Engineering the Substrate Specificity of Galactose Oxidase

Lucy Ann Chappell

BSc. (Hons)

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Molecular and Cellular Biology

September 2013

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

My greatest thanks go to Prof. Michael McPherson for the support, patience, expertise and time he has given me throughout my PhD. Many thanks to Dr Bruce Turnbull for always finding time to help me with the Chemistry side of the project, particularly the NMR work. Also thanks to Dr Arwen Pearson for her support, especially reading and correcting my final thesis.

Completion of the laboratory work would not have been possible without the help of numerous lab members and friends who have taught me so much, helped develop my scientific skills and provided much needed emotional support – thank you for your patience and friendship: Dr Thembi Gaule, Dr Sarah Deacon, Dr Christian Tiede, Dr Saskia Bakker, Ms Briony Yorke and Mrs Janice Robottom. Thank you to three students I supervised during my PhD for their friendship and contributions to the project: Mr Ciaran Doherty, Ms Lito Paraskevopoulou and Ms Upasana Mandal. Thanks also to Mr Daniel Williamson for help with the practical side of the NMR work.

I am very grateful to many staff members within the Astbury Centre and the Faculty of Biological Sciences for help during my time here, especially Dr Iain Manfield for many fruitful discussions, Dr Gareth Howell for help with imaging software and the ever friendly Stores staff. I greatly acknowledge the financial support from The Wellcome Trust.

Finally I offer my deepest gratitude to Dr Kevin Colgan, my future husband, for his love and support, for always being there for me and for keeping me in the real world.

Abstract

Biocatalysis, the use of enzymes to catalyse the generation of specific chemicals, has a number of advantages including high specificity, lower energy requirements and greater sustainability over an equivalent chemical process. Several methods exist for optimisation of enzymes for use in industrial processes, including introduction of mutations to generate libraries of variants which are then screened for the desired properties, such as stability or substrate specificity. Galactose oxidase catalyses the oxidation of primary alcohols to the corresponding aldehyde, with concomitant reduction of dioxygen to hydrogen peroxide. It has already been developed for use in a range of biotechnological processes and is an ideal candidate for further development due to features including high stability, a surface exposed active site displaying broad substrate specificity, and an autocatalytically-generated cofactor.

Research presented in this thesis investigates the effect on activity towards a range of alternative substrates of mutations at selected active site residues with the aim of expanding the biotechnological potential of galactose oxidase. Libraries of variants were designed and generated using high quality oligonucleotides constructed using trimer phosphoramidites. Screening assays used by other groups were optimised by varying different components. These assays were then used to identify a number of variants displaying enhanced activity towards D-arabinose, D-glucose, D-xylose or glycerol. A selection of these variants were then further characterised in order to understand the biochemical basis of the altered activities and determine some of the conditions required for potential industrial application of the variants. The most exciting results include identification of a variant displaying higher levels of activity towards glycerol than towards the native substrate D-galactose; determination of the position of oxidation of D-arabinose at the C-4 hydroxyl; and the observation that mutation of Phe194 significantly affects binding of D-glucose in the active site.

Table of Contents

x

List of Figures

List of Tables

Abbreviations

4-CN	4-chloronaphthol
ABTS	2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt
ACES	N-(2-acetamido)-2-aminoethanesulfonic acid
APS	Ammonium persulfate
ATP	Adenosine 5'-triphosphate
A_x	Absorbance at x nm
BLAST	Basic local alignment search tool
CAST	Combinatorial active-site saturation test
CFU	Colony forming units
CHES	3-(cyclohexylamino) ethanesulfonic acid
COSY	Correlation spectroscopy
CPEC	Circular polymerase extension cloning
CV	Column volume
DAB	3,3'-diaminobenzidine
ECAO	*Escherichia coli* amine oxidase
EDTA	Ethylenediaminetetraacetic acid
epPCR	Error prone polymerase chain reaction
FACS	Fluorescence activated cell sorting
GC	Gas chromatography
GO	Galactose oxidase
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
IPTG	Isopropyl-β-D-thiogalactopyranoside
ISM	Iterative saturation mutagenesis
LMCT	Ligand to metal charge transfer
MES	2-(N-morpholino) ethanesulfonic acid
NAD(H)	Nicotinamide adenine dinucleotide (reduced)
NEB	New England biolabs
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at 600 nm
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction

PEG	Polyethylene glycol
RBS	Ribosome binding site
RuBisCo	Ribulose-1,5-bisphosphate carboxylase
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SS	Secretion signal
StEP	Staggered extension process
TAE	Tris-acetate ethylenediaminetetraacetic acid
TEMED	N,N,N',N',-tetramethylethylenediamine

Chapter 1 : Introduction

1.1 Exploiting biology for industrial processes

The use of biology in industrial processes is not a new concept; yeast fermentation in beer and bread production and the use of various microorganisms in cheese manufacture has been documented for thousands of years. However, within the last century, significant advances have been made in the development of biotechnology. The discovery of a fungus which produces the antibiotic penicillin in 1928 and its development into a viable therapeutic resulted in the award of a Nobel Prize to Fleming, Chain and Florey in 1945. Meanwhile the use of mice, and later rabbits, in diagnosis of pregnancy by the Aschheim-Zondek test (Zondek and Aschheim, 1927) was used until the development of tests based on antibodies towards human chorionic gonadotropin in the 1960s.

The use of a variety of organisms in removal of environmental pollutants is known as bioremediation. While still an emerging area, bioremediation has great potential as a simple, cheap and environmentally safe option for removing contaminants such as petroleum hydrocarbons, halogenated solvents from industry, heavy metals, explosives and agricultural chemicals from the environment (Harms et al., 2011). Bioremediation processes can use bacteria, fungi (Harms et al., 2011) or even plants (Meagher, 2000) to either sequester the pollutant away from the surrounding environment or convert it to a less toxic species. Brim et al. (2000), engineered a bacterium for metal remediation in environments contaminated with radioactive material, for example as a result of nuclear weapons production. The *mer* operon, which encodes six proteins conferring mercury resistance, was cloned into the radiation resistant bacterium *Deinococcus radiodurans* and the bacterium was shown to reduce ionic Hg (II) to the less toxic Hg (O) while resisting high radiation levels as well as high mercury concentrations (Brim et al., 2000).

Metal and metal oxide nanoparticles have various functions including as antimicrobial agents (silver), in sunscreen (zinc oxide), or in electronic data storage (magnetic materials). Conventional methods for nanoparticle generation often use harsh methods such as thermal decomposition of organometallic compounds in high-boiling organic solvents. These consume large quantities of energy and result in significant waste product generation (Lu et al., 2007). Magnetic nanoparticles are naturally generated by magnetic bacteria to allow the organisms to align with external magnetic fields. Work is underway to optimise generation of these nanoparticles which often display better size consistency and biocompatibility than the

2

inorganically synthesised equivalents and can be generated under mild conditions with low environmental and economic cost (Galloway and Staniland, 2012).

1.2 Biocatalysis

Biocatalysis exploits the chemical reactions carried out by enzymes and microbes in order to synthesise chemicals for various industries including pharmaceuticals, food and agriculture as well as in the manufacture of specialist and commodity chemicals. Use of biological systems has a number of advantages over traditional chemical synthesis methods:

i.   High levels of specificity which means that starting products present in a mixture can be used to generate the final product. This feature is also exploited in the use of enzymes for diagnostic purposes, where the amount of a specific component within a mixture such as blood can be determined. The high selectivity, including chemoselectivity, stereoselectivity and regioselectivity (where a specific part of the substrate is modified), also generally results in low generation of unwanted by-products, such as different isomeric forms of the desired product, thus reducing waste and improving yields - an important consideration in large scale processes.

ii. Biocatalysts (enzymes and microbes) are biodegradable and the reactions catalysed are generally carried out under much milder conditions than the equivalent chemical process: neutral pH, aqueous solution, atmospheric pressures and ambient temperatures. These conditions result in lower energy and economic demands, increased safety and also reduce the impact on the local environment in terms of disposal of reaction solutions such as solvents or waste products. Where the process requires different conditions, such as the use of solvent for product solubility, enzymes can be genetically modified as detailed below.

iii. The use of enzymes and biological systems can be considered a sustainable resource; as a result the process is often cheaper than the use of chemical catalysts, such as rare earth metals which are generally expensive due to their rarity and complexity of mining. For example vanadium (V) oxide, used to catalyse sulfuric acid production at 400-620 °C, has to be purified from the extremely rare mineral shcherbinaite (Hughes and Finger, 1983, Thompson, 1931).

The use of microbes in biocatalysis exploits pathways within organisms which have evolved to efficiently generate complex metabolites. Metabolic engineering is generally carried out to

optimise production of the chemical via a pathway of enzymes within the host organism, such as *Escherichia coli* or *Saccharomyces cerevisiae*, or to combine different aspects of pathways from different organisms in order to generate a compound not created in nature. The product may then be secreted from the host in order to avoid toxicity effects and ease purification. Successful examples of metabolic engineering include the generation of potential biofuel components from microbial sources such as isopropanol (Hanai et al., 2007, Jojima et al., 2008) and 1-butanol (Atsumi et al., 2008, Inui et al., 2008) from *E. coli*. 1,3-propanediol, an important component in the polymer and cosmetic industries, is now exclusively produced by microorganisms either from glycerol, a by-product of bio-diesel production (Saxena et al., 2009), or from D-glucose contained in corn syrup in a process patented by DuPont, Tate & Lyle and Genencor International (Nakamura and Whited, 2003). A relatively famous example is that of artemisinic acid production by *S. cerevisiae*. Artemisinin is highly effective in treatment of malaria but it is difficult to purify large quantities from the host plant *Artemisia annua* resulting in a high cost of the drug. Metabolic engineering to enable efficient generation of the precursor artemisinic acid in *S. cerevisiae* involved optimisation of the farnesyl pyrophosphate biosynthetic pathway and introduction of genes for amorphadiene synthase and a novel cytochrome P450 from *A. annua* (Ro et al., 2006). This has opened the door to reducing the cost of the drug with potentially significant effects on global health.

The use of isolated enzymes in generation of chemicals is generally simpler than the enzyme pathways which need to be considered in metabolic engineering projects and there are a greater number of success stories. Early examples include the use of cellulases and proteases in laundry detergent (Estell et al., 1985) and the production of acrylamide by nitrile hydratase (Nagasawa and Yamada, 1989). Some enzymes exist which require little optimisation for use in a particular process, for example Sethi et al. (2013) identified an enantioselective ketoreductase which can be used to generate an intermediate for the pharmaceutical ingredient rivastigmine with >99% enantiomeric excess (Sethi et al., 2013). Extremophilic organisms provide a good source of robust enzymes which are often more resilient to the conditions required in industrial processes, such as high temperatures, extremes of pH and the presence of organic solvents. For example, an aldo-keto reductase from the thermophilic organism *Thermotoga maritima* is stable up to 80 °C and retains 70% activity in solutions containing 20% (v/ v) isopropanol or DMSO (Willies et al., 2010). However, where suitable enzymes are not available in nature, major developments in bioscience over the last 30 years

have led to significant advances in the optimisation of enzymes as discussed in the next section.

1.3 Optimisation of enzymes for biocatalysis

Numerous factors must be considered in development of enzymes for use in different applications as detailed below.

1.3.1 Immobilisation

In order to improve the stability of the enzymes, as well as facilitate their reuse in the industrial process, technology to immobilise enzymes was developed. Immobilisation of enzymes, generally by adsorption, entrapment or cross-linking (Singh et al., 2013) has a number of advantages. These include the enhancement of activity, as seen in immobilisation of glucose oxidase (Hashemifard et al., 2010, Pandey et al., 2007); improvement of thermostability, as seen with lipase from *Candida rugosa* (Matsumoto and Ohashi, 2003); and improved solvent stability, as seen with laccase activity in diethyl ether or ethyl acetate (Ruiz et al., 2000). Depending on the method of immobilisation, substrate selectivity may also be affected (Singh et al., 2013).

1.3.2 Traditional directed evolution

Since the 1980s, enzyme characteristics such as catalytic activity, thermostability, solvent resistance and substrate specificity have been targeted for improvement by the development of protein engineering technologies. The traditional method for improving enzyme activity is directed evolution whereby mutations are introduced at random throughout the DNA sequence using iterative cycles of error prone Polymerase Chain Reaction (epPCR) (You and Arnold, 1996). Variants displaying desirable changes in activity are identified at each stage by screening or selection methods (Section 1.4) and subjected to further rounds of mutagenesis. epPCR protocols contain modifications of the standard PCR protocol resulting in enhancement of the natural error rate of the polymerase, often *Taq* polymerase which already displays a naturally high error rate. These modifications include increased $MgCl_2$ concentrations, addition of $MnCl_2$ at different concentrations, alterations in the ratios of nucleotides in the reaction or inclusion of nucleotide analogues such as dITP (Cirino et al., 2003). Other methods of introducing mutations randomly across the gene of interest include the use of mutator strains

of *E. coli* such as XL1 Red (Agilent), chemical or physical mutagens such as UV light, or more complex methods such as random insertion/ deletion mutagenesis which simultaneously deletes part of the sequence at random and inserts a random sequence at the same place (Murakami et al., 2002).

1.3.3 The need for smaller libraries

The majority of changes in enzyme properties require mutation of more than one residue in combination to generate a library of variants. If two substitutions are to be made in a 200-residue protein, there are over 7×10^6 possibilities while three substitutions results in over 9×10^9 possibilities (Bornscheuer et al., 2012). In order to statistically screen all possible variants, or even a high proportion of variants, ultra high-throughput screening is required and this is not possible for many enzyme characteristics (Section 1.4). Introduction of mutations by epPCR can also lead to a significant bias towards the introduction of certain residues in a sequence for three distinct reasons (Neylon, 2004). The first is that the polymerase used can display a preference for AT to GC substitutions or *vice versa* and the position of the mutation can also be affected by surrounding sequence, meaning randomisation to every amino acid is not equally likely at each position. The second source of bias comes from the genetic code which displays a significant codon bias, for example there are six codons for arginine but only one for tryptophan. Mutation of a single nucleotide of the valine codon GTT can form a codon for Phe, Leu, Ile, Ala, Asp or Gly, however two substitutions are required for most other residues and three changes are necessary to form codons for Gln, Trp or Lys. The final source of bias is introduced via the mutagenesis protocol due to the exponential nature of PCR - mutations introduced early in the amplification process will be over-represented in the final library.

1.3.4 Reducing the bias problem

Bias introduced by the polymerase used in epPCR can be reduced by use of an engineered polymerase, such as Mutazyme (Stratagene), in combination with *Taq* polymerase (van Hercke et al., 2005). An alternative method of introducing mutations across the sequence without the bias of polymerases was developed by the Schwaneberg group (Wong et al., 2004). Sequence saturation mutagenesis (SeSaM) introduces a universal base such as inosine at random positions and then replaces this base with random nucleotides.

Site-directed mutagenesis, whereby mutation(s) are introduced at a specific position using oligonucleotides containing the mutated codon (Figure 1.1) was originally developed by Michael Smith, for which he shared the 1993 Nobel Prize in Chemistry. This method has proved invaluable in structure-function studies of proteins as well as in mutant library generation.

A	B	C	D	E
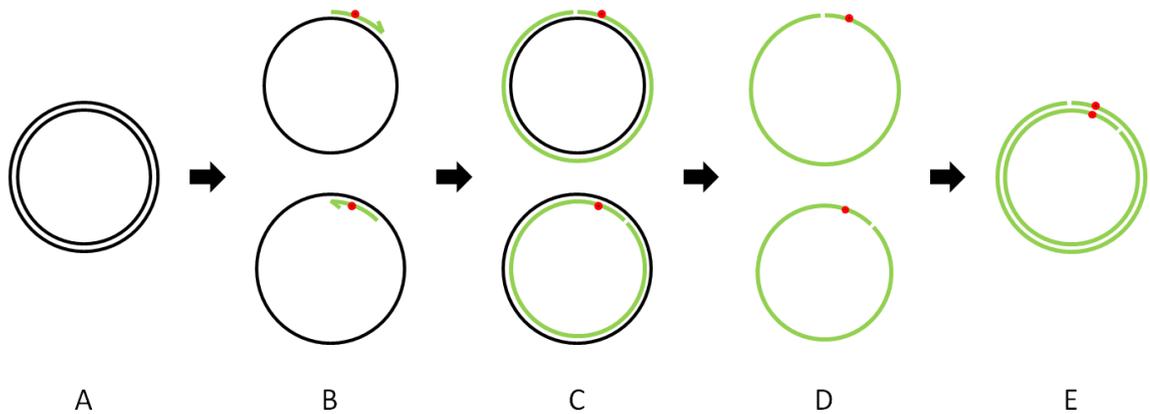
Figure 1.1. Schematic illustrating the Quikchange approach to site-directed mutagenesis

A plasmid containing the coding region of interest (**A**) is denatured and oligonucleotide primers, both containing the desired mutation (red), are annealed (**B**). A heat-stable DNA polymerase is used to extend and incorporate the mutagenic primers (**C**) to generate two circular strands which contain nicks (green). *Dpn*I digestion is then carried out to remove the parental DNA (black) (**D**) and the circular strands containing the mutation in both strands are transformed into *E. coli* (**E**).

Saturation mutagenesis, whereby particular residue(s) are mutated to each of the other 19 amino acids, is a well used technique in library construction. Oligonucleotides are generated as for site-directed mutagenesis but, instead of a specific codon being introduced at the position of interest, a mixture of nucleotides is provided during oligonucleotide synthesis. If the full mixture of four nucleotides is provided for all three positions the codon is written as NNN (where N = any nucleotide). As before, generation of libraries by this method will lead to significant codon bias and also results in a 3/ 64 chance of stop codon incorporation. It is therefore more common to limit the mixture of nucleotides at the third position by using NNK or NNS codons (where K = T or G and S = G or C). These degenerate codons encode all 20 amino acids but with only 32 codons, resulting in reduced over-representation of the most common residues. The probability of stop codon incorporation is also reduced to 1/ 32. Unfortunately, there is still some level of codon bias in the NNK/ S libraries which caused Reetz and co-workers to develop the NDT codon (where D = A, G or T) (Reetz et al., 2008). This degenerate codon encodes 12 amino acids using 12 codons with no stop codons: Phe, Leu, Ile,

Val, Tyr, His, Asn, Asp, Cys, Arg, Ser and Gly. The 12 residues include examples of all the different amino acid characteristics and exclude most cases of structural similarity. The NDT codon was shown to generate a much higher quality library (measured by frequency of positive variants and magnitude of improvement) than the equivalent NNK library in mutagenesis of the epoxide hydrolase from *Aspergillus niger* (Reetz et al., 2008). Subsequent development in the Reetz group has led to a further reduction in library redundancy by 'the 22-codon trick' which uses a combination of the degenerate codons NDT, VHG (where V = A, C or G and H = A, T or C) and TGG resulting in 22 possible codons encoding all 20 amino acids, no stop codons, and only Val and Leu encoded by more than one codon (Kille et al., 2013).

A method for completely removing the codon bias problem involves the use of trimer phosphoramidites in oligonucleotide synthesis. Instead of adding one nucleotide at a time, the oligonucleotide is constructed one codon at a time. This means that no stop codons are incorporated, the codons can be optimised for expression in the host organism (*i.e. E. coli*), and multiple randomised codons can be included in one oligonucleotide. By this method, it is possible to exquisitely control the mutation at the desired position(s) *i.e.* the proportion of WT to mutated sequence can be varied as well as inclusion of only a subset of codons if desired. Each residue is equally likely to appear at the randomised position. Despite a number of available protocols for production of randomised oligonucleotides by this method (Gaytan et al., 1998, Janczyk et al., 2012, Kayushin et al., 2000, Yagodkin et al., 2007), the complexity and relatively high cost of the process limits their use in mutagenesis projects. In laboratories without access to DNA synthesising equipment, oligonucleotides generated using trimer phosphoramidites can be sourced from companies such as Ella Biotech GmbH.

Another similar method, called the MAX method, which only requires simple primers was developed by the Hine group (Hughes et al., 2003) (Figure 1.2). A template oligonucleotide is generated covering the region to be mutated and the codons targeted for randomisation are replaced with the NNN codon. A set of 20 primers for each position targeted are synthesised, each containing a codon for a different residue optimised for expression in *E. coli* (termed MAX codons). These primers are then hybridised to the template containing the NNN codons and ligated together. PCR is then carried out to replace the template strand, resulting in a double stranded fragment containing mutations at the desired positions ready for cloning. Unfortunately this method cannot be used to randomise adjacent codons independently and, depending on the number of codons to be randomised, may require generation of a very large number of primers.
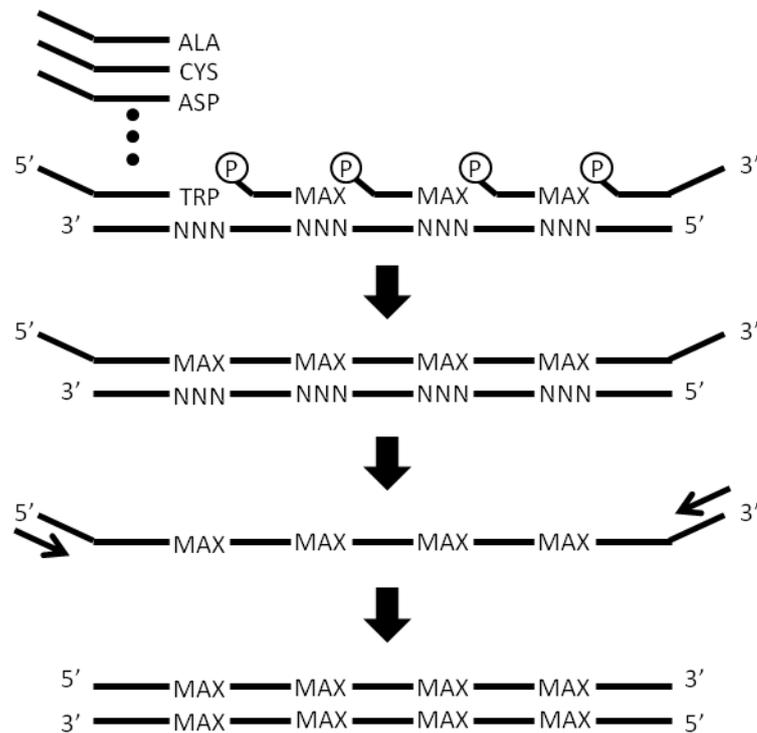
Figure 1.2. Schematic illustrating the MAX method of library generation

Copied from Neylon (2004). A description of the method is given in the text. MAX codons for Ala, Cys, Asp and Trp are shown as an example.

1.3.5 Combining beneficial mutations

A range of methods exist for combining selected point mutations in random combinations. These same methods are also used in generating libraries by recombination of similar genes from different sources. The most popular and straightforward recombination technique is DNA shuffling (Figure 1.3, left) which was pioneered by Willem P. C. Stemmer in the 1990s (Stemmer, 1994a, Stemmer, 1994b). The technique involves deoxyribonuclease I digestion of the DNA sequences to be combined to generate random fragments. The fragments are then size fractionated, mixed and repeated cycles of melting, annealing and extension carried out to reassemble the fragments. Recombination occurs at regions of high sequence identity causing fragments from different parents to anneal. PCR amplification of the recombined sequence is then carried out followed by cloning.
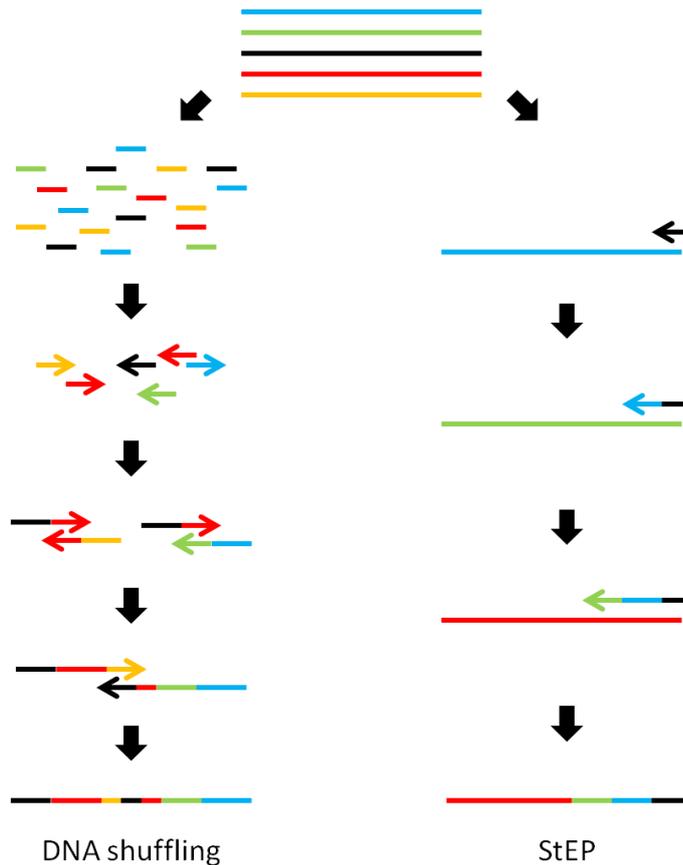
Figure 1.3. Schematic illustration of DNA shuffling (left) and StEP (right)

Copied from Neylon (2004). A description of each method is given in the text.

A related method developed in the lab of Frances Arnold (Zhao et al., 1998) is termed the Staggered Extension Process (StEP) (Figure 1.3, right). Here, gene fragments are extended by repeated cycles of melting, annealing and extension as before, but the duration and temperature of the extension steps results in only partial elongation of the growing strand before melting and annealing to a different strand with potentially different mutation(s). Libraries generated by both DNA shuffling and StEP are prone to bias as different mutations will recombine more easily than others based on their position within the sequence (Maheshri and Schaffer, 2003, Moore et al., 2001, Moore and Maranas, 2002). A number of variations have therefore been developed including Random Chimeragenesis on Transient Templates (RaChiTT) which is designed to enhance the number of cross-overs; and the three similar methods Degenerate Homoduplex Recombination (Coco et al., 2002), Synthetic Shuffling (Ness et al., 2002) and Assembly of Designed Oligonucleotides (Zha et al., 2003) which remove the potential for bias in the generated library. A number of methods also exist for generation of

10

libraries by homology-independent recombination, including Incremental Truncation for the Creation of Hybrid enzymes (ITCHy) (Ostermeier et al., 1999) and Sequence Homology-Independent Protein Recombination (SHIPRec) (Sieber et al., 2001).

1.3.6 Design of mutagenesis strategies

As discussed by Manfred Reetz (Reetz et al., 2008), there are a number of factors to consider when designing a mutagenesis strategy. The size of the library is often dictated by the screening or selection method which will be used to identify clones showing the desired activity (Section 1.4). Changes in certain enzyme characteristics such as enantioselectivity and regioselectivity can often only be detected by relatively low-throughput techniques such as gas chromatography (GC) and high performance liquid chromatography (HPLC) and so library sizes must be kept small. The consideration of the 'completeness' of library screening is also important. Work by different groups (Bosley and Ostermeier, 2005, Patrick et al., 2003, Patrick and Firth, 2005) has led to the development of algorithms which can be used to estimate the number of variants that need to be screened in order to statistically sample a given percentage of the possible variants. Regardless of codon usage or mutagenesis method, it is necessary to screen at least a three-fold excess of transformants in order to statistically cover 95% of the library (Reetz et al., 2008).

While it is possible to apply saturation mutagenesis to every residue in an enzyme individually (Funke et al., 2003), even for small proteins this still requires a large amount of time and effort and is perhaps not the most sensible use of resources. The rapidly increasing amount of information from sequencing, structural and biochemical studies can also aid in the design of so called 'smart' libraries where a selection of specific residues are chosen to be mutated. The residues they will be mutated to is also often limited based on what is considered most likely to have a positive effect on the characteristic to be altered (Reetz and Carballeira, 2007). Alignments of protein sequences of enzymes from the same family can be used to identify the residues present at a given position in the common ancestor which is likely to display greater thermostability (Miyazaki et al., 2001). An example of such an approach is 3-isopropylmalate dehydrogenase where six mutants showed greater thermal stability than the original enzyme after substitution with an ancestral amino acid (Watanabe et al., 2006). Sequence alignments of enzymes within a family have also been used by other groups to identify the residues permitted at specific positions, resulting in generation of higher quality libraries *i.e.* more

variants displaying improved thermostability (Amin et al., 2004), activity or enantioselectivity (Reetz and Wu, 2008).

The use of alignments in library design has also been used with structural data. In work presented by Jochens and Bornscheuer (2010) the amino acids present at four positions of interest in *Pseudomonas fluorescens* esterase were determined in 1751 structurally related enzymes identified using the 3DM software (Kuipers et al., 2010). Libraries were designed whereby the residues at each of the four positions were only randomised to those which appeared frequently at structurally equivalent positions. The resulting library contained significantly more variants showing improved activity and/ or enantioselectivity than control libraries which either included all 20 residues, or only those occurring rarely at the selected positions. Other structural analyses have also been used in library design, such as the B-FIT test (Reetz et al., 2006a) which identifies amino acids based on their flexibility in the crystal structure (higher flexibility is indicated by higher B-factors). Reetz and co-workers successfully demonstrated that saturation mutagenesis of the surface residues showing highest flexibility in *Bacillus subtilis* lipase A resulted in variants displaying significantly improved thermostability (Reetz et al., 2006a). Structural data and homology models, especially of enzyme-substrate complexes, are particularly useful in identifying which residues are involved in substrate binding and catalysis. Along with biochemical and functional studies, this information can be used in rational design of libraries which include or exclude particular residues. For example, solution of the crystal structure of the serine:pyruvate aminotransferase from *Sulfolobus sulfataricus* and comparison with the closely related alanine:glyoxylate aminotransferase, combined with biochemical studies, identified the structural features likely to determine substrate specificity which could be targeted for mutagenesis in the future with the aim of introducing activity against different substrate(s) (Sayer et al., 2012).

The use of computational techniques is an emerging area in the rational design of mutant libraries. Modelling of mutations within the enzyme active site using programs such as Coot (Emsley et al., 2010) and docking of different substrates, for example using Auto Dock 4.2 (Sanner, 1999) can greatly reduce the number of mutants which need to be synthesised in the lab. Substrate docking was used in the redesign of isomaltulose synthase to identify residues which could be changed to permit synthesis of isomelezitose, a potential sucrose alternative (Gorl et al., 2012). In another study, the generation of the potential anti-tumour substance perillyl alcohol by cytochrome P450 monoxygenase 102A1 was achieved following multiple rounds of molecular modelling using the Amber 9 molecular dynamics simulation program

(ambermd.org) (Seifert et al., 2011). Numerous computational tools and algorithms have been developed to help in the design of mutant libraries such as HotSpot wizard (Pavelka et al., 2009), ProSAR (Fox et al., 2007) and PoPMuSic 2.1 (Dehouck et al., 2011). There is also an increasing interest in design of *de novo* enzymes by computational methods (Kries et al., 2013) and some significant successes have been reported (Jiang et al., 2008, Rothlisberger et al., 2008, Siegel et al., 2010).

Numerous examples exist where mutation of multiple residues in combination leads to higher quality libraries than when only one residue is targeted at a time. This is not surprising given that residue changes often have synergistic effects on the environment and conformation of other nearby residues in the structure (Reetz et al., 2005). In a study to alter the enantioselectivity of P450-BM3 monoxygenase a series of single and double mutants were generated. The greatest shifts were observed with the double mutants in all but one case (Agudo et al., 2012). Unfortunately, mutation of multiple residues simultaneously can lead to large libraries which are difficult to screen effectively. In order to reduce this problem, Reetz et al. developed iterative saturation mutagenesis (ISM) (Figure 1.4) (Reetz et al., 2006b, Reetz et al., 2006a). This technique involves first identifying a number of regions or positions to target based on the characteristic to be altered – each site consists of one or more amino acids. Saturation mutagenesis is then carried out at each of the sites independently to generate a number of small libraries which are screened and the best hit identified from each library. Each of these hits is then used in another round of randomisation at one of the different sites (it is possible, but not essential to randomise all of the different sites in combination with the initial hit). The libraries are then screened and hits identified to take forward to the next round. The process can be continued as many times as desired. ISM was successfully used to increase the stability of lipase from *B. subtilis* towards the organic solvents acetonitrile, dimethylsulfoxide and dimethylformamide (Reetz et al., 2010).
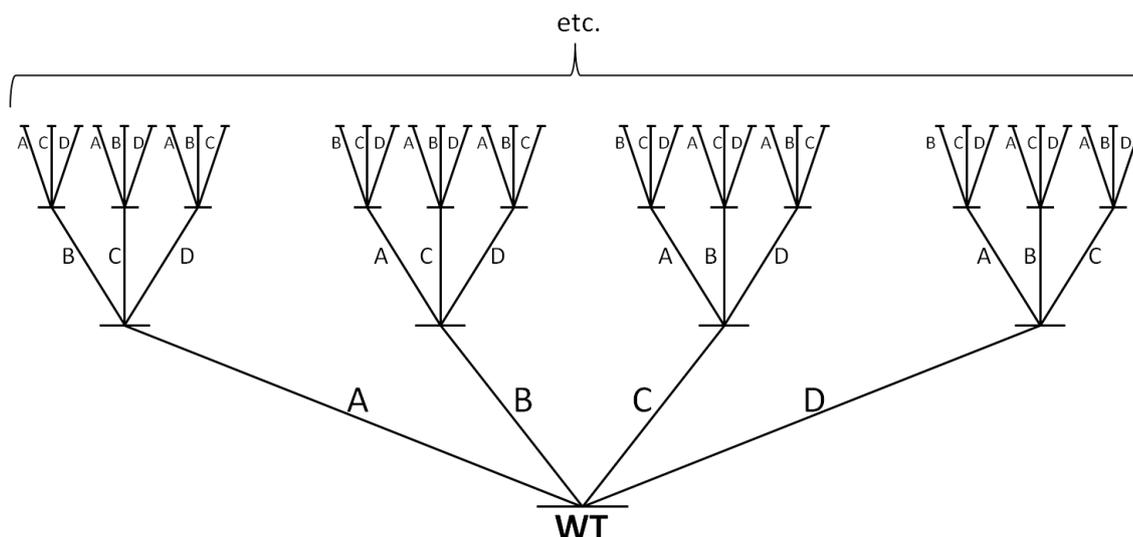
Figure 1.4. Schematic illustration of ISM whereby four sites were randomised: A, B, C and D

Copied from Reetz and Carballeira (2007). A description of the method is given in the text.

Another significant contribution from the Reetz group is the Combinatorial Active-site Saturation Test (CAST) (Reetz et al., 2005) which is designed to alter the substrate acceptance and/ or enantioselectivity of an enzyme. As mentioned above, randomisation of multiple residues at the same time allows for synergistic conformational effects from side chains of nearby residues which are otherwise difficult to predict in rational library design (Reetz et al., 2005). The CASTing procedure involves generation of a number of small libraries, each constructed by simultaneous saturation mutagenesis of two or three residues adjacent to each other in the 3D structure with side chains occurring next to the substrate binding site. Depending on the structural element (loop, β-sheet, α-helix, 3_{10} helix), the residues may or may not be adjacent in the protein sequence so structural analysis is essential in design of CASTing libraries. Following identification of hits within the CASTing libraries, the technique may be combined with ISM to further alter the substrate scope. Significant success has been reported by various groups using this method, such as enhancement of lipase activity towards various substrates (Reetz et al., 2005), and development of an amine dehydrogenase (Abrahamson et al., 2012).

There are numerous examples of rational design strategies where a combination of techniques have been employed, such as in the enhancement of the substrate scope of transketolase (Ranoux et al., 2012) where *in silico* docking experiments, sequence alignments and ISM were

all used to generate multiple variants with altered activities. epPCR is often used to identify residues of interest which are then incorporated into rational library design. A frequently quoted and relatively high profile example is the engineering of a transaminase to produce the antidiabetic compound sitagliptin (Savile et al., 2010) (Figure 1.5). The inability of the WT transaminase to bind pro-sitagliptin ketone was explored by docking studies and residues were identified for mutagenesis. Saturation mutagenesis was then carried out on a number of residues in the binding pocket, both individually and in a combinatorial approach, to generate variants showing a 75-fold enhancement in activity. Further evolution of the best variant by methods including random mutagenesis and DNA shuffling optimised the enzyme for use in industrial processes by improving its tolerance to high substrate and product concentrations, organic solvents and increased pH and temperature. The final biocatalyst contained 27 mutations and resulted in improved yield and productivity and reduced waste and cost compared to the previously used synthetic process as well as the elimination of heavy metals (rhodium) and the need for specialised high pressure hydrogenation equipment (Hansen et al., 2009).
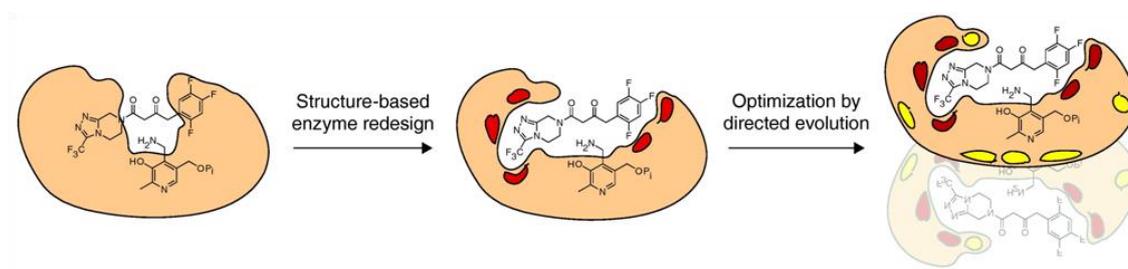
Figure 1.5. A schematic of the engineering of transaminase to catalyse production of sitagliptin

Taken from Lutz (2010). Initial stages of the enzyme engineering involved rational design and mutagenesis to introduce activity towards pro-sitagliptin ketone: mutations introduced are shown in red. The activity was then optimised for industrial manufacture of sitagliptin by numerous rounds of directed evolution: mutations introduced are shown in yellow, many occurred away from the substrate binding site at the dimer interface.

1.3.7 The argument for 'blind' evolution

Where limited information is available on the chosen enzyme, and particularly where there is no structural information, random mutagenesis methods such as epPCR are often the only option for introduction of new activities. The large number of successful studies where enzyme

activity was altered by such methods confirms the continuing value of 'blind' evolution, compared to rational design (Arnold, 1998). Arnold wrote in 1998 of our "dismally sparse understanding of protein structure-function" which, despite major advances in the last two decades, is still a limitation in rational design projects. Many protein features, especially resistance to high temperatures or organic solvents, are often dictated by residues found away from the substrate binding site and catalytic residues of the enzyme making them difficult to predict. Within the enzyme structure, more residues occur far from the active site than close to it so random mutagenesis will statistically target more distant mutations than close ones (Morley and Kazlauskas, 2005). In general, residues involved in determining enantioselectivity, regioselectivity and substrate selectivity occur in the first coordination sphere or have their $C\alpha$ within 10 Å of the substrate or key catalytic residues (Morley and Kazlauskas, 2005). However, there are examples where this is not the case. Distant residues can often affect catalysis due to major structural rearrangements, changes at the subunit interface or by affecting protein folding (Horsman et al., 2003). However, in many cases the changes are due to more subtle effects. For example, in β-lactamase the hydrogen bonding network which determines the position of the catalytic residue Arg244 was disrupted by mutation of the second sphere residue Asn276 (Bonomo et al., 1995). In mitochondrial creatine kinase a P207A mutation, which is over 15 Å from the active site, resulted in reduced activity and a shift in the pH optimum, presumably due to changes in long range electrostatic interactions (Forstner et al., 1998).

In summary, long range interactions can have significant effects on enzyme function which are difficult to predict with the current level of understanding of how enzyme structure relates to function. The easiest way to identify residues involved in these relatively unpredictable effects is by random mutagenesis methods such as epPCR.

1.3.8 Enzymes in industrial processes

Optimisation of catalysis of a specific reaction by an enzyme does not necessarily mean the enzyme is suitable for use in industrial processes. Numerous other factors, such as high turnover numbers, stability over longer time periods (hours, days or weeks) and resistance to harsh reaction conditions such as pH, temperature or organic solvents, also need to be considered. As mentioned in Section 1.3.1, immobilisation is often used to improve stability and other characteristics as well as reduce enzyme loss and many studies where various characteristics were optimised for industrial processes have been reported, such as for

sitagliptin (Section 1.3.6). If the enzyme is not optimised for the industrial process it may be necessary to increase the amount added which, even with cheaply available enzymes, will increase the cost of the process. As stated by Woodley (2013) typical examples of operating conditions that may be used in a biocatalytic process include:

- High concentrations of substrate and product (typically 50-400 g/ L, depending on application)
- High salt concentration
- Temperature and pH determined by substrate and product solubility and stability
- Gas-liquid interface (for reactions consuming or producing gas)
- Liquid-liquid interface (for aqueous-organic two-liquid phase reactions where organic phase is used as a carrier for the substrate and/ or product)
- Organic solvent (for assistance in solubilising substrates)
- Heterogeneity of temperature, pH and concentration of substrate and product (for large scale applications limited by mixing)

It is therefore generally important that the developed biocatalyst does not display substrate or product inhibition and is optimised for the required conditions of the process. While it is becoming increasingly possible to optimise the enzyme for the biocatalytic process, rather than *vice versa* (Bornscheuer et al., 2012), some limitations still exist such as the reaction vessel used. There are generally three types of reaction vessel: batch stirred tank reactor, continuous stirred tank reactor and continuous packed bed reactor, all of which require stable enzymes which are active at high substrate and/ or product concentrations (Woodley, 2013).

1.4 Screening for altered activities

A significant factor of any enzyme engineering project is how variants displaying the desired properties will be identified. The number of variants which can realistically be screened frequently determines the library generation method. For example, identifying changes in enantioselectivity can often only be carried out using HPLC and/ or GC and therefore two-stage screening is often carried out whereby the first screen identifies variants showing any activity and the second screen identifies changes compared to WT (van Leeuwen et al., 2012).

1.4.1 Selection techniques

Selection techniques provide a simple method for analysing large libraries efficiently, however the improved or acquired enzyme function must have a direct link with cell growth limiting the widespread use of these techniques (Leemhuis et al., 2009). Selection techniques are most frequently used in studies which aim to alter the substrate specificity of an enzyme (Griffiths et al., 2004, McLoughlin et al., 2005, Evnin et al., 1990). For example Bornscheuer and co-workers developed a selection method when introducing epoxide hydrolase activity into *Pseudomonas fluorescens* esterase (Jochens et al., 2009). Colonies expressing the variants could only grow on plates containing toxic glycidol butyrate if they displayed epoxide hydrolase activity which converted it to glycerol butyrate which could be used as a carbon source. Selection methods have also been used to identify changes in enantioselectivity, such as in the improvement of enantioselectivity of *Bacillus subtilis* lipase A. Aspartate auxotroph *E. coli* expressing variants showing activity against the (*R*)-enantiomer were unable to grow on plates containing the toxic phosphonate ester of (*R*)-(-)-1,2-*O*-isopropylidene-*sn*-glycerol while variants active against the (*S*)-enantiomer were selected for by inclusion of the aspartate ester of (*S*)-(+)-1,2-*O*-isopropylidene-*sn*-glycerol (Boersma et al., 2008).

1.4.2 Screening techniques

The use of screening techniques is much more common and there are a wide variety of examples with different benefits and limitations. The most simplistic screening format is agar plate based screening. Here colonies expressing the library variants are incubated with the enzyme substrate, or an analogue, resulting in a visual signal such as formation of colour or fluorescence. The signal may be generated by formation of a coloured product, either directly by the enzyme being screened (Parikh and Matsumura, 2005) or by another enzyme which uses a product of the reaction to convert a colourless substrate to a coloured product. An example of this was used by Alexeeva et al. (2002) to identify enantioselective amine oxidases. *E. coli* BL21 expressing the mutant library was plated out onto nitrocellulose filters on agar plates before cell lysis and addition of an assay mix containing the D- or L-enantiomer of the substrate, horseradish peroxidase (HRP) and 3,3'-diaminobenzidine (DAB). Hydrogen peroxide generated by any variants showing activity was used by HRP to oxidise DAB to a coloured product (Alexeeva et al., 2002) (Figure 1.6). A coloured product may also be generated as a result of pH change brought about by the catalysed reaction. A pH indicator is then used to identify colonies expressing active variants (Ben-David et al., 2008).
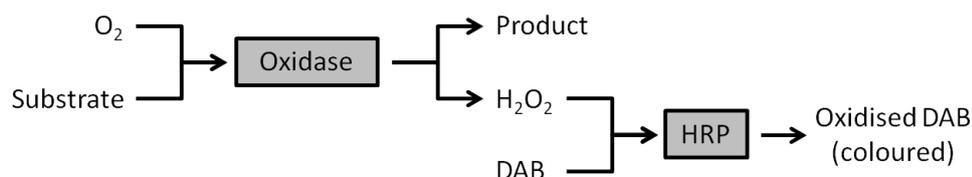
Figure 1.6. The coupled assay to detect oxidase activity used in many screening methods

Hydrogen peroxide produced as a by-product of the oxidation reaction is used to convert a colourless substrate such as DAB into a coloured product in a reaction catalysed by HRP.

The major advantage of agar plate-based screening is its simplicity. However, it does have the disadvantage of a limited dynamic range, *i.e.* while active versus inactive variants can be easily distinguished, *relative* levels of activity are generally not possible to measure. As a result, microtitre plate-based screening has developed as the most commonly applied technique in library screening (Leemhuis et al., 2009). Each well of the plate is inoculated with a single transformant expressing a different variant. The enzyme is either secreted into the growth medium or cells are lysed and, following centrifugation, the supernatant is added to a separate assay plate. Activity is detected by similar methods to the agar plate-based screens including use of pH indicators and colorimetric substrates which can be detected spectrophotometrically such as phenol red (Yi et al., 2012) and 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt (ABTS) (Torres-Salas et al., 2013). In directed evolution of a glycosyltransferase to expand the substrate promiscuity, a microtitre plate-based assay was used which measured the loss of fluorescence upon catalysis with various UDP-sugars as donors and the fluorescent coumarin 4-methylumbelliferone as acceptor (Williams et al., 2007). Microtitre plate-based screening has the advantage that analytical methods can be used to measure enzyme activity and relative activities of different variants compared to WT can accurately be measured. However, the relatively low throughput of ~10^4 variants per day has led to the development of various ultra-high throughput screening methods.

The most successful ultra-high throughput methods use fluorescence activated cell sorting (FACS) which permits screening of libraries of up to 10^9 or even 10^{10} variants (Leemhuis et al., 2009). Kolmar and co-workers labelled the two enantiomers of 2-methyldecanoic acid so that hydrolysis by *Pseudomonas aeruginosa* esterase A gave rise to green or red fluorescence. Esterase variants were expressed on the cell surface and, if hydrolysis of the substrate was catalysed, the red or green fluorescent product was immediately attached to a tyrosine

residue on the cell surface via radical formation catalysed by cell surface-expressed peroxidase (Figure 1.7) (Becker et al., 2007). Cells expressing variants active against each enantiomer could then be separated by FACS (Becker et al., 2008).
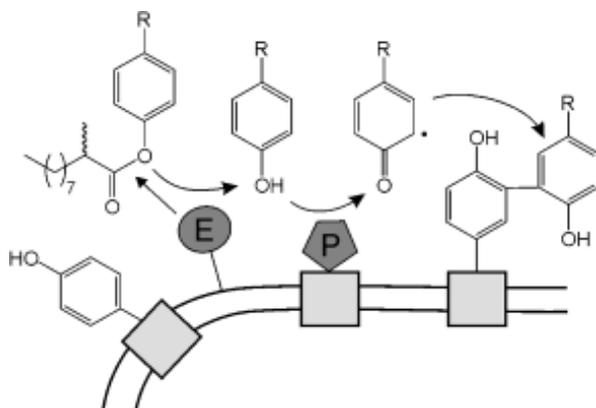
Figure 1.7. The reactions that lead to hydrolysis of the substrate and covalent attachment of the product to the cell surface

Taken from Becker et al. (2008). The esterase (E) cleaves the fluorescently labelled substrate. The product is then oxidised by peroxidase (P) to form a radical species which covalently attaches to a cell surface-displayed tyrosine.

Other methods which use FACS involve water-in-oil droplets which each contain a single bacterial cell expressing a different variant. The fluorescent product generated by variants active against the given substrate remains in the same droplet as the cell which can then be isolated by FACS (Aharoni et al., 2005). A similar method designed previously by the same group involved a cell-free water-in-oil droplet system whereby each droplet contained a different gene (as linear DNA) and a translation/ transcription mixture, resulting in multiple copies of the encoded variant which is retained within the droplet ready for FACS (Tawfik and Griffiths, 1998). There are numerous successful examples of this *in vitro* compartmentalisation being used in library screening (Taly et al., 2007, Leemhuis et al., 2005).

1.5 Enzyme cofactors

The functional groups of the 20 standard amino acids found within enzymes are capable of participating in acid-base reactions, charge-charge interactions and some group transfer reactions. However, a variety of reactions, including oxidation-reduction and group transfer processes, require the presence of small molecule cofactors (Voet and Voet, 1995). The majority of enzymes explored for biocatalytic processes utilise cofactors such as the organic compounds nicotinamide adenine dinucleotide (NAD(H)) and adenosine 5'-triphosphate (ATP) (Liu and Wang, 2007). The high cost of these cofactors can have significant implications for the economic feasibility of the biocatalytic process and, as such, the area of cofactor regeneration is a significant area of research (Zhao and van der Donk, 2003). A number of different regeneration processes have been developed, for example fusion of the biocatalyst with a reductase (Schuckel et al., 2012), development of robust cofactor-regenerating enzymes to include in the reaction vessel (Vazquez-Figueroa et al., 2007), and immobilisation-based techniques (Ngamsom et al., 2010).

There are a number of enzymes which contain cofactors generated by post-translational modification of encoded residues, rather than acquiring the cofactor from exogenous sources (Figure 1.8). A number of different residues can be modified to form these cofactors including serine in histidine decarboxylase (van Poelje and Snell, 1990), cysteine in various sulfatases (Schmidt et al., 1995) and lysine in ribulose-1,5-bisphosphate carboxylase (RuBisCo) (Taylor and Andersson, 1997). Numerous examples also exist of the formation of a cross-link between two residues to form protein-derived cofactors such as His-Tyr in catalase HPII from *E. coli* (Bravo et al., 1997) and cytochrome c oxidase (Ostermeier et al., 1997) and Tyr-Lys in lysyl oxidase (Wang et al., 1996).

Formation of protein-derived cofactors often requires additional proteins; for example, biogenesis of tryptophan tryptophylquinone from two active site tryptophans in methylamine dehydrogenase requires four gene products (van der Palen et al., 1997, van der Palen et al., 1995, Wilmot and Davidson, 2009). However, for some cofactors generation is self-catalysed and only requires components already present in the enzyme – this feature gives the enzymes great potential for development for the biocatalysis industry. An example of a self-processing enzyme is the copper amine oxidase family which contains a 2, 4, 5-trihydroxyphenylalanyl quinone cofactor. This is generated spontaneously in the presence of oxygen and copper from a highly conserved active site tyrosine (DuBois and Klinman, 2005).
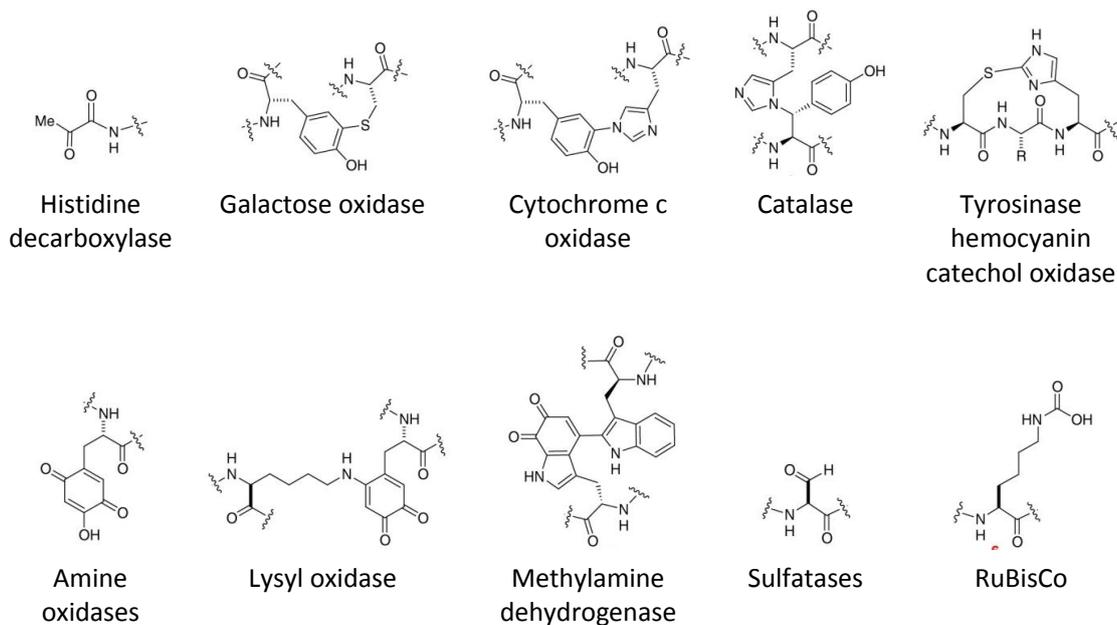
Figure 1.8. Structures of protein-derived cofactors from ten different enzymes

Adapted from Okeley and van der Donk (2000).

1.5.1 Metalloenzymes

Proteins incorporate metal ions for a number of different functions (Holm et al., 1996):

i)   Structural roles

ii)   Storage and/ or transport, such as metallothioneins which sequester zinc, cadmium and copper ions

iii)   Electron transfer, such as plastocyanin in photosynthesis

iv)   Dioxygen binding, such as haemoglobin

v)   Catalytic roles including substrate binding, activation and turnover

Up to a third of all enzymes use metal ion cofactors for catalytic roles (Voet and Voet, 1995) making the study of metalloenzymes a significant research area. Alkali and alkaline earth metal ions such as sodium, magnesium and calcium are generally loosely bound to the enzyme. Transition metal ions such as iron, cobalt, vanadium and copper are often found tightly bound at the active site via coordination by specifically orientated amino acid side chains. The ions may be directly incorporated into the protein, or may form part of prosthetic groups or complexes. Iron-containing enzymes are perhaps the most extensively studied of metalloenzymes due to their essential roles in generation of ATP via the haem-containing

cytochromes. The structure of the haem prosthetic group is shown in Figure 1.9. Other haem-containing enzymes include peroxidases and oxidoreductases, such as hydroxylamine oxidoreductase (Igarashi et al., 1997). There are also examples of iron-containing proteins where the metal ion is directly incorporated into the enzyme such as naphthalene 1,2-dioxygenase (Karlsson et al., 2003) and protocatechuate 3,4-dioxygenase (Vetting et al., 2000).
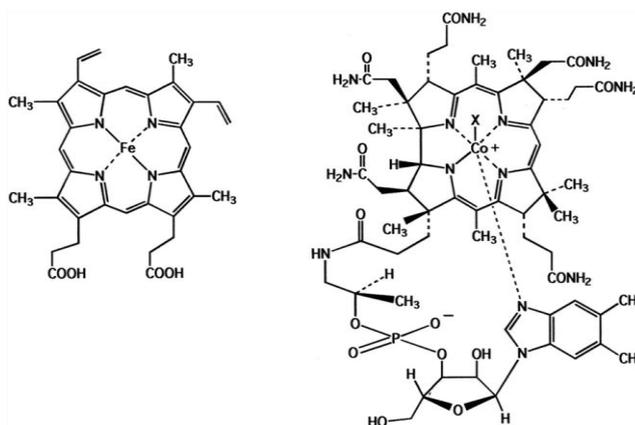
Figure 1.9. Structures of the haem (left) and cobalamin (right) prosthetic groups
  Taken from Roper et al. (2000).

Cobalt generally exists within enzymes as coenzyme B_{12} (cobalamin) (Figure 1.9) which is found in three different enzyme subfamilies: adenosylcobalamin-dependent isomerases, methylcobalamin-dependent methyltransferases and the dehalogenases (Banerjee and Ragsdale, 2003). The incorporation of vanadium into enzymes is relatively rare although a key example is the vanadium haloperoxidases, originally identified in seaweed (Vilter, 1984, de Boer et al., 1986). These enzymes catalyse the two electron oxidation of halides and may be of biotechnological interest due to their high stability and resistance to high substrate (hydrogen peroxide) and product (hypohalous acid) concentrations (Vollenbroek et al., 1995, van Schijndel et al., 1994, Liu et al., 1987) as well as their ability to halogenate compounds of potential industrial interest (Littlechild et al., 2002).

Copper-containing enzymes are widespread in nature due to their ability for involvement in redox reactions and the high solubility of copper oxides (MacPherson and Murphy, 2007). Fe(III)/ Fe(II) redox pairs are used for many enzyme-catalysed reactions, however, the Cu(II)/ Cu(I) redox potentials of copper proteins are generally higher than the potentials of functionally analogous iron proteins, permitting catalysis of more difficult reactions (Kaim and

23

Rall, 1996). Copper centres within proteins are generally classified into different types based on the coordinating ligands and the geometry of the metal centre which result in different spectroscopic properties and reduction potentials (Rubino and Franz, 2012, MacPherson and Murphy, 2007):

- Type 1 copper sites are typically coordinated by a Cys and two His residues in a trigonal planar arrangement. Additional axial ligands, such as Met, result in either tetrahedral or trigonal bipyramidal geometries. Charge transfer between the copper and the sulfur atom of the Cys ligand results in an intense blue-green colour. These sites are involved in single electron transfer reactions in electron transfer proteins, such as plastocyanin, and in intramolecular electron transfer in enzymes such as nitrite reductase.

- Type 2 copper sites take on a variety of geometries with a variety of different ligands; although most are three or four coordinate with His as at least one of the ligands. Coordination positions are often left vacant to be occupied by exogenous ligands such as enzyme substrates. Key examples of enzymes containing a type 2 copper centre include Cu, Zn superoxide dismutase and copper amine oxidase.

- Type 3 copper sites contain two coppers; each coordinated by three His residues in a trigonal planar geometry. Binding of dioxygen brings the copper atoms closer together to take on a trigonal bipyramidal geometry. A type 3 site is found in catechol oxidase and in the oxygen transporter haemocyanin.

- Additional types of copper sites include Cu_A and Cu_B sites which are both found in cytochrome c oxidase and contain two and one copper atoms, respectively. Cu_Z copper sites which contain four copper atoms are the most recent type to be identified and are found in N_2O reductase.

1.6 Galactose oxidase

The enzyme galactose oxidase (GO) (EC 1.1.3.9) contains a type 2 copper site and a protein derived cofactor which permit catalysis of the oxidation of primary alcohols to the corresponding aldehyde with concomitant reduction of dioxygen to hydrogen peroxide (Figure 1.10).

Figure 1.10. The reaction catalysed by GO

> Primary alcohol is oxidised to the corresponding aldehyde while molecular oxygen is reduced to hydrogen peroxide.

GO was first reported in the growth medium of *Polyporus circinatus* (Cooper et al., 1959), although the organism was later identified as *Dactylium dendroides* (Nobles and Madhosingh, 1963) and then *Fusarium graminearum* (Ogel et al., 1994). It is thought that the primary function of the enzyme is generation of hydrogen peroxide which is then used for either antibiotic defence roles, or for use by peroxidases in lignin degradation (Whittaker et al., 1998). This theory is backed up by the fact that GO displays a broad substrate specificity (Section 1.6.2) which may be a compromise with the need to achieve a high rate of turnover to generate hydrogen peroxide from a range of available substrates (Whittaker et al., 1998).

GO is an ideal candidate for development as an industrial biocatalyst due to a number of different features:

- An autocatalytically generated protein-derived cofactor removes the need for expensive exogenous cofactors or regeneration processes
- A broad substrate specificity provides a good starting point for introduction of novel activities or improving specificity for a particular substrate
- A high stability, demonstrated by activity after incubation for 1 hour in 8 M urea (Kelly-Falcoz et al., 1965)
- It has been extensively characterised since the 1950s, including solution of the crystal structure (Ito et al., 1991) (Section 1.6.1) and numerous mutagenesis studies (Section 1.6.2)
- It has a surface exposed active site, meaning modifications are unlikely to disrupt the overall protein fold
- There are established methods for recombinant expression and purification (Section 1.6.4)
- It is tolerant of high substrate and product concentrations.

1.6.1 Structure of GO

The crystal structure of GO was solved in 1991 at the University of Leeds (Ito et al., 1991) (Figure 1.11). The structure was particularly significant as it revealed the presence of a novel thioether bond in the active site which had not been previously observed (Section 1.6.1.2). The enzyme is monomeric, made up of 639 amino acids (McPherson et al., 1992) and has a relative molecular mass of 68.5 kDa. The structure consists of three domains made up of predominantly β-sheet (the only α-helix occurs between residues 327-332) which is likely to account for the high stability of the enzyme.

1.6.1.1 The three-domain structure

Domain 1 (residues 1 to 155) (Figure 1.11, yellow) is at the N-terminus of the protein and consists of eight β-strands in a β-sandwich structure where a five-stranded antiparallel β-sheet faces a three-stranded antiparallel β-sheet. Within Domain 1 a pair of β-strands is also rolled to form a jelly roll motif (Ito et al., 1994). The function of this domain is unclear, although the identification of a binding site showing some affinity for D-galactose led to the proposal that it may be involved in binding polysaccharides in cell walls thus anchoring the enzyme in position (Ito et al., 1994). The domain may also play a role in correct folding of the enzyme but has been shown to not be required for catalytic activity (Mahmoud et al., 2000).

Domain 2 (residues 156-532) (Figure 1.11, green) consists of one α-helix and 28 β-strands which are arranged in seven antiparallel β-sheets, each containing four β-strands. The β-sheets are aligned with pseudo seven-fold symmetry in an arrangement resembling a 'seven-petalled flower' (Ito et al., 1991, Ito et al., 1994), now known as a β-propeller (Murzin, 1992). The tight packing of the seven β-sheets is thought to contribute towards the high stability of the enzyme (Ito et al., 1994). Similar arrangements have been observed in other proteins such as methylamine dehydrogenase (Vellieux et al., 1989). Each of the seven 'blades' of the propeller are made up of the ~50 residue *kelch* sequence motif, originally identified in *Drosophila* mutants (Bork and Doolittle, 1994). Solution of the GO crystal structure provided the first evidence that each copy of the *kelch* sequence motif encodes a 'blade' of the β-propeller which has been seen to consist of four to eight blades in different proteins. The copper ion (Figure 1.11, cyan sphere) is found near the pseudo seven-fold axis of the β-propeller at the surface-exposed active site. Most of the residues important for substrate binding and catalysis are found within Domain 2 (Section 1.6.1.2).

Domain 3 (residues 533-639) (Figure 1.11, brown) is found on the opposite side of Domain 2 from the copper ion and consists of seven β-strands surrounding a hydrophobic core (Ito et al., 1994). Two of the β-strands are significantly longer than the others and form an antiparallel β-ribbon which penetrates the middle of Domain 2 along the pseudo seven-fold axis. The copper ligand His581 is found at the end of the β-ribbon. It is thought that Domain 3 acts to stabilise the β-propeller of Domain 2 (Ito et al., 1994).
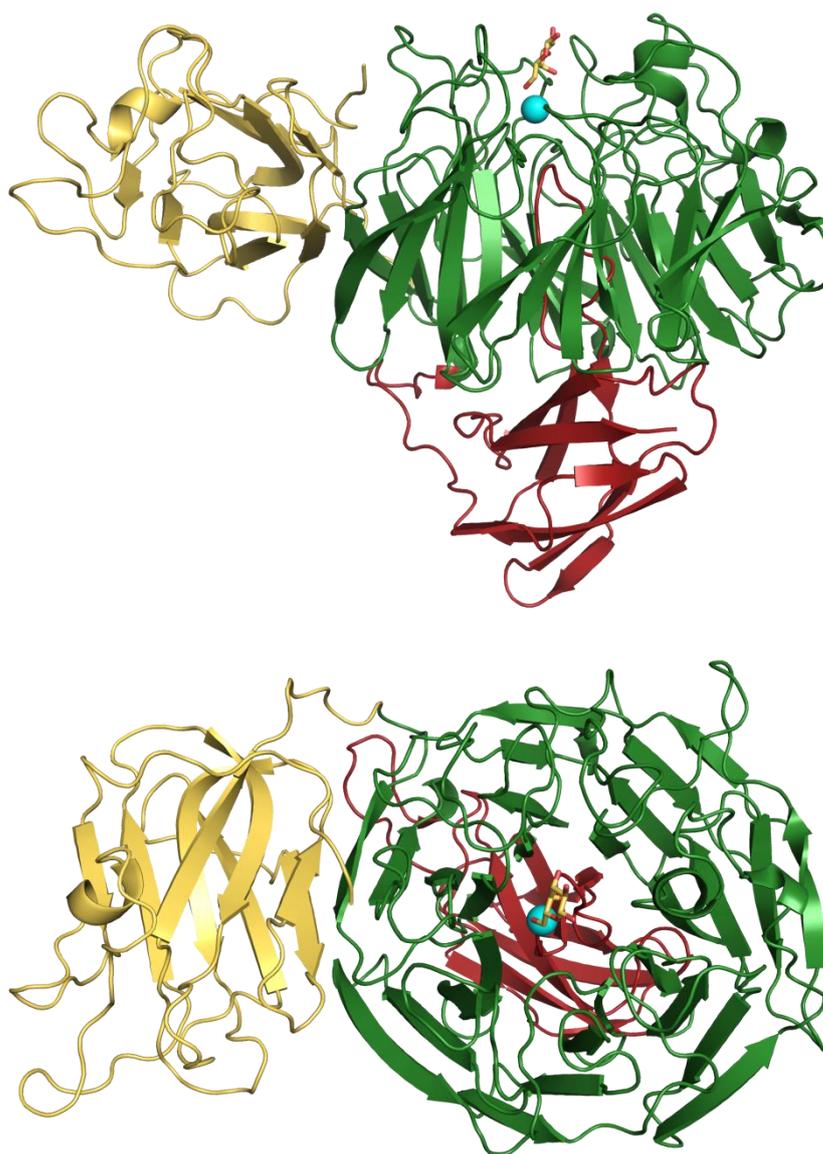
Figure 1.11. Representations of the crystal structure of GO

The top image is rotated 90 ° out of the plane of the paper to generate the bottom image. The figure was created with PyMOL using .pdb file 1GOF. Domain 1 is shown in yellow, domain 2 is shown in green and domain 3 is shown in brown. Copper is shown as a cyan sphere and the D-galactose molecule in yellow is modelled into the active site according to Wachter and Branchaud (1996).

1.6.1.2 The active site

Before determination of the crystal structure, a number of theories existed as to how the mononuclear metal ion active site of GO was able to catalyse the oxidation of the alcohol group and the reduction of dioxygen; both two-electron processes. The catalysed reaction is essentially the same as transfer of dihydrogen between the two substrates, a process which generally involves redox cofactors such as NAD(H) or quinones (Whittaker, 2005). An early theory suggested the involvement of a Cu(III)-Cu(I) redox couple (Hamilton et al., 1978) while the presence of a pyrroloquinoline quinone cofactor was implied by work carried out by Duine and co-workers (van der Meer et al., 1989). Around the same time, Whittaker and co-workers proposed that oxidation of GO resulted in a protein-based radical which was involved in catalysis, as well as the copper (Whittaker and Whittaker, 1988). They also suggested that a ligated tyrosine was involved in stabilisation of the radical within the active site (Whittaker et al., 1989).

