMB Express 4:86. doi: 10.1186/s13568-014-0086-z

Presnell SR, Benner SA (1988) Nucleic Acids Res16:1693-1702. PMID: 2451218

Richardson SM, Wheelan SJ, Yarrington RM, Boeke JD (2006) GeneDesign: rapid, automated design of multikilobase synthetic genes. Genome Res 16:550-556. doi: 10.1101/gr.4431306

Salvadó Z, Arroyo-López FN, Guillamón JM, Salazar G, Quero A, Barrio E (2011) Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. Appl Environ Microbiol 77:2292-2302. doi: 10.1128/AEM.01861-10

Shao Z, Zhao H, Zhao H (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37:1–10. doi: 10.1093/nar/gkn991

Snoek T, Picca Nicolino M, Van den Bremt S, Mertens S, Saels V, Verplaetse A, Steensels J, Verstrepen KJ (2015) Large-scale robot-assisted genome shuffling yields industrial *Saccharomyces cerevisiae* yeasts with increased ethanol tolerance. Biotechnol Biofuels 8:32. doi: 10.1186/s13068-015-0216-0

Stovicek V, Borja GM, Forster J, Borodina I (2015) EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial *Saccharomyces cerevisiae* strains. J Ind Microbiol Biotechnol 42:1519-1531. doi: 10.1007/s10295-015-1684-8

Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, Gustafsson C (2009A) Design parameters to control synthetic gene expression in *Escherichia coli*. PLoS One 4:e7002. doi: 10.1371/journal.pone.0007002

Woo JM, Yang KM, Kim SU, Blank LM, Park JB (2014) High temperature stimulates acetic acid accumulation and enhances the growth inhibition and ethanol production by *Saccharomyces cerevisiae* under fermenting conditions. Appl Microbiol Biotechnol 98:6085-6094. doi: 10.1007/s00253-014-5691-x

Wu G, Dress L, Freeland SJ (2007) Optimal encoding rules for synthetic genes: the need for a community effort. Mol Syst Biol 3:134. doi: 10.1038/msb4100176

# Addendum A

United Kingdom Patent Application No. GB 1 620 658.3, filed 5 December 2016, entitled “Recombinant yeast and use thereof”.

# Addendum A: RECOMBINANT YEAST AND USE THEREOF

## FIELD OF THE INVENTION

The invention relates to a recombinant yeast expressing at least one heterologous enzyme, wherein the heterologous enzyme is a glucoamylase of SEQ ID NO: 1, and to the use of the recombinant yeast in a process for converting sugars or starch to alcohol.

## BACKGROUND TO THE INVENTION

Cost effective, renewable and sustainable energy is a global concern, which has increased investigations into alternative fuel sources. Starch is an industrially important substrate for the production of biofuel, proteins and chemicals. It is produced by plants as an energy store and consists of α-1,4 linked glucose units with α-1,6 branching points. The amylose and amylopectin polymers are densely packed in starch granules forming a semi-crystalline structure with inter- and intra-molecular bonds.

A combination of α-amylases and glucoamylases is required for the complete hydrolysis of starch. Starch granules are insoluble in cold water and are often resistant to enzymatic hydrolysis (Uthumporn *et al*., 2010). The conventional process for the conversion of starch to ethanol requires a heat intensive liquefaction step to gelatinise the starch and thermostable α-amylases, followed by saccharification with glucoamylases. The high temperatures required for the initial processes usually account for approximately 30 - 40% of the total energy required for ethanol production (Szymanowska-Powałowska *et al*., 2012).

An alternative to this is a cold hydrolysis process at temperatures below the onset of starch gelatinisation (65°C for corn) (Robertson *et al*., 2006). Benefits of this process include reduced energy requirements and a higher nutritional content for the dried distillers' grains with solubles (DDGS) (Nkomba *et al*., 2016). DDGS are produced in large quantities during bioethanol production and represent a valuable ingredient for livestock feed (Brehmer *et al*., 2008).

Consolidated bioprocessing (CBP) combines enzyme production, hydrolysis and fermentation into a one-step process for bioethanol production at low temperatures. This technology represents a promising alternative for the economic production of biofuel from lignocellulosic

and starchy feedstocks. CBP could simplify operational processes (e.g. number of control steps and reaction vessels) and therefore reduce maintenance and production costs. CBP systems use a single organism that is able to produce the enzymes required for hydrolysis of starch at low temperatures, i.e. cold hydrolysis, as well as convert the resultant sugars to ethanol. The cold process requires amylases that have the ability to digest raw starch efficiently at fermentation conditions. A few raw starch hydrolysing amylases have been reported to date (Mamo and Gessesse, 1999; Robertson *et al*., 2006; Celińska *et al*., 2015). These amylases differ from conventional amylases in their affinity and interaction with the microcrystalline structures of starch granules. A starch binding domain (SBD) is a key characteristic of these enzymes and enables them to bind effectively to the surface of raw starch granules.

A comprehensive review on consolidated bioprocessing systems by Salehi Jouzani and Taherzadeh (2015) highlighted different CBP strategies, diversity in substrate types and the organisms involved in fermenting the sugars. One of the main challenges remains the simultaneous production of the amylases with high substrate affinities and specific activity (den Haan *et al*., 2013). In addition, fermentation requirements are ethanol concentrations in excess of 10 - 12% ($w.v^{-1}$) within 48 to 72 hours (Bothast and Schlicher, 2005). For example, raw starch amylase encoding genes from *Lipomyces kononenkoae* and *Saccharomycopsis fibuligera* (Eksteen *et al*., 2003; Knox *et al*., 2004), *Rhizopus arrhizus* (Yang *et al*., 2011), *Aspergillus tubingensis* (Viktor *et al*., 2013) and *Thermomyces lanuginoses* and *S. fibuligera* or *L. kononenkoae* (LKA1) protein (US 9,243,256) have been expressed in *Saccharomyces cerevisiae*, a yeast which is an efficient ethanol producer but which on its own lacks the ability to degrade starch. However, none of these transformed yeasts produce sufficient amylase to support efficient conversion of raw starch to ethanol in a single step at commercial scale. Although a bioengineered *S. cerevisiae* strain that secretes a glucoamylase is commercially available (TransFerm® from Lallemand, developed by Mascoma Corporation (http://www.ethanoltech.com/transferm)), it lacks the required α-amylase enzymes for starch liquefaction (den Haan *et al*., 2015) and is therefore only a semi-CBP yeast. The TransFerm® yeast strain is thus only suitable for the conventional (warm) process, as it only consolidates the saccharification and fermentation processes after starch liquefaction. CBP has therefore not yet been implemented on a commercial level, with the main challenge being the availability of an ideal microorganism that can express suitable enzymes and have a high fermentation capacity.

Other cold simultaneous saccharification and fermentation (SSF) processes have been developed for ethanol production from starchy substrates (Balcerek and Pielech-Przybylska,

2013; Szymanowska-Powałowska *et al*., 2014; Nkomba *et al*., 2016). In these processes, granular starch hydrolysing enzyme (GSHE) cocktails are added to the feedstock in addition to the yeast. Genencor's STARGEN 001™ and STARGEN 002™ cocktails (Dupont-Danisco, Itasca, Itasca) hydrolyse raw starch at low temperatures (48°C recommended for SSF), while POET (Sioux Falls, South Dekota, USA) uses a patented blend of Novozymes enzymes (POET BPX technology) in an SSF process (Görgens *et al*., 2015). However, these cold starch hydrolysis processes require high enzyme loadings and the cost of the commercial enzymes, e.g. STARGEN™ (Genencor International, California, USA), is high.

There thus remains a need for a yeast which can be used in a CBP process for producing ethanol from raw starch, without requiring the addition of amylases from a source other than the yeast.

## SUMMARY OF THE INVENTION

According to a first embodiment of the invention, there is provided a recombinant yeast which has been transformed with a heterologous gene which is capable of expressing a polypeptide comprising an amino acid sequence which is at least 70% identical to SEQ ID NO: 1.

The amino acid sequence of the polypeptide may be at least 80% identical to SEQ ID NO: 1; the amino acid sequence of the polypeptide may be at least 90% identical to SEQ ID NO: 1; or the amino acid sequence of the polypeptide may be identical to SEQ ID NO: 1.

The nucleic acid sequence of the heterologous gene may be at least 70% identical to either of SEQ ID NOS: 3 and 5; the nucleic acid sequence of the heterologous gene may be at least 80% identical to either of SEQ ID NOS: 3 and 5; the nucleic acid sequence of the heterologous gene may be at least 90% identical to either of SEQ ID NOS: 3 and 5; or the nucleic acid sequence of the heterologous gene may be identical to either of SEQ ID NOS: 3 and 5.

The recombinant yeast may have been further transformed with a second heterologous gene which is capable of expressing a second polypeptide comprising an amino acid sequence which is at least 70% identical to SEQ ID NO: 2.

The amino acid sequence of the second polypeptide may be at least 80% identical to SEQ ID NO: 2; the amino acid sequence of the second polypeptide may be at least 90% identical to

SEQ ID NO: 2; or the amino acid sequence of the second polypeptide may be identical to SEQ ID NO: 2.

The nucleic acid sequence of the second heterologous gene may be at least 70% identical to SEQ ID NO: 4; the nucleic acid sequence of the second heterologous gene may be at least 80% identical to SEQ ID NO: 4; the nucleic acid sequence of the second heterologous gene may be at least 90% identical to SEQ ID NO: 4; or the nucleic acid sequence of the second heterologous gene may be identical to SEQ ID NO: 4.

The yeast may be a yeast which is capable of converting sugars such as glucose to alcohol.

The yeast may be a *Saccharomyces* species, such as a *Saccharomyces cerevisiae* species.

The alcohol may be butanol or ethanol, and in particular is ethanol.

The recombinant yeast may be capable of hydrolysing raw starch in the absence of enzymes from a source other than the recombinant yeast. The raw starch may be hydrolysed at a temperature of below about 40°C.

According to a second embodiment of the invention, there is provided a process for producing an alcohol from sugars, the process comprising the step of using a recombinant yeast as described above to convert the sugars to alcohol.

The alcohol may be ethanol or butanol, and is typically ethanol.

According to a third embodiment of the invention, there is provided a process for producing an alcohol from starch, the process comprising the step of using a recombinant yeast as described above to convert the starch to alcohol.

The starch may be a grain starch.

The starch may be raw (granular) starch.
The raw starch may be hydrolysed by the recombinant yeast without requiring cooking of the starch, for example the raw starch may be hydrolysed by the recombinant yeast at a temperature of below about 40°C.

The starch may be cooked (processed or modified).

The alcohol may be produced from the starch without the addition of enzymes from a source other than the yeast.

Alternatively, amylolytic enzymes may be initially added to the process to reduce the time taken to convert the starch to alcohol, the amylolytic enzymes being added in an amount which is at least about 50% less than the amount added to cold hydrolysis processes which do not use the recombinant yeast as described above. The added amylolytic enzymes may be a cocktail of enzymes which together are capable of hydrolysing raw starch.

The alcohol may be ethanol or butanol, and is typically ethanol.

**BRIEF DESCRIPTION OF THE FIGURES**

Figure 2 Schematic representation of the vector constructs used in example 1. Amylase encoding genes were amplified using PCR and respectively cloned onto the yBBH1 and yBBH4 vectors (**a**, **b** and **c**). The *ENO1*$_P$-α-amylases-*ENO1*$_T$ cassettes were cloned onto the yBBH1-glucoamylase plasmids (**d**), to enable co-expression of the genes. *Bam*HI and *Bgl*II restriction enzyme sites were used for yeast mediated ligation (YML).

Figure 2 Extracellular α-amylase activity displayed by the *S. cerevisiae* Y294 strains expressing the (**a**) *ateA*, *amyA*, (**b**) *apuA* and (**c**) *temA* gene derivatives, respectively. The *S. cerevisiae* Y294[AmyA] strain was used for benchmark α-amylase production. Values represent the mean of three repeats and error bars represent the standard deviation. Supernatant from the *S. cerevisiae* Y294 strains (after 72 hours) was subjected to SDS-PAGE followed by silver staining. The arrows indicate the presence of the recombinant (**d**) AmyA, AteA, (**e**) ApuA and (**f**) TemA protein species, respectively. The *S. cerevisiae* Y294[BBH1] strain was used as the reference strain and the protein size marker (M) is depicted on the left hand side.

Figure 3 Extracellular glucoamylase activity displayed by the *S. cerevisiae* Y294 strains expressing the (**a**) *ateG, glaA* and (**b**) *temG* gene derivatives, respectively. The *S. cerevisiae* Y294[GlaA] strain was used for benchmark glucoamylase production. Values represent the

mean of three repeats and error bars represent the standard deviation. Supernatant from the *S. cerevisiae* Y294 strains (after 72 hours) was subjected to SDS-PAGE followed by silver staining. The arrows indicate the presence of the recombinant (**c**) AteG and (**d**) TemG protein species, respectively. The *S. cerevisiae* Y294[BBH1] strain was used as the reference strain and the protein size marker (M) is depicted on the left hand side.

Figure 4 The amylolytic *S. cerevisiae* Y294 strains were evaluated on 200 g.l$^{-1}$ raw corn starch and 5 g.l$^{-1}$ glucose as sole carbohydrate source. The (**a** and **b**) ethanol and (**c** and **d**) glucose production was monitored overtime. Results from the best performing strains (left panel) and suboptimal strains (right panel) came from the same fermentation. Values represent the mean of three repeats and error bars represent the standard deviation.

Figure 5 The performance of *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] in a 2 litre bioreactor. (**a**) Ethanol concentrations at 26°C (-●-) and 30°C (-■-) and residual glucose concentrations at 26°C (-○-) and at 30°C (-□-) and (**b**) product yield at 26°C (-●-) and 30°C (-■-), respectively, with 2×SC$^{-URA}$ broth supplemented with 5 g.l$^{-1}$ glucose and 200 g.l$^{-1}$ raw corn starch. Values represent the mean of triplicate repeats and error bars represent the standard deviation.

Figure 6 Schematic representation of the final vector and gene cassettes used in this study. The *TEF*$_P$-*amdSYM*-*TEF*$_T$ cassette (**a**) was cloned onto yBBH1 to generate the yBBH1-amdSYM expression vector. The *ENO1 temA_Nat* and *temG_Opt* gene cassettes (**b**) were amplified using PCR and contained flanking regions homologous to the δ integration sites.

Figure 7 Comparison of industrial transformants after integration of *temA* and *temG* gene cassettes. Ethanol produced (**a**) and percentage product yield (**b**) displayed by *S. cerevisiae* Ethanol Red (-□-) and M2n (-○-) parental strains and *S. cerevisiae* Ethanol Red T1 (-▲-), T12 (-■-), M2n T1 (-◆-) and Mn2 T2 (-●-) amylolytic transformants at a fermentation temperature of 30°C on 200 g.l$^{-1}$ raw corn starch. SC-Ac (**c**) and SC-Acr (**d**) plate assays confirmed the ability of recombinant *S. cerevisiae* Ethanol Red T12 and M2n T1 strains to utilise acetamide and acrylamide, respectively, whereas the parental *S. cerevisiae* Ethanol Red and M2n strains indicated no growth.

Figure 8 Comparison between the laboratory *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain (-●-) and the industrial amylolytic *S. cerevisiae* Ethanol Red™ T12 strain at

30°C (-■-) and 37°C (-▲-). The production of ethanol (**a**), glucose (**b**), maltose (**c**) and glycerol (**d**) were compared using 2×SC$^{-URA}$ fermentation media that contained 5 g.l$^{-1}$ glucose and 200 g.l$^{-1}$ raw corn starch. Data are the mean of 3 repeats showing standard deviation.

Figure 9 Different fermentation broth conditions during fermentation at 37°C on 200 g.l$^{-1}$ raw corn starch. *S. cerevisiae* Ethanol Red™ T12 in YP (-◆-), YP citrate-acid buffer pH 5 (-■-), SC citrate-acid buffer pH 5 (-●-) and SC citrate-acid buffer pH 5 with 10 g.l$^{-1}$ extra $(NH_4)_2SO_4$ (-▲-). Ethanol (**a**), glucose (**b**), glycerol concentrations (**c**) and product yield (**d**) were compared. Data are the mean of 3 repeats showing standard deviation.

Figure 10 Ethanol concentrations produced by *S. cerevisiae* Ethanol Red™ strains during fermentation with 200 g.l$^{-1}$ corn starch at 30°C (**a**), at 37°C, (**b**), product yield at 30°C (**c**) and product yield at 37°C (**d**). Untransformed Ethanol Red™ + 28 µl STARGEN™ (-▲-), Ethanol Red™ T12 (-■-), Ethanol Red™ T12 + 2.8 µl STARGEN™ (-◆-), Ethanol Red™ T12 + 4.6 µl STARGEN™ (-●-) and Ethanol Red™ T12 + 14 µl STARGEN™ (-▬-).

Figure 11 Ethanol concentrations produced by *S. cerevisiae* M2n strains during fermentation with 200 g.l$^{-1}$ at 30°C (**a**), at 37°C, (**b**), product yield at 30°C (**c**) and product yield at 37°C (**d**). The untransformed *S. cerevisiae* M2n strain + 28 µl STARGEN™ (-▲-), M2n T1 (-■-), M2n T1+ 2.8 µl STARGEN™ (-◆-) and M2n T1 + 4.6 µl STARGEN™ (-●-). Data are the mean of 3 repeats showing standard deviation.

Figure 12 SEQ ID NO: 1 - TemG protein. Protein sequence of *Rasamsonia emersoni* glucoamylase (secretion signal underlined). Sequence ID: CAC28076.1.

Figure 13 SEQ ID NO: 2: - TemA protein. Protein sequence of *Rasamsonia emersonii* alpha-amylase. [Rasamsonia emersonii CBS 393.64] Sequence ID: GenBank no. XP_013324946.

Figure 14 SEQ ID NO: 3 - *temG_Opt*. DNA sequence coding for the *Rasamsonia emersonii* glucoamylase (putative secretion signal underlined), optimised for expression in *S. cerevisiae* (by GenScript, USA).

Figure 15 SEQ ID NO: 4: *temA_Nat* gene. Synthetic DNA sequence coding for the *Rasamsonia emersoni* alpha-amylase (putative secretion signal underlined) used to produce TemA_Nat. This is 99% identical to *Rasamsonia emersonii* CBS 393.64 alpha-amylase mRNA

NCBI Reference Sequence: Genbank no. XM_013469492 (1 nucleotide was changed, compared to the original GenBank sequence, without affecting the protein sequence).

Figure 16 SEQ ID NO: 5: *temG_Nat* gene. Adapted native DNA sequence coding for the *Rasamsonia emersonii* glucoamylase (TemG_Nat). This sequence contained 3 nucleotide changes (bold and underlined) compared to the original GenBank sequence (introns removed) and the protein sequence is TemG_Nat.

Figure 17 SEQ ID NO: 6 *temA* – original Genbank sequence for native *Rasamsonia emersonii* CBS 393.64 Alpha-amylase mRNA NCBI Reference Sequence: XM_013469492.1

Figure 18 SEQ ID NO: 7: *temG* – original *Talaromyces emersonii* ga gene for glucoamylase, exons 1-5 (Genbank sequence including introns; GenBank: AJ304803.1).

**DETAILED DESCRIPTION OF THE INVENTION**

A recombinant yeast that expresses a glucoamylase from *Talaromyces emersonii* (recently re-named as *Rasamsonia emersonii*) is described. The glucoamylase comprises an amino acid sequence which is at least 70% identical to SEQ ID NO: 1. The recombinant yeast strain can be used for converting sugars to an alcohol, such as for biofuel.

Optionally, the recombinant yeast may also be transformed with a second gene expressing an α-amylase from *T. emersonii*. The amino acid sequence of the α-amylase is at least 70% identical to SEQ ID NO: 2.

The yeast can be transformed with the native genes for both of these enzymes, with codon-optimised genes for both of these enzymes, or with one native gene and one codon-optimised gene. Nucleotide changes may also be made to the native gene so as to disrupt restriction sites for cloning purposes, but without altering the protein sequence.

In one embodiment of the invention, the yeast is transformed with only the gene for the glucoamylase. This can be the native gene or the codon-optimised version, with the native secretion signal or another secrection signal (e.g. XYNSEC). In this embodiment, the yeast is transformed with a glucoamylase which comprises a nucleic acid sequence which is at least 68% similar to, at least 70% similar to, at least 80% similar to, at least 90% similar to, or identical to either of SEQ ID NOS: 3 or 5.

In another embodiment, the yeast is transformed with (i) a codon-optimised gene for the glucoamylase, which has 69% identity to the native sequence, and (ii) the native gene for the α-amylase. In this embodiment, the yeast is transformed with a glucoamylase which comprises a nucleic acid sequence which is at least 68% similar to, at least 70% similar to, at least 80% similar to, at least 90% similar to, or identical to SEQ ID NO: 3. The yeast is also transformed with an α-amylase which comprises a nucleic acid sequence with is at least 70% similar to, at least 80% similar to, at least 90% similar to, or identical to SEQ ID NO: 4.

In an alternative embodiment, instead of the yeast being transformed with the codon-optimised glucoamylase, it may be transformed with the native glucoamylase comprising a nucleic acid sequence which is at least 70% similar to, at least 80% similar to, at least 90% similar to, or identical to SEQ ID NO: 5.

Exemplary yeasts for the present invention are *Pichia* (*Hansenula*) spp. (e.g. *P. anomala*, *P. capsulate* and *P. angusta* (formerly *H. polymorpha*)), *Saccharomyces* spp. (e.g. *S. cerevisiae, S. italicus* and *S. rouxii*), *Yarrowia* (e.g. *Y. lipolytica*), *Kluyveromyces* spp. (e.g. *K. fragilis* and *K. lactis*), *Candida* spp. (e.g. *C. tropicales*), *Torulopsis* spp., *Torulaspora* spp., *Schizosaccharomyces* spp. (e.g *S. pombe*), *Citeromyces* spp., *Pachysolen* spp., *Debaromyces* spp., *Metschunikowia* spp., *Rhodosporidium* spp., *Leucosporidium* spp., *Botryoascus* spp., *Sporidiobolus* spp., *Endomycopsis* spp., *Schwanniomyces* spp. (e.g. *S. occidentalis*) and the like.

In one embodiment, the yeast is a *Saccharomyces* species, and in particular, *Saccharomyces cerevisiae*. The yeast is typically a *S. cerevisiae* strain which is capable of converting sugars to alcohol on an industrial scale. Such sugars could be derived from hydrolysed starch or other abundant hexose sugar-rich feedstocks.

The sugars from the yeast which are converted to alcohol can comprise glucose.

The alcohol can be butanol or ethanol. In one embodiment, the alcohol is ethanol.

The recombinant yeast of the invention is capable of hydrolysing raw starch in the absence of enzymes from a source other than the recombinant yeast. However, additional amylolytic enzymes can optionally be added to a cold fermentation process using the recombinant yeast, so as to reduce the time taken to convert the starch to alcohol. These enzymes are typically a

cocktail of enzymes which can hydrolyse raw starch, such as STARGEN™. When the enzymes are added to the cold fermentation process, they are required in a reduced amount compared to the dosage that would be required if a different yeast was being used, e.g. the Transferm™ yeast. For example, the enzymes can be added in an amount which is about 50% to about 95% less than the dosage which is used in commercial cold fermentation processes. In the examples below, the enzymes were added in an amount which was 90% less than the amount which is added in existing commercial processes.

The recombinant yeast can hydrolyse the raw starch at a temperature of below about 40°C, such as 37°C or even below 30°C. For example, the recombinant yeast was shown to be effective at 26°C.

A process for producing an alcohol from sugars is also provided, wherein the recombinant yeast of the invention is used to convert the sugars to alcohol. The sugars can be derived from hydrolysed starch, from other abundant hexose sugar-rich feedstocks (e.g. sugarcane) or from cellulose-derived sugar streams (i.e. with the addition of cellulase enzymes).

A process for producing an alcohol from starch is also provided, wherein the recombinant yeast described above is used to convert the starch to alcohol. The starch is preferably raw starch, although the yeast could also be used to convert soluble starch to alcohol. This can be done in a single step process, even when the feedstock is raw starch.

In the experiments detailed below, alpha-amylases and glucoamylases from *Aspergillus terreus*, *Aureobasidium pullulans*, *Chaetomium thermophilum*, *Humicola grisea*, *Neosartorya fischeri*, *Rhizomucor pusillus*, *Talaromyces emersonii*, *Talaromyces stipitatus* and *Thermomyces lanuginosus* were screened for activity on starch and compared to the *S. cerevisiae* Y294[AmyA] and Y294[GlaA] benchmark strains, respectively (Viktor *et al*., 2013). Thereafter, several different amylolytic *S. cerevisiae* Y294 strains (ATCC 201160) were constructed and compared to the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain (Viktor *et al*., 2013) for their ability to hydrolyse raw corn starch and ferment the resulting glucose to ethanol at a high substrate loading (200 $g.l^{-1}$ raw corn starch).