Solution of the crystal structure of GO identified a novel thioether bond between an active site tyrosine (272) and the cysteine at position 228 (Figure 1.12) (Ito et al., 1991). This covalent bond between the $C^{\varepsilon 1}$ of the tyrosine and the sulfur of the cysteine results in increased rigidity of the active site, in a similar way to the effect of a disulfide bond, although the thioether bond cannot be reduced in the same way. As a result, when analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) the 68.5 kDa protein migrates with an apparent size of ~65.5 kDa. This is because the presence of the thioether bond causes the 44 residues between Cys228 and Tyr272 to form a loop leading to a slightly more compact structure which migrates faster than the fully extended protein without the thioether bond (McPherson et al., 1993).
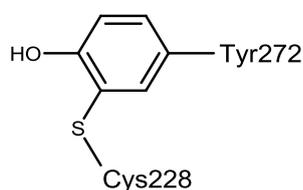
Figure 1.12. The thioether bond present in the GO active site

Along with other features of the protein environment within the active site, the thioether bond lowers the redox potential of Tyr272 from around 1000 to 400 mV making it more accessible for catalysis (Stubbe and van der Donk, 1998, Harriman, 1987, Hamilton et al.,

1978). The protein radical, formed upon oxidation of Tyr272, is stabilised by delocalisation across the thioether bond (Babcock et al., 1992, Gerfen et al., 1996). Other residues within the active site are thought to also contribute to radical stabilisation, particularly Trp290 which stacks over the thioether bond. Its six-membered ring is located exactly above the sulfur atom (Figure 1.13) (Ito et al., 1991). The side chain of Trp290 is in almost exactly the same plane as the two residues involved in the thioether bond (Figure 1.13, right) (Ito et al., 1994). It also appears likely that Trp290 protects the stabilised radical from solvent.
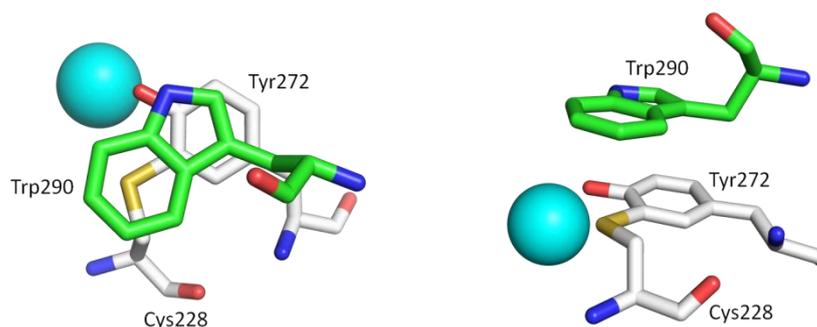
Figure 1.13. Trp290 (green) stacks over the sulfur atom of the thioether bond between Tyr272 and Cys228 (white)

The arrangement is shown looking down on the active site from bulk solvent (left) and looking from the side (right). Copper is shown as a cyan sphere. Figure created with PyMOL using .pdb file 1GOF.

Of particular interest for industrial development of GO, generation of the thioether bond cofactor does not require any additional proteins. Only copper and dioxygen are required to generate the catalytically active enzyme (Firbank et al., 2003, Rogers et al., 2000). In fact, the thioether bond can be generated under anaerobic conditions, although the enzyme is left with Cu(I) in the active site which must be oxidised to Cu(II) for catalysis to occur (Rogers et al., 2008). Since discovery of the thioether bond in GO, the same cofactor has been identified in other redox proteins, such as the iron-containing cysteine dioxygenase (McCoy et al., 2006, Simmons et al., 2006), sulfite reductase (Schnell et al., 2005) and nitrite reductase (Polyakov et al., 2009); and a copper-containing glyoxal oxidase which is closely related to GO (Whittaker et al., 1996).

Another essential component of the GO active site is the type 2 copper centre where the Cu(II) is coordinated by five ligands: Tyr272, Tyr496, His496, His581 and an exogenous ligand such as water or acetate (Ito et al., 1991). The position of the acetate molecule in the crystal structure

is proposed to mimic the binding site of the alcohol substrate during catalysis, discussed further in Section 1.6.3 (Whittaker, 2005). The site displays significant physical rigidity, due to the tight packing of Domains 2 and 3, which is suggested to result in the high specificity for copper as the metal cofactor (Ito et al., 1994). The arrangement of the copper ligands in the crystal structure at pH 4.5 is shown in Figure 1.14. In the crystal structure at pH 7.0 the acetate is replaced by a water molecule and the copper site takes on a distorted tetrahedral geometry (Ito et al., 1991).
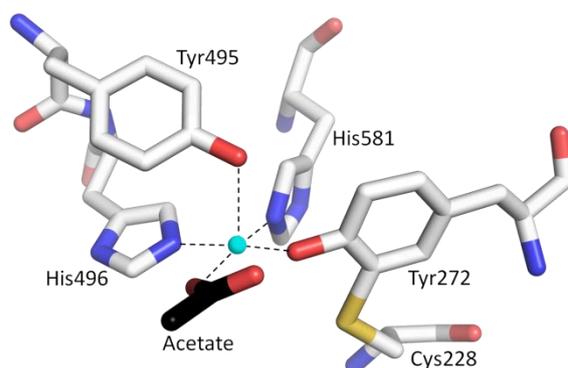
Figure 1.14. The type 2 copper centre of GO

The four equatorial ligands: acetate, Tyr272, His581 and His496 and the axial ligand: Tyr495, are shown with the copper (cyan sphere) in the centre. Figure created with PyMOL using .pdb file 1GOF.

1.6.1.3 GO family members

As mentioned above, the enzyme glyoxal oxidase is closely related to GO. Glyoxal oxidase is secreted by the fungus *Phanerochaete chrysosporium* and catalyses the oxidation of aldehydes to the corresponding carboxylic acid with reduction of dioxygen to hydrogen peroxide (Whittaker et al., 1996, Kersten and Kirk, 1987). Sequence alignments and mutagenesis studies have shown that GO and glyoxal oxidase have the same active site architecture with the copper coordinated by two histidines and two tyrosines, one of which forms a thioether bond with a cysteine. The key difference, which is likely to account for the biochemical differences in the two enzymes, is in the stacking tryptophan (W290 in GO) which is replaced with a histidine in glyoxal oxidase (Whittaker et al., 1999).

The increase in genomic data over the past decade has resulted in identification of numerous other potential GO family members. Analysis of the DNA sequence of GO from *F. graminearum* using the Basic Local Alignment Search Tool (BLAST) at www.ncbi.nlm.nih.gov identifies

sequences with >40% identity from over 50 different organisms. While most of these are from different fungal species such as *Cordyceps militaris* (Zheng et al., 2011) and *Beauveria bassiana* (Xiao et al., 2012), there are also examples of sequences showing as much as 47% sequence similarity from bacterial genera such as *Actinoplanes*, *Burkholderia* and *Streptomyces*. The vast majority of these gene products have not undergone any characterisation, but the ever increasing list of potential GO homologues may prove very useful in future development projects.

1.6.2 Substrate specificity & binding

GO displays broad substrate specificity with activity towards a wide range of primary alcohols reported (Table 1.1). This is a desirable trait in an industrial biocatalyst (Breuer et al., 2004). GO activity has also been reported towards polysaccharides where D-galactose occurs on the non-reducing end (Bretting and Jacobs, 1987, Avigad, 1985, Schlegel et al., 1968). The relative activities towards the different substrates, especially where activity was not detectable, have provided valuable information for modelling the enzyme-substrate interactions which occur during catalysis. Unfortunately, the crystal structure of an enzyme-substrate complex has not yet been solved, perhaps due to crystal packing limitations (Ito et al., 1994), or due to the relatively low affinity GO has for the substrate D-galactose, implied by a relatively high K_M.

Following solution of the crystal structure, Ito et al. (1994) used calculations of the water accessible surface (Connolly, 1983) to identify the active site at the position of the copper and thioether bond and demonstrated that the site is complementary to D-galactose in the chair conformation. They also proposed that the substrate binds at the equatorial copper ligand site occupied by acetate or water in the crystal structures. Detailed molecular modelling studies by Wachter and Branchaud (1996) predicted the most favourable conformation for D-galactose to bind in the active site. Seven hydrogen bonds were predicted (Figure 1.15): O-1 as a hydrogen bond donor to the O^η of Y329; O-3 as a hydrogen bond acceptor to $N^{\eta 1}$ of R330 and a donor to O^ε of Q406; O-4 as a hydrogen bond acceptor to both $N^{\eta 1}$ and $N^{\eta 2}$ of R330 and also a donor to O^η of Y495; and O-5 as a hydrogen bond acceptor to N^ε of W290. The B-face of the D-galactose molecule (defined as the side where numbering of the carbon atoms is anti-clockwise (Spurlino et al., 1991)) is much more hydrophobic than the A-face (where numbering is clockwise) due to protruding C-H groups rather than C-OH groups. Because of this feature, the D-galactose fits well against the hydrophobic residues F464 and F194, which may be important for positioning the substrate for hydrogen bonding interactions in other areas of the active site. It is proposed

that D-glucose binding is disfavoured due to its equatorial hydrophilic O-4 hydroxyl which would be too close to these hydrophobic residues if D-glucose bound in the same conformation as D-galactose in the WT enzyme (Wachter and Branchaud, 1996). The O-4 hydroxyl may also cause a steric clash with Tyr495, a theory which likely explains why substitutions at positions C-1, C-2 and C-3 are much better tolerated than those at positions C-4, C-5 or C-6 (Table 1.1 and (Schlegel et al., 1968)).
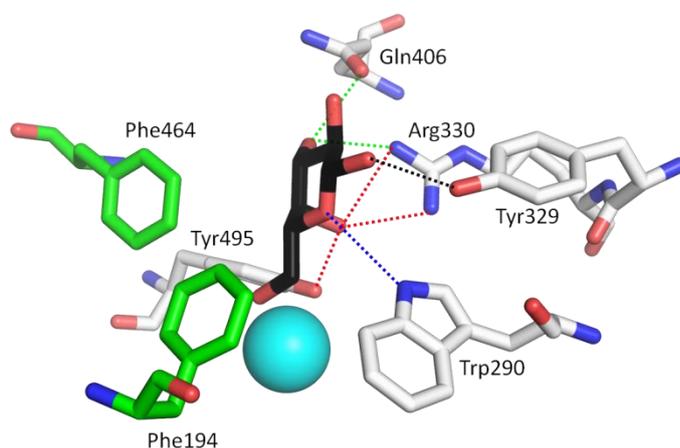
Figure 1.15. The proposed interactions of D-galactose in the active site of GO

> The D-galactose molecule (black) is shown docked into the active site according to Wachter and Branchaud (1996). F194 and F464 (green) are potentially involved in hydrophobic interactions with the B-face of the D-galactose molecule. Hydrogen bonds between O-1 (black), O-3 (green), O-4 (red) and O-5 (blue) with active site residues are shown as dotted lines. Copper is shown as a cyan sphere and is proposed to interact with the C-6 hydroxyl of D-galactose. Figure created with PyMOL using .pdb file 1GOF.

The strict regioselectivity of GO, where only the C-6 hydroxyl of D-galactose is oxidised (Whittaker and Whittaker, 1998), is important for potential industrial applications as use of the enzyme removes the need for protective group chemistry (Schoevaart and Kieboom, 2004). This regioselectivity is likely brought about by the multiple interactions between D-galactose and the active site, as well as the overall rigidity of the structure.

Table 1.1. Some of the substrates towards which GO activity has been explored

Activity higher than with D-galactose
1,5-Anhydrogalactitol (a)
Dihydroxyacetone (b)
2-Methylene-1,3-propanediol (b)
1,3-Dihyroxy-2-propanone (c)
Methyl α-D-galactopyranoside (d)
3-O-Methyl D-galactose (d)
1,5-Anhydro-D-galactitol (d)
Raffinose (d)
Guaran (d)

Activity lower than with D-galactose	
Dulcitol (a)	β-D-Lactose (c)
2-Deoxy-D-galactose (a)	2-Propene-1-ol (c)
D-Talose (a)	2-Pyridine methanol (c)
D-Galactosamine (a)	3-Pyridine methanol (c)
N-Acetyl-D-galactosamine (a)	4-Pyridine methanol (c)
D-Tagatose (a)	O-Nitrophenyl β-D-galactopyranoside (d)
D-Gulose (a)	1-O-(α-D-galactopyranosyl)-myo-inositol (d)
D-Lyxose (a)	2-O-Methyl D-galactose (d)
Methanol (b)	2,3-Di-O-Methyl D-galactose (d)
2-Chloroethanol (b)	O-α-D-Galactopyranosyl-(1-3)-D-galactose (d)
2-Nitroethanol (b)	2'-O-(α-D-fructopyranosyl) lactose (d)
1,3-Propanediol (b)	2-Amino-2-deoxy-D-galactose (d)
Ethylene glycol (b)	Methyl 2-acetamido-2-deoxy-3-O-methyl α-D-galactopyranoside (d)
Glycerol (b)	
Benzyl alcohol (b)	O-β-D-Galactopyranosyl-(1-3)-2-acetamido-2-deoxy-D-galactose (d)
Acetol (b)	
α-Hydroxyacetophenone (b)	Ovomucoid (d)

No detectable activity	
D-Galactonate (a)	2-Deoxy-D-glucopyranose (c)
D-Glucose (a, c)	Methyl β-D-glucopyranoside (c)
2-Deoxy-D-glucose (a)	D-Mannose (c)
D-Fucose (a)	D-Maltose (c)
L-Arabinose (a)	Amylose (c)
D-Galacturonic acid (a)	3-Buten-2-ol (c)
D-Glycero-D-galactoheptose (a)	4-O-Methyl D-galactose (d)
L-Galactose (a, c)	2,4-Di-O-Methyl D-galactose (d)
D-Fructose (a)	1,2:3,4-Di-O-isopropylidene-α-D-galactose (d)
L-Arabitol (a)	Chondrosine (d)
Ethanol (b)	O-β-D-Glucopyranosyl-(1-3)-2-acetamido-2-deoxy-D-galactose (d)
Propanol (b)	
Butanol (b, c)	Chondroitin 4-sulfate (d)
2,2-Dichloroethanol (b)	Dermatan sulfate (d)

Different relative rates of activity were reported in (Avigad et al., 1962) (a), (Kosman, 1984) (b), (Sun et al., 2002) (c) or (Schlegel et al., 1968) (d).

A number of mutagenesis studies have been carried out on GO with the aim of improving its activity or altering the substrate specificity. In the process, other residues involved in substrate binding and catalysis have been identified. Delagrave et al. (2001) carried out epPCR to evolve GO and identified mutations at three positions which resulted in improved activity towards methyl D-galactose: Cys383Ser, Tyr436His/ Asn and Val494Ala. These mutations were further explored by Wilkinson et al. (2004), revealing significant effects on $K_{M(D-galactose)}$ upon mutation of Cys383. Mutations of Tyr436 or Val494, however, only led to improvements in activity towards D-galactose when combined with other mutations, confirming the value of targeting multiple positions simultaneously when mutating enzymes (Section 1.3.6). Position 383 was later subjected to saturation mutagenesis (Deacon and McPherson, 2011) revealing significant effects on the kinetic parameters of the reaction with D-galactose (Table 1.2). C383E/ K/ M/ S showed reductions in K_M with only moderate or no effect on k_{cat}, while the C383T mutation greatly increased activity. It is unclear how the residue at position 383 is involved in substrate binding and catalysis – the crystal structure of the C383S variant revealed only minor changes compared to WT (Wilkinson et al., 2004). However, as the only unpaired cysteine in GO and with obvious effects on activity upon mutagenesis, this residue is of particular interest.

Table 1.2. Kinetic data for C383 variants

GO variant	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/ K_M (M^{-1} s^{-1})
WT	54 ± 6.1	1100 ± 41	20000 ± 3200
C383A	160 ± 21	1200 ± 50	7700 ± 1500
C383D	84 ± 2.6	440 ± 3.8	5200 ± 210
C383E	21 ± 1.8	550 ± 9.0	26000 ± 2700
C383F	130 ± 29	190 ± 13	1500 ± 450
C383G	83 ± 3.3	1100 ± 12	14000 ± 690
C383H	30 ± 3.7	210 ± 5.7	7000 ± 1100
C383I	510 ± 35.7	260 ± 8.5	510 ± 54
C383K	48 ± 6.7	1100 ± 30	23000 ± 3900
C383L	55 ± 8.3	450 ± 17	8200 ± 1600
C383M	33 ± 4.5	510 ± 15	15000 ± 2600
C383N	390 ± 38	410 ± 17	1100 ± 150
C383P	140 ± 3.9	490 ± 4.3	3500 ± 130
C383Q	64 ± 17	170 ± 12	2600 ± 960
C383R	290 ± 23	8.8 ± 0.3	30 ± 3.4
C383S	34 ± 3.6	1100 ± 30	32000 ± 4500
C383T	530 ± 99	3400 ± 300	6400 ± 1800
C383V	430 ± 21	360 ± 7.8	830 ± 59
C383W	60 ± 7.4	0.011 ± 0.0004	0.19 ± 0.033
C383Y	Not determined	Not determined	Not determined

Taken from Deacon and McPherson (2011).

Other studies have mutated residues known to be involved in substrate binding, such as R330K/ Q406T/ W290F which led to a 100-fold increase in D-glucose activity (Sun et al., 2002); R330M/ Q406T/ W290F which improved the activity and enantioselectivity towards secondary alcohols (Escalettes and Turner, 2008); and R330K which improved the catalytic efficiency of the reaction with D-fructose over eight-fold (Deacon et al., 2004). A directed evolution study also led to improved expression of GO in *E. coli* and enhanced the stability of the enzyme as discussed in Section 1.6.4 (Sun et al., 2001).

1.6.3 Mechanism of alcohol oxidation

1.6.3.1 The different redox states of GO

GO can exist in three distinct redox states: an oxidised state with Cu(II) and the tyrosyl radical which is the active form of the enzyme; a semi-reduced state with Cu(II) but no tyrosyl radical which is inactive; and a reduced state with Cu(I) and no tyrosyl radical which only occurs during catalysis or upon addition of substrate under anaerobic conditions (Figure 1.16) (Whittaker and Whittaker, 1988).
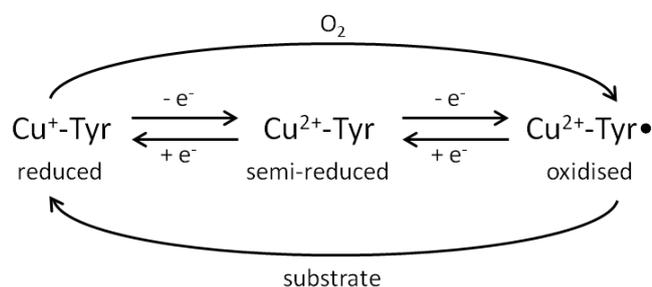
Figure 1.16. Schematic representation of the three redox states of GO
   Adapted from Deacon (2008).

GO is isolated as a mixture of oxidised and semi-reduced states, thus addition of a mild oxidant such as ferricyanide is required to ensure all of the GO is present in the oxidised, active form (Whittaker and Whittaker, 1988). HRP, commonly used in coupled assays of GO activity, has also been shown to oxidise GO to the active form (Kwiatkowski and Kosman, 1973, Cleveland et al., 1975, Tressel and Kosman, 1980). The semi-reduced form can be generated by addition of a mild reductant such a ferrocyanide. The three different states show characteristic UV/ vis spectra (Figure 1.17). The peak at 445 nm corresponds to a ligand to metal charge transfer (LMCT) between Tyr495 and Cu^{2+} (present in both the oxidised and semi-reduced states)

combined with a $\pi - \pi^*$ transition of the tyrosyl radical (present in the oxidised state only). The broad peak at ~810 nm has been attributed to ligand through LMCT coupling Tyr495, Tyr272 and the Cu^{2+}, combined with weaker Cu^{2+} d-orbital transitions (Whittaker et al., 1989, Whittaker and Whittaker, 1993, Messerschmidt et al., 2001). As the radical is not present in the semi-reduced form, the spectral feature associated with the stacking interactions is not visible.
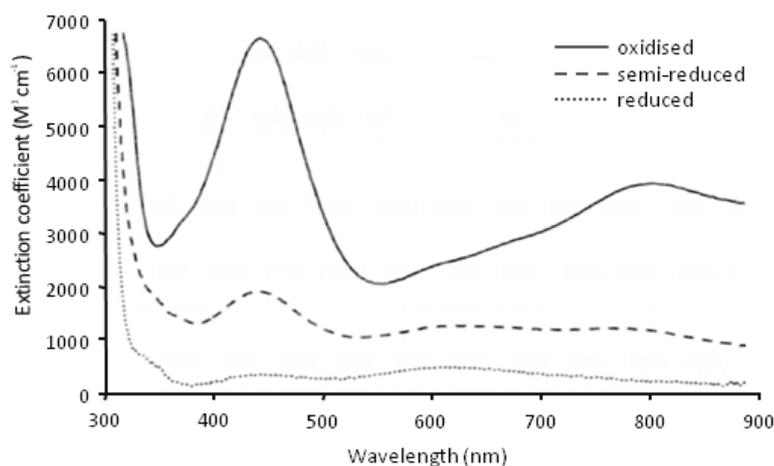
Figure 1.17. Typical UV/ vis spectra of the three different redox states of GO
Adapted from Messerschmidt et al. (2001).

1.6.3.2 The catalytic cycle of GO

It is currently accepted that oxidation of D-galactose by GO occurs by a 'bi-bi ping-pong' mechanism where the alcohol substrate binds, is oxidised, and aldehyde released and then dioxygen binds, is reduced, and hydrogen peroxide is released (Figure 1.18).

Figure 1.18. Schematic representation of the bi-bi ping-pong mechanism proposed for GO
Adapted from Deacon (2008).

In the reductive half reaction, substrate binds at the equatorial copper site normally occupied by solvent (Ito et al., 1994, Wachter and Branchaud, 1996), and a proton is abstracted from the C-6 hydroxyl by the axial copper ligand, Tyr495, which acts as the catalytic base (Figure 1.19, **(i)**). This step is supported by studies of binding of the inhibitor azide (Whittaker and Whittaker, 1993), as well as mutagenesis of Tyr495 to Phe which removes the ability of the residue to abstract protons (Reynolds et al., 1995). The radical present at Tyr272 then abstracts the *pro*-S hydrogen atom from the C-6 methylene group of the substrate intermediate, resulting in a substrate radical (Figure 1.19, **(ii)**). This transfer of a hydrogen atom is backed up by the observation of significant kinetic isotope effects and no detectable solvent isotope effects (Whittaker et al., 1998), while the stereoselective nature of hydrogen abstraction was shown by Whittaker and co-workers and proposed to be due to constraints imposed by the active site structure on the orientation of substrate binding (Wachter and Branchaud, 1996, Minasian et al., 2004). In the final step of the reductive half reaction, transfer of a single electron from Cu^{2+} to the substrate radical results in formation of a Cu^{+} centre and release of the aldehyde product (Figure 1.19, **(iii)**).

The oxidative half reaction is significantly less well characterised, perhaps due to the extremely fast reaction of Cu^{+} with dioxygen. Dioxygen is proposed to bind at the equatorial copper site in an end-on coordination and an electron is transferred from the Cu^{+} to yield a metal-bound superoxide and reform the Cu^{2+} centre (Figure 1.20, **(iv)**). The superoxide then abstracts a hydrogen atom from the tyrosyl group of the Tyr-Cys cofactor to form hydroperoxide and reform the cofactor radical at the same time (Figure 1.20, **(v)**). A proton is then transferred from Tyr495 resulting in hydrogen peroxide, which is released from the active site (Figure 1.20, **(vi)**). A hydrogen bond is reformed between Tyr495 and the Cu^{2+} and GO is restored to its fully reoxidised state (Humphreys et al., 2009).
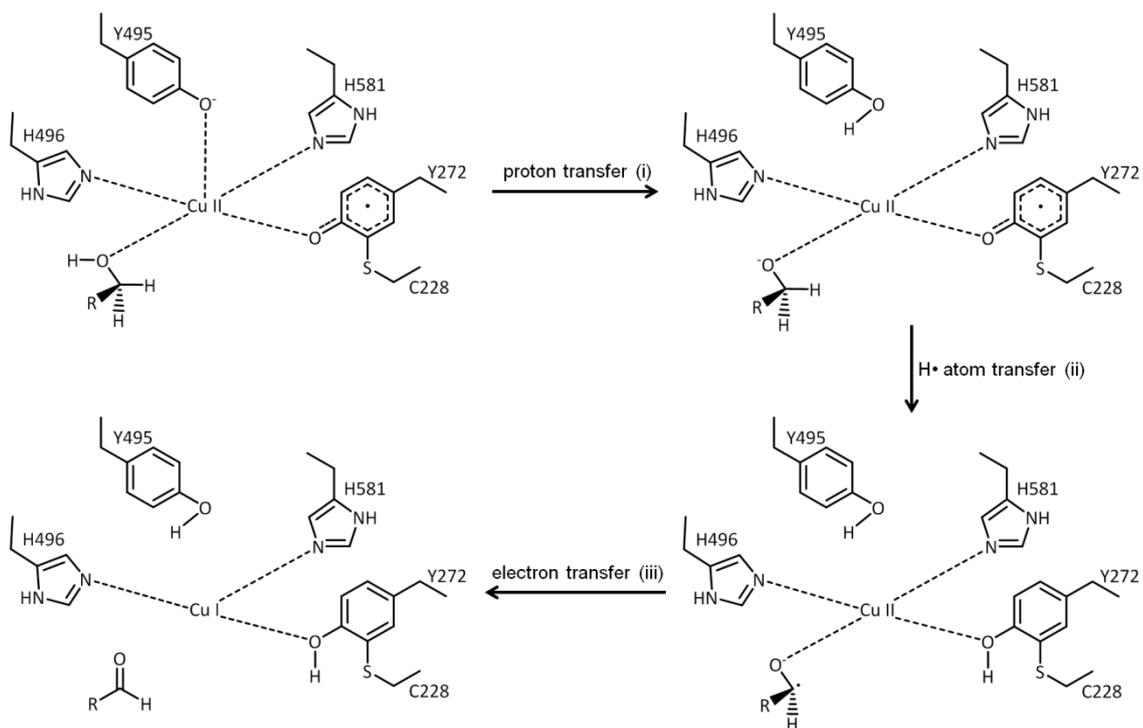
Figure 1.19. The reductive half reaction of GO

Adapted from Messerschmidt et al. (2001).
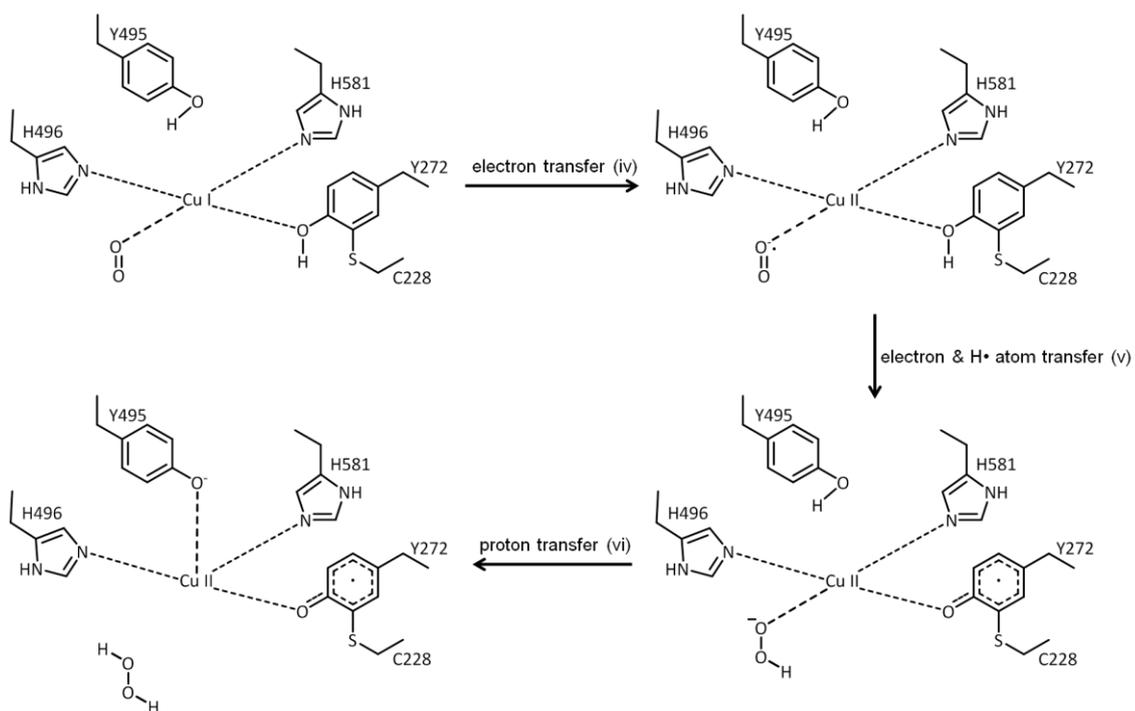
Figure 1.20. The oxidative half reaction of GO

Adapted from Messerschmidt et al. (2001).

1.6.4 Expression and purification of recombinant GO

In order to permit protein engineering studies, it is necessary to express GO in a host organism which does not contain endogenous GO activity. Expression has been previously optimised in the filamentous fungus *Aspergillus nidulans* (Baron et al., 1994) and in the methylotrophic yeast *Pichia pastoris* (Whittaker and Whittaker, 2000, Deacon et al., 2004). WT GO is initially translated as a precursor protein with an N-terminal pro region, including a secretion signal sequence, which is cleaved autocatalytically during maturation of the enzyme within the eukaryotic host (Rogers et al., 2000). In optimising the enzyme for expression in *E. coli*, Arnold and co-workers omitted the pro sequence from the initial clone and, following directed evolution, generated a variant which showed 18-fold increased expression compared to WT GO in *E. coli*, as well as improved stability and activity towards D-galactose (Sun et al., 2001). The mutant, later termed M1, contained six mutations: Ser10Pro and Met70Val in Domain 1; Pro136 (silent); Gly195Glu and Val494Ala in loops surrounding the active site; and Asn535Asp in a surface exposed loop of Domain 3.

Combination of these mutations with three silent mutations at positions 2, 3 and 5 (the 'N6' mutations) resulted in even greater enhancement of expression; this mutant was termed N6M1 (Deacon and McPherson, 2011). These silent mutations are predicted to reduce formation of mRNA secondary structures around the initiating codon resulting in more efficient initiation of translation. One of the many advantages of using *E. coli* as an expression host is that numerous different strains are available for optimising expression of the desired protein (Terpe, 2006). Deacon and McPherson (2011) compared expression levels in seven different strains and found BL21 Star (DE3) (Stratagene) gave the highest expression of soluble GO. Expression by autoinduction in 8ZYM-4x(5052) broth containing succinate (Studier, 2005) was also found to give the highest expression of all expression broths trialled (Deacon and McPherson, 2011).

Purification of GO from the host organism was traditionally carried out in a multi-step process involving ammonium sulfate precipitation (Avigad et al., 1962) and chromatographic separation using resins based on cellulose (Amaral et al., 1966) or agarose (Hatton and Regoeczi, 1976). Expression in *E. coli* enables use of polypeptide fusion partners, known as affinity tags, in protein purification. Examples include the widely used polyhistidine-tag or Strep II-tag which permit one-step purification procedures with minimal effect on the tertiary structure and biological activity of the protein (Terpe, 2003). Turner and co-workers first

reported purification of GO using a polyhistidine-tag added to the N-terminal end of the enzyme (Escalettes and Turner, 2008) enabling purification by nickel affinity chromatography. In order to avoid potential contamination of GO with metal ions other than copper, Deacon and McPherson (2011) added a Strep II-tag to the C-terminal end of GO-N6M1 permitting efficient purification using Strep-Tactin resin. Following optimisation of the cell lysis step, the expression of GO-N6M1 by autoinduction and purification by Strep-Tactin chromatography resulted in reported yields of up to 240 mg per litre of culture (Deacon and McPherson, 2011).

1.6.5 Uses of GO

1.6.5.1 Previous and existing applications

There are numerous examples of the use of WT GO in biotechnological applications:

- **Detection and quantification of D-galactose** is important in a number of different situations. For example, the condition galactosaemia, where one of the enzymes required for D-galactose metabolism is absent, requires the patient to closely monitor their D-galactose intake and blood and urine levels (Holton, 1990). There are also numerous examples where it is necessary to monitor the levels of D-galactose and other GO substrates in the food industry, for example in processing of sugar beet (Rorem and Lewis, 1962) and in the dairy industry (Adanyi et al., 1999, Mannino et al., 1999). Significant work has been undertaken to develop GO-based biosensors to enable accurate detection of D-galactose and there are a large variety of reported methods for immobilisation of GO within a biosensor (Taylor et al., 1977, Dicks et al., 1986, Vrbova et al., 1992, Sharma et al., 2006, Cevik et al., 2010, Charmantray et al., 2013, Kanyong et al., 2013, Kim et al., 2012).

- **Labelling of D-galactose-terminating glycans on cell surfaces** has been used by several groups in characterising and quantifying different cell types, such as in brain tissue (Singh and Kanfer, 1980). GO is also used in characterising changes in cell surface glycan expression, such as during human erythrocyte aging (Gattegno et al., 1981) or upon virus infection of fibroblasts (Critchley, 1974). The original methods involved oxidation of terminal galactosyl residues by GO followed by reduction of the resulting aldehyde by tritiated borohydride (Morell et al., 1966). However, more recently, methods have been developed where the aldehyde product is detected by fluorescent (Rannes et al., 2011) or chemiluminescent methods (Han et al., 2012).

- Labelling of cell surface glycans is also used in **cancer diagnosis** as the disaccharide D-galactose-β-(1-3)-*N*-acetyl-D-galactosamine (Gal-GalNAc) is expressed on the surface of colon cells in the early stages of carcinogenesis. After oxidation by GO, the aldehyde groups are detected using Schiff's reagent which produces a magenta coloration (Shamsuddin and Elsayed, 1988, Carter et al., 1997, Yang and Shamsuddin, 1996). The same method has also been used for detection of breast carcinoma (Chagpar et al., 2004).

- **Lymphocyte blastogenesis and generation of growth factors** is induced upon treatment with neuraminidase and GO (Novogrodsky and Katchalski, 1973) permitting research into lymphocytes and associated disorders such as leukaemia (Berzins et al., 1983, Zafar et al., 1981).

- **Modification of polysaccharides** by GO, often followed by further modifications by additional enzymes, provides access to a wide range of derivatives (Yalpani and Hall, 1982). For example, oxidation of the natural plant-based polymer guar has uses in paper strengthening (Delagrave et al., 2001) while derivatives of galactoxyloglucan have potential uses in abrasion resistant clothing or diagnostic strips (Parikka et al., 2012).

- **Generation of chemicals** such as unnatural sugars (Root et al., 1985); sugar-based polymers used in hydrogels, adsorbents and biorecognition agents (Liu and Dordick, 1999); and 5-*C*-(hydroxymethyl)hexoses, a potential group of artificial sweeteners (Mazur and Hiler, 1997), can all be carried out using WT GO as a catalyst which removes the need for protection of hydroxyl groups resulting in a more efficient process.

1.6.5.2 Potential future applications

In many ways GO represents an ideal enzyme for use in biotechnological applications as it is a stable enzyme, easily purified in high quantities and does not require expensive cofactors in order to function. Modification of the enzyme to introduce or improve activity towards substrates other than D-galactose has the potential to greatly expand the areas in which GO can be used. Introduction of activity towards D-monosaccharides other than D-galactose seems a realistic concept due to the similarity in structure to the native substrate.

D-glucose and D-xylose make up a significant proportion of the carbohydrate content of acid-treated corn stover, the parts of the maize plant remaining following harvest of the grain. The

sugars present in the acid-treated biomass are fermented to produce bioethanol. Measurement of the levels of D-glucose and D-xylose is important in monitoring the progress of sugar fermentation. This process is traditionally carried out using time-consuming HPLC-based methods. Quantification by an enzyme-based process using a variant of GO active against D-xylose and/ or D-glucose has the potential to significantly speed up this method and provide near real-time feedback information to help optimise bioethanol production and maximise yields. GO variants could also be used in the quantification of different sugars in the food industry as well as in biosensors, such as those for measuring blood glucose levels.

As well as D-galactose, cell surface glycans are frequently made up of D-glucose and D-mannose residues. GO variants showing activity towards these components could be used to label cell surfaces in a similar approach to that currently used for the labelling of D-galactose (Section 1.6.5.1) (Rannes et al., 2011). This has potential uses in diagnostic and research applications as different pathogens exploit different cell surface carbohydrates for entry into host cells (Disney and Seeberger, 2004, Doores et al., 2006).

D-arabinose has been identified as a building block in the generation of more complex molecules with potential roles in various pharmaceuticals such as noricumazole A, a potential drug against hepatitis C virus (Barbier et al., 2012b, Barbier et al., 2012a), and glycosidase inhibitors used in treatment of type 2 diabetes (Merino-Montiel et al., 2012). Regioselective oxidation of D-arabinose and other monosaccharides by GO variants has the potential to diversify the functionality of the sugar molecules facilitating their use in construction of more complex molecules and polymers. The oxidation of D-glucose and D-mannose at the C-6 position is not catalysed by any currently known enzymes (Rannes et al., 2011, Sun et al., 2002) so introduction of activity towards D-glucose or D-mannose into GO would provide access to enzyme-catalysed chemistry which is not currently available. Modification of polysaccharides containing D-glucose or D-mannose also has the potential to lead to new properties with various applications, for example in hair colouring and waving (Tsujino et al., 1991).

GO already shows some level of activity towards glycerol (Table 1.1) so enhancing this activity seems a realistic aim. The surge in biofuel development in the last decade has resulted in generation of large quantities of glycerol which forms the primary by-product in biodiesel production. As a result, glycerol prices have fallen significantly and the excess of available glycerol is now treated as a waste product. In order to reduce the economical and environmental impact of this waste storage and handling there is now great interest in

development of processes to convert glycerol into high-value products with uses in various industries, including generation of alternatives to petroleum-based chemicals (Johnson and Taconi, 2007). Oxidation of glycerol can lead to a variety of different products with a range of potential uses, some examples of which are shown in Table 1.3.

Table 1.3. Some products which can be generated by oxidation of glycerol

Product	Structure	Applications
Glyceraldehyde		Organic synthesis industry
Dihydroxyacetone		Organic synthesis and cosmetic industries (Hu et al., 2010)
Glyceric acid		Organic synthesis & pharmaceutical industries including generation of polymers
Malonic acid		Plastic manufacture, silver plating brightening agent (Paster et al., 2003)
Hydroxypyruvic acid		Production of amino acids
Tartronic acid		Used to make mesoxalic acid
Mesoxalic acid		Pharmaceuticals *e.g.* anti-HIV drug (Davis et al., 2000)
Lactic acid		Organic synthesis including petroleum product alternatives, food additives, textile/ leather treatment (Paster et al., 2003)
Pyruvic acid		Derivatives used as emulsifiers and in pharmaceuticals (Paster et al., 2003)
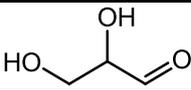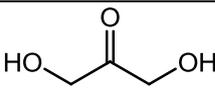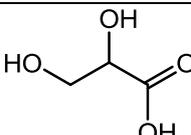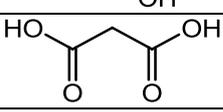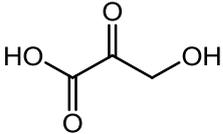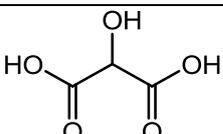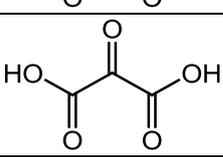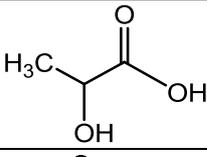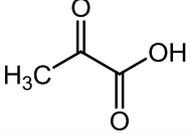
Many of these chemicals are currently generated from glycerol using metal catalysts such as gold or platinum (Garcia et al., 1995, Demirel-Gulen et al., 2005). Development of an efficient enzyme variant showing high activity and specificity would remove the need for these expensive catalysts. The oxidation of glycerol by WT GO has not been previously characterised so it is unclear where on the glycerol molecule oxidation occurs. However, it may be possible

to generate variants which produce different products depending on the mutations introduced.

Production of hydrogen peroxide through oxidation of a variety of different substrates also has potential industrial uses. These include the generation of bleaching agents in laundry detergents or toothpaste, or as a preservative in foodstuffs. Expanding the substrate repertoire of GO to include cheaply available substances such as glycerol and D-glucose is therefore also of interest in industrial hydrogen peroxide generation.

1.7 Aims of the project

The overall aim of this project was to identify mutations in GO which improve activity towards alternative substrates to D-galactose. Chapter 2 details the materials and methods which were used throughout the project. In Chapter 3 screening assays used by other groups to detect changes in GO activity were further optimised to improve a number of aspects, including sensitivity, to enable detection of activity towards all of the substrates used in screening (Figure 1.21).
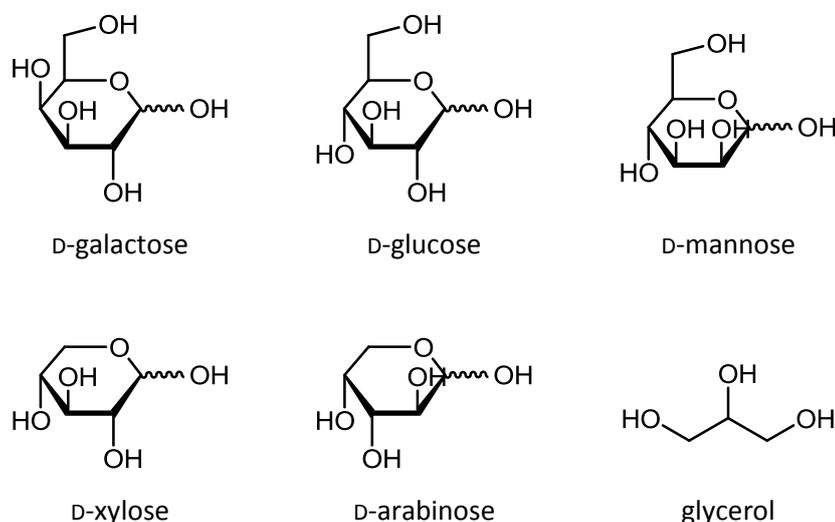
Figure 1.21. The six substrates for which activity will be screened
The structures of glycerol and the pyranose ring form of the five sugars are shown.

The substrate targets for this project were initially chosen due to their structural similarity to the native substrate D-galactose (D-glucose, D-mannose, D-arabinose and D-xylose) or because the WT enzyme already shows an appreciable level of activity towards the substrate (glycerol,

D-xylose). However, oxidation of each of the substrates also has potential biotechnological uses as detailed in Section 1.6.5.2.

Chapter 4 details the selection of amino acids around the GO active site which are thought to be involved in determining the substrate acceptance of GO; followed by design and generation of mutant libraries using randomised oligonucleotides generated using trimer phosphoramidites (Section 1.3.4). Screening of these libraries against the range of selected substrates (Figure 1.21) was then carried out using the assays developed in Chapter 3 and altered activities were confirmed with purified proteins.

Chapter 5 describes the further characterisation of nine of the variants identified in Chapter 4 with the aim of further understanding the basis behind the observed new activities. Also discussed is assessment of some of the conditions under which the new biocatalysts may function if further developed for use in biotechnological processes. The features of the enzymes explored include the effect of single versus double mutations, the pH optima of the variants, activity towards different forms of the substrate(s), the effect of increasing the oxygen concentration and Nuclear Magnetic Resonance (NMR) analysis of the oxidised products.

Finally, Chapter 6 compares the work presented here with that of other groups and discusses the future directions for this research.

Chapter 2 : Materials & Methods

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich or Fisher Scientific; randomised oligonucleotide primers were generated by Ella Biotech while other oligonucleotide primers were generated by Sigma-Aldrich.

2.1 Growth media

2.1.1 2TY broth

Tryptone (16 g/ L), yeast extract (10 g/ L) and NaCl (5 g/ L) were dissolved in deionised water and the solution was autoclaved at 121 °C, 15 psi for 20 minutes.

2.1.2 2TY agar plates

Tryptone (16 g/ L), yeast extract (10 g/ L) and NaCl (5 g/ L) were dissolved in deionised water, agar (12 g/ L) was added and the solution was autoclaved at 121 °C, 15 psi for 20 minutes. After autoclaving, the solution was cooled in a 50 °C water bath for 30 minutes before addition of antibiotic to the appropriate concentration (Section 2.1.7) followed by dispensing of approximately 25 ml into petri dishes, or 180 ml into 245 x 245 mm bio-assay plates (Thermo Scientific) under aseptic conditions.

2.1.3 ZYP-0.8G broth

ZY broth was made by dissolving tryptone (10 g/ L) and yeast extract (5 g/ L) in deionised water and autoclaving at 121 °C, 15 psi for 20 minutes. The following sterile solutions were added to 46.5 ml ZY broth before use (final concentrations are shown in brackets): 100 µl 1 M $MgSO_4$ (2 mM), 1 ml 40% (w/ v) D-glucose (0.8%), 2.5 ml NPSC component (50 mM NH_4Cl, 5 mM Na_2SO_4, 25 mM KH_2PO_4, 25 mM Na_2HPO_4).

2.1.4 8ZYM+5052+25 mM succinate broth

8ZY broth was made by dissolving tryptone (80 g/ L) and yeast extract (40 g/ L) in deionised water and autoclaving at 121 °C, 15 psi for 20 minutes. The following sterile solutions were added to 400 ml 8ZY broth before use (final concentrations are shown in brackets): 800 µl 1 M $MgSO_4$ (2 mM), 32 ml '5052 component' (0.5% (v/ v) glycerol, 0.05% (w/ v) D-glucose, 0.2% (w/ v) α-lactose monohydrate), 20 ml NPSC component (50 mM NH_4Cl, 5 mM Na_2SO_4, 25 mM KH_2PO_4, 25 mM Na_2HPO_4), 400 µl trace elements (50 µM $FeSO_4$, 10 µM $ZnSO_4$, 2 µM $CuSO_4$, 10 µM $MnSO_4$, 2 µM $Na_2B_4O_7$, 2 µM $(NH_4)_6MO_7O_{24}$), 20 ml 0.5 M succinate (25 mM).

2.1.5 BMMY broth

YES component was made by dissolving yeast extract (1 g/ L) and soytone (2 g/ L) in deionised water and autoclaving at 121 °C, 15 psi for 20 minutes. The following sterile solutions were added to 80 ml YES component before use (final concentrations shown in brackets): 10 ml 1 M potassium phosphate, pH 6.0 (100 mM), 10 ml 13.4% (w/ v) yeast nitrogen base with $(NH_4)_2SO_4$, without amino acids (1.34% (w/ v)), 200 μl 1 mM biotin (2 μM), 0.5 ml methanol (0.5% (v/ v)), 75 μl 100 mM $CuSO_4$ (75 μM).

2.1.6 SOC broth

Tryptone (20 g/ L), yeast extract (5 g/ L) and NaCl (0.5 g/ L) were dissolved in deionised water and autoclaved at 121 °C, 15 psi for 20 minutes. The following sterile solutions were added to 79 ml before use (final concentrations are shown in brackets): 10 ml 1 M $MgCl_2$ (0.1 M), 10 ml $MgSO_4$ (0.1 M), 1 ml 2 M D-glucose (20 mM).

2.1.7 Antibiotics

A 1000 x stock solution of each antibiotic was made up in deionised water, filter sterilised through a 0.2 μm syringe-end filter and stored in 1 ml aliquots at -20 °C. The final concentration used in cultures was 60 μg/ ml of carbenicillin or 50 μg/ ml kanamycin (Sambrook et al., 1989). Carbenicillin was used in place of ampicillin as it has the same resistance gene, is more stable, and the by-products generated upon its degradation are less toxic to cells, permitting denser cell growth.

2.2 *E. coli*

2.2.1 XL1 blue

Genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacI$_q$Z ΔM15 Tn10 (Tet$_r$)]*

Source: Stratagene (generated as detailed in Section 2.2.4)

XL1 blue cells were used for replication, purification and storage of plasmid DNA as they have been optimised by a number of mutations: (i) the *recA1* mutation maintains insert stability as unwanted recombination is reduced; (ii) non-specific digestion by endonuclease I is prevented

by incorporation of the *endA1* mutation; (iii) the *EcoK* endonuclease system is prevented from cleaving cloned DNA by incorporation of the *hsdR* mutation; (iv) termination of translation is reduced by the *supE44* mutation. The other mutations in the genotype were not necessary for the work presented here.

2.2.2 BL21 Star (DE3)

Genotype: *F- ompT hsdSB (rB-mB-) gal dcm rne131(DE3)*

Source: Invitrogen (generated as detailed in Section 2.2.4)

BL21 Star (DE3) cells were used for protein expression as they have been optimised by a number of mutations: (i) the *ompT* mutation in outer membrane protease VII as well as the absence of the *lon* protease reduces the proteolytic cleavage of the expressed protein; (ii) the *hsdSB* mutation reduces degradation of transformed plasmids which are 'foreign' to the host cell; (iii) mRNA degradation by ribonuclease E is reduced by the *rne131* mutation thereby increasing the yield of the recombinant protein; (iv) the presence of the DE3 lysogen carrying the gene for T7 RNA polymerase is required for expression in a T7 promoter-based system (as in the pET vectors used here, Section 2.3) which synthesises mRNA more rapidly than *E. coli* RNA polymerases and so permits higher expression levels.

2.2.3 Acella

Genotype: *F- ompT hsdSB (rB-mB-) gal dcm (DE3) ΔendA ΔrecA*

Source: Edge Bio (VH Bio)

Acella cells are based on BL21 Star (DE3) but with the additional deletion of the *endA* and *recA* genes resulting in no non-specific digestion by endonuclease I and reduced recombination. This means the cells are suitable for replication, purification and storage of plasmid DNA as well as protein expression.

2.2.4 Preparation of chemocompetent cells

A 2TY agar plate containing no antibiotic was streaked with a glycerol stock of the appropriate strain and incubated overnight in a static incubator at 37 °C. A single colony was picked and used to inoculate 5 ml 2TY broth which was then incubated overnight in a shaking incubator at 37 °C with shaking at 220 rpm. A 1 ml aliquot of the overnight culture was used to inoculate 50

50

ml pre-warmed 2TY broth which was then incubated in a shaking incubator at 37 °C with shaking at 220 rpm until an OD_{600} of ~0.4 was reached. The cells were then transferred to a sterile 50 ml Falcon tube and chilled on ice for 5 minutes. After centrifugation at 4 °C for 10 minutes at 4000 x g in a Hermle Z400 K centrifuge with a 220 97 VO2 rotor, the cell pellet was resuspended in 20 ml ice-cold TFB1 buffer (3 mM potassium acetate, 100 mM $RbCl_2$, 50 mM $MnCl_2$, 10 mM $CaCl_2$, 15% (v/ v) glycerol, pH adjusted to 5.8 with acetic acid, filter sterilised through a 0.2 μm membrane) and chilled on ice for 5 minutes. The cells were then centrifuged again for 10 minutes at 4000 x g, the supernatant discarded, and the tube wiped dry with tissue. The cells were resuspended in 2 ml ice-cold TFB2 buffer (10 mM $RbCl_2$, 75 mM $CaCl_2$, 10 mM MOPS, 15% (v/ v) glycerol, pH adjusted to 6.5 with potassium hydroxide, filter sterilised through a 0.2 μm membrane), chilled on ice for 15 minutes and then 50 μl aliquots pipetted into sterile pre-chilled 1.5 ml microcentrifuge tubes. The aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

2.3 The pET vectors

In this work the pET vectors pET101D (Invitrogen) and pET28c (Novagen) were used due to the presence of the T7 promoter upstream of the gene coding sequence. The BL21 Star (DE3) and Acella cells used for protein expression express T7 RNA polymerase which recognises the T7 promoter, resulting in transcription of the gene of interest (Studier and Moffatt, 1986). The pET vectors also carry a copy of the *lac* repressor gene which, when expressed, binds the *lac* operator inhibiting transcription of the gene of interest. Expression of T7 RNA polymerase is inhibited in the same way. To induce expression of T7 RNA polymerase and the gene of interest, the non-hydrolysable lactose analogue isopropyl-β-D-thiogalactopyranoside (IPTG) is added and displaces the *lac* repressor.

In the case of autoinduction, D-glucose is included in the broth as a primary carbon source as well as a repressor of induction until the culture has grown. Once levels of D-glucose are depleted, the lactose present in the broth can displace the repressor permitting protein expression without the requirement for monitoring of culture growth or addition of inducer (Studier, 2005).

2.4 DNA manipulation

2.4.1 PCR

In the majority of cases, PCR was carried out with Phusion DNA polymerase (New England Biolabs (NEB)) using the components supplied with the polymerase. Unless stated otherwise, the reaction was set up as detailed in Table 2.1 and thermal cycling carried out as detailed in Table 2.2 using a G-Storm GS2 thermal cycler.

Table 2.1. The reaction components of a PCR with Phusion DNA polymerase

Component	Final concentration
5 x HF buffer	1 x
dNTPs	200 µM of each
Forward primer	0.5 µM
Reverse primer	0.5 µM
Template DNA	1 ng/ µl
DMSO	3%
Deionised water	Up to final volume of 50 µl
Phusion DNA polymerase	0.02 U/ µl

Components were added in the order shown.

Table 2.2. The PCR reaction conditions

Stage	Temperature (˚C)	Time (seconds)
Hot start	98	120
25 cycles:		
Denature	98	20
Anneal	54	20
Extend	72	Depends on expected product
Final extension	72	600
Store	4	∞

When KOD DNA polymerase was used, the reaction conditions were the same except that no DMSO was used and 1.5 mM $MgCl_2$ was added along with the 1 x KOD reaction buffer.

2.4.2 Overlap extension

The overlap extension reaction used all of the reaction components shown in Table 2.1 apart from the template DNA and primers. Equimolar quantities of the two DNA fragments to be spliced were added to a total DNA concentration of 40 ng/ µl. The reaction conditions are shown in Table 2.3.

Table 2.3. The overlap extension reaction conditions

Stage	Temperature (˚C)	Time (seconds)
Hot start	98	120
10 cycles:		
Denature	98	20
Anneal	54	20
Extend	72	90
Addition of a further 0.02 U/ µg Phusion DNA polymerase		
10 cycles:		
Denature	98	20
Anneal	54	20
Extend	72	90
Final extension	72	600
Store	4	∞

2.4.3 Site-directed mutagenesis

In order to introduce a point mutation two complementary oligonucleotide primers containing the mutation, and designed to bind to opposite strands of the plasmid, were synthesised. Using the plasmid as a template, linear amplification temperature cycling was carried out in the same way as PCR with KOD or Phusion DNA polymerase (Section 2.4.1) to extend the two primers. Following temperature cycling, the methylated template DNA which did not contain the mutation was digested by *Dpn*I leaving the non-methylated nicked plasmid containing the introduced mutation on both strands. This was then transformed into XL1 blue cells, purified from transformants, and the mutation confirmed by DNA sequencing (Section 2.4.15).

2.4.4 Agarose gel electrophoresis

The agarose gel was prepared and run using a HU6 Mini (small gel) or HU10 Mini-Plus (large gel) horizontal gel unit (Scie-Plas, Harvard Bioscience). Agarose gels of the desired percentage were made by combining the appropriate volume of Tris-acetate ethylenediaminetetraacetic acid (TAE) buffer (40 mM Tris-acetate, 1 mM ethylenediaminetetraacetic acid (EDTA), made up as a 50 x stock, pH 8.0) with agarose in a 250 ml Duran bottle and heating in the microwave for approximately 2 minutes until the agarose completely dissolved and the solution was bubbling. The agarose solution was cooled in a 50 °C water bath for 30 minutes before addition of the appropriate volume of SYBRsafe DNA gel stain (Invitrogen). The molten agarose was then poured into the casting tray in a casting unit, a comb added and the gel left to set at room temperature.

53

To run, the casting tray containing the agarose gel was removed from the casting unit and placed into the running chamber. The comb was removed and TAE buffer was added until the gel was submerged. DNA samples were prepared by addition of 6 x gel loading dye (NEB) and then loaded into the wells of the agarose gel. A 1 µl aliquot of the appropriate DNA size ladder was also combined with 1 µl 6 x gel loading dye and added to a well: either 1 kb DNA size ladder or 100 bp DNA size ladder (NEB) were used (Figure 2.1).
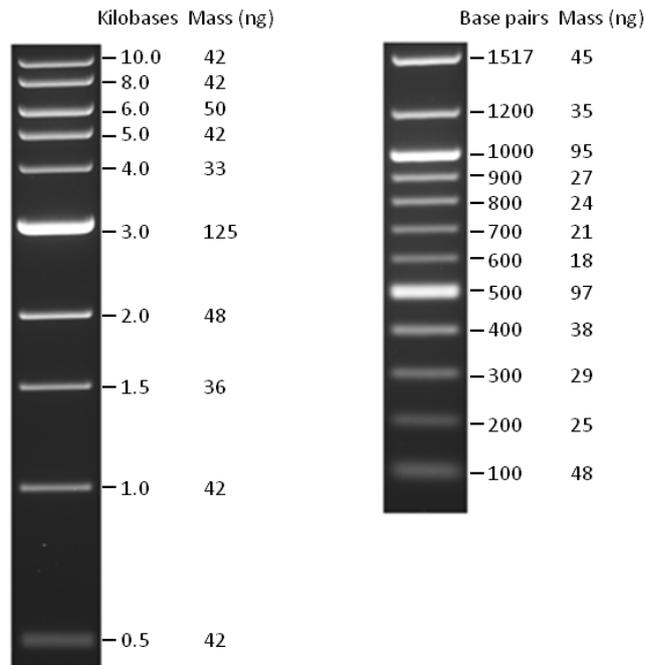
Figure 2.1. DNA size ladders used in agarose gel electrophoresis

The 1 kb (left) and 100 bp (right) DNA size ladders were used to estimate the size of DNA analysed by agarose gel electrophoresis as well as confirm the concentration measured using the Nanodrop-1000 spectrophotometer by comparison of the band intensities (Section 2.4.14).

Once samples were loaded the running chamber was connected to the power unit and a voltage of 100 V applied for 40 minutes, or longer if better resolution was required. DNA was either visualised under UV light and photographed using an AlphaImager system (Alpha Innotech) or if it was to be used for downstream applications, was visualised using a Safe Imager (Invitrogen) which is a blue light transilluminator and does not lead to DNA damage as UV light does.

2.4.5 Restriction digestion

Restriction digestion was carried out with *Mfe*I-HF™ and *Not*I-HF™ restriction enzymes (NEB). Each digestion reaction contained 1 x Buffer 4, 0.2 µg/ µl bovine serum albumin, 1 U/ µl of each of the restriction enzymes and 0.1 µg/ µl plasmid DNA and was incubated at 37 °C for >4 hours.

2.4.6 Dephosphorylation

By dephosphorylating the digested plasmid, the 5' phosphate is removed and self-ligation is prevented (Sambrook et al., 1989). Dephosphorylation was carried out by adding 0.1 U/ µl Antarctic Phosphatase and 1 x Antarctic Phosphatase reaction buffer (NEB) to the completed restriction digestion reaction (Section 2.4.5) and incubating at 37 °C for 15 minutes. The phosphatase enzyme was heat inactivated by incubation at 65 °C for 5 minutes in a G-Storm GS2 thermal cycler.

2.4.7 Ligation

Ligation of plasmid and insert was carried out using T4 DNA ligase from NEB or from Roche. The appropriate amounts of the digested, dephosphorylated plasmid and the digested insert were combined with the appropriate 1 x T4 DNA ligase reaction buffer and T4 DNA ligase and made up to a total reaction volume of 10 µl with deionised water. Incubations were carried out in 200 µl PCR tubes at 4 °C.

2.4.8 Transformation into *E. coli* cells by heatshock

A 50 µl aliquot of competent cells (Section 2.2.4) was defrosted on ice for at least 10 minutes before addition of 10-50 ng plasmid DNA in a total volume of no more than 5 µl. The mixture was gently mixed and incubated on ice for 30 minutes before transfer to a 42 °C water bath for 45 seconds. The microcentrifuge tube was then transferred immediately to ice for a minimum of two minutes before addition of 1 ml 2TY broth, pre-warmed to 37 °C. The culture was then placed in a shaking incubator at 37 °C with shaking at 220 rpm to permit recovery and expression of the antibiotic resistance protein. After one hour, a 100 µl aliquot was plated onto an agar plate containing the appropriate antibiotic and the remaining culture was centrifuged at 4000 rpm for 4 minutes in a microcentrifuge. All but 100 µl of the supernatant was discarded and the remaining broth used to resuspend the bacterial pellet which was then

plated out onto a second agar plate containing the appropriate antibiotic. Both plates were inverted and incubated at 37 °C in a static incubator overnight.

2.4.9 Transformation into *E. coli* cells by electroporation

A 25 µl aliquot of electrocompetent cells was defrosted on ice for 5-10 minutes and then added to 40 ng plasmid DNA in 1.6 µl deionised water in a chilled microcentrifuge tube. The mixture was then carefully transferred to a chilled, sterile electroporation cuvette with a 0.2 cm gap (BioRad). The samples were electroporated in a MicroPulser Electroporator (BioRad) at 2.5 V and 2 x 750 µl SOC broth, at room temperature, was added immediately. The culture was transferred to a 14 ml round-bottomed tube and placed in a shaking incubator at 37 °C with shaking at 220 rpm to permit recovery and expression of the antibiotic resistance protein. After one hour the culture was plated out onto 245 x 245 mm bio-assay agar plates containing the appropriate antibiotic, inverted and incubated at 37 °C in a static incubator overnight.

2.4.10 Colony PCR

A 200 µl (yellow) pipette tip was used to collect a single colony and swirled gently in 100 µl sterile deionised water before being used to inoculate 3 ml 2TY broth containing the appropriate antibiotic. This culture was incubated overnight at 37 °C with shaking at 220 rpm to enable purification of the DNA by miniprep (Section 2.4.11). Meanwhile the inoculated water was heated to 99 °C for 5 minutes in a G-Storm GS2 thermal cycler, cooled to room temperature and centrifuged at 13000 rpm in a microcentrifuge to pellet cell debris leaving the DNA in suspension. A 1 µl aliquot of this suspension was added to 24 µl of PCR mastermix (1 x green GoTaq Flexi buffer, 2 mM $MgCl_2$, 0.2 mM PCR nucleotide mix, 0.025 U/ µl GoTaq HotStart polymerase (Promega), 0.2 µM of each of the primers used), mixed and PCR carried out according to the parameters outlined in Table 2.4. The products were then analysed by agarose gel electrophoresis to confirm the presence or absence of the desired product.

Table 2.4. The reaction conditions for colony PCR

Stage	Temperature (˚C)	Time (seconds)
Hot start	95	300
30 cycles:		
Denature	95	30
Anneal	55	30
Extend	72	Depends on expected product
Final extension	72	300
Store	4	∞

2.4.11 Purification of DNA from *E. coli* by miniprep

The QIAprep® Spin miniprep kit (Qiagen) was used to purify DNA from *E. coli* cultures. A single colony containing the plasmid to be purified was used to inoculate 5 ml 2TY broth containing the appropriate antibiotic and incubated overnight at 37 °C in a shaking incubator at 220 rpm. The cells were harvested by centrifugation at 7500 x *g* for 5 minutes in a Hermle Z400 K centrifuge with a 220 97 VO2 rotor and the supernatant drained by inverting the tube. The pellet was completely resuspended in 250 μl buffer P1 (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 μg/ ml ribonuclease A, 1 μl/ ml LyseBlue reagent) using a vortex before transfer to a 1.5 ml microcentrifuge tube. 250 μl lysis buffer P2 (200 mM NaOH, 1% (w/ v) SDS) was added and the solution mixed by inversion until a homogeneous blue colour was achieved. 350 μl neutralisation buffer N3 (4.2 M guanidinium chloride, 0.9 M potassium acetate, pH 4.8) was then added and immediately mixed by inversion until all of the blue colour had disappeared indicating that the SDS had precipitated and the solution had been neutralised. The tube was then centrifuged at 13000 rpm in a microcentrifuge for 15 minutes to pellet all of the cell debris. The supernatant was applied to a QIAprep spin column, which contains a silica gel membrane which can bind up to 20 μg DNA in the presence of high concentrations of chaotropic salts such as guanidinium chloride, and centrifuged for 45 seconds. The flow-through was discarded and the silica gel membrane was washed to remove impurities by addition of 750 μl buffer PE (10 mM Tris-HCl, pH 7.5, 80% (v/ v) ethanol) and centrifugation for 45 seconds. The flow-through was discarded and the spin column with DNA adsorbed was spun for a further 60 seconds to remove any residual ethanol which would prevent loading of the sample onto an agarose gel and inhibit any future enzymatic reactions. The column was transferred to a sterile 1.5 ml microcentrifuge tube, 30 μl of deionised water was carefully

pipetted onto the membrane and after 60 seconds the column was centrifuged for 2 minutes to elute the DNA which was then stored at -20 °C.

2.4.12 Purification of DNA from an agarose gel

The QIAquick® Gel Extraction kit (Qiagen) was used to purify DNA from agarose gels. The DNA sample was first separated by agarose gel electrophoresis and visualised using a Safe Imager (Invitrogen) (Section 2.4.4). DNA was not visualised by UV light as this can cause damage which significantly reduces the efficiency of downstream applications such as overlap extension, and ligation (lifetechnologies.com). The band of interest was excised using a clean, sharp scalpel and placed in a 1.5 ml microcentrifuge tube or 7 ml bijou container depending on the size of the gel slice. The agarose gel was solubilised by addition of 3 gel volumes of buffer QG (5.5 M guanidine thiocyanate, 20 mM Tris-HCl, pH 6.6) at 50 ˚C, which contains high levels of the chaotropic salt guanidine thiocyanate to ensure binding of DNA to the QIAquick membrane. To ensure the highest efficiency binding the pH of the solution was brought below pH 7.5 by addition of 10 µl of 3 M sodium acetate, pH 5.0 and 1 gel volume of isopropanol was added. The sample was then applied to the QIAquick column and centrifuged for 1 minute at 13000 rpm in a microcentrifuge. The flow-through was discarded and the column was washed to remove impurities by addition of 750 µl of buffer PE (10 mM Tris-HCl, pH 7.5, 80% (v/ v) ethanol) and centrifugation for 1 minute at 13000 rpm. To elute the DNA the column was placed in a clean 1.5 ml microcentrifuge tube and 30 µl sterile, deionised water added to the centre of the membrane. After 60 seconds at room temperature this was centrifuged for 2 minutes at 13000 rpm, the spin column discarded and the DNA solution stored at -20 ˚C.

2.4.13 Purification of PCR products

The QIAquick® PCR Purification kit (Qiagen) was used to purify DNA products following PCR, overlap extension or ligation where separation from other DNA species was not necessary but removal of reaction components was required for the highest efficiency of downstream processes, such as digestion or transformation. The protocol was the same for that of the QIAquick® Gel Extraction kit except that the PCR reaction was mixed with 5 volumes of buffer PB (5 M guanidinium chloride, 30% (v/ v) isopropanol) containing pH indicator, rather than buffer QG, before being applied to the QIAquick column.

2.4.14 Determination of DNA concentration

DNA concentration was determined using a Nanodrop-1000 spectrophotometer (Thermo-Scientific). The spectrophotometer was initialised using 10 µl deionised water, then deionised water was used to take a blank reading. A 2 µl aliquot of the DNA solution was then applied to the sample pedestal and the concentration determined from the A_{260} reading according to the Beer-Lambert equation: $A_{260} = \varepsilon Cl$ where ε is the extinction coefficient (50 (ng/ µl)$^{-1}$ cm^{-1}), C is the DNA concentration in ng/ µl and l is the path length in cm.

DNA concentration was also confirmed by comparing band intensity on agarose gels with that of the 1 kb or 100 bp DNA size ladders (NEB) (Figure 2.1).

2.4.15 DNA sequencing

DNA sequencing was carried out by GATC Biotech or Beckman Coulter Genomics. Plasmid DNA and sequencing primers (Table 2.5) were provided and sequencing data was returned as .ab1 files. Results were analysed using BioEdit Sequence Alignment Editor v. 7.0.5.3 and Vector NTI Advance 11.5.2 (Invitrogen).

Table 2.5. Primers used for sequencing GO

Primer name	Sequence
T7F	5' taatacgactcactataggg 3'
T7R	5' ccgctgagcaataactag 3'
GOF1	5' ccgacatcgggacgagtcc 3'
GOF2	5' gccacaaccaacgcccac 3'
GOR1	5' gtgggcgttggttgtggc 3'
GOR2	5' ggactcgtcccgatgtcgg 3'

2.4.16 Preparation of glycerol stocks

A single colony of XL1 blue cells containing the plasmid of interest was used to inoculate 5 ml of 2TY broth containing the antibiotic. This culture was incubated at 37 ˚C with shaking at 220 rpm overnight. The following morning the culture was centrifuged at 4000 x *g* in a Hermle Z400 K centrifuge with a 220 97 VO2 rotor at 4 ˚C for 10 minutes to pellet the cells but not damage them. The pellet was then gently resuspended in 3 ml sterile 2TY broth containing 80% (v/ v) glycerol and aliquoted into three cryovials. These were then flash frozen in liquid nitrogen and stored at -80 ˚C.

2.5 Protein expression and purification

2.5.1 Expression by autoinduction

A single colony of BL21 Star (DE3) cells containing the plasmid of interest was used to inoculate 2 ml of ZYP-0.8G broth containing the appropriate antibiotic. This culture was incubated at 37 ˚C with shaking at 220 rpm for >6 hours after which time a 200 µl aliquot was used to inoculate 400 ml 8ZYM+5052+25 mM succinate broth in a 2 L baffled flask. This autoinduction culture was then incubated at 25 ˚C with shaking at 220 rpm for 64 hours.

2.5.2 Preparation of the crude extract

Protein-expressing cells were harvested by centrifugation at 7500 x g at 4 ˚C for 15 minutes in a Sorvall RC-5B centrifuge and SLA-1500 rotor and the pellet stored at -80 ˚C. Lysis buffer was prepared by dissolving 50 mM HEPES pH 8.0, 25% sucrose (w/ v), 5 mM $MgCl_2$ and 1% Triton X-100 in deionised water, filter sterilising through a 0.2 µm membrane and chilling to 4 °C. Directly before use 1 mg/ ml hen egg white lysozyme, 20 U/ ml Omnicleave endonuclease (Epicentre, USA) and two c*O*mplete® protease inhibitor cocktail tablets (Roche) per 100 ml were added. The cells were lysed by dispersing the frozen pellet in 20 ml lysis buffer and stirring in the cold room for 45-60 minutes. The lysate was then centrifuged at 30000 x g at 4 ˚C for 45 minutes in a Sorvall RC-5B centrifuge and Fibrelite F21 8x50 rotor to separate the insoluble fraction (the pellet) from the soluble fraction (the supernatant). The soluble fraction was dialysed in 10 kDa MWCO dialysis tubing against 5 L phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4) for 1 hour at 4 °C and then 5 L PBS for 2-3 hours at 4 °C and then centrifuged at 30000 x g at 4 ˚C for 20 minutes in a Sorvall RC-5B centrifuge and Fibrelite F21 8x50 rotor to remove any precipitate. Following centrifugation, the crude extract was filtered through a 0.2 µm syringe-end filter before loading onto the Strep-Tactin resin.

2.5.3 Strep-Tactin affinity chromatography

The entire purification was carried out in the cold room. The four buffers used are shown in Table 2.6.

Table 2.6. Buffers used in Strep-Tactin chromatography purification

Buffer	Components
Wash buffer 1 (PBS)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4
Elution buffer	wash buffer 1 containing 5 mM desthiobiotin, pH 7.4
Regeneration buffer	100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM 2-(4-hydroxyphenylazo)benzoic acid
Wash buffer 2	100 mM Tris-HCl, pH 10.5, 150 mM NaCl

All buffers were filtered through a 0.2 µm membrane and chilled to 4 °C before use.

A 0.5-2.0 ml aliquot of Strep-Tactin Superflow high capacity resin (IBA) was dispensed with a Pasteur pipette into a 2 ml disposable polystyrene column (Thermo Scientific) and equilibrated with >10 column volumes (CV) of wash buffer 1. The filtered crude lysate was then applied to the column and the flow-through collected. Unbound proteins were washed off the column with >10 CV of wash buffer 1 and the end cap of the polystyrene column was attached to pause the flow. Elution buffer was then added to fill the polystyrene column and the resin mixed with a Pasteur pipette to ensure maximal elution of the Strep II-tagged GO. After the Strep-Tactin resin had settled, the end cap was removed and elution fractions collected in 2 CV volumes. Extra elution buffer was added to the column and collected in the same way to yield a total elution volume of 12 CV.

The following day, the six elution fractions were pooled and dialysed in 10 kDa MWCO tubing against 1 L 20 mM PIPES, pH 6.1 at 4 °C for 1 hour to remove the desthiobiotin. Dialysis was then carried out for 3 hours against 3 L 20 mM PIPES, pH 6.1 containing 1 mM $Cu(NO_3)_2$ at 4 °C to ensure complete incorporation of copper into the active site of the enzyme. The protein solution was finally dialysed against 5 L 100 mM sodium phosphate, pH 7.0 at 4 °C for 1 hour and then for a further 2-3 hours after a buffer change.

The Strep-Tactin resin was regenerated by addition of >10 CV regeneration buffer which was removed by addition of 5 CV wash buffer 2 before equilibration with >10 CV wash buffer 1 ready for the next purification. In order to avoid any potential contamination of one mutant with another, fresh resin was used for purification of each variant.

2.5.4 Concentration and storage of protein samples

Following dialysis, protein samples were filtered through a 0.2 μm syringe-end filter and concentrated using a centrifugal concentrator (Vivaspin) with a polyethersulfone membrane and 10 kDa MWCO. Centrifugation was carried out at 8000 x g in a Hermle Z400 K centrifuge with a 220 97 VO2 rotor at 4 ˚C for 10 minutes at a time until the protein solution volume was less than 1 ml. The solution was then divided into aliquots of 50-200 μl, flash frozen in liquid nitrogen and stored at -80 °C.

2.6 Protein analysis

2.6.1 Determination of protein concentration

Protein concentration was determined using a UV-2401PC spectrophotometer (Shimadzu) controlled by UV Probe v.2.2 software. A baseline reading was recorded with 100 mM sodium phosphate, pH 7.0 before a spectrum was recorded from 360 to 250 nm for a 10-fold dilution of the purified protein. A spectrum was recorded rather than a single wavelength reading at 280 nm in order to confirm that the λ_{max} was at 280 nm and not at another wavelength such as 260 nm which would suggest DNA contamination. To determine the molar concentration of GO the Beer-Lambert equation was used: $A_{280} = \epsilon_{280}Cl$ where ϵ is the extinction coefficient of GO at 280 nm (104900 M^{-1} cm^{-1}) (Kosman et al., 1974), C is the molar protein concentration and l is the path length in cm. Therefore, for an absorbance of 1:

$$[GO] \quad = A_{280} / \epsilon l$$

$$= 1/ (104900 \ M^{-1} \ cm^{-1} \ x \ 1 \ cm)$$

$$= 9.53 \ x \ 10^{-6} \ M$$

To convert to units of mg/ ml of enzyme this is multiplied by the molecular mass of the Strep II-tagged protein (70435), as mass (g) = number of moles x molecular weight. As the Strep II-tag contains a Trp residue, the value then needs to be corrected as Trp residues have a significant effect on A_{280} measurements (Baron et al., 1994, Rannes et al., 2011). GO contains 16 Trp residues so the concentration was multiplied by a correction factor of 16/ 17. Therefore, the concentration in mg/ ml is determined by multiplying the absorbance reading by 9.53 x 10^{-6} x 70435 x 16/ 17 = **0.63**.

2.6.2 Analysis by SDS-PAGE

Protein samples were separated according to their motility through a SDS-polyacrylamide gel by electrophoresis following the commonly used protocol developed by Laemmli (1970).

2.6.2.1 Preparation of soluble samples

To ensure visualisation, denaturation and easy loading, samples were mixed with 4 x loading buffer (200 mM Tris-HCl, pH 6.8, 20% (v/ v) glycerol, 8% (w/ v) SDS, 0.4% (w/ v) bromophenol blue, 20% (v/ v) β-mercaptoethanol) and heated at >70 °C for 5 minutes.

2.6.2.2 Preparation of insoluble samples

The insoluble fraction obtained from the lysis of the cell pellet in 20 ml lysis buffer (Section 2.5.2) was resuspended in 20 ml 10% (v/ v) lysis buffer. A 1 ml aliquot of the resuspended mixture was then centrifuged at 13000 rpm in a microcentrifuge for 5 minutes, the supernatant was discarded and the pellet resuspended once more in 10% (v/ v) lysis buffer. This was repeated three more times to wash the insoluble pellet and remove any soluble proteins before the pellet was resuspended in enough 4 x loading buffer (Section 2.6.2.1) to make the total volume up to 1 ml. The resuspended pellet was then heated at >70 °C for 5 minutes.

2.6.2.3 Preparing and running the gel

Using a Bio-Rad PROTEAN casting system a 15% resolving gel (15% (v/ v) acrylamide (Severn Biotech Ltd.), 375 mM Tris-HCl, pH 8.8, 0.1% (w/ v) SDS, 0.1% (w/ v) ammonium persulfate (APS), 0.04% (v/ v) N,N,N',N',-tetramethylethylenediamine (TEMED)) was poured between casting plates using a Pasteur pipette and allowed to polymerise overlaid with ethanol to ensure a flat interface. The ethanol was then removed by rinsing with deionised water and a stacking gel (5% (v/ v) acrylamide, 125 mM Tris-HCl, pH 6.8, 0.1% (w/ v) SDS, 0.1% (w/ v) APS, 0.1% (v/ v) TEMED) poured using a Pasteur pipette before addition of a 10- or 15-well comb to create sample wells. Protein samples were loaded along with PageRuler™ Unstained Protein Size Ladder (Fermentas) (Figure 2.2) to permit estimation of the size of the protein bands. Gels were electrophoresed at 200 V in SDS-PAGE running buffer (25 mM Tris, 250 mM glycine, 0.1% (w/ v) SDS, pH 8.3) until the dye front had just run off the bottom of the gel (approximately 45 minutes). The gel was stained with Instant*Blue* (Expedeon); bands were visible within about 10

minutes and the gel was photographed after 1 hour using an AlphaImager system (Alpha Innotech).
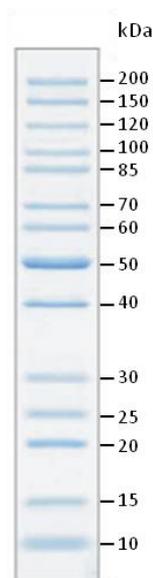
Figure 2.2. PageRuler™ unstained protein size ladder used in SDS-PAGE

2.6.3 Kinetic analysis of enzyme activity

In order to measure the activity of GO, a coupled assay was carried out in which hydrogen peroxide generated by the GO-catalysed reaction is used by HRP to oxidise two molecules of ABTS to generate a coloured product. This product has a λ_{max} of 414 nm so generation can be followed spectrophotometrically. The assay mix shown in Table 2.7 was equilibrated to 25 °C in a water bath before addition of 950 µl to a 1 ml plastic cuvette.

Table 2.7. The assay mix used to measure GO activity spectrophotometrically

Component	Concentration in assay mix
Sodium phosphate, pH 7.0	100 mM
ABTS	1 mM
HRP	30 U/ ml
Sugar substrate	varied

A 50 µl aliquot of GO at an appropriate concentration to give a measurable rate was added. The cuvette was immediately covered with Parafilm and mixed by inversion and then immediately placed into the UV-2401PC spectrophotometer (Shimadzu). Absorbance at 414 nm was typically measured for 30-60 seconds using the UV Probe v.2.2 software and the plot

of A_{414} against time was used to determine the rate in absorbance units per second. This rate was then divided by two as one molecule of hydrogen peroxide is used to generate two molecules of coloured product. According to the Beer-Lambert equation $A_{414} = \varepsilon_{414}Cl$, where the extinction coefficient of ABTS radical at 414 nm is 31300 M^{-1} cm^{-1}, the rate in absorbance units per second was divided by 31300 and then multiplied by 10^{6} to convert the units to μM per second.

Kinetic parameters were determined by measuring activity over a range of substrate concentrations and fitting the plot of rate against substrate concentration to the Michaelis-Menten equation using non-linear regression in OriginPro 9.0:

$$\text{Rate} = (V_{max}[S])/ (K_M+[S])$$

Where V_{max} (μM hydrogen peroxide/ second) is the maximum rate of the enzyme, [S] (mM) is the substrate concentration and the Michaelis constant, K_M (mM) is the [S] at half V_{max} and can give an indication of the affinity of the enzyme for the substrate. The turnover number, k_{cat} (s^{-1}) was determined by dividing V_{max} by the enzyme concentration (μM).

2.6.4 Determination of specific activity

The assay to determine specific activities of the variants with the different substrates was carried out in the same way as kinetic analysis (Section 2.6.3) except that activity was only measured at one substrate concentration (Table 2.8). The rate given in units of μM per minute was divided by the mg of GO added and then divided by 1000 ml to convert to units of μmol hydrogen peroxide per minute per mg GO.

Table 2.8. Concentrations of substrates used to measure substrate specificity

Substrate	Concentrations used (M)
D-galactose	0.6, 1.0
D-arabinose	2.0, 1.0
D-glucose	2.0, 1.0
D-mannose	2.0, 1.0
D-xylose	1.5, 1.0
Glycerol	1.5, 1.0

pH profiles were determined by measuring specific activities in buffers covering a range of pH. The protocol was the same except that the sodium phosphate, pH 7.0 was substituted for different buffers as detailed in Section 5.6.

2.6.4.1 Determination of specific activity at varied oxygen concentrations

In order to measure the specific activities at different oxygen concentrations, the assay mixes were prepared differently. For measuring activity with a reduced oxygen concentration, the assay mix (Table 2.7), in a round bottomed flask sealed with a Suba Seal rubber septum, was degassed using a vacuum and nitrogen bubbled through overnight to remove all oxygen. The cuvette to be used for the activity measurement had a round opening and was sealed with a Suba Seal rubber septum before nitrogen was used to purge the cuvette of oxygen. The degassed assay mix was removed from the round bottomed flask with a Hamilton syringe and immediately transferred to the nitrogen-filled assay cuvette. The enzyme solution, which had been briefly bubbled with nitrogen to reduce the oxygen concentration, was transferred to the assay mix using a Hamilton syringe. After mixing by inversion, product generation was followed spectrophotometrically.

In order to measure activity at increased oxygen concentration, the assay mix (Table 2.7) was transferred to a cuvette with a round opening and the cuvette sealed with a Suba Seal rubber septum. Oxygen was bubbled though the assay mix for a minimum of one hour, with an additional needle to permit exit of gases, before addition of the enzyme solution with a Hamilton syringe. After mixing by inversion, product generation was followed spectrophotometrically.

2.6.5 NMR spectroscopy

2.6.5.1 Preparation of samples

The complete reaction mix, or 200 µl aliquots, were freeze-dried overnight to remove water. The dried samples were then redissolved in an appropriate volume of deuterium oxide (700 µl for the 200 µl aliquots) and transferred to 5 mm NMR tubes.

2.6.5.2 Acquisition and analysis of ^{1}H-NMR and ^{13}C-NMR spectra

^{1}H and ^{13}C NMR spectra were recorded on a Bruker Avance 500 instrument (at 500 MHz and 125 MHz respectively) at 25 °C. The data were processed using MestReNova software (mestrelab.com). Signals were assigned by Dr Bruce Turnbull using a combination of COSY and HMQC 2D spectra.

Chapter 3 : Development and Optimisation of Screening Assays to Detect Oxidase Activity

3.1 Introduction

In order to identify variants showing alterations in activity compared to WT it is necessary to screen libraries of mutants. Many examples of different assays exist to screen for changes in enzyme activities. In screening assays it is desirable to use selection techniques where the novel activity is essential for cell survival or growth. However, this is not possible in the majority of cases and therefore high-throughput screens must be developed. A number of factors must be considered when choosing or designing a screening method:

- Reproducibility and accuracy so that results are reliable and minimal repeats are required.

- Affordability of the assay components to reduce the financial impact of the process on laboratory budgets. For example, use of antibodies or expensive substrates will significantly increase the cost of screening.

- Minimal use of expensive equipment such as NMR or LC-MS which can significantly reduce the feasibility of screening larger libraries.

- Maximal throughput, for example, screening involving NMR or LC-MS can be considered very low-throughput compared to FACS or phage display methods which are ultra-high throughput.

- Sensitivity of the screen so that small changes in activity can be detected.

- Safety of the components and equipment used, for example avoiding use of radioactive or carcinogenic substances where possible to remove unnecessary risk of exposure to researchers.

Unfortunately it is rarely possible to fulfil all of these criteria in development of an assay. For screening GO libraries an agar plate-based screen has already been used successfully by different groups (Delagrave et al., 2001, Escalettes and Turner, 2008). The basic screen uses the coupled assay detailed in Figure 1.6. The mutant library is transformed into an expression strain of *E. coli* which is plated out onto membranes and grown on agar plates containing the appropriate antibiotic. Cytoplasmic expression of the library variants is either induced by transfer of the membrane to an induction plate (Delagrave et al., 2001), or occurs by autoinduction during a longer growth period at lower temperature (Escalettes and Turner, 2008). Unlike Delagrave et al*.* (2001), Escalettes et al*.* (2008) duplicated the plate of colonies onto an additional membrane. The duplicate colonies were used in later inoculations based on

the results of screening. Both protocols exposed the colonies to chloroform vapour in order to disrupt the cell membranes and release GO. The colonies, on the membrane support, were transferred to an assay mix containing the substrate of interest, CuSO$_4$, HRP, and a chromogenic substrate which generates a coloured product when oxidised by HRP in the presence of hydrogen peroxide. Colour generation was then observed visually for colonies expressing variants active against the substrate provided. The bulk of this chapter details experiments to optimise the agar plate-based assay for maximal sensitivity as well as reducing the cost and improving the safety of the procedure.

Optimisation of a microtitre plate-based screen for confirmation of novel activities identified in initial agar plate-based screening is also described. Microtitre plate-based screening is very widely used in various bioscience applications including drug discovery (Brik et al., 2006) and screening conditions required for different cellular processes (Tomasini-Johansson et al., 2012). It also has the potential for automation if robotic systems are available. The basic assay involves growth of a different library variant in each well of the microtitre plate, followed by induction using IPTG and harvesting of the cell pellet by centrifugation. The pellet is then lysed and aliquots of the lysate added to different assay plates containing the same assay mix as used in the agar plate-based screen. Positive results can be detected visually, or spectrophotometrically.

Some of the cloning described in this chapter (Section 3.2.3) is carried out using a method called Circular Polymerase Extension Cloning (CPEC) which was designed by Quan and Tian (2009). This sequence-independent method replaces the restriction digestion and ligation procedure commonly used to insert a coding sequence into a plasmid. It is convenient to use when it is not possible (or desirable) to engineer restriction sites at specific positions in a sequence. This relatively simple method extends overlapping regions between the insert and vector in much the same way as overlap extension (McPherson and Moller, 2006), but the overlaps occur at both ends of the linearised plasmid and insert resulting in a complete circular plasmid which can then be used to transform cells (Figure 3.1). This technique was selected to avoid the need to engineer restriction sites at the desired position.

69
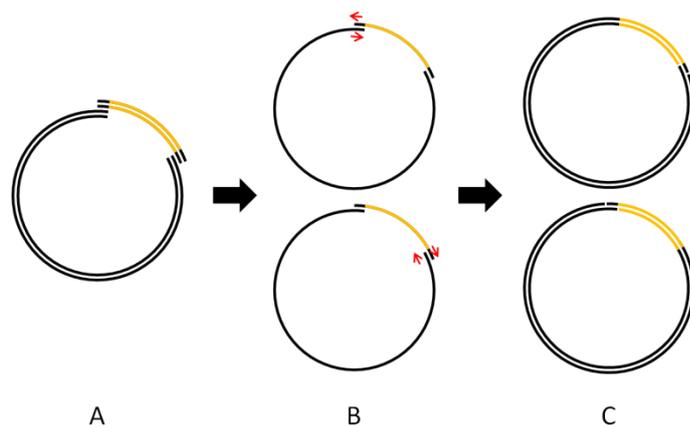
Figure 3.1. Schematic illustrating CPEC

Adapted from Quan and Tian (2009). Appropriate PCR primers are designed and used to generate the linear plasmid and insert with 20-30 bp overlapping regions (**A**). The plasmid and insert are denatured and annealed and the sequence is extended using a thermostable polymerase (**B**). This results in plasmids containing a single nick in each strand which are ready for transformation of *E. coli* (**C**).

3.2 Optimisation of the agar plate-based assay

3.2.1 The best membrane for growth of colonies

As the agar plate-based assay was to be used for large scale screening of multiple libraries, it was desirable to use cheap components where possible to reduce the overall cost of the process. Other groups using this format of assay have plated colonies onto polyester (Delagrave et al., 2001), nitrocellulose (Alexeeva et al., 2002) or nylon (Escalettes and Turner, 2008) membranes. A trial was carried out to see if these membranes could be substituted with filter paper, which is significantly cheaper. Unfortunately it is clear from Figure 3.2 that filter paper is not an appropriate substitute as the coloured ABTS product diffused significantly across the membrane whereas with the nitrocellulose membrane the product remained localised, enabling identification of positive colonies. Filter paper would only be usable in screening if colonies were plated out at very low density and very few displayed colour change.
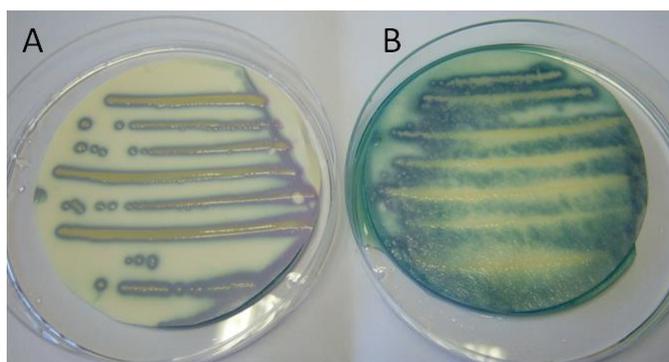
Figure 3.2. Comparison of nitrocellulose membrane and filter paper for growth of colonies

E. coli expressing WT GO-N6M1 were grown on A: nitrocellulose membrane or B: filter paper and the assay carried out. After 30 minutes, the coloured product has diffused across the filter paper to a significantly greater extent than the nitrocellulose membrane.

The sensitivity of the assay was compared when nitrocellulose membrane of two different thicknesses or nylon membrane was used. Nitrocellulose membrane is available from different suppliers at thicknesses of approximately 125 μm or 190 μm while the nylon membrane used (Hybond-N, GE LifeSciences) is between 100 and 200 μm according to the manufacturer. As shown in Figure 3.3, when the assay was carried out with nitrocellulose membrane of 125 μm, colour formation occurred much more rapidly than with the other two membranes implying a more sensitive assay with the thinner membrane.
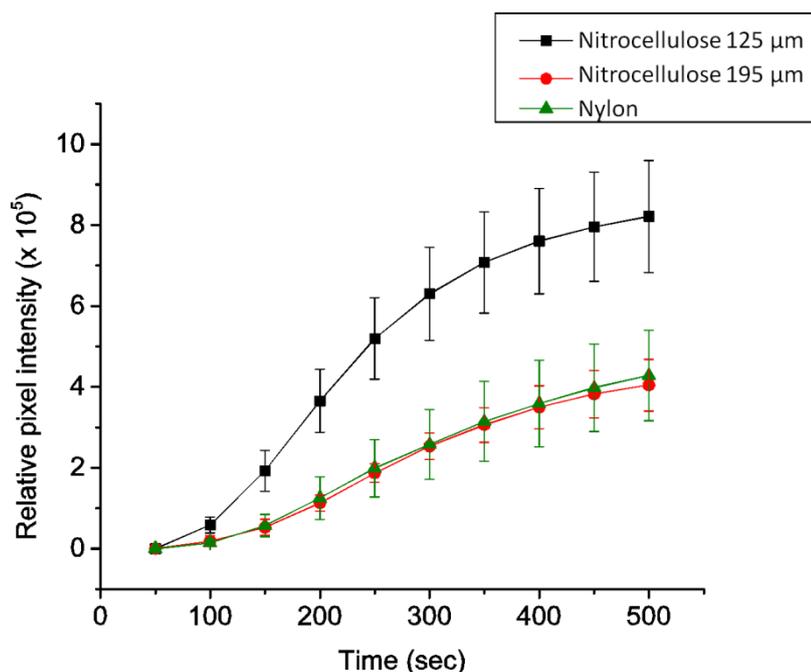
Figure 3.3. Effect on rate of colour formation with three different screening membranes

> The sensitivity of the assay was compared by measuring the rate of colour formation when nitrocellulose of two different thicknesses or nylon membrane was used for plating out the *E. coli*.

The final optimisation for the membrane used was to compare the sensitivity of nitrocellulose membrane with different pore sizes. Nitrocellulose membrane is available with different pore sizes due to its primary function in blotting experiments. Membranes with smaller pore sizes have a larger surface area and so show better retention of proteins. As shown in Figure 3.4, there was no significant difference in screen sensitivity when the pore size was 0.45 or 0.2 µm implying that the difference in surface area does not affect the retention of any assay components such as the protein HRP.
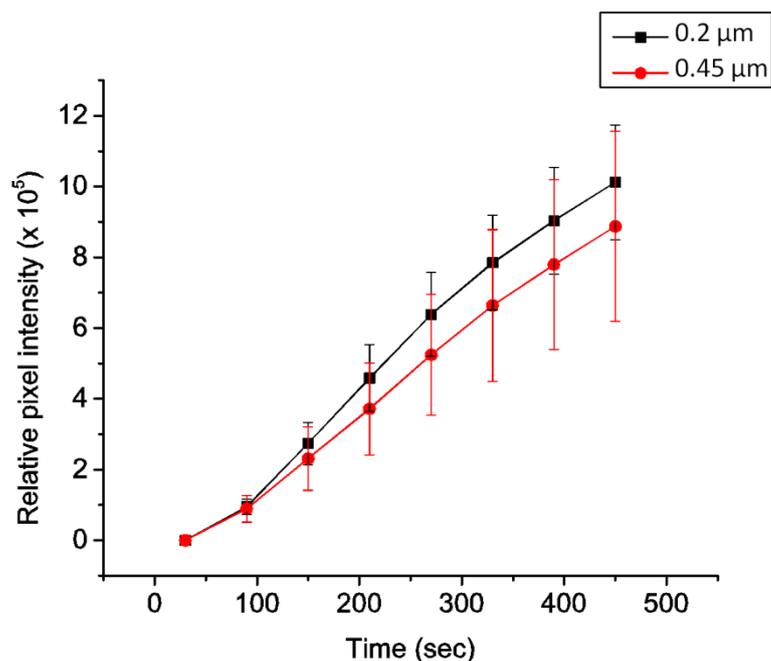
Figure 3.4. Effect on rate of colour formation of nitrocellulose of different pore sizes

The sensitivity of the assay was compared by measuring the rate of colour formation when nitrocellulose membranes with two different pore sizes were used for plating out the *E. coli*.

3.2.2 The most appropriate chromogenic peroxidase substrate for the assay

Other published work using agar plate-based assays has used either 4-chloronaphthol (4-CN) (Delagrave et al., 2001, Escalettes and Turner, 2008, Rannes et al., 2011) or DAB (Alexeeva et al., 2002) as the substrate for HRP to generate a coloured product in the presence of hydrogen peroxide. ABTS is a highly sensitive chromogenic peroxidase substrate already used in GO assays for kinetic analysis (Baron et al., 1994). Assays were carried out with GO-N6M1-expressing *E. coli* to compare ABTS or DAB and 4-CN. As shown in Figure 3.5, use of ABTS produced much better results than a combination of the other two substrates as limited colouration of the nitrocellulose membrane occurred and the two negative controls (with no D-galactose and with no HRP) resulted in essentially no colour change. Additionally, DAB is a toxic compound which is carcinogenic so its use should be avoided if there is an alternative. 4-CN is less stable than ABTS and is light sensitive which may explain the colouration of the membranes in the assays containing this compound. Overall, ABTS seems to be the most sensitive and safest chromogenic substrate for use in the agar plate-based assay.
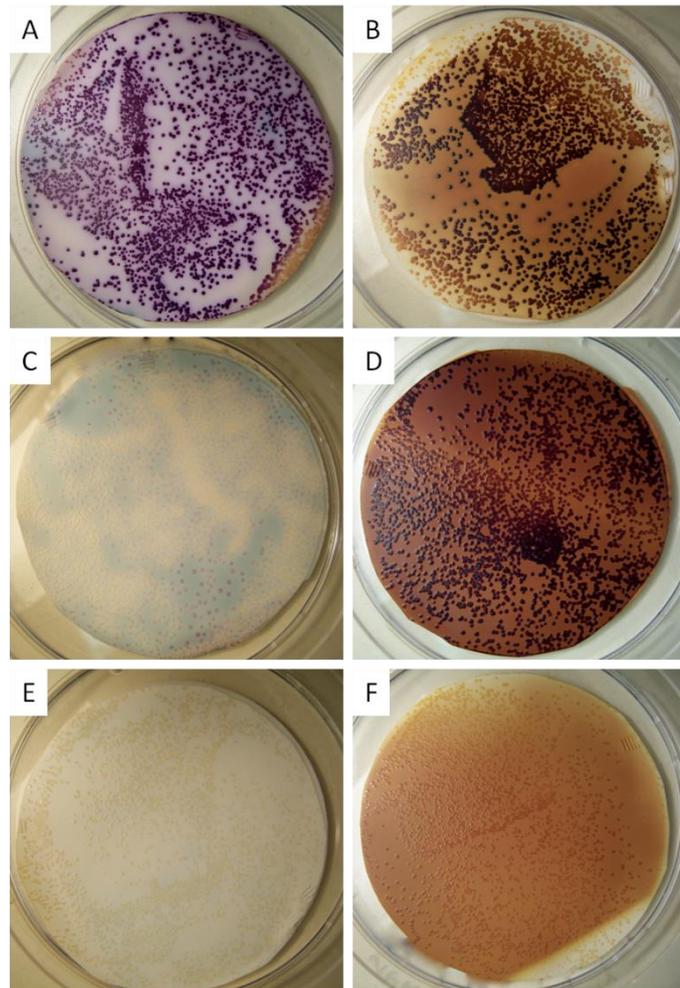
Figure 3.5. Comparison of ABTS and DAB/ 4-CN as the coupled assay substrate

The assay was carried out with GO-N6M1-expressing *E. coli* with ABTS (A, C and E) or DAB & 4-CN (B, D and F) as the chromogenic substrate. A and B show the complete assay; C and D show controls where no D-galactose was included in the assay mix; and E and F show controls where no HRP was included in the assay mix.

Trials were carried out to ensure that colonies expressing variants active against the substrate provided could be distinguished from those expressing inactive variants. A mixture of *E. coli* expressing GO-N6M1 or *E. coli* amine oxidase (ECAO) were mixed and plated onto the nitrocellulose membrane. ECAO shows no activity against D-galactose so no hydrogen peroxide would be produced by these colonies. Figure 3.6 clearly shows that colonies expressing active variants could be easily identified in the screen using ABTS. When DAB and 4-CN were used as coupled substrates it was more difficult to identify the colonies, but still possible.
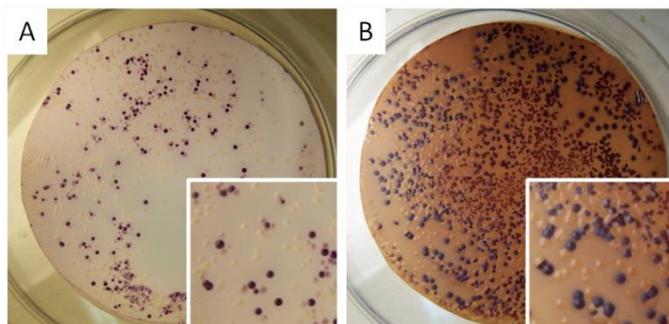
Figure 3.6. Distinguishing colonies expressing/ not expressing active GO

> Screens were carried out using a mixture of colonies expressing GO or ECAO (which shows no galactose oxidase activity). In the screen using ABTS (A) colonies expressing active GO are purple while those expressing ECAO remain white. In the screen using DAB & 4-CN (B) colonies expressing active GO are dark brown while those expressing ECAO are light brown. In each case, the inset shows a magnified view of part of the plate.

3.2.3 Secretion of GO into the periplasm

In all published examples of this screen except Alexeeva et al. (2002), where the *E. coli* were plated directly onto agar, after growing on the membrane colonies were duplicated and lysed by chloroform vapour before addition of assay components. Any colonies displaying the desired activity were then recovered from the duplicate plate. With the aim of reducing the number of steps in the screening process and also the chance of errors introduced by replica plating, cloning was carried out to generate a version of GO which is expressed in the periplasm. The hypothesis was that lysing cells would not be required as the assay mix components should be much more accessible to periplasmic GO.

WT ECAO is expressed in the periplasm of *E. coli* due to a 31-residue secretion signal sequence at the N-terminus and had previously been cloned into the pET28c plasmid by Dr M. Smith, University of Leeds. Primers were designed to amplify the pET28c plasmid including the

secretion signal sequence, N-terminal Strep II-tag and a TEV cleavage site, but excluding the ECAO coding region. Another primer pair was designed to amplify the GO coding region from the pET101D-N6M1 plasmid and add overlaps of 23 and 26 nucleotides complementary to the appropriate regions of the pET28c construct (Figure 3.7) according to the CPEC method detailed in Quan and Tian (2009) (Section 3.1). Primer sequences are shown in Table 3.1.
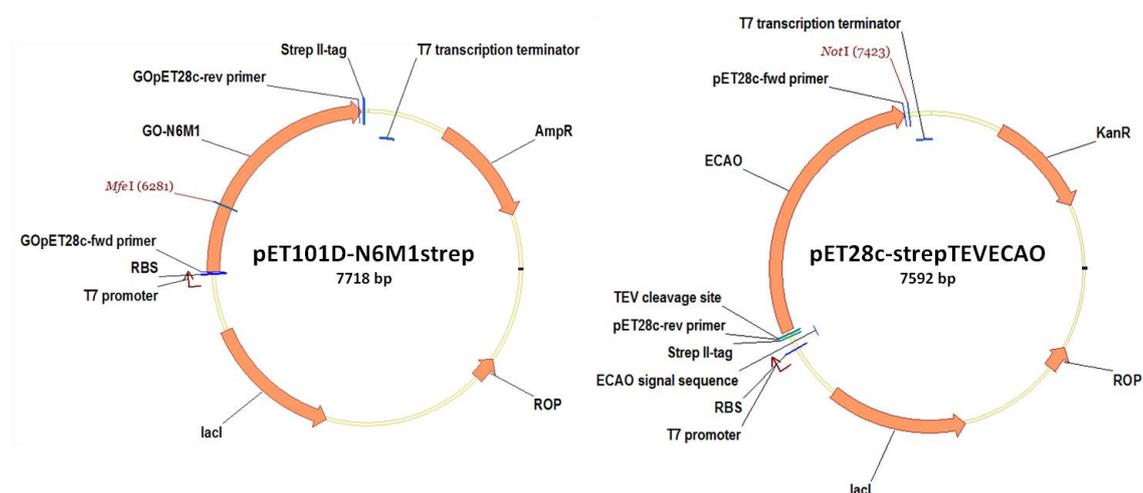
Figure 3.7. The two plasmids used in cloning to transfer the ECAO secretion signal sequence to the GO-N6M1 coding region

The features of each plasmid are shown including the ribosome binding site (RBS), primer binding sites required for the cloning described here and the coding regions of interest

Table 3.1. Primers used to transfer the GO coding region into the pET28c plasmid containing the ECAO secretion signal sequence, Strep II-tag and TEV cleavage site

Primer name	Sequence
pET28c-fwd	5' ggatccgaattcgagctccgtcg 3'
pET28c-rev	5' accctggaagtacaggttttctttctc 3'
GOpET28c-fwd	5' agaaagaaaacctgtacttccagggtgcatctgcacctattggaa gcgcc 3'
GOpET28c-rev	5' cgacggagctcgaattcggatccctgagtaacgcgaatcgtcgaa gcc 3'

Overlapping regions are underlined.

PCR (Section 2.4.1) was carried out to amplify the plasmid and insert using extension times of 5 and 2 minutes, respectively. Following purification of the products using the QIAquick PCR Purification kit (Section 2.4.13), 200 ng of the plasmid and an equimolar quantity of the insert (72 ng) were combined with KOD polymerase reaction components and thermal cycling was

carried out for either 5, 10 or 15 cycles: denaturation at 98 °C for 30 seconds, annealing at 68 °C for 30 seconds, extension at 72 °C for 8 minutes. The products (Figure 3.8) were transformed into XL1 Blue cells (Section 2.4.8) and plated out onto agar plates containing kanamycin which would only support growth of cells containing the pET28c plasmid, which contains the Kan[R] gene. Correct incorporation of the insert into the plasmid was confirmed by DNA sequencing. Unfortunately the primers used to amplify the GO-N6M1 coding region from the pET101D plasmid did not include a stop codon at the 3' end of the coding region which was necessary as the existing C-terminal Strep II-tag was not amplified. Site-directed mutagenesis was therefore carried out using KOD DNA polymerase to insert a stop codon using the primers shown in (Table 3.2).

Table 3.2. Primers used to introduce a stop codon at the end of the GO-N6M1 coding region within the pET28c plasmid by site directed mutagenesis

Primer name	Sequence
pET28c-GO-STOP-fwd	5' cgcgttactcag**t**gatccgaattcgag 3'
pET28c-GO-STOP-rev	5' ctcgaattcggatc**a**ctgagtaacgcg 3'

The mutated nucleotide is shown in red and bold.

78
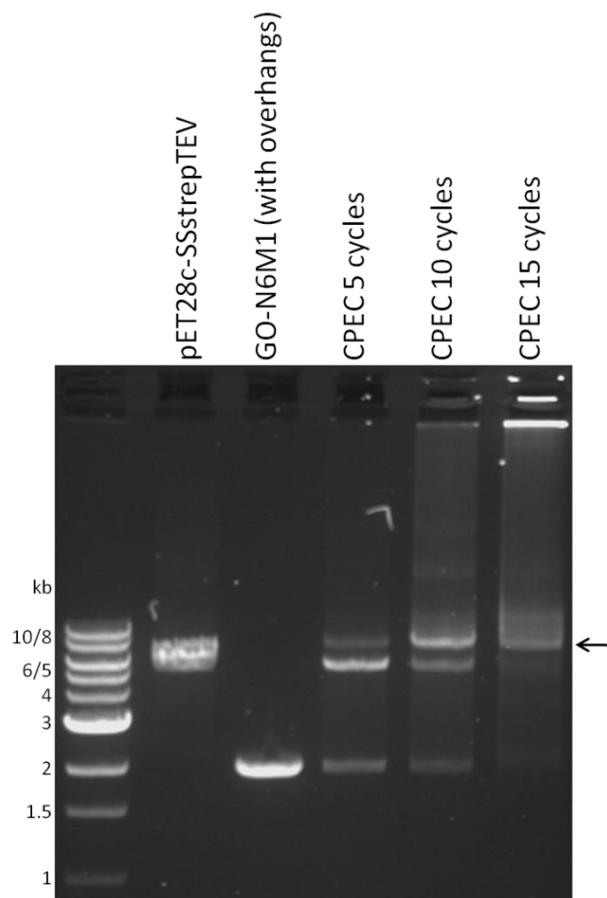
Figure 3.8. Insertion of the GO-N6M1 coding region into pET28c by CPEC

The pET28c plasmid including the ECAO secretion signal sequence, Strep II-tag and TEV cleavage site and the GO coding region were amplified by PCR. The CPEC reaction was carried out for 5, 10 or 15 cycles and the products transformed into XL1 Blue cells. 10 cycles of CPEC appears to have generated the most of the desired product (indicated by the arrow).

Once the construct with the secretion signal sequence upstream of the GO coding region had been generated, it was used to transform BL21 Star (DE3) cells and a screen of colonies expressing secreted or cytoplasmic GO was performed (Figure 3.9). Secretion to the periplasm appears to have no significant impact on the rate of formation of the coloured product on those colonies giving rise to a coloured signal. However, with the periplasmically-expressed GO, all colonies on the plate showed colour change in the assay whereas for the cytoplasmically-expressed GO only approximately 20% of the colonies showed colour change (Figure 3.10), despite all colonies containing the same plasmid with no secretion signal sequence (as confirmed by DNA sequencing). This observation supported the view that the periplasmically-expressed version was the most appropriate for future screening.
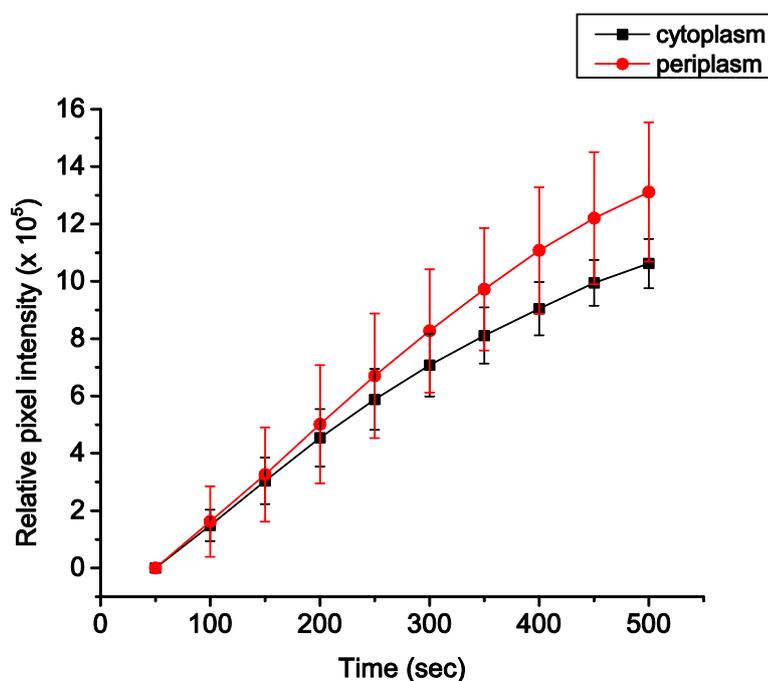
Figure 3.9. Rate of colour formation by GO expressed in the periplasm (red) or the cytoplasm (black)

The sensitivity of the assay was compared by measuring the rate of colour formation when GO was expressed in the cytoplasm or in the periplasm.
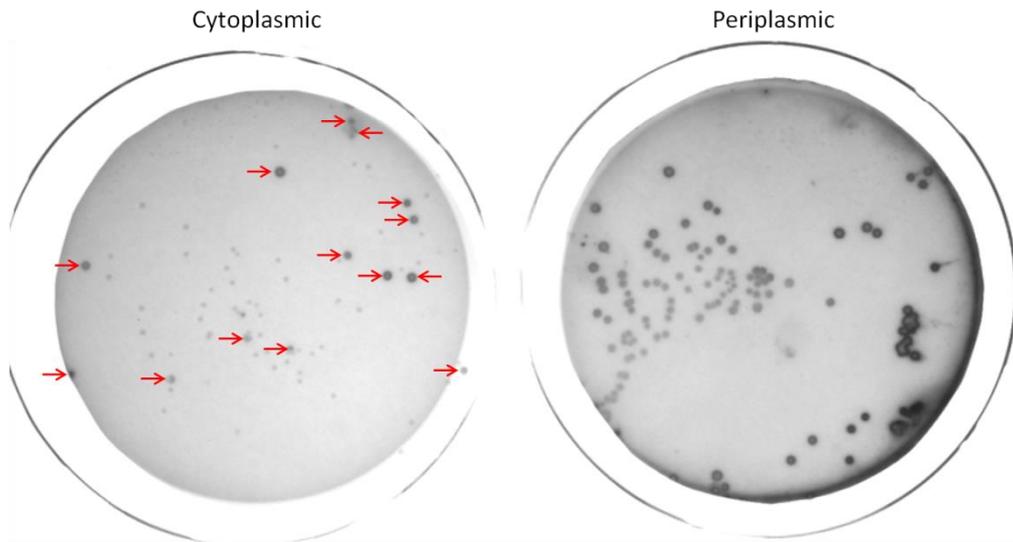
Figure 3.10. The proportion of colonies showing colour change varies if GO is expressed in the cytoplasm or periplasm

Left: when GO was expressed cytoplasmically, only the 14 colonies indicated with the red arrows showed colour change while the 63 other colonies did not. Right: when GO was expressed periplasmically, all colonies showed colour change.

3.2.4 Lysis of colonies by freeze-thaw

When colonies were lysed by chloroform vapour (Delagrave et al., 2001, Alexeeva et al., 2002, Escalettes and Turner, 2008), duplicate plating was required to permit recovery of *E. coli* cells expressing the target variant to allow plasmid purification. It was hypothesised that freezing the membranes at -20 °C and then allowing them to thaw may be a less disruptive method and therefore only lyse a proportion of the cells within the colony. After the assay, the colony could then be used directly to inoculate a liquid culture for plasmid purification.

Freezing overnight, or for 15 minutes (Figure 3.11) resulted in all colonies (expressing GO in the periplasm) subsequently used to inoculate liquid cultures successfully growing, confirming that not all cells in the colonies had lysed.
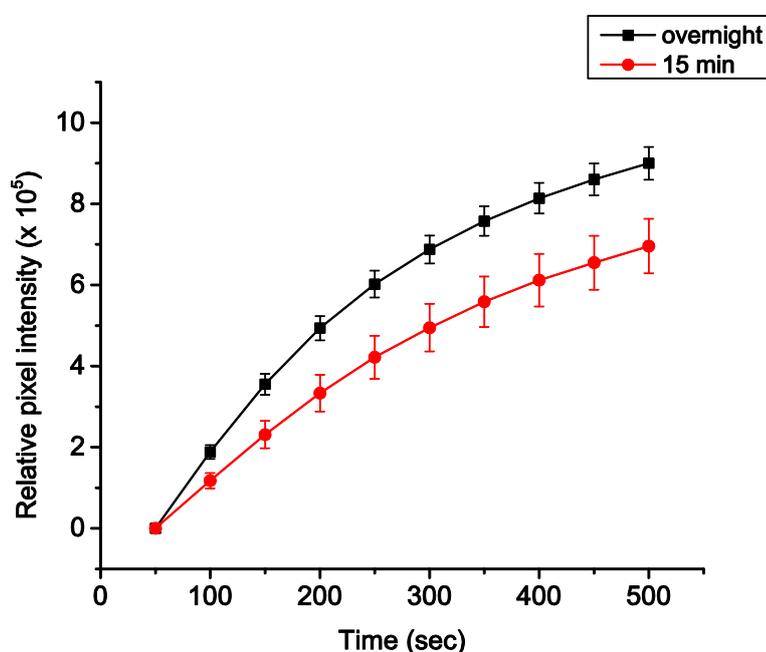
Figure 3.11. The rate of colour formation when the membrane is frozen overnight versus for 15 minutes

The sensitivity of the assay was compared by measuring the rate of colour formation when colonies were frozen for 15 minutes, or overnight.

Overall, it would appear that freezing the colonies overnight improved the sensitivity of the assay compared to freezing for just 15 minutes. However, as the increase in sensitivity was moderate and the addition of an extra day to the screening process would be quite inconvenient, it was decided to use the 15 minute freezing time in future screens. Further

trials were then carried out to determine if carrying out the freeze-thaw cycle multiple times led to further increases in assay sensitivity (Figure 3.12). Interestingly, performing multiple cycles reduced the assay sensitivity. This could be because the freeze-thaw process led to disruption of GO structure as well as the *E. coli* cells or that the extra time of the cell disruption process allowed degradation of the recombinant GO by proteolytic enzymes in the *E. coli* cells.
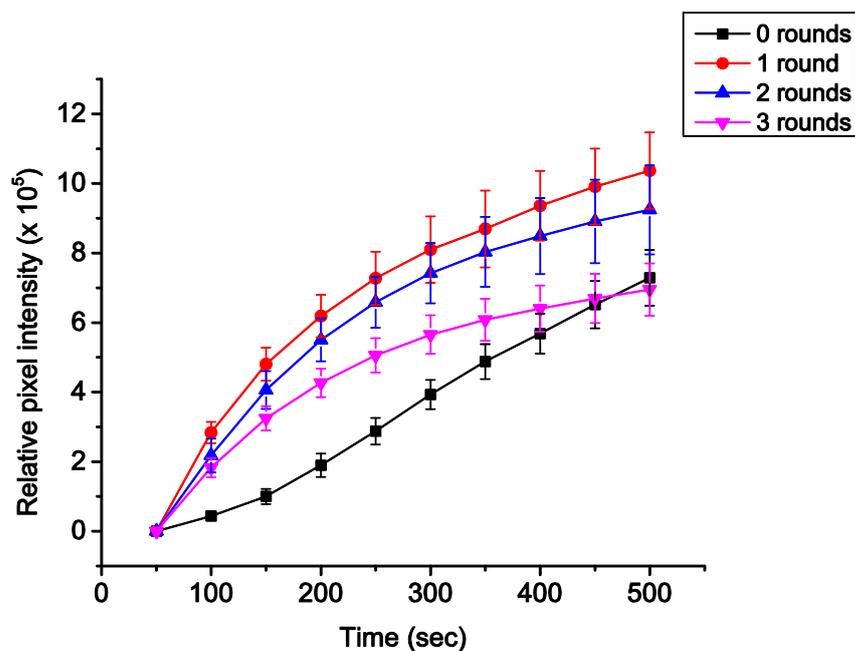
Figure 3.12. The rate of colour formation when multiple rounds of freeze-thaw are carried out

The sensitivity of the assay was compared by measuring the rate of colour formation when freeze-thaw was carried out 0, 1, 2 or 3 times.

3.2.5 Form of copper supplied in the assay

Copper sulfate ($CuSO_4$) is commonly added to GO during purification or in activity screens (Sun et al., 2001, Escalettes and Turner, 2008, Spadiut et al., 2010). The alternative copper (II) donating compound copper nitrate ($Cu(NO_3)_2$) is significantly more soluble than $CuSO_4$: 145 g/ 100 g H_2O compared to 22 g/ 100 g H_2O. This may affect the accessibility of the copper to GO and may have an effect on assay sensitivity. Excess copper may inhibit GO so it is important to determine an appropriate concentration of the copper compound to be used. Assays were carried out where copper was included in the induction mix in the two different forms and at two different concentrations (Figure 3.13). While no huge differences were seen in the sensitivity of the assay, 10 mM $Cu(NO_3)_2$ seemed to be slightly better and so was selected for future screens.
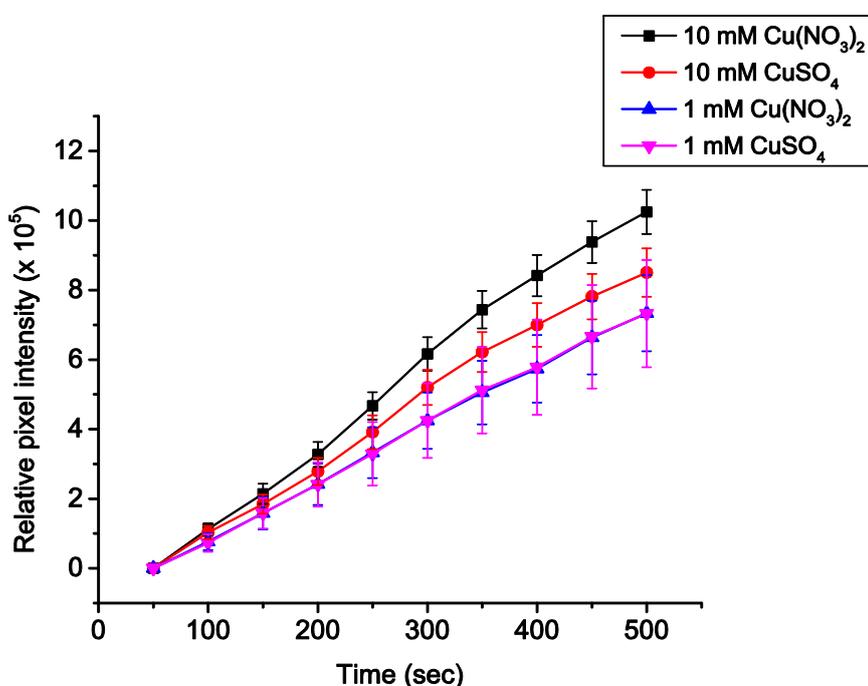
Figure 3.13. The effect of different forms of copper and different concentrations on screen sensitivity

The sensitivity of the assay was compared by measuring the rate of colour formation when copper (II) was provided in two different forms and at two different concentrations.

3.2.6 Optimisation of induction conditions

Induction of GO expression in the colonies plated onto the membrane was based on the method used by Delagrave et al. (2001) where the membrane was transferred to 'induction mix' before exposure to the 'assay mix'. While that method supplied the inducer (L-arabinose) in LB-agar, the screens detailed here were found to work when the membrane was placed onto filter paper soaked in the induction mix (containing IPTG). In order to ensure maximal expression and therefore maximal assay sensitivity, IPTG concentration (Figure 3.14) and the duration of induction (Figure 3.15) were varied. Surprisingly, no significant difference in assay sensitivity was seen at any IPTG concentration, including 0 mM. This implies that GO is being expressed without induction, a phenomenon known as 'leaky expression'. As addition of IPTG appears to have no effect, it implies that the level of leaky expression is adequate for the screen, however as the effect of various factors on leaky expression is not known, it was decided to include 0.1 mM IPTG in the induction mix to ensure equal induction in all screens.
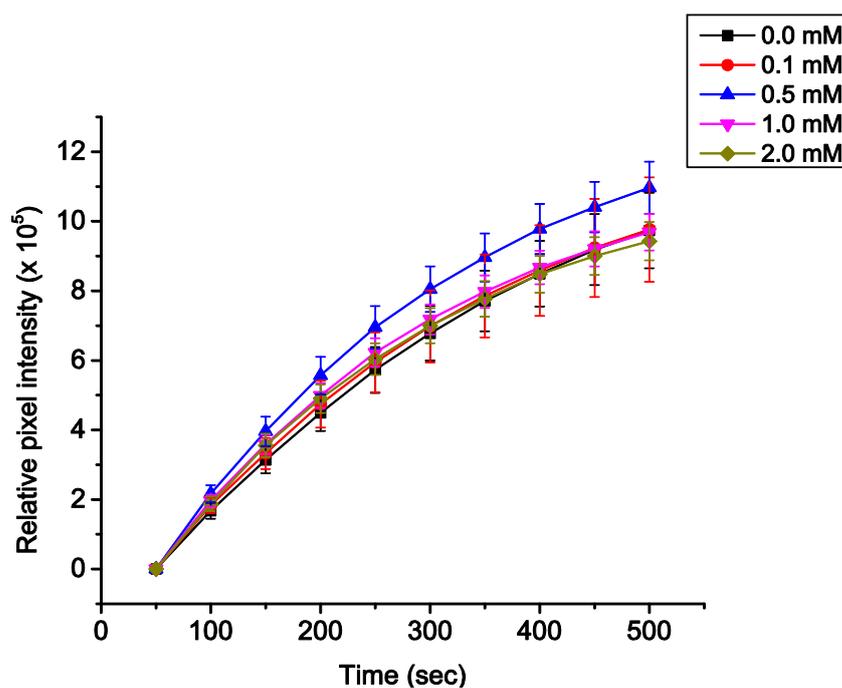
Figure 3.14. The effect of varied IPTG concentrations on screen sensitivity

The sensitivity of the assay was compared by measuring the rate of colour formation after induction with different IPTG concentrations.

Trials where induction time was varied (Figure 3.15) showed only small differences with 1 hour induction resulting in a slightly reduced assay sensitivity while there was no noticeable difference between 2 and 3 hour inductions. It was therefore decided to expose the colonies to the induction mix (containing 10 mM $Cu(NO_3)_2$ and 0.1 mM IPTG) for 2 hours.
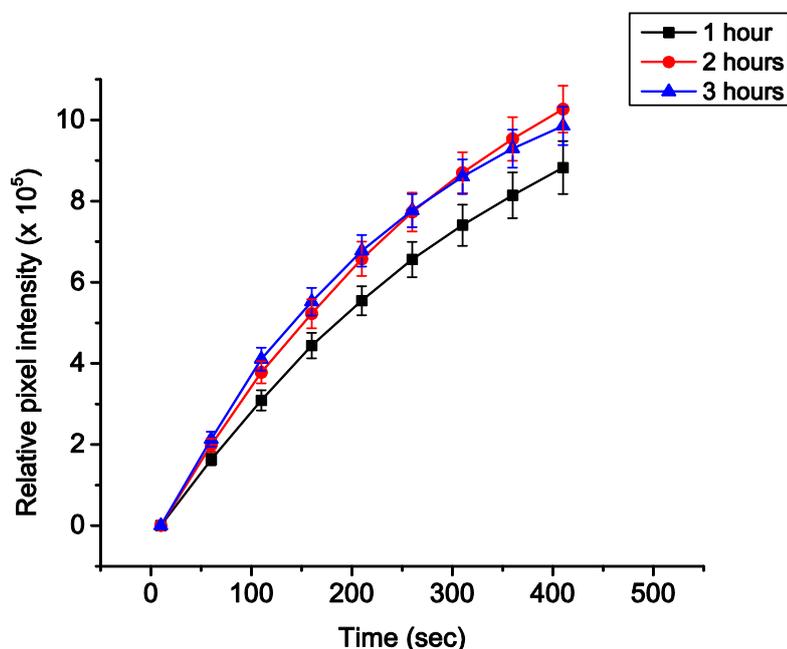
Figure 3.15. The effect of different induction times on screen sensitivity at 0.1 mM IPTG

The sensitivity of the assay was compared by measuring the rate of colour formation after incubation with 'induction mix' for 1, 2 or 3 hours.

In a separate experiment, the screen was carried out following a) incubation with 'induction mix' containing 0.1 mM IPTG but no copper and b) no incubation with 'induction mix'. Even when 10 mM $Cu(NO_3)_2$ was included in the assay mix, no colour change was observed after 500 seconds implying that the incubation with the 'induction mix' is perhaps most important for the copper-dependent processing of the enzyme (Section 1.6.1.2).

3.2.7 The concentration of substrate used

In order to give a qualitative estimation of the sensitivity of the assay, screens were carried out at four different D-galactose concentrations. As shown in Figure 3.16, after 12 minutes some colour change was visible in screens where only 10 mM of D-galactose was supplied, at least 5-fold lower than the K_M of the enzyme. No significant difference was visible between the screens carried out at 100 and 600 mM D-galactose which was as expected as both concentrations are greater than the K_M and therefore saturating. It was decided to carry out all future screens with substrate concentrations of 200 mM, regardless of which substrate was being screened.
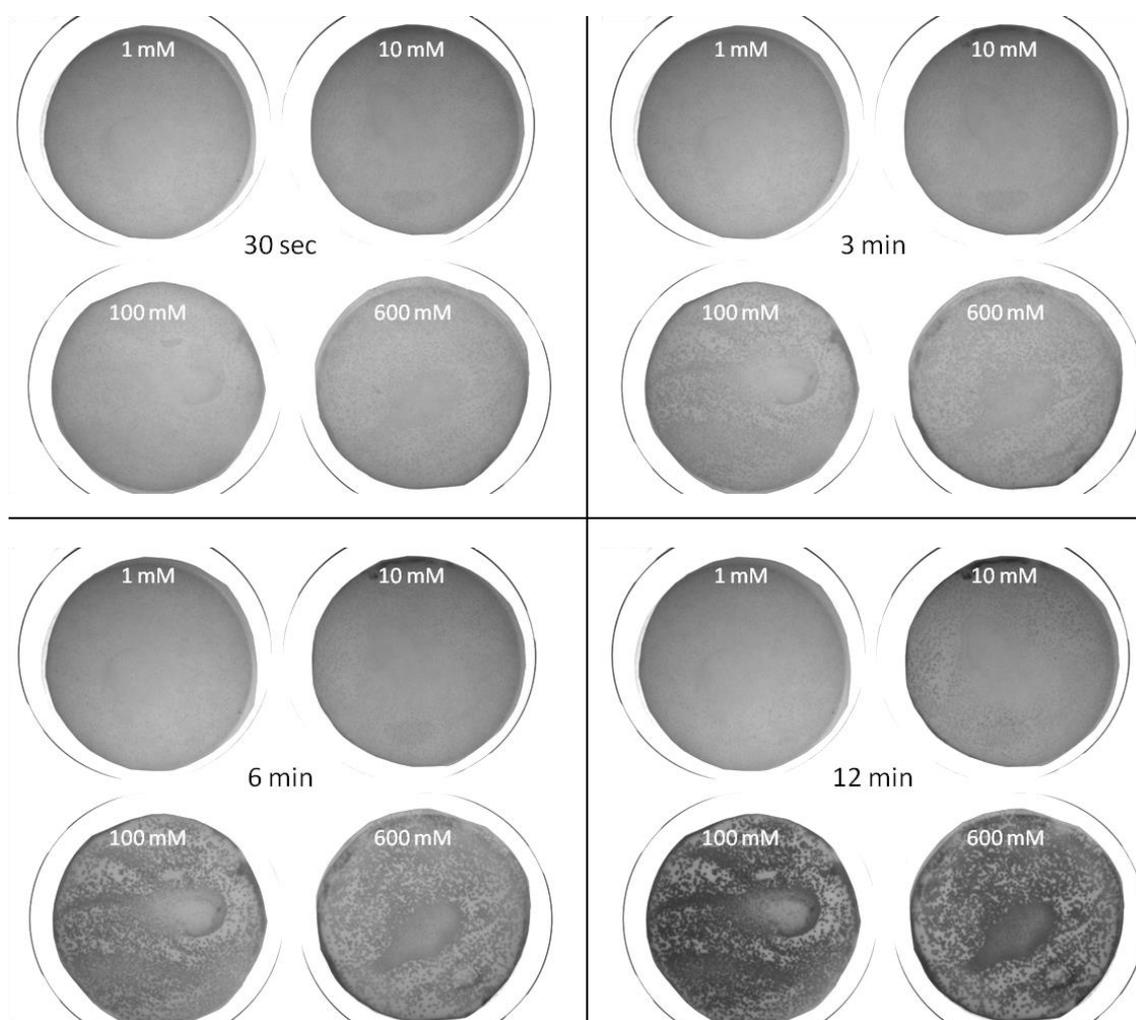
Figure 3.16. The sensitivity of the screen with different D-galactose concentrations

3.3 Optimisation of the microtitre plate-based screen

In order to achieve the highest levels of expression of GO in the wells of the microtitre plate and thereby achieve the maximal screen sensitivity, a number of parameters were varied: IPTG concentration, induction time and whether the pelleted cells were resuspended in lysis buffer or in sodium phosphate buffer, pH 7.0, in case the lysis buffer interfered with the coupled enzyme reaction. As shown in Figure 3.17, the screen showed higher sensitivity when the cell pellets were dissolved in lysis buffer. This implies that, if the lysis buffer did interfere with the assay reaction, this effect was not as great as the effect of not lysing the cells. In contrast to the agar plate-based screen, addition of IPTG did seem to improve the microtitre plate-based screen, implying that the leaky expression was not adequate in this case. When IPTG was added, the length of the induction period had a variable effect on the screen with longer induction periods resulting in lower sensitivity at IPTG concentrations of 1.0 and 0.5 mM but with the opposite effect seen when 0.1 mM IPTG was added. Appropriate results were seen when induction was carried out overnight with 0.1 mM IPTG and these conditions were therefore selected as this reduces cost while overnight induction permits completion of the subsequent screening stage within the next working day.

To determine whether the changes in activity observed in the screen correlated with expression of GO-N6M1, the cell lysates were analysed by SDS-PAGE (Figure 3.18). In most cases, it was not possible to visualise differences in expression levels and Western blot analysis may have been more appropriate. However, in some cases, expression levels did not seem to correlate with the screen results. An example of this is the trials at 2 and 4 hours of induction with 0.1 mM IPTG with lysis buffer; while the screen data were very similar, expression levels seem to be much lower after 2 hours of induction.
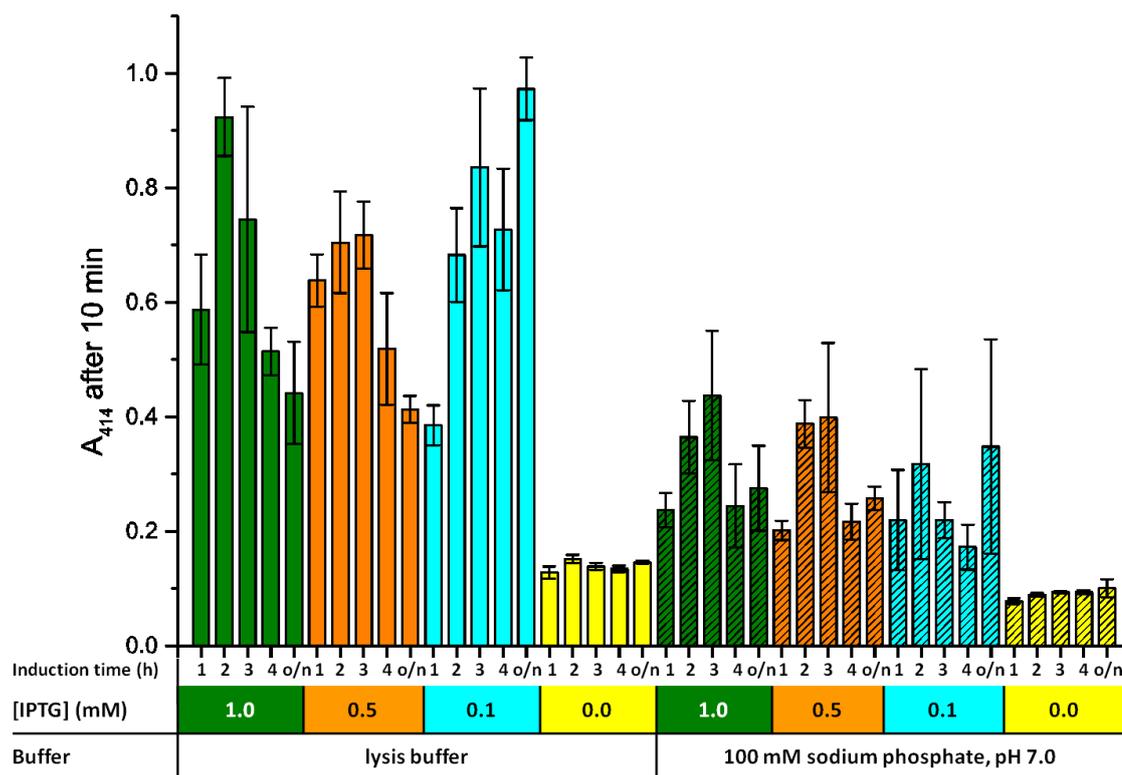
Figure 3.17. Trials to maximise the sensitivity of the microtitre plate-based assay
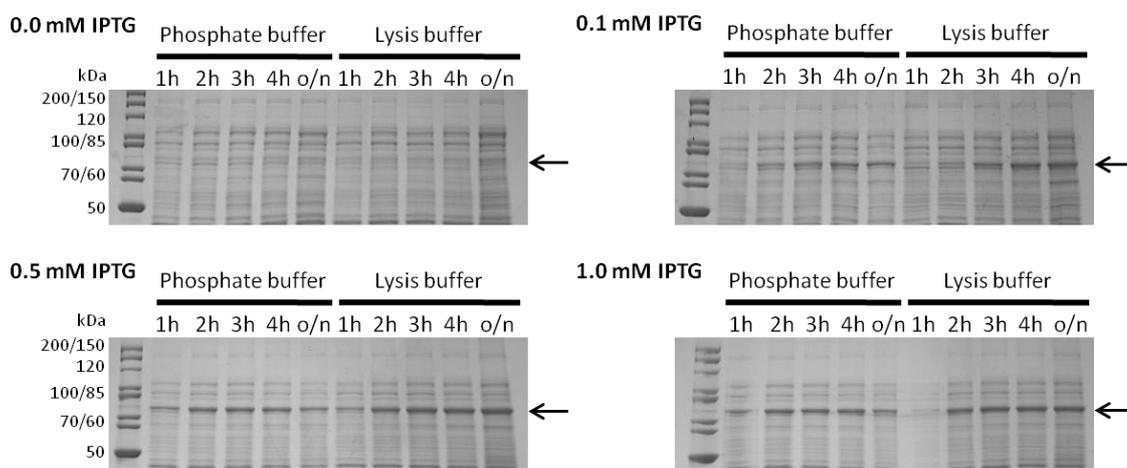
Figure 3.18. SDS-PAGE gels of the lysates from the microtitre plate-based assay trials

The lysates from one of the screens used to generate Figure 3.17 were analysed by SDS-PAGE to visualise relative expression levels of GO.

Following these optimisations with D-galactose, screens were then carried out with GO-N6M1 and alternative substrates. After 60 minutes, only screens with glycerol or D-xylose showed any visible colour change while screens with D-arabinose, D-glucose and D-mannose showed very small increases in absorbance, only detectable by plate reader (Figure 3.19).
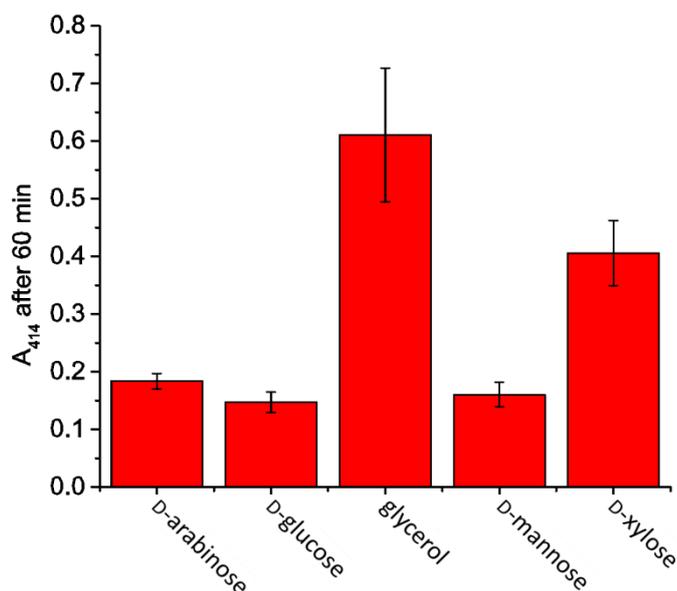
Figure 3.19. Typical A$_{414}$ values in screens against the different substrates with GO-N6M1 after 60 minutes

Only screens with glycerol or D-xylose resulted in a visible colour change while the smaller changes in D-arabinose, D-glucose and D-mannose screens were detected by plate reader.