A glucoamylase from *T. emersonii* (TemG) (SEQ ID NO: 1) was found to be the best enzyme for converting sugars to alcohol.

A combination of a glucoamylase from *T. emersonii* (TemG) (SEQ ID NO: 1)) and an α-amylase from *T. emersonii* (TemA (SEQ ID NO: 2)) was found to be the most efficient at hydrolysing raw corn starch at fermentation conditions. Further investigations showed that when these enzymes were expressed in yeast, a combination of the codon optimised glucoamylase gene (*temG_Opt* (SEQ ID NO: 3)) and native α-amylase gene (*temA_Nat* (SEQ ID NO: 4)) provided even better results than when the native glucoamylase gene (*temG_Nat* (SEQ ID NO: 5)) was used or when both genes had been codon optimised. For example, the recombinant *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain expressing the codon optimised glucoamylase and native α-amylase from *T. emersonii* produced 51.71 $g.l^{-1}$ ethanol from raw starch after 120 hours of fermentation compared to 33.14 $g.l^{-1}$ produced by the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain. The *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain displayed a 85% product yield after 192 hours, compared to the 54% by the benchmark strain.

The codon optimised *T. emersonii* glucoamylase gene (*temG_Opt* (SEQ ID NO: 3)) and native *T. emersonii* α-amylase gene (*temA_Nat* (SEQ ID NO: 4)) were then transformed into two commercially available industrial *S. cerevisiae* strains, namely Ethanol Red™ and the M2n (MH-1000) distillery yeast (Favaro *et al.*, 2015). Ethanol Red™ is one of the most widely used yeast strains for first generation bioethanol production (Stovicek *et al*., 2015).

Two δ-integration gene cassettes were constructed to allow for the simultaneous multiple integration of the codon-optimised *T. emersonii* glucoamylase gene (*temG_Opt*) and the native *T. emersonii* α-amylase gene (*temA_Nat*) into the genomes of the yeasts. The *T. emersonii* amylases were both constitutively expressed under the control of the *ENO1* promoter, using the δ-integration DNA transformation system. The amylolytic industrial strains were evaluated at high solids loadings and were able to ferment starch to ethanol in a single step with ethanol yields close to the theoretical maximum yield. After 192 hours at 30°C, the *S. cerevisiae* Ethanol Red™ T12 and M2n T1 strains (containing integrated *temA_Nat* and *temG_Opt* gene cassettes) produced 86.45 $g.l^{-1}$ and 99.40 $g.l^{-1}$ ethanol, respectively, corresponding to a product yield of 83.98% and 95.56%, respectively.

The addition of STARGEN 002™ in combination with the recombinant amylolytic yeast allowed for a 90% reduction in the enzyme dosage, compared to the conventional simultaneous saccharification (SSF) process with the untransformed host strains.

It is envisaged that the amylolytic strains of the present invention could be used as a drop-in candidate for existing cold fermentation processes.

*Glossary of terms*

As used herein, the singular forms "a", "an" and "the" include the plural references unless the content clearly dictates otherwise. Thus for example, reference to a composition containing "a compound" includes a reference to a mixture of two or more compounds. It should be noted that the term "or" is generally employed in the sense including "and/or" unless the context dictates otherwise.

The term "about" as used in relation to a numerical value means, for example, within 50% (±50%) of the numerical value, preferably ±30%, ±20%, ±15%, ±10%, ±7%, ±5%, or ±1%. Where necessary, the word “about” may be omitted from the definition of the invention.

The term “comprising” means “including”. Thus, for example, a composition or polypeptide “comprising” X may consist exclusively of X or may include one or more additional components. In some embodiments, “comprising” means “including principally, but not necessarily solely”.

As used herein, "heterologous" in reference to a nucleic acid or protein includes a molecule that has been manipulated by human intervention so that it is located in a place other than the place in which it is naturally found. For example, a nucleic acid sequence from one organism (e.g. from one strain or species) may be introduced into the genome of another organism (e.g. of another strain or species). A heterologous protein includes, for example, a protein expressed from a heterologous coding sequence or a protein expressed from a recombinant gene in a cell that would not naturally express the protein.

The terms "polypeptide" and "protein" are used interchangeably.

The term “alpha-amylase” refers to the EC 3.2.1.1 class of enzymes (1,4-alpha-D-glucan glucanohydrolase) which catalyse the hydrolysis of alpha-1,4-glucosidic linkages. The enzymes are endohydrolases, employ a retaining mechanism for hydrolysis (Enzyme Nomenclature, 1992) and belong to the glycoside hydrolase (GH) Family 13 and clan GH-H (MacGregor *et al*., 2001). They hydrolyse the 1,4-alpha-D-glucosidic linkages in polysaccharides containing three or more 1,4-alpha-linked D-glucose units. Hydrolysis reduces the molecular size of starch and therefore the viscosity of the starch solution. The alpha-amylases have considerably low sequence similarity.

Glucoamylases (glucan α-1,4-glucosidase, EC 3.2.1.3) belong to GH Family 15. Glucoamylases are exo-acting enzymes which catalyse the hydrolysis of α-1,4- and α-1,6-glucosidic linkages, thereby releasing the inverted β-d-glucose from the non-reducing ends of starch.

Further information of the structure and function of glucoamylases and alpha-amylases may be found in Christiansen *et al.* FEBS Journal 276 (2009) 5006–5029.

The phrases "percent identity", "% identity," “protein identity”, “sequence identity” etc. as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences. Percent identity may be determined using one or more computer algorithms or programs known in the art. For example the UWGCG Package provides the BESTFIT program which can be used to calculate sequence identity (for example used on its default settings) (Devereux *et al*. (1984) Nucleic Acids Research 12, p387-395). The PILEUP and BLAST (Basic Local Alignment Search Tool) algorithms can be used to calculate sequence identity or line up sequences (typically on their default settings), for example as described in Altschul S. F. (1993) J Mol Evol 36:290-300 and in Altschul, S, F *et al.* (1990) J Mol Biol 215:403. Software for performing BLAST analyses is available from several sources, including the National Center for Biotechnology Information (NCBI), Bethesda, MD, and on the internet at, for example, "www.ncbi.nlm.nih.gov/". Preferably, the default settings of the aforementioned algorithms / programs are used.

Whether an amino acid can be substituted at all (or deleted), or whether it can only be substituted by a conserved amino acid can be determined by comparing the amino acid sequence of one or more members of the protein family. Amino acids that are identical in all the members of a protein family often cannot be substituted. Amino acids which are conserved can usually be substituted by other conserved amino acids without significantly affecting the protein's function. Amino acids which are not conserved within a family can usually be freely substituted. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological activity may also be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR). Guidance concerning how to make phenotypically silent amino acid substitutions is provided, for example, in J. U. Bowie *et al*., "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," Science 247:1306-1310 (1990). Also, it will be recognized by those skilled in

the art that there may be critical areas on the protein which determine activity, such as the starch binding domain (SBD) and catalytic domain. The skilled person will appreciate that it may be desirable to take into account these areas when determining what changes to the amino acid sequence can be made. A detailed overview of SBDs may be found in Machovič and Janeč, 2006. Amino acid residues essential to activity of the polypeptide, and therefore preferably not subject to alteration e.g. by substitution or deletion (or if substituted only substituted by conservative substitutions), may be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (see, e.g., Cunningham and Wells, 1989, Science 244: 1081-1085). Sites of substrate-enzyme interaction can also be determined by analysis of the three-dimensional structure as determined by such techniques as nuclear magnetic resonance analysis, crystallography or photoaffinity labelling (see, e.g., de Vos *et al*., 1992, Science 255: 306-312; Smith *et al*., 1992, Journal of Molecular Biology 224: 899-904; Wlodaver *et al*., 1992, FEBS Letters 309: 59-64). Amino acid deletions, substitutions or additions remote from an active or binding site of a protein are generally more easily tolerated. In general, it is often possible to replace residues which form the tertiary structure, provided that residues performing a similar function are used. In other instances, the type of residue may be completely unimportant if the alteration occurs at a non-critical region of the protein.

The term "starch" refers to any material comprised of the complex polysaccharide carbohydrates of plant, comprised of amylose and amylopectin with the formula (C6H10O5)x, wherein X can be any number. In some embodiments, the starch-containing material may comprise xylan. Examples of “starch-containing” material include plant-based substrates (which may be fractionated plant material, for example a cereal grain such as corn, which is fractionated into components such as fiber, germ, protein and starch (endosperm)), tubers, roots, stems, whole grains, grains, corms, cobs, tall grasses, wheat, barley, rye, milo, sago, tapioca, rice peas, beans, arrow root, cassava, sweet potatoes, cereals, sugar-containing raw materials (e.g. molasses, fruit materials, sugar cane or sugar beet), potatoes, cellulose-containing materials (e.g. wood, wood residues, lignocelluloses, plant residues), wastes from agriculture (e.g. corn stover, rice straw, cereal, bran, damaged cereals, damaged potatoes, potato peel), non-cellulosic feed stocks such as sorghum, municipal waste (e.g. newspaper, waste paper), manure biomass, and agricultural residues etc.

The term "raw starch" refers to granular (unmodified) uncooked starch that has not been subjected to gelatinisation. At about 25°C, starch granules start absorbing water, and as the temperature increases, the granules start to vibrate vigorously. Crystallinity decreases, and

when the starch and water suspension is heated above a critical point, designated the pasting or gelatinisation temperature, the granules disintegrate to make a paste.

The term “hydrolysis of starch” refers to the chemical breakdown of glucosidic bonds with the addition of water molecules.

The terms “liquefaction,” “liquefy,” “liquefact,” and variations thereof refer to the process or product of converting starch to soluble dextrinized substrates (e.g. smaller polysaccharides). Liquefact can also be referred to as “mash.”

The term “gelatinisation” refers to the alteration of the starch granule from ordered, semi-crystalline granules to an amorphous state and occurs in the presence of water. This is generally done by heating the treated starch (typically treated with alpha amylase) to temperatures up to 100°C. The exact temperature of gelatinisation depends on the specific starch, and can readily be determined by the skilled person.

The term “gelatinisation temperature” refers to the lowest temperature at which gelatinisation of a starch containing substrate begins.

The term “soluble starch” refers to starch resulting from the hydrolysis of insoluble starch (e.g. granular/raw starch).

The terms “granular starch hydrolysing (GSH) enzyme” and “enzymes having granular starch hydrolysing (GSH) activity” refer to enzymes that are able to hydrolyse uncooked/granular starch.

The terms “saccharifying enzyme” and “starch hydrolysing enzyme” refer to any enzyme that is capable of converting starch to mono- or oligosaccharides (eg a hexose or pentose).

The phrase “consolidated bioprocessing” refers to a one-step process involving the use of a single organism that is able to achieve liquefaction, hydrolysis and fermentation of starch in a single fermentation vessel

The phrase “simultaneous saccharification and fermentation (SSF)” refers to a process in the production of end products in which a fermenting organism, such as an ethanol producing microorganism and at least one enzyme, such as a saccharifying enzyme, are combined in the same process step in the same vessel.

Yeasts do not form an exact taxonomic or phylogenetic grouping, but rather it is the colloquial name for single-celled members of the fungal divisions Ascomycota and Basidiomycota. The budding yeasts ("true yeasts") are classified in the order Saccharomycetales. Most reproduce asexually by budding, although a few do so by binary fission. Yeasts are unicellular, although some species with yeast forms may become multicellular through the formation of a string of connected budding cells known as pseudohyphae, or false hyphae as seen in most molds.

The invention will now be described in more detail by way of the following non-limiting examples.

**Example 1: Evaluation of α-amylases and glucoamylases and combinations thereof for raw starch hydrolysis**

**Materials and methods**

*Media and cultivation conditions*

All chemicals were of analytical grade and were obtained from Merck (Darmstadt, Germany), unless otherwise stated. *Escherichia coli* DH5α (Takara Bio Inc.) was used for vector propagation. The *E. coli* transformants were selected for on Luria Bertani agar (Sigma-Aldrich, Germany), containing 100 µg.ml$^{-1}$ ampicillin and cultivated at 37°C in Terrific Broth (12 g.l$^{-1}$ tryptone, 24 g.l$^{-1}$ yeast extract, 4 ml.l$^{-1}$ glycerol, 0.1 M potassium phosphate buffer) containing 100 µg.ml$^{-1}$ ampicillin for selective pressure (Sambrook *et al*., 1989).

The *S. cerevisiae* Y294 strain was maintained on YPD plates (10 g.l$^{-1}$ yeast extract, 20 g.l$^{-1}$ peptone and 20 g.l$^{-1}$ glucose and 15 g.l$^{-1}$ agar ) and amylolytic transformants were selected and maintained on SC$^{-URA}$ plates (containing 6.7 g.l$^{-1}$ yeast nitrogen base without amino acids (BD-Diagnostic Systems, Sparks, MD), 20 g.l$^{-1}$ glucose, 1.5 g.l$^{-1}$ yeast synthetic drop-out medium supplements (Sigma-Aldrich, Germany), 2% corn starch (Sigma-Aldrich, Germany) and 15 g.l$^{-1}$ agar). The *S. cerevisiae* strains were aerobically cultivated on a rotary shaker (200 rpm) at 30°C, in 125 ml Erlenmeyer flasks containing 20 ml double strength SC$^{-URA}$ medium (2×SC$^{-URA}$ containing 13.4 g.l$^{-1}$ yeast nitrogen base without amino acids (BD-Diagnostic Systems, Sparks, MD), 20 g.l$^{-1}$ glucose and 3 g.l$^{-1}$ yeast synthetic drop-out medium

supplements (Sigma-Aldrich, Germany). All cultures were inoculated to a concentration of $1\times10^6$ cells.ml$^{-1}$.

*Strains and plasmids*

The genotypes of the bacterial and fungal strains, as well as the plasmids used in this example, are summarised in Table 1.

Table 2. Strains and plasmids used in this example

| Strains and plasmids | Genotype | Reference |
|---|---|---|
| *E. coli DH5α* | *supE44 ΔlacU169 (ϕ80lacZ*ΔM15) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1* | Sambrook *et al.* (1989) |
| ***S. cerevisiae* strains** | | |
| Y294 | *α leu2*-3,112 *ura3*-52 *his3 trp1-289* | ATCC 201160 |
| Y294[BBH1] | *URA3 ENO1*$_P$-*ENO1*$_T$ | Viktor *et al*. (2013) |
| Y294[AmyA][1] | *URA3 ENO1*$_P$-*amyA-ENO1*$_T$ | Viktor *et al*. (2013) |
| Y294[GlaA][1] | *URA3 ENO1*$_P$-*glaA-ENO1*$_T$ | Viktor *et al*. (2013) |
| Y294[AmyA-GlaA][1] | *URA3 ENO1*$_P$-*amyA-ENO1*$_T$; *ENO1*$_P$-*glaA-ENO1*$_T$ | Viktor *et al*. (2013) |
| Y294[ApuA_Nat][1] | *URA3 ENO1*$_P$-*apuA_Nat-ENO1*$_T$ | This study |
| Y294[ApuA_Opt-NatSS][1] | *URA3 ENO1*$_P$-*NatSS-apuA_Opt-ENO1*$_T$ | This study |
| Y294[ApuA_Nat-XYNSEC] | *URA3 ENO1*$_P$- *XYNSEC-apuA_Nat-ENO1*$_T$ | This laboratory |
| Y294[ApuA_Opt-XYNSEC] | *URA3 ENO1*$_P$- Opt*XYNSEC-apuA_Opt-ENO1*$_T$ | This study |
| Y294[ApuA_Opt-OptXYNSEC] | *URA3 ENO1*$_P$- Opt*XYNSEC-apuA_Opt-ENO1*$_T$ | This study |
| Y294[AteA_Nat] [1] | *URA3 ENO1*$_P$-*ateA_Nat-ENO1*$_T$ | This study |
| Y294[TemA_Nat][1] | *URA3 ENO1*$_P$-*temA_Nat-ENO1*$_T$ | This study |
| Y294[TemA_Opt] | *URA3 ENO1*$_P$-*temA_Opt-ENO1*$_T$ | This study |
| Y294[TemA_Opt-XYNSEC] | *URA3 ENO1*$_P$-*XYNSEC-temA_Opt-ENO1*$_T$ | This study |
| Y294[TemA_Nat- XYNSEC] | *URA3 ENO1*$_P$-*XYNSEC-temA_Nat-ENO1*$_T$ | This study |
| Y294[TemA_Opt-NatSS][1] | *URA3 ENO1*$_P$-*NatSS-temA_Opt-ENO1*$_T$ | This study |
| Y294[AteG_Nat][1] | *URA3 ENO1*$_P$-*ateG_Nat-ENO1*$_T$ | This study |
| Y294[AteG_Nat-XYNSEC] | *URA3 ENO1*$_P$-*XYNSEC-ateG_Nat-ENO1*$_T$ | This study |
| Y294[AteG_Opt-XYNSEC] | *URA3 ENO1*$_P$-*XYNSEC-ateG_Opt-ENO1*$_T$ | This study |
| Y294[AteG_Opt-NatSS] | *URA3 ENO1*$_P$-*NatSS-ateG_opt-ENO1*$_T$ | This study |

| | | |
|---|---|---|
| Y294[TemG_Nat][1] | *URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$ | This study |
| Y294[TemG_Opt] | *URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$ | This study |
| Y294[TemG_Opt-XYNSEC][1] | *URA3 ENO1*$_P$*-XYNSEC-temG_Opt-ENO1*$_T$ | This study |
| Y294[TemG_Nat-XYNSEC][1] | *URA3 ENO1*$_P$*-XYNSEC-temG_Nat-ENO1*$_T$ | This study |
| Y294[TemG_Opt-NatSS][1] | *URA3 ENO1*$_P$*-NatSS-temG_Opt-ENO1*$_T$ | This study |
| Y294[TemG_Opt-AmyA] | *URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-amyA-ENO1*$_T$ | This study |
| Y294[TemG_Opt-TemA_Nat] | *URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| Y294[TemG_Opt-TemA_Opt] | *URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Opt-ENO1*$_T$ | This study |
| Y294[TemG_Opt-AteA_Nat] | *URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-ateA_Nat-ENO1*$_T$ | This study |
| Y294[TemG_Opt-ApuA_Nat] | *URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-apuA_Nat-ENO1*$_T$ | This study |
| Y294[GlaA_Nat-TemA_Nat] | *URA3 ENO1*$_P$*-glaA-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| Y294[TemG_Nat-AmyA] | *URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-amyA-ENO1*$_T$ | This study |
| Y294[TemG_Nat-AteA_Nat] | *URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-ateA_Nat-ENO1*$_T$ | This study |
| Y294[TemG_Nat-ApuA_Nat] | *URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-apuA_Nat-ENO1*$_T$ | This study |
| Y294[AteG_Nat-XYNSEC-AmyA] | *URA3 ENO1*$_P$*-ateG_Nat-ENO1*$_T$; *ENO1*$_P$*-amyA-ENO1*$_T$ | This study |
| **Plasmids** | | |
| yBBH1 | *bla URA3 ENO1*$_P$*-ENO1*$_T$ | Njokweni *et al*. (2012) |
| yBBH4 | *bla URA3 ENO1*$_P$*-XYNSEC-ENO1*$_T$ | Njokweni *et al*. (2012) |
| yBBH1-AmyA | *bla URA3 ENO1*$_P$*-amyA-ENO1*$_T$ | Viktor *et al*. (2013) |
| yBBH1-GlaA | *bla URA3 ENO1*$_P$*-glaA-ENO1*$_T$ | Viktor *et al*. (2013) |
| yBBH1-AteA_Nat | *bla URA3 ENO1*$_P$*-ateA_Nat-ENO1*$_T$ | This study |
| yBBH1-ApuA_Nat | *bla URA3 ENO1*$_P$*-apuA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemA_Nat | *bla URA3 ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemA_Opt | *bla URA3 ENO1*$_P$*-temA_Opt-ENO1*$_T$ | This study |
| yBBH1-AteG_Nat-XYNSEC | *bla URA3 ENO1*$_P$*-ateG_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Nat | *bla URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$ | This study |
| yBBH1-TemG_Nat-ApuA_Nat | *bla URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-apuA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Nat-AmyA | *bla URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-amyA-ENO1*$_T$ | This study |
| yBBH1-TemG_Nat-AteA_Nat | *bla URA3 ENO1*$_P$*-temG_Nat-ENO1*$_T$; *ENO1*$_P$*-ateA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt-ApuA_Nat | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-apuA_Nat-ENO1*$_T$ | This study |

| | | |
|---|---|---|
| yBBH1-TemG_Opt-AmyA | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-amyA-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt-AteA_Nat | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-ateA-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt-TemA_Nat | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt-TemA_Opt | *bla URA3 ENO1*$_P$*-temG_Opt -ENO1*$_T$; *ENO1*$_P$*-temA_Opt-ENO1*$_T$ | This study |
| yBBH1-GlaA-TemA_Nat | *bla URA3 ENO1*$_P$*-glaA-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| yBBH4-AteG_Nat-XYNSEC-AmyA | *bla URA3 ENO1*$_P$*-XYNSEC-ateG_Nat-ENO1*$_T$; *ENO1*$_P$*-amyA-ENO1*$_T$ | This study |

[1] native secretion signal
_Nat: native coding sequence;
_Opt: codon optimised coding sequences (GenScript);
-NatSS: native secretion signal;
-XYNSEC: native secretion signal from *Trichoderma reesei Xyn2* gene,
-OptXYNSEC: codon optimised-XYNSEC secretion signal

*DNA manipulations*

Standard protocols were followed for all DNA manipulations and *E. coli* transformations (Sambrook *et al*., 1989). All genes were synthesised by GenScript (Piscataway, NJ, USA), based on the nucleotide accession numbers listed below. The internal *Eco*RI, *Xho*I, *Bam*HI and *Bgl*II restriction sites were avoided, but the amino acid sequence remained unaffected. The polymerase chain reaction (PCR) was performed using a Perkin Elmer Gene Amp® PCR System 2400 and TaKaRa Ex Taq™ (Takara Bio Inc, Japan) as per the manufacturer's recommendations. The amylase genes were amplified using primers (Inqaba Biotec, South Africa) (Table 2) designed for yeast mediated ligation (YML) and visualised on a 0.8% agarose gel. DNA was eluted from agarose gels with the Zymoclean™ Gel Recovery Kit (Zymo Research, USA).

Table 2. PCR oligo-primers used in this study with the relevant restriction sites underlined (*Eco*RI = gaattc; *Nru*I = tcgcga; *Xho*I = ctcgag)

| Gene name (host organism) | Sequence (5'-3') | SEQ ID NO: | Signal peptide[1] |
|---|---|---|---|
| *apuA* (*A. pullulans)* | ApuA_Nat-L: tgcttatcaacacacaaacactaaatcaaagaattcatggcagccaactacgtttctcgattgttg | 8 | 22 |
| | ApuA_N-R:gactagaaggcttaatcaaaagctctcgagtcacccctgccaagtattgctgaccgatgc | 9 | |
| | ApuA_Opt-NatSS-L: tctctacttgaccgggttggtgcagtgtttgactccagctcaatggagaagtcaatctat | 10 | |
| | ApuA_Opt-R: ggactagaaggcttaatcaaaagctctcgagctaaccttgccatgtattggagactgagg | 11 | |
| | ApuA_optXynSec-L: gaacccgtggctgtggagaagcgctcgcgattgactccagctcaatggagaagtc | 12 | |
| | ApuA_Opt-R: ggactagaaggcttaatcaaaagctctcgagctaaccttgccatgtattggagactgagg | 13 | |
| *ateA* (*A. terreus)* | AteA_Nat-L: tgcttatcaacacacaaacactaaatcaaagaattcatgaagtggacctcctcgctcctcctctta | 14 | 20 |
| | AteA_Nat-R: gactagaaggcttaatcaaaagctctcgagtcacctccaagtatcagcaactgtcaccgt | 15 | |
| *temA* (*T. emersonii)* | TemA_Nat-L: tgcttatcaacacacaaacactaaatcaaagaattcatgacgcctttcgtcctcacggcc | 16 | 19 |
| | TemA_Nat-R ggactagaaggcttaatcaaaagctctcgagctatctccatgtgtcgacaatcgtctccg | 17 | |
| | TemA_Opt-NatOptSS-L: tgcttatcaacacacaaacactaaatcaaagaattcatgacccctttgttttgacagcc | 18 | |
| | TemA_Opt-R: ggactagaaggcttaatcaaaagctctcgagctatctccaagtgtcaacaatagtttcag | 19 | |
| | TemA_Nat-xynsecSS-L: gaacccgtggctgtggagaagcgctcgcgattgaccccggccgaatggcgcaaacaat | 20 | |
| | TemA_Opt-xynsecSS-L gaacccgtggctgtggagaagcgctcgcgattgacaccagccgaatggagaaagcaatc | 21 | |
| | TemA_Opt-NatSS-L: tcttgctggggaatgccgtgttggccttgacaccagccgaatggagaaagc | 22 | |
| *ateG* (*A. terreus)* | AteG_Nat-L: tgcttatcaacacacaaacactaaatcaaagaattcatgacgcgcattctcaccctcgcccttcat | 23 | 20 |
| | AteG_Nat-R: ggactagaaggcttaatcaaaagctctcgagctagcgccaagtggtgttcaccaccgcggt | 24 | |
| | AteG_Opt-NatSS-L: gggctggctcttgtccaaagtgttgttggggcaccacaattggctcctagagcaactaca | 25 | |
| | AteG_Opt-R: tggactagaaggcttaatcaaaagctctcgagctatctccaggttgtgttgacaacggcg | 26 | |
| | AteG_Nat-xynSS-L: gaacccgtggctgtggagaagcgctcgcgagctccccaattggcccccagagcgacaacc | 27 | |
| *temG* (*T. emersonii)* | TemG_Nat-L: tgcttatcaacacacaaacactaaatcaaagaattcatggcgtccctcgttgctggcgctctctgc | 28 | 20 |
| | TemG_Nat-R: ggactagaaggcttaatcaaaagctctcgagtcactgccaactatcgtcaagaatggcggt | 29 | |
| | TemG_Nat-xynsecSS-L:gaacccgtggctgtggagaagcgctcgcgacgagcgcccgttgcagcgcgagccaccggt | 30 | |
| | TemG_Opt-xynsecSS-L: gaacccgtggctgtggagaagcgctcgcgaagagccccagtcgcagccagagcaacaggt | 31 | |
| | TemG_Opt-R: gactagaaggcttaatcaaaagctctcgagtcattgccaagagtcgtccaagattgcggt | 32 | |
| | TemG_Opt-NatOptSS-L: ttatcaacacacaaacactaaatcaaagaattcatggcctccttagtcgcaggtgcctta | 33 | |
| | TemG_Opt-NatSS-L: atcctgggcctgacgcctgctgcatttgcaagagccccagtcgcagccagagcaacaggt | 34 | |

[1]The length (amino acids) of putative signal peptides was analysed using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP).