In an attempt to increase the sensitivity of the screen, the amount of lysate added to the assay plate was increased for the glycerol screen. As shown in Figure 3.20, adding 20 or 40 µl of lysate had the same effect on screen sensitivity while higher volumes resulted in a lower level of colour change. It seems likely that this is due to inhibition of the coupled assay by component(s) of the lysis buffer but it was not considered necessary to explore which components are responsible for this. For all subsequent screening, 20 µl lysate was used.
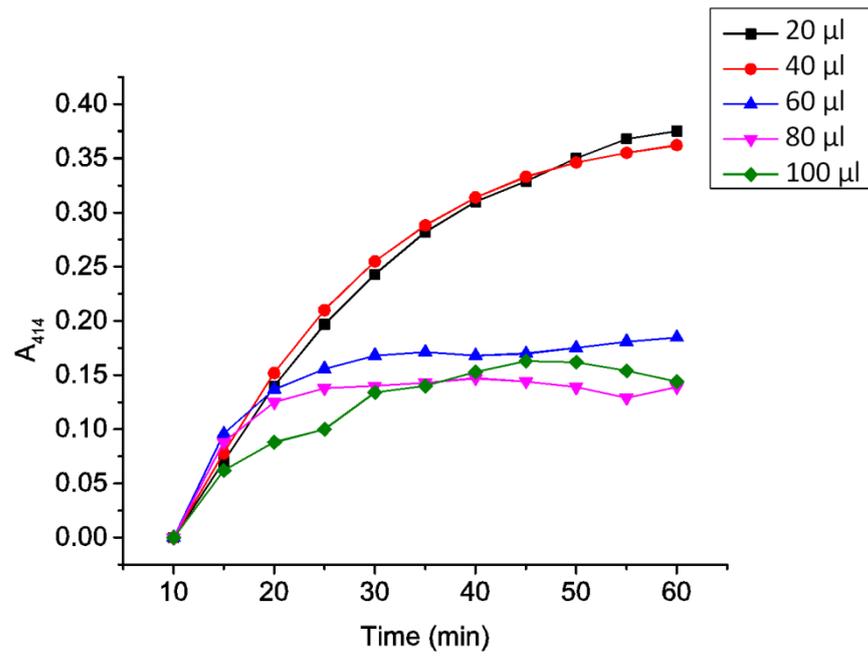
Figure 3.20. The effect of adding different lysate volumes on rate of colour formation

Adding lower amounts of lysate appeared to lead to greater screen sensitivity as the final degree of colour change was greater.

3.4 Removal of the secretion signal sequence and comparison of GO expression levels in the periplasm or cytoplasm

Following screening, large scale expression and purification of any variants of interest is required to enable confirmation and characterisation of putative novel activities. Expression and purification from *E. coli* was optimised previously (Deacon and McPherson, 2011). As detailed in Section 3.2.3, expression of GO in the *E. coli* periplasm led to improvement of the agar plate-based screen. However, it was important to determine whether this had an effect on expression levels in large scale cultures. A comparison was therefore conducted of secreted versus cytoplasmic protein production and purification. Primers were designed to amplify the pET28c plasmid including the GO-N6M1 coding region, the Strep II-tag and TEV cleavage site but excluding the ECAO secretion signal sequence (Table 3.3). The forward primer incorporated a 20 nucleotide overlap so that, following amplification, the product would spontaneously circularise under the PCR conditions.

Table 3.3. Primers used to remove the ECAO secretion signal sequence from the pET28c-SSstrepTEVN6M1 construct

Primer name	Sequence
GODelSS_F	5' gaaggagatataccatgggttggagccatccg 3'
GODelSS_R	5' catggtatatctccttc 3'

The overlapping region to permit re-circularisation is underlined.

PCR was carried out with an extension time of 3 minutes and deletion of the secretion signal sequence was confirmed by colony PCR using the T7F and GOR2 primers and a 60 second extension time (Section 2.4.10). The original sequence containing the secretion signal sequence resulted in a 778 bp product, while successful removal of the secretion signal sequence resulted in a 682 bp product. Agarose gel electrophoresis was carried out to determine the size of the product generated (Figure 3.21). Plasmid DNA was purified from positive clones and the absence of the secretion signal sequence was confirmed by DNA sequencing.
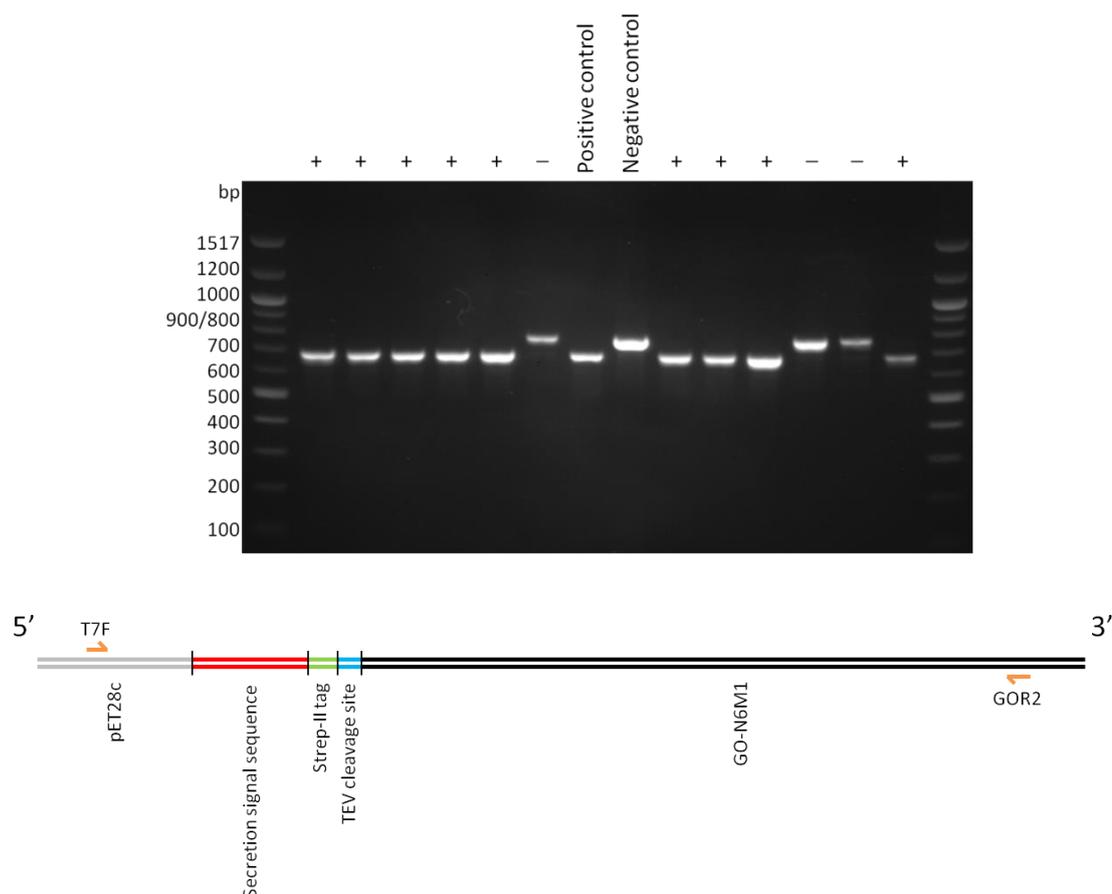
Figure 3.21. Agarose gel of colony PCR products showing positive (+) and negative (−) results

TOP: Following colony PCR with the T7F and GOR2 primers, plasmids still containing the secretion signal sequence (-) generated a 778 bp product while plasmids where removal of the secretion signal sequence was successful generated a 682 bp product (+). **BOTTOM**: the position of the primers with respect to the secretion signal sequence.

Table 3.4. The three different constructs used in the project

Construct	Secretion signal sequence	TEV cleavage site	Strep II-tag
pET101D-N6M1strep	No	None	C-terminal
pET28c-SSstrepTEVN6M1	Yes	N-terminal	N-terminal
pET28c-strepTEVN6M1	No	N-terminal	N-terminal

pET101D-N6M1strep was generated by Dr S. Deacon (Deacon and McPherson, 2011), pET28c-SSstrepTEVN6M1 was generated as described in Section 3.2.3.

Each of the three constructs generated were expressed in BL21 Star (DE3) cells by autoinduction and purified by Strep-Tactin chromatography using 2 ml of Strep-Tactin resin (Section 2.5). The two constructs lacking the ECAO secretion signal sequence yielded 8-10 mg

purified protein per litre of culture while periplasmic expression yielded less than 3 mg per litre of culture. Comparison of the insoluble fractions from these purifications indicated that the reduced yield of the periplasmically expressed protein was due to reduced solubility of GO-N6M1. The amount of GO-N6M1 in the insoluble fraction correlates inversely with the yield of soluble purified protein (Figure 3.22).
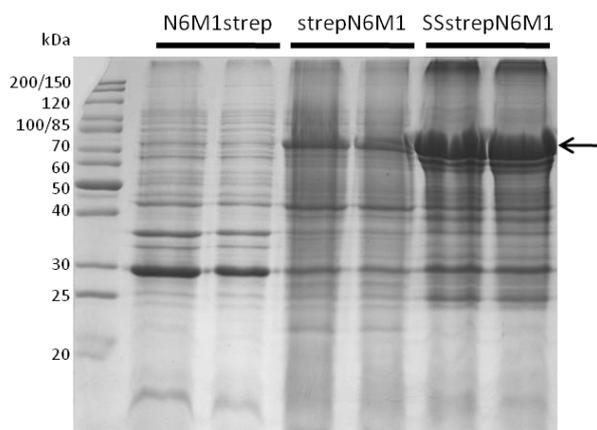
Figure 3.22. SDS-PAGE gel of the insoluble fractions following purification of the three constructs

Insoluble fractions from the purifications of the three different constructs: pET101D-N6M1strep, pET28c-strepN6M1 and pET28c-SSstrepN6M1 were analysed by SDS-PAGE. The amount of GO (marked with an arrow) in the insoluble fraction increased when the Strep II-tag occurred on the N- rather than C-terminus, and was substantially greater in the presence of the ECAO secretion signal sequence (SS).

Following screening and identification of hits the strategy adopted was to remove the secretion signal sequence before protein production and purification. To ensure maximal efficiency, this process was optimised by varying the annealing temperature of the reaction and the initial plasmid concentration (Table 3.5).

Table 3.5. Annealing temperature and plasmid concentration were varied in order to optimise the cloning to remove the secretion signal sequence

Annealing temperature (°C)	Plasmid concentration (ng/ μl)	Number of colonies
45	0.5	20
45	1.0	70
45	2.0	68
55	0.5	27
55	1.0	36
55	2.0	26
65	0.5	1
65	1.0	7
65	2.0	12

The success of the reaction was determined by the number of colonies which grew following transformation of XL1 Blue cells.

While varying the plasmid concentration only appears to have a minimal effect, an annealing temperature of 45 °C led to significantly more colonies than the other temperatures tested. The most appropriate conditions used to delete the secretion signal sequence from clones of interest were therefore:

- Annealing temperature: 45 °C
- Plasmid concentration: 1 ng/ μl

3.5 Discussion

When optimising a process there are normally multiple components which can be tested by sampling a variety of different conditions. In order to save time and resources, and as the two screens optimised here have already been successfully used by other groups, only limited optimisation was carried out. A number of further optimisation experiments are of course possible, but the improvements detailed above were deemed adequate for the screening strategy adopted in this project.

There are a number of further immobilisation membranes available for plating out *E. coli* cells besides those sampled here. These include the polyester membranes used by Delagrave et al. (2001), the polyvinylidene difluoride used in Western blotting experiments, ceramic membranes and cellulose acetate. Each membrane displays different properties such as temperature or pressure resistance, thickness, pore size, hydrophobicity or chemical compatibility which are important for processes such as filtration, blotting experiments and

assaying mixtures. For the agar plate-based assay described here, most of these characteristics are not important and only five membranes were compared (filter paper, nylon membrane and nitrocellulose membrane of two thicknesses and pore sizes). The best membrane from this selection was identified as 125 μm thick nitrocellulose membrane and shown to be adequate for the assay.

The chromogenic peroxidase substrate used is an important component as it is essential in the identification of enzyme activity. Chromogenic peroxidase substrates are commonly used in immunoblotting procedures to identify HRP-conjugated antibodies as well as in other immunochemical techniques such as tissue staining. Previous screens by others have used either DAB or 4-CN as the chromogenic substrate. DAB displays relatively high sensitivity, essential for the assay, but is toxic; while 4-CN shows a lower sensitivity but the blue-purple colour generated is easily visible in screens. Combining these two substrates has been shown previously to result in high sensitivity and generation of a clearly visible black product (Young, 1989) so this mixture was compared with ABTS. The most suitable was ABTS as DAB combined with 4-CN resulted in a brown product with high background levels. However, the trial was only done once and trials using DAB or 4-CN independently were not carried out so the data generated are not conclusive. The trial did however, confirm that ABTS is a suitable substrate for the screen and as this substrate is stable, safe and affordable it was selected for future screening. ABTS is also now used as the chromogenic peroxidise substrate in the Turner and Flitsch groups (Dr M. de Abreu, personal communication). Another chromogenic peroxidase substrate is 10-acetyl-3,7-dihydroxyphenoxazine (Amplex Red®) (Mohanty et al., 1997, Zhou et al., 1997) which is oxidised by HRP to form the fluorescent product resorufin in the presence of hydrogen peroxide. Amplex Red® is claimed to show much greater sensitivity than ABTS, although it is less stable, and it would be interesting in the future to see if the use of Amplex Red® results in increased sensitivity for both the agar plate- and microtitre plate-based assays. Development of further safe and ultrasensitive chromogenic substrates for HRP is still underway with the recent report of another fluorescent probe based on 4-amino-1,8-naphthalimide (Groegel et al., 2011).

Expression of GO in the *E. coli* periplasm was hypothesised to increase the rate of colour formation in the agar plate-based assay. Whilst this was not the case, assay sensitivity was improved by increasing the number of colonies showing colour change, compared to cytoplasmic expression which only led to colour change in a fraction of the colonies. A potential explanation is that these colonies were in fact expressing GO in the periplasm.

However, this explanation is difficult to rationalise as sequencing revealed that the secretion signal sequence was absent from all plasmids in the 'cytoplasmic expression' trial, regardless of the screen result. While the environment for the different colonies on the plate appears to be identical, subtle variations in hydration or nutrient levels may affect the colonies resulting in different levels of lysis, leading to the release of GO from the colonies; or secretion of GO into the periplasm, despite the absence of the secretion signal sequence. Further trials would be necessary to determine the reasons for this odd result but these were deemed unnecessary as the assay works adequately with the periplasmically-expressed GO.

Whilst replica plating is a widely used technique, it is relatively easy to introduce errors into a library screening process. Some of the colonies on the first plate may not grow successfully on the replica plate, with the risk that variants of interest may not be recoverable. It is also sometimes difficult to identify corresponding colonies on the replica plate, leading to selection of the incorrect colony. In order to remove the need for replica plating, colonies were *partially* lysed by freeze-thaw rather than the more destructive chloroform vapour technique used by others. One round of freeze-thaw was shown to improve assay sensitivity, presumably by lysing a proportion of the cells in each colony due to the formation of ice crystals within the cells. Following freeze-thaw, sufficient cells remained viable as each colony could be used to inoculate liquid culture leading to culture growth. As well as adding to the simplicity of the assay, the freeze-thaw technique removed the safety concerns over use of chloroform, and the cost of the chloroform, making the assay safer and cheaper.

Only limited trials were carried out to observe the effect of changing substrate concentration on the agar plate-based screen and the trials were only carried out with D-galactose. It would have perhaps been better to have determined the effect of varying the concentration of each of the substrates to be screened to identify the substrate concentration at which WT GO-N6M1 shows colour change. However, it was known from initial experiments that the substrate concentrations required to observe activity against the alternative substrates (particularly D-arabinose, D-mannose and D-glucose) would be greater than 1 M which would result in the use of large quantities of substrates as well as potential issues with the viscosity of the assay mixes, particularly as some screens were to be carried out with mixtures of substrates. It is also important to consider the first law of directed evolution: 'you get what you screen for' (You and Arnold, 1996), as screening with high substrate concentrations is more likely to identify variants showing activity against these concentrations which are

97

important in potential downstream applications such as chemical synthesis. The issue of substrate concentrations used in screening is discussed further in Chapter 6.

Interestingly, inclusion of IPTG as an inducer in the agar plate-based assay had no effect on the screen while it significantly increased the sensitivity of the microtitre plate-based screen. It would appear that the level of uninduced expression of GO is adequate for the agar plate-based screen but not for the microtitre plate-based screen. Uninduced, or leaky expression is a common occurrence in expression of genes under the control of the *lac* operon and is primarily due to three factors: inefficiency in binding of the repressor protein to the *lac* operator (Glascock and Weickert, 1998); the use of high copy number plasmids which increase the number of promoters and operators requiring repression (Anthony et al., 2004); and the potential for read-through transcription from other promoters present on the plasmid (Krebber et al., 1996, Nishihara et al., 1994, Brown and Campbell, 1993). Leaky expression can be reduced with additional copies of the *lac*I (repressor) coding region or suppressed by inclusion of D-glucose in the growth medium; however this is only necessary if the recombinant protein is toxic to the host cell, which GO is not. Unfortunately it was not possible to determine the relative expression levels in the two screens so it is not possible to say whether the uninduced expression seen in the agar plate-assay occurs at the same level in the microtitre plate-based assay. It could be that the conditions of the agar plate-based assay result in a higher level of uninduced expression or it may be that the microtitre plate-based assay requires a higher expression level, above that generated without IPTG.

As stated in Section 3.1, the ideal assay for identifying hits in a mutant library must be affordable, safe, sensitive, high-throughput, reproducible, accurate and minimise the use of expensive equipment. Where possible, both screens have been optimised to use the cheapest and safest components available such as the use of ABTS which replaces the carcinogenic DAB as a chromogenic substrate. Optimisations have been carried out to improve the sensitivity of both assays, although these optimisations could be continued if required; also neither assay requires the use of expensive equipment such as NMR in the screening process. If carried out on large agar plates (245 x 245 mm bio-assay plates (Thermo Scientific)) the agar plate-based assay can be used to screen up to 2 x 10^5 colonies per plate which can be considered high-throughput screening. The microtitre plate-based screen can be used to screen 88 variants per plate (if eight wells are used for positive and negative controls) and can be considered medium throughput, although it has the added advantage of being quantifiable to an extent so that the relative levels of activity compared to WT can be observed. Unfortunately the screens do not

show as high reproducibility as desired, particularly the microtitre plate-based screen, as seen, for example, in the large error bars in Figure 3.17. This is most likely due to differences in growth of the cultures and expression of GO. Despite growing each culture to saturation, growth rates will vary in each well of the microtitre plate due to differences in the colonies used for inoculation. As a result, effective IPTG concentrations as well as the local environment will vary in each well affecting expression levels. As shown by Figure 3.18, increased expression of GO does not necessarily result in a higher level of activity in the screen. This could be because other components present in the cell lysate, which are not necessarily present in the same proportion as GO, may interfere with the assay components. It would be much easier to compare expression levels of GO if a Western blot was carried out on the samples used in Figure 3.18 as the other bands corresponding to other proteins in the lysate would not be visible. Because of the variation in the microtitre plate-based assay, the optimisations carried out may not be optimal as many more repeats may be required to generate reliable data. When screening against D-arabinose, D-mannose or D-glucose where only low levels of activity are initially expected, any increases in absorbance compared to WT may be within the error of the plate reader. It is therefore essential that repeats are carried out on the microtitre plate-based screens, particularly for these three substrates, so that only genuine hits are taken forwards.

Periplasmic expression often leads to improved solubility of recombinant proteins due to the oxidising environment and presence of chaperones which promote correct formation of disulfide bonds (Berkmen, 2012). However, in the case of GO-N6M1, it would appear that periplasmic expression results in generation of large quantities of insoluble enzyme. This could be because the periplasmic environment contains factors which promote aggregation of GO, however it seems more likely that the issue lies with the secretion process from the cytoplasm into the periplasm. Perhaps the levels of expression are too high for the secretion system used by the ECAO secretion signal sequence and so GO-N6M1 forms aggregates in the cytoplasm instead. Little work has been done on the secretion process of ECAO, however, analysis of the secretion signal sequence using the Signal P method (Nielsen et al., 1999) identified key features of a sec signal peptide: 31 residues long; hydrophilic N-terminal region containing two positively charged residues (lys and arg); central hydrophobic region composed of ten hydrophobic residues; and polar residues (his and gln) at the C-terminal region near the cleavage site. Sec-dependent secretion is often post-translational and so there is greater opportunity for the translated, unfolded protein to aggregate before translocation across the

membrane than if the process was co-translational. In order to improve expression of the periplasmically-expressed GO, the ECAO secretion signal sequence could be substituted for one which directs co-translational secretion such as DsbA (Schierle et al., 2003). However, as sufficient expression levels can be obtained by cytoplasmic expression this was not deemed necessary.

Conclusions

After various optimisation experiments, it was determined that the best membrane for the agar plate-based assay is nitrocellulose at a thickness of 125 μm and the best chromogenic peroxidase substrate is ABTS. Secretion of GO into the periplasm improves the reproducibility of the assay while colony lysis by freeze-thaw improves sensitivity without the need for replica plating. Induction of GO expression by IPTG is not required as leaky expression occurs at sufficient levels, but incubation with $Cu(NO_3)_2$ is essential for enzyme activity. The induction and lysis procedures for the microtitre plate-based assay were optimised and the screen tested with the substrates of interest. Unfortunately, expression of GO in the periplasm leads to significantly reduced yields upon scaled-up purification. Therefore the process for removal of the ECAO secretion signal sequence was optimised so that it can be efficiently carried out for each library variant to be taken forwards.

Chapter 4 : Generation of Mutant Libraries and Screening to Identify Altered Enzyme Activities

4.1 Introduction

This chapter details the process of library generation where residues in loops surrounding the active site are selected for randomisation; oligonucleotides are designed which will be used to combine the randomisations in different combinations; and the range of libraries to be generated is described. The process of library generation and optimisation of the different steps involved is then described, followed by the results of library screening carried out using the assays developed in Chapter 3. Finally, the variants showing significant changes in activity are purified and the novel activities confirmed.

Substrate specificity is often determined by residues occurring in loops, perhaps due to the inherent flexibility of these regions which is important in the 'induced fit' theory of enzyme specificity (Herschlag, 1988, Koshland, 1994). For many enzymes, however, the substrate-induced conformational changes associated with this theory do not occur upon substrate binding (Johnson, 2008). For example, there is no evidence of a conformational change upon binding of substrate to the relatively rigid GO. Examples of loop regions determining substrate specificity include the 'trigger loop' in RNA polymerase which is responsible for the high specificity for the correct ribonucleotide during transcription (Wang et al., 2006); and the active site loops of D-hydantoinase which, when mutated, resulted in significant shifts in substrate specificity (Cheon et al., 2004). Loop regions are also frequently targeted in mutagenesis studies as they are likely to withstand mutation better than secondary structure elements such as α-helices and β-sheets, particularly when surface-exposed. In the case of GO, substrate modelling studies (Ito et al., 1994, Wachter and Branchaud, 1996) and previous mutagenesis projects (Section 1.6.2) (Sun et al., 2002, Delagrave et al., 2001, Deacon et al., 2004) have identified several residues involved in substrate binding, all of which occur in loop regions surrounding the active site.

The libraries created and screened in this chapter are generated using randomised oligonucleotides constructed using trimer phosphoramidites (Section 1.3.4). This rational design method was selected as GO has already been subjected to traditional directed evolution with only limited success (Sun et al., 2001) and there is a large body of research which can be used to select specific residues to target with the aim of altering substrate acceptance. Libraries generated using oligonucleotides generated in this way are generally of a

higher quality than similar methods such as use of degenerate codons (Section 1.3.4) due to a number of factors:

- Each residue has an equal likelihood of being incorporated at each randomised position
- The randomisation can include all 20 residues, all 19 excluding cysteine, or specific subsets as required
- No stop codons are incorporated
- The codons introduced are optimal for usage in *E. coli*
- Neighbouring positions can be randomised independently

The ability to randomise multiple positions in combination and independently of each other is particularly desirable as it allows for potential synergistic conformational effects brought about by the different residues in the different positions. This concept forms the backbone of the CASTing technique developed in the Reetz group (Reetz et al., 2005) (Section 1.3.6). In the work presented here, libraries were generated randomising between one and six residues in combination, some of which are in distinct regions of the active site. Targeting multiple regions of the active site simultaneously has the potential to alter activity in favour of very different substrates from those accepted by the WT enzyme.

4.2 Library design

4.2.1   Selection of residues to mutate

By visually analysing the crystal structure of GO, 13 loops surrounding the active site were identified (Figure 4.1). Two thirds of the residues on these loops point away from the active site copper or have their α-carbon atom at a distance of greater than 12 Å so were disregarded as these are unlikely to be involved in substrate binding.
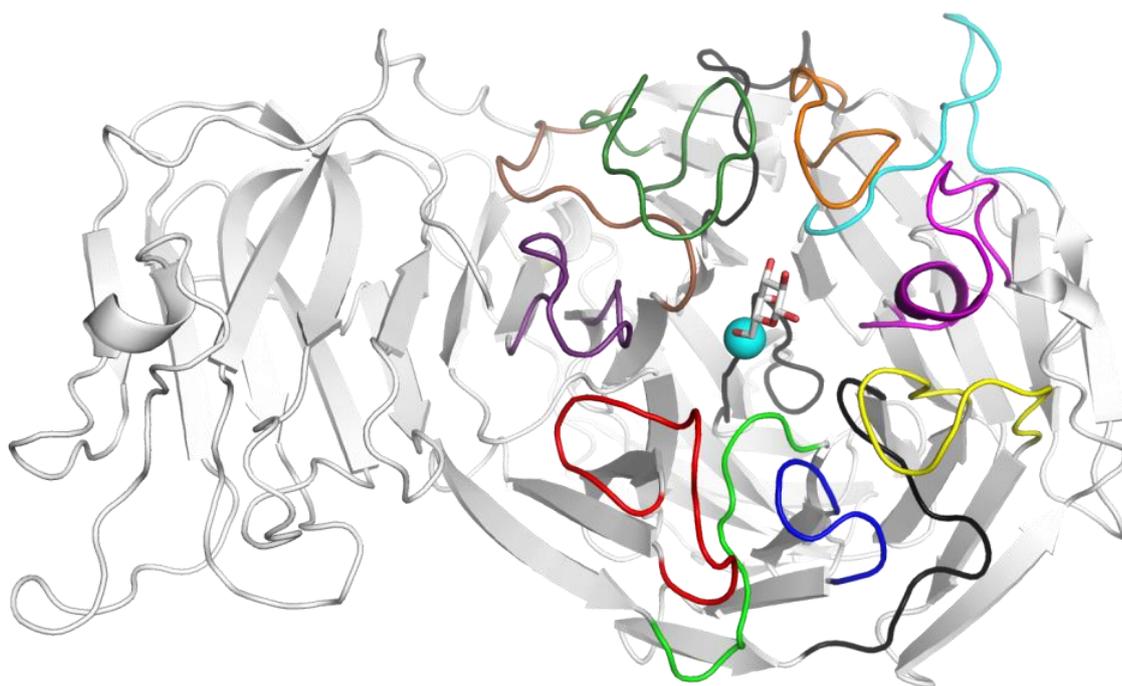
Figure 4.1. Loops surrounding the active site which appear likely to contain residues involved in substrate binding

> The 13 loops surrounding the active site are shown with the active site copper (cyan sphere) and D-galactose modelled in according to Wachter and Branchaud (1996). Loops shown in black are those on which no residues were selected for mutagenesis. Loops 189-201 (red), 220-229 (light green), 243-249 (blue), 287-298 (yellow), 321-335 (magenta), 370-384 (cyan), 400-413 (orange), 456-473 (dark green), 486-496 (brown) and 511-523 (purple) surround the active site. Figure created with PyMOL using .pdb file 1GOF.

Information from previous molecular modelling (Wachter and Branchaud, 1996) as well as mutagenesis studies (Deacon et al., 2004, Sun et al., 2002) was then used to identify residues likely to be involved in substrate interactions (Section 1.6.2): F194 and F464 interact with C4, C5 and C6, Y329 interacts with the C1 hydroxyl, R330 interacts with the C3 and C4 hydroxyls

and Q406 interacts with the C2 hydroxyl. The positions of other residues within the same loops and their chemical characteristics were then assessed in terms of their likelihood of being involved in substrate binding. In all cases the small non-polar residues, glycine and alanine, were thought to be less likely to form important interactions so were disregarded. This resulted in selection of Q326 and N333, two polar residues close to Y329 and R330; Y405, which may be involved in interactions close to Q406; and P463, which may affect the positioning of F464. In order to avoid potential disruption of the thioether linkage between C228 and Y272, residues in the loop containing Y272 were not selected. In the loop containing C228 however, F227 was selected for mutagenesis. This residue is close to the hydrophobic pocket formed by F194 and F464 which is thought to be an important binding surface in the active site. The two loops containing residues 431-442 and 577-587 are relatively buried within the enzyme structure and were not targeted as changes may disrupt the fold of the protein. Finally, residues N245, L514 and the disulfide bond-forming residues C515 and C518 were selected based on their positions close to the active site copper.

Table 4.1. Loops of interest and residues selected for mutagenesis

Loop	Residue(s) selected
189-201	F194
220-229	F227
243-249	N245
263-273	
287-298	W290*
321-335	Q326, Y329, R330, N333
370-384	C383*
400-413	Y405, Q406
431-442	
456-473	P463, F464
486-496	V494*
511-523	L514, C515, C518
577-587	

Loops are coloured according to Figure 4.1. Residues marked * were only randomised to a subset of residues based on previous studies.

Most of the selected residues were randomised to any of the 19 amino acids excluding cysteine. Cysteine is commonly omitted from mutagenesis studies due to the potential formation of disulfide bonds which can result in oxidised forms, cause significant structural disruption, or aggregation of monomers. Residues 515 and 518 which are cysteines in the WT

structure were randomised to any of the 20 amino acids *including* cysteine. Three residues were selected to be mutated to a subset of residues due to previous characterisation of mutants. The stacking tryptophan Trp290 was randomised to only the small non-polar residue Ala and the aromatic residues Phe and Tyr. Histidine was also included at this position as the W290H mutation resulted in decreased K_M in a previous study (Rogers et al., 2007) and His is present at the corresponding position in glyoxal oxidase (Whittaker et al., 1999). Trp290 was also identified as a target for mutation in a previous study where a mutant showing increased D-glucose activity contained, amongst others, the W290F mutation (Sun et al., 2002). Position 383 was previously subjected to saturation mutagenesis as mentioned in Section 1.6.2 (Deacon and McPherson, 2011) which revealed a number of interesting variants, most notably mutations to Glu, Lys, Met and Ser which all reduced K_M whilst having a limited effect on k_{cat} and mutation to Thr which significantly increased k_{cat}. Mutations at position C383 also had varied effects on the pH optimum of the enzyme. In this study therefore, C383 was randomised to these five residues, or Cys. Finally, position 494, which is sited close to the copper ligands Y495 and H496, was changed to valine (as in WT) or alanine (as in the M1 mutant (Section 1.6.4) (Sun et al., 2001)).

4.2.2 Design of randomised oligonucleotides

The positions of the residues selected for mutagenesis within GO are shown in Figure 4.2. A mutant oligonucleotide was designed for each loop so that some oligonucleotides contained only one mutated residue whereas others contained two, three or four residues. The oligonucleotides were designed so that the mutations occurred towards the centre of the sequence with 13-21 bases on each side. Each oligonucleotide also started and ended with at least one guanine or cytosine as these bases form stronger interactions due to the presence of three, rather than two hydrogen bonds as found in the adenine/ thymine interaction, and so improve specificity of binding. This is particularly important at the 3' end of the oligonucleotide. The designed oligonucleotides were generated by Ella Biotech GmbH and are shown in Table 4.2. Their positions within the GO sequence are shown in Figure 4.3.
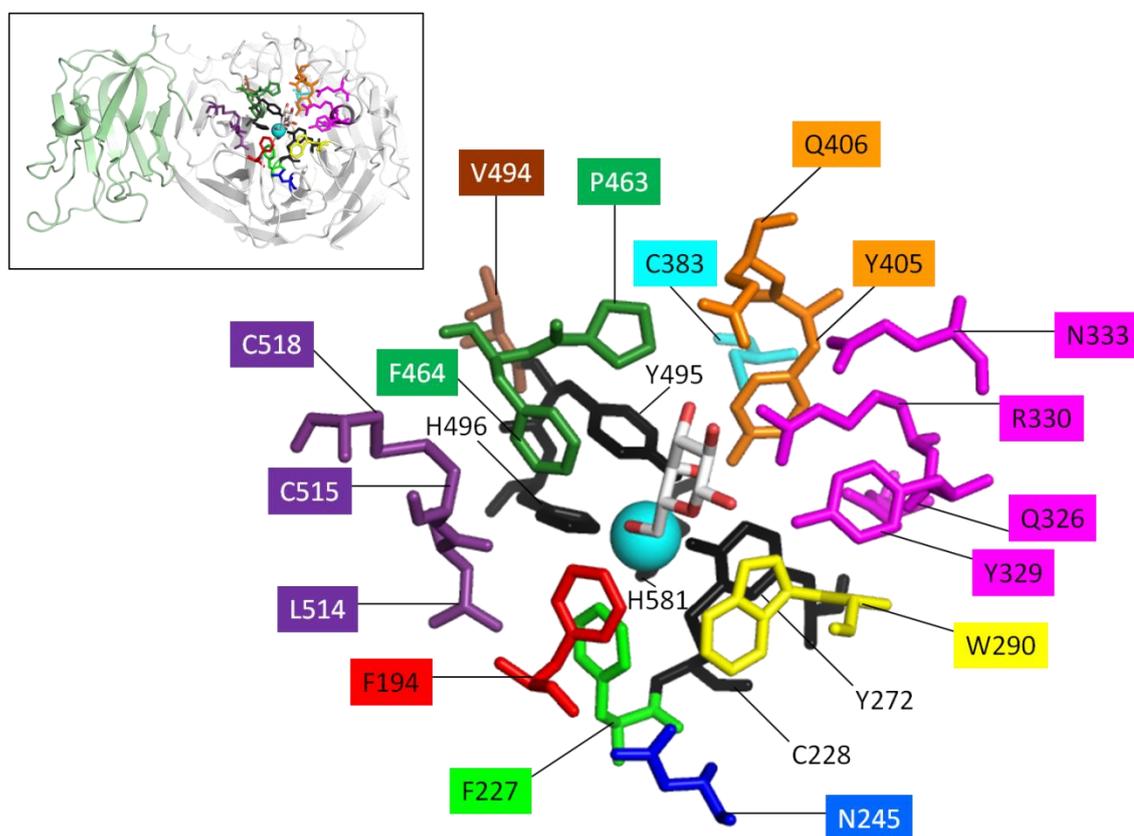
Figure 4.2. Residues selected for mutagenesis

> Residues selected for mutagenesis are shown looking down on the substrate binding region from solvent (see inset for reference); colours correspond to the loops surrounding the active site (as in Figure 4.1). Copper ligands and thioether bond residues are shown in black, copper is shown as a cyan sphere and D-galactose (white) is modelled in according to Wachter and Branchaud (1996). Figure created with PyMOL using .pdb file 1GOF.

Table 4.2. Oligonucleotides for library construction

Name	Residues randomised	Sequence
F194X	F194	5' catatcgcaatgatgcaXXXgaaggatcccctgg 3'
F227X	F227	5' caccaagcatgatatgXXXtgccctggtatctcc 3'
N245X	N245	5' cgtagtcacaggtggcXXXgatgccaagaagacc 3'
W290AFHWY	W290 (subset)	5' ccattggaggctccXXXagcggtggcgtatttg 3'
QYRN326-333	Q326, Y329, R330, N333	5' caatgttgacggctgacaagXXXggattgXXXXXXtcagacXXXcacgcgtggctctttggatgg 3'
C383CEKMST	C383 (subset)	5' cccctgatgccatgXXXggaaacgctgtcatg 3'
Y405XQ406X	Y405, Q406	5' ggcggctccccagatXXXXXXgactctgacgccacaac 3'
P463XF464X	P463, F464	5' gccaacgacgtggaattXXXXXXgaggattcaaccccgg 3'
V494AMV	V494 (subset)	5' ctccattgttcgcXXXtaccatagcatttcc 3'
L514XC515XC518X	L514, C515, C518	5' cggtggtggtggtXXXXXXggcgatXXXaccacgaatcatttcg 3'
GOLibF1		5' gattgttcctgcggctgcag 3'
GOLibR1		5' tgcatcattgcgatatgaag 3'
GOLibR2		5' catatcatgcttggtgac 3'
GOLibR3		5' gccacctgtgactacg 3'
GOLibR4		5' ggagcctccaatggtaaaaac 3'
GOLibR5		5' cttgtcagccgtcaac 3'
GOLibR6		5' catggcatcaggggctac 3'
GOLibR7		5' atctggggagccgcc 3'
GOLibR8		5' aattccacgtcgttggc 3'
GOLibR9		5' gcgaacaatggagttgg 3'
GOLibR10		5' accaccaccaccgttaaatac 3'
GOLibR11		5' ggtggtggtgctcgagtg 3'

TOP: The ten oligonucleotides designed for library generation - colours correspond to the loops surrounding the active site. The codons to be randomised are shown as XXX. BOTTOM: The reverse primers required for generating the double stranded fragments.

```
                              GOLibF1           MfeI                                                              F194X
490  ACTATTGACT TACCGATTGT TCCTGCGGCT GCAGCAATTG AACCGACATC GGGACGAGTC CTTATGTGGT CTTCATATCG CAATGATGCA TTTGAAGGAT
     TGATAACTGA ATGGCTAACA AGGACGCCGA CGTCGTTAAC TTGGCTGTAG CCCTGCTCAG GAATACACCA GAAGTATAGC GTTACTACGT AAACTTCCTA
                                                                                            GOLibR1

                                                                                                                 F227X
590  CCCCTGGTGG TATCACTTTG ACGTCTTCCT GGGATCCATC CACTGGTATT GTTTCCGACC GCACTGTGAC AGTCACCAAG CATGATATGT TCTGCCCTGG
     GGGGACCACC ATAGTGAAAC TGCAGAAGGA CCCTAGGTAG GTGACCATAA CAAAGGCTGG CGTGACACTG TCAGTGGTTC GTACTATACA AGACGGGACC
                                                                                            GOLibR2

                                                        N245X
690  TATCTCCATG GATGGTAACG GTCAGATCGT AGTCACAGGT GGCAACGATG CCAAGAAGAC CAGTTTGTAT GATTCATCTA GCGATAGCTG GATCCCGGGA
     ATAGAGGTAC CTACCATTGC CAGTCTAGCA TCAGTGTCCA CCGTTGCTAC GGTTCTTCTG GTCAAACATA CTAAGTAGAT CGCTATCGAC CTAGGGCCCT
                                     GOLibR3

                                                                                     W290AFHWY
790  CCTGACATGC AAGTGGCTCG TGGGTATCAG TCATCAGCTA CCATGTCAGA CGGTCGTGTT TTTACCATTG GAGGCTCCTG GAGCGGTGGC GTATTTGAGA
     GGACTGTACG TTCACCGAGC ACCCATAGTC AGTAGTCGAT GGTACAGTCT GCCAGCACAA AAATGGTAAC CTCCGAGGAC CTCGCCACCG CATAAACTCT
                                                                       GOLibR4

                                                                                           QYRN326-333
890  AGAATGGCGA AGTCTATAGC CCATCTTCAA AGACATGGAC GTCCCTACCC AATGCCAAGG TCAACCCAAT GTTGACGGCT GACAAGCAAG GATTGTACCG
     TCTTACCGCT TCAGATATCG GGTAGAAGTT TCTGTACCTG CAGGGATGGG TTACGGTTCC AGTTGGGTTA CAACTGCCGA CTGTTCGTTC CTAACATGGC
                                                                       GOLibR5

990  TTCAGACAAC CACGCGTGGC TCTTTGGATG GAAGAAGGGT TCGGTGTTCC AAGCGGGACC TAGCACAGCC ATGAACTGGT ACTATACCAG TGGAAGTGGT
     AAGTCTGTTG GTGCGCACCG AGAAACCTAC CTTCTTCCCA AGCCACAAGG TTCGCCCTGG ATCGTGTCGG TACTTGACCA TGATATGGTC ACCTTCACCA

                                                                      C383CEKMST
1090 GATGTGAAGT CAGCCGGAAA ACGCCAGTCT AACCGTGGTG TAGCCCCTGA TGCCATGTGC GGAAACGCTG TCATGTACGA CGCCGTTAAA GGAAAGATCC
     CTACACTTCA GTCGGCCTTT TGCGGTCAGA TTGGCACCAC ATCGGGGACT ACGGTACACG CCTTTGCGAC AGTACATGCT GCGGCAATTT CCTTTCTAGG
                                                  GOLibR6

                          Y405XQ406X
1190 TGACCTTTGG CGGCTCCCCA GATTATCAAG ACTCTGACGC CACAACCAAC GCCCACATCA TCACCCTCGG TGAACCCGGA ACATCTCCCA ACACTGTCTT
     ACTGGAAACC GCCGAGGGGT CTAATAGTTC TGAGACTGCG GTGTTGGTTG CGGGTGTAGT AGTGGGAGCC ACTTGGGCCT TGTAGAGGGT TGTGACAGAA
                GOLibR7
```

```
                                                                                                         P463XF464X
1290 TGCTAGCAAT GGGTTGTACT TTGCCCGAAC GTTTCACACC TCTGTTGTTC TTCCAGACGG AAGCACGTTT ATTACAGGAG GCCAACGACG TGGAATTCCG
     ACGATCGTTA CCCAACATGA AACGGGCTTG CAAAGTGTGG AGACAACAAG AAGGTCTGCC TTCGTGCAAA TAATGTCCTC CGGTTGCTGC ACCTTAAGGC
                                                                                                    GOLibR8

                                                                                          V494AMV
1390 TTCGAGGATT CAACCCCGGT ATTTACACCT GAGATCTACG TCCCTGAACA AGACACTTTC TACAAGCAGA ACCCCAACTC CATTGTTCGC GCTTACCATA
     AAGCTCCTAA GTTGGGGCCA TAAATGTGGA CTCTAGATGC AGGGACTTGT TCTGTGAAAG ATGTTCGTCT TGGGGTTGAG GTAACAAGCG CGAATGGTAT
                                                                                        GOLibR9

                                            L514XC515C518X
1490 GCATTTCCCT TTTGTTACCT GATGGCAGGG TATTTAACGG TGGTGGTGGT CTTTGTGGCG ATTGTACCAC GAATCATTTC GACGCGCAAA TCTTTACGCC
     CGTAAAGGGA AAACAATGGA CTACCGTCCC ATAAATTGCC ACCACCACCA GAAACACCGC TAACATGGTG CTTAGTAAAG CTGCGCGTTT AGAAATGCGG
                                                 GOLibR10

1590 AAACTATCTT TACGATAGCA ACGGCAATCT CGCGACACGT CCCAAGATTA CCAGAACCTC TACACAGAGC GTCAATGTCG GTGGCAGAAT TACAATCTCG
     TTTGATAGAA ATGCTATCGT TGCCGTTAGA GCGCTGTGCA GGGTTCTAAT GGTCTTGGAG ATGTGTCTCG CAGTTACAGC CACCGTCTTA ATGTTAGAGC

1690 ACGGATTCTT CGATTAGCAA GGCGTCGTTG ATTCGCTATG GTACAGCGAC ACACACGGTT AATACTGACC AGCGCCGCAT TCCCCTGACT CTGACAAACA
     TGCCTAAGAA GCTAATCGTT CCGCAGCAAC TAAGCGATAC CATGTCGCTG TGTGTGCCAA TTATGACTGG TCGCGGCGTA AGGGGACTGA GACTGTTTGT

1790 ATGGAGGAAA TAGCTATTCT TTCCAAGTTC CTAGCGACTC TGGTGTTGCT TTGCCTGGCT ACTGGATGTT GTTCGTGATG AACTCGGCCG GTGTTCCTAG
     TACCTCCTTT ATCGATAAGA AAGGTTCAAG GATCGCTGAG ACCACAACGA AACGGACCGA TGACCTACAA CAAGCACTAC TTGAGCCGGC CACAAGGATC

1890 TGTGGCTTCG ACGATTCGCG TTACTCAGTG ATCCGAATTC GAGCTCCGTC GACAAGCTTG CGGCCGCACT CGAGCACCAC CACCACCACC ACTGAGA
     ACACCGAAGC TGCTAAGCGC AATGAGTCAC TAGGCTTAAG CTCGAGGCAG CTGTTCGAAC GCCGGCGTGA GCTCGTGGTG GTGGTGGTGG TGACTCT
                                                                   NotI          GOLibR11
```

Figure 4.3. The positions of the oligonucleotides shown in Table 4.2 within the GO sequence

Oligonucleotides containing randomisations are bold and underlined while the non-mutated reverse oligonucleotides are shown in green. Randomised positions are shown in magenta and the *Mfe*I and *Not*I restriction sites are shown in red text.

The basic protocol for library construction is shown in Figure 4.4. The oligonucleotides containing the randomisation(s) (grey) were all designed as forward (coding strand) primers. For each of these primers, there is a corresponding reverse primer (blue) (Table 4.2) which, in conjunction with an upstream forward primer permits amplification of a region of DNA (Figure 4.4A). The reverse primers are designed so that the neighbouring amplified fragments will overlap by 13-17 base pairs permitting joining of the fragments by overlap extension (Section 2.4.2) (McPherson and Moller, 2006) (Figure 4.4B). The forward and reverse primers shown in red do not contain mutations and ensure that the randomised region (Figure 4.4C) can be digested by restriction enzymes *Mfe*I (magenta) and *Not*I (cyan) (Section 2.4.5) and ligated into the digested plasmid (Section 2.4.7) ready for transformation into *E. coli* (Sections 2.4.8 and 2.4.9).
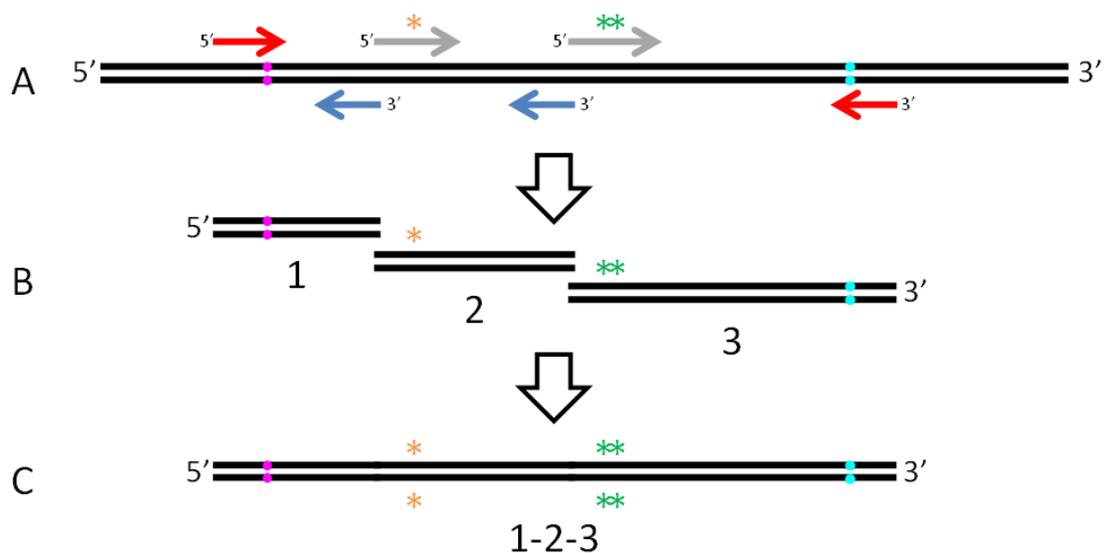
Figure 4.4. Schematic showing the protocol for construction of a library using two randomised oligonucleotides

> The randomised primers (grey) with mutations shown as green or yellow asterisks; reverse primers for generating each fragment are shown in blue; primers for amplifying from each end of the mutated region are shown in red; and the *Mfe*I and *Not*I restriction sites are shown as magenta and cyan circles.

With the ten different oligonucleotides that were designed, it is possible to generate a number of different libraries of varying sizes. Initially, it was decided to create 16 libraries which varied in complexity as well as the areas of the active site targeted (Table 4.3).

Table 4.3. Libraries generated using the randomised oligonucleotides

Library	Mutations introduced	Number of residues targeted	Number of mutants in library	Screening effort required for 95% completeness	Agar plate-based screens	Microtitre plate-based screens
A	Q326X/ Y329X/ R330X/ N333X/ P463X/ F464X	6	4.70×10^7	1.4×10^8	7047	
C	W290AFHWY/ L514X/ C515X/ C518X	4*	38000	1.1×10^5	6	
D	F194X/ P463X/ F464X	3	6859	20548	1	
E	Q326X/ Y329X/ R330X/ N333X	4	1.30×10^5	3.9×10^5	20	
F	F194X/ F227X/ W290AFHWY	3*	1805	5407	<1	
H	F194X/ F227X/ Y405X/ Q406X/ P463X/ F464X	6	4.70×10^7	1.4×10^8	7047	
I	F194X	1	19	57		<1
J	F227X	1	19	57		<1
K	N245X	1	19	57		<1
L	Y405X/ Q406X	2	361	1081	<1	
M	P463X/ F464X	2	361	1081	<1	
N	L514X/ C515X/ C518X	3	6859	20548	1	
O	N245X/ C383CEKMST (*n.b.* same as R)	2*	95	285		3
P	F194X/ C383CEKMST	2*	95	285		3
Q	F227X/ C383CEKMST	2*	95	285		3
R	N245X/ C383CEKMST	2*	95	285		3

The residues targeted in the 16 different libraries are shown along with the theoretical; number of different variants present in the library. Where the number of residues targeted is shown with an asterisk, one of the residues is only mutated to a subset of residues. The screening effort required is determined using the equation: T = -V ln (1-P) where T is the number of transformants to screen, V is the number of gene mutants in the library and P is the % completeness desired (in this case 95%) (Reetz et al., 2008).

Library A was designed to target two areas on opposite sides of the active site. However, after generation, based on the assay approach used, it was determined that the screening effort to statistically cover even 5% of this library was unrealistic. The same was true for **Library H** which also randomised six residues. This library was designed to target an entire side of the active site as four neighbouring loops were selected. However, to statistically cover only 5% of the library would require screening of over 120 large agar plates. **Library D** was designed to

target the 'hydrophobic surface' (Sun et al., 2002) which potentially interacts with C4, C5 and C6 of the D-galactose molecule while **Library E** was made with a single randomised oligonucleotide to explore the effect of randomising four residues in the loop containing Arg330. **Library F** combined mutation of the stacking tryptophan W290 with randomisation of two nearby residues F194 and F227; while **Library C** also involved mutation of the stacking tryptophan but this time combined with the relatively uncharacterised L514, C515 and C518.

Oligonucleotides which contained residues not previously subjected to mutagenesis studies (*i.e*. excluding W290, C383 and V494) were used individually to generate relatively small libraries which could easily be screened to 95% completeness: **Libraries I – N**. **Libraries I, J** and **K** were also combined with mutation at position 383 to produce **Libraries P, Q** and **R**, respectively. **Library O**, which was originally designed to mutate position 383 alone, actually included randomisation at position 245 as well (in error), effectively duplicating **Library R**.

4.3 Library generation

Library generation was optimised with **Library A** (Q326X/ Y329X/ R330X/ N333X/ P463X/ F464X) and then repeated for generation of each of the other libraries.

4.3.1 Generation of the mutant insert

As the library uses two randomised oligonucleotides, three fragments were initially generated: fragment 1 using GOlibF1 and GOLibR5 primers, fragment 2 using QYRN326-333 and GOLibR8 primers; and fragment 3 using P463XF464X and GOLibR11 primers (Figure 4.3 and Figure 4.4). PCRs were carried out using Phusion polymerase (Section 2.4.1) and an extension time of 20 seconds. The expected fragment sizes from the three PCRs were 473 bp, 431 bp and 604 bp, respectively, and these were confirmed by agarose gel electrophoresis (Section 2.4.4) (Figure 4.5).
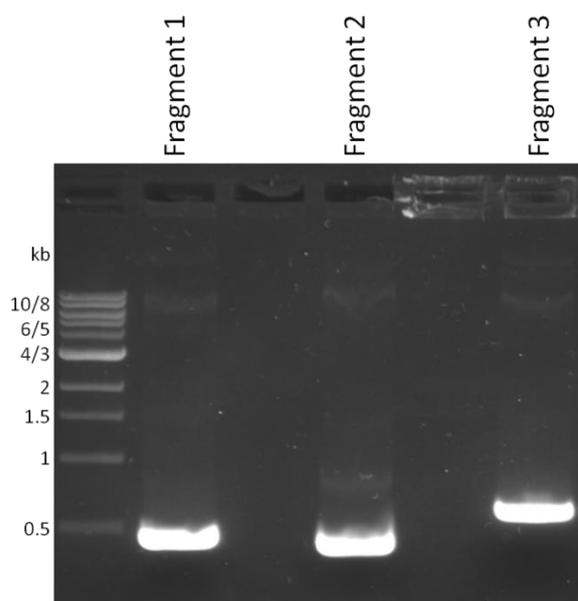
Figure 4.5. Amplification of fragments required for Library A

> The three fragments were generated by PCR using GOLibF1 and GOLibR5; QYRN326-333 and GOLibR8; and P463XF464X and GOLibR11 primers, respectively, and then purified by gel extraction.

In order to combine the fragments by overlap extension (Section 2.4.2), equimolar quantities of each fragment were combined to give a total amount of 2 µg. The online tool at www.encorbio.com/protocols/Nuc-MW.htm was used to determine the molecular mass and,

114

following gel extraction using the QIAquick Gel Extraction kit (Qiagen) (Section 2.4.12), appropriate volumes of each fragment were combined with Phusion PCR components and the overlap extension reaction carried out with an extension time of 1.5 or 3 minutes. As shown in Figure 4.6, the overlap extension successfully joined fragments 1 and 2 as well as fragments 2 and 3, but there is no band corresponding to the combination of all three fragments.
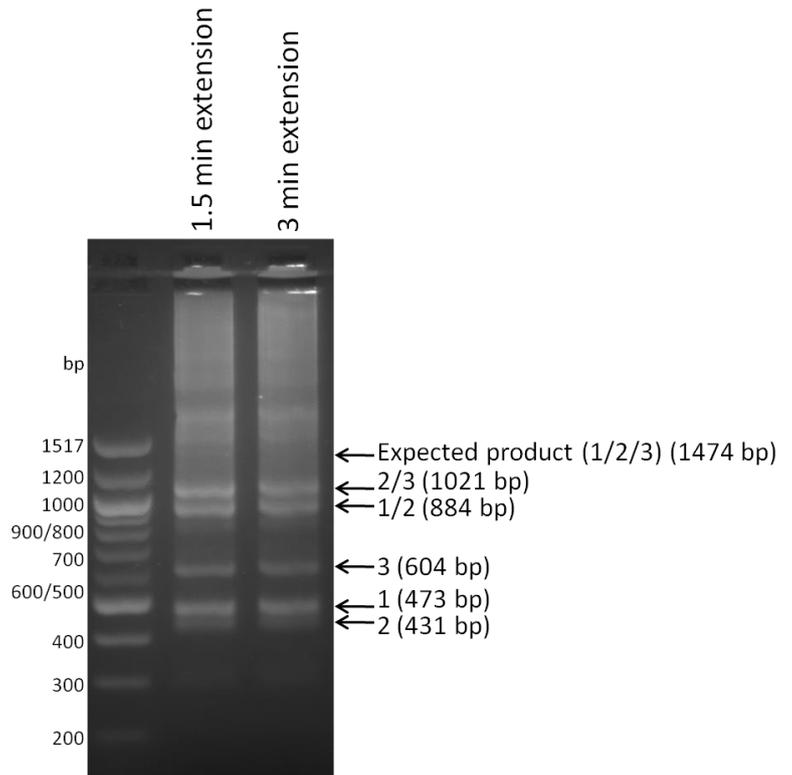
Figure 4.6. Attempts to combine all three fragments in a 'one pot' reaction

> Bands are clearly visible at sizes corresponding to each fragment individually as well as fragment 2 combined with each of the other fragments, however there is no band for all three fragments combined.

It was therefore decided to combine the fragments in a stepwise manner, first combining fragments 1 and 2 to create fragment 1/2 (Figure 4.7A), then, following gel extraction, combining this with fragment 3 in a separate overlap extension. As shown in Figure 4.7B, two annealing temperatures were trialled in the second overlap extension and 64 °C resulted in more product than 54 °C.
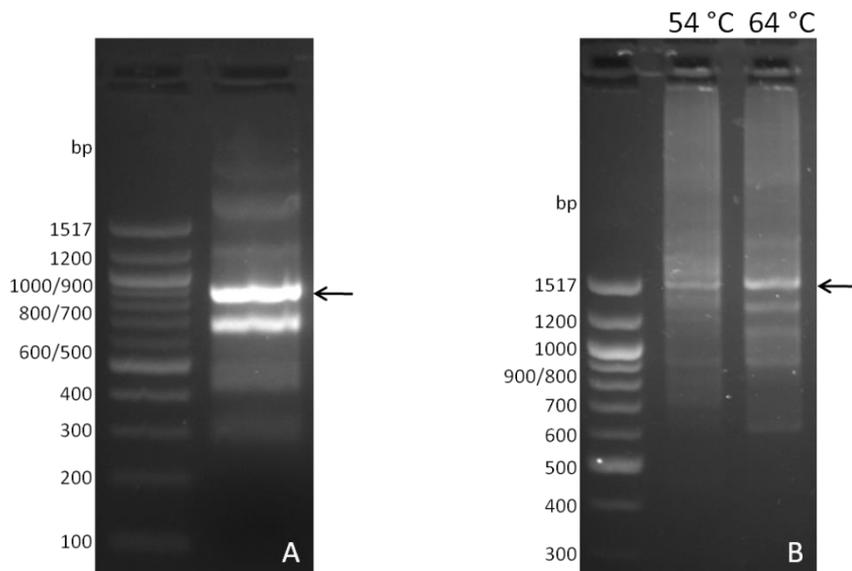
Figure 4.7. Combining fragments by overlap extension

Overlap extension was carried out with **A**: fragments 1 and 2 to create an 884 bp product and **B**: fragments 1/2 and 3 to create a 1474 bp product. For the second reaction two annealing temperatures were tested and 64 °C appeared to generate more of the desired product. Both products were purified by gel extraction.

4.3.2 Ligation of the mutated insert into the pET28c-N6M1 plasmid

In order to ligate the combined fragments (insert) into the rest of the GO-N6M1 gene within the pET28c plasmid, both were double digested with the same restriction enzymes (Section 2.4.5). The plasmid DNA was double digested with *Mfe*I and *Not*I for >4 hours. This was followed by dephosphorylation (Section 2.4.6) to prevent spontaneous re-ligation before separation from the smaller WT fragment by agarose gel electrophoresis (Section 2.4.4) (Figure 4.8).
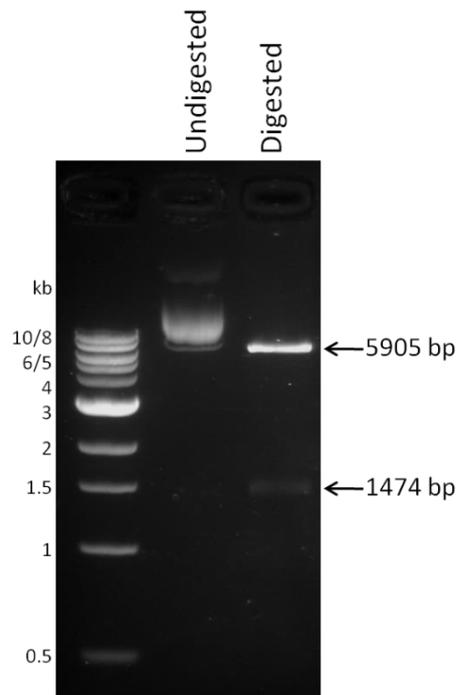
Figure 4.8. pET28c-N6M1 was double digested ready for insertion of the mutant insert

The agarose gel shows the pET28c-N6M1 before and after digestion with *Mfe*I and *Not*I to remove the 1474 bp region of the GO gene, which will be replaced by the mutated version upon ligation. The 5905 bp species was purified by gel extraction.

In order to optimise ligation of the insert into the plasmid and not waste the mutated fragments, WT insert was generated by PCR using GOLibF1 and GOLibR11 primers and an extension time of 2 minutes (Section 2.4.1). The correct band (1474 bp) was purified from an agarose gel by gel extraction (Section 2.4.12). Double digestion was carried out with *Mfe*I and *Not*I for >4 hours (Section 2.4.5) before removal of the restriction enzymes and the digested ends using the QIAquick PCR Purification kit (Qiagen) (Section 2.4.13).

In optimising the ligation reaction, four conditions were varied: the incubation time, the plasmid to insert ratio, the amount of DNA present in the 10 µl reaction and the manufacturer of the T4 DNA ligase. Ligases from Roche and from NEB were selected; the instructions for the Roche ligase recommend total DNA amounts of 30 – 300 ng while NEB recommends 10 – 100 ng. Amounts of 30, 100 and 300 ng were therefore selected for the Roche ligase and 10, 50 and 100 ng for NEB. As recommended by the manufacturers, plasmid to insert ratios of 1:1 and 1:2 were trialled for the Roche ligase while 1:2 and 1:4 were trialled for the NEB ligase. Finally, for both ligases, incubation times at 4 °C of 2 hours and overnight were compared. From each

reaction, 5 ng of DNA was transformed into *E. coli* (Section 2.4.8) and the number of colonies that grew used to measure the success of the ligation reaction (Figure 4.9).
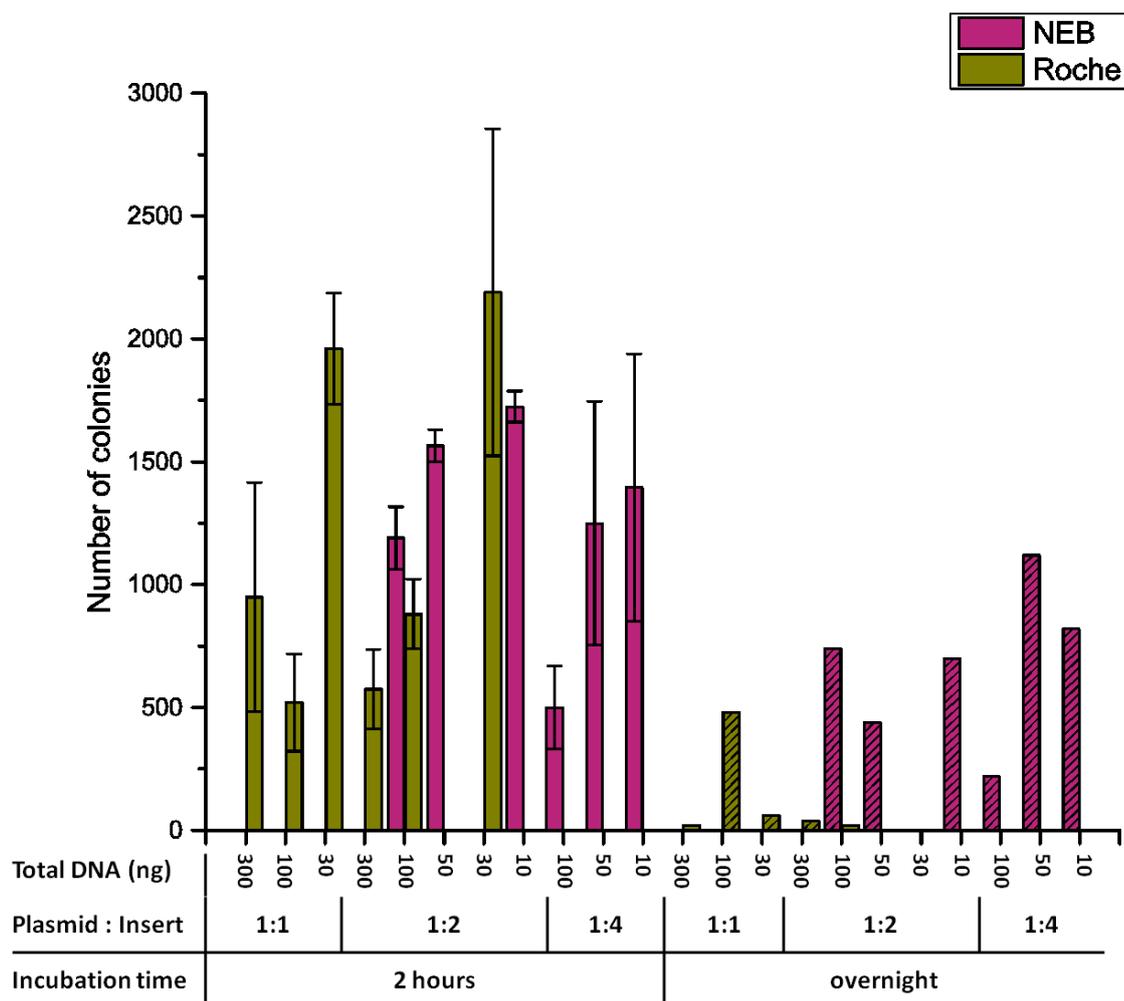
Figure 4.9. A comparison of different ligation conditions

Different ligation reactions were carried out where the manufacturer of the T4 DNA ligase, the incubation time, the plasmid to insert ratio and the total amount of DNA present were varied. 5 ng of DNA from each reaction was transformed into *E. coli* and the number of colonies counted.

In general incubating the ligation reaction for 2 hours is more efficient than incubating overnight, especially for the Roche enzyme; and smaller quantities of DNA in the reaction led to higher transformant colony numbers. A 1:2 ratio of plasmid to insert seemed slightly better than the other ratios trialled. However, in order to save materials it is desirable to use less of the mutant insert so a 1:1 ratio was selected as more optimal overall. The conditions selected for use in ligations of the mutant fragment into the plasmid were therefore:

- T4 DNA ligase: Roche

- total DNA in 10 µl reaction: 10 ng

- plasmid to insert ratio: 1:1

- incubation time: 2 hours

4.3.3 Transformation into an appropriate expression strain

Following the ligation reaction the plasmid must be transformed into an appropriate strain of *E. coli* for the agar plate-based assay developed in Chapter 3. In order to permit screening of as many library variants as possible, very high transformation efficiencies are desirable. To this end, the transformation efficiencies of six different competent cell strains were compared (Table 4.4). Transformations were carried out according to the manufacturers' protocols using 40 ng of the ligated plasmid and 25 µl of cells (Section 2.4.8 and 2.4.9).

Table 4.4. Comparison of different competent cell strains for transformation with the library construct

Strain	Supplier	Electro-/ chemocompetent	CFU/ µg
Acella	VH Bio	Electrocompetent	6×10^7
XL1 Blue	Stratagene	Electrocompetent	2×10^7
ER2738	Lucigen	Electrocompetent	9×10^6
T7 Express	NEB	Chemocompetent	3×10^4
BL21 Gold (DE3)	Stratagene	Chemocompetent	1×10^4
BL21 Star (DE3)	Invitrogen	Chemocompetent	2×10^2

40 ng of the ligated plasmid was transformed into six different strains of *E. coli* and plated out at an appropriate dilution. The number of colony forming units (CFU) was counted the following day.

Acella cells from VH Bio showed the highest transformation efficiency. This was ideal as these are the only *electrocompetent* cells trialled which also carry a copy of the T7 RNA polymerase permitting expression of the GO protein. If either XL1 Blue or ER2738 were to be used, extra steps would have to be added to express the library variants. The chemocompetent cells trialled showed significantly lower transformation efficiencies, as expected with this transformation method.

4.3.4 Library construction and confirmation of diversity

Despite scaling up the initial PCR to amplify the fragments and introduce mutations, due to loss of DNA in the gel extraction and PCR purification stages and inefficiencies in the overlap extensions, in some libraries there was inadequate insert produced for the ligation reaction and subsequent transformation. It was therefore necessary to amplify the insert by PCR before the double digestion. This was carried out using KOD polymerase, GOLibF1 and GOLibR11 primers and an extension time of 20 seconds (Section 2.4.1). In order to reduce the possibility of bias being introduced into the library, only five cycles of PCR were carried out and this produced sufficient insert in each case for the ligation reaction. The final library construction procedure is shown in Figure 4.10 and was used to construct the 16 libraries shown in Table 4.3.

Before the large scale transformation of the library into Acella cells, a small aliquot was transformed into XL1 Blue cells (Section 2.4.8). Plasmids were purified from 10 colonies by DNA miniprep (Section 2.4.11) and sent for sequencing (Section 2.4.15) to assess library quality. While data from 10 variants is not adequate to confirm the expected diversity, it was sufficient to confirm that mutation(s) had been introduced at the correct position(s) and that there was not a significant level of errors such as frameshifts, insertions or deletions. In all cases, at least seven out of the 10 sequenced variants contained no errors and confirmed introduction of different residues at the desired position(s).