The amylase genes were subcloned individually onto the yBBH1 or yBBH4 plasmid (Figures 1a, b and c) in order to construct the expression vectors listed in Table 1. The yBBH4 vector (Figure 1c) contained the sequence encoding for the *XYNSEC* secretion signal of the *Trichoderma reesei xyn2* (Den Haan *et al*., 2007) for directing the secretion of the amylases. The *ENO1*$_P$-α-amylase-*ENO1*$_T$ cassettes were amplified from the yBBH1-α-amylase vectors using YML cassette primers: ENOCASS-L: gtgcggtatttcacaccgcataggagatcgatcccaattaatgtgagttacctcactc (SEQ ID NO: 35) and ENOCASS-R: cgggcctcttcgctattacgccagagcttagatct (SEQ ID NO: 36) and cloned on the *Bgl*II site of yBBH1-glucoamylase or yBBH4-glucoamylase vectors (Figures 1c and d). Sequence verification of the final vector constructs was performed by the dideoxy chain termination method, with an ABI PRISM™ 3100 Genetic Analyser (CAF, Stellenbosch University).

*Amylase genes and GenBank Accession numbers*

The following amylases were cloned and expressed in *S. cerevisiae* Y294. The native glucoamylases from *A. pullulans (*Accession no. HM246718), *A. terreus* (Accession no. XP_001213553), *H. grisea (*Accession no. M89475), *T. emersonii* (Accession no. AJ304803) and *T. lanuginosus (*Accession no. EF545003), as well as the native α-amylases from *A. pullulans* (Accession no. AEH03024), *A. terreus* (Accession no. XM_001209405), *N. fischeri* (Accession no. XP_001265628), *R. pusillus* (Accession no. AGJ52081) and *T. emersonii* (Accession no. XM_013469492). Coding sequences for the glucoamylases from *C. thermophilum* (Accession no. ABD96025), *T. stipitatus* (Accession no. XP_002484948), *A. terreus* and *T. emersonii,* as well for α-amylases from *A. pullulans* and *T. emersonii* were codon optimised for expression in *S. cerevisiae* (GenScript, Piscataway, NJ, USA). *T. emersonii* has recently been classified as *Rasamsonia emersonii* (Houbraken *et al*., 2012).

*Yeast transformations*

The *S. cerevisiae* Y294 strain was grown overnight in 5 ml YPD broth and prepared according to Cho *et al*. (1999). After electroporation, 1 ml of YPDS was immediately added to the cuvette. Cultures were incubated at 30°C for 1 hour prior to plating out onto SC$^{-URA}$ plates containing 2% starch. Plates were incubated at 30°C for 2 - 3 days and then transferred to 4°C for 24 hours to allow the starch to precipitate.

*Activity assays*

For quantitative assays, yeast transformants were cultured in 20 ml 2×SC$^{-URA}$ medium in 125 ml Erlenmeyer flasks with agitation at 200 rpm and sampling at 24 hour intervals. The supernatant was harvested and extracellular enzymatic activity levels were assessed colourimetrically (xMark$^{TM}$ Microplate Spectrophotometre, Bio-Rad, San Francisco, USA) using the reducing sugar assay with glucose as standard (Miller 1959). The α-amylase activities were determined after a 5 minute incubation with 0.2% soluble corn starch in 0.05 M citrate-acid buffer (pH 5) at 37°C.

Glucoamylase activity was determined by incubating 50 µl supernatant with 450 µl of 0.2% soluble corn starch in 0.05 M citrate-acid buffer (pH 5) at 37°C for 15 minutes. The glucose concentration was determined using the D-Glucose Assay Kit (Megazyme, Ireland) with absorbance measured at 510 nm (xMark$^{TM}$ Microplate Spectrophotometre, Bio-Rad, San Francisco, USA). Enzymatic activities were expressed as nano-katals per ml (nkat.ml$^{-1}$), with nkat defined as the enzyme activity needed to produce 1 nmol of glucose per second under the described assay conditions.

*Protein analysis*

Recombinant *S. cerevisiae* Y294 strains were cultivated in 125 ml Erlenmeyer flasks containing 20 ml 2×SC$^{-URA}$ medium for 3 days. Twenty microliters of supernatant was added to protein loading buffer and the samples boiled for 3 minutes to denature the proteins. The recombinant proteins were separated on an 8% SDS-polyacrylamide gel using a 5% stacking gel and Tris-glycine buffer (Sambrook *et al*., 1989). Electrophoresis was carried out at 100 V for ± 90 minutes at ambient temperature and protein species were visualised using the silver staining method (O'Connell and Stults, 1997). The broad-range Page Ruler Prestained SM0671 Protein Ladder (Fermentas, China) was used as a molecular mass marker.

*Fermentation studies*

Precultures were cultured in 60 ml 2×SC$^{-URA}$ media in 250 ml Erlenmeyer flasks and incubated at 30°C with agitation of 200 rpm. Fermentations were performed with 2×SC$^{-URA}$ media containing 200 g.l$^{-1}$ raw corn starch and 5 g.l$^{-1}$ glucose and inoculated with a 10% (v.v$^{-1}$) inoculum. Ampicillin (100 µg.ml$^{-1}$) and streptomycin (50 µg.ml$^{-1}$) were added to inhibit bacterial contamination. Agitation and incubation were performed on a magnetic multi-stirrer at 30°C, with daily sampling through a syringe needle pierced through the rubber stopper.

For bioreactor experiments, precultures were cultivated in 120 ml 2×$SC^{-URA}$ media in 500 ml Erlenmeyer flasks at 30°C with agitation at 200 rpm. Bioreactor fermentations were performed in a 2 litre MultiGen Bioreactor (New Brunswick Scientific Corporation, Edison, New Jersey) containing 2×$SC^{-URA}$ media supplemented with 200 g.$l^{-1}$ raw corn starch and 5 g.$l^{-1}$ glucose as carbohydrate source. A 10% (v.$v^{-1}$) inoculum was used in a total working volume of 1 litre. Fermentations were carried out at 26°C and 30°C with stirring at 300 rpm and daily sampling through a designated sampling port. All fermentation experiments were performed in triplicate.

*High Performance Liquid Chromatography (HPLC) analysis*

Ethanol, glucose, maltose, glycerol and acetic acid concentrations were quantified with HPLC using a Surveyor Plus liquid chromatograph (Thermo Scientific) consisting of a liquid chromatography pump, autosampler and refractive index (RI) detector. The compounds were separated on a Rezex RHM Monosaccharide 7.8×300 mm column (00H0132-K0, Phenomenex) at 60°C with 5 mM $H_2SO_4$ as mobile phase at a flow rate of 0.6 ml.$min^{-1}$.

*Analytical methods and calculations*

The theoretical $CO_2$ yields were calculated according to Favaro *et al*. (2015). The product yield (percentage starch converted to products) was calculated from ethanol, glucose, maltose, glycerol, acetic acid and $CO_2$ concentrations. The ethanol yield (% of the theoretical yield) was calculated as the amount of ethanol produced per gram of consumed sugar. The ethanol rate of productivity was calculated based on ethanol titres produced per hour (g.$l^{-1}$.$h^{-1}$).

*Statistical analysis*

Data was analysed using the Student's t-test.

## Results

*Functional expression of recombinant amylases*

The *S. cerevisiae* Y294 strain was used as host for the heterologous gene expression of recombinant amylases. Recombinant strains were constructed to express either an α-amylase or glucoamylase encoding gene (Table 1) and evaluated for their ability to hydrolyse corn starch using the *S. cerevisiae* Y294[AmyA] and Y294[GlaA] strains, respectively, as

benchmarks strains (Viktor *et al*., 2013). All of the recombinant strains evaluated in this study were able to hydrolyse soluble starch. The different native and optimised counterparts for the *ateA*, *apuA*, *temA*, *ateG* and *temG* genes contained different DNA sequences, but encoded for the same amino acid sequence.

*α-Amylases*

The *ateA_Nat* gene was efficiently expressed by the *S. cerevisiae* Y294[AteA_Nat] strain, but the extracellular levels of activity were consistently lower than that of the *S. cerevisiae* Y294[AmyA] benchmark strain (Figure 2a). Replacing the native secretion signal with the native *XYNSEC* (*S. cerevisiae* Y294[AteA_Nat-XYNSEC]) did not result in significant differences in either extracellular activity or the amount of AteA secreted (Figures 2a and 2d). The extracellular protein levels of AmyA and AteA were similar (Figure 2d).

The *S. cerevisiae* Y294[ApuA_Nat] and Y294[TemA_Nat] strains displayed more extracellular α-amylase activity on soluble starch (Figures 2b and 2c) than the *S. cerevisiae* Y294[AmyA] benchmark strain. Codon optimisation of the *apuA_Nat* and *temA_Nat* genes resulted in less extracellular activity due to a decrease in enzyme concentration (Figures 2e and 2f). Changing the secretion signal also resulted in a decrease in extracellular enzyme concentration, with a negative impact on extracellular activity (Figures 2c and 2d).

SDS-PAGE analysis of the supernatant indicated that most of these α-amylases are glycosylated. ApuA and AteA protein species (calculated molecular weights of 65.25 kDa and 64.14 kDa, respectfully) (Figures 2b and 2d) are the least glycosylated with a putative recombinant size of around 70 kDa, while TemA (calculated molecular weight of 66.29 kDa) had a higher degree of glycosylation (Figure 2f) and a putative size of around 90 kDa. The large heterogeneous smear between 110 and 150 kDa for the AmyA protein is consistent with that of a previous report (Viktor *et al*., 2013).

*Glucoamylases*

The replacement of the *ateG_Nat* secretion signal with the *XYNSEC* sequence improved extracellular glucoamylase activity, albeit less than the activity displayed by the *S. cerevisiae* Y294[GlaA] strain (Figure 3a). The *S. cerevisiae* Y294[AteG_Opt-XYNSEC] and Y294[AteG_Nat-XYNSEC] strains produced similar levels of activity, which exceeded the activity by the strains containing the native *ateG* secretion signal. The *S. cerevisiae* Y294[AteG_Opt-NatSS] strain secreted no visible protein (Figure 3c) confirming that the native *ateG* secretion signal negatively affected protein secretion. Codon optimisation did not have

a visible effect on the extracellular amount of AteG protein produced, despite the increase in extracellular activity (Figures 3a and 3c).

A significant increase in extracellular glucoamylase activity was observed when the *temG* gene was codon optimised (Figure 3b). At 72 hours, extracellular activity for the *S. cerevisiae* Y294[TemG_Opt] strain was > 3-fold higher than the *S. cerevisiae* Y294[TemG_Nat] strain and > 10-fold higher than the Y294[GlaA] benchmark strain. Changing secretion signals for the expression of the *temG* indicated that the optimised *temG* secretion signal contributed to enhanced protein secretion and extracellular activity (Figures 3b and 3d), whereas replacement with the XYNSEC had a negative impact.

SDS-PAGE analysis of the supernatant indicated that these glucoamylases are glycosylated. The AteG protein species (calculated molecular weight of 65.73 kDa) (Figure 3c) had a putative size of around 95 kDa, while the TemG protein (calculated molecular weight of 63.57 kDa) is less glycosylated with a putative size of around 85 kDa (Figure 3d). Moreover, the intensity of the recombinant protein species visualised using SDS-PAGE showed correlation with the extracellular enzyme activity levels for all amylases.

*Raw corn starch fermentations*

The amylase encoding genes that resulted in the highest levels of extracellular activity when expressed in *S. cerevisiae* Y294 (*apuA_Nat*, *ateA_Nat*, *temA_Nat, temA_Opt*, *ateG_Nat-XYNSEC*, *temG_Nat* and *temG_Opt*), together with the reference (*amyA* and *glaA*) genes, were then used to construct amylolytic strains that produced an α-amylase and glucoamylase combination (Table 1). The recombinant yeast strains were evaluated for their ability to hydrolyse raw starch and ferment glucose at a high substrate loading under oxygen-limited conditions.

At 192 hours, the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain produced the highest ethanol concentration (62.20 g. $l^{-1}$), which is 59.67% of the theoretical value (Figure 4a). After 120 hours, this strain produced 51.71 g. $l^{-1}$ ethanol, which represents a 1.6-fold improvement on the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain ($p$ = 0.0013). Ethanol levels of 38.57 g.$l^{-1}$ and 39.40 g.$l^{-1}$ produced by the *S. cerevisiae* Y294[TemG_Opt-ApuA_Nat] and Y294[TemG_Opt-AteA_Nat] strains, respectively, were also higher than the benchmark strain (at 120 hours). The *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain accumulated 46.30 g.$l^{-1}$ residual glucose after 192 hours of fermentation (Figure 4c).

The *S. cerevisiae* Y294 strains expressing the TemG_Nat-AmyA, TemG_Nat-AteA_Nat, TemG_Nat-ApuA_Nat and AteG_Nat-XYNSEC-AmyA enzyme combinations produced less ethanol compared to the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain (Figures 4a and 4b), with little to no residual glucose detected (Figure 4d). Overall, results depicted in Figure 4c indicated that the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain was superior to the other strains and this enzyme combination was effective in hydrolysing raw corn starch. At 192 hours, product yield obtained by the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain was 57% higher than that of the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain, whereas the Y294[TemG_Opt-AteA_Nat] strain produced comparable results to that of the benchmark strain (Table 3).

Table 3. Products formed by *S. cerevisiae* Y294 strains after 192 hours of fermentation at 30°C in 2×SC$^{-URA}$ broth with glucose (5 g.l$^{-1}$) and raw corn starch (200 g.l$^{-1}$)

| ***S. cerevisiae* Y294 strains** | **[TemG_Opt-AmyA]** | **[TemG_Opt-TemA_Nat]** | **[TemG_Opt-TemA_Opt]** | **[TemG_Opt-AteA_Nat]** | **[TemG_Opt-ApuA_Nat]** | **[GlaA-AmyA]** | **[GlaA-TemA_Nat]** |
|---|---|---|---|---|---|---|---|
| **Substrate (g.l$^{-1}$)** | | | | | | | |
| Raw starch (dry weight) | 185 | 185 | 185 | 185 | 185 | 185 | 185 |
| Glucose equivalent | 208.5 | 208.5 | 208.5 | 208.5 | 208.5 | 208.5 | 208.5 |
| **Products (g.l$^{-1}$)** | | | | | | | |
| Glucose | 2.72 | 46.30 | 1.67 | 1.94 | 1.21 | 5.30 | 4.12 |
| Glycerol | 4.76 | 6.64 | 2.40 | 3.43 | 2.45 | 2.46 | 2.26 |
| Maltose | 1.09 | 1.03 | 1.07 | 1.14 | 0.95 | 1.02 | 1.17 |
| Acetic acid | 1.91 | 1.66 | 0.60 | 0.85 | 0.61 | 0.61 | 0.56 |
| Ethanol | 47.40 | 62.20 | 48.71 | 53.46 | 43.12 | 52.78 | 46.56 |
| $CO_2$[1] | 45.33 | 59.50 | 46.59 | 51.13 | 41.25 | 50.48 | 44.53 |
| **Total** | 103.21 | 177.33 | 101.04 | 111.95 | 89.60 | 112.65 | 99.20 |
| Product yield (%) | 49.50 | 85.05 | 48.46 | 53.69 | 42.97 | 54.03 | 47.58 |
| Ethanol yield[2] (% of theoretical yield) | 45.46 | 59.67 | 46.72 | 51.28 | 41.36 | 50.63 | 44.66 |
| Ethanol rate of productivity[3] | 0.247 | 0.324 | 0.254 | 0.278 | 0.225 | 0.275 | 0.242 |

[1]$CO_2$ yields were deduced from the ethanol produced.
[2]Ethanol yield (% of the theoretical yield) was calculated as the amount of ethanol produced per gram of consumed sugar

[3]Ethanol rate of productivity was calculated based ethanol titres produced per hour ($g.l^{-1}.h^{-1}$)

The *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain was evaluated in a 2 litre bioreactor (1 litre working volume) under two fermentation temperatures (26°C and 30°C) (Figure 5). The ethanol concentrations increased significantly at a fermentation temperature of 26°C (Figure 5a). The percentage product yield was the similar at 144 hours (around 77%), but a decrease in fermentation temperature resulted in 66% and 83% improvement in ethanol concentration after 144 and 192 hours, respectively. No glucose was detected in the fermentation broth after 192 hours at a fermentation temperature of 26°C (Figure 5a). After 192 hours, the total product yield displayed by the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain was similar for both fermentation types (100 ml bottle fermentations and bioreactor), 85% and 81%, respectively (Figures 4a and 5b).

**Discussion**

A selection of amylases from various fungi have been investigated independently by several research groups, with raw starch hydrolysing enzymes being favoured for starch conversion to ethanol (Robertson *et al*., 2006; Viktor *et al*., 2013; Favaro *et al*., 2015; Celińska *et al*., 2015). Approximately 10% of all amylases contain a starch binding domain (SBD) (Sun *et al*., 2010), which is classically associated with the adsorption of these enzymes to raw starch granules, thereby enhancing the amylolytic rate and the subsequent hydrolysis (Santiago *et al*., 2005; Mitsuiki *et al*., 2005). Thus, for this study, the presence of a SBD was a prerequisite when selecting amylases for expression in *S. cerevisiae*.

Amylase genes were heterologously expressed in order to choose the enzymes with the highest extracellular enzyme activity and to investigate the effect of synonymous codon usage on gene expression (Table 1). In this study, several amylase candidates showed significantly low levels of extracellular activity, compared to the benchmark strain (data not shown). Thus, the following genes were omitted from further studies: native glucoamylases from *A. pullulans, H. grisea* and *T. lanuginosus,* as well as the optimised α-amylases from *N. fischeri, R. pusillus* and codon optimised glucoamylases from *C. thermophilum* and *T. stipitatus.*

High levels of protein expression can be correlated to the codon adaptation index (CAI) (Carbone *et al*., 2003). A CAI value of 1.0 is considered to be ideal, while GenScript recommends that a CAI of >0.8 is rated as good for expression in the desired expression organism. Analysis of the genes' CAI values using GenScript's OptimumGeneTM

(http://www.genscript.com/cgi-bin/tools/rare_codon_analysis) indicated that all CAI values increased when the genes were optimised. GenScript's algorithm for gene optimisation aims to improve gene expression and therefore the synthetic amylase genes in this study were codon optimised for expression in *S. cerevisiae*. However, results from this study indicated that increased gene expression and protein secretion was not guaranteed by codon optimisation (Figures 2 and 3).

The strains expressing the *apuA_Nat* and *temA_Nat* genes were superior to the strains expressing the codon optimised counterparts *apuA_Opt-NatSS*/*apuA_Opt-OptXYNSEC* and *temA_Opt*, respectively (Figures 2b and 2c), while optimisation of the *temG* coding sequence resulted in a significant increase in TemG_Opt protein secreted by the *S. cerevisiae* Y294[TemG_Opt] strain (Figure 3d). Increased recombinant protein secretion correlated with enhanced levels of extracellular activity, which suggested similar specific activities (Figures 2e and 2f) and SDS-PAGE analysis indicated that codon optimisation did not affect amylase protein size (Figures 2 and 3). Based on the deduced amino acid sequences, the predicted molecular weights of the unglycosylated amylases are around 64 - 70 kDa, which is in agreement with previous reports on similar amylases (Gupta *et al*., 2003).

The *temA_Nat* had a CAI of 0.61 compared to *temA_Opt* with a CAI of 0.91. However, the *S. cerevisiae* Y294[TemA_Nat] strain produced 59% more extracellular α-amylase activity than the *S. cerevisiae* Y294[TemA_Opt] strain after 72 hours. The *temG_Nat* gene had a CAI of 0.58 compared to *temG_Opt*, which had a CAI of 0.91. The extracellular glucoamylase activity for the *S. cerevisiae* Y294[TemG_Nat] and Y294[TemG_Opt] strains represented a > 3-fold and 10-fold fold improvement, respectively, compared to the *S. cerevisiae* Y294[GlaA] benchmark strain. Therefore, even for genes originating from the same species (in this case *T. emersonii*), significant differences in protein secretion and extracellular enzyme activities were observed between native and codon optimised genes. Thus, CAI values alone cannot be relied upon for improving gene expression.

The secretion of recombinant proteins into the culture medium simplifies downstream purification methods (Damasceno *et al*., 2012). Secretion signals are used to direct the propeptide to the endoplasmic reticulum (ER) and then through the secretory pathway (Futatsumori-Sugai and Tsumoto, 2010). Once in the ER, the mature peptide is folded into its native structure and there are a number of factors that effect this folding process (Tyo *et al*., 2012). The secretion of recombinant proteins by yeast is a key industrial objective for the biotechnology field, and significant efforts have gone into improving protein secretion. This process is dependent on the target protein, host strain and secretion signal sequence

(Hashimoto *et al*., 1998). Therefore, signal peptides represented an important factor to consider when improving the concentration of secreted protein.

The XYNSEC secretion signal from *Trichoderma reesei*'s β-xylanase 2 gene has been used successfully for the secretion of a number of proteins (van Wyk *et al*., 2010; van Rensburg *et al*., 2012; Favaro *et al*., 2013) and was used in this study for comparative purposes. All of the native enzymes selected for this study were successfully secreted using their native secretion peptides, and the replacement of the native *ateG* signal peptide encoding sequence with the *XYNSEC* sequence resulted in enhanced extracellular activity (Figure 3a). However, in general, the XYNSEC secretion signal was less effective than the proteins' native secretion signals.

Following the identification of successful amylase candidates, novel gene combinations were expressed in *S. cerevisiae* Y294 in order to obtain an amylolytic yeast suitable for raw starch CBP. It was previously reported that starch fermentation by genetically engineered strains is limited by the glucoamylase activity (Inlow *et al*., 1988), but in a more recent review the limiting factor in raw starch hydrolysis was attributed to α-amylase activity (Görgens *et al*., 2015). The type of starchy biomass (used as substrate) is likely to affect the ratio of amylases, but if a recombinant amylolytic yeast is able to produce highly active enzymes, an exact ratio should not be a limiting factor.

During cultivation on 200 $g.l^{-1}$ raw corn starch, simultaneous expression of the α-amylase and glucoamylase combinations in *S. cerevisiae* resulted in varying ethanol yields (Figures 4a and b). After 72 hours, the product yield displayed by the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain was 2.7-fold higher than the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain. The *S. cerevisiae* Y294[TemG_Opt-ApuA_Nat] and Y294[TemG_Opt-AteA_Nat] strains also outperformed the *S. cerevisiae* Y294[AmyA-GlaA] benchmark strain (Figure 4a) in the early stages of fermentation (> 2.4-fold higher ethanol concentrations after 48 hours). Substantially higher ethanol concentrations were obtained, compared to the modified amylolytic yeast strain constructed by Yamakawa *et al*. (2012), which produced 46.5 $g.l^{-1}$ ethanol from 200 $g.l^{-1}$ of raw corn starch. Furthermore, these results showed considerable improvements when compared to amylolytic CBP systems listed in a recent review by Salehi Jouzani and Taherzadeh, (2015). The product yield displayed by the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain on raw corn starch (Table 3) represented the highest reported for amylolytic *S. cerevisiae* Y294 strains in fermentations with high substrate loading and low inoculums.

Overall, *S. cerevisiae* recombinant strains with higher levels of glucoamylase, i.e. those expressing the *temG_Opt* glucoamylase, hydrolysed starch better than the *S. cerevisiae* Y294 strains with the *temG_Nat* glucoamylase. However, *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] displayed a significantly higher product yield (~1.6-2.0 fold) compared any of the other recombinant *S. cerevisiae* Y294 strains expressing the *temG_Opt* glucoamylase (Table 3). This suggested that there was a unique synergistic effect between the *T. emersonii* TemG_*Opt* and TemA_*Nat* enzymes that outperformed the other TemG_Opt-α-amylase combinations.

A synergistic effect was also observed for the *A. tubingensis* enzyme combination. At 192 hours, the product yield displayed by the *S. cerevisiae* Y294[GlaA-AmyA] strain (54%) was 9% higher than the product yield displayed by the *S. cerevisiae* Y294[TemG_Opt-AmyA] strain (49%) (Table 3), even though TemG_Opt was superior to GlaA in terms of activity (Figure 3). This highlights the importance of comparing different enzyme combinations in the chosen expression host. Even though extracellular amylase activities differed (Figures 2 and 3), enzymes originating from the same host may have a superior synergistic hydrolytic effect as a result of their modes of action and affinity for raw starch. Presečki *et al*. (2013) developed a mathematical model to explain the synergism between a glucoamylase and two α-amylases (in different combinations) and showed that the type and combinations of amylases affected enzyme synergy. Furthermore, whether an α-amylases is classified as “liquefying” or “saccharifying” may also attribute to the synergist relationship (Liakopoulou-Kyriakides *et al*., 2001).

The AmyA α-amylase displayed a greater extracellular activity on soluble starch, compared to the AteA_Nat enzyme (Figure 2a). However, during fermentation studies the AteA_Nat α-amylase combinations facilitated a faster rate of raw starch conversion compared to the enzyme combinations with AmyA (Figure 4). AteA_Nat also contributed to higher ethanol productivity levels (compared to AmyA) when combined with the TemG_Opt and TemG_Nat glucoamylases, respectively (Figure 4 and Table 3). This suggested that AteA_Nat may have performed better on raw starch compared to the AmyA enzyme, or it had a superior synergistic effect with the TemG glucoamylase.

Dissimilarly, the extracellular activity produced by the *S. cerevisiae* Y294[ApuA_Nat] strain (expressing the native α-amylase from *A. pullulans*) was 2.7-fold higher than that of the *S. cerevisiae* Y294[AmyA] benchmark strain (Figures 2a and b), but overall the product yield by the amylolytic *S. cerevisiae* Y294[TemG_Opt-ApuA_Nat] strain was 13% lower than the *S. cerevisiae* Y294[TemG_Opt-AmyA] strain (Table 3). Therefore, AmyA may either have had

improved raw starch converting ability, or a better synergistic relationship with TemG_Opt, compared to ApuA_Nat (Figure 4a). Chi *et al*. (2009) demonstrated that the glucoamylase from *A. pullulans* hydrolysed potato starch granules (type-B crystallinity) better than raw corn starch granules (type-A crystallinity), although type-B starch structures are usually more resistant to enzyme hydrolysis (Man *et al*., 2013). Corn starch has a higher amylose content and smaller granule diameter compared to potato starch (Hii *et al*., 2012) and the combination of these properties are known to influence the rate and extent of starch hydrolysis (Naguleswaran *et al*., 2013). Results from this study (Figures 2 and 4) highlighted a prime example where starch structure affected the action of different amylolytic enzymes.

Although *S. cerevisiae* is known for its ethanol tolerance, the Y294 strains were inhibited by fermentation conditions at a cultivation temperature of 30°C and thus ethanol concentrations did not exceed 63 $g.l^{-1}$ (Figures 5 and 6). The poor fermentative performance by the *S. cerevisiae* Y294 laboratory strain was not as a result of inadequate recombinant protein secretion or low enzymatic activity, since glucose concentrations increased rapidly throughout the fermentation with the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain (Figure 4b).

Raw starch fermentation by recombinant *S. cerevisiae* strains is often disadvantaged by long cultivations times required for sufficient enzyme secretion. However, it was clear from the fermentation results for the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain (Figure 5) that volumetric productivity and starch conversion rates were high. Furthermore, a cultivation temperature of 26°C relieved physiological stress on the yeast cells, allowing for improved glucose conversion. The product yields by the *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain were similar (4% difference after 192 hours) for the 100 ml serum bottles and bioreactor fermentations, thus suggesting that the lower temperature was the main factor to favour glucose fermentation and that the extracellular enzyme activity was not significantly affected by a lower fermentation temperature. Therefore, decreasing the fermentation temperature confirmed that it was possible to increase the conversion of glucose to ethanol and improve the theoretical ethanol yield.

Schmidt *et al*. (2006) provided several definitions for ethanol tolerance, one of which was the effect of ethanol concentrations on the ability of a cell to metabolise sugar. Biochemical and physiological responses occur when yeast are exposed to accumulating ethanol concentrations (Schmidt *et al*., 2006) and as a result *S. cerevisiae* Y294 strains were likely to experience compromised membrane structure and protein function. The presence of ethanol changes the composition of the phospholipid bilayer making it permeable to small molecules. Since many cellular functions rely on membrane integrity, high ethanol concentrations can

have a number of adverse effects on the yeast cell. In this study, the negative effects of ethanol accumulation could be avoided by lowering the fermentation temperature to 26°C.

Few genes have been cloned and sequenced from thermophilic fungi. Glucoamylases from *T. lanuginosus* (Thorsen *et al*., 2006), *T. emersonii* (Nielsen *et al*., 2002), *H. grisea* var. thermoidea (Allison *et al*., 1992) and *C. thermophilum* (Chen *et al*., 2007) have been expressed in fungal host strains. However, *T. emersonii* amylases have not been expressed in *S. cerevisiae*. *T. emersonii* is an acidicothermophilic fungus that is industrially important and well recognised for its production of glycoside hydrolases (GHs) with special enzymatic properties, especially cellulases (Amore and Faraco, 2012; Wang *et al*., 2014). However, few studies have investigated its starch hydrolysing enzymes.

**Conclusion**

Currently, industry lacks the implementation of an amylolytic CBP yeast that expresses both an α-amylase and glucoamylase (den Haan *et al*., 2013). This study focused on the selection of highly active amylases with the ability to convert raw starch to glucose. This led to the identification and evaluation of novel amylase combinations for the hydrolysis of raw starch. The recombinant *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain was superior in its ability to convert 85% of the available carbon in 200 $g.l^{-1}$ raw corn starch within 192 hours. Thus, this unique TemG_Opt-TemA_Nat enzyme combination represents a promising candidate for the industrial conversion of uncooked starch.

## Example 2: Construction of amylolytic CBP *S. cerevisiae* Ethanol Red™ and M2n strains

**Materials and methods**

*Media and cultivation conditions*

All chemicals were of analytical grade and were obtained from Merck (Darmstadt, Germany), unless otherwise stated. *Escherichia coli* DH5α (Takara Bio Inc.) was used for vector propagation. The *E. coli* transformants were selected for on Luria Bertani agar (Sigma-Aldrich, Germany), containing 100 $\mu g.ml^{-1}$ ampicillin and cultivated at 37°C in Terrific Broth (12 $g.l^{-1}$ tryptone, 24 $g.l^{-1}$ yeast extract, 4 $ml.l^{-1}$ glycerol, 0.1 M potassium phosphate buffer) containing 100 $\mu g.ml^{-1}$ ampicillin for selective pressure (Sambrook *et al*., 1989).

The *S. cerevisiae* parental strains were maintained on YPD agar plates (10 g.l$^{-1}$ yeast extract, 20 g.l$^{-1}$ peptone, 20 g.l$^{-1}$ glucose and 20 g.l$^{-1}$ agar). The *S. cerevisiae* Y294 transformants were selected for and maintained on SC$^{-URA}$ agar plates (6.7 g.l$^{-1}$ yeast nitrogen base without amino acids (BD-Diagnostic Systems, Sparks, MD), 20 g.l$^{-1}$ glucose and 1.5 g.l$^{-1}$ yeast synthetic drop-out medium supplements (Sigma-Aldrich, Germany) and 20 g.l$^{-1}$ agar). *S. cerevisiae* strains were aerobically cultivated on a rotary shaker (200 rpm) at 30°C, in 125 ml Erlenmeyer flasks containing 20 ml double strength SC$^{-URA}$ medium (2×SC$^{-URA}$ containing 13.4 g.l$^{-1}$ yeast nitrogen base without amino acids (BD-Diagnostic Systems, Sparks, MD), 20 g.l$^{-1}$ glucose and 3 g.l$^{-1}$ yeast synthetic drop-out medium supplements). Fermentation media for *S. cerevisiae* Y294 strains comprised of 2×SC$^{-URA}$ containing 5 g.l$^{-1}$ glucose and 200 g.l$^{-1}$ raw corn starch, whereas the medium for *S. cerevisiae* Ethanol Red™ and M2n strains was YP containing 5 g.l$^{-1}$ glucose and 200 g.l$^{-1}$ raw corn starch. Ampicillin (100 µg.ml$^{-1}$) and streptomycin (50 µg.ml$^{-1}$) were added to inhibit bacterial contamination. All cultures were inoculated to a concentration of 1×10$^{6}$ cells.ml$^{-1}$, unless otherwise stated.

SC media (yeast synthetic drop-out medium omitted) containing 2% starch was used to maintain industrial transformants. The *S. cerevisiae* Ethanol Red™ and M2n transformants were selected for on SC-Ac plats (SC plates with $(NH_4)_2SO_4$ replaced by 0.6 g.l$^{-1}$ acetamide and 6.6 g.l$^{-1}$ $K_2SO_4$) and transferred to SC-Acr plates (SC-Ac with 0.71 g.l$^{-1}$ acrylamide replacing the acetamide). For plate assays, 2% soluble starch was added to SC-Ac and SC-Acr plates. SC-Fac plates (SC media containing 2.3 g.l$^{-1}$ fluoroacetamide) was used to remove the yBBH1-amdSYM vector from the transformants. The pH in all the media was adjusted to 6.0 with NAOH.

*Strains and plasmids*

The genotypes of the bacterial and yeast strains, as well as the plasmids used in this study, are summarised in Table 4.

Table 4. Strains and plasmids used in this study

| **Strains and plasmids** | **Genotype** | **Reference/ Source** |
|---|---|---|
| *E. coli DH5α* | *supE44 ΔlacU169 (ϕ80lacZ*ΔM15*) hsdR17 recA1 endA1 gyrA96 thi-1 relA1* | Sambrook *et al.* (1989) |
| ***S. cerevisiae* strains** | | |
| Y294 | *α leu2*-3,112 *ura3*-52 *his3 trp1-289* | ATCC 201160 |
| Y294[amdSYM] | *URA3 TEF$_P$-amdS-TEF$_T$* | This study |

| | | |
|---|---|---|
| Y294[TemG_Opt-TemA_Nat] | *URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Opt-ENO1*$_T$ | This study |
| Ethanol Red™[1] | *MATa/α* prototroph | Fermentis, Lesaffre, France |
| M2n | *MATa/α* prototroph | Favaro *et al*. (2015) |
| Ethanol Red™ T1[2] | δ-integration of *ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| Ethanol Red™ T12[2] | δ-integration of *ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| M2n T1[2] | δ-integration of *ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| M2n T2[2] | δ-integration of *ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| **Plasmids** | | |
| yBBH1 | *bla URA3 ENO1*$_P$*-ENO1*$_T$ | Njokweni *et al*. (2012) |
| yBBH1-TemA_Nat[3] | *bla URA3 ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt[4] | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$ | This study |
| yBBH1-TemG_Opt-TemA_Nat | *bla URA3 ENO1*$_P$*-temG_Opt-ENO1*$_T$; *ENO1*$_P$*-temA_Nat-ENO1*$_T$ | This study |
| pUG-amdSYM[5] | *bla TEF*$_P$*-amdS-TEF*$_T$ | Solis-Escalante *et al*. (2013) |
| yBBH1-amdSYM | *bla URA3 TEF*$_P$*-amdS-TEF*$_T$ | This study |

[1]Ethanol Red™ Version 1, referred to as Ethanol Red™
[2]Amylolytic transformants (T) contain integrated copies of *ENO1*$_P$*-temA_Nat-ENO1*$_T$ and *ENO1*$_P$*-temG_Opt-ENO1*$_T$ gene cassettes, the number indicates the transformant number during the screening process
[3]Accession no. XM_013469492 for the native *T. emersonii α-amylase* (*temG_Nat*)
[4]Accession no. AJ304803 for the native *T. emersonii glucoamylase* (*temG_Opt* encodes for the codon-optimised gene)
[5]Assession no. P30669 for pUG-amdSYM plasmid

*DNA manipulations*

Standard protocols were followed for all DNA manipulations and *E. coli* transformations (Sambrook *et al*., 1989). The enzymes used for restriction digests and ligations were purchased from Inqaba Biotec and used as recommended by the supplier. Digested DNA was eluted from 0.8% agarose gels using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, USA). The *temA_Nat* and *temG_Opt* gene cassettes (*ENO1* promoter and terminator) (Figure 1b) were amplified through PCR using Delta-ENO1 primers (Table 5), together with the yBBH1[TemA_Nat] and yBBH1[TemG_Opt] vectors (see Example 1), respectively, as template.

*Plasmid construction*

The *TEF*$_P$*-amdSYM-TEF*$_T$ gene cassette was amplified from pUG-amdSYM through PCR using amdSYMCas primers (Table 5) and cloned onto yBBH1 using yeast-mediated ligation

(YML) yielding plasmid yBBH1-amdSYM (Figure 6a). The *Ashbya gossypii TEF* promoter regulated the expression of the acetamidase-encoding gene (*amdS*) for the selection of transformants on SC-Ac plates. The yBBH1-amdSYM plasmid was retrieved from the *S. cerevisiae* Y294[amdSym] strain and transformed into *E. coli* DH5α in order to obtain a high concentration of plasmid DNA. Plasmid DNA was isolated using the High Pure Plasmid Isolation kit (Roche, Germany) and sequence verification was performed by the dideoxy chain termination method, with an ABI PRISM™ 3100 Genetic Analyser (CAF, Stellenbosch University).

Table 5. PCR primers designed and used in this study with the relevant restriction sites underlined (*Eco*RI = gaattc; *Xho*I =ctcgag, *Bam*HI = ggatcc, *Bgl*II = agatct)

| Primer name | Sequence (5'-3') | SEQ ID NO: |
|---|---|---|
| amdSYMCas:L | ccgcgcgttggccgattcattaatccaggatccacatggaggcccagaataccctccttgac | 37 |
| amdSYMCas:R | gggcctcttcgctattacgccagagcttagatctcagtatagcgaccagcattcacatacttaa | 38 |
| Delta-ENO1p:L | tggaataaaaatccactatcgtctatcaactaatagttatattatcaatatattatcatatacggtgtta agatgatgacataagttatgagaagctgtcggatcccaattaatgtgagttacctcac | 39 |
| Delta-ENO1t:R | tgagatatatgtgggtaattagataattgttgggattccattgttgataaaggctataatattaggtata cagaatatactagaagttctcctcgaggatagatctcctatgcggtgtgaaataccgc | 40 |
| TemG_Opt:L | ttatcaacacacaaacactaaatcaaagaattcatggcctccttagtcgcaggtgcctta | 41 |
| TemG_Opt:R | gactagaaggcttaatcaaaagctctcgagtcattgccaagagtcgtccaagattgcggt | 42 |
| TemA_Nat:L | tgcttatcaacacacaaacactaaatcaaagaattcatgacgcctttcgtcctcacggcc | 43 |
| TemA_Nat:R | ggactagaaggcttaatcaaaagctctcgagctatctccatgtgtcgacaatcgtctccg | 44 |

*Yeast transformations*

Electro-competent *S. cerevisiae* Y294, Ethanol Red™ and M2n cells were prepared according to Cho *et al*. (1999) and transformed by means of electroporation using a BioRad system (GenePluserXcell TM, Bio-Rad, Hercules, California). For the transformation of industrial strains, amylases (*temA_Nat* and *temG_Opt ENO1* linear DNA cassettes) and the yBBH1-amdSYM vector containing the selection marker (Figure 6) were simultaneously transformed into the genomes of the yeasts. After electroporation, 1 ml of YPDS was immediately added to the cuvettes. Cells were incubated at 30°C for 3 hours. Transformants were selected for by plating the transformation mix on to SC-Ac plates containing 2% starch (adapted from Solis-Escalante *et al*., 2013) and incubated at 30°C for 4 days. The integration of the linear

expression cassette DNA into the yeast genome was confirmed by PCR using gene specific primers (Table 5).

*Marker recycling*

Plasmid curing was performed on the industrial recombinant strains as described by Solis-Escalante *et al*. (2013). The removal of the yBBH1-amdSYM containing the acetamide marker was achieved by growing cells overnight in 5 ml liquid YPD and transferring 20 µl to a 125 ml Erlenmeyer flask containing 10 mL SC-Fac. Marker-free single colonies were obtained by plating 100 µl of culture on SC-Fac solid media containing 2% starch and confirmed by colony PCR. The genomic DNA of the amylolytic strains was isolated using the ZR fungal/bacterial DNA miniprep kit (Zymo Research, USA) and it was then used as a template for real-time PCR.

*Quantitative PCR*

Oligo primers for real-time PCR were designed using IDT's PrimerQuest Tool (http://eu.idtdna.com/PrimerQuest/Home/Index). Special attention was given to primer length (18 – 22 bp), annealing temperature (58 – 62°C), base composition, 3′-end stability and amplicon size (75 – 100 bp). All primers were synthesised by Inqaba Biotech (South Africa) with reverse phase cartridge purification and are listed in Table 6. The performance of all primers was experimentally confirmed by conventional PCR to ensure that there was no formation of primer dimers and confirm the amplification of a single region with the correct amplicon length.

Table 6. List of candidate reference genes and target genes including details of primers and amplicons for each gene

| Gene name | Amplicon length (bp) | Primers (5'-3') | SEQ ID NO: |
|---|---|---|---|
| *URA3* | 92 | L: cgtggatgatgtggtctctac | 45 |
| | | R: gttcaccctctaccttagcatc | 46 |
| *temA_Nat* | 100 | L: gcgatgtcactgagaggatcta | 47 |
| | | R: gaaatccagatggccgtgaa | 48 |
| *temG_Opt* | 95 | L: tacaggtggtttgggtgaac | 49 |
| | | R: ctctcaatgctggaccatctc | 50 |

Real-time PCR was carried out on a StepOne real time Polymerase Chain Reaction (PCR) instrument (Applied Biosystems) using white-walled PCR plates (96 wells). A × 2 KAPA HRM Fast Master Mix (containing a fast proof-reading polymerase, dNTPs, stabilisers and

EvaGreen® dye) was used according to the manufacturer's instructions (KAPA Biosystems). Reactions were prepared in a total volume of 20 µl containing, 2.5 mM $MgCl_2$, 0.2 µM of each primer and 1 - 10 ng DNA. The cycle conditions were set as follows: initial template denaturation at 95°C for 30 seconds, followed by 45 cycles of denaturation at 95°C for 5 seconds and combined primer annealing/elongation at 60 °C for 20 seconds and a final denaturation at 95°C for 1 minute to ensure all amplicons were fully melted. The yBBH1-TemG_Opt-TemA_Nat plasmid DNA was used to set up the standard curves (starting with $1\times10^7$ copies and making a 1:10 serial dilution) using primer pairs listed in Table 6. Genomic DNA concentrations were standardised to 10 ng for all samples. The PCR efficiency for each of the primer sets was calculated using StepOne software (Applied Biosystems). The number of copies of the *temG_Opt* and *temA_Nat* genes was calculated using the standard curve method using the *URA3* gene as reference gene.

*Fermentations*

The *S. cerevisiae* Y294 precultures were cultured in 100 ml $2\times SC^{-URA}$ medium in 500 ml Erlenmeyer flasks, and the *S. cerevisiae* Ethanol Red™ and M2n precultures were cultivated similarly in YPD medium. All precultures were incubated at 30°C with agitation at 200 rpms until stationary phase. *S. cerevisiae* Y294 fermentations were performed in $2\times SC^{-URA}$ media, whereas *S. cerevisiae* Ethanol Red™ and M2n fermentations were performed in YP starch media and inoculated with a 10% ($v.v^{-1}$) inoculum from the stationary preculture. Agitation and incubation were performed on a magnetic multi-stirrer platform (Velp Scientifica, Italy) at 30°C and 37°C, with daily sampling through a syringe needle pierced through the rubber stopper.

Exogenous enzymes used in the fermentation processes were STARGEN™ 002 GSHE (now referred to as STARGEN™), obtained from Dupont Industrial Biosciences (Palo Alto, California), with an activity minimum of 570 $GAU.gm^{-1}$ (http://www.genencor.com) and used according to the manufacturer's instructions. STARGEN™ contained *Aspergillus kawachii* α-amylase expressed in *Trichoderma reesei* and a glucoamylase from *T. reesei* that works synergistically to hydrolyse granular starch to glucose (Huang *et al*., 2015).

*HPLC and analytical methods*

Ethanol, glucose, maltose, glycerol and acetic acid concentrations were quantified with high performance liquid chromatography (HPLC) using a Surveyor Plus liquid chromatograph (Thermo Scientific) consisting of a liquid chromatography pump, autosampler and refractive index (RI) detector. The compounds were separated on a Rezex RHM Monosaccharide 7.8 ×

300 mm column (00H0132-K0, Phenomenex) at 60°C with 5 mM $H_2SO_4$ as mobile phase at a flow rate of 0.6 ml.min$^{-1}$. The theoretical $CO_2$ yields were calculated according to Favaro *et al*. (2015). The product yield (percentage starch converted to products) was calculated from ethanol, glucose, maltose, glycerol, acetic acid and $CO_2$ concentrations.

*Statistical analysis*

Data was analysed using the Student's t-test.

## Results

The *T. emersonii temA_Nat* and *temG_Opt* genes encode for valuable amylase enzymes for use in the production of biofuel and are produced and secreted during cultivation on raw corn starch. The linear *ENO1*$_P$-*temA_Nat*-*ENO1*$_T$ and *ENO1*$_P$-*temG_Opt*-*ENO1*$_T$ DNA gene cassettes (Figure 6b), flanked by the δ sequence, were amplified and integrated into the δ-integration sites in the *S. cerevisiae* Ethanol Red™ and M2n industrial strains' genomes, in order to generate multi-copy integrants (Kim *et al*., 2011). The *amdS* gene was present on an episomal vector (Figure 6a) to enable plasmid curing for easy recycling of the marker.

*Industrial strain screening*

*S. cerevisiae* transformants were screened on SC plates containing 2% corn starch and those producing zones of hydrolysis were selected for further testing. PCR was used to confirm the integration of both *ENO1*$_P$-*temA_Nat*-*ENO1*$_T$ and *ENO1*$_P$-*temG_Opt*-*ENO1*$_T$ gene cassettes. The four strains showing the highest extracellular amylase activity were then evaluated under fermentative conditions (Figures 7a and 7b). Significant differences in the amount of product yield were noted during the early stages of fermentation (Figures 7b). However, after 192 hours product yields started to plateau, representing an approximate 85% conversion of starch. The *S. cerevisiae* Ethanol Red™ T12 and M2n T1 strains hydrolysed starch and fermented the sugars quicker than the *S. cerevisiae* Ethanol Red™ T1 and M2n T2 strains (Figure 7b and Table 7). They were therefore selected for further evaluation under different fermentation conditions.

Plasmid curing of the strains was performed by plating cultures onto SC-FAc plates containing 2% soluble corn starch. Quantitative PCR assays were performed using the genomic DNA from the cured amylolytic *S. cerevisiae* transformants, in order to determine the number of integrated copies of both *temA_Nat* and *temG_Opt* genes, respectively (Figure 7d). The

*S. cerevisiae* Ethanol Red™ T1, M2n T1 and M2n T2 strains contained single copies of *temA_Nat* and *temG_Opt* gene cassettes, whereas the *S. cerevisiae* Ethanol Red™ T12 contained 1 copy of *temA_Nat* and 2 copies of *temG_Opt*.