1. Amplify fragments 1, 2 & 3 by PCR

2. Gel extract correct bands

3. Combine fragments 1 & 2 by overlap extension

4. Gel extract correct bands

5. Combine fragments 1/2 & 3 by overlap extension

6. PCR purification to remove Phusion reaction components

7. Five cycles of PCR to amplify the insert

8. PCR purification to remove KOD reaction components

9. Double digestion for > 4 hours

10. Gel extract band at 1474 bp

11. Ligate into double digested, dephosphorylated plasmid

12. PCR purification to remove ligation reaction components

13. Transform a small aliquot into XL1 Blue cells

14. Miniprep and sequence ten variants

15. Transform library into Acella cells and plate onto large agar plates

16. Scrape off colonies, resuspend in 2TY containing 20% glycerol

17. Aliquot and freeze at -80 °C ready for screening

Figure 4.10. General protocol for library construction

The protocol for construction of a library using two randomised oligonucleotides is shown. If only one randomised oligonucleotide is used, steps 3 and 4 would be omitted, if more than two oligonucleotides are used, additional overlap extensions would be carried out.

The primary limitation on sequencing more than 10 variants from every library produced was the cost of the sequencing reactions. However, in order to confirm that the expected diversity was being introduced and that the statistical approach to determining the screening effort required was appropriate, 88 variants from **Library K** (N245X) were sequenced (Figure 4.11).
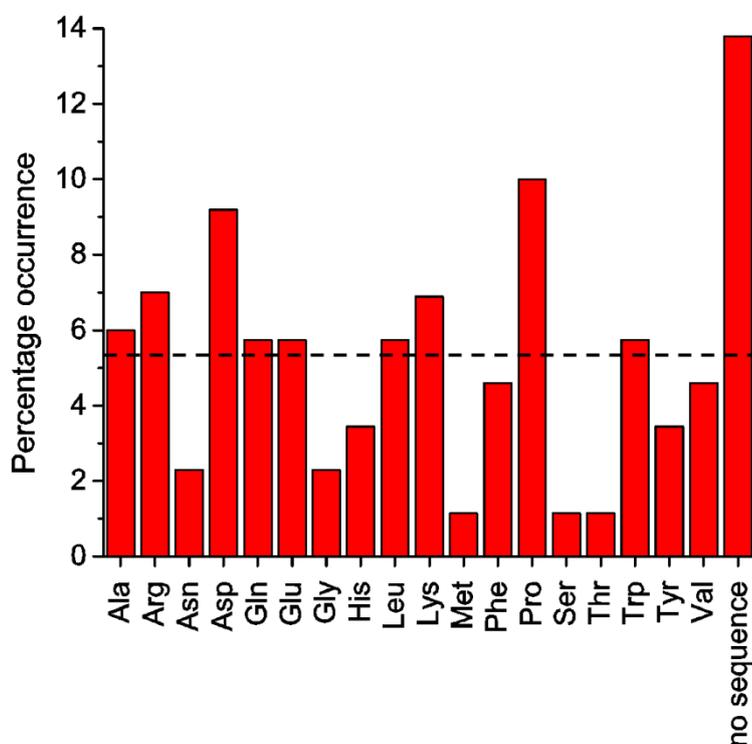
Figure 4.11. The results of sequencing 88 variants from Library K (N245X)

In order to assess library diversity in an example library, 88 variants from **Library K** were sequenced. The dashed line at 5.33% shows the expected proportion of each residue if randomisation was perfect. The proportion of some residues shows significant variation from this value but all amino acids except Ile are represented.

Unfortunately 12 of the variants failed to sequence which could be due to errors in the library construction which led to disruption of the primer binding site, but could also be due to issues in plasmid purification or the sequencing process. The sequencing data from the remaining 76 variants revealed a diverse library with only Ile not represented at position 245.

4.4 Library screening

4.4.1 Agar plate-based screening

WT GO-N6M1 shows high enough levels of activity against glycerol and D-xylose that a colour change is visible with the agar plate-based assay. On the other hand, the level of activity with D-arabinose, D-mannose and D-glucose is so low that a colour change is not visible. When screening the different mutant libraries, the substrates were divided into two groups: a) glycerol & D-xylose and b) D-arabinose, D-mannose & D-glucose. In screening against the first set of substrates, any colonies showing a colour change could be expressing WT so sequencing was required to identify active mutants and confirmation of results by the quantifiable microtitre plate-based screen was essential to identify any enhancements in activity compared to WT. When screening against the substrate mix of D-arabinose, D-mannose and D-glucose however, any colony showing a colour change is likely to be expressing a mutant showing higher activity than WT against one or more of the substrates provided, unless the mutation has resulted in significant increases in expression levels as discussed later.

The libraries screened using the agar plate-based assay are shown in Table 4.5. Due to the size of libraries and therefore the screening effort required, it was not always possible to screen to 95% completeness and two of the libraries were not screened at all (Section 4.2.2). These are available for future screening efforts. It was also not possible to carry out microtitre plate-based screening on all mutants identified in the agar plate-based screening but, again, these are available for future projects.

Table 4.5. The libraries screened using the agar plate-based screen

Library	Mutations introduced	Number screened (approximately)	Completeness (%)
A	Q326X/ Y329X/ R330X/ N333X/ P463X/ F464X	6.0×10^5	1.3
C	W290AFHWY/ L514X/ C515X/ C518X	0	0
D	F194X/ P463X/ F464X	2.0×10^5	>95
E	Q326X/ Y329X/ R330X/ N333X	4.0×10^4	27
F	F194X/ F227X/ W290AFHWY	2.0×10^5	>95
H	F194X/ F227X/ Y405X/ Q406X/ P463X/ F464X	0	0
L	Y405X/ Q406X	6.0×10^5	>95
M	P463X/ F464X	6.4×10^5	>95
N	L514X/ C515X/ C518X	6.0×10^5	>95

The percentage completeness in the screening effort was determined using the equation: T = -V ln (1-P) where T is the number of transformants to screen, V is the number of gene mutants in the library and P is the % completeness (Reetz et al., 2008). **Library D** was screened by Ms Lito Paraskevopoulou; **Library F** was screened by Mr Ciaran Doherty.

No enhancements in activities were identified upon screening of **Library A**. This could be due to the fact that a very small proportion of the library was screened, but it could also be that mutating these six residues in combination leads to a large number of deleterious mutations. The two oligonucleotides used to generate this library were used separately to generate two smaller libraries: **Library E**: Q326X/ Y329X/ R330X/ N333X and **Library M**: P463X/ F464X.

Library E contained a number of mutants showing activity against glycerol and/ or D-xylose (Table 4.6) although whether these activities are greater than WT would have to be determined using the microtitre plate-based screen. However, no activity was seen against D-arabinose, D-mannose or D-glucose. This could be because statistically, only 27% of the library was screened. From the sequencing data of the glycerol/ D-xylose hits it can be seen that the charged residues lysine and particularly arginine at position 333 are significantly favoured over other residues but no other trends are seen at the other three positions.

Table 4.6. Variants from Library E showing activity against glycerol and/ or D-xylose

Q326	Y329	R330	N333
Q	N	Y	R
V	M	F	R
S	W	M	R
M	Q	Q	R
R	Q	E	K
L	L	A	R
E	Q	S	K
Y	V	F	R
M	A	I	R
N	K	S	S

Library M, which was screened to greater than 95% completeness, revealed variants active against both sets of substrates (Table 4.7). The only trend visible in the mutants identified is that a hydrophobic residue (Ile, Val or Trp) seems to be favoured at position 463 in variants showing activity against glycerol and/ or D-xylose. As before, the microtitre plate-based screen would need to be carried out on these variants to determine if any show higher activity than WT against glycerol or D-xylose.

Table 4.7. Variants from Library M showing activity against alternative substrates

Active against glycerol and/ or D-xylose			Active against D-arabinose, D-mannose and/ or D-glucose	
P463	F464		P463	F464
I	G		H	Q
V	S		P	E
V	Y			
W	K			
W	W			

Library L: Y405X/ Q406X was also screened to >95% using the agar plate-based screen. In an initial screen of 200 variants the only variants showing activity against D-galactose contained a tyrosine at position 405, as in the WT enzyme. Substitution at this position with a threonine, lysine or tryptophan appeared to abolish galactose oxidase activity, although sequencing of a number of active variants revealed substitutions at position 406. The screens of **Library L** against alternative substrates produced similar results with only two hits identified: Q406E

active against glycerol and/ or D-xylose and Q406K active against D-arabinose, D-mannose and/ or D-glucose.

Screening of **Library N**: L514X/ C515X/ C518X to >95% completeness revealed five mutants with altered activities (Table 4.8). From this small dataset it appears that a hydrophobic residue is favoured at position 514 (Leu, Val or Met) in variants showing activity towards alternative substrates, and that the disulfide bond between Cys515 and Cys518 in the WT structure is not required for activity against these substrates as various substitutions were observed at these positions.

Table 4.8. Variants from Library N showing activity against alternative substrates

Active against glycerol and/ or D-xylose				Active against D-arabinose, D-mannose, and/ or D-glucose		
L514	C515	C518		L514	C515	C518
V	A	A		V	A	A
L	G	F		M	M	F
L	W	E		W	W	A

Approximately 2 x 10^5 colonies from **Library D**: F194X/ P463X/ F464X and from **Library F**: F194X/ F227W/ W290AFHWY were screened by Ms L Paraskevopoulou and Mr C Doherty, final year undergraduate project students under my supervision. 21 variants from **Library D** and 14 variants from **Library F** showed activity against glycerol and/ or D-xylose while only one variant from each library was identified with activity against D-glucose, D-arabinose and/ or D-mannose. These 38 variants were isolated and the screens repeated using the microtitre plate-based assay (Section 3.3).

4.4.2 Microtitre plate-based screening

Data from the microtitre plate-based screens was analysed in different ways depending on which substrate was being screened. WT GO-N6M1 shows sufficiently high activity against glycerol and D-xylose that the colour change was clearly visible by eye (Figure 4.12). In the case of D-glucose, D-arabinose and D-mannose however, the colour change was much more subtle and thus a plate reader was used to measure A_{414} values. Blank readings were then subtracted and the A_{414} values ranked in order of magnitude and plotted against A_{414} (Figure 4.13). This clearly revealed any wells showing a higher level of absorbance than others, thus identifying mutants to be taken forward for further analysis.
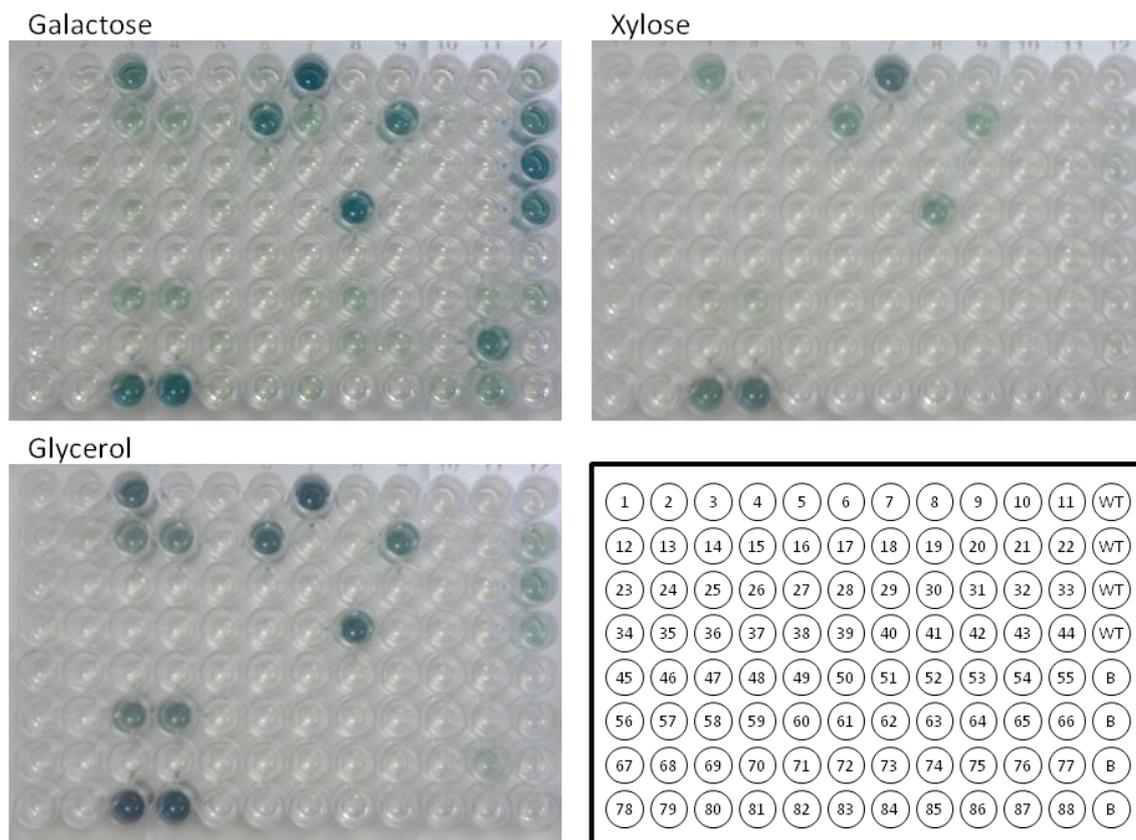
Figure 4.12. Photographs of the microtitre plate-based screens of Library P: F194X/ C383CEKMST

Screens against D-galactose (top left), glycerol (bottom left) and D-xylose (top right) are shown as an example of the screening methodology. Variants are numbered 1 to 88 with positive and negative controls in the final lane, where WT is WT GO-N6M1 enzyme with the full assay mix and B is WT GO-N6M1 with HRP excluded (bottom right).
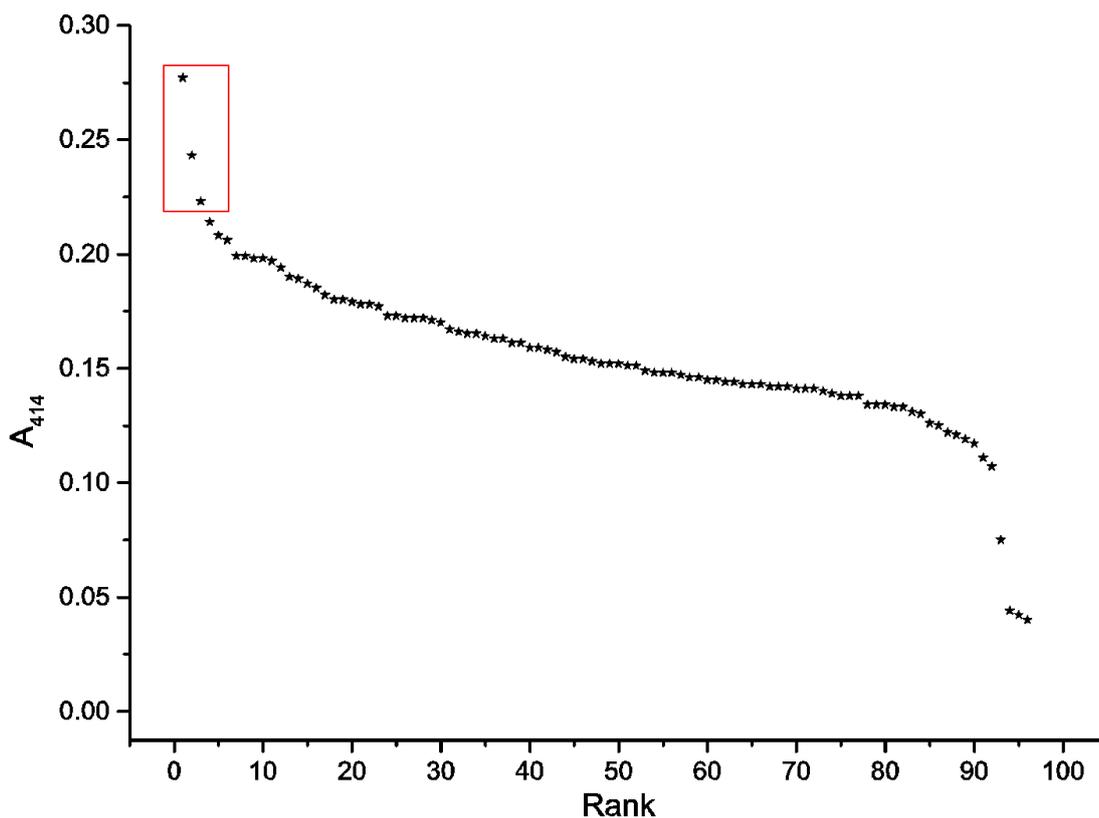
Figure 4.13. Absorbance at 414 nm readings ranked according to magnitude

The data show absorbance readings for variants from **Library P**: F194X/ C383CEKMST in a screen with D-glucose as the substrate. While no colour change was visible by eye, analysing the data by this method identified three potential hits: P72, P83 and P74 (red box).

All microtitre plate-based assays were carried out on two separate days to ensure reproducibility of the results and that anomalous results (positive and negative) did not prevent identification of variants of interest. Plasmids encoding any variants showing increased activity compared to WT were purified by DNA miniprep and sent for DNA sequencing to identify the mutations present.

Although the majority of sequencing reactions allowed identification of the mutations present, reactions for a number of the mutants failed, despite repeats, implying that the primer binding site was not present in the sample provided. In some cases, sequencing data revealed duplicates of mutants, which was reassuring as this confirmed the assay results. Almost 4% of variants were also shown to contain frameshifts, insertions, deletions and mutations at multiple positions, rather than just those which had been targeted. This is most likely due to

errors in the proof-reading mechanism of the polymerases used in library construction. Most of the mutants containing these errors displayed very low or no activity towards D-galactose, implying that the activity observed with other substrates may be an artefact of the screens. An example of this is the mutant P72 which appeared to have increased D-glucose activity (Figure 4.13) but no D-galactose activity – sequencing revealed a frameshift mutation near the beginning of the gene. The observed D-glucose activity for this variant is possibly due to a screening artefact such as contamination with another variant.

The libraries screened using the microtitre plate-based assay and the number of hits identified from each library are shown in Table 4.9. The seven smaller libraries were only screened using this assay, whereas hits from agar plate screening of **Libraries D and F** were confirmed using the microtitre plate-based screen. Unfortunately none of the variants from **Libraries D and F** which showed activity against D-arabinose, D-glucose and/ or D-mannose showed such activities in the microtitre plate-based screen. However, some of the other hits proved to be reproducible: P463V from **Library D** and F194W/ W290F from **Library F** both showed enhanced activity towards glycerol.

Table 4.9. The libraries screened using the microtitre plate-based screen

Library	Mutations introduced	Number screened	Completeness (%)	No. hits identified
D	F194X/ P463X/ F464X	19 (confirmation)	n/a	1
F	F194X/ F227X/ W290AFHWY	15 (confirmation)	n/a	1
I	F194X	88	>95	3
J	F227X	88	>95	1
K	N245X	88	>95	6
O	N245X/ C383CEKMST	88	60	2
P	F194X/ C383CEKMST	88	60	10
Q	F227X/ C383CEKMST	88	60	7
R	N245X/ C383CEKMST	88	60	6

The percentage completeness in the screening effort was determined as before. The number of hits identified excludes duplicates and only includes those for which sequencing data was attainable. **Library D** was screened by Ms Lito Paraskevopoulou; **Library F** was screened by Mr Ciaran Doherty.

In order to explore the basis for some of the failed sequencing reactions, analytical digests with *Mfe*I and *Not*I were carried out on a selection of the plasmids for which sequencing reactions had failed. Analysis by agarose gel electrophoresis (Figure 4.14) revealed that, in all

but one case, the digestion did not result in fragments of the expected size. This suggests that a significant insertion or deletion has occurred during construction of these plasmids which may have deleted the binding site of the sequencing primer. Positive results seen in screens with these mutants seem to be artefacts.
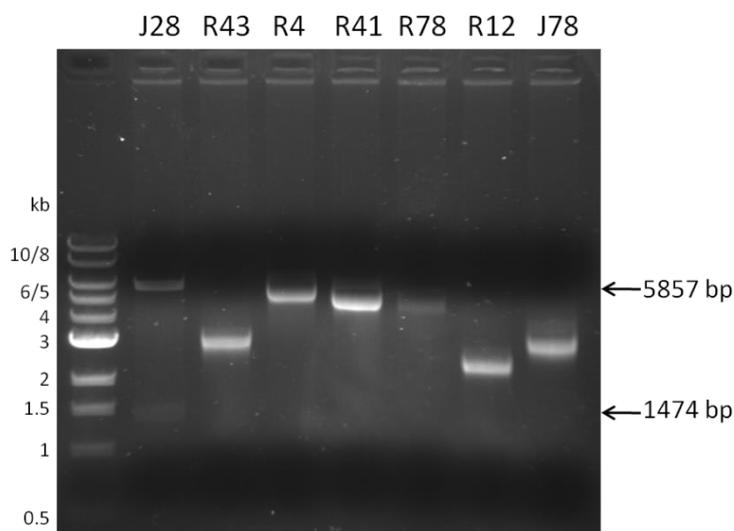
Figure 4.14. Analytical digests of a selection of plasmids for which sequencing reactions failed on more than one occasion

Double digests were carried out with *Mfe*I and *Not*I. The expected products for the intact pET28c-GO-N6M1 plasmid are 5857 and 1474 bp. Only the plasmid containing mutant J28 displays both of these bands.

4.5 Confirmation of screen results

4.5.1 Purification of the mutants

Before large scale expression of variant proteins showing enhanced activities from **Libraries D, F, I, J, K, O, P, Q and R** it was necessary to delete the secretion signal sequence coding region to allow subsequent intracellular protein expression as detailed in Section 3.4. The resulting plasmids were used to transform BL21 Star (DE3) cells and the variant proteins were expressed by autoinduction before purification by Strep-Tactin chromatography.
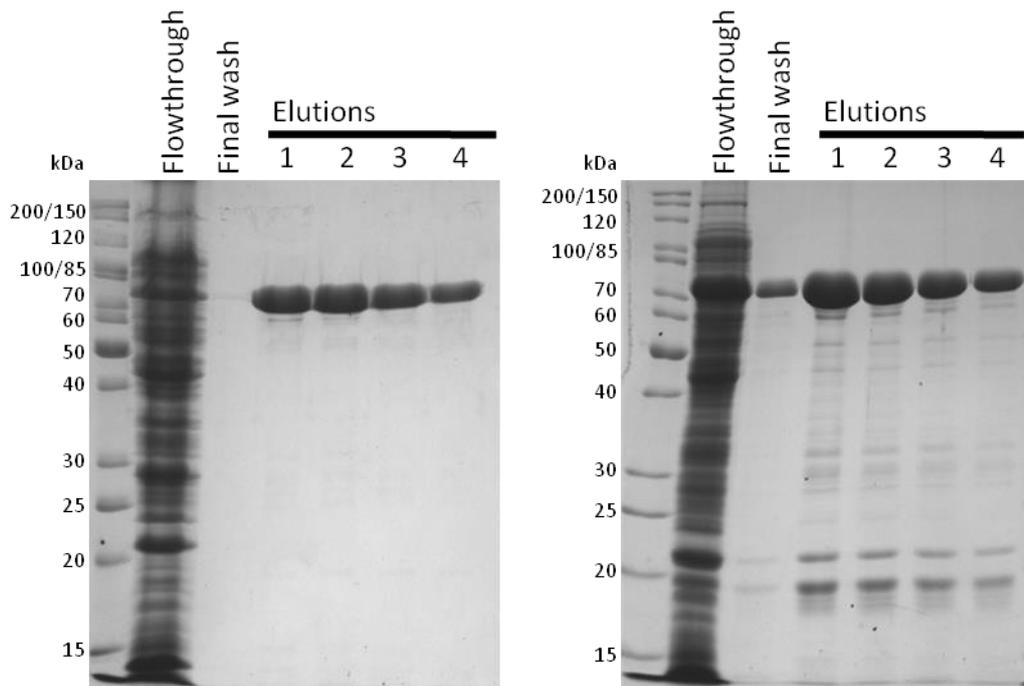
Figure 4.15. SDS PAGE gels of samples taken during two different purifications of WT GO-N6M1

Samples taken at different stages of the purification process were analysed by SDS-PAGE. The strepTEVN6M1 construct runs at approximately 70 kDa while the lower molecular weight bands seen in the elution fractions were unidentified.

Despite an identical protocol being carried out each time, on several occasions multiple bands of lower molecular mass were visible by SDS-PAGE analysis of the purified GO (Figure 4.15). The presence or absence of the bands showed no correlation with activity and even two purifications of the same construct could show marked differences in the levels of 'contaminating' bands. WT N6M1, which showed high levels of the lower molecular weight bands, still demonstrated kinetic behaviour within the range of published values. These bands could be a) contaminating proteins not removed during the purification process; b) degradation products of the GO protein; or c) due to errors in translation leading to truncated products. The first possibility seems unlikely due to the high affinity of the Strep-Tactin resin for the Strep II-tag present on GO (Voss and Skerra, 1997). There are no other proteins present in the cell lysate with a Strep II-tag so, unless these proteins are attached to GO itself, this does not explain how the low molecular weight bands elute at the same time as GO. The second possibility also seems unlikely as the cell lysate is filtered through a 0.2 µm filter to remove any bacterial contamination and protease inhibitor cocktail tablets are included in the lysis buffer to prevent proteolysis. The only way proteolysis could occur is if it is a) autocatalytic (which

only occurs in the WT enzyme to cleave the pro-region (Section 1.6.4) (Rogers et al., 2000); or b) catalysed by a protease not inhibited by the protease inhibitor cocktail used, which is unlikely. The final possibility also seems unlikely due to the maintenance of galactose oxidase activity (see below) which would not be expected in a version of GO lacking large amounts of the C-terminal end of the protein.

When carrying out analyses on the purified protein, concentration was determined from the A_{280} reading. If the 'contaminating' bands were not part of the GO protein, the A_{280} readings would be inaccurate as this method measures total protein in the solution, not just the protein of interest. As k_{cat} values for the WT enzyme, where these bands are also observed, are within the range of published values, and the relative proportion of these bands show no correlation with activity, it implies that the bands are not contaminants and are in fact part of GO. The most reasonable explanation is that the bands are products of cleavage of the protein backbone but when folded, the protein is held together with intramolecular interactions such as hydrogen bonds. When the protein is denatured to be analysed by SDS-PAGE these intramolecular interactions are disrupted causing the fragments to dissociate. This hypothesis would be most easily explored by native PAGE, however due to the high negative charge on its surface, it is not possible to analyse GO in this way. The only remaining question is what causes the cleavage in the first place but as the presence of the low molecular weight bands does not appear to affect activity, exploration of this problem was not considered a priority. In future, the gel bands could be identified using trypsin digestion and MALDI mass spectrometry in order to try and identify the cause of the proposed cleavage and try to prevent it.

4.5.2 Determination of specific activities

Following purification, specific activities were determined for the 37 variants identified by screening against the substrates with which enhanced activities had been detected, as well as D-galactose (Table 4.10, Table 4.11, Table 4.12 and Table 4.13). Substrate concentrations were selected for D-galactose (0.6 M), glycerol (1.5 M) and D-xylose (1.5 M) based on concentrations which appear to be saturating for the WT enzyme with these substrates. D-arabinose, D-mannose and D-glucose were used at 2 M as, whilst this does not appear to be saturating for WT GO, concentrations higher than this are likely to be inhibitory due to viscosity and reduced solubility of the sugars. Each variant is labelled with a letter and a number whereby the letter refers to the library number (I, J, K, P, Q or R) as detailed in Table 4.3, and the number refers to the position of the variant in the microtitre plate as shown in the bottom right of Figure 4.12. In the case of **Libraries D and F**, as only one variant was identified from each library, no number was used.

Table 4.10. Specific activities of mutants from Libraries D and F

Variant	Mutations	0.6 M D-GALACTOSE	1.5 M GLYCEROL
	WT	550 ± 25	64 ± 4.5
D	P463V	370 ± 50	95 ± 10
F	F194W/ W290F	28 ± 1.8	140 ± 10

For each variant, specific activity values were compared to WT activity with that substrate (100%): values <25% are coloured dark blue, 25-90% are pale blue and 110-1000% are yellow.

Table 4.11. Specific activities of mutants from Libraries I and P (containing mutations at position 194)

Variant	Mutations	0.6 M D-GALACTOSE	1.5 M GLYCEROL	1.5 M D-XYLOSE	2.0 M D-MANNOSE	2.0 M D-ARABINOSE	2.0 M D-GLUCOSE
	WT	550 ± 25	64 ± 4.5	3.3 ± 0.14	89 ± 4.6	1.2 ± 0.04	0.15 ± 0.02
P19	F194D/ C383E	28 ± 0.84		3.5 ± 0.16			0.51 ± 0.03
P28	F194E/ C383S	69 ± 2.4				0.62 ± 0.02	
I8	F194G	29 ± 0.73		0.78 ± 0.03			
P88	F194G/ C383S	52 ± 0.85		5.2 ± 0.21		3.9 ± 0.16	2.0 ± 0.07
P76	F194H/ C383E	74 ± 5.2		12 ± 0.74		6.0 ± 0.20	0.02 ± 0.01
P74	F194I/ C383S	76 ± 2.6		12 ± 0.23		13 ± 0.39	2.4 ± 0.08
P60	F194K/ C383M	8.2 ± 0.38		2.1 ± 0.14			
P20	F194Q/ C383S	200 ± 5.5	68 ± 2.2	19 ± 1.0		13 ± 1.4	
P83	F194T/ C383E	72 ± 1.1			8.6 ± 0.00		3.8 ± 0.12
I5	F194V	19 ± 0.72				0.51 ± 0.03	0.30 ± 0.00
P7	F194W/ C383S	470 ± 24	230 ± 7.6	32 ± 1.1	85 ± 8.7	12 ± 0.29	0.47 ± 0.01
P77/ I81	F194Y	470 ± 29	69 ± 1.9	15 ± 0.29	37 ± 3.1	0.49 ± 0.01	0.09 ± 0.00

For each variant, specific activity values were compared to WT activity with that substrate (100%): values <25% are coloured dark blue, 25-90% are pale blue, 90-110% are green, 110-1000% are yellow and >1000% (*i.e.* greater than 10-fold increase) are red.

Table 4.12. Specific activities of mutants from Libraries J and Q (containing mutations at position 227)

Variant	Mutations	0.6 M D-GALACTOSE	1.5 M GLYCEROL	1.5 M D-XYLOSE	2.0 M D-MANNOSE	2.0 M D-ARABINOSE	2.0 M D-GLUCOSE
	WT	550 ± 25	64 ± 4.5	3.3 ± 0.14	89 ± 4.6	1.2 ± 0.04	0.15 ± 0.02
J16	F227D	0.99 ± 0.03	0.20 ± 0.01	0.01 ± 0.00			not detectable
Q16	F227G/ C383E	0.33 ± 0.01				0.01 ± 0.00	
Q48	F227L/ C383E	2.9 ± 0.05	2.4 ± 0.08		1.1 ± 0.36	not detectable	
Q61	F227M/ C383E	18 ± 1.1	14 ± 0.62	2.7 ± 0.23			0.26 ± 0.02
Q75	F227N/ C383E	0.98 ± 0.00					not detectable
Q20	F227P/ C383S	0.65 ± 0.01	0.19 ± 0.00				not detectable
Q3	F227Q/ C383K	1.5 ± 0.09	0.22 ± 0.00	0.02 ± 0.00			
Q14	F227W/ C383E	18 ± 1.6	18 ± 1.3	12 ± 0.56		5.9 ± 0.30	0.73 ± 0.02

For each variant, specific activity values were compared to WT activity with that substrate (100%): values <25% are coloured dark blue, 25-90% are pale blue and 110-1000% are yellow.

Table 4.13. Specific activities of mutants from Libraries K and R (containing mutations at position 245)

Variant	Mutations	0.6 M D-GALACTOSE	1.5 M GLYCEROL	1.5 M D-XYLOSE	2.0 M D-MANNOSE	2.0 M D-ARABINOSE	2.0 M D-GLUCOSE
	WT	550 ± 25	64 ± 4.5	3.3 ± 0.14	89 ± 4.6	1.2 ± 0.04	0.15 ± 0.02
R24	N245A/ C383M	170 ± 17	110 ± 3.8	43 ± 0.69		7.0 ± 0.36	1.1 ± 0.04
K64	N245F	600 ± 20	98 ± 5.8	28 ± 2.2			0.28 ± 0.01
K50	N245G	480 ± 33	83 ± 5.6	12 ± 1.7			
K58	N245H	490 ± 33	87 ± 2.4	14 ± 0.45			
R38	N245H/ C383E	130 ± 14	140 ± 3.6	67 ± 13			
R69	N245H/ C383M	110 ± 3.8	84 ± 6.1	33 ± 2.2	49 ± 1.1	5.4 ± 0.19	1.2 ± 0.11
R85	N245H/ C383S	410 ± 7.8	250 ± 16	62 ± 3.1	110 ± 1.9	7.5 ± 0.32	0.88 ± 0.03
R28	N245L/ C383S	520 ± 12	180 ± 12	30 ± 2.2		6.8 ± 0.49	0.59 ± 0.02
K72	N245P	500 ± 32		12 ± 0.41	87 ± 7.8	1.8 ± 0.05	0.24 ± 0.02
K60	N245R	600 ± 37	140 ± 9.9	22 ± 2.6			
R62	N245R/ C383E	100 ± 6.1	97 ± 12	77 ± 6.0	9.7 ± 2.0	12 ± 0.57	3.0 ± 0.08
O62	N245R/ C383S	460 ± 21	270 ± 12	91 ± 1.8		8.9 ± 0.36	1.0 ± 0.03
K21	N245S	510 ± 5.9		3.5 ± 0.22			0.19 ± 0.01
O52	N245W/ C383S	430 ± 10	280 ± 6.9	93 ± 3.7	120 ± 20	8.5 ± 0.32	1.2 ± 0.31

For each variant, specific activity values were compared to WT activity with that substrate (100%): values <25% are coloured dark blue, 25-90% are pale blue, 90-110% are green, 110-1000% are yellow and >1000% (*i.e.* greater than 10-fold increase) are red.

As the data are grouped by library, it is easy to visualise trends. For example, mutations at positions 194 or 227 almost always lead to significant reductions in D-galactose activity while this is not the case for most mutations at position 245. Mutations at position 245 generally increase the activity against most of the alternative substrates screened, particularly D-xylose.

In some cases, enhanced activities seen during the initial screens were not observed with the purified protein. For the eleven examples of this in libraries containing F194 or N245 mutations, this is most likely due to errors in the screening. These could include pipetting errors, contamination between wells of the microtitre plate, or contamination with another microorganism, such as fungus, which can catalyse one of the reactions in the coupled assay. It is also possible that the mutation(s) have resulted in increased expression levels rather than increased activity which would not be detected with the purified proteins. This was observed by the Ollis group when evolving a bacterial phosphotriesterase (McLoughlin et al., 2005).

In the case of libraries containing mutations at position 227 however, only four of the 19 observed enhancements in activity were confirmed with the purified protein (Table 4.12). F227 is a large hydrophobic residue which occurs directly next to the thioether bond-forming residue C228 in the active site of the enzyme. When the residue was selected it was thought that it may play a role in substrate binding through the hydrophobic surface with F194 and F464. This may be the case, given the enhancement in activities seen with two of the mutants (F227M/ C383E and F227W/ C383E, Table 4.12). However, it also seems likely that mutating this residue has, in some cases, had a negative impact on the environment around the thioether bond, which is essential for the enzyme mechanism. Most of the variants taken forward from these screens were selected because they showed the highest change in absorbance. However, it now seems likely that many of these absorbance readings may have been within the error of the plate reader used to measure the absorbance as the relative increases compared to WT GO-N6M1 were smaller than in other screens. Nevertheless, these screens have still proved worthwhile as two variants showing enhancements in activity were confirmed: F227M/ C383E with enhanced D-glucose activity and F227W/ C383E with enhanced activity towards D-xylose, D-arabinose and D-glucose.

In some potential applications of variants generated in this project, the specificity of the enzyme for the new substrate will be an important factor. In order to help the selection of variants to take forward for further characterisation, the shifts in specificity away from D-galactose and towards the new substrate were determined (Table 4.14). While some variants

stand out due to large increases in specific activity, others stand out due to shifts in specificity towards the new substrate. In all cases where the specificity has shifted by over 100-fold, apart from Q14: F227W/ C383E, activity towards D-galactose is reduced by more than 4-fold at the same time as activity against the alternative substrate(s) is increased more than 10-fold.

Table 4.14. Shifts in specificity away from D-galactose and towards the alternative substrate

Variant	Mutations	1.5 M GLYCEROL	1.5 M D-XYLOSE	2.0 M D-MANNOSE	2.0 M D-ARABINOSE	2.0 M D-GLUCOSE
P19	F194D/ C383E		21			64
P88	F194G/ C383S		17		32	127
P76	F194H/ C383E		27		37	1
P74	F194I/ C383S		27		77	108
I39	F194K		9		10	
P60	F194K/ C383M		44			
P20	F194Q/ C383S	3	16		29	
P83	F194T/ C383E					183
I5	F194V				12	56
P7	F194W/ C383S	4	12	1	11	3
Q16	F227G/ C383E				17	
Q61	F227M/ C383E	7	25			50
Q14	F227W/ C383E	8	113		145	139
R24	N245A/ C383M	6	42		18	22
K64	N245F	1	8			2
R38	N245H/ C383E	9	88			
R69	N245H/ C383M	7	51	3	23	39
R85	N245H/ C383S	6	25	2	7	7
R28	N245L/ C383S	3	10		6	4
K60	N245R	2	6			
R62	N245R/ C383E	8	129		53	103
O62	N245R/ C383S	5	34		9	8
O52	N245W/ C383S	6	36	2	9	10
C383E	C383E	8	33		16	21
C383S	C383S	4	7	2	5	5
F	F194W/ W290F	42				

Only variants displaying a shift in specificity of greater than 5-fold towards at least one substrate are shown. Shifts in specificity were determined by multiplying the shift away from D-galactose by the shift towards the alternative substrate, for example if there is no overall shift, the value would be 1.

The nine variants showing the most interesting changes in activity and which were taken forward for further characterisation are shown in Table 4.15. Full analysis of activity increases and specificity shifts is carried out in Chapter 5.

Table 4.15. The nine variants to be taken forwards for further characterisation

Variant	Mutations	Substrate towards which activity is altered	Increased activity?	Shift in specificity?
P88	F194G/ C383S	D-glucose	✓	✓
P74	F194I/ C383S	D-arabinose	✓	✓
P74	F194I/ C383S	D-glucose	✓	✓
P83	F194T/ C383E	D-glucose	✓	✓
Q14	F227W/ C383E	D-arabinose		✓
Q14	F227W/ C383E	D-glucose		✓
Q14	F227W/ C383E	D-xylose		✓
R38	N245H/ C383E	D-xylose	✓	✓
R62	N245R/ C383E	D-arabinose	✓	
R62	N245R/ C383E	D-glucose	✓	✓
R62	N245R/ C383E	D-xylose	✓	✓
O62	N245R/ C383S	D-xylose	✓	
O52	N245W/ C383S	D-xylose	✓	
F	F194W/ W290F	Glycerol	✓	

4.6 Discussion

Deciding which residues to randomise in projects such as this is crucial. Until more is known about the precise roles of each residue, designing libraries will always be a form of educated guess work. Fortunately, each project which aims to alter the substrate specificity of GO, regardless of its success, provides information on the effects of different residues at different positions which can then feed into future projects (Delagrave et al., 2001, Sun et al., 2002). Selecting the residues for randomisation in this project was done with great care, following analysis of all the previous studies on GO. It was reassuring that another group working on a similar project simultaneously selected very similar residues (Rannes et al., 2011). There are, no doubt, other residues within GO which play a role in determining substrate specificity which have not been included here. C383 was identified by epPCR and it is still unclear how this residue, ~12 Å from the active site copper, can have the observed significant effects on catalysis. Perhaps further studies like the one by Delagrave et al. (2001) would identify additional residues which would not be considered important upon simple structural analysis, although no mutants displaying enhanced activity were identified in a directed evolution study

by the Arnold group (Sun et al., 2001). This is perhaps a consequence of the proposed rigidity of the GO structure (Section 1.6.1). As our knowledge of enzyme structures increases, *in silico* tools will also prove invaluable in design of mutant libraries.

Designing the mutant oligonucleotides and the libraries is another important step. Due to the nature of the library generation method selected for this project, it was necessary to include multiple randomisations on one oligonucleotide. In some cases this proved advantageous as generation of **Libraries L** (Y405X/ Q406X) and **M** (P463X/ F464X) was very straightforward. However, in the case of the oligonucleotide QYRN326-333 where four residues were randomised in combination, this meant any library with one of these mutations also had to include the other three, instantly increasing the size of the library and screening effort required.

As some of the substrates used in screening seem to structurally differ quite considerably from D-galactose, it is easy to make the mistake of assuming that a large number of mutations must be required to enhance this activity. However, this can lead to significant disruption of the active site meaning no catalysis of any substrates can occur. There is also the requirement for an increased screening effort in order to stand a chance of finding the mutant(s) with the desired activity. In this project, similar numbers of variants showing enhanced activity were observed in small and large libraries implying that it may have been better to focus only on small libraries initially and use the data gathered to design the larger libraries by more rational approaches.

As noted in Chapter 3, it is easy to spend large amounts of time optimising a process beyond what is necessarily required. If a process is adequately efficient for its purpose it is often wise to cease optimisation experiments. In the case of library generation, it would have avoided extra purification and overlap extension steps if the DNA fragments containing the different mutations could have been combined in a 'one pot' reaction. It is not clear why this did not work but it was more efficient to simply combine the fragments in multiple steps than spend an unknown amount of time exploring why the 'one pot' reaction did not work. Likewise with optimisation of the ligation reaction and transformation into *E. coli*; various additional conditions such as incubation times longer than 2 hours, different incubation temperatures and a wider choice of cells to transform into could have been included. However, the conditions selected from the optimisations carried out were completely adequate for the library generation here, as demonstrated by the diverse libraries generated.

140

Testing a variety of different conditions is often required to optimise ligation of an insert and plasmid as the kinetics of the ligation reaction are very sensitive to the molecular weight of the two species and the reaction conditions. In the conditions sampled here it was found that a low plasmid to insert ratio, a low DNA concentration and an incubation time of 2 hours gave the best results. However, there may have been additional conditions which were not sampled which would have improved the efficiency of the ligation further. Ligation of different inserts into different plasmids may proceed most efficiently under totally different conditions than those used here and further optimisation should be carried out if a higher efficiency is required.

A large number of variants were screened for altered activities using the agar plate-based screen. While it was not possible to confirm many of the hits using the microtitre plate-based screen, a useful set of initial data is now available for future studies. However, without further analysis of the hits from the agar plate-based screening and confirmation of the new activities, it is necessary to be cautious when analysing these datasets. In most cases, the number of hits is so small that any trends seen in the residues represented could easily be down to coincidence. However, two conclusions can be drawn from the agar plate screens with some confidence. A charged residue at position 333 (arginine or lysine) is significantly favoured in variants showing enhanced glycerol and/ or D-xylose activity (**Library E**); and the disulfide bond between C515 and C518 in the WT enzyme is not essential for structural integrity or catalytic activity as removal of this bond has not inactivated the enzyme (**Library N**). The substitution of an Asn residue at position 333 with Arg or Lys not only introduces a charge at this position, but also introduces a larger residue. This may reduce the size of the active site to permit more efficient binding of the smaller glycerol molecule or introduce a charge at an optimal position to promote binding of glycerol or D-xylose. Without confirmation of the enhanced activity this mutation was not explored further here, but this is a potential target for future work.

The identification and confirmation of 27 variants showing at least some enhancement in activity towards one or more of the substrates screened is a valuable achievement. It is, however, impossible to fully characterise all of the variants in order to understand the basis for the altered activity so the most promising subset of variants was selected for more in-depth characterisation. As well as showing an enhancement in activity towards the new substrate, it is often desirable to reduce the activity towards the native substrate at the same time. Specific activities were therefore determined at different substrate concentrations, depending on the substrate used due to the different saturation kinetics of WT GO with each substrate. While

this data is not ideal for identifying shifts in specificity, it still provides an indication of which variants show the greatest shifts. Calculation of specificity shifts using the same substrate concentrations is shown in Section 5.3. Only nine of the variants were taken forwards for further characterisation. However, those not selected may still be helpful in analysis of the new activities. All nine of the variants are double mutants so it will, of course, be important to analyse the effect of each mutation individually. Unselected variants which contain one of the same mutations, but show different activities will help in the analysis. For example, combined with Cys383Ser, substitution of Phe194 with smaller hydrophobic residues Gly or Ile results in reduced D-galactose activity but increased activity with other substrates. Substitution with the larger hydrophobic residue Trp however, leads to similar increases in D-xylose and D-arabinose activities but a much smaller loss in D-galactose activity. Screening of **Library R** (N245X/ C383CEKMST) identified multiple mutants containing His at position 245 combined with different residues at position 383. While N245H/ C383S (R85) showed higher activities against D-galactose and glycerol than the N245H/ C383E (R38) mutant, activities against D-xylose were about the same. With a Met at position 383 the results were different again. Only N245H/ C383E (R38) was selected for further characterisation but initial data from the two similar mutants may help in analysis of this data. The two similar variants N245R/ C383E (R62) and N245R/ C383S (O62) were both selected for further characterisation so the differences in activity will be explored in more detail.

Conclusions

Based on analysis of previous data, 17 residues surrounding the active site of GO were selected for randomisation. Libraries were designed to randomise different combinations of these residues and, following optimisation of the library generation procedure, 16 libraries were produced. The two screens developed in Chapter 3 were used effectively to identify a number of variants showing altered activities and 37 of these variants were purified by Strep-Tactin chromatography. Specific activities were measured against the different substrates and 27 variants were shown to display enhanced activities. Using this data, nine variants were selected for further characterisation.

Chapter 5 : Characterisation of Altered Enzyme Activities

5.1 Introduction

In the previous chapter, 27 variants of GO were identified with enhanced activity towards one or more substrates (Table 4.15). Many of these variants also displayed shifts in specificity towards the alternative substrate and away from D-galactose. Other studies which have introduced novel activities into GO have presented further data, either detailing characterisation of the activities, or demonstrated the use of the variant(s) in a targeted application. For example, Arnold and co-workers characterised the products of D-glucose oxidation by the W290F/ R330K/ Q406T variant using thin-layer chromatography and ^{13}C- and ^{1}H-NMR, and demonstrated that oxidation occurs on the C-6 hydroxyl of the D-glucose molecule (Sun et al., 2002). Turner and co-workers successfully demonstrated labelling of mannosylated glycoproteins on the surface of *P. pastoris* cells using a GO variant with increased activity towards mannose (Rannes et al., 2011).

In this chapter, the nine variants selected during this study were subjected to more extensive analysis and characterisation with the aim to further understand the enhanced activities, and to determine more optimal reaction conditions for the variants for potential future applications. As each of the nine variants contains two mutations, the contribution of each single mutation on activity towards D-galactose and the alternative substrates was also assessed. Determination of the specific activities at pH 7.0 with D-galactose and the alternative substrates all at the same concentration makes it possible to determine the shift in specificity more accurately than in Chapter 4, where activities were measured at different substrate concentrations depending on the substrate used. The simplest way to visualise changes in specificity is by a graphical method which can demonstrate the true extent to which any enzyme substrate specificity switch has been achieved. As WT GO-N6M1 displays such a high level of activity towards D-galactose (k_{cat} = 889 ± 23) and such low levels of activity towards D-arabinose (k_{cat} = 1.87 ± 0.08) and D-glucose (k_{cat} = 0.39 ± 0.01) it is a major challenge to generate a variant displaying higher activity towards D-arabinose or D-glucose than D-galactose. WT GO-N6M1 displays a higher level of activity towards D-xylose (k_{cat} = 6.76 ± 0.17) than the other substrates and so generation of a 'D-xylose oxidase' showing a preference for D-xylose over D-galactose does seem more feasible. Values for K_M are not considered in this project due to the high substrate concentrations used which are likely to be higher than the K_M for all substrates, this is discussed further in Section 6.3.

Determination of kinetic parameters is carried out in many enzyme evolution projects (Delagrave et al., 2001, Deacon et al., 2004) as the different parameters provide specific information on the catalysed reaction. k_{cat} represents the number of turnovers catalysed by the enzyme per unit time. In the case of WT GO with D-galactose, k_{cat} is around 1000 s^{-1}. K_M represents the substrate concentration at which the enzyme rate is half the maximum rate (V_{max}) and is often interpreted as an indication of the affinity of the enzyme for the substrate – a high K_M indicates a lower affinity. However, it is important to realise that K_M and affinity are not necessarily directly correlated as the value of K_M is also dependent on the rate of product release. A standard one-substrate enzyme reaction is shown below where E is enzyme, S is substrate, ES is the enzyme-substrate complex and P is product:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$
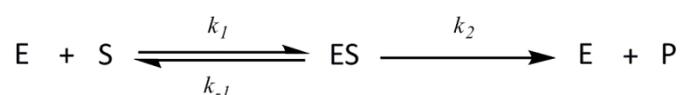
Here K_M is defined as,

$$K_M = (k_{-1} + k_2)/ k_1$$

while the dissociation constant (K_S) of the first step of the reaction (formation/ dissociation of ES) defines the affinity of the enzyme for the substrate and is defined as,

$$K_S = k_{-1}/ k_1$$

Therefore, K_M is only a direct measure of the enzymes affinity for substrate when k_2/ k_1 is very small *i.e.* $k_2 < k_1$. This is not always the case in enzyme catalysed reactions and the relative rate constants may be altered by the mutations introduced so comparison of K_M values as a measure of affinity should be treated with caution. For WT GO with D-galactose, the K_M is very high at around 50-70 mM.

The analysis of activity towards different forms of the substrate can provide valuable information on the binding of substrate within the active site. The activities of the GO variants towards methylated versions of the substrates were investigated and the products of some of the reactions were analysed by ^{1}H-NMR in order to determine the regioselectivity of the variants. pH can have a significant effect on important characteristics of certain substrates and products such as solubility, oligomeric state and stability. The pH optima of the different variants were measured for two reasons:

- to identify shifts in the pH profile and optima which may help explain the altered activity

- to identify the pH range over which the variants are most active, an important consideration for potential future development.

The final characterisation aimed to observe changes to the oxidative half reaction as measurements so far have focussed on changes in the reductive half reaction. As detailed in Section 1.6.3.2, following oxidation of the alcohol group and release of the aldehyde product, the enzyme undergoes an oxidative half reaction whereby dioxygen binds, the Cu^{2+} centre and thioether bond radical is regenerated with hydrogen peroxide released. This regeneration process is essential for continued use of the enzyme over long periods and it is important to consider whether catalytic rate could be improved by increasing the oxygen concentration in the reaction. This is not necessarily desirable in industrial processes as addition of oxygen can greatly increase the cost of the process as well as adding extra safety concerns.

A single variant showing significant change in activity towards glycerol, F194W/ W290F, was discovered towards the end of the project and so complete characterisation was not possible, but further investigation of this variant would be of value. The bulk of this chapter focuses on variants showing enhanced activity towards D-arabinose, D-glucose or D-xylose while Section 5.9 presents the preliminary characterisation of variant F194W/ W290F.

5.2 Alignments with galactose oxidases from other species

While structural alignments with similar proteins can greatly aid the generation of so-called 'smart' libraries (Jochens and Bornscheuer, 2010), they can also prove very useful in rationalising the results of screening more diverse libraries, as created here. The BLAST tool at www.uniprot.org was used to identify GO-like sequences from 50 different species (42 fungal and eight bacterial) that show the highest levels of sequence identity to the WT *F. graminearum* GO sequence (Table 5.1).

Table 5.1 (Next page). Details of the 50 sequences showing highest identity to GO from *F. graminearum*

Fungi are shown in orange, bacteria are shown in yellow. Entries shown in bold were used in the alignment shown in Figure 5.2.

Entry	Protein name	Organism	Identity (%)
I1S2N3	**Galactose oxidase**	**Fusarium graminearum**	**100**
K3VPS0	GAOA	Fusarium pseudograminearum	98
E6PBN5	**Galactose oxidase**	**Fusarium oxysporum**	**79**
E6PBN6	**Galactose oxidase**	**Fusarium verticillioides**	**79**
C7YSK0	**Putative uncharacterized protein**	**Fusarium solani**	**68**
E6PBN7	Galactose oxidase	Fusarium verticillioides	68
E6PBN9	Galactose oxidase	Fusarium graminearum	66
C0KZ93	**Putative galactose oxidase**	**Epichloe festucae**	**61**
E7CHE7	**Galactose oxidase**	**Fusarium subglutinans**	**60**
G3JJG5	**Galactose oxidase**	**Cordyceps militaris**	**60**
J5K2D6	**Galactose oxidase**	**Beauveria bassiana**	**60**
E6PBN4	Galactose oxidase	Fusarium oxysporum	59
E6PBP1	Galactose oxidase	Fusarium graminearum	59
E9E9T3	**Putative galactose oxidase**	**Metarhizium acridum**	**58**
B8MVM2	**Galactose oxidase, putative**	**Aspergillus flavus**	**55**
I8ACL7	**Galactose oxidase**	**Aspergillus oryzae**	**55**
Q0TYM4	Putative uncharacterized protein	Phaeosphaeria nodorum	53
A1DD28	Galactose oxidase, putative	Aspergillus fischerianus	52
E3RTD1	Putative uncharacterized protein	Pyrenophora teres	52
F8N1M2	Galactose oxidase	Neurospora tetrasperma	52
Q1K6D3	Galactose oxidase	Neurospora crassa	52
Q4WH00	Galactose oxidase, putative	Aspergillus fumigatus	52
A7F712	Putative uncharacterized protein	Sclerotinia sclerotiorum	51
B6HHT0	Pc21g18600 protein	Penicillium chrysogenum	51
F9XG69	Putative uncharacterized protein	Mycosphaerella graminicola	51
G2XTJ0	Carbohydrate-Binding Module family 32	Botryotinia fuckeliana	50
F7WCC9	WGS project CABT0000 data, contig 2.115	Sordaria macrospora	49
A1C8B2	F5/8 type C domain protein	Aspergillus clavatus	48
B2W528	Galactose oxidase	Pyrenophora tritici-repentis	48
E5AB94	Galactose oxidase domain protein	Leptosphaeria maculans	48
B6QHP5	Galactose oxidase, putative	Penicillium marneffei	47
E3QHV8	Kelch domain-containing protein	Colletotrichum graminicola	47
G8S9Z0	Putative uncharacterized protein	Actinoplanes sp.	47
H1VE61	Kelch domain-containing protein	Colletotrichum higginsianum	47
Q0CJ74	Galactose oxidase	Aspergillus terreus	47
C9SWV4	Galactose oxidase	Verticillium albo-atrum	46
D5WMI0	Galactose oxidase	Burkholderia sp.	46
G2XFJ0	Galactose oxidase	Verticillium dahliae	46
B2AU87	Predicted CDS Pa_1_18310	Podospora anserina	45
C5CYJ5	Galactose oxidase	Variovorax paradoxus	45
G4NG45	Galactose oxidase	Magnaporthe oryzae	45
J3PC95	Galactose oxidase	Gaeumannomyces graminis	45
A2S3I5	Galactose oxidase-related protein	Burkholderia mallei	44
C0Y524	Galactose oxidase domain protein	Burkholderia pseudomallei	44
B5WUJ3	Galactose oxidase	Burkholderia sp.	43
B8MB95	Galactose oxidase, putative	Talaromyces stipitatus	43
I2N9Q4	Galactose oxidase domain protein	Streptomyces tsukubaensis	43
K1XJ48	Galactose oxidase	Marssonina brunnea	42
Q2T7C2	Lectin repeat domain protein	Burkholderia thailandensis	42
Q4PAQ4	Putative uncharacterized protein	Ustilago maydis	42

An alignment was carried out using the Vector NTI program and the residues occurring at each of the randomised positions was assessed. As shown in Figure 5.1, of the 16 residues randomised in libraries, the majority show relatively high levels of conservation. Position 227, for example contains Phe in all 50 sequences while positions 405 and 514 only differ from the input sequence in one case. Some of the other randomised positions, however, show greater variation, such as position 326 where 11 different residues occur. Indeed, only 11 of the 50 sequences contain Gln at this position as in the WT *F. graminearum* sequence.

Residues showing high levels of conservation are more likely to play an important role within the enzyme, such as substrate binding or catalysis. However, it is important to remember that the engineering carried out here aims to introduce binding to an alternate substrate to D-galactose. Indeed, when initial screening data from many of the libraries (Chapter 4) is compared with the alignment data shown here the results do not always correlate. In the case of **Library E** (Q326X/ Y329X/ R330X/ N333X) for example, the range of residues seen in the first three positions for variants showing enhanced glycerol or D-xylose activity is not limited to those residues shown in the alignments. At position 333, the prevalence of the charged residues Lys and Arg bears no resemblance to the uncharged polar residues Asn or Ser seen in the alignment. Screening of **Library M** (P463X/ F464X) revealed a preference for hydrophobic residues at position 463 which was also displayed in the alignments. However, the high conservation of Phe at position 464 is not seen in screening data. It would be interesting to see the effect of mutations at position 464 on activity against D-galactose. Screening of **Library L** (Y405X/ Q406X) showed that mutation of Y405 had a negative effect on activity against D-galactose. This is in agreement with the alignment data which shows a high level of conservation at this position. Tyr405 is obviously an important residue and will no doubt be the subject of future studies on GO.

Mutation of F194 in **Library I** led to significant reductions in D-galactose activity in most cases which is in agreement with the relatively high conservation of this residue in the different species. The absolute conservation of F227 in the GO sequences from the 50 organisms combined with the deleterious effect on enzyme activity towards all substrates seen upon screening **Libraries J** (F227X) and **Q** (F227X/ C383X) implies an essential role for this residue in the enzyme.
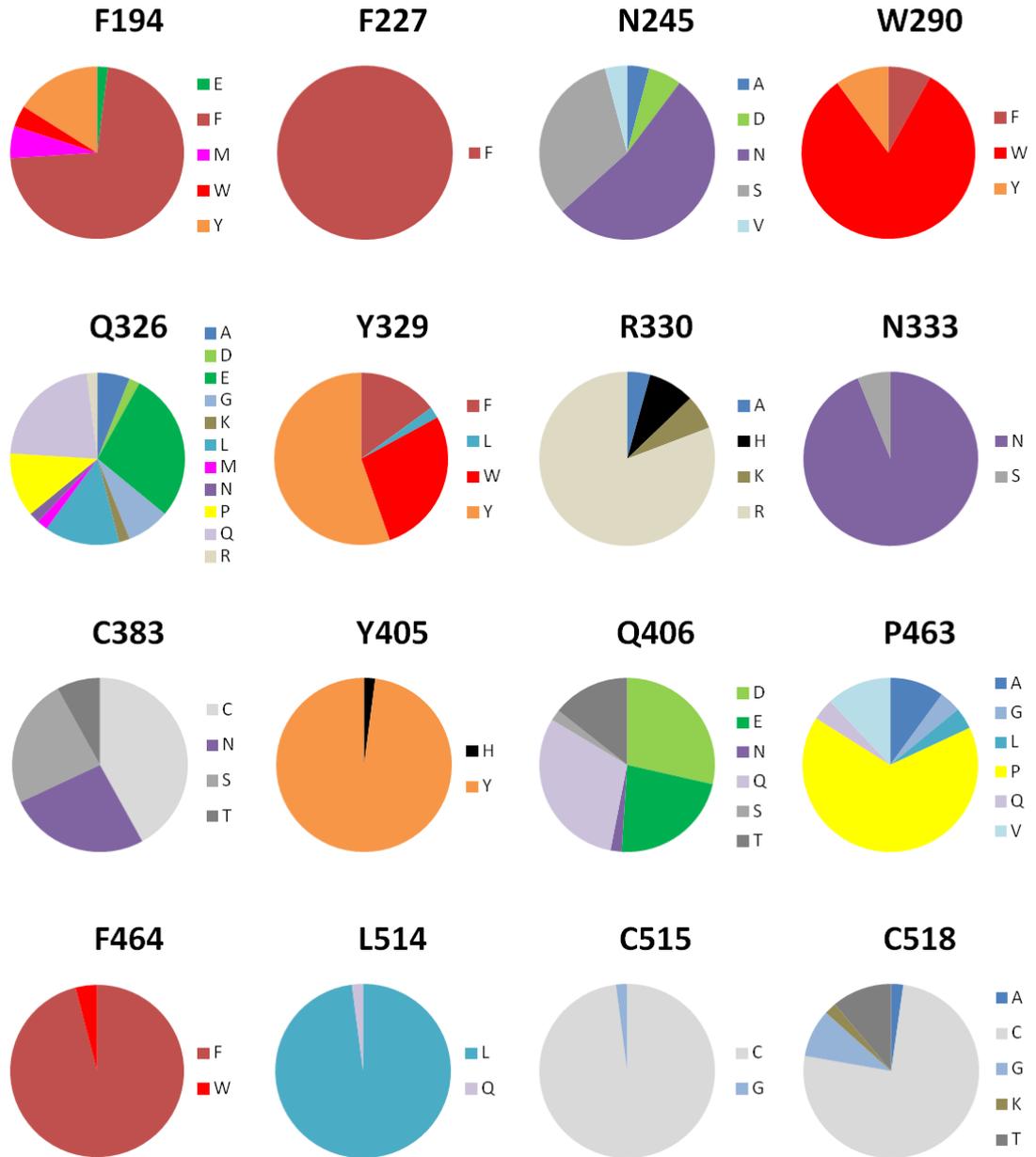
Figure 5.1. Pie charts showing the prevalence of different residues at the positions selected in this study

Figure 5.2 shows an alignment of the sequences showing highest identity to GO from *F. graminearum* from ten different organisms (Table 5.1). All of the sequences share over 55% sequence identity with GO from *F. graminearum* and Domain 1 appears to be slightly less highly conserved than domains 2 and 3. This is unsurprising given the role of these two domains in enzyme activity.