Table 7. Product formation by *S. cerevisiae* strains after 144 hours of fermentation at 30°C

| ***S. cerevisiae*** | **Ethanol Red™ T1** | **Ethanol Red™ T12** | **M2n T1** | **M2n T2** |
|---|---|---|---|---|
| **Substrate (g.l$^{-1}$)** | | | | |
| Raw starch weighed | 200 | 200 | 200 | 200 |
| Glucose weighed | 5 | 5 | 5 | 5 |
| Raw starch (dry weight) | 185 | 185 | 185 | 185 |
| Glucose equivalent | 208.5 | 208.5 | 208.5 | 208.5 |
| **Products (g.l$^{-1}$)** | | | | |
| Glucose | 0.82 | 0.67 | 0.60 | 0.72 |
| Glycerol | 2.39 | 3.40 | 1.92 | 2.29 |
| Acetic acid | 0.49 | 0.46 | 0.76 | 0.35 |
| Ethanol | 57.76 | 74.19 | 72.19 | 64.68 |
| Maltose | 0.99 | 1.09 | 1.01 | 1.08 |
| $CO_2$[1] | 55.25 | 70.94 | 69.05 | 61.87 |
| **Total** | 117.68 | 150.76 | 145.53 | 131.00 |
| Product yield (%) | 56.44 | 72.31 | 69.80 | 62.83 |
| Ethanol yield[2] (% of theoretical yield) | 55.41 | 71.17 | 69.25 | 62.05 |
| Ethanol rate of productivity[3] | 0.40 | 0.52 | 0.50 | 0.45 |

[1]$CO_2$ yields were deduced from the ethanol produced.
[2]Ethanol yield (% of the theoretical yield) was calculated as the amount of ethanol produced per gram of consumed sugar
[3]Ethanol rate of productivity was calculated based ethanol titres produced per hour (g.l$^{-1}$.h$^{-1}$)

The fermentation vigour of the amylolytic *S. cerevisiae* Ethanol Red™ T12 strain at 30°C and 37°C was compared to the laboratory *S. cerevisiae* Y294[TemG_Opt-TemA_Nat] strain at 30°C (Figure 8). The *S. cerevisiae* Ethanol Red™ T12 strain was able to ferment all the available glucose (Figure 8b) at a fermentation temperature of 30°C and produced significantly less glycerol during the fermentation (Figure 8d). This indicated a more efficient carbon conversion for ethanol (Bideaux *et al*., 2006). However, at a temperature of 37°C, ethanol levels did not increase significantly after 144 hours (Figure 8a) and high level of residual glucose were present (> 40 g.l$^{-1}$ after 264 hours). Maltose concentrations were similar at both fermentation temperatures (Figure 8c).

The evaluation of different media conditions (Figure 9) was subsequently undertaken in order to determine whether buffered fermentation media (pH 5), the type of media (YP versus SC) or the addition of extra nitrogen (in the form of $(NH_4)_2SO_4$) could increase the efficiency of

glucose fermentation by the *S. cerevisiae* Ethanol Red™ T12 strain at a fermentation temperature of 37°C.

YP starch media (unbuffered) had a pH lower than 5 and this was more favourable for ethanol production, compared to the buffered YP broth (pH 5) (Figure 9a). The addition of extra ammonium sulphate (10 $g.l^{-1}$) to the SC buffered fermentation broth did not increase ethanol production or product yield (Figures 9a and 9d), indicating sufficient nitrogen levels in the fermentation broth. Increased residual glucose concentrations were observed when YP media was used (Figure 9b), while higher glycerol concentrations were noted when the fermentation was performed in SC media (Figure 9c). The latter contributed to increased product yields, especially during the first 120 hours of fermentation. Overall, the results in Figure 9 showed that the media composition (SC vs YP and the pH) affected the ethanol and glycerol production. However, changes in the type of media only affected the product yield during the first 120 hours of fermentation. After 192 hours, the differences in product yields were less apparent (between 92 – 100%).

*Fermentations with STARGEN™*

The recommended STARGEN™ dosage was calculated as 1.42 $\mu l.g^{-1}$ starch, according to the manufacturer's specifications. The amylolytic S. *cerevisiae* Ethanol Red™ T12 and M2n T1 strains were compared to a simulated conventional SSF process (parental *S. cerevisiae* Ethanol Red™/M2n strains + STARGEN™) at 200 $g.l^{-1}$ corn starch. Three different enzyme dosages were evaluated based on the percentage of the recommended enzyme loading: 2.8 µl (10%), 5.6 µl (20%) and 14 µl (50%) and compared to the SSF, which had 28 µl STARGEN™ per 100 ml (representing 100% of the recommended dosage). The addition of exogenous enzymes significantly increased ethanol concentrations and enhanced the rate of ethanol productivity (ethanol $g.l^{-1}.h^{-1}$) during the first 72 hours of fermentation (Figures 10 and 11).

At a fermentation temperature of 30°C the ethanol profiles for the *S. cerevisiae* Ethanol Red™ and M2n parental strains were similar for the respective condition (Figures 10a and 11a). By 48 hours, the *S. cerevisiae* Ethanol Red™ T1 strain supplemented with 2.8 ul STARGEN™ produced the same amount of ethanol and produced a similar product yield to that of the control SSF process with untransformed *S. cerevisiae* Ethanol Red™ supplemented with 28 µl STARGEN™ (Table 8). A similar trend was observed for the *S. cerevisiae* M2n T1 strain supplemented with 2.8 µl STARGEN™ compared to the *S. cerevisiae* M2n parental strain (Figures 11a and c).

After 96 hours, ethanol produced by the *S. cerevisiae* Ethanol Red™ T12 strain supplemented with 2.8 µl STARGEN™ (90.41 $g.l^{-1}$) was similar to the amount of ethanol produced by the *S. cerevisiae* Ethanol Red™ T12 strain supplemented with 5.6 µl STARGEN™ (92.04 $g.l^{-1}$) (Figure 10a). The product yield displayed by these two strains was also similar (between 88 – 90%), at 96 hours (Figure 10c). This demonstrated that 2.8 µl (10% of the recommended dosage) was sufficient for increasing the carbon conversion to end products (% product yield) and suggested that dosages higher than this would be unnecessary for industrial scale-up fermentations.. Furthermore, an enzyme dosage representing 50% (14 µl of STARGEN™) did not improve results (Figure 10a) and was therefore unnecessary.

Similar results and trends were observed for the *S. cerevisiae* M2n strains at a fermentation temperature of 30°C, compared to the *S. cerevisiae* Ethanol Red™ equivalent strains (Figures 10 and 11). However, the final ethanol concentration for the *S. cerevisiae* M2n T1 transformant was higher > 10 $g.l^{-1}$ after 192 hours ($p$ = 0.0392). At 30°C, the low residual levels of glucose and maltose in the fermentation broth (Table 8) indicated a rapid sugar uptake by all the amylolytic strains.

Table 8. Product formation by *S. cerevisiae* Ethanol Red™ and M2n strains after 192 hours of fermentation at 30°C in YP media, supplemented with different STARGEN™ dosages

| ***S. cerevisiae* strains** | **Ethanol Red™** | **M2n** | **Ethanol Red™ T12** | **M2n T1** | **Ethanol Red™ T12** |
|---|---|---|---|---|---|
| STARGEN™ added (µl) | 28 | 28 | 2.8 | 2.8 | 5.6 |
| **Substrate ($g.l^{-1}$)** | | | | | |
| Raw starch weighed | 200 | 200 | 200 | 200 | 200 |
| Glucose weighed | 5 | 5 | 5 | 5 | 5 |
| Raw starch (dry weight) | 185 | 185 | 185 | 185 | 185 |
| Glucose equivalent | 208.5 | 208.5 | 208.5 | 208.5 | 208.5 |
| **Products ($g.l^{-1}$)** | | | | | |
| Glucose | 0.02 | 0.31 | 0.02 | 3.28 | 0.12 |
| Glycerol | 4.07 | 4.30 | 4.76 | 4.59 | 5.22 |
| Acetic acid | 0.00 | 0 | 0.90 | 0.31 | 0.96 |
| Ethanol | 97.23 | 98.49 | 98.37 | 99.08 | 100.32 |
| Maltose | 0.79 | 0.71 | 0.31 | 0.37 | 0.26 |
| $CO_2$[1] | 93.00 | 94.21 | 94.09 | 94.77 | 95.96 |
| **Total** | 195.11 | 198.02 | 198.44 | 202.40 | 202.85 |
| Product yield (%) | 93.58 | 94.98 | 95.17 | 97.07 | 97.29 |

| | | | | | |
|---|---|---|---|---|---|
| Ethanol yield (% of theoretical yield)[2] | 93.26 | 94.48 | 94.36 | 95.04 | 96.23 |
| Ethanol rate of productivity[3] | 0.51 | 0.51 | 0.51 | 0.52 | 0.52 |

[1] $CO_2$ yields were deduced from the ethanol produced
[2]Ethanol yield (% of the theoretical yield) was calculated as the amount of ethanol produced per gram of consumed sugar
[3]Ethanol rate of productivity was calculated based ethanol titres produced per hour ($g.l^{-1}.h^{-1}$)

The extent of product yield was similar at the two fermentation temperatures (Figures 10 and 11). However, at 37°C both the amylolytic *S. cerevisiae* Ethanol Red™ T12 and M2n T1 strains had lower ethanol productivity (Figures 10d and 10d). At 37°C, the ethanol concentration plateaued after 48 hours for all *S. cerevisiae* M2n strains supplemented with STARGEN™ (Figure 11b). The Ethanol Red™ T12 transformant had a higher ethanol tolerance at 37°C and was able to ferment for longer compared to the *S. cerevisiae* M2n T1 transformant, producing a 2.3-fold increase in ethanol concentration at 192 hours (Figures 10a and 11a). Residual glucose levels were > 40 $g.l^{-1}$ at 37°C (data not shown), which represented a large amount of unfermented glucose. Overall, results showed that temperature played a major role on the fermentation vigour of industrial *S. cerevisiae* Ethanol Red™ T12 and M2n T1 strains. The addition of STARGEN™ was able to reduce the fermentation time and increase the theoretical ethanol yield. At 30°C the recombinant *S. cerevisiae* M2n T1 strain produced more ethanol, but was severely affected at a higher fermentation temperature (Figures 11a and b).

**Discussion**

Starch-rich biomass is currently the main substrate for bioethanol production and can be efficiently hydrolysed by α-amylases and glucoamylases (Viktor *et al*., 2013). The *S. cerevisiae* Ethanol Red™ strain is a widely used industrial yeast, predominantly applied in first-generation bioethanol production from corn and wheat and was the primary expression host used in this study. It is characterised by excellent fermentation capacity and yield, high robustness and stress tolerance (Demeke *et al*., 2013). The *S. cerevisiae* M2n strain is a South African distillery yeast and was also used in this study for comparative purposes. The construction of a CBP yeast that can simultaneously express heterologous amylases and produce ethanol could yield more cost-effective ethanol production from starchy feedstocks.

***Gene integration***

Industrial strains can be used as the platform for heterologous amylase expression intended for first generation bioethanol production (Favaro *et al*., 2013). After the initial screening process, four recombinant strains expressing the *temG_Opt* and *temA_Nat* gene cassettes (the *S. cerevisiae* Ethanol Red™ T1/T12 and *S. cerevisiae* M2n T1/T2 strains) were selected for further evaluation (Figure 7). The *S. cerevisiae* M2n T1 strain performed slightly better than the *S. cerevisiae* Ethanol Red™ T12 strain at 30°C and achieved a maximum ethanol titre of 99.40 $g.l^{-1}$, which was 15% higher than the *S. cerevisiae* Ethanol Red™ T12 strain, at 192 hours (Figures 10a and 11a). However, at 37°C it was clear that the *S. cerevisiae* Ethanol Red™ T12 transformant had a greater fermentation vigour and was more ethanol and temperature tolerant (Figures 10b and 11b), compared to the *S. cerevisiae* M2n strain.

Results from this study showed significant improvements when compared to the industrial *S. cerevisiae* M2n[TLG1 SFA1] and MEL2[TLG1 SFA1] amylolytic strains (Favaro *et al*., 2015) that produced 64 $g.l^{-1}$ ethanol from 200 $g.l^{-1}$ raw corn starch, corresponding to 55% of the maximum theoretical ethanol yield, as well as the *S. cerevisiae* Mnuα1[AmyA-GlaA] strain (Viktor *et al*., 2013) that produced 65.83 $g.l^{-1}$ ethanol (after 10 days) representing 57% of the maximum theoretical ethanol yield. Theoretical ethanol yields obtained from the recombinant industrial strains in this study were > 90% and thus represented a significant improvement on previously constructed amylolytic strains.

Ethanol concentrations were also higher than those reported for the amylolytic yeast strain, which produced 46.5 $g.l^{-1}$ of ethanol from 200 $g.l^{-1}$ of raw corn starch after 120 hours of fermentation (Yamakawa *et al*., 2012). The amylolytic yeast strains in this study were superior in their ethanol production, producing > 50 $g.l^{-1}$ and > 60 $g.l^{-1}$ ethanol for the *S. cerevisiae* Ethanol Red™ T12 and M2n T1 strains, respectively, after 120 hours (Figures 10a and 11a). Furthermore, since the amylases were secreted into the fermentation broth they had increased physical contact with the starch granules, compared to recombinant yeast that displayed amylases on the cell's surface (Yamakawa *et al*., 2012). This eliminated potential bottlenecks and facilitated improved starch hydrolysis because the enzymes were able to penetrate into the granules and create pores more quickly.

***STARGEN™ addition***

During fermentation with the amylolytic *S. cerevisiae* Ethanol Red™ and M2n strains, there was an initial lag phase in product yield, up until 48 hours (Figures 10c and 11c). This was

expected, since the strains first had to adjust to the fermentation conditions and produce amylases *de novo*. On the other hand, during the SSF process with STARGEN™ (Figures 10a and 11a), the enzymes were in abundance at the start of the fermentation and rapidly produced glucose upon addition. Therefore, although *S. cerevisiae* Ethanol Red™ T12 and M2n T1 strains were able to achieve high product yields (Figure 7b), supplementation with STARGEN™ (Figures 10 and 11) increased ethanol productivity at the start of the fermentation.

In the industrial cold hydrolysis set-up for bioethanol production, commercial amylase enzymes are only added at the beginning of the process and therefore their overall efficiency will decrease over time. However, the amylolytic CBP yeasts of the present invention were able to continually replenish the recombinant enzymes in the fermentation broth and thus had increased overall product yields when the fermentation was supplemented with STARGEN™ (Figures 10c,10d, 11c and 11d). The cost of commercial enzyme addition was estimated at 4.8 US cents per gallon, representing 8.3% of the total possessing costs in ethanol production from corn (Wong *et al*., 2010). The recombinant amylolytic *S. cerevisiae* Ethanol Red ™T12 and M2n T1 strains described herein thus represent a novel alternative for lowering the enzyme dosage required for raw starch hydrolysis, as well as being able to provide continuous amylolytic activity for a continuous cold fermentations process. Furthermore, the use of amylolytic yeasts of the present invention allow for a simplified fermentation design, since pretreatment steps and costs can be bypassed (Salehi Jouzani and Taherzadeh, 2015).

***Fermentation temperature***

There are a number of other factors that are commonly associated with a stuck fermentation, including the yeast strain, nitrogen availability and glucose concentration (Henderson and Block, 2014). However, fermentation temperature is considered as one of the main bottlenecks with regards to ethanol production by SSF and CBP strategies. Figure 9 showed the performance of the *S. cerevisiae* Ethanol Red™ T12 strain in different fermentation media and results confirmed that extra nitrogen (in the form of $(NH_4)_2SO_4$)) did not increase the fermentation of glucose to ethanol at a temperature of 37°C. Furthermore, increasing the pH of the conventional YP fermentation medium (to pH 5) did not improve fermentation conditions. Therefore, a lower pH was more favourable for starch conversion when using the TemG_Opt and TemA_Nat enzymes from *T. emersonii*, which have a pH optimum around 4 - 4.5 (Nielsen *et al*., 2002).

The demand for higher temperature fermentations began in the 1980s (Abdel-Banat *et al*., 2010). High-temperature fermentations may assist in making the simultaneous fermentation and ethanol extraction process more suitable for fuel ethanol production. Operational costs can be decreased (especially in regions with hot climates where cooling of fermentation vessels is required) and hydrolysis conditions improved. Ethanol production at high temperatures has several advantages, namely reduced risk of contamination, increased ethanol recovery, as well as decreased volumes of cooling waste-water effluent (Banat *et al*., 1998). Currently, the fermentation temperatures used in industry are between 30 - 34°C (Mukhtar *et al*., 2010). However, the internal temperature of a fermentation vessel exceeds these temperatures due to exogenic metabolic activities, as well as environmental temperatures in higher-temperature regions. This subsequently lowers the efficiency of ethanol production.

The effect of temperature on fermentation products has been described by a number of different research groups (Favaro *et al*., 2013b; Woo *et al*. 2014). Although *S. cerevisiae* is known for its high ethanol tolerance and relatively high ethanol concentrations, it still lacks the ability to ferment at higher than normal temperatures (Figure 10b). Moreover, ethanol concentrations of approximately 10% ($wt.vol^{-1}$) will reduce the fermentative activity of yeast by approximately 50% (Henderson and Block, 2014) and inhibit cell growth and viability. This leads to lower productivity and lower ethanol yields (Stanley *et al*., 2010). In order to improve ethanol tolerance of yeasts, the understanding of the cellular impact of ethanol toxicity needs to be explored.

Results for the comparison of ethanol production by recombinant *S. cerevisiae* Y294 and Ethanol Red™ T12 strains were in agreement with a study by Favaro *et al*. (2013b). They showed that at 30°C the laboratory *S. cerevisiae* Y294 strain had low fermentation vigour compared to the industrial strain at 30°C. The decreased ability to consume glucose could be explained by the *S. cerevisiae* Y294 strain displaying an optimum cultivation temperature around 25°C and not 30°C. Similarly, the amylolytic *S. cerevisiae* Ethanol Red™ T12 strain had reduced fermentation vigour at 37°C compared to 30°C (Figure 10b).

***Glycerol***

Reduced glycerol concentrations were observed when lower fermentation temperatures were used, indicating that better carbon conversion to ethanol occurred at a fermentation temperature of 30°C compared to 37°C (Figure 8d). Carbon source utilisation was important for the optimisation of ethanol production (Navarrete *et al*., 2014) and results showed that the

fermentation media influenced glycerol production (Figure 8). The commercially available TransFerm™ Yield+ yeast (Mascoma and Lallemand Biofuels and Distilled Spirts) was engineered to produce significantly less glycerol during fermentations so that more carbon can be utilised for ethanol production. In this study, the accumulating glycerol concentrations were below the conventional concentration (10 $g.l^{-1}$) (Huang *et al*., 2015) and therefore would not have had a significant effect on the yeast cells.

## Conclusion

Few studies have engineered *S. cerevisiae* Ethanol Red™ for the expression of gene cassettes or adapted it for desired characteristics. Demeke *et al*. (2013b) developed a D-xylose fermenting strain, Wallace-Salinas and Gorwa-Grauslund (2013) developed a strain capable of growing and fermenting spruce hydrolysate and Stovicek *et al*. (2015b) introduced a xylose consumption pathway into Ethanol Red™. To the applicant's knowledge, this study is the first to engineer *S. cerevisiae* Ethanol Red™ for the expression of both an α-amylase and glucoamylase for efficient raw starch conversion. Complete starch hydrolysis was achieved, although a fermentation temperature of 30°C enabled higher ethanol titres.

It is envisaged that recombinant amylolytic *S. cerevisiae* strains expressing an α-amylase and glucoamylase from *T. emersonii* could be used in commercial cold fermentation processes that are currently used by ethanol producers.

## References

Abdel-Banat BMA, Hoshida H, Ano A, Nonklang S, Akada R (2010) High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast? Appl Microbiol Biotechnol 85:861-867. doi: 10.1007/s00253-009-2248-5

Allison DS, Rey MW, Berka RM, Armstrong G, Dunn-Coleman NS (1992) Transformation of the thermophilic fungus *Humicola grisea* var. thermoidea and over- production of *Humicola* glucoamylase. Curr Genet 21:225–229. doi: 10.1007/BF00336845

Amore A, Faraco V (2012) Potential of fungi as category I Consolidated BioProcessing organisms for cellulosic ethanol production. Renew Sust Energ Rev 16:3286–3301. doi.10.1016/j.rser.2012.02.050

Balcerek M, Pielech-Przybylska K (2013) Effect of simultaneous saccharification and fermentation conditions of native triticale starch on the dynamics and efficiency of process and composition of the distillates obtained. J Chem Technol Biotechnol 88:615–622. doi: 10.1002/jctb.3873

Banat IM, Nigam P, Singh D, Marchant R, McHale AP (1998) Ethanol production at elevated temperatures and alcohol concentrations: Part I – Yeasts in general. World J Microbiol Biotechnol 14:809-821. doi: 10.1023/A:1008802704374

Bideaux C, Alfenore S, Cameleyre X, Molina-Jouve C, Uribelarrea JL, Guillouet SE (2006) Minimization of glycerol production during the high-performance fed-batch ethanolic fermentation process in *Saccharomyces cerevisiae*, using a metabolic model as a prediction tool. Appl Environ Microbiol 72:2134–2140. doi: 10.1128/AEM.72.3.2134-2140.2006

Bothast RJ, Schlicher MA (2005) Biotechnological processes for conversion of corn into ethanol. Appl Microbiol Biotechnol 67:19–25. doi:10.1007/s00253-004-1819-8

Brehmer B, Bals B, Sanders J, Dale B (2008) Improving the corn-ethanol industry: Studying protein separation techniques to obtain higher value-added product options for distillers grains. Biotechnol Bioeng 101:49–61. doi: 10.1002/bit.21881

Carbone A, Zinovyev A, Képès F (2003) Codon adaptation index as a measure of dominating codon bias. Bioinformatics 19:2005–2015. doi: 10.1093/bioinformatics/btg272

Celińska E, Białas W, Borkowska M, Grajek W (2015) Cloning, expression, and purification of insect (*Sitophilus oryzae*) alpha-amylase, able to digest granular starch, in *Yarrowia lipolytica* host. Appl Microbiol Biotechnol 99:2727–2739. doi: 10.1007/s00253-014-6314-2

Chen J, Zhang Y, Zhao C, Li A, Zhou Q, Li D (2007) Cloning of a gene encoding thermostable glucoamylase from *Chaetomium thermophilum* and its expression in *Pichia pastoris*. J Appl Microbiol 103:2277–2284. doi: 10.1111/j.1365-2672.2007.03475.x

Chi Z, Wang F, Chi Z, Yue L, Liu G, Zhang T (2009) Bioproducts from *Aureobasidium pullulans*, a biotechnologically important yeast. Appl Microbiol Biotechnol 82:793–804. doi: 10.1007/s00253-009-1882-2

Cho KM, Yoo YJ, Kang HS (1999) δ-Integration of endo/exoglucanase and β-glucosidase genes into the yeast chromosomes for direct conversion of cellulose to ethanol. Enzyme Microb Technol 25:23–30. doi: 10.1016/S0141-0229(99)00011-3

Damasceno LM, Huang CJ, Batt CA (2012) Protein secretion in *Pichia pastoris* and advances in protein production. Appl Microbiol Biotechnol 93:31–39. doi: 10.1007/s00253-011-3654-z

Demeke MM, Dumortier F, Li Y, Broeckx T, Foulquié-Moreno MR, Thevelein JM (2013) Combining inhibitor tolerance and D-xylose fermentation in industrial *Saccharomyces cerevisiae* for efficient lignocellulose-based bioethanol production. Biotechnol Biofuels 6:120. doi: 10.1186/1754-6834-6-120

Demeke MM, Dietz H, Li Y, Foulquié-Moreno MR, Mutturi S, Deprez S, Den Abt T, Bonini BM, Liden G, Dumortier F, Verplaetse A, Boles E, Thevelein JM (2013b) Development of a D-xylose fermenting and inhibitor tolerant industrial *Saccharomyces cerevisiae* strain with high performance in lignocellulose

hydrolysates using metabolic and evolutionary engineering. Biotechnol Biofuels 6:89. doi: 10.1186/1754-6834-6-89

Den Haan R, Rose SH, Lynd LR, van Zyl WH (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. Metab Eng 9:87-94. doi: 10.1016/j.ymben.2006.08.005

den Haan R, Kroukamp H, Mert M, Bloom M, Görgens JF, van Zyl WH (2013) Engineering *Saccharomyces cerevisiae* for next generation ethanol production. J Chem Technol Biotechnol 88:983–991. doi: 10.1002/jctb.4068

den Haan R, van Rensburg E, Rose SH, Görgens J, van Zyl WH (2015) Progress and challenges in the engineering of non-cellulolytic microorganisms for consolidated bioprocessing. Curr Opin Biotechnol 33:32–38. doi: 10.1016/j.copbio.2014.10.003

Eksteen JM, van Rensburg P, Cordero Otero RR, Pretorius IS (2003) Starch fermentation by recombinant *Saccharomyces cerevisiae* strains expressing the α-amylase and glucoamylase genes from *Lipomyces kononenkoae* and *Saccharomycopsis fibuligera*. Biotechnol Bioeng 84:639–646. doi: 10.1002/bit.10797

Favaro L, Jooste T, Basaglia M, Rose SH, Saayman M, Görgens JF, Casella S, van Zyl WH (2013) Designing industrial yeasts for consolidated bioprocessing of starchy biomass to ethanol. Bioengineered 4:97–102. doi: 10.4161/bioe.22268

Favaro L, Basaglia M, Trento A, van Rensburg E, García-Aparicio M, van Zyl WH, Casella S (2013b) Exploring grape marc as trove for new thermotolerant and inhibitor-tolerant *Saccharomyces cerevisiae* strains for second-generation bioethanol production. Biotechnol Biofuels 6:168. doi: 10.1186/1754-6834-6-168

Favaro L, Viktor M, Rose SH, Viljoen-Bloom M, van Zyl WH, Basaglia M, Cagnin L, Casella S (2015) Consolidated bioprocessing of starchy substrates into ethanol by industrial *Saccharomyces cerevisiae* strains secreting fungal amylases. Biotechnol Bioeng 112:1751–1760. doi: 10.1002/bit.25591

Futatsumori-Sugai M, Tsumoto K (2010) Signal peptide design for improving recombinant protein secretion in the baculovirus expression vector system. Biochem Biophys Res Commun 391:931–935. doi: 10.1016/j.bbrc.2009.11.167

Görgens JF, Bressler DC, van Rensburg E (2015) Engineering *Saccharomyces cerevisiae* for direct conversion of raw, uncooked or granular starch to ethanol. Crit Rev Biotechnol 35:369–391. doi: 10.3109/07388551.2014.888048

Gupta R, Gigras P, Mohapatra H, Goswami VK, Chauhan B (2003) Microbial α-amylases: a biotechnological perspective. Process Biochem 38:1599–1616. doi: 10.1016/S0032-9592(03)00053-0

Hashimoto Y, Koyabu N, Imoto T (1998) Effects of signal sequences on the secretion of hen lysozyme

by yeast: construction of four secretion cassette vectors. Protein Eng 11:75–77. doi: 10.1093/protein/11.2.75

Henderson CM, Block DE (2014) Examining the role of membrane lipid composition in determining the ethanol tolerance of *Saccharomyces cerevisiae*. Appl Environ Microbiol 80:2966–2972. doi: 10.1128/AEM.04151-13

Hii SL, Tan JS, Ling TC, Ariff AB (2012) Pullulanase: role in starch hydrolysis and potential industrial applications. Enzyme Res 2012:921362. doi: 10.1155/2012/921362

Houbraken J, Spierenburg H, Frisvad JC (2012) *Rasamsonia*, a new genus comprising thermotolerant and thermophilic *Talaromyces* and *Geosmithia* species. Antonie Van Leeuwenhoek 101: 403–421. doi: 10.1007/s10482-011-9647-1

Huang H, Qureshi N, Chen MH, Liu W, Singh V (2015) Ethanol production from food waste at high solids content with vacuum recovery technology. J Agric Food Chem 63:2760–2766. doi: 10.1021/jf5054029

Inlow D, McRae J, Ben-Bassat A (1988) Fermentation of corn starch to ethanol with genetically engineered yeast. Biotechnol Bioeng 32:227–234. doi: 10.1002/bit.260320213

Kim HR, Im YK, Ko HM, Chin, JE, Kim IC, Lee HB, Bai S (2011) Raw starch fermentation to ethanol by an industrial distiller's yeast strain of *Saccharomyces cerevisiae* expressing glucoamylase and α-amylase genes. Biotechnol Lett 33:1643–1648. doi: 10.1007/s10529-011-0613-9

Knox AM, Du Preez JC, Kilian SG (2004) Starch fermentation characteristics of Saccharomyces cerevisiae strains transformed with amylase genes from *Lipomyces kononenkoae* and *Saccharomycopsis fibuligera*. Enzyme Microb Technol 34:453–460.

Liakopoulou-Kyriakides M, Karakatsanis A, Stamatoudis M, Psomas S (2001) Synergistic hydrolysis of crude corn starch by α-amylases and glucoamylases of various origins. Cereal Chem 78:603–607. doi: 10.1094/CCHEM.2001.78.5.603

Machovic M, Janecek S (2006) Starch-binding domains in the post-genome era. Cell Mol Life Sci 63:2710–2724Mamo G, Gessesse A (1999) Purification and characterization of two raw-starch-digesting thermostable α-amylases from a thermophilic *Bacillus*. Enzyme Microb Technol 25:433–438. doi: 10.1016/S0141-0229(99)00068-X

Man J, Yang Y, Huang J, Zhang C, Zhang F, Wang Y, Gu M, Liu Q, Wei C (2013) Morphology and structural properties of high-amylose rice starch residues hydrolysed by amyloglucosidase. Food Chem 138:2089–2098. doi: 10.1016/j.foodchem.2012.12.009

Miller GL (1959) Use of dinitrosalicyclic acid reagent for determination of reducing sugars. Anal Chem 31:426–428. doi: 10.1021/ac60147a030

Mitsuiki S, Mukae K, Sakai M, Goto M, Hayashida S, Furukawa K (2005) Comparative characterization of raw starch hydrolyzing α-amylases from various *Bacillus* strains. Enzyme Microb Technol 37:410–416. doi: 10.1016/j.enzmictec.2005.02.022

Mukhtar K, Asgher M, Afghan S, Hussain K, Zia-Ul-Hussnain S (2010) Comparative study on two commercial strains of *Saccharomyces cerevisiae* for optimum ethanol production on industrial scale. J Biomed Biotechnol 210:419586. doi: 10.1155/2010/419586

Naguleswaran S, Vasanthan T, Hoover R, Bressler D (2013) The susceptibility of large and small granules of waxy, normal and high-amylose genotypes of barley and corn starches toward amylolysis at sub-gelatinization temperatures. Food Res Int 51:771–782. doi: 10.1016/j.foodres.2013.01.057

Navarrete C, Nielsen J, Siewers V (2014) Enhanced ethanol production and reduced glycerol formation in fps1Δ mutants of *Saccharomyces cerevisiae* engineered for improved redox balancing. AMB Express 4:86. doi: 10.1186/s13568-014-0086-z

Nielsen BR, Lehmbeck J, Frandsen TP (2002) Cloning, heterologous expression, and enzymatic characterization of a thermostable glucoamylase from *Talaromyces emersonii*. Protein Expr Purif 26:1–8. doi: 10.1016/S1046-5928(02)00505-3

Njokweni A, Rose SH, van Zyl WH (2012) Fungal β-glucosidase expression in *Saccharomyces cerevisiae*. J Mol Microbiol Biotechnol 39:1445–1452. doi: 10.1007/s10295-012-1150-9

Nkomba EY, van Rensburg E, Chimphango AF, Görgens JF (2016) The influence of sorghum grain decortication on bioethanol production and quality of the distillers' dried grains with solubles using cold and conventional warm starch processing. Bioresour Technol 203:181–189. doi: 10.1016/j.biortech.2015.12.045

O'Connell K, Stults JT (1997) Identification of mouse liver proteins on two-dimensional electrophoresis gels by matrix-assisted laser desorption/ionization mass spectrometry of in situ enzymatic digests. Electrophoresis 18:349–359. doi: 10.1186/1754-6834-6-167

Presečki AV, Blažević ZF, Vasić-Rački D (2013) Complete starch hydrolysis by the synergistic action of amylase and glucoamylase: impact of calcium ions. Bioprocess Biosyst Eng 36:1555–1562. doi: 10.1007/s00449-013-0926-2

Robertson GH, Wong DW, Lee CC, Wagschal K, Smith MR, Orts WJ (2006) Native or raw starch digestion: a key step in energy efficient biorefining of grain. J Agric Food Chem 54:353–365. doi: 10.1021/jf051883m

Salehi Jouzani G, Taherzadeh MJ (2015) Advances in consolidated bioprocessing systems for bioethanol and butanol production from biomass: a comprehensive review. Biofuel Res J 2:152–195. doi: 10.18331/BRJ2015.2.1.4

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor,

New York: Cold Spring Harbor Laboratory Press

Santiago M, Linares L, Sánchez S, Rodríguez-Sanoja R. (2005) Functional characteristics of the starch-binding domain of *Lactobacillus amylovorus* α-amylase. Biologia (Bratislava) 60:111–114.

Schmidt SA, Tran T, Chambers PJ, Herderich MJ, Pretorius IS (2006) Developing indicators of wine yeast performance: an overview of the impact of ethanol stress. Austral New Zealand Wine Industry J 21:24–30.

Solis-Escalante D, Kuijpers NGA, Bongaerts N, Bolat I, Bosman L, Pronk JT, Daran JM, Daran-Lapujade P (2013) amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. FEMS Yeast Res 13: 126–139. doi: 10.1111/1567-1364.12024

Stanley D, Bandara A, Fraser S, Chambers PJ, Stanley GA (2010) The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*. J Appl Microbiol 109:13–24. doi: 10.1111/j.1365-2672.2009.04657.x

Stovicek V, Borodina I, Forster J (2015) CRISPR–Cas system enables fast and simple genome editing of industrial *Saccharomyces cerevisiae strains*. Metab Eng Commun 2:13–22. doi: 10.1016/j.meteno.2015.03.00

Stovicek V, Borja GM, Forster J, Borodina I (2015b) EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial *Saccharomyces cerevisiae* strains. J Ind Microbiol Biotechnol 42:1519-1531. doi: 10.1007/s10295-015-1684-8

Sun H, Zhao P, Ge X, Xia Y, Hao Z, Liu J, Peng M (2010) Recent advances in microbial raw starch degrading enzymes. Appl Biochem and Biotechnol 160:988–1003. doi: 10.1007/s12010-009-8579-y

Szymanowska-Powałowska D, Lewandowicz G, Kubiak P, Błaszczak W (2014) Stability of the process of simultaneous saccharification and fermentation of corn flour. The effect of structural changes of starch by stillage recycling and scaling up of the process. Fuel 119:328–334. doi.org/10.1016/j.fuel.2013.11.034

Thorsen TS, Johnsen AH, Josefsen K, Jensen B (2006) Identification and characterization of glucoamylase from the fungus *Thermomyces lanuginosus*. Biochim Biophys Acta 1764:671–676. doi: 10.1016/j.bbapap.2006.01.009

Tyo KE J, Liu Z, Petranovic D, Nielsen J (2012) Imbalance of heterologous protein folding and disulfide bond formation rates yields runaway oxidative stress. BMC Biol 10:16. doi: 10.1186/1741-7007-10-16

Uthumporn U, Zaidul ISM, Karim AA (2010) Hydrolysis of granular starch at sub-gelatinization temperature using a mixture of amylolytic enzymes. Food and Bioprod Process 88:47-54. doi: 10.1016/j.fbp.2009.10.001

van Rensburg E, den Haan R, Smith J, van Zyl WH, Görgens JF (2012) The metabolic burden of

cellulase expression by recombinant *Saccharomyces cerevisiae* Y294 in aerobic batch culture. Appl Microbiol and Biotechnol 96:197–209. doi: 10.1007/s00253-012-4037-9

van Wyk N, den Haan R, van Zyl WH (2010) Heterologous production of NpCel6A from *Neocallimastix patriciarum* in *Saccharomyces cerevisiae*. Enzyme Microb Technol 46:378–383. doi: 10.1016/j.enzmictec.2009.11.005

van Zyl WH, Bloom M, Viktor MJ (2012) Engineering yeasts for raw starch conversion. Appl Microbiol Biotechnol 95:1377-1388. doi: 10.1007/s00253-012-4248-0

Viktor MJ, Rose SH, van Zyl WH, Viljoen-Bloom M (2013) Raw starch conversion by *Saccharomyces cerevisiae* expressing *Aspergillus tubingensis* amylases. Biotechnol Biofuels 6:167. doi: 10.1186/1754-6834-6-167

Wallace-Salinas V, Gorwa-Grauslund MF (2013) Adaptive evolution of an industrial strain of *Saccharomyces cerevisiae* for combined tolerance to inhibitors and temperature. Biotechnol Biofuels 6:151. doi: 10.1186/1754-6834-6-151

Wang K, Luo H, Shi P, Huang H, Bai Y, Yao B (2014) A highly-active endo-1,3-1,4-beta-glucanase from thermophilic *Talaromyces emersonii* CBS394.64 with application potential in the brewing and feed

Wong D, Batt Throne SB, Robertson GH, Lee CC, Wagschal KC (2010) Chromosomal integration of both an alpha-amylase and a glucoamylase gene in *Saccharomyces cerevisiae* for starch conversion. Ind Biotechnol 6:112–119. doi: 10.1089/ind.2010.0010.

Woo JM, Yang KM, Kim SU, Blank LM, Park JB (2014) High temperature stimulates acetic acid accumulation and enhances the growth inhibition and ethanol production by *Saccharomyces cerevisiae* under fermenting conditions. Appl Microbiol Biotechnol 98:6085–6094. doi: 10.1007/s00253-014-5691-x

Yamakawa S, Yamada R, Tanaka T, Ogino C, Kondo A (2012) Repeated fermentation from raw starch using *Saccharomyces cerevisiae* displaying both glucoamylase and α-amylase. Enzyme Microb Technol 50:343–347. doi: 10.1016/j.enzmictec.2012.03.005

Yang S, Jia N, Li M, Wang J (2011) Heterologous expression and efficient ethanol production of a *Rhizopus* glucoamylase gene in *Saccharomyces cerevisiae*. Mol Biol Rep 38:59–64. doi: 10.1007/s11033-010-0077-3

**CLAIMS:**

1. A recombinant yeast which has been transformed with a heterologous gene which is capable of expressing a polypeptide comprising an amino acid sequence which is at least 70% identical to SEQ ID NO: 1.

2. A recombinant yeast according to claim 1, wherein the amino acid sequence of the polypeptide is at least 80% identical to SEQ ID NO: 1.

3. A recombinant yeast according to claim 1, wherein the amino acid sequence of the polypeptide is at least 90% identical to SEQ ID NO: 1.

4. A recombinant yeast according to claim 1, wherein the amino acid sequence of the polypeptide is identical to SEQ ID NO: 1.

5. A recombinant yeast according to claim 1, wherein the nucleic acid sequence of the heterologous gene is at least 70% identical to either of SEQ ID NOS: 3 and 5.

6. A recombinant yeast according to claim 5, wherein the nucleic acid sequence of the heterologous gene is at least 70% identical SEQ ID NO: 3.

7. A recombinant yeast according to claim 5, wherein the nucleic acid sequence of the heterologous gene is at least 70% identical to SEQ ID NO: 5.

8. A recombinant yeast according to claim 5, wherein the nucleic acid sequence of the heterologous gene is at least 80% identical to either of SEQ ID NOS: 3 and 5.

9. A recombinant yeast according to claim 5, wherein the nucleic acid sequence of the heterologous gene is at least 90% identical to either of SEQ ID NOS: 3 and 5.

10. A recombinant yeast according to claim 5, wherein the nucleic acid sequence of the heterologous gene is identical to either of SEQ ID NOS: 3 and 5.

11. A recombinant yeast according to any one of claims 1 to 10, which has further been transformed with a second heterologous gene which is capable of expressing a second polypeptide comprising an amino acid sequence which is at least 70% identical to SEQ ID NO: 2.

12. A recombinant yeast according to claim 11, wherein the amino acid sequence of the second polypeptide is at least 80% identical to SEQ ID NO: 2.

13. A recombinant yeast according to claim 11, wherein the amino acid sequence of the second polypeptide is at least 90% identical to SEQ ID NO: 2.

14. A recombinant yeast according to claim 11, wherein the amino acid sequence of the second polypeptide is identical to SEQ ID NO: 2.

15. A recombinant yeast according to claim 11, wherein the nucleic acid sequence of the second heterologous gene is at least 70% identical to SEQ ID NO: 4.

16. A recombinant yeast according to claim 15, wherein the nucleic acid sequence of the second heterologous gene is at least 80% identical to SEQ ID NO: 4.

17. A recombinant yeast according to claim 15, wherein the nucleic acid sequence of the second heterologous gene is at least 90% identical to SEQ ID NO: 4.

18. A recombinant yeast according to claim 15, wherein the nucleic acid sequence of the second heterologous gene is identical to SEQ ID NO: 4.

19. A recombinant yeast according to any one of claims claim 1 to 18, wherein the yeast is a yeast which is capable of converting sugars to alcohol.

20. A recombinant yeast according to claim 19, wherein the sugars include glucose.

21. A recombinant yeast according to any one of claims 1 to 20, wherein the yeast is a *Saccharomyces* species.

22. A recombinant yeast according to claim 21, wherein the yeast is a *Saccharomyces cerevisiae* species.

23. A recombinant yeast according to any one of claims 19 to 22, wherein the alcohol is selected from the group consisting of butanol and ethanol.

24. A recombinant yeast according to claim 23, wherein the alcohol is ethanol.

25. A recombinant yeast according to any one of claims 1 to 24, which is capable of hydrolysing raw starch in the absence of enzymes from a source other than the recombinant yeast.

26. A recombinant yeast according to claim 25, which is capable of hydrolysing the raw starch at a temperature of below about 40°C.

27. A process for producing an alcohol from sugars, the process comprising the step of using a recombinant yeast according to any one of claims 1 to 26 to convert the sugars to alcohol.

28. A process according to claim 27, wherein the sugars comprise glucose.

29. A process according to either one of claims 27 or 28, wherein the alcohol is selected from the group consisting of ethanol and butanol.

30. A process according to claim 29, wherein the alcohol is ethanol.

31. A process for producing an alcohol from starch, the process comprising the step of using a recombinant yeast according to any one of claims 1 to 26 to convert the starch to alcohol.

32. A process according to claim 31, wherein the starch is raw starch.

33. A process according to either one of claims 31 or 32, wherein the starch is grain starch.

34. A process according to any one of claims 31 to 33, wherein the alcohol is selected from the group consisting of ethanol and butanol.

35. A process according to claim 34, wherein the alcohol is ethanol.

36. A process according to any one of claims 32 to 35, wherein the raw starch is hydrolysed by the recombinant yeast without requiring cooking of the starch.

37. A process according to claim 36, wherein the raw starch is hydrolysed by the recombinant yeast at a temperature of below about 40°C.

38. A process according to any one of claims 31 to 37, wherein the alcohol is produced from the starch without the addition of enzymes from a source other than the yeast.

39. A process according to any one of claims 31 to 37, wherein amylolytic enzymes are initially added to the process to reduce the time taken to convert the starch to alcohol, the amylolytic enzymes being added in an amount which is at least about 50% less than the amount added to cold hydrolysis processes which do not use the recombinant yeast of claim 1.

40. A process according to claim 39, wherein the added amylolytic enzymes are a cocktail of enzymes which together are capable of hydrolysing raw starch.