Identity

```
                1         9        19        27        37        47
1. F. graminearum   ASAPIG-SAISRNNWAVTCDSAQ--SGNECNKAIDGNKDTFWHTFYGANG-
                                        Domain 1
```

```
2. F. oxysporum       ASAPLG-TAINRDKWKVTCDSQH--EGDECSKAIDGDRNTFWHTNWAAGGT
3. F. verticillioides ASAPLG-TAINRDKWKVTCDSQH--EGDECSKAIDGDRNTFWHTNWAAGGT
4. F. solani          ATGPLG-IAISRNNWKVTCDSDNPDSDNACQKAIDGDVNSFWHTAWFEDTS
5. E. festucae        AAPPDGN-EIISAGWKVTCDSFE---PGNECVKAIDGNNGSFWHTAINNAN-
6. F. subglutinans    AAPPVGSNPIDRKGWTVKCSSQA--PNFPCGRAIDGDKNTFWQTPYGTTN-
7. C. militaris       AAPPTNGRIIDRTGWGVTCDSFA--AGNECWKAIDGDNTTFWHTAWSGSAP
8. B. bassiana        AAPPTNGRFIDRAGWGVYCDSVE--PGNECWKAIDGDNNTVWHTQWSGTSP
9. M. acridum         AAPPIANGEIDRAGWKARCDSFE---PGNECDKAIDGNNDTFWHTKFEDSN-
10. A. flavus         QSPPYNSARIDRSSWIATCDSEL--VGHECINAIDGDNSTYWHSGDDTNG-
11. A. oryzae         QSPPYNSARIDRSSWIATCDSEL--VGHECINAIDGDNSTYWHSGDDTNG-
```

Identity

```
                       55        65        75        85        95
1. F. graminearum   -DPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSDGTNWG
                                        Domain 1
```

```
2. F. oxysporum       NDPKPPHTITIDMGSSQNVNGLSVLPRQDGSNHGWIGRHNVFLSTNGKNWG
3. F. verticillioides NDPKPPHTITIDMGSSQNVNGLSVLPRQDGSNHGWIGRHNVFLSTNGKNWG
4. F. solani          KDPTLPHTLTVDMKTVQNVNGISALPRQDGTTHGWIASHDIFLSTDGKTWG
5. E. festucae        ----LPHQIVVDLGLTKNVNGISALPRQDGNSHGWITAHEVAVSADNKNW-
6. F. subglutinans    --TPPPHTITIDMKQTQYVSGLQITPRQDGNTRNWIGRHEVYLSSDGTNWG
7. C. militaris       ---PPPHTITLDMGSTYNVNGLSVLPRQDGNQNGWIARHEIAVSTDNTNW-
8. B. bassiana        ---GPPHSLTLDMRTTHNINGISVLPRQDNSKNGWIARHEISVSEGGDNW-
9. M. acridum         ----VPHQIVVDFGTTHNIGISALPRQDGNDHGFIAQHDVAVSTDGSNW-
10. A. flavus         -IASLPHNITINLGTVQNVSGIAVWPR--AVEDGWIGTHDVSLSTDGVTWG
11. A. oryzae         -IASLPHNITINLGTVQNVSGIAVWPR--AVEDGWIGTHDVSLSTDGVTWG
```

Identity

```
                      105       114       124       134       143
1. F. graminearum   SPVASGSWF-ADSTTKYSNFETRPARYVRLVAVTEAN-GQPWTSIAEINVF
                                        Domain 1
```

```
2. F. oxysporum       DAVATGTWF-ADNTEKYSNFETRPARYVRLVAVTEAN-DQAWTSIAEINVF
3. F. verticillioides DAVATGTWF-ADNTEKYSNFETRPARYVRLVAVTEAN-DQAWTSIAEINVF
4. F. solani          SPVATGTWY-ADGTEKFSNFETKRARYVRIVATTEAY-DGPWTSIAEFNVY
5. E. festucae        ETVALGAWYGGDETPKTANFETRSIRYLRLRATGEAG-GNQWTSLAELKVY
6. F. subglutinans    KPVAFGTYW-GDKYPWITNFETHPARYLRFVALSNVNSDYPWIAIADFQVY
7. C. militaris       EVVAFGHWP-KDALPKYANFETRKVRYARVRALTEVA-GNPWTSITDLQVF
8. B. bassiana        EVVAMGNWP-ADALIKYANFETKRARYVRIKAMSETD-GNAWTSIADLQVY
9. M. acridum         ETVAAGTWYGGDKTLKYANFETRTVRYVRIRATSEAN-GGPWTSLAELKAY
10. A. flavus         DPVAHGAWW-PDSTVKLAVFEPKAVQYVRLIARSSFN-GDNATSIADLQIW
11. A. oryzae         DPVAHGAWW-PDSTVKLAVFEPKAVQYVRLIARSSSN-GDNATSIADLQIW
```

Identity

```
                      153       158       168       178       188
1. F. graminearum   QASSYTAPQPG-----LGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRND
                    Domain 1          Domain 2
```

```
2. F. oxysporum       KAASYTSPQPG-----LGRWGPTLDFPIVPVAAAVEPTSGKVLVWSSYKND
3. F. verticillioides KAASYTSPQPG-----LGRWGPTLDFPIVPVAAAVEPTSGKVLVWSSYKND
4. F. solani          KAASYTAPKAG-----IGMWGPTLDFPIVPVAGAVDPETGKVLVWSSYYHD
5. E. festucae        EAKAGPTAYNG-----KGQWGPTINFPTVPVAGTVDPLSGQVLVWSSYTYD
6. F. subglutinans    NALKYNPPAKG-----LGKWGPTIDFPVIPVAGAVEPVSGKVVIWSAYRYD
7. C. militaris       AAGADPAPYAG-----LGKWGPTINFPTVPVAGMVDPLTSKITIWSAYAYN
8. B. bassiana        DANAEPTPYAG-----LGKWGPTINFPTVPVAGMVDPLTGKITIWSAYAYN
9. M. acridum         EAKSGPAPYNR-----VGKWGPTINFPTVPVAATVDPVSGKVIVWSSYTYD
10. A. flavus         SANSIPTAPQGKRLSEVGAWGPTIDFPLVPASAAIEPSSGKVLVWSSYRKN
11. A. oryzae         SANSIPTAPQGKRLSEVGAWGPTIDFPLVPASAAIEPSSGKVLVWSSYRKN
```

Identity

```
                198         206         216         226         236
1. F. graminearum   AFGGSP--GGITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVV
                                        Domain 2

2. F. oxysporum     EFGGSP--GGVTLTSTWDPSTGVISQRTVTNTKHDMFCPGISMDGNGQVVV
3. F. verticillioides EFGGSP--GGVTLTSTWDPSTGVISQRTVTNTKHDMFCPGISMDGNGQVVV
4. F. solani        TMNGSP--GGMTLTSLWVPATGNITQREVFETNHDMFCPGISMDGTGQIVV
5. E. festucae      NFQGSP--NDRVFTALWDPATNIVTPKLVDSTDHDMFCPGISIDGTGQLVV
6. F. subglutinans  AFQGTTPRGGFTLTSIWDPKTNVISNRNVSNNHHDMFCPGISMDGEGQIVV
7. C. militaris     NYYGSS--WDRVFTSIWDPATNDVQPKVVDNTDHDMFCPGISIDGRGQVIV
8. B. bassiana      NYLGSS--WDRVFTSVWDPSTNDVEPKIVDDTDHDMFCPGISIDGKGQVIV
9. M. acridum       NYLGSS--QDRVFTSIWDPATGLVTPKLVDNTDHDMFCPGISIDGTGKMVV
10. A. flavus       QYGGTS--GGLTQTATWDPNTGVVSRREVSDTEHDMFCSGISMDVNGRVIV
11. A. oryzae       QYGGTS--GGLTQTATWDPNTGVVSRREVSDTEHDMFCSGISMDVNGRVIV
```

Identity

```
                246         256         266         276         286
1. F. graminearum   TGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATMSDGRVFTIGGSWSG
                                        Domain 2

2. F. oxysporum     TGGNNAEKTSLYDSSSDSWIPGPDMKVARGYQSSATLSNGRVFTIGGSWSG
3. F. verticillioides TGGNNAEKTSLYDSSSDSWIPGPDMKVARGYQSSATLSNGRVFTIGGSWSG
4. F. solani        TGGNSAARTSTWDPVKNQWVSGPDMRIARGYQASATLSNGKVFTIGGSWAG
5. E. festucae      TGGNSASKTTLYDFRSQTWIPSADMNVARGYQSSATLSDGRVFTIGGSWSG
6. F. subglutinans  TGGNDAKKTTILMPDGN-WVPGPDMQIARGYQSSATCSDGRVFTIGGSWSG
7. C. militaris     TGGNSKSKTTLYDFPSQQWIAGPDMKIPRGYQSSATCSDGRVFTIGGSWSG
8. B. bassiana      TGGNSKLKTTIYDFPSQRWNPGPDMHVPRGYQSSATCSDGRVFTIGGSWCG
9. M. acridum       TGGNSASKTTLYDFASGTWAPGPDMNLPRGYQASATLSDGRVFTIGGCWSG
10. A. flavus       TGGNDDTMTSIYDSFSDSWIAGAPMNVERGYQASTILSDGNMFVLGGSWNG
11. A. oryzae       TGGNDDTMTSIYDSFSDSWIAGAPMNVERGYQASTILSDGNMFVLGGSWNG
```

Identity

```
                296         305         315         325         335
1. F. graminearum   GVFE-KNGEVYSPSSKTWTSLPNAKVNPMLTADKQGLYRSDNHAWLFGWKK
                                        Domain 2

2. F. oxysporum     GISE-KNGEVYDPSSKTWTSLSGALVKPMLTADQQGLYRSDNHGWLFGWKK
3. F. verticillioides GISE-KNGEVYDPSSKTWTSLSGALVKPMLTADQQGLYRSDNHGWLFGWKK
4. F. solani        GEDF-KPGEIFDPSSRKWTLLPNAKVQPMLTADAQGLFRSDNHAWLFGWKG
5. E. festucae      GWFK-KNGEVYDPRTKKWTLLNGADVTPMLTNDARGIYRSDNHGWLFGWKK
6. F. subglutinans  ARGG-KNGEIYDPRAKTWTSLPKCLVGPMLTHDKEGVYKADNHAWLFGWKK
7. C. militaris     GEVEPKDGEIYDPQAKTWTMLPGAKVANLLTQDAQGVYRSDNHAWLFGWKD
8. B. bassiana      QEVQPKDGEIYDFRSNAWTNLPGAKVANLLTQDAQGIYRSDNHAWLFGWTN
9. M. acridum       GWFD-KNGEVYDPKARTWTNLTRALVRPMLTNDTQGIYRADNHGWLFGWRN
10. A. flavus       PQLQNKNSEVYNVTADTWTQLPNAGSQPMLTHDNLGPYHADNHGWIFGWKN
11. A. oryzae       PQLQNKNSEVYNVTADTWTQLPNAGSQPMLTHDNLGPYHADNHGWIFGWKN
```

Identity

```
                345         355         363         373         382
1. F. graminearum   GSVFQAGPSTAMNWYYTSG--SGDVKSAGKRQSNRGVAP-DAMCGNAVMYD
                                        Domain 2

2. F. oxysporum     GSVFQAGPSTAMNWYYTSG--NGNTKSAGKRQSSRGNDP-DAMCGNAVMYD
3. F. verticillioides GSVFQAGPSTAMNWYYTSG--NGNTKSAGKRQSSRGNDP-DAMCGNAVMYD
4. F. solani        ETIFQAGPSSAMNWYYTGG--KGTVKPAGKRQSSRGVDP-DSMCGNAVMFD
5. E. festucae      GSVFQAGPSTAMNWYTTSG--AGGVTPAGKRSSSRGADP-DSMNGNAVMYD
6. F. subglutinans  GSVFQAGPSTAMNWYYTDRGTQGNTKAAGTRRKNGRVDP-DSMNGNVAMFD
7. C. militaris     GSVFQAGPSTAMNWYTTAG--TGAVSAAGKRTTAGRGDDPDSMCGIAVMYD
8. B. bassiana      GTVFQAGPSTAMNWYETNG--NGNVRSAGKRTSN-RGDDLDSMCGIAVMYD
9. M. acridum       GSVFQAGPSVAMNWYTTAG--DGSVTPAGQRRPGG-----DSMNDNAVMFD
10. A. flavus       LSIFHAGPSQAMHWYFAQG--VGNVTNAGNRSTDY-----DQMSGNAVMFD
11. A. oryzae       LSIFHAGPSQAMHWYFAQG--EGNVTNAGNRSTDY-----DQMSGNAVMFD
```

Identity

```
        392        402        412        422        432
1. F. graminearum    AVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNG--LYFAR
                                          Domain 2
```

```
2. F. oxysporum       AVKGKILTFGGSPSYQDSDATTNAHIITIGEPGSTPKTVFASNG--LYYPR
3. F. verticillioides AVKGKILTFGGSPSYQDSDATTNAHIITIGEPGSTPKTVFASNG--LYYPR
4. F. solani          AVKGKIVTFGGTPNYQDSYATTNAHIITIGAPGTQANVVFASDG--MYYPR
5. E. festucae        AAQGKILTVGGSPSYDDSSATAHAHIITIGDVGTQAQVKFASNG--MYSAR
6. F. subglutinans    AVKGKILTFGGATSYQQAPATANAHVLTIDQPGAIAQTALVGNNGAGIHAR
7. C. militaris       ATQGKILTAGGSPSYQNSDAHTNAHIITLAGAGQKPTVKFASTG--MRFPR
8. B. bassiana        ATAGKILTAGGAPSYQNSQAHTNAHIITLGRPGDRPSVRFASNG--MWSAR
9. M. acridum         AARGRILTFGGAPSYQNSPATAHAHLVAVGDPGSPADVRLASNG--MWSAR
10. A. flavus         ATGGRILTFGGSPNYEDSDATKNATLITIGDPNTPPVTVKAGGD--MGYAR
11. A. oryzae         ATGGRILTFGGSPNYEDSDATKNATLITIGDPNTPPVTVKAGGD--MGYAR
```

Identity

```
        440        450        460        470        480        490
1. F. graminearum    TFHTSVVLPDGSTFITGGQRRGIPFBDSTPVFTPBIYVPBQDTFYKQNPNS
                                          Domain 2
```

```
2. F. oxysporum       TFHTSVILPDGNVFITGGQQRGIPFADSTPQLTPBLYVPNDDTFYKQQPNS
3. F. verticillioides TFHTSVILPDGNVFITGGQQRGIPFADSTPQLTPBLYVPNDDTFYKQQPNS
4. F. solani          VFHTSVLLPDGTVFITGGQEYAIPFBDSTPQLTPBLYIPDSDTFVKQQPNS
5. E. festucae        AFHSSVVLPDGTTFITGGQSYAVPFSDENAQLTPBLYDPAADAFTQQQPNS
6. F. subglutinans    VFATSTILPDGNVFITGGQSYSNPFTDTNAQLEPBMFISSSNTFTKQQPNT
7. C. militaris       AFGTATVLPDGQTFITGGQAYAIPFBDSTSQLTPBMYDPARDTFTAMAPNS
8. B. bassiana        SFATATLLPNGQTFITGGQSYAIPFBDSTAQLTPBLYDPBQDSFRQQAPNA
9. M. acridum         SFHTSVVLPDGTVFITGGQSWAVPFSDDAAQLTPBLYDPAADSFRQQQPNS
10. A. flavus         TFHTSVVLPDGSVFITGGQAHGLPFNEDTAQLTPBRYIPBEDRFVEHFPNN
11. A. oryzae         TFHTSVVLPDGSVFITGGQAHGLPFNEDTAQLTPBRYIPBEDRFVEHFPNN
```

Identity

```
        500        510        519        529        539
1. F. graminearum    IVRVYHSISLLLPDGRVFNGGGGLCG-DCTTNHFDAQIFTPNYLYNSNGNL
                                          Domain 2                   Doma...
```

```
2. F. oxysporum       IVRVYHSVSLLLPDGRVFNGGGGLCG-GCTTNHFDAQIYTPNNLYDSNGKL
3. F. verticillioides IVRVYHSVSLLLPDGRVFNGGGGLCG-GCTTNHFDAQIYTPNNLYDSNGKL
4. F. solani          IVRTYHSMSILLPDARVFNGGGGLCG-DCSTNHFDAQIFTPSYLLTKDGKP
5. E. festucae        IVRVYHSIALLMHDGRVFSAGGGLCGGGCKVNHFDGQIFTPQYLLTSSGQP
6. F. subglutinans    IPRTYHSMSLLLPDGTVFNGGGGLCG-SCKSNHFDAQIFTPQYLLDGNGNL
7. C. militaris       IPRNYHSISLLMPDARVFNAGGGLCG-DCATNHFDGQIYTPSYLLKADGSP
8. B. bassiana        IPRTYHSISLLMPDARVFNAGGGLCG-DCNTNHFDGQVFTPSYLLNRDGSP
9. M. acridum         IIRVYHSVALLLPDARVLSAGGGLCG-DCNTNHFDGQVFTPQYLLTADGAP
10. A. flavus         IVRVYHSWSLLLPDATVINGGGGLCA-NCTANHYDAQIYTPPYLFDADGNR
11. A. oryzae         IVRVYHSWSLLLPDATVINGGGGLCA-NCTANHYDAQIYTPPYLFDADGNR
```

Identity

```
        549        559        569        579        589
1. F. graminearum    ATRPKITRTSTQSVKVGGRITISTDSSITKASLIRYGTATHTVNTDQRRIP
                                          Domain 3
```

```
2. F. oxysporum       ATRPKITKVSAKSVKVGGKITISTDSSIKQASLIRYGTSTHTVNTDQRRIP
3. F. verticillioides ATRPKITKVSAKSVKVGGKITISTDSSIKQASLIRYGTSTHTVNTDQRRIP
4. F. solani          AARPKIVSVSATTIKVGGSITVTTGGAINTASLIRYGTATHTVNTDQRRIP
5. E. festucae        ATRPVIQSVTQS----DRSITIATDSAVESASLVRFGTATHAVDTDQRRIP
6. F. subglutinans    ATRPKITAVSATTAKVGSTITVTANSAIKSASIMRYGTATHVVNTDQRRIP
7. C. militaris       AARPAIASATFS----YGRLITGTDGAVASAALMRVGTSTHTINTDQRRVP
8. B. bassiana        AVRPAITSADVN---AGRITIGTDGAVSSASTIRVGTSTHTVNTDQRRIP
9. M. acridum         AARPAIRSAALS----GRTITIATDSPVASASLVRFGTATHTVNTDQRRVP
10. A. flavus         APRPHIETVAPASLRYGGQITITADSPISNASLIRYGTTHTVNTDQRRIE
11. A. oryzae         APRPHIETVAPASLRYGGQITITADSPISNASLIRYGTTHTVNTDQRRIE
```

Figure 5.2. Sequence alignment of the sequences from ten different organisms showing highest identity to GO from *F. graminearum*

Residues identical to those in GO from *F. graminearum* are coloured while other residues are shown in grey. The relative identity for each residue is shown as a bar chart above the alignment: 100% conserved (green), >50% conserved (lime), <50% conserved (red). Figure created using Geneious 7.0.6 software.

5.3 Contributions of single mutations to the altered activities

5.3.1 Generation of single mutants

As all of the variants selected at the end of Chapter 4 are double mutants, analysing the activity of each mutation individually is likely to provide information on the contribution of each mutation towards the altered activity. The single mutants F194G, N245H and N245R had already been isolated from library screening, but the other single mutants needed to be generated. Each double mutant combined a mutation at position 194, 227 or 245 with a mutation at position 383 so the single mutants at positions 194, 227 and 245 were easily created by site-directed mutagenesis using primers which introduced a Cys at position 383, as in the WT enzyme. The primers are shown in Table 5.2 and Phusion DNA polymerase was used for the primer extension (Section 2.4.3). Initial attempts to introduce the tgc codon were unsuccessful; however, following optimisation of annealing temperature, number of cycles and buffer, success was achieved using 16 cycles and the 'GC buffer' provided with the Phusion DNA polymerase. The C383E and C383S mutants had been generated previously (Deacon and McPherson, 2011).

153

Table 5.2. Primers used to reverse the C383 mutation to WT by site directed mutagenesis

Primer name	Sequence
C383Xreversal_fwd	5' ctgatgccatg**tgc**ggaaacgctgtc 3'
C383Xreversal_rev	5' gacagcgtttcc**gca**catggcatcag 3'

The mutated codon is shown in red and bold.

All single and double mutant proteins were expressed, purified and had their specific activities determined with the relevant substrates (D-galactose, D-xylose, D-arabinose or D-glucose) at 1 M substrate concentrations (Sections 2.5 and 2.6.4).

5.3.2 Analysis of activity towards D-galactose

As shown in Figure 5.3, at pH 7.0 all of the double mutants showed reduced D-galactose activity compared to WT GO-N6M1 or the relevant C383 single mutant; however the magnitude of this reduction varied.

C383S displayed similar activity to WT, while C383E activity was reduced about two-fold. However, the four double mutants containing C383E displayed a greater than ten-fold reduction in D-galactose activity compared to WT, with F227W/ C383E showing >40-fold reduction.

N245R and N245W combined with C383S showed much smaller reductions in D-galactose activity compared to WT of two- and three-fold, respectively. By contrast F194G and F194I with C383S showed activity reduction of 27- and nine-fold, respectively.
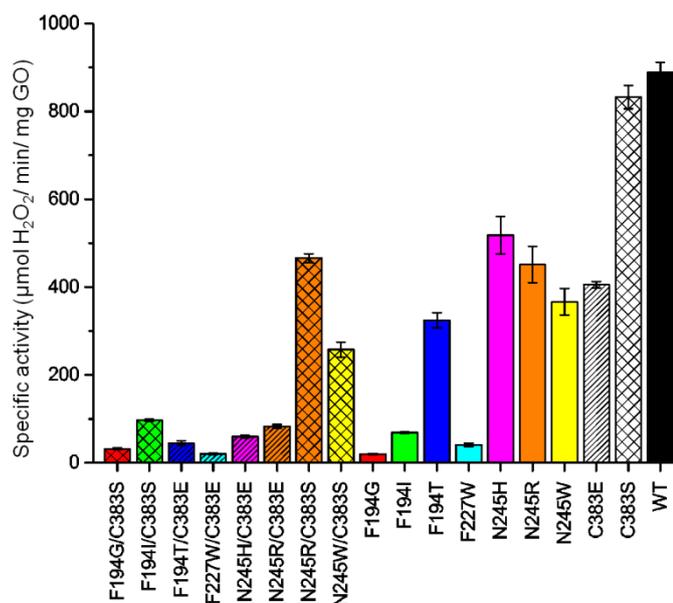
Figure 5.3. Specific activities at pH 7.0 of double and single mutants with 1 M D-galactose

The F194X, F227X or N245X single mutant is the same colour as the relevant double mutant. Double mutants containing the C383E mutant are shown ▨ while those containing C383S are shown ⊠.

For F194X, F227X and N245X double mutants compared with single mutants, the results can be divided into two groups. When combined with C383S, neither F194X nor N245X single mutant activity was significantly different from the double mutant. However, with C383E, the F194X, F227X and N245X single mutant activity was significantly greater than for the double mutants. F227W displayed two-fold greater activity than F227W/ C383E while the other three single mutants displayed five- to nine-fold greater activity than their respective double mutant. The N245X single mutants all showed a two-fold reduction in D-galactose activity compared to WT while F227W showed a reduction of >20-fold compared to WT. Mutation of F194 to the polar residue Thr resulted in a three-fold reduction in activity while the non-polar residues Gly or Ile resulted in much greater reduction in activity of 45- and 13-fold, respectively.

In summary, the F194X, F227X and N245X single mutations all led to a reduction in D-galactose activity compared to WT. Addition of C383E led to a further reduction in activity while C383S had a more limited effect on D-galactose activity.

5.3.3 Analysis of activity towards D-glucose

Figure 5.4 shows the specific activities, at pH 7.0, of the double mutants which exhibited enhanced activity towards D-glucose, and the corresponding single mutants. F194T/ C383E showed a four-fold enhancement while F194G/ C383S and F194I/ C383S showed smaller enhancements of two- to three-fold. N245R/ C383E also showed a four-fold enhancement and F227W/ C383E was double that seen with WT GO-N6M1. These enhancements in activity are smaller than those measured in Chapter 4 as the D-glucose concentration was different at 2 M (Chapter 4) and 1 M (Chapter 5). The specific activities for all of the mutants are consequently reduced by approximately 50%, whereas the specific activity of the WT enzyme doubles, resulting in a less significant enhancement at 1 M D-glucose.

Figure 5.4. Specific activities at pH 7.0 of double and single mutants with 1 M D-glucose
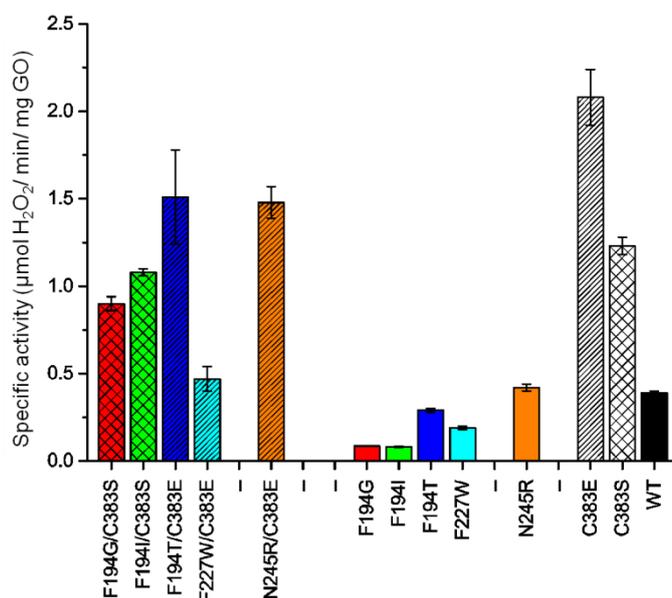
> The F194X, F227X or N245X single mutant is the same colour as the relevant double mutant. Double mutants containing the C383E mutant are shown ▨ while those containing C383S are shown ⊠.

The most significant increase in D-glucose activity compared to WT was seen with the C383E single mutant which displayed a five-fold increase. However as this mutant retained ~50% of the D-galactose activity (Figure 5.3), the shift in specificity towards D-glucose is much smaller than for some of the double mutants. C383S activity towards D-glucose was enhanced three-fold – within a similar range to the four double mutants, excluding F227W/ C383E, although C383S displayed no reduction in D-galactose activity.

All of the F194X, F227X and N245X single mutants showed significantly lower activity towards D-glucose than their respective double mutants and, with the exception of N245R, activity was also reduced compared to WT.

In summary, C383E and C383S mutations led to significant increases in D-glucose activity but, as outlined in Section 5.3.2, only limited or no loss in D-galactose activity. Combining the C383 mutation with an additional mutation at position 194, 227 or 245 resulted in a moderate decrease in activity towards D-glucose but a significant reduction in activity towards D-galactose, compared to the C383 mutation alone.

As shown in Figure 5.5, despite significant reductions in D-galactose activity and increases in D-glucose activity it is not possible to visualise any activity towards D-glucose on the same scale as D-galactose activity. All variants still show significantly higher activity towards D-galactose than D-glucose.

Figure 5.5. Graphical representation to visualise any shift in specificity towards D-glucose

> Specific activities at pH 7.0 towards 1 M D-galactose are shown in red on the left of the axis. Specific activities at pH 7.0 towards 1 M D-glucose are shown in yellow on the right of the axis – as the activities are so small relative those of D-galactose they are not possible to visualise on the same axis.

157

5.3.4 Analysis of activity towards D-xylose

Specific activities against D-xylose are shown in Figure 5.6. As seen with D-glucose, the double mutant F227W/ C383E only showed a moderate increase in D-xylose activity compared to WT. However, as the reduction in D-galactose activity was so great, a significant shift in specificity was observed and the variant is almost as active towards D-xylose as D-galactose (Figure 5.7). The double mutants N245H/ C383E, N245R/ C383E and N245W/ C383S showed increases in D-xylose activity compared to WT of five- to eight-fold; while N245R/ C383S displayed a much greater increase of 20-fold.
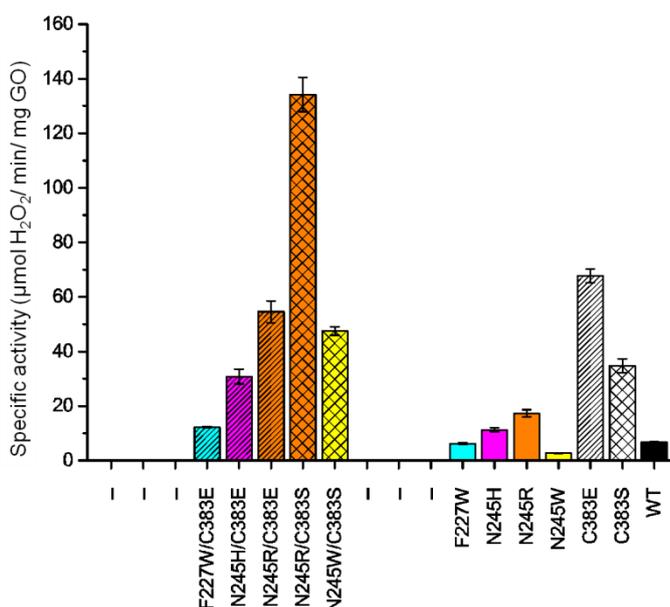
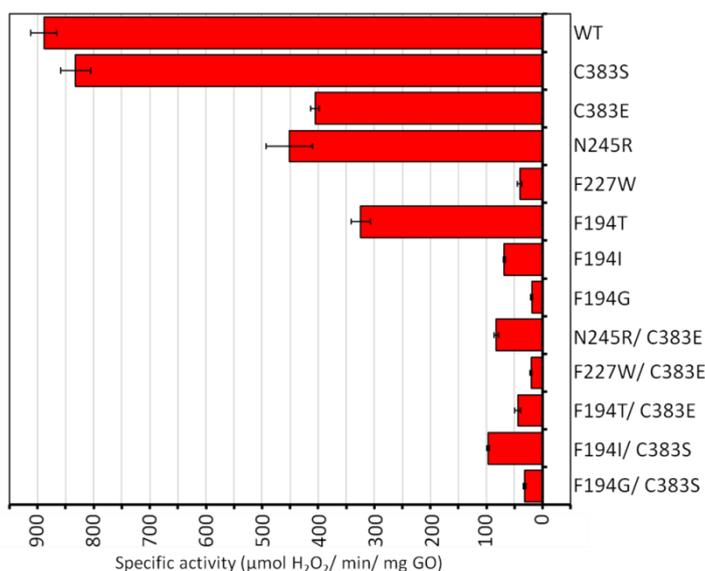**Figure 5.6. Specific activities at pH 7.0 of double and single mutants with 1 M D-xylose**

> The F227X or N245X single mutant is the same colour as the relevant double mutant. Double mutants containing the C383E mutant are shown ⬜ while those containing C383S are shown ⊠.

The variant showing the greatest potential is N245R/ C383E which couples a ten-fold reduction in D-galactose activity with an eight-fold increase in D-xylose activity.

As seen with D-glucose, the C383E mutation led to a significant (10-fold) increase in activity against D-xylose. When C383E was combined with F227W, N245H or N245R, the activity towards D-xylose was reduced compared to C383E alone but the specificity for D-xylose was increased compared to C383E or WT (Figure 5.7). In fact with activity towards 1 M D-xylose of 54.5 µmol H_2O_2/ min/ mg and towards 1 M D-galactose of 82.9 µmol H_2O_2/ min/ mg, the

double mutant N245R/ C383E almost represents a non-selective enzyme in terms of activity with these two substrates. The three single mutants (F227W, N245H or N245R) showed increased or equal D-xylose activity to WT but lower activity than the double mutants containing the C383E mutation.

The C383S mutation increased D-xylose activity five-fold compared to WT, however when combined with the N245W mutation, which alone reduced D-xylose activity by 2.5-fold, actually displayed seven-fold higher activity than WT. A similar effect was seen with N245R, which displayed 2.6-fold greater D-xylose activity than WT, but when combined with C383S, enhanced activity by 20-fold.

In summary, as seen with D-glucose, C383E and C383S mutations led to significant increases in D-xylose activity but, as outlined in Section 5.3.2, only limited or no loss in D-galactose activity. Combining C383E with F227W, N245H or N245R resulted in decreased D-xylose activity, but a greater reduction in D-galactose activity, compared to the C383E single mutant. Combining C383S with N245R or N245W resulted in larger increases in D-xylose activity but less of an effect on D-galactose activity.
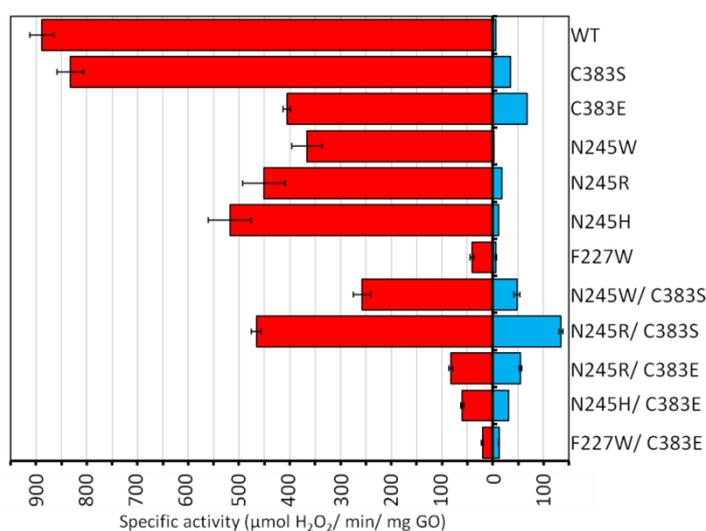

Figure 5.7. Graphical representation to visualise any shift in specificity towards D-xylose

Specific activities at pH 7.0 towards 1 M D-galactose are shown in red on the left of the axis. Specific activities at pH 7.0 towards 1 M D-xylose are shown in blue on the right of the axis.

5.3.5 Analysis of activity towards D-arabinose

Figure 5.8 shows the specific activity measurements towards D-arabinose compared to WT. F227W/ C383E showed a two-fold increase in D-arabinose activity compared to WT while F194I/ C383S and N245R/ C383E showed a three-fold increase. As seen with the D-glucose data, the enhancement in activity is reduced compared to previous measurements at 2 M D-arabinose (Chapter 4) due to the different effect of changing substrate concentration on WT versus the mutants. As also observed with D-glucose, the C383E mutation led to the greatest enhancement in D-arabinose activity of over six-fold. Combination with an additional mutation (F227W or N245R) reduced this activity but, due to the effect on D-galactose activity, specificity was increased. The F194I single mutant displayed a 2.5-fold reduction in D-arabinose activity compared to WT but, when combined with the C383S mutation, the double mutant retained the D-arabinose activity of the C383S mutant while the D-galactose activity was reduced nine-fold compared to WT (Figure 5.3).

Figure 5.8. Specific activities at pH 7.0 of double and single mutants with 1 M D-arabinose

> The F194X, F227X or N245X single mutant is the same colour as the relevant double mutant. Double mutants containing the C383E mutant are shown ▨ while those containing C383S are shown ⊠.

As shown in Figure 5.9, as with the data for D-glucose, for many of the variants it is not possible to visualise activity towards D-arabinose on the same scale as D-galactose activity. All

variants displaying increased activity towards D-arabinose activity still show greater levels of activity towards D-galactose.
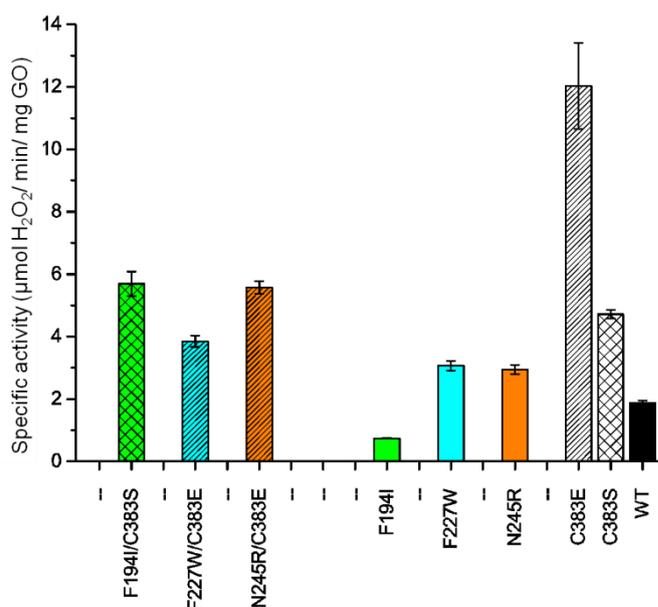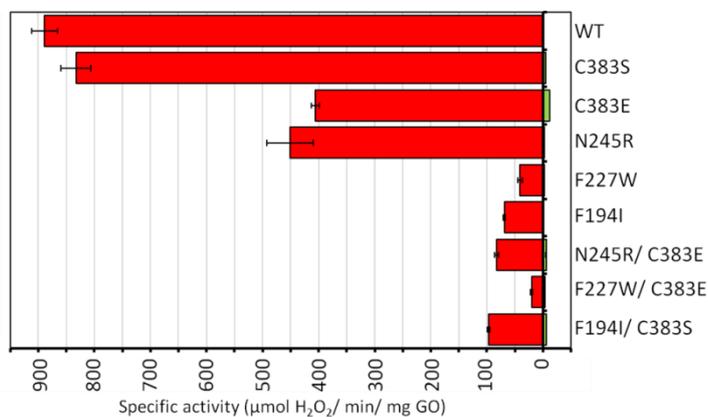

Figure 5.9. Graphical representation to visualise any shift in specificity towards D-arabinose

Specific activities at pH 7.0 towards 1 M D-galactose are shown in red on the left of the axis. Specific activities at pH 7.0 towards 1 M D-arabinose are shown in green on the right of the axis – some variants do not display high enough levels to visualise on the same axis as D-galactose activity.

5.4 Kinetic analysis of variants

5.4.1 Analysis with D-galactose

In order to further characterise the oxidation of D-galactose and the alternative substrates by the different variants, activity was measured over a range of substrate concentrations with the aim of determining the kinetic parameters k_{cat} and K_M. While kinetic parameters with D-galactose were determined in all but one case (Table 5.3), very few of the reactions with the alternative substrates followed a hyperbolic rate equation. It was therefore not possible to determine meaningful kinetic parameters for the reactions with alternative substrates.

Table 5.3. Kinetic parameters for each single and double mutant with D-galactose at pH 7.0

Mutant	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/ K_M (s^{-1} M^{-1})
WT	1000 ± 16	47 ± 6.1	22000 ± 1200
F194G/ C383S	69 ± 4.2	160 ± 25	420 ± 70
F194G	33 ± 0.3	260 ± 5.5	130 ± 3.0
F194I/ C383S	130 ± 4.8	120 ± 12	1100 ± 110
F194I	110 ± 4.0	480 ± 32	230 ± 18
F194T/ C383E	97 ± 13	160 ± 54	590 ± 210
F194T	510 ± 16	40 ± 4.5	13000 ± 1500
F227W/ C383E	ND	ND	ND
F227W	50 ± 4.5	21 ± 8.6	2400 ± 1000
N245H/ C383E	160 ± 4.1	7.9 ± 1.4	21000 ± 3700
N245H	1300 ± 39	35 ± 3.9	38000 ± 4300
N245R/ C383E	140 ± 0.8	5.6 ± 0.3	25000 ± 1300
N245R/ C383S	570 ± 24	18 ± 3.8	32000 ± 6900
N245R	560 ± 8.2	37 ± 2.1	15000 ± 880
N245W/ C383S	460 ± 26	27 ± 6.3	17000 ± 4000
N245W	600 ± 9.6	48 ± 2.6	123000 ± 710
C383E	730 ± 11	10 ± 1.0	71000 ± 7100
C383S	1100 ± 26	15 ± 1.8	72000 ± 8600

Kinetic parameters were determined with D-galactose for each double and single mutant and compared to WT (100%): values <25% are coloured dark blue, 25-90% are pale blue, 90-110% are green and 110-1000% are yellow. It was not possible to generate reproducible data for F227W/ C383E so kinetic parameters could not be determined.

Enzyme efficiency is measured by k_{cat}/ K_M and therefore higher values for k_{cat} and lower values for K_M indicate a more efficient enzyme for the reaction measured. As expected, the value determined for k_{cat} correlates with the previously determined specific activities with D-galactose (Section 5.3.2) in all cases except N245H, which displays a slightly increased value for k_{cat} compared to WT. However, the enzyme efficiency with D-galactose does not always correlate with the specific activity measurements due to the values determined for K_M. Apart from the F194T single mutant, all of the mutants containing a mutation at position 194 show significant increases in K_M for D-galactose. However, all of the other single and double mutants show similar or decreased values for K_M, some of which result in increased enzyme efficiency. If it were possible to determine kinetic parameters for the alternative substrates, the specificity of the enzyme for each substrate could be determined by comparing the values for k_{cat}/ K_M.

5.4.2 Analysis with alternative substrates

The plot of rate versus substrate concentration took on different forms with the different variants and substrates. As seen in the examples shown in Figure 5.10, these included sigmoidal plots (N245R/ C383E with D-arabinose and possibly F194I/ C383S with D-glucose); linear plots (F194G/ C383S with D-glucose); and plots where the rate increased, and then decreased and levelled off (N245R/ C383E with D-xylose). In order to explain the non-hyperbolic nature of the plots, further kinetic characterisation would be required which was beyond the scope of this project.
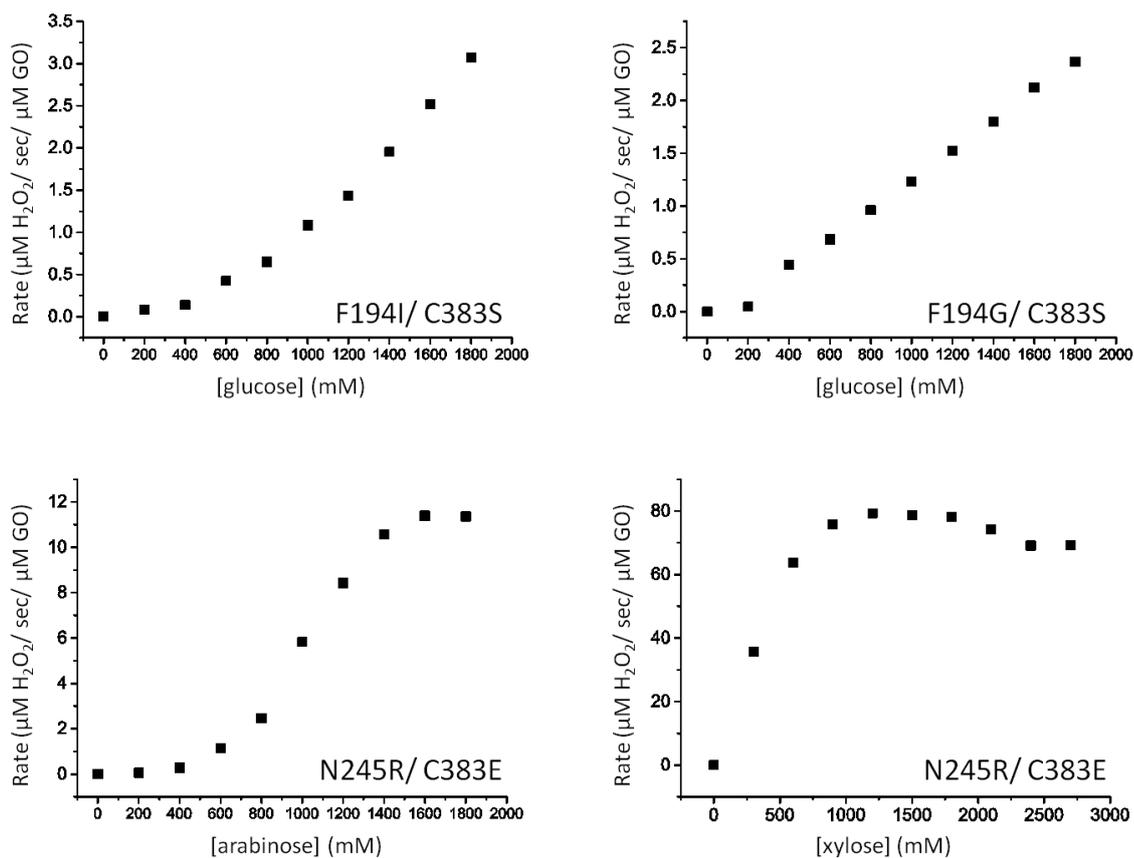
Figure 5.10. Examples of kinetic plots at pH 7.0 which do not follow a hyperbolic rate equation

Kinetic plots took on various different forms for the different substrates and variants. Four examples are shown: F194I/ C383S with D-glucose may be the beginning of a sigmoidal plot, F194G/ C383S with D-glucose is almost linear, N245R/ C383E with D-arabinose is sigmoidal, and N245R/ C383E with D-xylose shows increasing rate which then decreases slightly and levels off as substrate concentration was increased.

5.5 Activities of the variants against methyl glycosides

When dissolved in water, the hemiacetal ring of the monosaccharides D-galactose, D-glucose, D-xylose and D-arabinose opens and reforms to give up to five different isomers: α-pyranose (six-membered ring), β-pyranose, α-furanose (five-membered ring), β-furanose and the acyclic carbonyl form (Isbell and Pigman, 1968, Isbell and Pigman, 1969) (Figure 5.11).

D-galactose   α-D-galactopyranose β-D-galactopyranose   α-D-galactofuranose   β-D-galactofuranose
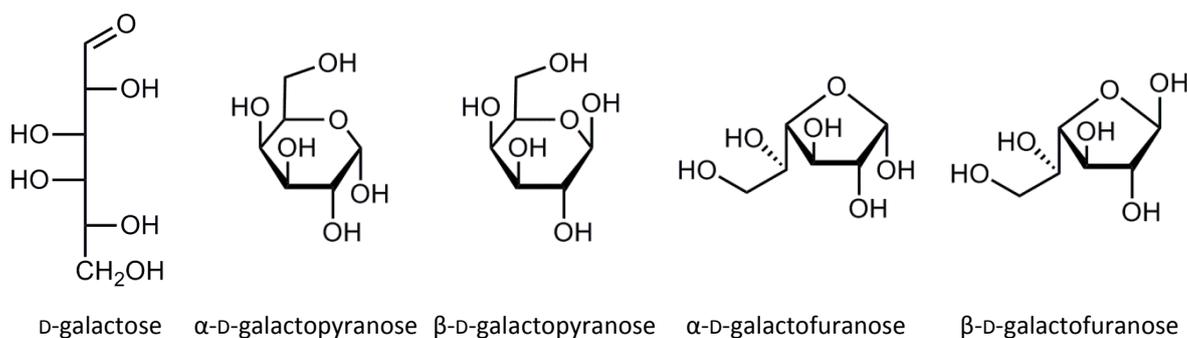
Figure 5.11. The five forms of D-galactose

The mechanism of interconversion of α- and β-D-galactopyranose is shown in Figure 5.12; interconversion of the other sugars occurs by a similar mechanism.
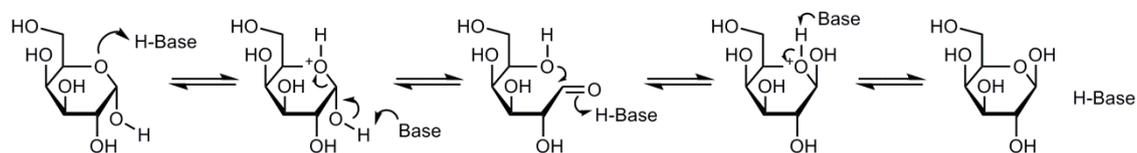
Figure 5.12. The mechanism of interconversion of α- and β-D-galactopyranose
     Adapted from Collins and Ferrier (1995).

In general, the six-membered pyranose rings are more stable than the five-membered furanose rings and the acyclic form and so exist at higher proportions at equilibrium (Table 5.4).

165

Table 5.4. The percentage compositions of sugars in aqueous solution at equilibrium at 31 °C

Sugar	α-pyranose	β-pyranose	α-furanose	β-furanose	Acyclic form
D-arabinose	60	35.5	2.5	2	0.03
D-galactose	30	64	2.5	3.5	0.02
D-glucose	38	62	-	0.14	0.02
D-xylose	36.5	63	< 1	< 1	0.02

Taken from Collins and Ferrier (1995).

In order to prevent isomerisation a different group can be added to C-1 of the hemiacetal ring, such as a methyl group to form the methyl glycoside. In order to determine the relative activities of each mutant against the different isomers and observe the effect (if any) of a methyl group at position C-1, specific activities at pH 7.0 were determined with the methyl glycosides. Unfortunately, the methyl D-galactofuranosides, methyl D-glucofuranosides and methyl D-xylofuranosides were not commercially available and the only methyl D-arabinoglycosides available were methyl α-D-arabinofuranoside and methyl β-D-arabinopyranoside. Specific activities were determined with those glycosides which were commercially available (Figure 5.13) and compared with the activity with the isomeric mixture of the unmethylated form.

Methyl α-D-galactopyranoside   Methyl β-D-galactopyranoside   Methyl α-D-glucopyranoside   Methyl β-D-glucopyranoside

Methyl α-D-xylopyranoside   Methyl β-D-xylopyranoside   Methyl α-D-arabinofuranoside   Methyl β-D-arabinopyranoside
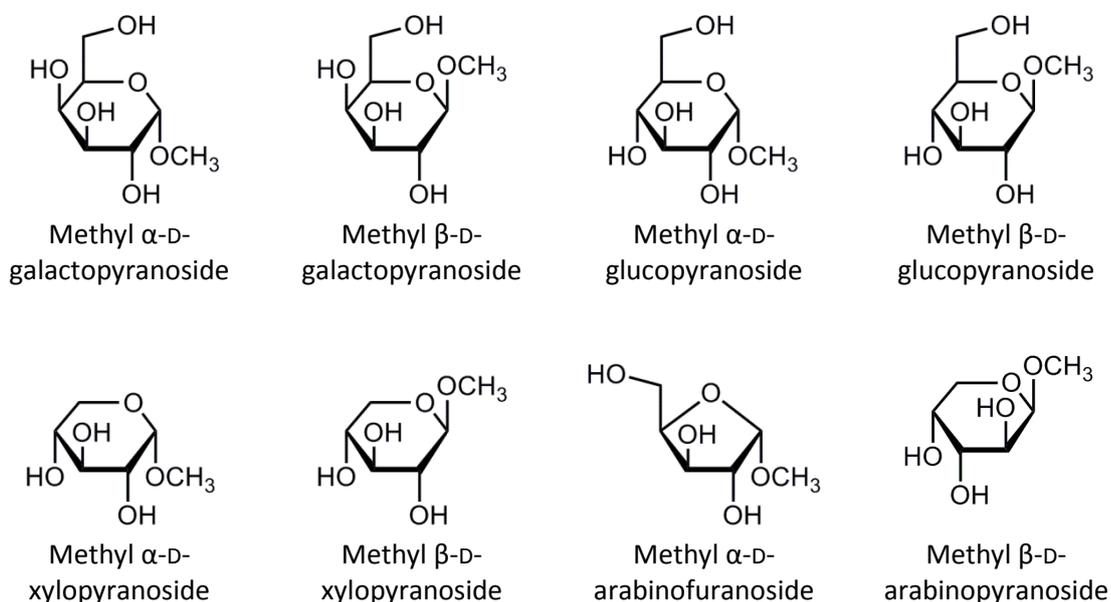
Figure 5.13. The eight commercially available methyl glycosides

5.5.1 ᴅ-Galactopyranosides

When analysing the specific activities of each mutant with ᴅ-galactose and the methyl ᴅ-galactopyranosides (Table 5.5) most show a slight reduction in activity compared to the isomeric mixture. This implies that the presence of the methyl group may have a minor effect on substrate binding. A proportion of the mutants show a slight preference for methyl β-ᴅ-galactopyranoside over methyl α-ᴅ-galactopyranoside but not more than two-fold. There are no obvious trends linking the mutation(s) with the different activities.

The C1-hydroxyl of ᴅ-galactose is predicted to act as a hydrogen bond donor to Tyr329 (Wachter and Branchaud, 1996) so replacement of the hydroxyl with a methoxy group would disrupt this interaction. This may result in a lower binding affinity for methyl ᴅ-galactopyranoside compared to ᴅ-galactose, thus reducing the overall efficiency of catalysis. It is possible that new enzyme-substrate interactions may be made with the methyl ᴅ-galactopyranoside and that these interactions are modestly more favourable with the β-anomer than the α-anomer. Further structural analysis would be required to predict these potential interactions.

Table 5.5. Specific activities at pH 7.0 of all selected mutants and corresponding single mutants against 1 M D-galactose and the methyl D-galactopyranosides

Variant	D-galactose (isomeric mixture)	Methyl α-D-galactopyranoside	Methyl β-D-galactopyranoside	Anomer preference
F194G/ C383S	32 ± 2.2	31 ± 1.0	30 ± 1.5	None
F194G	20 ± 1.2	20 ± 0.5	20 ± 1.2	None
F194I/ C383S	97 ± 2.3	48 ± 2.1	66 ± 4.0	β
F194I	69 ± 1.9	50 ± 1.3	52 ± 2.5	None
F194T/ C383E	45 ± 5.1	21 ± 0.3	19 ± 0.0	None
F194T	320 ± 17	150 ± 12	330 ± 0.0	β
F227W/ C383E	20 ± 1.9	8.5 ± 0.13	12 ± 0.0	β
F227W	41 ± 4.3	21 ± 0.9	25 ± 0.0	β
N245H/ C383E	60 ± 3.5	66 ± 1.4	63 ± 1.6	None
N245H	520 ± 42	250 ± 40	310 ± 5.4	β
N245R/ C383E	83 ± 4.0	51 ± 1.3	61 ± 0.0	β
N245R/ C383S	470 ± 9.8	270 ± 16	350 ± 14	β
N245R	450 ± 41	240 ± 17	280 ± 0.0	β
N245W/ C383S	260 ± 17	270 ± 12	260 ± 0.5	None
N245W	370 ± 31	320 ± 28	370 ± 22	β
C383E	410 ± 7.3	220 ± 15	340 ± 20	β
C383S	830 ± 27	420 ± 5.4	640 ± 58	β
WT	890 ± 23	410 ± 5.2	620 ± 8.8	β

Values are given in units of μmol H_2O_2/ min/ mg GO. For each mutant, specific activity values for the methyl D-galactopyranosides were compared to the value for the isomeric mixture (100%): values 25-90% are pale blue while those within 10% of the isomeric mixture are green.

5.5.2 D-Glucopyranosides

Analysis of the specific activities of each of the D-glucose hits, and respective single mutants, against the isomeric D-glucose mixture and the methyl D-glucopyranosides revealed some interesting trends (Table 5.6). The methyl D-glucopyranoside activities of the F194G and F194I single mutants were only 6-15% of the isomeric D-glucose mixture activities. When combined with the C383S mutation, these activities were reduced even further to 0.6-3.2%. The F194T/ C383E mutant shows only 0.7-1.0% of the isomeric D-glucose mixture activity against the methyl D-glucosides but the F194T single mutant shows essentially no difference in activity. It is possible that the F194T data is incorrect due to contamination or incorrect labelling of the protein sample.

Table 5.6. Specific activities at pH 7.0 of all selected mutants and corresponding single mutants against 1 M D-glucose and the methyl D-glucopyranosides

Variant	D-glucose (isomeric mixture)	Methyl α-D-glucopyranoside	Methyl β-D-glucopyranoside	Anomer preference
F194G/ C383S	0.90 ± 0.04	0.006 ± 0.000	0.005 ± 0.000	None
F194G	0.09 ± 0.00	0.005 ± 0.000	0.006 ± 0.000	None
F194I/ C383S	1.08 ± 0.02	0.035 ± 0.007	0.030 ± 0.005	None
F194I	0.08 ± 0.00	0.009 ± 0.000	0.012 ± 0.001	β
F194T/ C383E	1.5 ± 0.27	0.011 ± 0.003	0.015 ± 0.002	β
F194T	0.29 ± 0.01	0.26 ± 0.003	0.21 ± 0.015	α
F227W/ C383E	0.47 ± 0.07	1.0 ± 0.05	0.037 ± 0.003	α
F227W	0.19 ± 0.02	0.20 ± 0.01	0.10 ± 0.000	α
N245R/ C383E	1.5 ± 0.09	2.0 ± 0.13	0.48 ± 0.024	α
N245R	0.42 ± 0.02	0.81 ± 0.003	0.22 ± 0.000	α
C383E	2.1 ± 0.16	2.9 ± 0.25	1.3 ± 0.03	α
C383S	1.2 ± 0.05	1.0 ± 0.04	1.2 ± 0.13	β
WT	0.39 ± 0.01	0.32 ± 0.01	0.35 ± 0.011	None

Values are given in units of µmol H_2O_2/ min/ mg GO. For each mutant, specific activity values for the methyl D-glucopyranosides were compared to the value for the isomeric mixture (100%): values < 25% are dark blue, 25-90% are pale blue, 90-110% are green and >110% are yellow.

The F227W mutant displays similar activity for methyl α-D-glucopyranoside as the isomeric mixture but activity towards the β-anomer is 2-fold lower. When combined with the C383E mutation, this preference for the α-anomer is greatly enhanced.

In the case of N245R, the C383E mutation increases the moderate preference for the α-anomer by increasing activity towards the α-anomer while reducing activity towards the β-

anomer. This appears to be a cumulative effect as the C383E single mutant also shows a moderate preference for the α-anomer.

Without the structure of an enzyme-substrate complex, it is difficult to explain the different observations with sugars other than D-galactose. However, it is clear from the data presented for D-glucose and the methyl D-glucopyranosides that variants containing mutations at position 194 bind D-glucose in a different way from the WT enzyme or F227 or N245 mutants as addition of a methyl group at position C-1 led to significant loss of activity for F194 mutants. WT GO-N6M1 shows no anomeric preference while F227W, N245R and C383E all showed a modest preference for the α-anomer. This implies that position C-1 of the sugar ring influences substrate binding in some minor manner in the mutants which may lead to a slight improvement in the efficiency of catalysis compared to WT. The major disruption observed upon addition of a methyl group to the sugar ring with the variants containing the F194 mutation suggests that D-glucose may bind in a very different orientation in these mutants, perhaps with the C1 position facing into the active site and making more contacts with the enzyme. It is, of course, possible that the activity observed with the isomeric mixture is actually activity against β-D-glucofuranose. However, as this isomer only exists as 0.14% at equilibrium, this seems very unlikely.

5.5.3 D-Xylopyranosides

The specific activity measurements against D-xylose and the methyl D-xylopyranosides (Table 5.7) revealed the surprising result that WT and mutant GO is active against the methyl D-xylopyranosides as these substrates only contain secondary alcohol groups against which it has previously been reported GO does not show activity (Escalettes and Turner, 2008). If the methyl group indeed has no effect on catalysis, the specific activities recorded for each of the methyl D-xylopyranosides multiplied by the proportions of each anomer in solution (Table 5.4) should generate a value close to the activity recorded for the isomeric mixture (as less than 1% of D-xylose exists in the furanose or acyclic species at equilibrium). This is the case for the WT enzyme, the N245W single mutant and the three mutants containing the N245R mutation, implying that binding of the methyl D-xylopyranosides to the other mutants is affected by the methyl group. Most of the mutants show a preference for one or other of the anomeric forms, although the difference in values is no more than two-fold in all cases. Variants containing the F227W mutation show a slight preference for the α-anomer while those containing mutations at N245 show a slight preference for the β-anomer. This implies that substrate binding is different in the two cases, which is not surprising given the nature of the mutated residues.

Table 5.7. Specific activities at pH 7.0 of all selected mutants and corresponding single mutants against 1 M D-xylose and the methyl D-xylopyranosides

Variant	D-xylose (isomeric mixture)	Methyl α-D-xylopyranoside	Methyl β-D-xylopyranoside	Anomer preference
F227W/ C383E	12 ± 0.23	22 ± 0.62	13 ± 0.32	α
F227W	6.2 ± 0.29	15 ± 0.58	12 ± 0.70	α
N245H/ C383E	31 ± 2.7	33 ± 2.4	43 ± 0.03	β
N245H	11 ± 0.73	11 ± 0.07	14 ± 0.84	β
N245R/ C383E	55 ± 4.0	62 ± 2.0	55 ± 1.3	None
N245R/ C383S	130 ± 6.3	120 ± 5.0	170 ± 7.9	β
N245R	17 ± 1.3	14 ± 0.77	22 ± 1.5	β
N245W/ C383S	48 ± 1.5	50 ± 0.43	110 ± 2.0	β
N245W	2.7 ± 0.12	2.4 ± 0.00	2.6 ± 0.01	None
C383E	68 ± 2.5	93 ± 2.1	100 ± 6.2	None
C383S	35 ± 2.5	38 ± 0.88	57 ± 0.94	β
WT	6.8 ± 0.17	6.2 ± 0.06	6.9 ± 0.00	None

Values are given in units of μmol H_2O_2/ min/ mg GO. For each mutant, specific activity values for the methyl D-xylopyranosides were compared to the value for the isomeric mixture (100%): values 25-90% are pale blue, 90-110% are green and >110% are yellow.

5.5.4 D-Arabinosides

When analysing the specific activity measurements for D-arabinose and the two methyl D-arabinosides (Table 5.8), the data cannot be compared in the same way as for the other sugars as the α-anomer used is the furanoside while the β-anomer used is the pyranoside. WT GO-N6M1 and both of the C383 single mutants show a moderate preference for methyl α-D-arabinofuranoside over methyl β-D-arabinopyranoside. The F194I variant shows 18-fold greater activity towards methyl β-D-arabinopyranoside compared to the α-furanoside. When this is combined with the C383S mutation, the overall activity towards both species increases but the relative preference is unchanged. A similar trend was observed with N245R and its combination with C383E although the difference in activity between the two D-arabinosides is not as great and the preference is for methyl α-D-arabinofuranoside. When F227W and C383E are combined, both of which display a preference for the α-furanoside, the specificity shifts to the β-pyranoside. This is difficult to rationalise without knowing where oxidation is occurring on the two substrates. It is surprising that in a number of cases, the enzyme shows similar levels of activity towards the D-arabinofuranoside and the D-arabinopyranoside. This implies that the enzyme is either oxidising a secondary alcohol group in both cases, or that the mechanisms are different with the two substrates as only the furanoside contains a primary alcohol group.

Table 5.8. Specific activities at pH 7.0 of all selected mutants and corresponding single mutants against 1 M D-arabinose, methyl α-D-arabinofuranoside and methyl β-D-arabinopyranoside

Variant	D-arabinose (isomeric mixture)	Methyl α-D-arabinofuranoside	Methyl β-D-arabinopyranoside	Preference
F194I/ C383S	5.7 ± 0.39	0.45 ± 0.02	7.7 ± 0.10	β-pyr
F194I	0.74 ± 0.01	0.037 ± 0.000	0.68 ± 0.032	β-pyr
F227W/ C383E	3.9 ± 0.18	9.8 ± 0.82	14 ± 1.3	β-pyr
F227W	3.1 ± 0.15	2.0 ± 0.038	1.0 ± 0.076	α-fur
N245R/ C383E	5.6 ± 0.20	37 ± 0.14	7.8 ± 0.043	α-fur
N245R	3.0 ± 0.15	3.4 ± 0.036	0.58 ± 0.063	α-fur
C383E	12 ± 1.4	60 ± 0.44	13 ± 0.89	α-fur
C383S	4.7 ± 0.14	11 ± 0.40	4.1 ± 0.032	α-fur
WT	1.9 ± 0.08	1.5 ± 0.079	0.66 ± 0.016	α-fur

Values are given in units of μmol H_2O_2/ min/ mg GO. For each mutant, specific activity values for the methyl D-arabinosides were compared to the value for the isomeric mixture (100%): values < 25% are dark blue, 25-90% are pale blue, 90-110% are green and >110% are yellow.

5.6   pH dependence of the enhanced activities

5.6.1   Determination of an appropriate three-component buffer system for pH profile measurements

When measuring the effect of pH on enzyme activity, it is important to ensure that the effect seen is solely due to the variation in pH and not another factor such as ionic strength or buffer composition (Ellis and Morrison, 1982). With this in mind, it was decided to determine the pH profile of GO activity using a three-component buffer system designed so that ionic strength is maintained across the pH range. Three-component buffer systems also use the same buffer mix across the pH range so errors will not be introduced by using different buffers for different pH values.

The first three-component buffer trialled was 0.05 M acetic acid, 0.05 M 2-(*N*-morpholino) ethanesulfonic acid (MES) and 0.1 M Tris and covered a pH range of 3.7 to 9.0 (Ellis and Morrison, 1982). The specific activity of WT GO-N6M1 with 600 mM D-galactose was measured in this buffer at pH 7.0 and found to be within 10% of the activity measured in 100 mM sodium phosphate, pH 7.0. However, when specific activities with 600 mM D-galactose were measured over a pH range of 4.0 to 9.0 with WT GO-N6M1, it was not possible to identify the highest pH at which the enzyme is active as pH 9.0 does not appear to inhibit activity (Figure 5.14, red line). For pH values above 9.0, specific activities were determined using 100 mM 3-(cyclohexylamino) ethanesulfonic acid (CHES) (Figure 5.14, blue line), however it is not clear whether the rapid drop off in activity above pH 9.0 is due to the pH of the solution, or inhibition by CHES, or a combination.
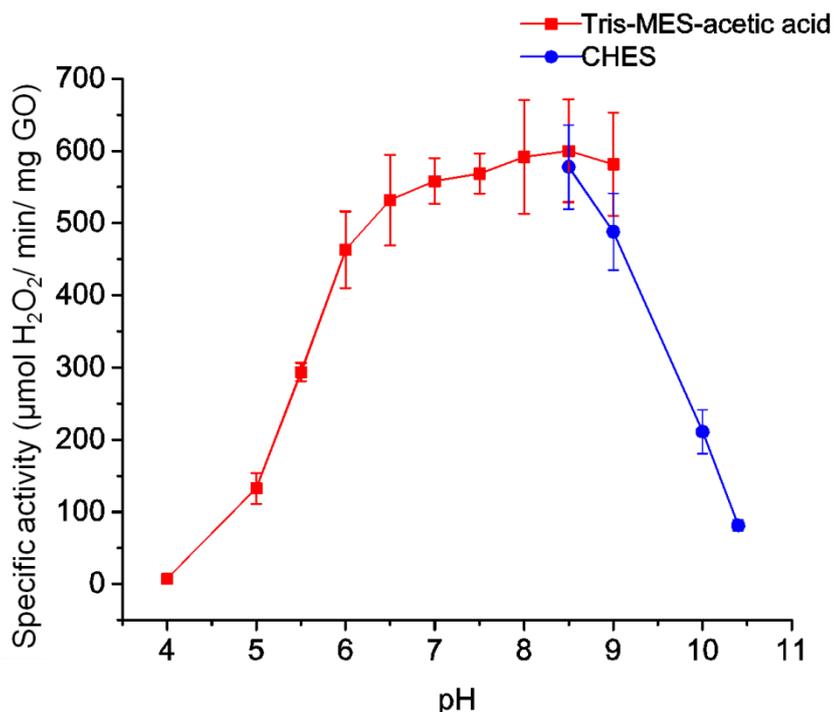
Figure 5.14. pH profile of WT GO-N6M1 with D-galactose determined using Tris-MES-acetic acid and CHES as buffers

A different three-component buffer was therefore trialled: 0.1 M *N*-(2-acetamido)-2-aminoethanesulfonic acid (ACES), 0.052 M Tris and 0.052 M ethanolamine which covers a pH range of 5.6 to 10.4 (Ellis and Morrison, 1982). Again, the specific activity with 600 mM D-galactose of WT GO-N6M1 in this buffer at pH 7.0 was found to be within 10% of the activity measured in 100 mM sodium phosphate, pH 7.0 so activity over the full pH range of the buffer was measured. As can be seen in Figure 5.15, this three-component buffer is much more suitable as activity clearly goes down towards the lowest and highest pH values measured allowing visualisation of the pH profile.
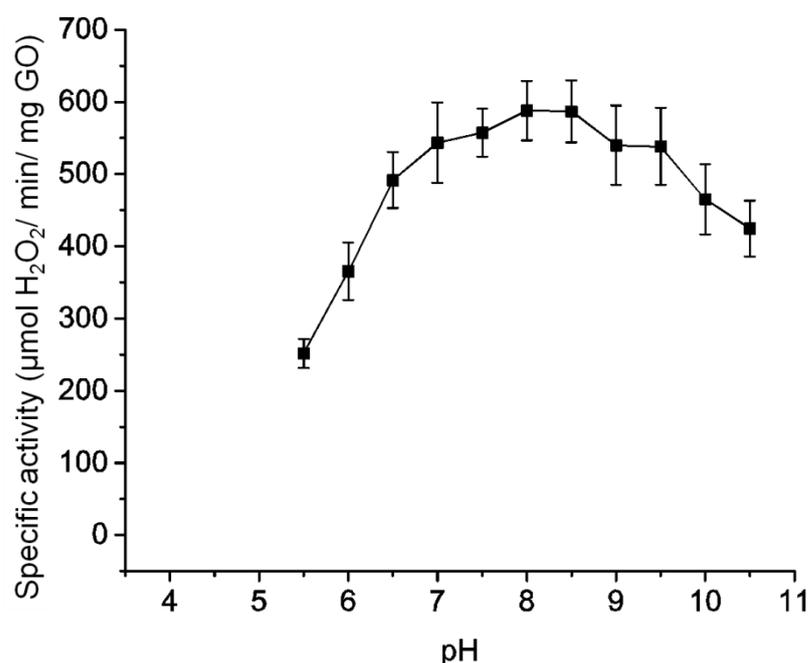
Figure 5.15. pH profile of WT GO with D-galactose determined using ACES-Tris-ethanolamine as a buffer

5.6.2 The effect of high pH on D-xylose and D-arabinose

The ACES-Tris-ethanolamine buffer was then used to make up solutions of 1 M D-glucose, D-xylose or D-arabinose across the pH range in order to observe the effect of pH on oxidation of these substrates. However, with the pentoses D-xylose and D-arabinose, a coloured substance was generated upon dissolving in the buffer which was more pronounced at higher pH values (Figure 5.16). Colour was visible after a few hours and increased over the course of several days.
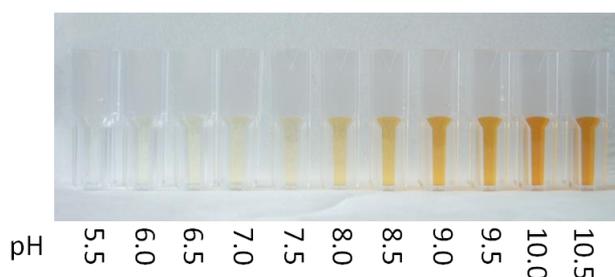
Figure 5.16. D-xylose dissolved in ACES-Tris-ethanolamine buffer at different pH values

Similar colour changes were observed for D-arabinose.

175

Without further analysis such as gas-liquid chromatography or thin-layer chromatography, it is not possible to identify the coloured substance. The instability of sugars in alkaline solutions, and subsequent conversion or degradation to a number of other products, has been previously reported (Forsskahl et al., 1976, Shaffer and Friedemann, 1930).

Despite the obvious presence of something other than the pentose in the D-xylose and D-arabinose solutions, activity was measured with WT GO-N6M1 across the pH range. As shown in Figure 5.17, the measurements showed a marked shift in pH optimum of the enzyme towards lower pH compared to the reaction with D-galactose (Figure 5.15); activity was barely detectable above pH 8.0. However, it is unclear whether this shift in pH optimum is solely due to the pH of the buffer, or if the coloured substance (or another substance also present in the higher pH solutions) inhibits the oxidation reaction. Without the further analysis and perhaps comparing the effect of yet more three-component buffers, the pH profiles with D-xylose and D-arabinose could not be reliably measured.
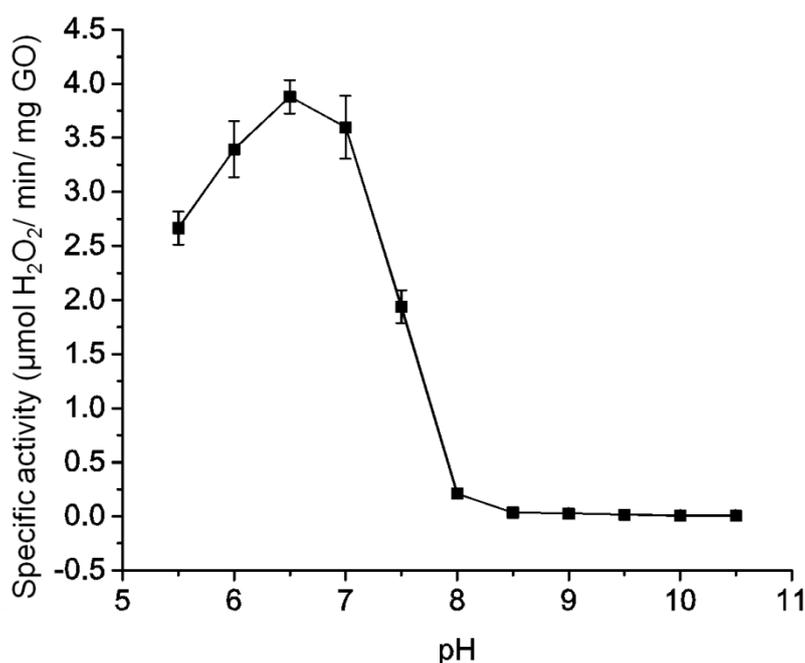
Figure 5.17. pH profile of WT GO with D-xylose determined using ACES-Tris-ethanolamine as a buffer

The corresponding pH profile with D-arabinose showed a similar shift in pH optimum towards acidic pH.

5.6.3 pH profiles of D-glucose activity

As no colour change was observed upon dissolution of 1 M D-glucose in ACES-Tris-ethanolamine across the pH range, and the pH profile with WT GO-N6M1 appeared similar to that with D-galactose, pH profiles were measured for each of the variants showing enhanced D-glucose activity as well as the corresponding single mutants.

Figure 5.18 shows the pH profiles of variants containing mutations at position 194. When the hydrophobic residues Gly or Ile were introduced the pH optimum shifted from pH 7.0-8.0 for WT to pH 6.0-7.0 for the mutants; meanwhile the C383S single mutant displays a pH optimum of pH 9.0-9.5. The double mutant in each case seemed to take on much more of the character of the C383S mutation than the F194 mutation as, compared to WT, activity was reduced at low pH and increased at high pH with a pH optimum of pH 8.5-9.0 in both cases. In the case of the F194T/ C383E double mutant, similar trends were observed: the F194T single mutant has a pH optimum of pH 7.0-7.5 and a narrower pH profile while C383E has a pH optimum of 9.5-10.0 and greatly reduced activity at low pH compared to WT. The double mutant took on much more of the C383E character with a pH optimum of 8.5-9.0 and reduced activity at low pH.

As shown in Figure 5.19, the pH profile for the F227W mutant revealed a very similar pH optimum to WT of pH 7.5, but as activity dropped off drastically as pH was increased or decreased the profile is very different. The F227W/ C383E double mutant also displayed a relatively narrow pH profile but with a pH optimum of pH 9.5, the same as the C383E single mutant. Again the C383X mutation had the greatest effect on the pH profile.

Measurement of the pH profiles for the N245R/ C383E variant and associated single mutants (Figure 5.20) did not reveal such dramatic variation as with variants containing F227W: the N245R single mutant has a pH profile almost identical to WT and the double mutant has a pH profile almost identical to C383E.