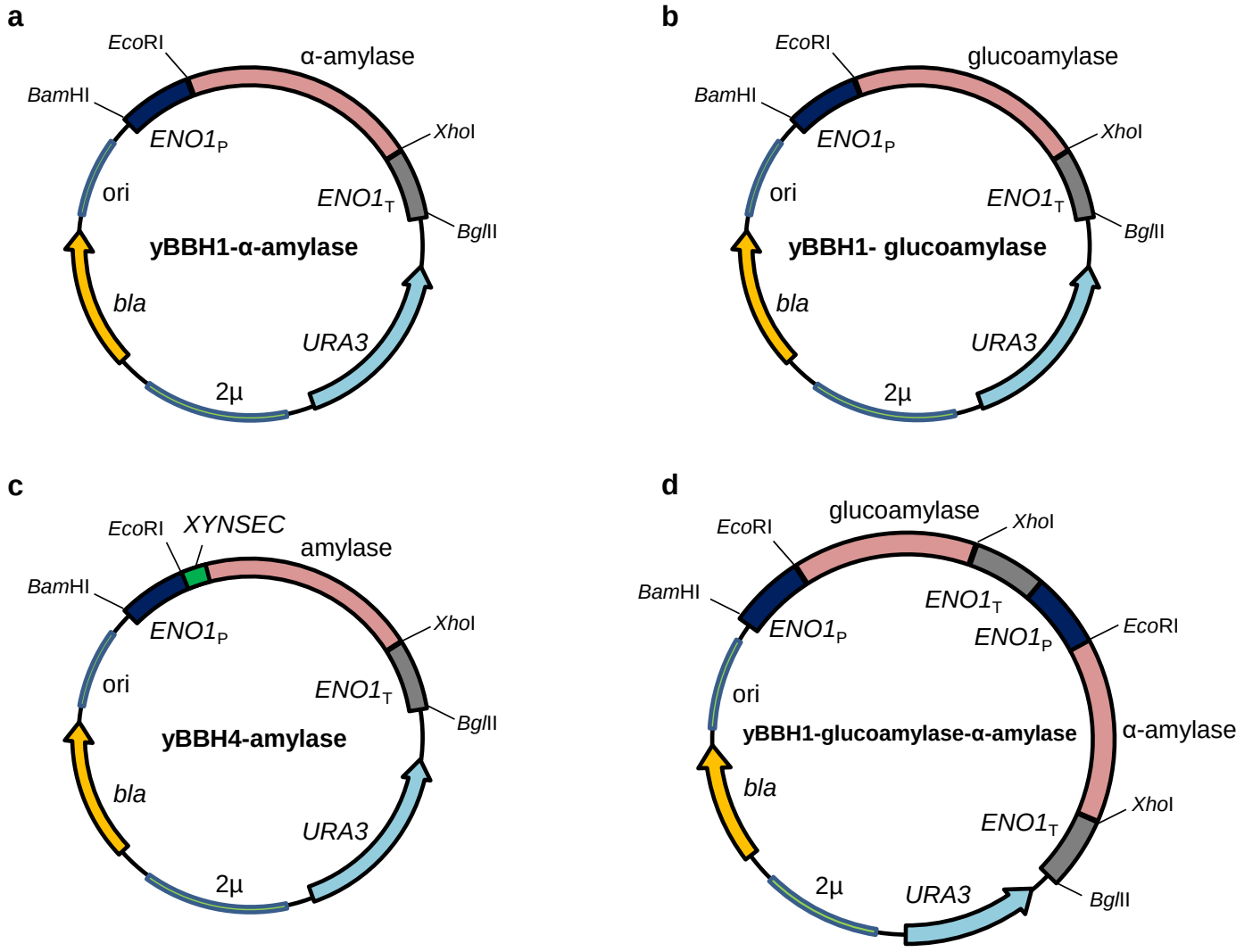


**Fig. 1**

**Fig. 2**

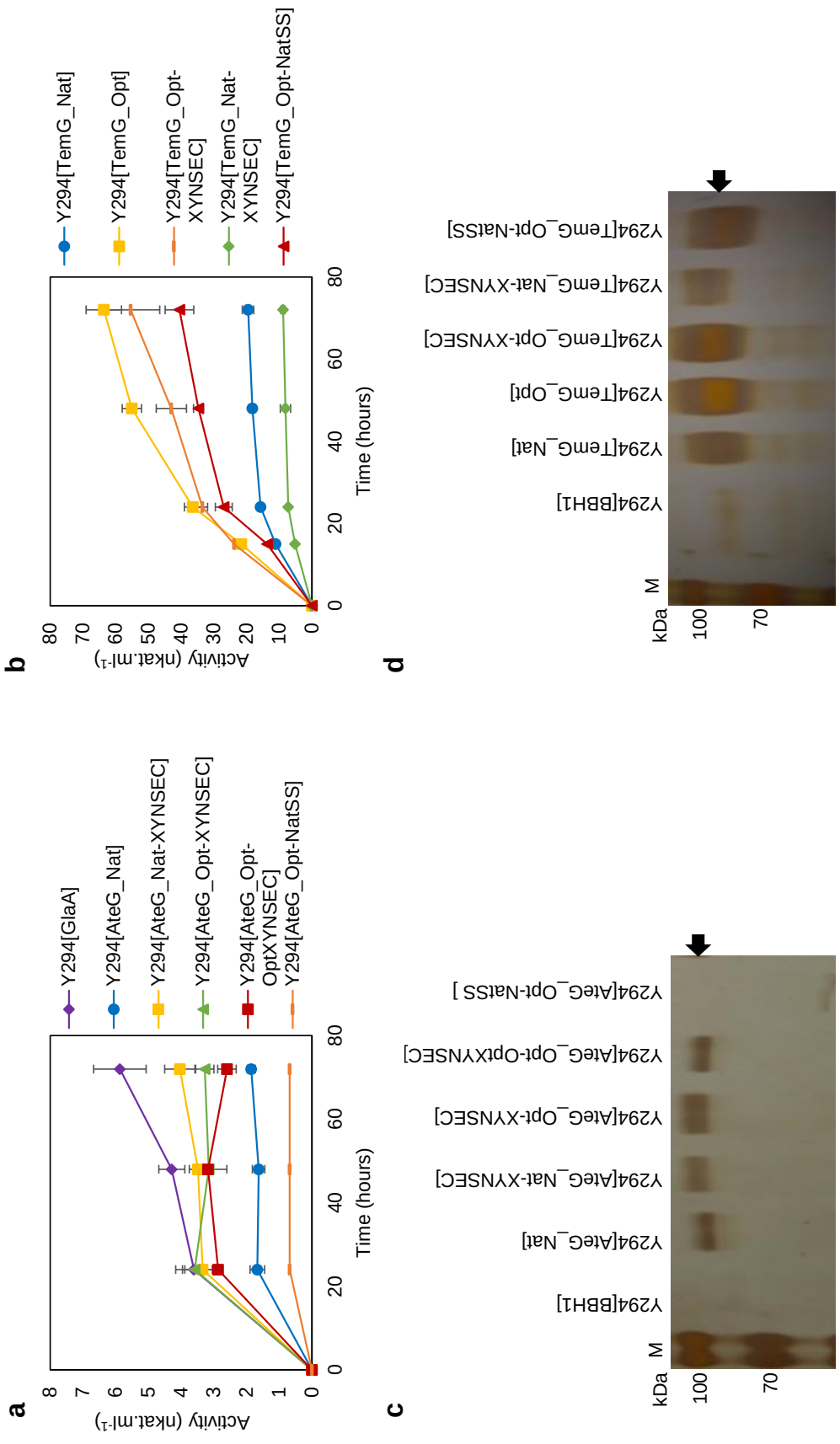


**Fig. 3**

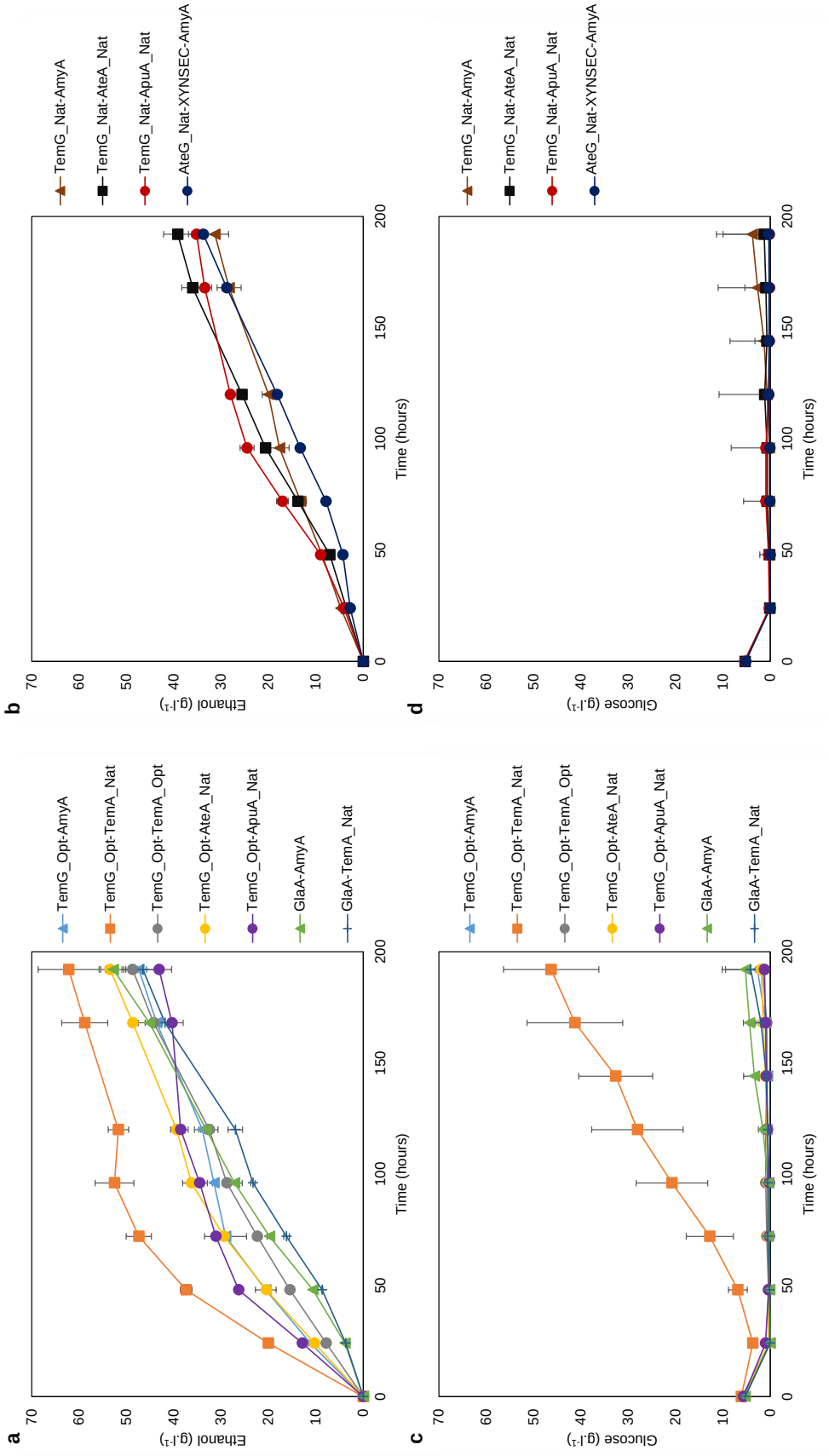


**Fig. 4**

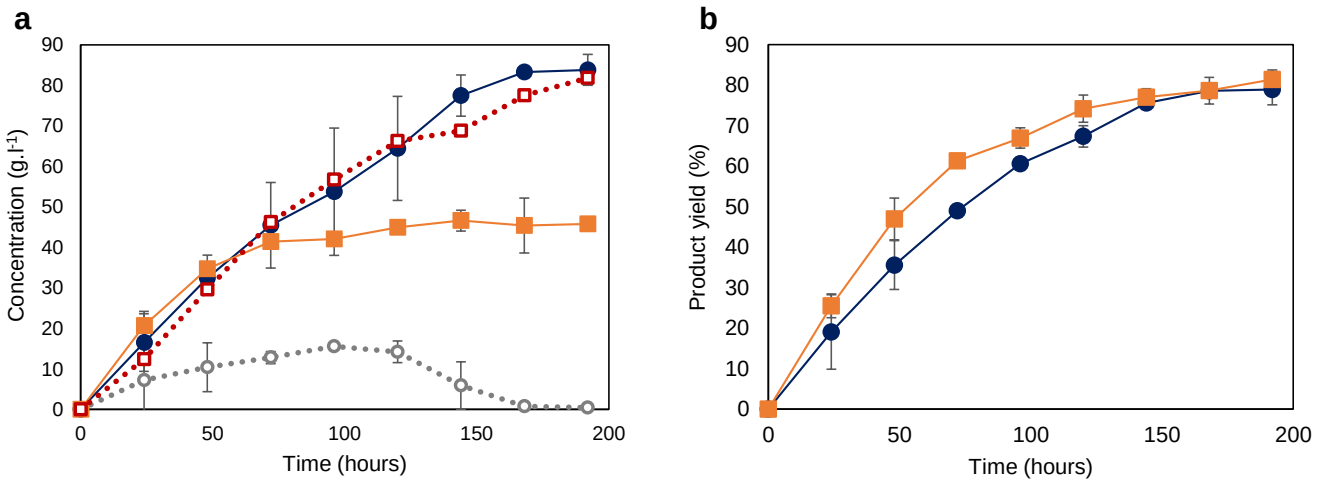


**Fig. 5**

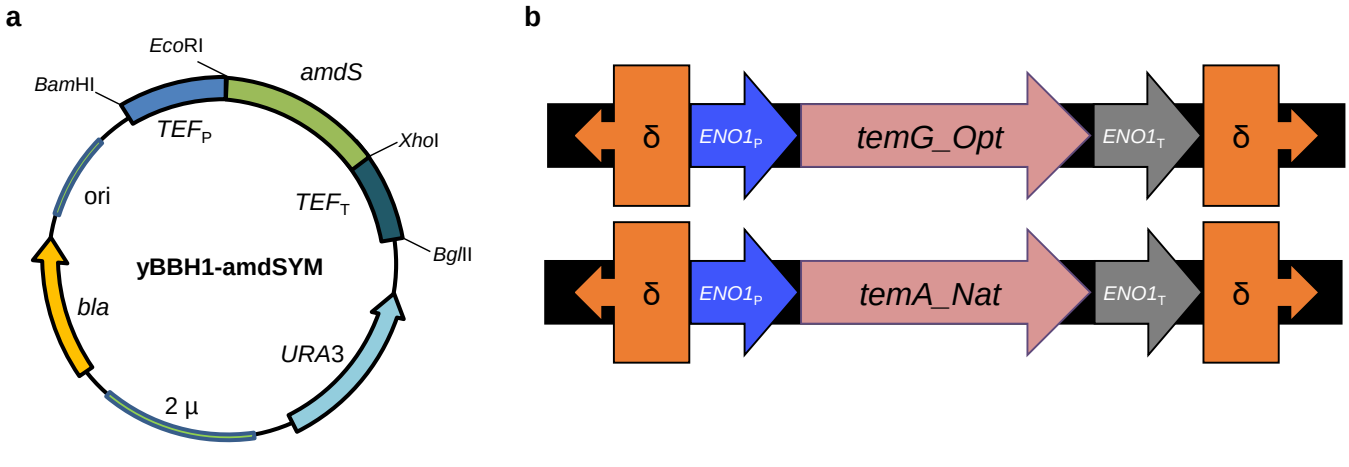


**Fig. 6**

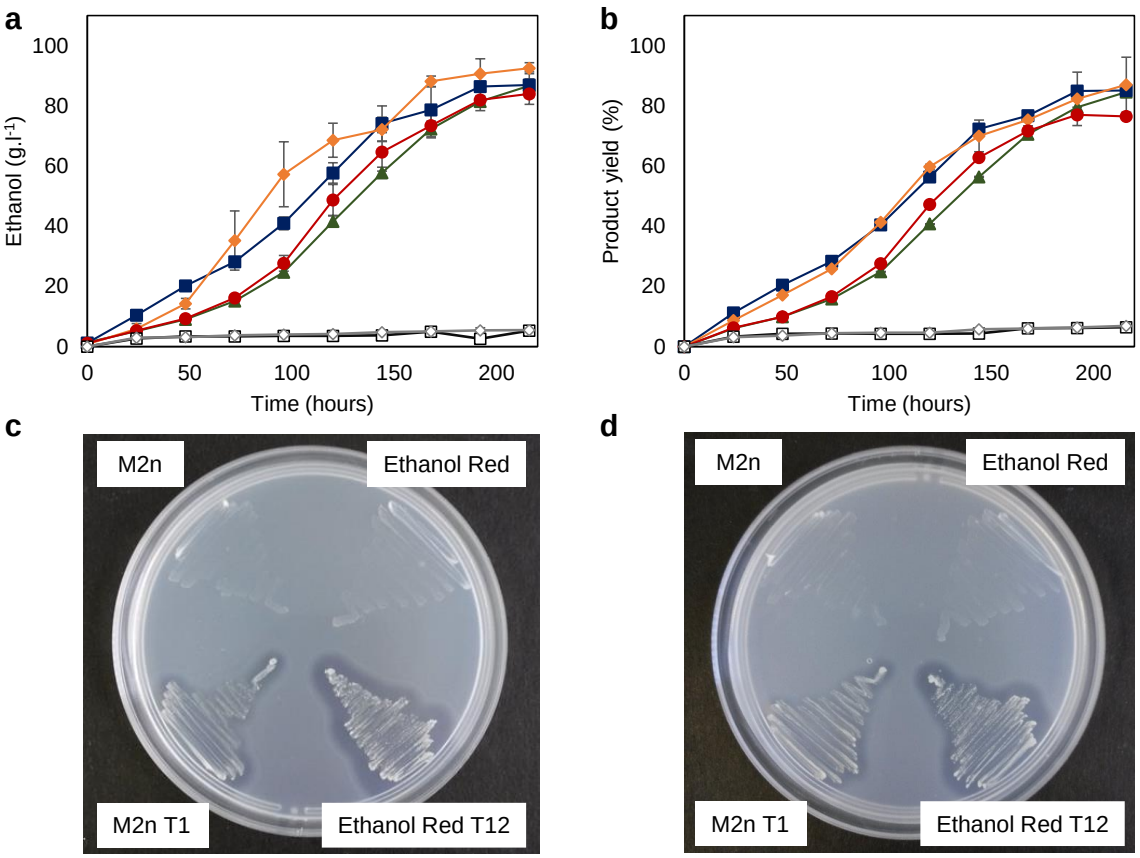


**Fig. 7**

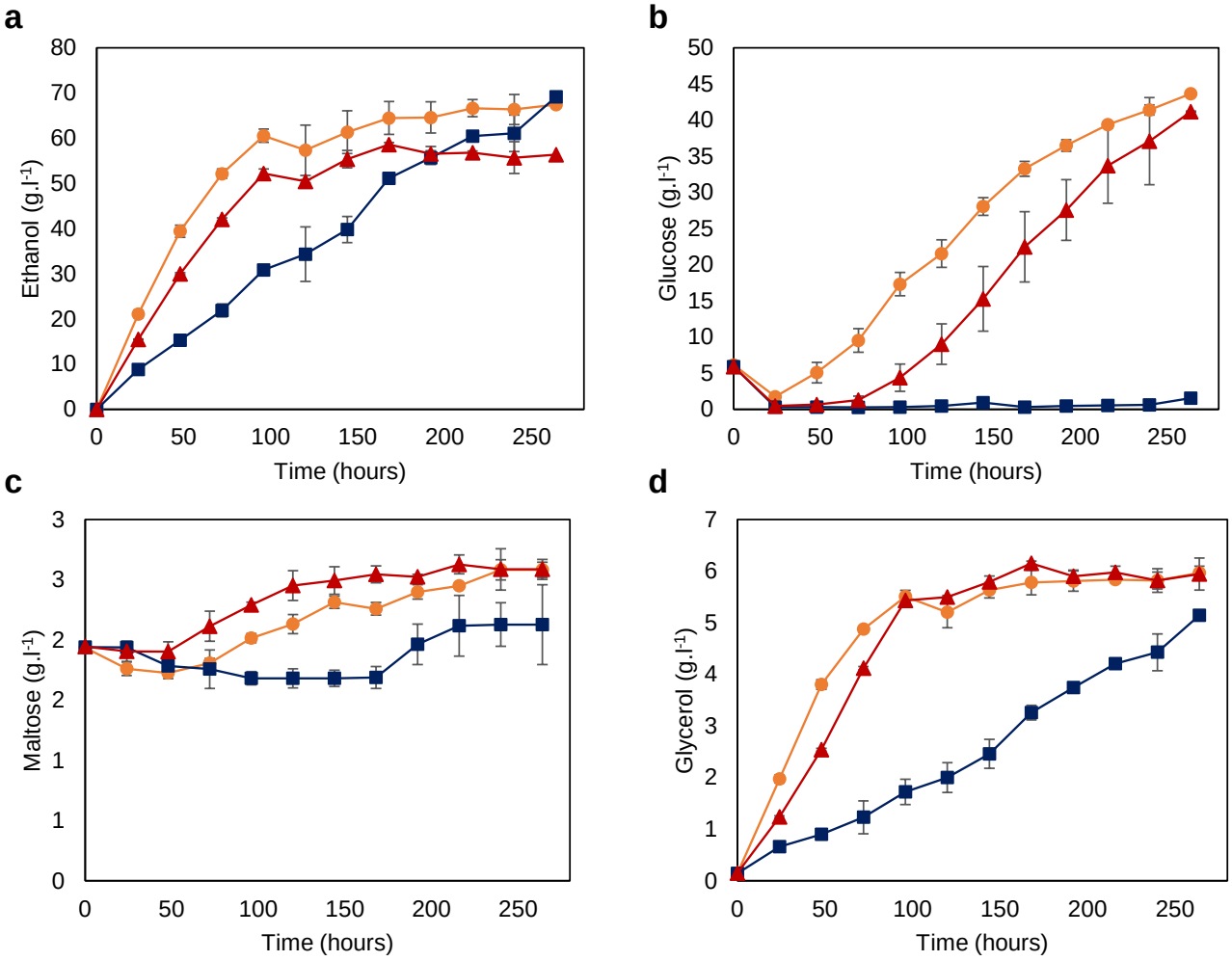


**Fig. 8**

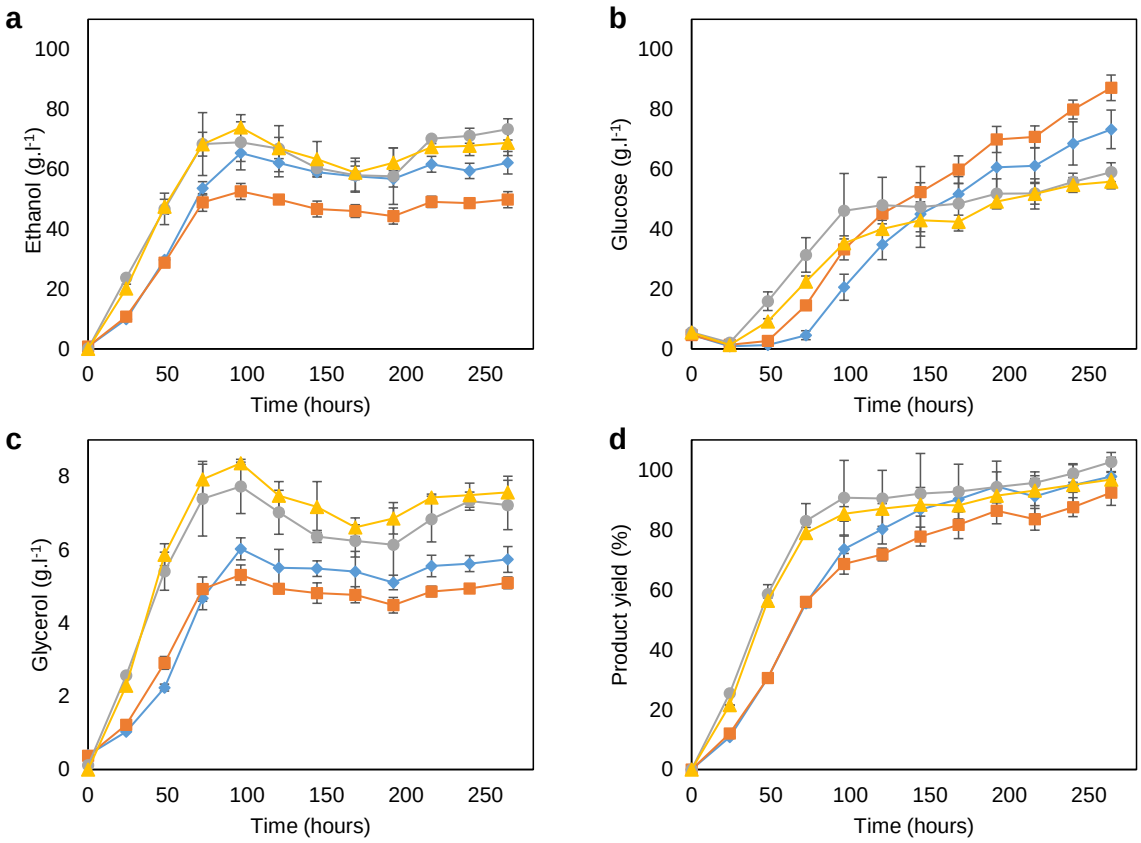


**Fig. 9**

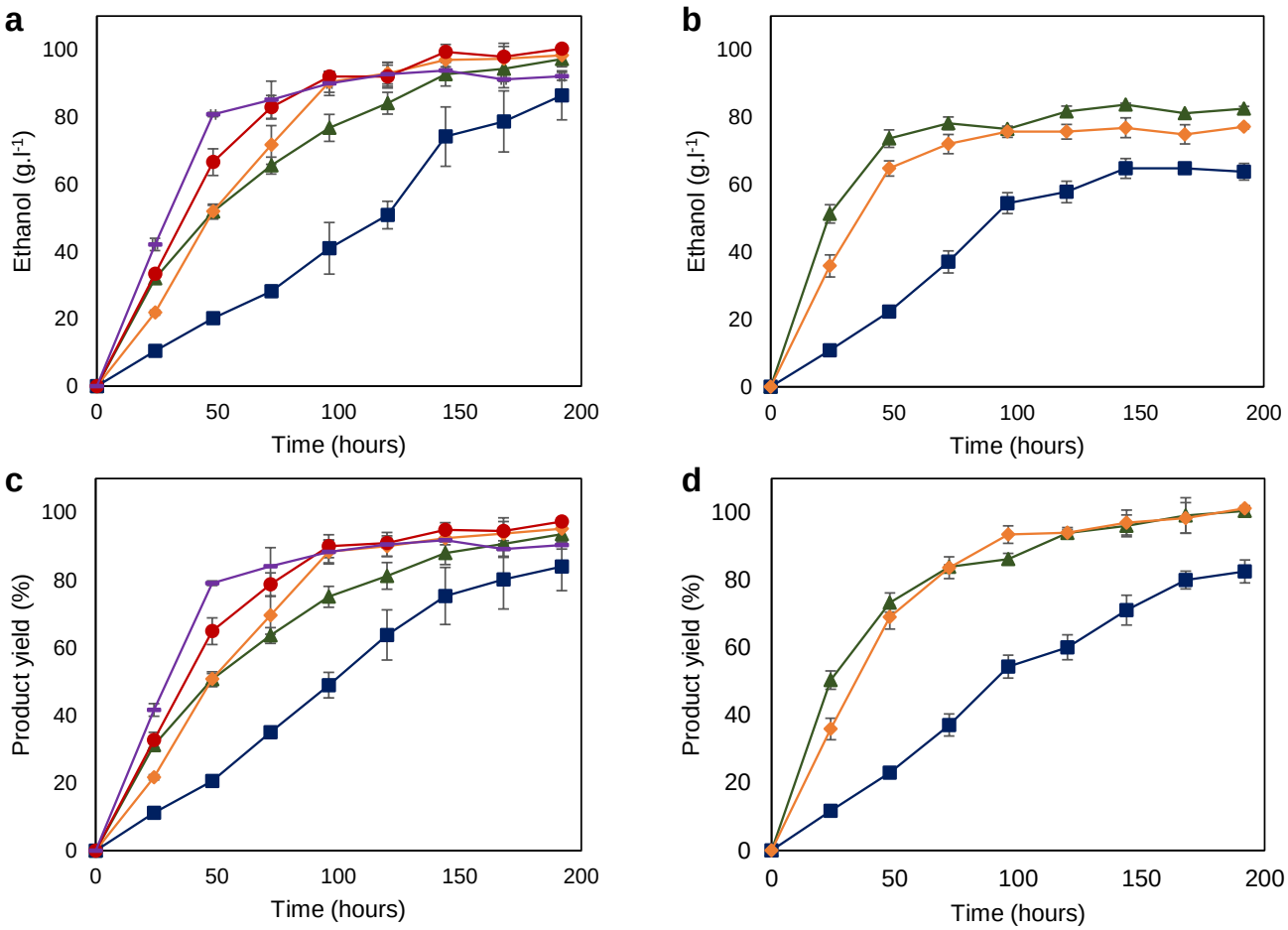


**Fig. 10**

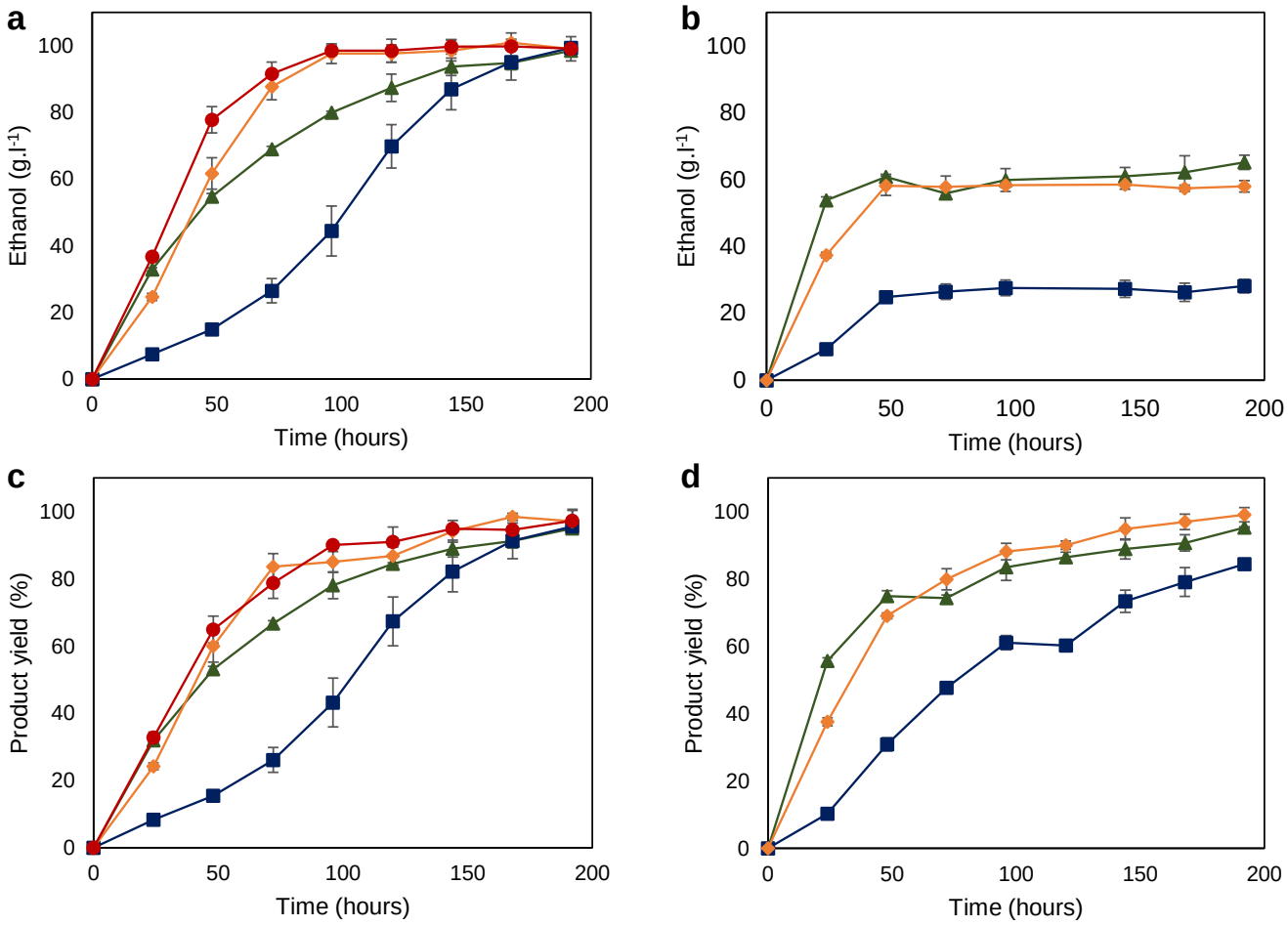


**Fig. 11**

**MASLVAGALCILGLTPAAFA**RAPVAARATGSLDSFLATETPIALQGVLNNIGPNGADVAGASAGIVVA
SPSRSDPNYFYSWTRDAALTAKYLVDAFIAGNKDLEQTIQQYISAQAKVQTISNPSGDLSTGGLGEPK
FNVNETAFTGPWGRPQRDGPALRATALIAYANYLIDNGEASTADEIIWPIVQNDLSYITQYWNSSTFD
LWEEVEGSSFFTTAVQHRALVEGNALATRLNHTCSNCVSQAPQVLCFLQSYWTGSYVLANFGGSGRSG
KDVNSILGSIHTFDPAGGCDDSTFQPCSARALANHKVVTDSFRSIYAINSGIAEGSAVAVGRYPEDVY
QGGNPWYLATAAAAEQLYDAIYQWKKIGSISITDVSLPFFQDIYPSAAVGTYNSGSTTFNDIISAVQT
YGDGYLSIVEKYTPSDGSLTEQFSRTDGTPLSASALTWSYASLLTASARRQSVVPASWGESSASSVPA
VCSATSATGPYSTATNTVWPSSGSGSSTTTSSAPCTTPTSVAVTFDEIVSTSYGETIYLAGSIPELGN
WSTASAIPLRADAYTNSNPLWYVTVNLPPGTSFEYKFFKNQTDGTIVWEDDPNRSYTVPAYCGQTTAI
LDDSWQ

**Fig. 12**

**MTPFVLTAVLFLLGNAVLA**LTPAEWRKQSIYFLLTDRFGRADNSTTAACDVTERIYCGGSWQGIINHL
DYIQGMGFTAIWISPVTEQLPQNTGEGEAYHGYWQQEIYTVNSNFGTSDDLLALSKALHDRGMYLMVD
VVANHMGYDGDGDSVDYSVFNPFNSSSYFHPYCLITDYSNQTDVEDCWLGDTTVSLPDLNTTETVVRT
IWYDWVADLVSNYSIDGLRIDTVKHVEKSFWPGYNSAAGVYCVGEVLDGDPSYTCPYQDYLDGVLNYP
IYYQLLYAFESSSGSISNLYNMINSVASECSDPTLLGNFIENHDNPRFASYTSDYSLAKNVIAFIFFS
DGIPIVYAGQEQHYNGGNDPYNREATWLSGYSTTAELYTFIATTNAIRSLAISVDSEYLTYKNDPFYY
DSNTLAMRKGSDGLQVITVLSNLGADGSSYTLTLSGSGYSSGTELVEAYTCTTVTVDSNGDIPVPMES
GLPRVFLPASSFSGSSLCSSSPSPTTTTSTSTSTTSTACTTATAVAVLFEELVTTTYGENVYLSGSIS
QLGDWNTDDAVALSAANYTSSNPLWYVTVTLPVGTSFEYKFIKKEENGDVEWESDPNRSYTVPTACTG
ATETIVDTWR

**Fig. 13**

**ATGGCCTCCTTAGTCGCAGGTGCCTTATGTATTTTAGGTTTGACCCCAGCAGCCTTCGCA**AGAGCCCC
AGTCGCAGCCAGAGCAACAGGTTCATTGGATTCATTTTTGGCTACAGAAACTCCAATCGCATTGCAAG
GTGTTTTGAACAACATCGGTCCAAACGGTGCTGATGTTGCTGGTGCATCTGCTGGTATTGTTGTTGCA
TCTCCATCTAGATCAGATCCAAACTACTTCTACTCTTGGACTAGAGATGCTGCATTGACTGCTAAGTA
TTTGGTTGATGCTTTTATTGCAGGTAATAAGGATTTGGAACAAACTATCCAACAATACATCTCTGCAC
AAGCTAAGGTTCAAACTATCTCAAACCCATCTGGTGACTTGTCTACAGGTGGTTTGGGTGAACCAAAG
TTTAATGTTAACGAAACTGCTTTTACAGGTCCATGGGGTAGACCACAAAGAGATGGTCCAGCATTGAG
AGCAACTGCTTTGATCGCATACGCTAACTACTTGATCGATAACGGTGAAGCTTCTACAGCAGATGAAA
TCATCTGGCCAATCGTTCAAAACGATTTGTCATACATCACTCAATACTGGAACTCTTCTACATTTGAT
TTGTGGGAAGAAGTTGAAGGTTCTTCTTTCTTTACTACAGCTGTTCAACATAGAGCATTAGTTGAGGG
TAATGCATTGGCTACTAGATTGAACCATACATGTTCAAACTGTGTTTCTCAAGCTCCACAAGTCTTGT
GTTTCTTGCAATCATATTGGACTGGTTCTTACGTTTTGGCTAATTTTGGTGGTTCAGGTAGATCAGGT
AAAGATGTTAATTCAATCTTGGGTTCTATTCATACTTTTGATCCAGCTGGTGGTTGTGATGATTCTAC
ATTTCAACCATGTTCAGCAAGAGCTTTGGCAAACCATAAGGTTGTTACTGATTCTTTTAGATCAATCT
ATGCTATTAATTCTGGTATTGCAGAAGGTTCAGCTGTTGCAGTTGGTAGATATCCAGAAGATGTTTAC
CAAGGTGGTAATCCATGGTACTTGGCTACTGCTGCAGCTGCAGAACAATTGTACGATGCAATCTATCA
ATGGAAGAAAATTGGTTCAATCTCTATCACAGATGTTTCTTTGCCATTTTTCCAAGATATCTATCCAT
CAGCTGCAGTTGGTACTTACAACTCAGGTTCTACTACTTTTAATGATATCATTTCTGCTGTTCAAACA
TATGGTGACGGTTACTTGTCAATCGTTGAAAAGTACACTCCATCAGATGGTTCTTTGACAGAACAATT
TTCTAGAACTGATGGTACACCATTGTCAGCTTCTGCATTAACTTGGTCATACGCTTCTTTGTTAACAG
CTTCAGCAAGAAGACAATCTGTTGTTCCAGCATCATGGGGTGAATCTTCAGCTTCTTCAGTTCCAGCA
GTTTGTTCAGCTACTTCTGCAACAGGTCCATATTCTACAGCTACTAATACAGTTTGGCCATCTTCAGG
TTCAGGTTCTTCAACTACAACTTCTTCAGCTCCATGTACAACTCCAACTTCTGTTGCAGTTACATTCG
ATGAAATCGTTTCAACTTCTTACGGTGAAACAATATATTTGGCTGGTTCTATTCCAGAATTGGGTAAT
TGGTCAACTGCTTCTGCAATTCCATTGAGAGCTGATGCATACACAAATTCTAATCCATTGTGGTATGT
TACTGTTAATTTGCCACCAGGTACATCATTCGAATACAAGTTTTTCAAGAATCAAACTGATGGTACAA
TTGTTTGGGAAGATGATCCAAATAGATCCTACACCGTTCCTGCTTACTGTGGTCAAACTACCGCAATC
TTGGACGACTCTTGGCAA

**Fig. 14**

**ATGACGCCTTTCGTCCTCACGGCCGTGCTGTTCTTGCTGGGGAATGCCGTGTTGGCC**TTGACCCCGGC
CGAATGGCGCAAACAATCTATCTACTTTCTCCTCACGGACCGCTTTGGCAGGGCAGATAACTCGACCA
CTGCTGCCTGCGATGTCACTGAGAGGATCTACTGTGGCGGGAGTTGGCAAGGAATCATCAACCATCTC
GACTATATCCAAGGCATGGGGTTCACGGCCATCTGGATTTCACCGGTGACCGAGCAGCTGCCGCAAAA
TACGGGTGAGGGAGAAGCCTATCATGGGTATTGGCAGCAGGAAATATACACGGTCAACTCCAACTTTG
GGACATCAGACGATCTCTTAGCCCTGTCAAAGGCGCTCCATGACCGTGGCATGTACCTCATGGTCGAT
GTGGTTGCGAATCACATGGGATACGATGGAGATGGCGACTCCGTTGATTACAGCGTCTTCAATCCATT
TAATTCCTC**T**AGTTATTTCCATCCCTATTGCCTGATTACAGACTACAGCAATCAGACCGATGTGGAAG
ACTGTTGGCTGGGCGATACGACTGTCTCGTTGCCCGATCTCAACACCACGGAGACTGTTGTGAGGACT
ATATGGTATGACTGGGTGGCGGATCTCGTCTCCAATTACTCTATTGATGGGCTTCGCATCGACACGGT
GAAACACGTAGAAAAGTCATTCTGGCCTGGTTACAACAGTGCTGCGGGTGTCTACTGTGTTGGCGAGG
TCCTCGATGGAGATCCGTCTTACACTTGTCCCTACCAGGATTATCTGGACGGTGTATTAAACTATCCA
ATATACTATCAACTACTGTATGCGTTTGAATCCTCTAGCGGCAGCATCAGCAATCTTTACAACATGAT
CAACTCTGTCGCCTCTGAATGTTCCGATCCCACTCTGTTGGGCAACTTTATCGAGAACCATGACAACC
CTAGATTTGCCTCCTATACAAGTGATTATTCTCTTGCTAAAAATGTGATTGCTTTCATCTTCTTCTCT
GACGGCATCCCTATCGTCTATGCCGGTCAGGAGCAGCATTACAACGGGGGAAATGACCCCTACAACCG
CGAGGCCACCTGGCTGTCAGGATACTCGACGACGGCCGAACTGTACACGTTCATTGCGACCACCAACG
CGATCCGTAGCTTGGCGATCTCCGTCGACTCGGAGTATTTGACGTACAAGAATGACCCATTCTACTAC
GACAGCAATACCCTCGCTATGCGCAAGGGTTCGGATGGCCTGCAGGTCATCACTGTTCTGTCCAATCT
GGGCGCCGATGGTAGCTCGTACACGTTGACTCTGAGTGGCAGTGGCTATTCGTCAGGCACGGAGCTGG
TGGAAGCTTACACCTGCACAACGGTCACTGTTGACTCTAATGGCGATATTCCAGTTCCCATGGAGTCC
GGACTGCCGCGCGTTTTCCTACCAGCATCCTCATTCAGTGGTAGCAGTCTATGCAGTTCTTCTCCTAG
CCCTACTACTACAACATCGACATCGACATCGACAACGTCGACGGCCTGCACCACCGCCACCGCTGTGG
CGGTCCTCTTCGAAGAGTTGGTGACAACGACCTACGGTGAAAATGTCTACCTCAGCGGATCGATCAGC
CAACTCGGGGACTGGAACACGGACGACGCCGTGGCCCTGTCCGCAGCTAATTACACTTCTTCGAATCC
CCTGTGGTATGTGACAGTCACATTGCCGGTTGGGACGTCCTTTGAGTACAAGTTCATCAAGAAGGAAG
AGAACGGCGATGTCGAGTGGGAGAGCGATCCCAATCGGTCGTATACTGTGCCGACGGCCTGCACGGGA
GCGACGGAGACGATTGTCGACACATGGAGATAG

**Fig. 15**

ATGGCGTCCCTCGTTGCTGGCGCTCTCTGCATCCTGGGCCTGACGCCTGCTGCATTTGCACGAGCGCC
CGTTGCAGCGCGAGCCACCGGTTCCCTGGACTCCTTTCTCGCAACCGAAACTCCAATTGCCCTCCAAG
GCGTGCTGAACAACATCGGGCCCAATGGTGCTGATGTGGCAGGAGCAAGCGCCGGCATTGTGGTTGCC
AGTCCGAGCAGGAGCGACCCAAATTATTTCTACTCCTGGACACGTGACGCAGCGCTCACGGCCAAATA
CCTCGT**T**GACGCCTTCATCGCGGGCAACAAGGACCTAGAGCAGACCATCCAGCAGTACATCAGCGCGC
AGGCGAAGGTGCAAACTATCTCCAATCCGTCCGGAGATTTATCCACCGGTGGCTTAGGTGAGCCCAAG
TTCAATGTGAATGAGACGGCTTTTACCGGGCCCTGGGGTCGTCCACAGAGGGACGGACCAGCGTTGAG
AGCGACGGCCCTCATTGCGTATGCGAACTATCTCATCGACAACGGCGAGGCTTCGACTGCCGATGAGA
TCATCTGGCCGATTGTCCAGAATGATCTGTCCTACATCACCCAATACTGGAACTCATCCACCTTCGAC
CTCTGGGAAGAAGTAGAAGG**T**TCCTCATTCTTCACAACCGCCGTGCAACACCGCGCCCTGGTCGAAGG
CAATGCACTGGCAACAAGGCTGAACCACACGTGCTCCAACTGCGTCTCTCAGGCCCCTCAGGTCCTGT
GTTTCCTGCAGTCATACTGGACCGGATCGTATGTTCTGGCCAACTTTGGTGGCAGCGGTCGTTCCGGC
AAGGACGTGAA**C**TCGATTCTGGGCAGCATCCACACCTTTGATCCCGCCGGAGGCTGTGACGACTCGAC
CTTCCAGCCGTGTTCGGCCCGTGCCTTGGCAAATCACAAGGTGGTCACCGACTCGTTCCGGAGTATCT
ATGCGATCAACTCAGGCATCGCAGAGGGATCTGCCGTGGCAGTCGGCCGCTACCCTGAGGATGTCTAC
CAGGGCGGGAACCCCTGGTACCTGGCCACAGCAGCGGCTGCAGAGCAGCTTTACGACGCCATCTACCA
GTGGAAGAAGATCGGCTCGATAAGTATCACGGACGTTAGTCTGCCATTTTTCCAGGATATCTACCCTT
CTGCCGCGGTGGGCACCTATAACTCTGGCTCCACGACTTTCAACGACATCATCTCGGCCGTCCAGACG
TATGGTGATGGATATCTGAGTATTGTCGAGAAATATACTCCCTCAGACGGCTCTCTTACCGAACAATT
CTCCCGTACAGACGGCACTCCGCTTTCTGCCTCTGCCCTGACTTGGTCGTACGCTTCTCTCCTAACCG
CTTCGGCCCGCAGACAGTCCGTCGTCCCTGCTTCCTGGGGCGAAAGCTCCGCAAGCAGCGTCCCTGCC
GTCTGCTCTGCCACCTCTGCCACGGGCCCATACAGCACGGCTACCAACACCGTCTGGCCAAGCTCTGG
CTCTGGCAGCTCAACAACCACCAGTAGCGCCCCATGCACCACTCCTACCTCTGTGGCTGTGACCTTCG
ACGAAATCGTCAGCACCAGTTACGGGGAGACAATCTACCTGGCCGGCTCGATCCCCGAGCTGGGCAAC
TGGTCCACGGCCAGCGCGATCCCCCTCCGCGCGGATGCTTACACCAACAGCAACCCGCTCTGGTACGT
GACCGTCAATCTGCCCCCTGGCACCAGCTTCGAGTACAAGTTCTTCAAGAACCAGACGGACGGGACCA
TCGTCTGGGAAGACGACCCGAACCGGTCGTACACGGTCCCAGCGTACTGTGGGCAGACTACCGCCATT
CTTGACGATAGTTGGCAGTGA

**Fig. 16**

```
   1 atgacgcctt tcgtcctcac ggccgtgctg ttcttgctgg ggaatgccgt gttggccttg
  61 accccggccg aatggcgcaa acaatctatc tactttctcc tcacggaccg ctttggcagg
 121 gcagataact cgaccactgc tgcctgcgat gtcactgaga ggatctactg tggcgggagt
 181 tggcaaggaa tcatcaacca tctcgactat atccaaggca tggggttcac ggccatctgg
 241 atttcaccgg tgaccgagca gctgccgcaa aatacgggtg agggagaagc ctatcatggg
 301 tattggcagc aggaaatata cacggtcaac tccaactttg ggacatcaga cgatctctta
 361 gccctgtcaa aggcgctcca tgaccgtggc atgtacctca tggtcgatgt ggttgcgaat
 421 cacatgggat acgatggaga tggcgactcc gttgattaca gcgtcttcaa tccatttaat
 481 tcctcgagtt atttccatcc ctattgcctg attacagact acagcaatca gaccgatgtg
 541 gaagactgtt ggctgggcga tacgactgtc tcgttgcccg atctcaacac cacggagact
 601 gttgtgagga ctatatggta tgactgggtg gcggatctcg tctccaatta ctctattgat
 661 gggcttcgca tcgacacggt gaaacacgta gaaaagtcat tctggcctgg ttacaacagt
 721 gctgcgggtg tctactgtgt tggcgaggtc ctcgatggag atccgtctta cacttgtccc
 781 taccaggatt atctggacgg tgtattaaac tatccaatat actatcaact actgtatgcg
 841 tttgaatcct ctagcggcag catcagcaat ctttacaaca tgatcaactc tgtcgcctct
 901 gaatgttccg atcccactct gttgggcaac tttatcgaga accatgacaa ccctagattt
 961 gcctcctata caagtgatta ttctcttgct aaaaatgtga ttgctttcat cttcttctct
1021 gacggcatcc ctatcgtcta tgccggtcag gagcagcatt acaacggggg aaatgacccc
1081 tacaaccgcg aggccacctg gctgtcagga tactcgacga cggccgaact gtacacgttc
1141 attgcgacca ccaacgcgat ccgtagcttg gcgatctccg tcgactcgga gtatttgacg
1201 tacaagaatg acccattcta ctacgacagc aataccctcg ctatgcgcaa gggttcggat
1261 ggcctgcagg tcatcactgt tctgtccaat ctgggcgccg atggtagctc gtacacgttg
1321 actctgagtg gcagtggcta ttcgtcaggc acggagctgg tggaagctta cacctgcaca
1381 acggtcactg ttgactctaa tggcgatatt ccagttccca tggagtccgg actgccgcgc
1441 gttttcctac cagcatcctc attcagtggt agcagtctat gcagttcttc tcctagccct
1501 actactacaa catcgacatc gacatcgaca acgtcgacgg cctgcaccac cgccaccgct
1561 gtggcggtcc tcttcgaaga gttggtgaca acgacctacg gtgaaaatgt ctacctcagc
1621 ggatcgatca gccaactcgg ggactggaac acggacgacg ccgtggccct gtccgcagct
1681 aattacactt cttcgaatcc cctgtggtat gtgacagtca cattgccggt tgggacgtcc
1741 tttgagtaca agttcatcaa gaaggaagag aacggcgatg tcgagtggga gagcgatccc
1801 aatcggtcgt atactgtgcc gacggcctgc acgggagcga cggagacgat tgtcgacaca
1861 tggagatag
```

**Fig. 17**

1acgagatgtgtatatactgtgaaccaaactagatgatgtcagttatgctggtctgagaac
61tcatagaagcccttgaaaataccccaagctagcactccaaccctaactctgttgctctac
121tagatcaagacgagtactctgattgagctgcaggcttggaatatatgattagcagaaaaa
181gggttaaaacttgtatgacaatcagtttgtcagtactccgtagtgatgccatgtctatag
241agtcgacactaaggcagcatgtgaatgagtcggaaatgacaggaagcagattccttaaca
301gtcatgttctccgtgcctgcatccccacgtcacctgcaaagatgcgacgctactccacac
361cggcgccttgatgtctgctgttcctggcctagtggagccccatgcgctgctagctcgtgg
421tcttcgaataaatcagaataaaaaacggagtaattaattgcgcccgcaacaaactaagca
481atgtaactcaatgccaagcttccgctgatgctcttgacatctccgtagtggcttctttcg
541taatttcagacgtatatatagtagtaatgcccagcaggccgggataatgatggggatttc
601tgaactctcagcttccgtacgctgaacagtttgcttgcgttgtcaaccatggcgtccctc
661gttgctggcgctctctgcatcctgggcctgacgcctgctgcatttgcacgagcgcccgtt
721gcagcgcgagccaccggttccctggactcctttctcgcaaccgaaactccaattgccctc
781caaggcgtgctgaacaacatcgggcccaatggtgctgatgtggcaggagcaagcgccggc
841attgtggttgccagtccgagcaggagcgacccaaattgtaggttctttcccaccagaaat
901tacttatttaaatcagccctctgacaggttgaagatttctactcctggacacgtgacgca
961gcgctcacggccaaatacctcgtcgacgccttcatcgcgggcaacaaggacctagagcag
1021accatccagcagtacatcagcgcgcaggcgaaggtgcaaactatctccaatccgtccgga
1081gatttatccaccggtggcttaggtgagcccaagttcaatgtgaatgagacggcttttacc
1141gggccctggggtcgtccacagagggacggaccagcgttgagagcgacggccctcattgcg
1201tatgcgaactatctcatcgtaagcttctgctcgctgcccttctctctgctcgtatgctaa
1261gtagtcctgtcaggacaacggcgaggcttcgactgccgatgagatcatctggccgattgt
1321ccagaatgatctgtcctacatcacccaatactggaactcatccaccttcggtaggcaaat
1381gaatattcccgacacagcgtggtactaatttgattcagacctctgggaagaagtagaagg
1441atcctcattcttcacaaccgccgtgcaacaccgcgccctggtcgaaggcaatgcactggc
1501aacaaggctgaaccacacgtgctccaactgcgtctctcaggcccctcaggtcctgtgttt
1561cctgcagtcatactggaccggatcgtatgttctggccaactttggtggcagcggtcgttc
1621cggcaaggacgtgaattcgattctgggcagcatccacacctttgatcccgccggaggctg
1681tgacgactcgaccttccagccgtgttcggcccgtgccttggcaaatcacaaggtggtcac
1741cgactcgttccggagtatctatgcgatcaactcaggcatcgcagagggatctgccgtggc
1801agtcggccgctaccctgaggatgtctaccagggcgggaacccctggtacctggccacagc
1861agcggctgcagagcagctttacgacgccatctaccagtggaagaagatcggctcgataag
1921tatcacggacgttagtctgccatttttccaggatatctacccttctgccgcggtgggcac
1981ctataactctggctccacgactttcaacgacatcatctcggccgtccagacgtatggtga
2041tggatatctgagtattgtcgtacgttttgccttagattctcaggtgtaaagaaaaaaatg
2101gaactaactcagttctaggagaaatatactccctcagacggctctcttaccgaacaattc
2161tcccgtacagacggcactccgctttctgcctctgccctgacttggtcgtacgcttctctc
2221ctaaccgcttcggcccgcagacagtccgtcgtccctgcttcctggggcgaaagctccgca
2281agcagcgtccctgccgtctgctctgccacctctgccacgggcccatacagcacggctacc
2341aacaccgtctggccaagctctggctctggcagctcaacaaccaccagtagcgccccatgc
2401accactcctacctctgtggctgtgaccttcgacgaaatcgtcagcaccagttacggggag
2461acaatctacctggccggctcgatccccgagctgggcaactggtccacggccagcgcgatc
2521cccctccgcgcggatgcttacaccaacagcaacccgctctggtacgtgaccgtcaatctg
2581ccccctggcaccagcttcgagtacaagttcttcaagaaccagacggacgggaccatcgtc
2641tgggaagacgacccgaaccggtcgtacacggtcccagcgtactgtgggcagactaccgcc
2701attcttgacgatagttggcagtgagataacatccacccttctgtttta

**Fig. 18**