Overall, it would seem that mutations at positions 194 or 227 in a WT background had a significant effect on the pH profile of the enzyme while the same mutations in the background of the C383E or C383S mutations had a much less significant effect. This implies that the C383X mutations may induce change(s) in the active site which make it less sensitive to mutation at position 194 or 227. The N245R mutation had a very limited effect on the pH profile regardless of the background into which it was introduced.
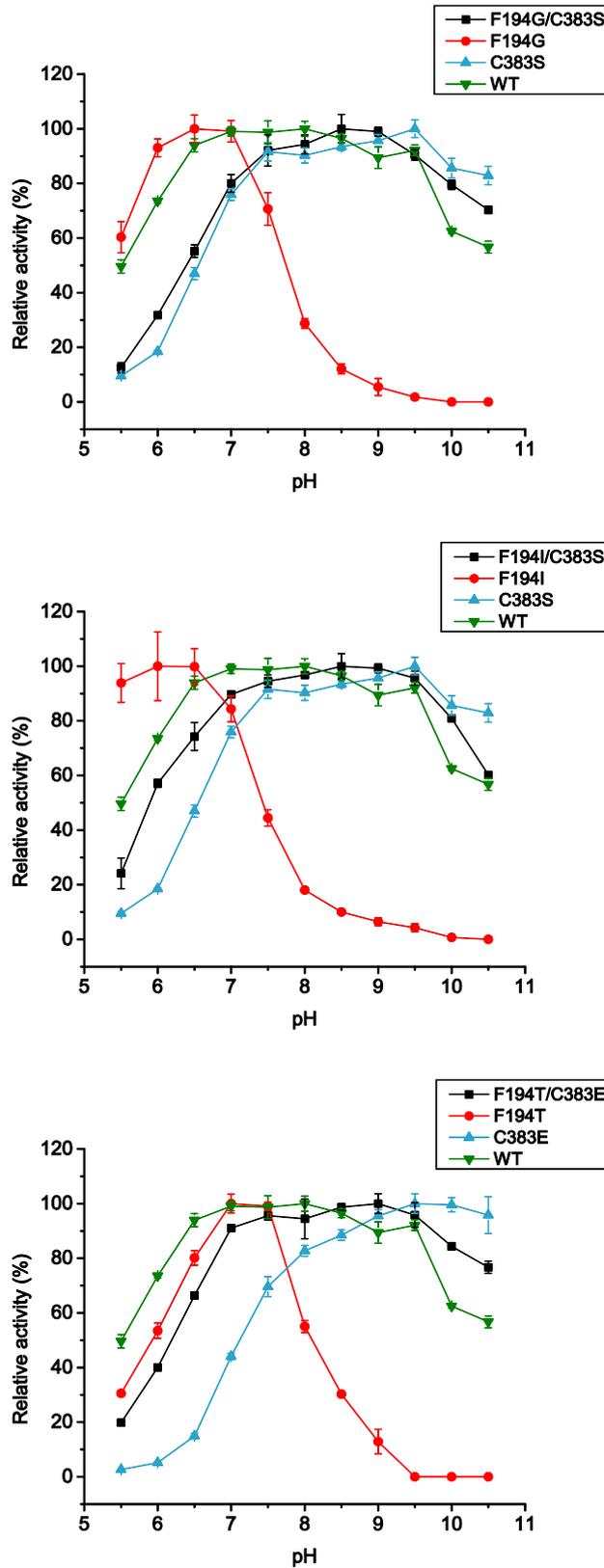
Figure 5.18. pH profiles of double and single mutants from F194X libraries with D-glucose

For each variant, activities are presented as a percentage of the highest activity to permit comparison of the pH profiles regardless of relative activities towards D-glucose.
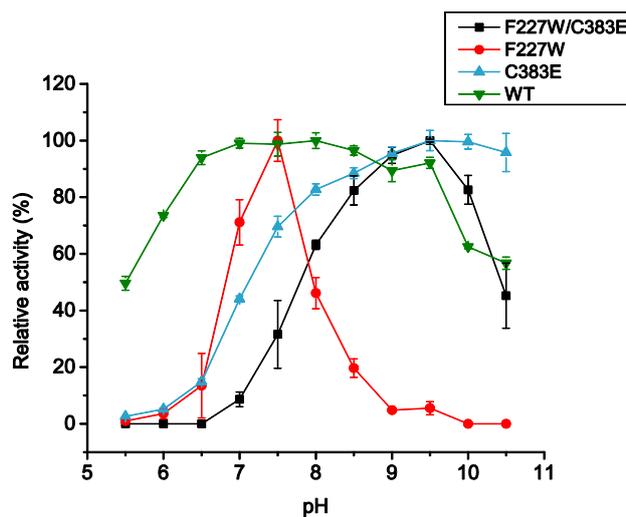
Figure 5.19. pH profiles of F227W/ C383E, WT and single mutants with D-glucose

For each variant, activities are presented as a percentage of the highest activity to permit comparison of the pH profiles regardless of relative activities towards D-glucose.
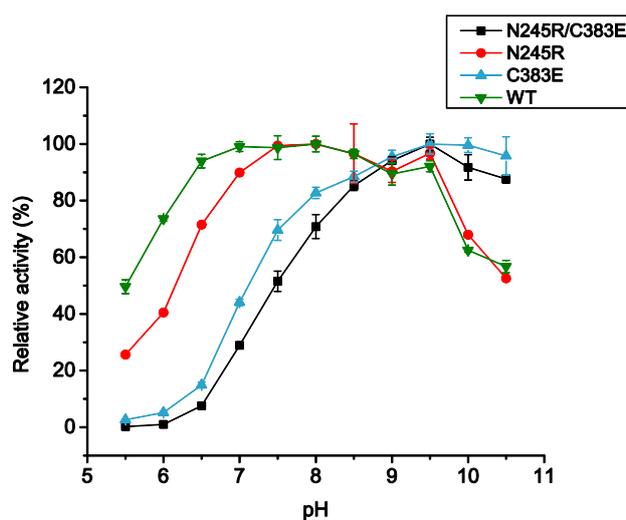
Figure 5.20. pH profiles of N245R/ C383E, WT and single mutants with D-glucose

For each variant, activities are presented as a percentage of the highest activity to permit comparison of the pH profiles regardless of relative activities towards D-glucose.

5.7 The effect of the mutations on the oxidative half reaction

By measuring the effect of varying sugar substrate concentration (Section 5.4.2), only the effect of the mutations on the reductive half reaction was measured. In order to see whether the oxidative half reaction (where oxygen binds, the copper (II) and thioether bond cofactors are regenerated and hydrogen peroxide is released (Section 1.6.3.2)) is affected by the mutations, specific activities at pH 7.0 were measured for some of the variants at three different oxygen concentrations: air saturated, 'increased oxygen' and 'decreased oxygen'. The measurements only represent initial data as molar oxygen concentrations were not determined. While oxygen was bubbled through the 'increased oxygen' samples for over an hour, it was not confirmed that the solution was saturated with oxygen. Likewise, while every effort was made to purge the 'decreased oxygen' samples of oxygen, small quantities of oxygen no doubt entered the assay mix during transfer to the reaction cuvette and upon addition of the enzyme solution, which was only partially degassed (Section 2.6.4.1).

As shown in Figure 5.21, in the case of WT GO-N6M1 with D-arabinose, increased or decreased oxygen levels had essentially no effect on the activity of the enzyme. This implies that the very low level of oxygen present in the 'decreased oxygen' sample was adequate for maximal activity. When the C383E or C383S mutations were introduced however, the effect was quite different. Increasing the oxygen concentration resulted in a small increase in activity, while in the 'decreased oxygen' sample activity was barely detectable. This could imply that the $K_{M(oxygen)}$ for the mutants is higher than for WT as the oxygen concentration in the air saturated samples was not saturating for the enzyme. However, the observed effect could also be explained by the more rapid turnover of the variant which depletes local oxygen concentrations faster than the WT, making it more sensitive to overall oxygen availability. A similar effect was observed for the F227W/ C383E variant while F194I/ C383S displayed no difference in D-arabinose activity between air saturated and 'increased oxygen' samples implying that the $K_{M(oxygen)}$ may lie somewhere in-between the oxygen concentrations of the oxygen depleted and air saturated samples.
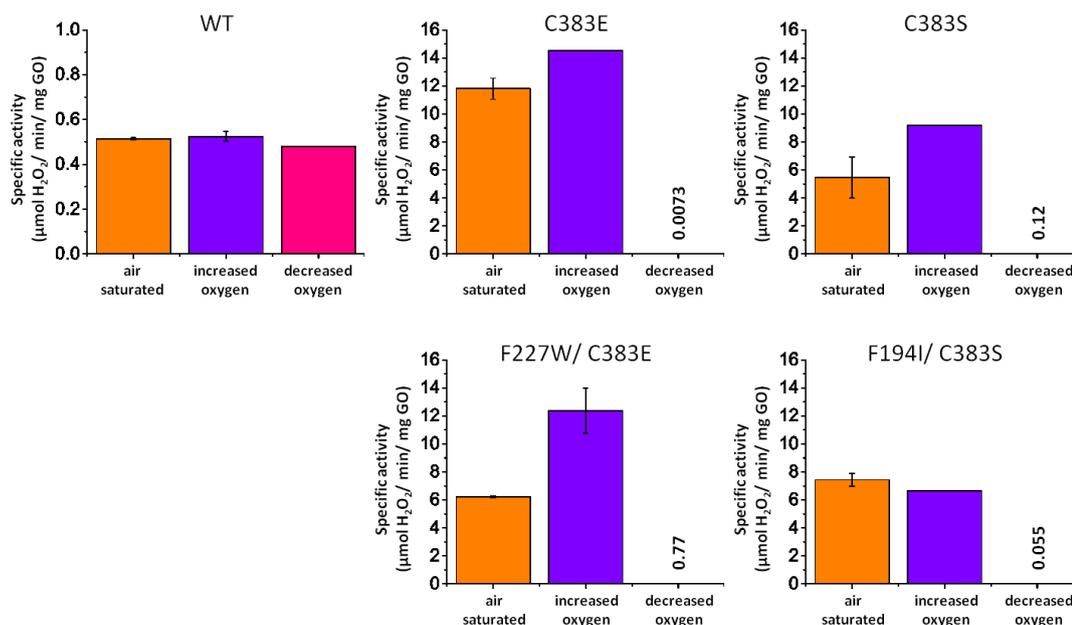
Figure 5.21. Specific activities at pH 7.0 with 1 M D-arabinose at three different oxygen concentrations

Specific activities were measured in three different assay mixes: i) air saturated, ii) increased oxygen concentration and iii) decreased oxygen concentration for a selection of variants, as detailed in Section 2.6.4.1. Note the different y-axis scales.

When oxygen levels were varied in the measurement of activity with 1 M D-glucose (Figure 5.22), more moderate effects were observed. For WT GO-N6M1, increasing the oxygen concentration had no effect on activity towards 1 M D-glucose whereas reducing the oxygen concentration reduced activity compared to the air saturated level. This implies that the $K_{M(oxygen)}$ for activity with D-glucose as the other substrate is higher than with D-arabinose (Figure 5.21). For the C383E, F194G/ C383S and F194I/ C383S variants, increasing oxygen concentration increased activity slightly, although reducing oxygen concentration had a smaller effect than for WT. The C383S single mutation resulted in an enzyme which showed equal activities at all three oxygen concentrations trialled. All four of the variants display changes in the effect of oxygen concentration compared to WT; however these changes are small in all cases. Given the limited accuracy of the measurements due to variation in oxygen concentrations, the observed differences between the variants are unlikely to be significant.
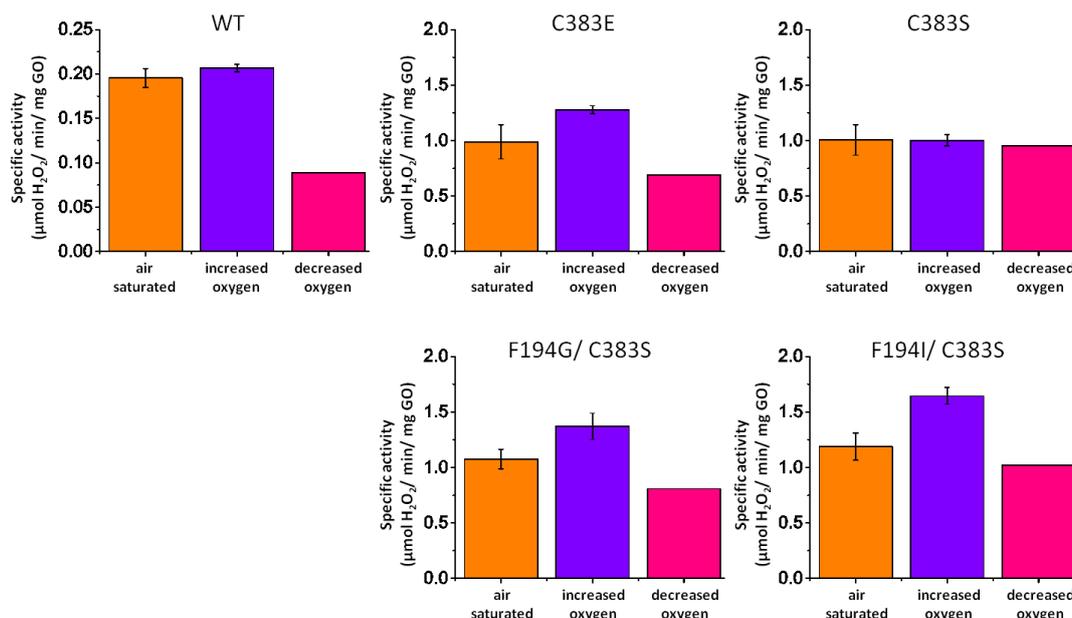
Figure 5.22. Specific activities at pH 7.0 with 1 M D-glucose at three different oxygen concentrations

Specific activities were measured in three different assay mixes: i) air saturated, ii) increased oxygen concentration and iii) decreased oxygen concentration for a selection of variants, as detailed in Section 2.6.4.1. Note the different y-axis scales.

As observed for 1 M D-arabinose, altering the oxygen concentration had no effect on WT GO-N6M1 activity against 1 M D-xylose (Figure 5.23). The C383E mutation led to a reduction in activity when oxygen levels were depleted and this effect was enhanced by addition of the F227W or N245H mutations. The N245H/ C383E variant also showed increased activity when oxygen concentration was increased, implying that the level of oxygen in the air saturated sample was not saturating for the enzyme. C383S also showed a reduction in activity towards 1 M D-xylose when oxygen was reduced, but to a lesser extent than for C383E. When combined with the N245R or N245W mutations however, activity was significantly reduced when oxygen levels were reduced and increased at higher oxygen concentrations. The data imply that mutations at position 245 may increase $K_{M(oxygen)}$, although analysis of the single mutants would have to be carried out to confirm this.

Figure 5.23. Specific activities at pH 7.0 with 1 M D-xylose at three different oxygen concentrations

Specific activities were measured in three different assay mixes: i) air saturated, ii) increased oxygen concentration and iii) decreased oxygen concentration for a selection of variants, as detailed in Section 2.6.4.1. Note the different y-axis scales.
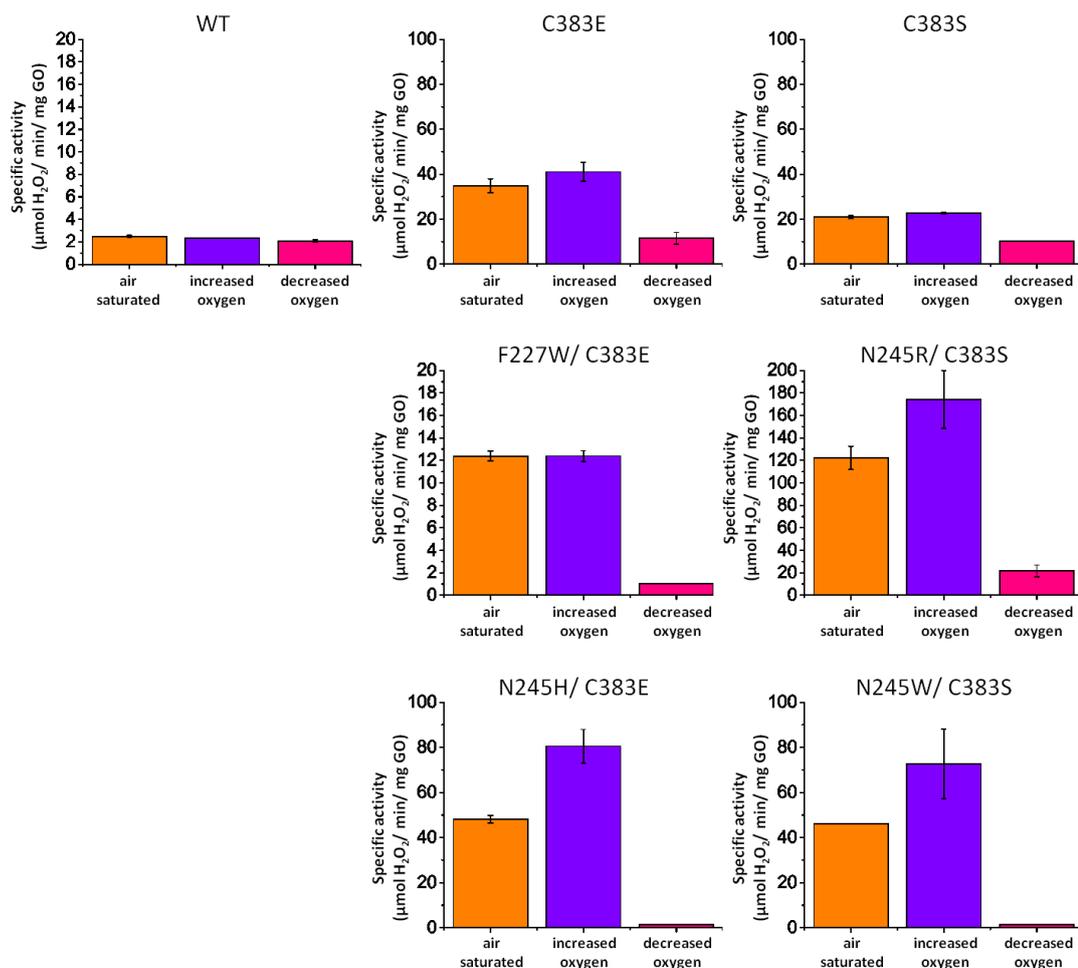
In summary, many of the mutations do appear to have an effect on the oxidative half reaction. However, without more detailed analyses, including accurate determination of the oxygen concentrations, it is difficult to confidently identify the effect(s). The different trends observed with the different sugar substrates support the possibility of different catalytic mechanisms for the three substrates. If the mechanisms were the same, altering oxygen concentration would be expected to have similar effects.

5.8 Determination of the regioselectivity of some variants

The oxidation of methyl D-galactopyranoside to the corresponding aldehyde can be followed by proton NMR spectroscopy (Parikka and Tenkanen, 2009). 1D and 2D NMR experiments were carried out by Parikka and Tenkanen (2009) with methyl α-D-galactopyranoside in order to assign the ^{1}H chemical shifts of the aldehyde product, which exists as a hydrate when dissolved in D_2O (Figure 5.24). If a mixture of isomeric forms of the sugars was used it would be very difficult to assign the chemical shifts as many would overlap. It is therefore essential to use only one isomer, such as the methyl α-pyranoside used for D-galactose.
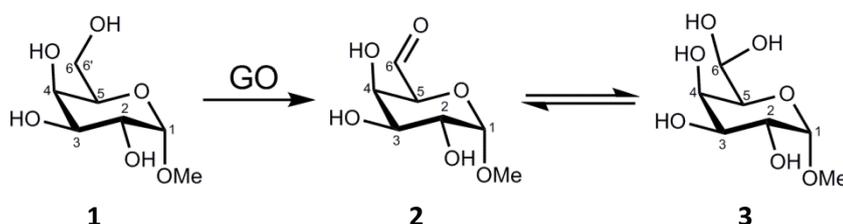
Figure 5.24. Oxidation of methyl α-D-galactopyranoside to form the aldehyde

Methyl α-D-galactopyranoside (**1**) is oxidised by GO to form the aldehyde (**2**) which occurs as a hydrate (**3**) in H_2O (or D_2O). Adapted from Parikka and Tenkanen (2009).

Figure 5.25 shows the assignments of ^{1}H chemical shifts. The percentage conversion can be estimated from the relative sizes of the peaks corresponding to H-4 in the substrate (**1**) and product (**3**) as these peaks are clearly distinct from other chemical shifts.
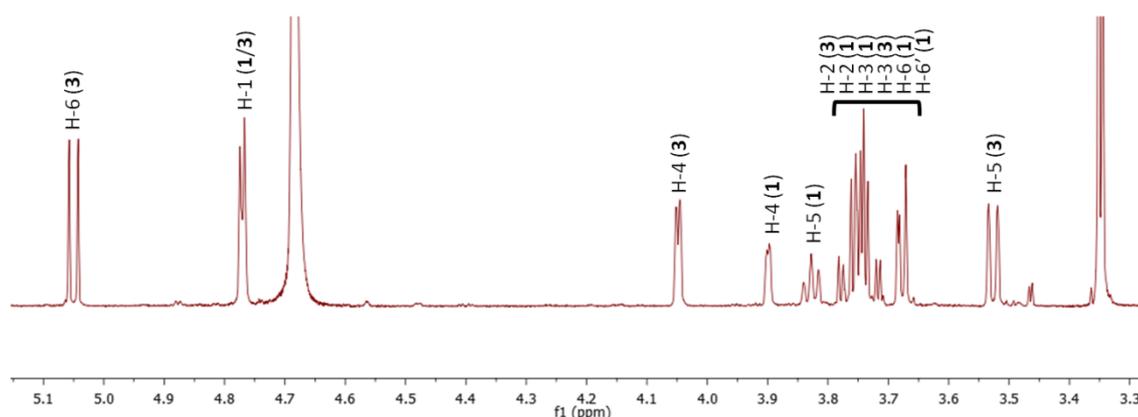
Figure 5.25. Chemical shift assignments determined by Parikka and Tenkanen (2009)

The molecule numbers **1** and **3** refer to the structures shown in Figure 5.24.

The initial aim of the NMR study was to characterise the reaction over several days, to provide information on enzyme stability and rate of conversion over time. This information is important for development of any of the variants for use in industrial scale syntheses. The hydrogen peroxide generated is known to inactivate the enzyme at high concentrations (Cooper et al., 1959) so catalase was included in the reaction to convert this to water. The inclusion of HRP in the reaction mixture maintains GO in the fully oxidised form (Section 1.6.3.1) and therefore enhances activity. The reaction conditions were optimised by Parikka and Tenkanen (2009): 7.5 ml water containing 70 mM methyl α-D-galactopyranoside, 45 U HRP, 11200 U catalase and 5.2 U GO, stirred vigorously at room temperature for up to 96 hours.

When the water was substituted with 25 mM sodium phosphate buffer, pH 7.0, it was reported by Parikka and Tenkanen (2009) that generation of the aldehyde product was significantly reduced. However, when this optimisation was repeated here, this was not the case. In water 76% conversion was observed in 24 hours, while in sodium phosphate, pH 7.0 79% conversion was observed. It was therefore decided to carry out all future reactions in buffer as maintenance of a constant pH is most likely to maintain enzyme activity. When a similar reaction was carried out by Sun et al. (2002), copper was included in the reaction mix. Addition of 0.1 mM $Cu(NO_3)_2$ to the 7.5 ml reaction was therefore trialled and resulted in a 5-10% increase in product generation over 24 hours. Neither Sun et al. (2002) or Parikka and Tenkanen (2009) give any information on the reaction vessel used for the reaction. The 7.5 ml reaction was therefore carried out in either a 10 ml beaker or a 14 ml glass bijou container, both with vigorous stirring. Surprisingly, this made a significant difference to the progression of the reaction as after 24 hours, the reaction carried out in the beaker had reached >80% conversion while the reaction in the bijou had only reached 34% conversion. This implies that the level of oxygenation of the reaction is important as the most significant difference between the two reaction vessels is the diameter which determines the surface area of the reaction mix. It was decided not to further optimise the reaction vessel or explore the effect of oxygenation as any use of the GO variants in industrial processes would be carried out in much larger volumes than 7.5 ml, so the reaction vessel and associated conditions would need to be optimised in a different way. Accurate determination of oxygen kinetics would also be useful for this optimisation.

Oxidation of D-galactose by GO has already been well characterised and is exploited in a number of commercial applications (Section 1.6.5.1). The oxidation of D-arabinose, D-glucose

and D-xylose by the variants showing enhanced activities was therefore of interest. Unfortunately, the reaction conditions optimised for oxidation of methyl α-D-galactopyranoside did not result in any generation of product in the oxidation of methyl β-D-xylopyranoside by 5 U of WT, C383E or N245R/ C383S. Only when 28 U of N245W/ C383S was used was a change in the NMR spectrum observed. Unfortunately, as shown in Figure 5.26, a large number of peaks were present in the spectrum of the product mixture and it was not possible to identify the position(s) of oxidation. This could imply oxidation at multiple positions to generate a number of different products, or the product could have undergone degradation before analysis by NMR.
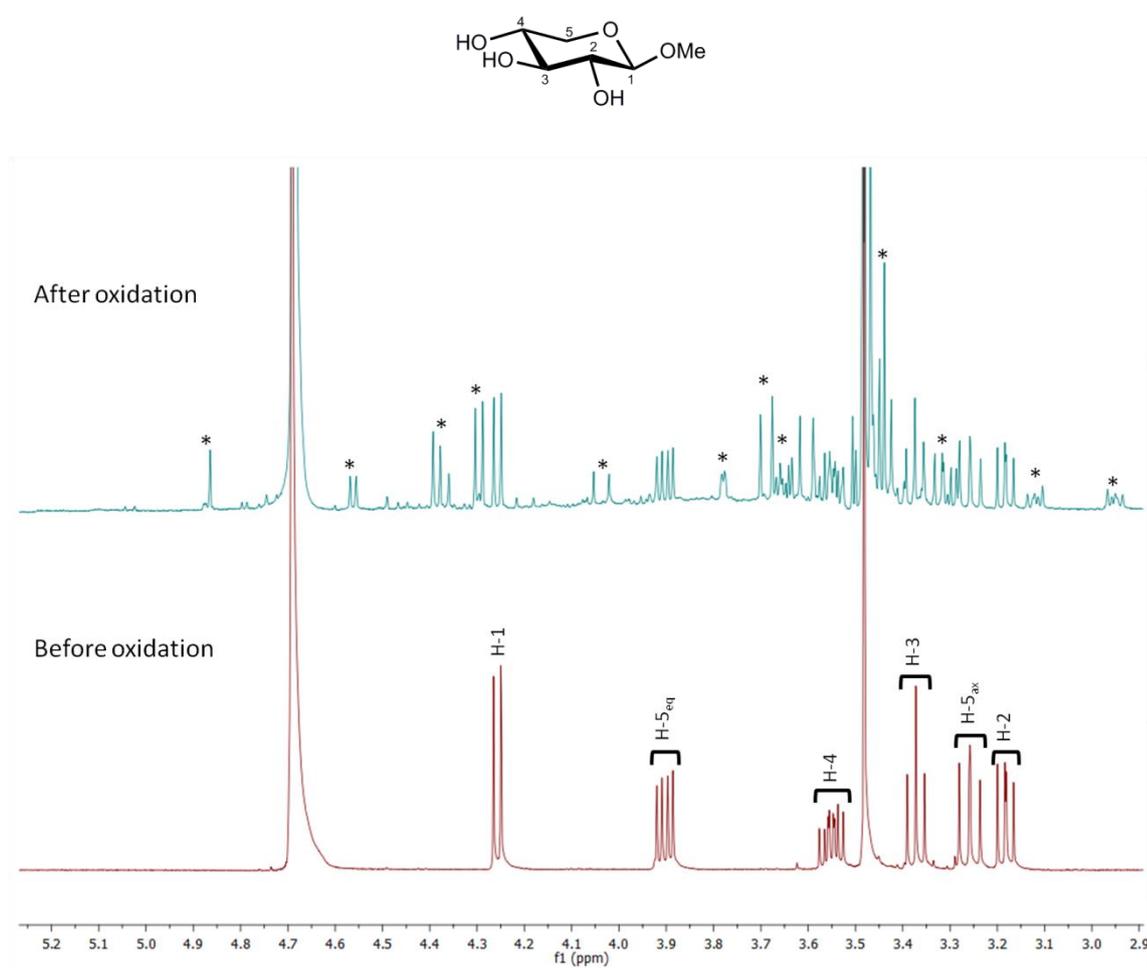
Figure 5.26. The structure of methyl β-D-xylopyranoside and NMR spectra before and after oxidation

Chemical shifts for each of the protons in methyl β-D-xylopyranoside have been assigned for the reaction before oxidation. However, after oxidation a large number of additional chemical shifts have appeared which were not possible to assign (marked *). These could indicate the presence of multiple oxidation products, or degradation of the product(s) before NMR analysis.

With methyl β-D-arabinopyranoside, it was not possible to add 5 U of the variant GO as the number of units per mg was too low. A reaction with 0.8 U C383E resulted in no detectable product after 72 hours, however 1.2 U of F227W/ C383E resulted in formation of product which was analysed by ^{1}H-NMR, ^{13}C-NMR (Figure 5.27 and Figure 5.28) and 2D correlation spectroscopy (COSY) experiments (Figure 5.29).

The data were analysed by Dr W. B. Turnbull. 2D COSY spectra (Figure 5.29) were used to assign signals in the ^{1}H-NMR spectra (Figure 5.27) to specific atoms in the starting material and product. The presence of cross-peaks correlating H-1, H-2 and H-3 implied that the oxidation reaction had occurred at position 4. This hypothesis was further supported by the lack of couplings from H-5 to an H-4 proton signal. The ^{13}C-NMR spectrum (Figure 5.28) showed an extra signal at 92.8 ppm which is consistent with the formation of a hydrate at the oxidised carbon as a ketone would be expected to have a chemical shift above 200 ppm.

The same changes in spectra were observed upon oxidation by F194I/ C383S and N245R/ C383S implying that all three variants oxidise methyl β-D-arabinopyranoside at the same position.
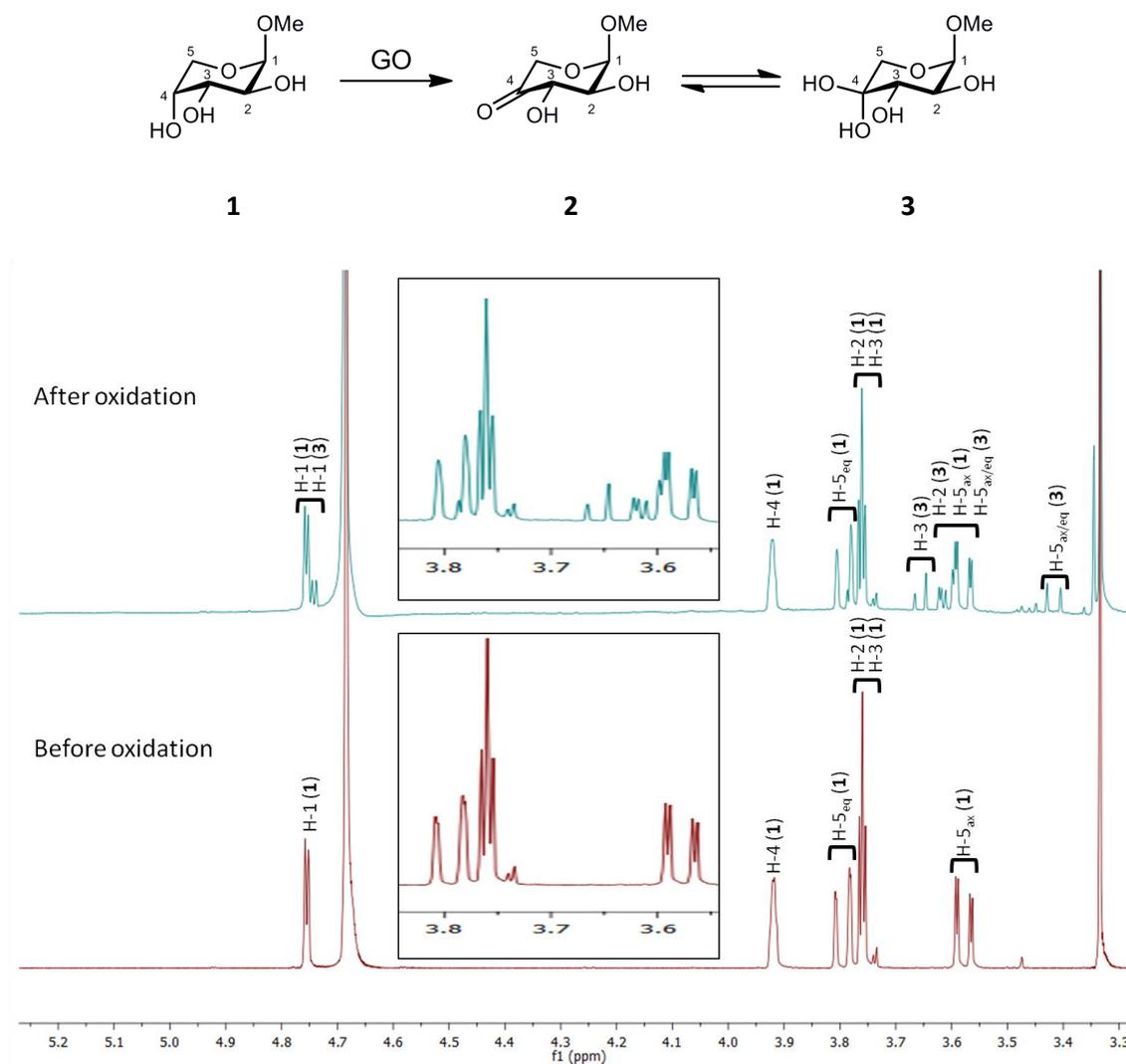
Figure 5.27. The proposed oxidation of methyl β-D-arabinopyranoside at the C-4 position and ¹H-NMR spectra before and after the reaction

The data support the reaction shown at the top where the GO variant oxidises methyl β-D-arabinopyranoside (**1**) specifically at the C-4 position. The insets show an expansion of the region 3.84-3.54 ppm.
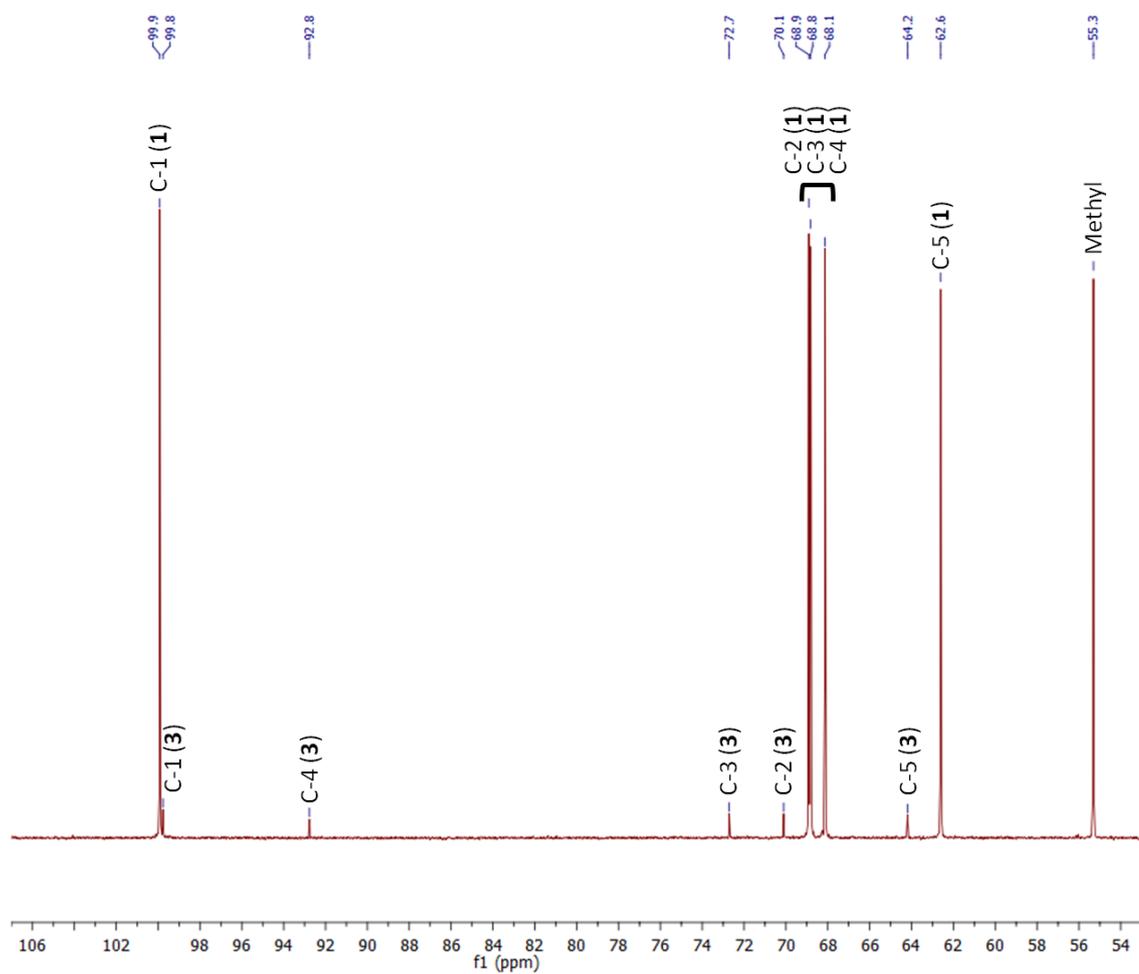
Figure 5.28. ^{13}C-NMR spectra after oxidation of methyl β-D-arabinopyranoside

Molecule numbering refers to the scheme shown at the top of Figure 5.27. The signal at 92.8 ppm is consistent with formation of a hydrate at position 4.
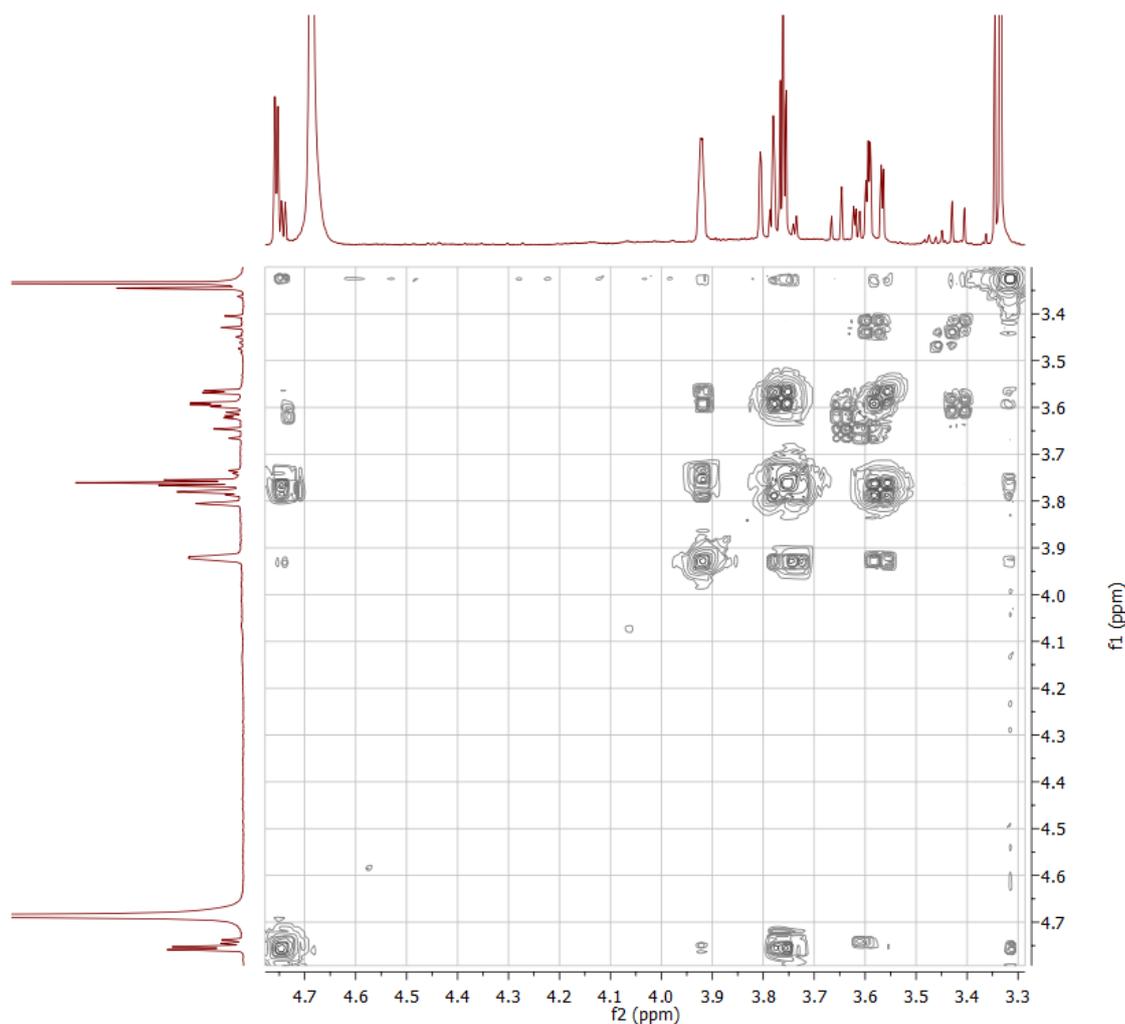
Figure 5.29. 2D COSY spectra after oxidation of methyl β-D-arabinopyranoside

The spectra were used to assign assign signals in the ^{1}H-NMR spectra (Figure 5.27).

As with methyl β-D-arabinopyranoside, it was not possible to add 5 U of the GO variants in the reaction with methyl α-D-glucopyranoside as the number of units per mg was too low, therefore 1.4 U of N245R/ C383E was used. As shown in Figure 5.30, the NMR spectrum after oxidation shows appearance of a chemical shift at approximately 5.2 ppm which corresponds to the proton at C-6 in the species with an aldehyde group at C-6 (**2**) (Sun et al., 2002). As this is the only difference in the spectra, it would appear that the N245R/ C383E variant oxidises specifically at the C-6 position of D-glucose, the same position that is oxidised by WT GO in the reaction with D-galactose.
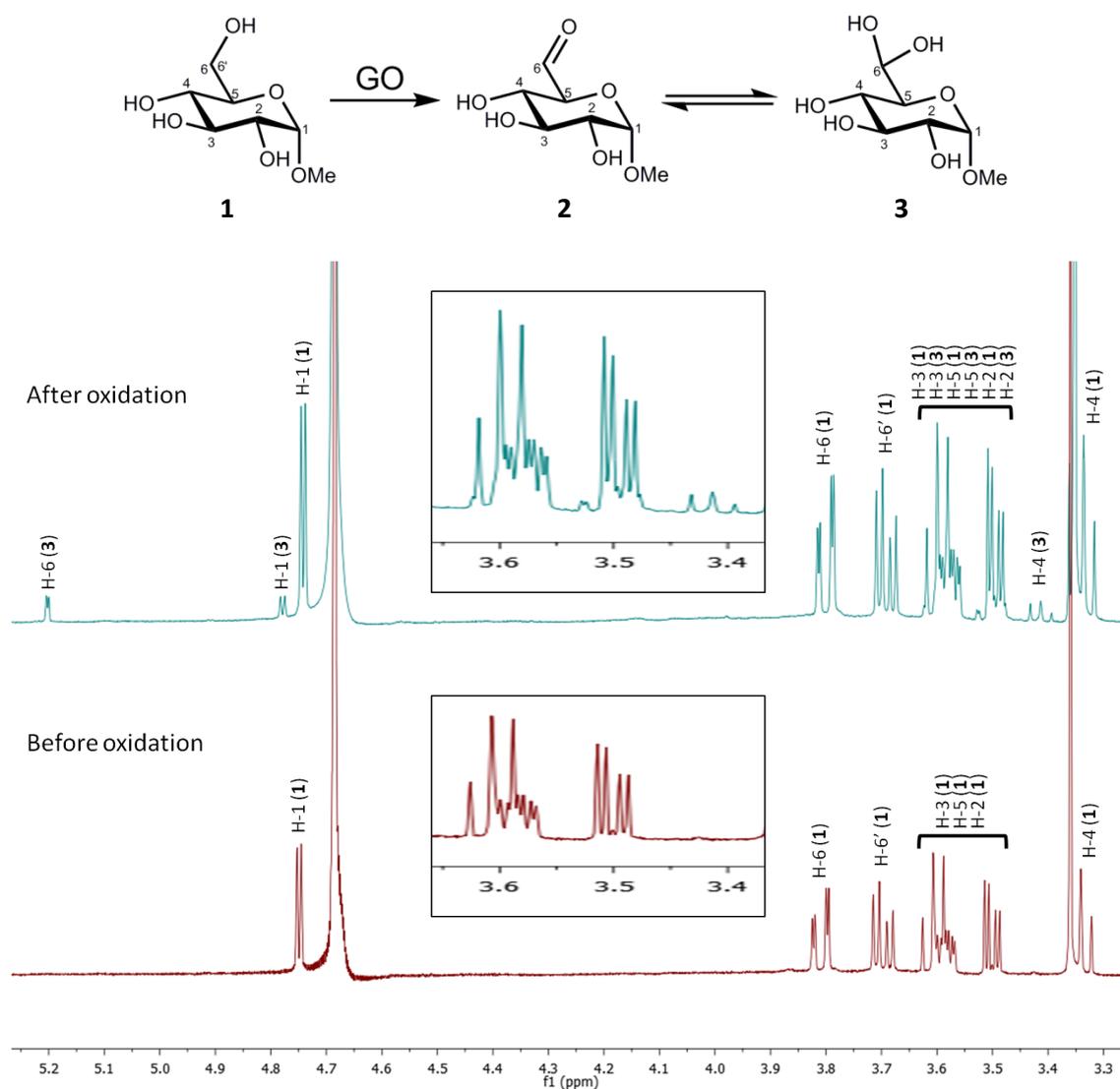
Figure 5.30. The proposed oxidation of methyl α-D-glucopyranoside at the C-6 position and NMR spectra before and after the reaction

Appearance of the chemical shift at ~5.2 ppm after oxidation supports the reaction shown at the top where the GO variant oxidises methyl α-D-glucopyranoside (**1**) specifically at the C-6 position. The insets show an expansion of the region 3.66-3.37 ppm.

5.9 The F194W/ W290F variant: a glycerol oxidase

The F194W/ W290F variant displayed a non-hyperbolic relationship between reaction velocity and substrate concentration for activity towards D-galactose and glycerol, so it was not possible to determine kinetic parameters. However, the activity measurements taken over a range of substrate concentrations revealed that the mutant displayed two- to three-fold higher activity towards glycerol than D-galactose (Table 5.9). As the WT enzyme is already saturated at 200 mM D-galactose, the decrease in D-galactose activity at lower substrate concentrations upon addition of the two mutations is much more significant than at higher substrate concentrations: D-galactose activity is reduced 92-fold at 200 mM compared to 23-fold at 800 mM. When combined with the four- to five-fold increase in glycerol activity the shift in specificity is around 100-fold at 800 mM substrate but over 200-fold at 400 mM substrate (Figure 5.31), the largest shift in specificity of any variant identified.

Table 5.9. Specific activities at pH 7.0 of WT and F194W/ W290F towards D-galactose and glycerol

Substrate concentration (mM)	Specific activity (μM H_2O_2/ sec/ μM GO)			
	WT		F194W/ W290F	
	D-galactose	Glycerol	D-galactose	Glycerol
200	830 ± 50	Not measured	9.1 ± 1.0	29 ± 0.50
400	790 ± 71	12 ± 0.88	19 ± 0.94	63 ± 6.0
600	880 ± 33	Not measured	28 ± 1.8	82 ± 2.7
800	930 ± 59	25 ± 2.5	41 ± 0.86	100 ± 1.6
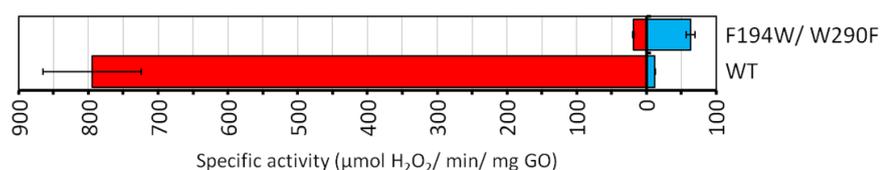
Figure 5.31. Graphical representations to visualise the substrate specificity of F194W/ W290F compared to WT at 400 mM

Unfortunately it was not possible to assign the chemical shifts of the oxidation reaction followed by ^{1}H-NMR (Figure 5.32). This is an important consideration for future continuation of this project as the product generated will determine the industrial potential of this variant.
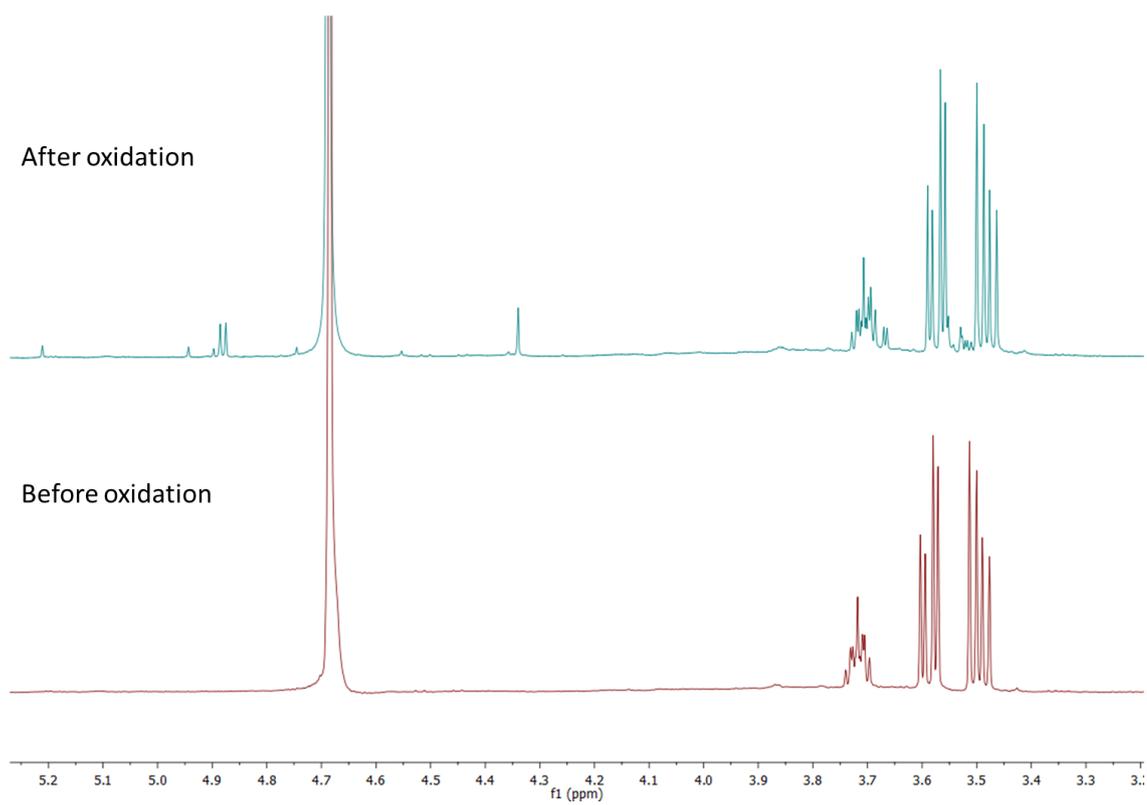
Figure 5.32. NMR spectra before and after oxidation of glycerol by the F194W/ W290F variant

5.10 Attempts to crystallise *E. coli*-expressed GO for structural analysis of variants

Previous attempts in the group to crystallise the *E. coli*-expressed N6M1 version of GO using the published conditions for the *P. pastoris*-expressed WT were unsuccessful. This could be due to a number of different factors, including effects of the M1 mutations (S20P, M70V, G195E, V494A and N535D), the Strep II-tag, or perhaps some unidentified factor which is essential for crystallisation picked up by the protein during expression and purification.

The process of protein crystallisation is still poorly understood and it is common to trial a large number of buffers, precipitants and additives in order to identify conditions which promote formation of crystals of suitable quality for diffraction experiments. Crystallisation trials were carried out with the C-terminally Strep II-tagged N6M1 (Deacon and McPherson, 2011) using crystallisation screens from Hampton Research: Crystal Screen 1, Crystal Screen 2, MembFac and SaltRx2. A number of crystals grew in the different screens but following analysis by UV light and diffraction experiments on the University of Leeds home X-ray source diffractometer, were identified as salt crystals containing no protein.

WT GO has been successfully crystallised following expression in *P. pastoris* (Wilkinson et al., 2004) or *A. nidulans* (Baron et al., 1994) and purification by ammonium sulfate precipitation followed by chromatography using cellulose phosphate or Sepharose 6B resin. The C-terminally Strep II-tagged N6M1 was expressed and purified as detailed in Section 2.5 and then incubated in the *P. pastoris* induction medium BMMY at 4 °C overnight. In order to purify the protein from the BMMY medium the mixture was applied to a cellulose phosphate column pre-equilibrated with 10 mM sodium phosphate, pH 7.3. To elute the GO from the column, the concentration of sodium phosphate, pH 7.3 was gradually increased to 500 mM. However, GO appeared exclusively in the column flow-through implying that it had not adsorbed to the cellulose phosphate resin.

The presence of the Strep II-tag could be responsible for the lack of binding to the cellulose phosphate column and also the lack of crystallisation. As detailed in Section 3.2.3, a TEV cleavage site was added to the construct in the pET28c plasmid to permit cleavage of the Strep II-tag following purification. This construct is available for further crystallisation studies if required.

Following the failure of the initial screening for crystallisation conditions using the Hampton screens, trials were carried out using crystallisation conditions based on the different published conditions (Table 5.10) (Firbank et al., 2001, Rogers et al., 2007, Deacon et al., 2004, Ito et al., 1991, Wilkinson et al., 2004). The N6M1strep, strepTEVN6M1 and SSstrepTEVN6M1 constructs were each dialysed against 20 mM MES, pH 7.0 and crystallisation trials set down using the sitting drop method. Drops containing 0.5 μl protein solution at 2, 4 or 6 mg/ ml and 0.5 μl mother liquor were used. Needle-like crystals grew in the screen using SSstrepTEVN6M1 at 4 mg/ ml in 10-24% polyethylene glycol (PEG) 8000, 100 mM MES, pH 5.8-6.4 and 100-200 mM calcium acetate (Figure 5.33). The crystals were confirmed to contain protein by analysis under UV light and following transfer to mother liquor containing 20% (v/ v) glycerol as a cryoprotectant and flash cooling in liquid nitrogen, were used in diffraction experiments at Diamond Light Source, Oxfordshire. Unfortunately none of the crystals showed diffraction to high resolution so further optimisation will need to be carried out.

After the trials detailed above were completed, another group published a study including a crystal structure of *E. coli*-expressed GO containing the M1 mutations as well as W290F, R330K, Q406T and P463I (Rannes et al., 2011). The structure of this variant, which contained multiple mutations around the active site, and the C383S structure from *P. pastoris*-expressed protein solved previously (Wilkinson et al., 2004), revealed no noticeable changes in the structure compared to the WT enzyme apart from the substituted residues. Neither structure provided an explanation for the observed changes in activity. It was therefore decided in this project not to focus further on obtaining crystal structures of the variants. However, if this were to be carried out in the future, the crystallisation conditions used by Rannes et al. (2011) would be used as a starting point.
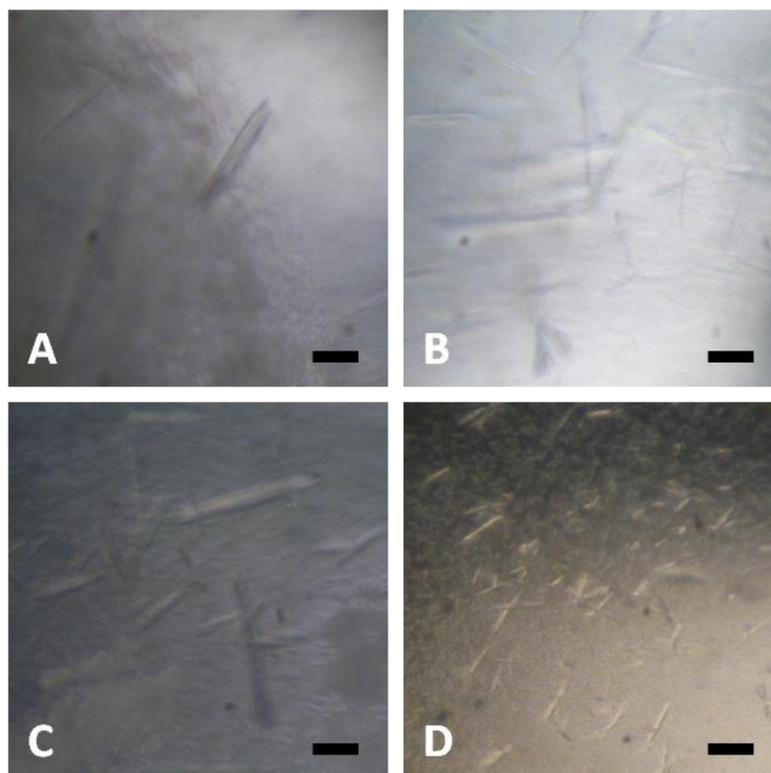
Figure 5.33. Some of the GO crystals grown under different conditions

In each photo the bar indicates 100 μm. The crystallisation conditions were as follows: **A**: 16% PEG8000, 100 mM MES, pH 6.0, 200 mM calcium acetate; **B**: 20% PEG8000, 100 mM MES, pH 6.0, 100 mM calcium acetate; **C**: 18% PEG8000, 100 mM MES, pH 6.0, 150 mM calcium acetate; **D**: 10% PEG8000, 100 mM MES, pH 5.8, 100 mM calcium acetate.

Table 5.10 (Next page). Crystallisation trials based on published conditions

The 96 different combinations of precipitant (PEG8000, PEG4000 or ammonium sulfate), buffer (MES, pH 5.8-6.4 or sodium acetate, pH 4.4-5.0) and additive (100-200 mM calcium acetate or 100-200 mM ammonium sulfate) were based on previous published conditions from Firbank et al. (2001) and Deacon et al. (2004): green; Wilkinson et al. (2004): white; or Ito et al. (1991) and Rogers et al. (2007): blue. Conditions shown in bold are those which resulted in crystal growth.

	1	2	3	4	5	6	7	8	9	10	11	12
A	**10% PEG8000 100 mM MES pH 5.8 100 mM Ca(OAc)$_2$**	10% PEG8000 100 mM MES pH 6.0 150 mM Ca(OAc)$_2$	10% PEG8000 100 mM MES pH 6.2 200 mM Ca(OAc)$_2$	10% PEG8000 100 mM MES pH 6.4 100 mM Ca(OAc)$_2$	20% PEG4000 100 mM NaOAc pH 4.4 100 mM (NH$_4$)$_2$SO$_4$	20% PEG4000 100 mM NaOAc pH 4.6 150 mM (NH$_4$)$_2$SO$_4$	20% PEG4000 100 mM NaOAc pH 4.8 200 mM (NH$_4$)$_2$SO$_4$	20% PEG4000 100 mM NaOAc pH 5.0 100 mM (NH$_4$)$_2$SO$_4$	1.0 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.4	1.0 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.6	1.0 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.8	1.0 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 5.0
B	12% PEG8000 100 mM MES pH 5.8 150 mM Ca(OAc)$_2$	12% PEG8000 100 mM MES pH 6.0 100 mM Ca(OAc)$_2$	12% PEG8000 100 mM MES pH 6.2 150 mM Ca(OAc)$_2$	12% PEG8000 100 mM MES pH 6.4 200 mM Ca(OAc)$_2$	21% PEG4000 100 mM NaOAc pH 4.4 150 mM (NH$_4$)$_2$SO$_4$	21% PEG4000 100 mM NaOAc pH 4.6 100 mM (NH$_4$)$_2$SO$_4$	21% PEG4000 100mM NaOAc pH 4.8 150 mM (NH$_4$)$_2$SO$_4$	21% PEG4000 100mM NaOAc pH 5.0 200 mM (NH$_4$)$_2$SO$_4$	1.1 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.4	1.1 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.6	1.1 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.8	1.1 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 5.0
C	14% PEG8000 100mM MES pH 5.8 200 mM Ca(OAc)$_2$	14% PEG8000 100mM MES pH 6.0 200 mM Ca(OAc)$_2$	14% PEG8000 100mM MES pH 6.2 100 mM Ca(OAc)$_2$	14% PEG8000 100mM MES pH 6.4 150 mM Ca(OAc)$_2$	23% PEG4000 100mM NaOAc pH 4.4 200 mM (NH$_4$)$_2$SO$_4$	23% PEG4000 100mM NaOAc pH 4.6 200 mM (NH$_4$)$_2$SO$_4$	23% PEG4000 100mM NaOAc pH 4.8 100 mM (NH$_4$)$_2$SO$_4$	23% PEG4000 100mM NaOAc pH 5.0 150 mM (NH$_4$)$_2$SO$_4$	1.3 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.4	1.3 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.6	1.3 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.8	1.3 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 5.0
D	**16% PEG8000 100 mM MES pH 5.8 150 mM Ca(OAc)$_2$**	**16% PEG8000 100 mM MES pH 6.0 200 mM Ca(OAc)$_2$**	**16% PEG8000 100 mM MES pH 6.2 100 mM Ca(OAc)$_2$**	**16% PEG8000 100 mM MES pH 6.4 100 mM Ca(OAc)$_2$**	24% PEG4000 100 mM NaOAc pH 4.4 150 mM (NH$_4$)$_2$SO$_4$	24% PEG4000 100 mM NaOAc pH 4.6 200 mM (NH$_4$)$_2$SO$_4$	24% PEG4000 100 mM NaOAc pH 4.8 100 mM (NH$_4$)$_2$SO$_4$	24% PEG4000 100 mM NaOAc pH 5.0 100 mM (NH$_4$)$_2$SO$_4$	1.4 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.4	1.4 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.6	1.4 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.8	1.4 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 5.0
E	**18% PEG8000 100 mM MES pH 5.8 100 mM Ca(OAc)$_2$**	**18% PEG8000 100 mM MES pH 6.0 150 mM Ca(OAc)$_2$**	**18% PEG8000 100 mM MES pH 6.2 200 mM Ca(OAc)$_2$**	**18% PEG8000 100 mM MES pH 6.4 150 mM Ca(OAc)$_2$**	26% PEG4000 100 mM NaOAc pH 4.4 100 mM (NH$_4$)$_2$SO$_4$	26% PEG4000 100 mM NaOAc pH 4.6 150 mM (NH$_4$)$_2$SO$_4$	26% PEG4000 100 mM NaOAc pH 4.8 200 mM (NH$_4$)$_2$SO$_4$	26% PEG4000 100 mM NaOAc pH 5.0 150 mM (NH$_4$)$_2$SO$_4$	1.6 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.4	1.6 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.6	1.6 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.8	1.6 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 5.0
F	**20% PEG8000 100 mM MES pH 5.8 200 mM Ca(OAc)$_2$**	**20% PEG8000 100 mM MES pH 6.0 100 mM Ca(OAc)$_2$**	**20% PEG8000 100 mM MES pH 6.2 150 mM Ca(OAc)$_2$**	**20% PEG8000 100 mM MES pH 6.4 200 mM Ca(OAc)$_2$**	27% PEG4000 100 mM NaOAc pH 4.4 200 mM (NH$_4$)$_2$SO$_4$	27% PEG4000 100 mM NaOAc pH 4.6 100 mM (NH$_4$)$_2$SO$_4$	27% PEG4000 100 mM NaOAc pH 4.8 150 mM (NH$_4$)$_2$SO$_4$	27% PEG4000 100 mM NaOAc pH 5.0 200 mM (NH$_4$)$_2$SO$_4$	1.7 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.4	1.7 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.6	1.7 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.8	1.7 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 5.0
G	**22% PEG8000 100 mM MES pH 5.8 200 mM Ca(OAc)$_2$**	**22% PEG8000 100 mM MES pH 6.0 100 mM Ca(OAc)$_2$**	22% PEG8000 100 mM MES pH 6.2 100 mM Ca(OAc)$_2$	**22% PEG8000 100 mM MES pH 6.4 150 mM Ca(OAc)$_2$**	29% PEG4000 100 mM NaOAc pH 4.4 200 mM (NH$_4$)$_2$SO$_4$	29% PEG4000 100 mM NaOAc pH 4.6 100 mM (NH$_4$)$_2$SO$_4$	29% PEG4000 100 mM NaOAc pH 4.8 100 mM (NH$_4$)$_2$SO$_4$	29% PEG4000 100 mM NaOAc pH 5.0 150 mM (NH$_4$)$_2$SO$_4$	1.9 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.4	1.9 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.6	1.9 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.8	1.9 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 5.0
H	**24% PEG8000 100 mM MES pH 5.8 150 mM Ca(OAc)$_2$**	**24% PEG8000 100 mM MES pH 6.0 200 mM Ca(OAc)$_2$**	**24% PEG8000 100 mM MES pH 6.2 150 mM Ca(OAc)$_2$**	**24% PEG8000 100 mM MES pH 6.4 100 mM Ca(OAc)$_2$**	30% PEG4000 100 mM NaOAc pH 4.4 150 mM (NH$_4$)$_2$SO$_4$	30% PEG4000 100 mM NaOAc pH 4.6 200 mM (NH$_4$)$_2$SO$_4$	30% PEG4000 100 mM NaOAc pH 4.8 150 mM (NH$_4$)$_2$SO$_4$	30% PEG4000 100 mM NaOAc pH 5.0 100 mM (NH$_4$)$_2$SO$_4$	2.0 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.4	2.0 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.6	2.0 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 4.8	2.0 M (NH$_4$)$_2$SO$_4$ 100 mM NaOAc pH 5.0

5.11 Discussion

The data presented in this chapter provide valuable insight into the alterations in activity identified in Chapter 4. Further experiments would be required to deduce the detailed mechanisms of binding and catalysis for each variant and each substrate as discussed in Chapter 6. However, based on the observation that the published structures of GO variants show minimal structural rearrangements, it is possible to model the variants identified in this study *in silico* using the program COOT (Emsley et al., 2010). The rotamer which resulted in the least disruption to the surrounding molecule was selected for each mutation. When considered with the biochemical data presented here, it is possible to use these structures to postulate how the new activities might be brought about. For each mutation, it is important to consider both the effect on activity with the alternative substrate (D-glucose, D-arabinose or D-xylose) and with the native substrate D-galactose.

There are four main points to be considered for the variants showing altered D-arabinose, D-glucose or D-xylose activity:

1. The C383E and C383S mutations result in significant enhancements in activity towards the alternative substrates D-glucose, D-arabinose and D-xylose compared to WT; however the effect on D-galactose activity at pH 7.0 is limited

2. F194G/ I/ T and F227W mutations result in significant reductions in activity towards D-galactose at pH 7.0, compared to WT; however the effect on activity towards alternative substrates is limited

3. N245H/ R/ W mutations result in moderate reductions in activity towards D-galactose at pH 7.0, compared to WT; activity towards alternative substrates is enhanced in most cases

4. In most cases, the double mutants combine the reduced D-galactose activity of the F194, F227 or N245 mutation with the increased activity of the C383 mutation.

<u>The C383 mutations</u>

Despite previous work on mutations at position 383 (Section 1.6.2), including solution of the crystal structure of C383S (Wilkinson et al., 2004), the effect on catalysis of the different mutations at this position is yet to be explained. C383S shows similar specific activity values with 1 M D-galactose as WT GO, while specific activity for C383E is reduced by around half.

Despite this, both enzymes show an improved efficiency (k_{cat}/K_M) compared to WT due to significantly reduced K_M values, implying more efficient substrate binding. Activity towards methyl D-galactopyranosides shows exactly the same trends as the WT enzyme: activity towards the α-anomer is reduced by half and activity towards the β-anomer is reduced by around a quarter compared to the isomeric, non-methylated mixture. Therefore it would seem that, whatever effect the C383E/ S mutation has on D-galactose binding and oxidation, the area around C-1 of the D-galactose molecule is not involved.

The C383S mutation is thought to bring about the observed changes due to subtle changes in active site structure or flexibility caused by the introduction of the stronger hydrogen bonding group (–OH compared to –SH) which affects the hydrogen bonding network around the active site (Wilkinson et al., 2004). The C383E mutation (Figure 5.34) introduces a much more significant change as glutamate is much larger than cysteine and is a charged residue. A number of interactions around the active site may be affected. For example, Thr440 and Phe441 may move away from the Glu383 due to unfavourable interactions and the hydrogen bond between Cys383 and Tyr405 via a water molecule may be replaced by a direct hydrogen bond between the two residues. Tyr405 has been shown previously to play an important role within the active site (Deacon, 2008) which is supported by the high level of conservation of this residue in the 50 sequences showing highest identity to *F. graminearum* GO (Figure 5.1). Movement of Tyr405 or changes in the hydrogen bond interactions due to changes at positions 383, 440 and 441 may affect copper ligation by Tyr495 and maybe Tyr272, resulting in major changes in substrate binding and catalysis.
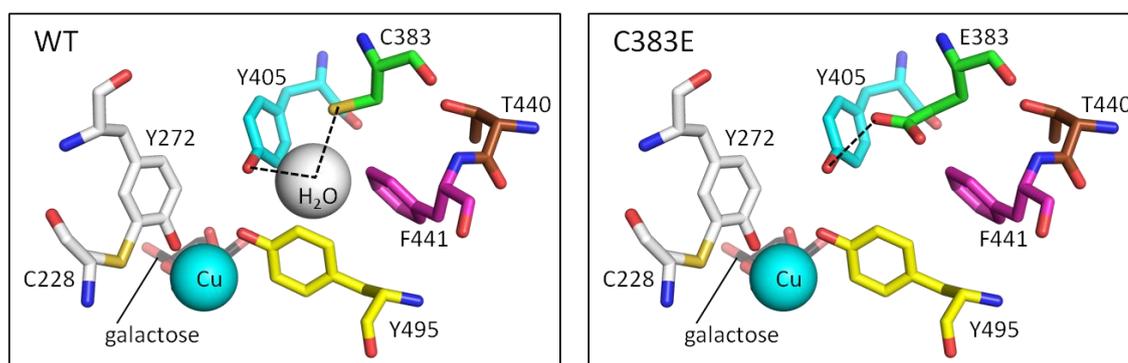
Figure 5.34. Comparison of WT and C383E structures

> The figure was created with PyMol using .pdb file 1GOF and shows all residues which may be affected by the C383E substitution. Hydrogen bonds are shown as dashed black lines: Cys383 and Tyr405 both form hydrogen bonds with a water molecule in the WT structure, Glu383 may form a direct hydrogen bond with Tyr405 in the mutant.

With D-arabinose, D-glucose and D-xylose, activity compared to WT was greatly increased with the C383E variant and, to a lesser extent, the C383S variant. In the case of D-glucose and D-xylose, addition of the methyl group at position C-1 did not affect activity significantly, implying that the pyranose substrate binds in a similar conformation to WT, although the C-6 hydroxyl is not present in D-xylose. With D-arabinose however, C383E and C383S show a strong preference for methyl α-D-arabinofuranoside over methyl β-D-arabinopyranoside and activity is much higher with the α-furanoside compared to the isomeric mixture. As this effect is not seen with the WT, D-arabinose may bind differently in the mutants than in the WT active site. Measurements with D-arabinose also displayed significantly greater changes in the oxidative half reaction than D-glucose or D-xylose, again implying major changes to substrate binding and/ or catalysis.

Comparing the pH profiles of D-glucose activity of C383E and C383S with WT reveals an increase in the optimum pH of around 2.0 units in both cases. While this matches previous data for the D-galactose pH profile for C383E, C383S has been previously reported to show a decrease in the pH optimum with D-galactose (Deacon, 2008). The different results are likely due to variations in preparation of the protein or different experimental conditions. The changes in pH optimum for D-galactose and D-glucose activity cannot be directly explained by the data presented here but may be due to alterations in the environment of residues involved in binding or catalysis resulting in a change in pK_a. It is, perhaps, surprising that introduction of the negatively charged glutamate results in an *increased* pH optimum as the opposite would normally be expected. As the C383E mutation has a similar effect on the pH profile of D-glucose and D-galactose activities and the N245R/ C383E variant oxidises D-glucose at the C-6 hydroxyl, the same as WT oxidation of D-galactose, it is likely that D-glucose binds the C383E active site in a similar way to D-galactose binding to WT. How the steric clash with the equatorial C-4 hydroxyl in D-glucose is accommodated (Section 1.6.2) cannot be explained without further knowledge of the structural changes induced by C383E, for which a crystal structure is absolutely required. The increased activity for the three alternative substrates is most likely due to changes in the shape of the active site and/ or increased flexibility, brought about by the C383X mutations, leading to more efficient binding of substrates which are less complementary to the shape of the active site. Molecular modelling or analysis of the crystal structure of an enzyme-substrate complex will provide more information on the basis of the increased activities.

The F194 and F227 mutations

The three mutations at position 194 resulted in a significant disruption of activity towards D-galactose, particularly F194G and F194I. The data imply that these two mutations have a negative effect on D-galactose binding in the active site as the $K_{M(galactose)}$, which gives an indication of binding affinity, is significantly increased. Relative activities with the α- and β-anomers of methyl D-galactopyranoside reveal another difference which can be attributed to substrate binding, as WT GO shows a preference for the β-anomer while the F194G and F194I mutants show no preference. This may suggest that there is more space around the C-1 position in the active site of the mutants and that the hydrogen bonding interaction with Tyr329 (Section 1.6.2) may not be present. In 42 of the 50 GO sequences showing highest identity with GO from *F. graminearum* a bulky aromatic residue is present at position 194. While it is thought that the hydrophobic nature of Phe194 may be important in substrate positioning through hydrophobic interactions with the B-face of D-galactose (Section 1.6.2) (Wachter and Branchaud, 1996), it is possible that the size of this residue is a more important feature. Glycine and isoleucine are both hydrophobic residues but are significantly smaller than phenylalanine. Perhaps substitution with these residues increases the space around position 194 resulting in less optimal positioning of D-galactose for hydrogen bonding interactions with other residues in the active site (Section 1.6.2) (Figure 5.35).
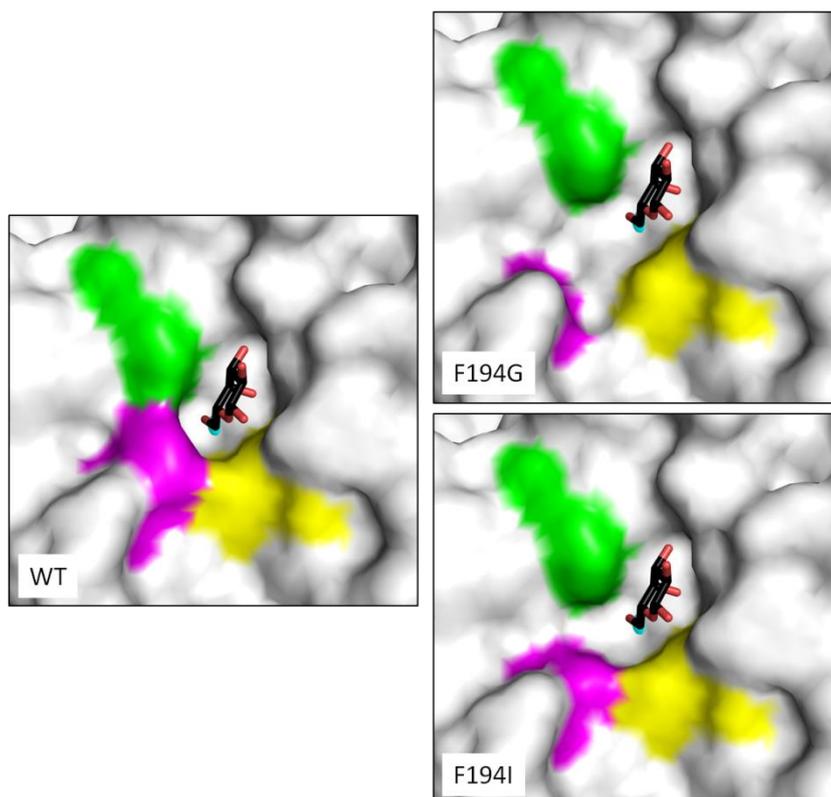
Figure 5.35. Surface view of WT GO and the F194G and F194I variants

F194 mutations were modelled using COOT and the rotamer which led to the least potential structural disruption was selected. The figure was created with PyMOL using .pdb file 1GOF and shows the view looking down from bulk solvent into the active site. F464 is green, W290 is yellow and the residue at position 194 is magenta. It is clear that substitution with a Gly or Ile at position 194 results in much more space in this area of the active site compared to the Phe194 in the WT structure.

The F194T mutation results in reduced activity towards D-galactose but to a lesser extent than the F194G or F194I mutations. The variant shows a preference for methyl β-D-galactopyranoside over the α-anomer, similar to WT GO. $K_{M(galactose)}$ is also within a similar range to WT, implying that this mutation has not led to as significant a disruption of D-galactose binding as the other two mutations. Introduction of the polar threonine at position 194 (Figure 5.36) is likely to have a significant effect on substrate binding in this area. As well as reducing the hydrophobicity of the region, Thr194 may interact unfavourably with Phe227. Due to its packing against Trp290 and proximity to the copper and the thioether bond, Phe227 seems unlikely to be able to move away from Thr194. It is more likely that the loop containing Thr194 (residues 192-196) would instead move away from the copper resulting in a greater increase in space in this area of the active site than was seen with the F194G and F194I

mutations (Figure 5.36). The polar Thr may also permit waters to occupy this portion of the active site and form hydrogen bonds with Thr194, as well as the substrate hydroxyl(s). However, the only moderate reduction in D-galactose activity that was observed is difficult to explain.
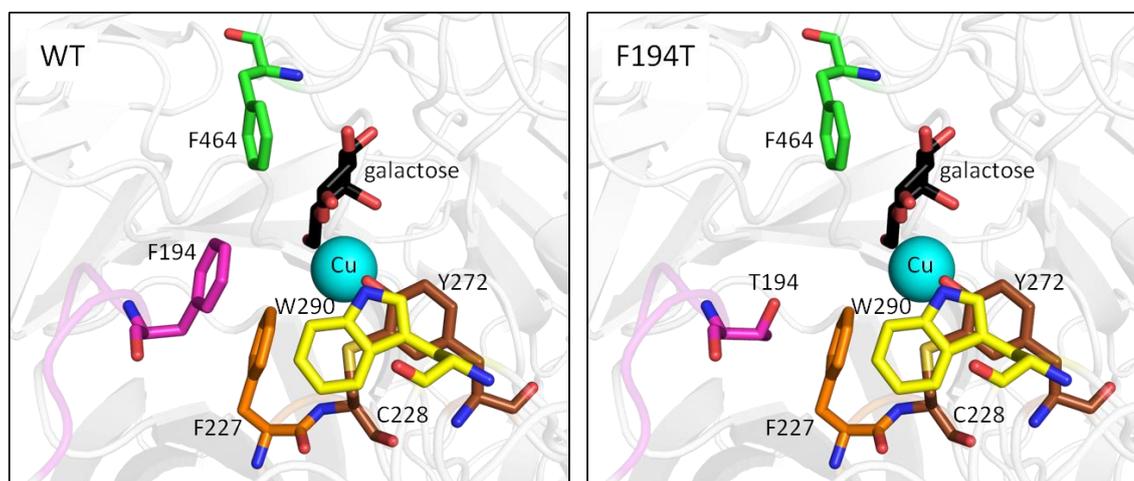
Figure 5.36. Comparison of WT and F194T structures

The F194T mutation was modelled using COOT and the rotamer which led to the least potential structural disruption was selected. The loop containing residue 194, which may move away from the copper upon mutation to Thr, is shown in magenta. Figure created with PyMOL using .pdb file 1GOF.

The reductions in activity of the three F194 mutants towards D-glucose were not as significant as the reductions in D-galactose activity. The significant reductions in activity upon addition of a methyl group at position C-1 of D-glucose in the F194G and F194I mutants imply that binding of D-glucose occurs quite differently in these mutants than in WT. This is likely due to the increased space around the area of the active site normally involved in preventing D-glucose binding caused by the substitution of Phe194 for smaller residues (Figure 5.35). Perhaps the C-1 end of the D-glucose molecule points into the active site rather than being relatively solvent exposed as predicted for D-galactose binding to the WT enzyme (Wachter and Branchaud, 1996). The likelihood of a different binding conformation was further backed up by the significant change in the pH profile of D-glucose activity to a much more acidic pH optimum. The oxidative half of the reaction does not seem to be significantly affected in the F194G/ C383S and F194I/ C383S variants, although it would be interesting to see the effect of the F194 single mutants. Unfortunately due to the extremely low activity towards the methyl D-glucopyranosides, it was not possible to analyse the products of the oxidation by NMR to

determine the position of oxidation on the D-glucose molecule for these variants. However, given the differences observed with methyl D-glucopyranosides, it would not be surprising if oxidation by F194 mutants occurred at a different position on the D-glucose molecule compared to the N245R/ C383E variant which was shown to oxidise at C-6 (Section 5.8).

F194I shows a moderate decrease in activity towards D-arabinose compared to WT. The mutant is barely active against the methyl α-D-furanoside, unlike WT, although activity towards the methyl β-D-pyranoside is the same as WT. This could imply that WT GO binds and oxidises the two methyl D-arabinosides by different mechanisms and that only one is disrupted by the F194I mutation. However, it could also be the case that binding of methyl β-D-arabinopyranoside is quite different in WT and F194I.

The F227W mutation leads to a significant reduction in activity towards D-galactose but the only slightly reduced K_M, compared to WT, implies that D-galactose is still able to bind in the active site. Without QM/ MM calculations or a crystal structure, it is not possible to estimate the structure of the mutant active site as the bulky tryptophan residue exists in close proximity to key residues including Trp290, Tyr272, Cys228 and the active site copper and copper ligand His496. It is likely this proximity that explains the lack of activity with any substrate for the majority of variants in libraries containing mutations at position 227. F227W shows only small reductions in activity towards D-glucose and D-xylose, compared to WT, while D-arabinose activity is increased. The dramatically different pH profile implies significant differences in binding and/ or oxidation of D-glucose but this is unsurprising given the significant structural change required to accommodate the tryptophan residue.

<u>The N245 mutations</u>

The N245H, N245R and N245W mutations all result in reduction of D-galactose activity, but not to the same degree as for the F194 and F227 mutations. The K_M values remain in a similar range to WT, implying that substrate binding is not majorly affected. The effect of adding a methyl group at position C-1 is also very similar to WT for the N245H and N245R mutants, while there is barely any effect for the N245W mutant compared to activity with the non-methylated isomeric mixture. The N245H and N245R mutations result in increased activity towards D-xylose while N245W reduces D-xylose activity. No effect is observed when a methyl group is added to D-xylopyranose, as with WT. N245R also has minor effects on D-glucose and D-arabinose activities with the most significant change being a relatively high level of activity towards methyl α-D-arabinofuranoside while methyl β-D-arabinopyranoside activity is similar

to WT. These minor changes may be explained by changes in the surface charge in this area of the active site by addition of the arginine which may affect how different substrates access the active site (Figure 5.37).
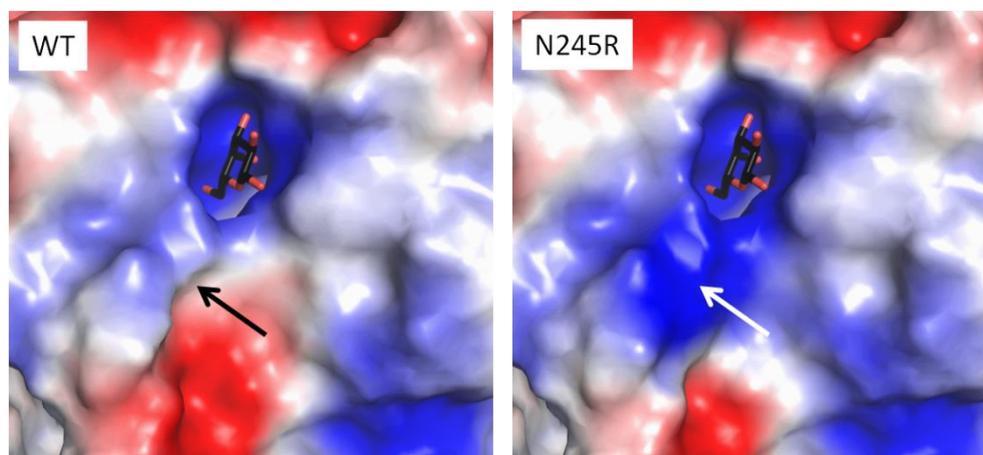
Figure 5.37. Surface view of the active site showing relative charge

Positively charged regions are shown in blue while negatively charged regions are red. The image was generated with PyMOL using .pdb file 1GOF. It is clear that the N245R substitution increases the positive charge at a region of the protein surface (indicated by an arrow) which may affect substrate access to the active site. D-galactose is shown in black, modelled according to Wachter and Branchaud (1996).

Changes to hydrogen bonding interactions with Asp225 or Ser197 brought about by the N245R or N245H mutations or changes in the environment close to Phe227 and Trp290 may also account for the differences (Figure 5.38). It is, however, surprising that substitution of Asn with the much larger, more hydrophobic Trp did not result in more dramatic changes to enzyme activity.
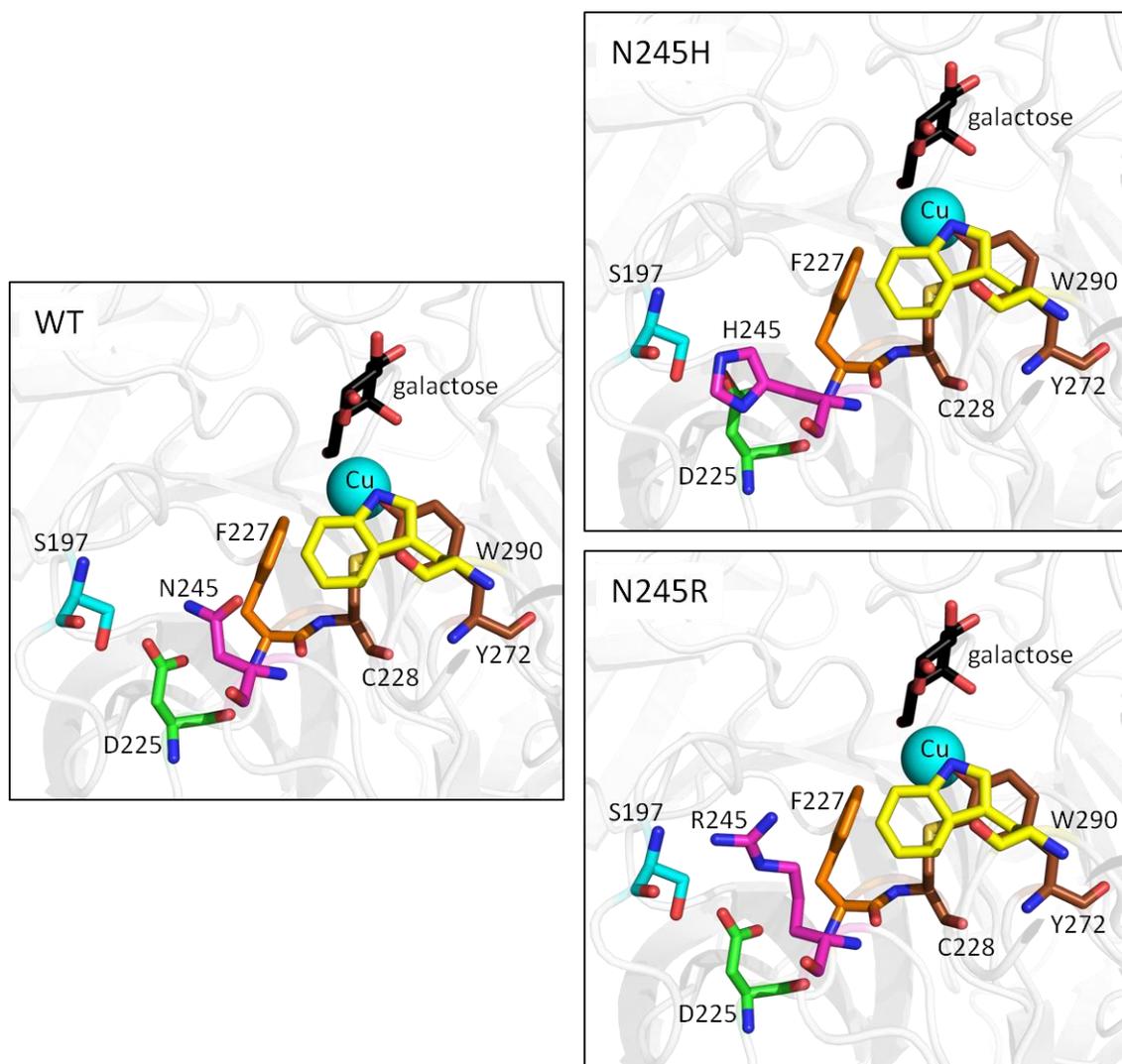
Figure 5.38. The predicted structures of the N245H and N245R variants compared to WT

> N245 mutations were modelled using COOT and the rotamer which led to the least potential structural disruption was selected. Introduction of a histidine or arginine may alter the environment of the surrounding residues (shown). Figure created with PyMOL using .pdb file 1GOF.

Double mutants

Phe194, Phe227 and Asn245 occur close to the active site copper and are more surface exposed than Cys383. This makes it more straightforward to postulate how mutations at these positions can change enzyme activity. As explaining the changes observed with the Cys383Glu and Cys383Ser mutations is much more difficult, it is not possible to fully rationalise the novel activities of the double mutants which contain a mutation at position 194, 227 or 245 combined with C383E or C383S.

With the exception of the results for F194T which appear to be anomalous, when the F194 or F227 mutations are combined with C383E/ S, D-galactose activity shows similar trends to the F194X/ F227X single mutants, *i.e.* reduced k_{cat}, increased K_M and limited or no preference for the α- or β-anomer of the methyl D-galactopyranoside. This is likely due to the changes in the area of the active site around position 194 or 227 which reduce the efficiency of D-galactose binding to such an extent that the increased flexibility or changes induced by the C383 mutations cannot counteract the effect.

Addition of the C383S mutation to the F194I mutant led to an increase in D-arabinose activity with essentially no effect on the specific activity with 1 M D-galactose. The maintenance of the strong preference for methyl β-D-arabinopyranoside over methyl α-D-arabinofuranoside seen in F194I/ C383S implies that D-arabinose binds in the same way in the single and double mutants, but that the increased flexibility induced by the C383S mutation improves the efficiency of binding and therefore catalysis. NMR data reveals that oxidation occurs at the C-4 position which implies the methyl β-D-arabinopyranoside binds in a different orientation to how D-galactose binds to WT GO, where the axial C-4 hydroxyl forms hydrogen bond interactions with R330 and Y495 (Section 1.6.2). It is surprising that improvement in D-arabinose activity of the F194I/ C383S variant is so much greater than for the F194G/ C383S mutant (Table 4.11) and implies that the larger size of the Ile at this position is important in correct positioning of the D-arabinose molecule. When C383S was added to mutations of F194 the strong inhibition of D-glucose activity upon addition of a methyl group at C-1 was maintained. However, the pH profile of the double mutant is much more similar to the C383 mutant or WT. Taken together, these results suggest that if the binding conformation has significantly changed as suggested above, this does not explain the change in pH profile seen for the F194 single mutants. Perhaps the F194 mutation has changed the environment of one or more ionisable residues resulting in a shift in the pH profile of D-glucose activity, but the increased flexibility of the C383S mutation restored the environment of the residue(s) meaning pH is less important in the reaction.

In order to accommodate the bulky Trp at position 227, significant changes in a number of surrounding residues would need to occur for each of the possible rotamers. It is not possible to estimate which is the most likely rotamer from visual analysis of the structure as with the other mutations. Addition of the C383E mutation to F227W led to small enhancements in the D-arabinose, D-glucose and D-xylose activities compared to F227W alone. Changes in the relative activities towards the different substrates and corresponding methyl glycosides

compared to both the F227W and C383E single mutants implies that the two mutations are having combined effects on binding of the different substrates. However, the pH profile for D-glucose activity and the oxidative half reaction for D-xylose or D-arabinose activities both follow similar trends to C383E. Further analysis of this double mutant is required before the novel activities can be explained, most notably determination of the position of the tryptophan at position 227 within the structure by QM/ MM calculations and/ or solution of the crystal structure.

When combined with the C383E mutation, N245H and N245R result in a significant loss of activity towards D-galactose, which is surprising as all three single mutants display only moderate reduction in activity. The relative activities with the methyl D-galactopyranosides show similar trends to WT and each of the single mutants, however substrate binding does appear to be improved as K_M values for both single mutants are significantly reduced. Perhaps this reveals an increased affinity of the enzyme for D-galactose meaning catalysis is limited by slow release of the aldehyde product. Both double mutants show enhanced activity towards D-xylose compared to WT or the N245 single mutants although relative activity towards the methyl D-xylopyranosides does not reveal any large differences. The N245H/ C383E variant displays an increased sensitivity to oxygen concentration, compared to C383E or WT, implying that the combined mutations have an effect on the oxidative half reaction. Although it is possible that this effect is as a result of the higher turnover rate of this variant towards D-xylose compared to WT GO-N6M1. The two mutations appear to have effects on both D-xylose binding and catalysis making it difficult to deduce the precise structural reasons behind the change in activity. The ^{1}H-NMR data reveals the possibility of oxidation of methyl β-D-xylopyranoside at multiple positions. This seems more likely for D-xylose due to the presence of three equatorial hydroxyl groups than for D-arabinose which contains two equatorial and one axial hydroxyl group. The different shape of the two molecules may affect the relative accessibility of the different hydroxyls for oxidation (Figure 5.39).
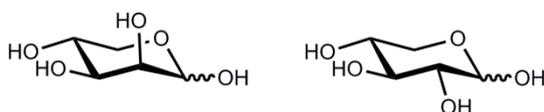
Figure 5.39. Structures of D-arabinose (left) and D-xylose (right)

The enhanced D-glucose activity of N245R/ C383E follows a similar trend to that seen with D-xylose. However, with D-arabinose the variant shows much greater activity for methyl α-D-arabinofuranoside than the β-pyranoside, as also seen with the C383E single mutant. This implies that the C383E mutation plays a significant role in binding of D-arabinose in the active site. It is interesting that N245R/ C383E shows a preference for the α-furanoside over the β-pyranoside, unlike F194I/ C383S and F227W/ C383E, as all three variants oxidise methyl β-D-arabinopyranoside at the C-4 position. It seems likely that binding of methyl α-D-arabinofuranoside may be different in N245R/ C383E than the other two variants showing enhanced D-arabinose activity.

Addition of the C383S mutation to N245R or N245W has essentially no effect on D-galactose activity compared to the N245X single mutants, apart from a small reduction in K_M as seen for the C383S single mutant. However, D-xylose activity is enhanced more than either of the single mutants, compared to WT, implying a synergistic effect of the two mutations, *i.e.* one mutation enhances the effect of the other. N245R/ C383S shows the highest level of activity towards 1 M D-xylose of all variants identified, while N245W/ C383S is interesting because the N245W single mutant displays such low levels of D-xylose activity. All single and double mutants show a moderate preference for methyl β-D-xylopyranoside over the α-anomer and the double mutants both display increased sensitivity to oxygen concentration. While the enhancement in D-xylose activity in these two variants is probably due to different features, given the different characteristics of arginine and tryptophan, it is only possible to estimate how either mutation brings about the observed effects with additional structural data.

Of all the variants identified, only the F194W/ W290F variant displays higher activity with an alternative substrate than with D-galactose. It is unfortunate that further analysis of this activity was not possible within the time frame of the project as this variant likely has the greatest potential for further biotechnological development. Of particular interest is identification of the product generated upon oxidation of glycerol. This would be a priority if the project were to be continued. It was not possible to estimate the positions of the residues upon mutation, however their relative positions in the WT structure are shown in Figure 5.40.
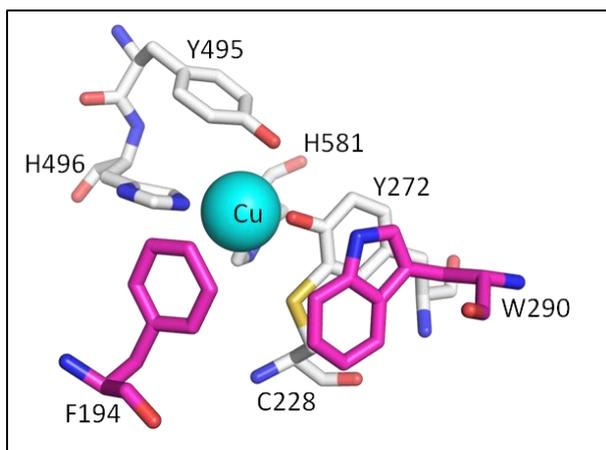
Figure 5.40. Positions of F194 and W290 in relation to key active site residues

F194 and W290 are shown in magenta, other residues are shown in white, copper is shown as a cyan sphere. Figure created with PyMOL using .pdb file 1GOF.

The non-hyperbolic plots of reaction velocity versus substrate concentration seen for most of the mutants with the alternative substrates (Section 5.4.2) could have a number of different meanings. The linear plots could simply indicate that the enzyme displays a K_M for the substrate which is much higher than the substrate concentrations that activity was measured at. In these experiments however, very high substrate concentrations of up to 3 M were used so if the K_M is above this value, the enzyme must have extremely low affinity for the substrate. Sigmoidal velocity curves can often be attributed to cooperativity between multiple binding sites on an enzyme, a relatively common occurrence with multimeric enzymes such as pyruvate dehydrogenase (Bisswanger, 1984). However, as GO is a monomeric protein, cooperativity seems unlikely. If cooperativity was occurring with the alternative substrates it could possibly involve binding at the postulated binding site in domain 1 (Section 1.6.1.1). Removal of this domain has been shown previously to have no effect on catalysis for WT GO with galactose, but that does not rule out a role in the oxidation of alternate substrates (Mahmoud et al., 2000). Another possibility for the non-hyperbolic dependence of reaction velocity on substrate concentration includes a change in the reaction mechanism from bi bi ping pong (Section 1.6.3.2) to a random bi bi reaction mechanism which can result in a variety of different velocity curves depending on a number of different factors (Segel, 1975). Various hybrid mechanisms, such as a ping pong-ordered mechanism can also lead to generation of non-hyperbolic plots (Segel, 1975). It has been previously shown that substitution of as little as one residue can alter an enzyme mechanism (Berry et al., 1989). However, the most likely

explanation for the non-hyperbolic relationship requires recognition that the reaction catalysed by GO is a two-substrate reaction (galactose and oxygen). In Section 5.7 it was identified that many of the mutations have altered the sensitivity of the enzyme to oxygen so atmospheric oxygen concentrations (used in activity measurements in Section 5.4.2) will not be saturating in many cases. It therefore seems likely that if the same measurements were carried out under saturating oxygen concentrations, hyperbolic plots of reaction velocity versus substrate concentration would be seen for most of the mutants.

Unfortunately, in the time-scale of this project, it was not possible to fully characterise each new activity, even by the selection of methods shown in this chapter. The effect of varying oxygen concentration on enzyme activity and the determination of the oxidation position by NMR were only carried out for a subset of the variants. As the oxygen concentrations were not determined, it cannot be ruled out that the differences observed in sensitivity to oxygen are due to inaccuracies during the experimental procedure. Oxygen may also display different solubility in the different sugar solutions, which would complicate any comparisons between the different substrates. Determination of the position of oxidation is only valid for those variants studied. Even variants carrying similar mutations may oxidise the same substrate at a different position. Optimisation of the NMR procedure could, however, be useful in future studies to refine the selectivity of the enzyme as well as characterising and optimising the experiment over long time periods.

Conclusions

The nine variants identified in Chapter 4 were characterised using a variety of methods and some significant changes were identified. C383E and, to a lesser extent C383S, greatly enhanced the activity of GO towards D-arabinose, D-glucose and D-xylose due to alterations of the active site which appear to generally improve binding; D-arabinose binding seemed most significantly altered. Combination of C383E/ S with mutations of either F194, F227 or N245 led to reductions in activity towards D-galactose, but maintained the enhanced activity towards the other substrates compared to WT. Variants containing F194X mutations appeared to make different interactions with D-glucose and/ or D-arabinose and may bind the substrates in a different orientation from WT. Oxidation of methyl β-D-arabinopyranoside was shown to occur at the C-4 hydroxyl. The F227W mutation had significant effects on several aspects of catalysis, most notably the pH profile, but the altered activity is difficult to rationalise without structural data. Mutations of N245 may have altered the enzyme specificity by altering the surface charge around the active site, or by changing the environment of key active site residues. The most significant results were seen with the N245R/ C383S and N245W/ C383S variants which showed high levels of D-xylose activity due to synergistic effects between the two mutations. Finally, the F194W/ W290F variant displays higher activity towards glycerol than D-galactose across a range of substrate concentrations. This variant is probably the most important identified in this study in terms of development of GO for industrial applications although further characterisation of the new activity is required.

Chapter 6 : General Discussion

6.1 Comparison with similar studies by other groups

GO represents a promising target for enzyme engineering due to the existence of high throughput screening methods, which have already been used successfully; and a wealth of information on the roles of specific residues in substrate binding and catalysis. In this project, screening assays were further optimised (Chapter 3) and libraries were designed and generated incorporating mutations at *all* previously identified positions of interest in the first and second coordination sphere, excluding essential catalytic residues (Chapter 4).

It is interesting to compare the residues selected for mutagenesis in the study presented here with that of two similar projects from the Arnold and Turner groups. The Arnold group demonstrated a small level of improved activity towards D-glucose by mutagenesis at positions 290, 330 and 406 (194 and 464 were also targeted but did not improve activity) (Sun et al., 2002); while the Turner group constructed a series of CAST libraries (Section 1.3.6) targeting residues surrounding the active site to improve activity towards D-glucose, D-mannose and D-N-acetyl glucosamine (Rannes et al., 2011). The residues selected in the two studies are shown in Table 6.1. All of the residues which led to altered activity upon mutagenesis by the other groups were included in the project described here, although not all of the relevant libraries were screened to >95% completion. This may be why the same variants were not identified. The libraries designed in Chapter 4 also did not contain the same combinations of residues as some of the successful combinations identified by Turner and co-workers. This is something to consider if the randomised oligonucleotides generated here are to be used in future library generation.

Table 6.1. The amino acid positions targeted in studies aiming to alter the substrate specificity of GO

Residue	This project	Sun et. al. (2002)	Rannes et. al. (2011)
F194	✓	✓	✓
G195			✓
F227	✓		
N245	✓		
W290	✓	✓	✓
S291			✓
Q326	✓		✓
Y329	✓		✓
R330	✓	✓	✓
N333	✓		
C383	✓		
Y405	✓		✓
Q406	✓	✓	✓
E407			✓
P463	✓		✓
F464	✓	✓	✓
V494	✓		
L514	✓		
C515	✓		
C518	✓		

Mutations shown in red and green were randomised in this project and alterations in substrate specificity identified. Those in red were further characterised while those in green were not. Mutations shown in yellow are those variants identified and characterised in other projects and shown to lead to altered substrate specificity.

The W290F/ R330K/ Q406T variant was reported to show 100-fold increased activity towards 250 mM D-glucose compared to the M1 variant (Sun et al., 2002) although, when reported by the Turner group, the increase in activity was only four-fold when 50 mM D-glucose was used (Rannes et al., 2011). This demonstrates the importance of the substrate concentration used in screening and characterisation of variants as discussed below. The specific activities in this project were measured using higher concentrations of substrate than the other studies, so only limited comparisons can be carried out. The increases in D-glucose activity compared to WT GO-N6M1 of two- to five-fold are within a similar range to those reported by the Turner group, who identified variants displaying increases of up to four-fold compared to the parent enzyme (Rannes et al., 2011). The mutations introduced by the Turner group only had a limited effect on activity towards D-galactose. However, as most libraries were constructed using the W290F/ R330K/ Q406T variant as the parent sequence, the D-galactose activity was already

reduced by around 70-fold compared to the M1 variant. This resulted in a much greater shift in specificity away from D-galactose and towards D-glucose than for the variants reported here. It would be interesting to introduce these three mutations into the variants identified in the study presented here. Arnold and co-workers reported a reduction in activity towards 250 mM D-galactose of 1000-fold compared to native WT GO for the W290F/ R330K/ Q406T variant (Sun et al., 2002). It is interesting to note that all previously reported variants showing significant shifts in specificity contain the W290F mutation. It would seem that this mutation leads to a loss in activity towards D-galactose but has neutral or positive effects on activity towards other substrates, as seen here with the glycerol oxidase F194W/ W290F. This is likely due to altered hydrogen bonding interactions with substrates but further study would be required to confirm this.

The increase in activity towards D-xylose is much greater than the increases seen towards D-glucose, perhaps because GO-N6M1 already shows a significant level of activity towards D-xylose so less significant changes to the enzyme were required to improve activity. The 20-fold enhancement in D-xylose activity upon introduction of the N245R/ C383S mutations is greater than the enhancement in activity towards any substrate reported by Turner and co-workers (activity towards D-xylose and D-arabinose was not investigated by Turner and co-workers). The greatest increase in activity towards D-arabinose was for the C383E mutant which is six-fold more active than GO-N6M1. Again, this is greater than any increase observed with D-glucose, perhaps because of the already higher level of activity of the parent enzyme.

While shifts in specificity away from D-galactose and towards the alternative substrate(s) have been achieved, it was unfortunately not possible to introduce mutations which reduced D-galactose activity to such an extent that the variant displayed higher activity for D-arabinose, D-glucose or D-xylose than for D-galactose. This has not been reported by any groups working on GO and is probably due to the fact that GO displays a very high level of activity for D-galactose (the k_{cat} is around 1000 s^{-1}). The enzyme also displays a relatively low specificity for the D-galactose substrate, implied by a high K_M of 50-70 mM, as well as broad substrate specificity. As D-galactose is not tightly or very specifically bound within the active site, it is difficult to significantly disrupt this binding. The overall rigidity of the structure may mean that introduction of mutations only has a very limited effect on the shape of the binding site, as seen with the crystal structure of the C383S (Wilkinson et al., 2004) and W290F/ R330K/ Q406T/ P463I (Rannes et al., 2011) variants. Removal of the D-galactose activity is not an essential requirement for development of GO for industrial uses. This would only be required

where substrate is provided in a mixture including D-galactose and pure product is required (which is unlikely). In fact, enzymes displaying broad substrate specificity are often desirable in the chemical synthesis industry as they can be used in various different transformations, removing the need for development of multiple enzymes (Breuer et al., 2004).

No characterisation has been previously reported for the oxidation of D-arabinose, D-xylose or glycerol by GO. However, ^{13}C- and ^{1}H-NMR on the products of GO catalysed reactions has shown that variants active towards D-glucose maintain the WT regioselectivity for the C-6 hydroxyl (Sun et al., 2002, Rannes et al., 2011). This matches the observation in this project with variants showing enhanced activity towards D-glucose.

6.2 Evaluation of the library generation method

As discussed in Section 1.3.4, the use of trimer phosphoramidites in generation of randomised oligonucleotides has a number of advantages. However, the significant financial cost of this library generation method should be reflected in the generation of high quality libraries. In all variants sequenced, no stop codons were identified at the randomised positions and no sequence bias was detected. In order to fully evaluate the quality of the libraries, significantly more sequence analysis would be required, but the identification of 27 variants showing enhanced activity confirms that the method was successful. Unfortunately, the process of library generation using the high quality oligonucleotides was not optimal, as shown by a number of sequencing variants with frameshifts and random insertions and deletions (Section 4.4.2). However, generation of correct clones in >70% of cases (and in some cases >90%) is still a good achievement. The library generation process consisted of a number of steps (Figure 4.10), increasing the opportunity for errors to be introduced. Despite extremely high fidelity, it would appear that the Phusion and KOD polymerases used in the overlap extension and PCR stages of library generation may have introduced errors. This could be due to inclusion of contaminants in the reaction mix, such as those used in epPCR (Section 1.3.2), but could also be due to damage to the DNA during manipulations such as gel extraction and purification. It is likely that a reduction in the number of steps required for library generation would reduce the number of errors incorporated, although this would require further optimisation of the process. Alternative methods for introducing mutations at defined positions all involve use of oligonucleotides (Section 1.3.4). While it may reduce the overall cost of library generation, use of degenerate codons (NNK/ NDT/ '22-codon trick') or MAX codons would not prevent the

errors observed here and would likely lead to libraries with lower diversity. If further libraries are to be generated it would therefore make sense to further optimise the process to increase the chance of generating variants with the desired activities with minimal introduction of errors.

Another disadvantage of the library generation process was that it was not possible to target certain residues individually due to the close proximity of another selected residue which was therefore randomised on the same oligonucleotide. This is unfortunately a drawback of oligonucleotide-based mutagenesis and could only be avoided by generation of more oligonucleotides. Libraries would need to be designed before the oligonucleotides were generated as, for example, it would not be economical to generate oligonucleotides for each of the 15 possible combinations of Q326, Y329, R330 and N333 found on oligonucleotide QYRN326-333.

6.3 Selection of an appropriate substrate concentration

As mentioned in Section 6.1, similar studies carried out by other groups have used substrate concentrations of 50 mM (Rannes et al., 2011) or 250 mM (Sun et al., 2002) in screening libraries and characterising GO variants. This makes comparison of results difficult as the effect of increasing substrate concentrations is different for the different variants (as shown by attempts to determine kinetic parameters in Section 5.4.2). In this project, much higher substrate concentrations, up to 2 M, were used for microtitre plate-based screening and characterisation of variants due to the potential future use of variants in synthesis of different chemicals or pharmaceutical intermediates (Section 1.6.5.2) where the substrate is generally provided at high concentrations (Section 1.3.8) (Woodley, 2013). However, use of such high concentrations also has disadvantages:

- As the concentrations used are higher than the K_M for most, if not all substrates screened, it is not possible to quickly identify improvements in K_M as the concentration used is likely to be saturating in all cases.
- High substrate concentrations may result in enzyme inhibition. While it is perhaps desirable to exclude variants showing inhibition at high substrate concentrations, it may prevent identification of variants showing significant improvements in activity at lower concentrations which could be selected for further development and

optimisation. Higher substrate concentrations may also result in product inhibition effects.

- High concentrations of the sugar substrates used increases the viscosity of the assay mixtures which may affect diffusion and reduce the apparent rate. Variants may appear less active than they are. Viscosity effects could be reduced by altering the temperature or by continuous mixing.

More variants may have been identified if multiple screening rounds were carried out using different substrate concentrations. For example, the first round could identify variants with increased activity compared to WT at lower substrate concentrations, while later rounds could determine whether any of the selected variants also displayed improved activity at the higher concentrations likely required for industrial development of the enzyme. However, this would greatly increase the amount of time required to identify variants of interest.

As mentioned in Chapter 3, the substrate concentration used in screening was selected upon consideration of Arnold's law: 'you get what you screen for' (You and Arnold, 1996). High substrate concentrations were used in order to find variants active under these conditions. However, upon screening the mutant libraries with the two assays, some of the enhanced activities were not reproduced with the purified protein, suggesting that the observed enhancements in activity may have actually been due to enhanced expression levels. The agar plate- and microtitre plate-based assays had therefore not only been developed to identify higher levels of enzyme activity but also, unintentionally, to identify higher expression levels, as long as activity still remained. Fortunately, this was not a significant problem as only a small number of the selected variants did not show the expected activity enhancement. Development of a high-throughput screen to detect enhanced activity regardless of expression levels would need to involve some level of purification of the enzyme from the *E. coli* culture and quantification of the enzyme before measurement of the relative activity. Such an assay is likely to be more time-consuming and expensive than the assays used here.

In Chapter 4, substrate concentrations were selected based on concentrations which are saturating for the WT enzyme (Section 4.5.2). However, as seen with the data for the glycerol oxidase (F194W/ W290F, Section 5.9), activity is enhanced to different extents at different substrate concentrations. For example, the enhancements in D-glucose activity were much less significant with 1 M than with 2 M D-glucose due to different effects of substrate concentration on WT and on the variants (Sections 4.5.2 and 5.3). Ideally, an appropriate

substrate concentration for the potential downstream application(s) of the variants would have been determined at the beginning of the project and only this concentration used in screening, confirmation of screening data, and subsequent characterisation. If 1 M substrate had been used for the measurements presented in Chapter 4, it is possible that a different selection of variants may have been taken forward. The different behaviour of the variants towards different substrate concentrations may reflect substrate/ product inhibition in some cases. Future exploration of this would be essential if the variants were to be developed for use in the biotechnology industry.

6.4 Future directions

6.4.1 Continuation of the project

This project and the results displayed in this thesis could be improved in various ways. Further repeats of some of the measurements taken would increase the confidence in the conclusions drawn. For example the different activity measurements with the F194T single mutant are hard to interpret; repeating the data with freshly purified protein would rule out the possibility of contamination, for example with WT GO-N6M1. Within the timescale of the project, it was only possible to carry out the more comprehensive characterisations detailed in Chapter 5 with some of the variants. It would be very interesting to complete this characterisation for all nine of the variants selected. For example different variants may show different regioselectivity. This seems particularly likely with the variants showing enhanced D-glucose activity, as addition of the methyl group at position C-1 had a significant effect on activity for some variants and not for others (Table 5.6) implying differences in binding of D-glucose within the active site. Identification of the product(s) of glycerol oxidation would also be a priority as the potential products have numerous industrial applications (Section 1.6.5.2). In order to confirm the assignments of chemical shifts in the oxidation of methyl β-D-arabinopyranoside, the predicted product could be generated by a different route and the ^{1}H-NMR spectra compared. For example, chromium trioxide or ruthenium tetroxide could be used to oxidise methyl β-D-arabinopyranoside with protecting groups at the C-2 and C-3 hydroxyls; this would result in oxidation exclusively at the C-4 hydroxyl. Similar methods could also be used to aid in the assignments of chemical shifts for the other reactions.

The identification of changes in the oxidative half reaction in some variants by measuring the effect on activity of increased or decreased oxygen concentrations is relatively qualitative data

as the oxygen concentration was not measured and may have shown some variation between different measurements. More complete characterisation of the effect of mutations on the oxidative half reaction could be carried out by determining the full oxygen kinetics for each variant, where possible, by measuring activity over a range of defined oxygen concentrations while keeping the concentration of sugar substrate (D-galactose/ D-arabinose/ D-xylose/ D-glucose/ glycerol) constant.

Unfortunately diffraction quality crystals of *E. coli*-expressed GO were not obtained in this project. Frustratingly, the available structures of GO mutants (Wilkinson et al., 2004, Rannes et al., 2011) did not provide any clear explanations for the observed changes in activity in this study. If this work is continued however, crystal structures will prove very useful in explaining the changes in activity:

- The position of the introduced tryptophan residues at position 227 or 245 could not be modelled using COOT alone, as significant structural changes would be required to accommodate the large side chain. Crystal structures of F227X mutations would be particularly interesting to observe the effect on the copper environment and thioether bond which are very close to this residue in the structure.

- Where it has not been possible to determine the regioselectivity by ^{1}H-NMR, solution of the structure of an enzyme-substrate complex could provide this information, although no such structure has ever been published for GO, perhaps due to crystal packing constraints, but more likely due to the relatively low affinity of WT for all substrates. Variants displaying increased affinity for substrate may overcome this problem. In addition, the different space group observed in crystals of the *E. coli*-expressed GO (Rannes et al., 2011) compared to that previously reported for GO (Ito et al., 1991), may permit entry of substrate into the active site of GO within the crystal (the *E. coli*-expressed GO crystallised in space group $P6_522$ compared to space group C2 for previous structures (Ito et al., 1991)).

- While the C383S mutation did not result in any major structural changes in the crystal structure (Wilkinson et al., 2004), introduction of a glutamate at this position is a more significant change as the residue is larger than cysteine or serine and also carries a negative charge. Solution of the crystal structure of the C383E mutant may reveal structural differences which could help explain the activity and substrate specificity changes induced upon introduction of this mutation.

221

Modelling studies similar to (Wachter and Branchaud, 1996) could also be carried out instead of, or in addition to, crystallisation studies. These studies may prove very useful as there have been significant advances in computing resources for molecular modelling since 1996 - significant additional information, such as long range interactions, can now be predicted much more easily. Information on regioselectivity and changes in substrate specificity determined here could be used to guide such studies. Changes in the secondary structure of the enzyme, and in the active site environment, brought about by the mutations could be identified by circular dichroism and UV/ vis spectroscopy (Section 1.6.3.1).

Determining the mechanism by which the GO variants catalyse oxidation of the alternative substrates will require a significant amount of work, especially given the possibility that some mutations may have altered the catalytic mechanism (Sections 5.4.2 and 5.7). Determination of the mechanism would require several techniques, including the use of substrate analogues, analysis of the kinetic isotope effects (as carried out by (Whittaker et al., 1998)) or even following the reaction crystallographically, although such methods are still in relatively early stages of development (Pearson and Owen, 2009).

Purification and characterisation of GO family members, such as glyoxal oxidase (Section 1.6.1.3) and GO from other species (Section 5.2) would likely aid in the explanation of changes observed with the different mutations. It is also likely that family members may show different substrate specificities due to the different niches inhabited by the host organisms - this information could prove invaluable in design of future mutagenesis projects.

6.4.2 Other potential directions

As discussed earlier, the randomised oligonucleotides generated for this project can be used to generate a wide variety of further libraries. Screening could be carried out towards the same potential substrates used here, or also a range of others including different monosaccharides such as D-fructose, D-ribose or L-isomers of sugars; polysaccharides such as guar, already targeted by (Delagrave et al., 2001); or even biomarkers which could be developed for use in diagnostic assays. One of the original targets considered for this project was glycated albumin, a biomarker for hyperglycaemia in the blood of people suffering from diabetes. Glycation of serum proteins occurs under hyperglycaemic conditions and can alter protein function causing health problems such as atherosclerosis and retinopathy (Nathan, 1993). Detection of the level of glycation by mass spectrometry- and HPLC-based techniques is used in diagnosis of diabetes and in management of the condition (Barnaby et al., 2010, Frolov et al., 2006, Frolov and

Hoffmann, 2010). Development of a biosensor using a GO variant could provide a more rapid and cost-effective alternative.

Generation of variants by less random approaches such as addition of specific mutations (*i.e.* C383S or W290F) into existing variants or combination of advantageous mutations from different libraries by methods such as DNA shuffling, detailed in Section 1.3.5, may also improve the activity and specificity of GO towards different substrates. Many variants were identified upon initial screening of **Libraries E, L, M and N** (Section 4.4) which, following confirmation of the enhanced activities, could be combined with the variants which were characterised in Chapter 5. Design and screening of mutant libraries *in silico* followed by confirmation of results in the laboratory could also lead to increased efficiency in finding variants with the desired activities.

6.4.3 Potential future development of variants

Unfortunately, as also observed by the Turner group (Rannes et al., 2011), the broad specificity of the GO variants identified thus far precludes their use in specific labelling of cell surface glycans. However, following further development, all of the variants have the potential for use as biocatalysts in chemical synthetic processes, generation of complex molecules and process monitoring as detailed in Section 1.6.5.2. Further development of the variants would have to consider the process they would be used in and the conditions they would be exposed to. As detailed in Section 1.3.8 these may include high substrate and product concentrations, specific pH and temperatures, the use of organic solvents and heterogeneity in the reaction conditions (Woodley, 2013). Different areas of the molecule than the active site would likely be also targeted to improve the already high stability of GO at different temperatures and resistance to solvents (Kelly-Falcoz et al., 1965). Libraries may be designed following analysis of sequence alignments (Jochens et al., 2010) and/ or using the B-FIT method (Reetz et al., 2006a) as detailed in Section 1.3.6. Many of the potential problems associated with traditional reaction conditions, such as product inhibition, are likely easily overcome by use of appropriate reaction vessels, such as batch stirred tank reactors, continuous stirred tank reactors or continuous packed bed reactors (Woodley, 2013).

The pH optima of the variants can be determined by the targeted residues close to the active site, as shown in Section 5.6.3. Unfortunately it was only possible to measure the pH profile for activity towards D-glucose. However, all of the variants showing enhanced activity towards D-glucose (double mutants and C383S single mutants) showed an increase in pH optimum

compared to WT. GO has been shown to also catalyse oxidation of alcohols to form the carboxylic acid as well as the aldehyde (Kelleher and Bhavanandan, 1986, Parikka and Tenkanen, 2009) which may lead to acidification of the reaction medium within the reaction vessel. Therefore, ideally the engineered variants should show better resistance to acidic conditions, although the current activities at low pH may be acceptable. Parikka and Tenkanen (2009) demonstrated that oxidation of D-galactose by WT GO can lead to formation of multiple products (Figure 6.1) and demonstrated that reaction conditions can be optimised to reduce formation of unwanted by-products. This will be important in developing an efficient process for use on larger scales.
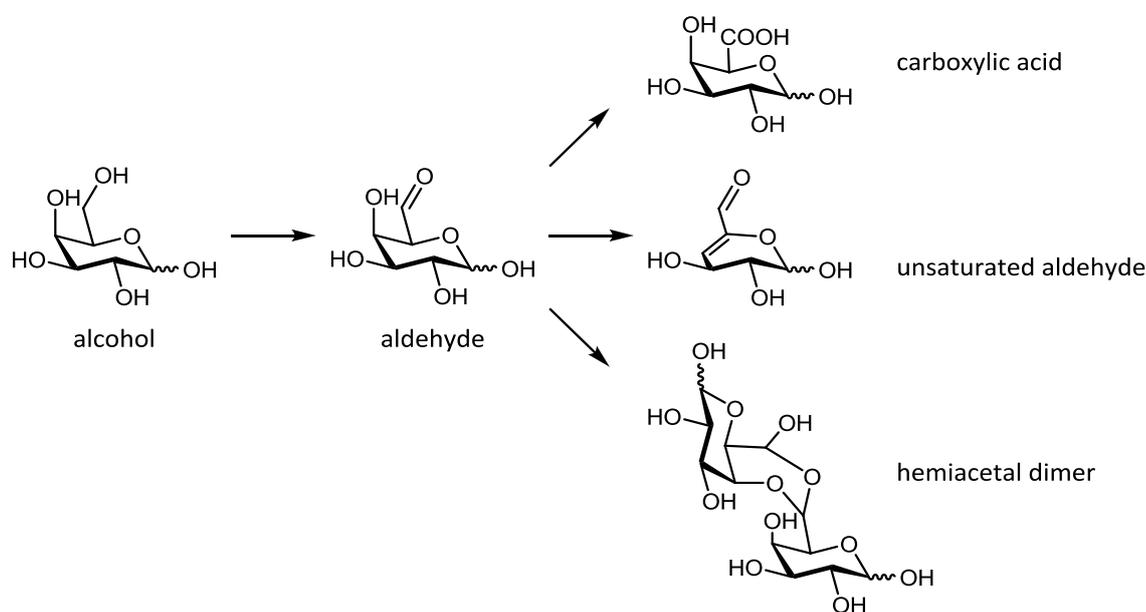
Figure 6.1. The different products which can be generated upon oxidation of D-galactose by GO

Copied from Parikka and Tenkanen (2009).

Another consideration in development of variants for industrial uses is the processivity, or residual rate of activity, of the enzyme, *i.e.* the number of times it will turnover before inactivation due to a number of possible factors. In the case of GO this is largely determined by the reaction conditions. For example, catalase is included here in the reaction mix to remove the hydrogen peroxide generated which has been shown to damage the enzyme, while the presence of an oxidising agent such as ferricyanide or HRP is essential to maintain GO in the

oxidised, active form. The relative quantities of these components may vary with different variants and substrates.

Conclusions

This work has contributed to the already significant area of research into the enzyme GO and further expanded and displayed its potential for biotechnological development. High throughput assays, already successfully used by other groups, have been improved by optimisation of a number of aspects. Mutant libraries were designed based on the wealth of information on the role of different residues within the GO structure and generated using high quality oligonucleotides to prevent the introduction of any sequence bias. The libraries were screened using the optimised assays and a number of variants identified with enhanced activity towards different substrates with potentially important applications in biotechnology. The most significant achievement is the identification of a variant displaying higher activity towards glycerol than towards D-galactose. Other noteworthy observations include the significant effect on glucose binding of mutations at position 194 and the regioselectivity for the C-4 hydroxyl of methyl β-D-arabinopyranoside displayed by some of the GO variants. The activities of the variants were characterised by a variety of methods to provide some understanding of the reaction catalysed as well as the optimal conditions under which the enzyme variants function. The variants, particularly the glycerol oxidase, show great promise for future development as valuable biocatalysts for the chemical synthesis and pharmaceutical industries and as tools in process monitoring.

Chapter 7 References

ABRAHAMSON, M. J., VAZQUEZ-FIGUEROA, E., WOODALL, N. B., MOORE, J. C. & BOMMARIUS, A. S. 2012. Development of an Amine Dehydrogenase for Synthesis of Chiral Amines. *Angewandte Chemie-International Edition,* 51**,** 3969-3972.

ADANYI, N., SZABO, E. E. & VARADI, M. 1999. Multi-enzyme biosensors with amperometric detection for determination of lactose in milk and dairy products. *European Food Research and Technology,* 209**,** 220-226.

AGUDO, R., ROIBAN, G.-D. & REETZ, M. T. 2012. Achieving Regio- and Enantioselectivity of P450-Catalyzed Oxidative CH Activation of Small Functionalized Molecules by Structure-Guided Directed Evolution. *Chembiochem,* 13**,** 1465-1473.

AHARONI, A., AMITAI, G., BERNATH, K., MAGDASSI, S. & TAWFIK, D. S. 2005. High-throughput screening of enzyme libraries: Thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chemistry & Biology,* 12**,** 1281-1289.

ALEXEEVA, M., ENRIGHT, A., DAWSON, M. J., MAHMOUDIAN, M. & TURNER, N. J. 2002. Deracemization of alpha-methylbenzylamine using an enzyme obtained by in vitro evolution. *Angewandte Chemie-International Edition,* 41**,** 3177-3180.

AMARAL, D., KELLY-FALCOZ, F. & HORECKER, B. L. 1966. Galactose Oxidase of Polyporus circinatus. *Methods in Enzymology,* 9**,** 87-92.

AMIN, N., LIU, A. D., RAMER, S., AEHLE, W., MEIJER, D., METIN, M., WONG, S., GUALFETTI, P. & SCHELLENBERGER, V. 2004. Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Engineering Design & Selection,* 17**,** 787-793.

ANTHONY, L. C., SUZUKI, H. & FILUTOWICZ, M. 2004. Tightly regulated vectors for the cloning and expression of toxic genes. *Journal of Microbiological Methods,* 58**,** 243-250.

ARNOLD, F. H. 1998. When blind is better: Protein design by evolution. *Nature Biotechnology,* 16**,** 617-618.

ATSUMI, S., CANN, A. F., CONNOR, M. R., SHEN, C. R., SMITH, K. M., BRYNILDSEN, M. P., CHOU, K. J. Y., HANAI, T. & LIAO, J. C. 2008. Metabolic engineering of Escherichia coli for 1-butanol production. *Metabolic Engineering,* 10**,** 305-311.

AVIGAD, G. 1985. Oxidation Rates Of Some Desialylated Glycoproteins By Galactose-Oxidase. *Archives of Biochemistry and Biophysics,* 239**,** 531-537.

AVIGAD, G., ASENSIO, C., HORECKER, B. L. & AMARAL, D. 1962. D-galactose oxidase of Polyporus circinatus. *Journal of Biological Chemistry,* 237**,** 2736-2743.

BABCOCK, G. T., ELDEEB, M. K., SANDUSKY, P. O., WHITTAKER, M. M. & WHITTAKER, J. W. 1992. Electron-Paramagnetic Resonance And Electron Nuclear Double-Resonance Spectroscopies Of The Radical Site In Galactose-Oxidase And Of Thioether-Substituted Phenol Model Compounds. *Journal of the American Chemical Society,* 114**,** 3727-3734.

BANERJEE, R. & RAGSDALE, S. W. 2003. The many faces of vitamin B-12: Catalysis by cobalamin-dependent enzymes. *Annual Review of Biochemistry,* 72**,** 209-247.

BARBIER, J., GERTH, K., JANSEN, R. & KIRSCHNING, A. 2012a. Total synthesis of noricumazole B establishes D-arabinose as glycan unit. *Organic & Biomolecular Chemistry,* 10**,** 8298-8307.

BARBIER, J., WEGNER, J., BENSON, S., GENTZSCH, J., PIETSCHMANN, T. & KIRSCHNING, A. 2012b. Total Synthesis of a Noricumazole A Library and Evaluation of HCV Inhibition. *Chemistry-a European Journal,* 18**,** 9083-9090.

BARNABY, O. S., WA, C. L., CERNY, R. L., CLARKE, W. & HAGE, D. S. 2010. Quantitative analysis of glycation sites on human serum albumin using 16O/18O-labeling and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Chim. Acta,* 411**,** 1102-1110.

BARON, A. J., STEVENS, C., WILMOT, C., SENEVIRATNE, K. D., BLAKELEY, V., DOOLEY, D. M., PHILLIPS, S. E. V., KNOWLES, P. F. & MCPHERSON, M. J. 1994. Structure and

mechanism of galactose-oxidase - The free-radical site. *Journal of Biological Chemistry,* 269**,** 25095-25105.

BECKER, S., HOBENREICH, H., VOGEL, A., KNORR, J., WILHELM, S., ROSENAU, F., JAEGER, K. E., REETZ, M. T. & KOLMAR, H. 2008. Single-cell high-throughput screening to identify enantioselective hydrolytic enzymes. *Angewandte Chemie-International Edition,* 47**,** 5085-5088.

BECKER, S., MICHALCZYK, A., WILHELM, S., JAEGER, K.-E. & KOLMAR, H. 2007. Ultrahigh-throughput screening to identify E-coli cells expressing functionally active enzymes on their surface. *Chembiochem,* 8**,** 943-949.

BEN-DAVID, A., SHOHAM, G. & SHOHAM, Y. 2008. A universal screening assay for glycosynthases: Directed evolution of glycosynthase XynB2(E335G) suggests a general path to enhance activity. *Chemistry & Biology,* 15**,** 546-551.

BERKMEN, M. 2012. Production of disulfide-bonded proteins in Escherichia coli. *Protein Expression and Purification,* 82**,** 240-251.

BERRY, A., SCRUTTON, N. S. & PERHAM, R. N. 1989. Switching Kinetic Mechanism And Putative Proton Donor By Directed Mutagenesis Of Glutathione-Reductase. *Biochemistry,* 28**,** 1264-1269.

BERZINS, T., WIKEN, M., HELLSTROM, U. & PERLMANN, P. 1983. The Use Of Supernatants From Neuraminidase And Galactose-Oxidase (Nago)-Treated Lymphocytes As A Source Of T-Cell Growth-Factor. *Journal of Immunological Methods,* 63**,** 309-319.

BISSWANGER, H. 1984. Cooperativity In Highly Aggregated Enzyme-Systems - A Slow Transition Model For The Pyruvate-Dehydrogenase Complex From Escherichia-Coli. *Journal of Biological Chemistry,* 259**,** 2457-2465.

BOERSMA, Y. L., DROGE, M. J., VAN DER SLOOT, A. M., PIJNING, T., COOL, R. H., DIJKSTRA, B. W. & QUAX, W. J. 2008. A novel genetic selection system for improved enantioselectivity of Bacillus subtilis lipase A. *Chembiochem,* 9**,** 1110-1115.

BONOMO, R. A., DAWES, C. G., KNOX, J. R. & SHLAES, D. M. 1995. Beta-Lactamase Mutations Far From The Active-Site Influence Inhibitor Binding. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology,* 1247**,** 121-125.

BORK, P. & DOOLITTLE, R. F. 1994. Drosophila Kelch Motif Is Derived From A Common Enzyme Fold. *Journal of Molecular Biology,* 236**,** 1277-1282.

BORNSCHEUER, U. T., HUISMAN, G. W., KAZLAUSKAS, R. J., LUTZ, S., MOORE, J. C. & ROBINS, K. 2012. Engineering the third wave of biocatalysis. *Nature,* 485**,** 185-194.

BOSLEY, A. D. & OSTERMEIER, M. 2005. Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomolecular Engineering,* 22**,** 57-61.

BRAVO, J., FITA, I., FERRER, J. C., ENS, W., HILLAR, A., SWITALA, J. & LOEWEN, P. C. 1997. Identification of a novel bond between a histidine and the essential tyrosine in catalase HPII of Escherichia coli. *Protein Science,* 6**,** 1016-1023.

BRETTING, H. & JACOBS, G. 1987. The Reactivity Of Galactose-Oxidase With Snail Galactans, Galactosides And D-Galactose-Composed Oligosaccharides. *Biochimica Et Biophysica Acta,* 913**,** 342-348.

BREUER, M., DITRICH, K., HABICHER, T., HAUER, B., KESSELER, M., STURMER, R. & ZELINSKI, T. 2004. Industrial methods for the production of optically active intermediates. *Angewandte Chemie-International Edition,* 43**,** 788-824.

BRIK, A., WU, C. Y. & WONG, C. H. 2006. Microtiter plate based chemistry and in situ screening: a useful approach for rapid inhibitor discovery. *Organic & Biomolecular Chemistry,* 4**,** 1446-1457.

BRIM, H., MCFARLAN, S. C., FREDRICKSON, J. K., MINTON, K. W., ZHAI, M., WACKETT, L. P. & DALY, M. J. 2000. Engineering Deinococcus radiodurans for metal remediation in radioactive mixed waste environments. *Nature Biotechnology,* 18**,** 85-90.

BROWN, W. C. & CAMPBELL, J. L. 1993. A New Cloning Vector And Expression Strategy For Genes Encoding Proteins Toxic To Escherichia-Coli. *Gene,* 127**,** 99-103.

CARTER, J. H., DEDDENS, J. A., PULLMAN, J. L., COLLIGAN, B. M., WHITELEY, L. O. & CARTER, H. W. 1997. Validation of the galactose oxidase Schiff's reagent sequence for early detection and prognosis in human colorectal adenocarcinoma. *Clinical Cancer Research,* 3**,** 1479-1489.

CEVIK, E., SENEL, M. & ABASIYANIK, M. F. 2010. Construction of biosensor for determination of galactose with galactose oxidase immobilized on polymeric mediator contains ferrocene. *Current Applied Physics,* 10**,** 1313-1316.

CHAGPAR, A., EVELEGH, M., FRITSCHE, H. A., KRISHNAMURTHY, S., HUNT, K. K. & KUERER, H. M. 2004. Prospective evaluation of a novel approach for the use of a quantitative galactose oxidase-Schiff reaction in ductal fluid samples from women with breast carcinoma. *Cancer,* 100**,** 2549-2554.

CHARMANTRAY, F., TOUISNI, N., HECQUET, L. & MOUSTY, C. 2013. Amperometric Biosensor Based on Galactose Oxidase Immobilized in Clay Matrix. *Electroanalysis,* 25**,** 630-635.

CHEON, Y. H., PARK, H. S., KIM, J. H., KIM, Y. & KIM, H. S. 2004. Manipulation of the active site loops of D-hydantoinase, a (beta/alpha)(8)-barrel protein, for modulation of the substrate specificity. *Biochemistry,* 43**,** 7413-7420.

CIRINO, P. C., MAYER, K. M. & UMENO, D. 2003. Generating Mutant Libraries Using Error-Prone PCR. *In:* ARNOLD, F. H. & GEORGIOU, G. (eds.) *Directed Evolution Library Creation: Methods and Protocols.* Totowa: Humana Press Inc.

CLEVELAND, L., COFFMAN, R. E., COON, P. & DAVIS, L. 1975. Investigation Of Role Of Copper In Galactose Oxidase. *Biochemistry,* 14**,** 1108-1115.

COCO, W. M., ENCELL, L. P., LEVINSON, W. E., CRIST, M. J., LOOMIS, A. K., LICATO, L. L., ARENSDORF, J. J., SICA, N., PIENKOS, P. T. & MONTICELLO, D. J. 2002. Growth factor engineering by degenerate homoduplex gene family recombination. *Nature Biotechnology,* 20**,** 1246-1250.

COLLINS, P. & FERRIER, R. 1995. *Monosaccharides. Their Chemistry and Their Roles in Natural Products*, John Wiley & Sons Ltd.

CONNOLLY, M. L. 1983. Analytical Molecular-Surface Calculation. *Journal of Applied Crystallography,* 16**,** 548-558.

COOPER, J. A. D., SMITH, W., BACILA, M. & MEDINA, H. 1959. Galactose oxidase from Polyporus circinatus, FR. *Journal of Biological Chemistry,* 234**,** 445-448.

CRITCHLEY, D. R. 1974. Cell-Surface Proteins Of Nil1 Hamster Fibroblasts Labeled By A Galactose Oxidase, Tritiated Borohydride Method. *Cell,* 3**,** 121-125.

DAVIS, W. R., TOMSHO, J., NIKAM, S., COOK, E. M., SOMAND, D. & PELISKA, J. A. 2000. Inhibition of HIV-1 reverse transcriptase-catalyzed DNA strand transfer reactions by 4-chlorophenylhydrazone of mesoxalic acid. *Biochemistry,* 39**,** 14279-14291.

DE BOER, E., VAN KOOYK, Y., TROMP, M. G. M., PLAT, H. & WEVER, R. 1986. Bromoperoxidase From Ascophyllum-Nodosum - A Novel Class Of Enzymes Containing Vanadium As A Prosthetic Group. *Biochimica Et Biophysica Acta,* 869**,** 48-53.

DEACON, S. E. 2008. *Protein Expression and Engineering of Galactose Oxidase.* PhD thesis, University of Leeds.

DEACON, S. E., MAHMOUD, K., SPOONER, R. K., FIRBANK, S. J., KNOWLES, P. F., PHILLIPS, S. E. V. & MCPHERSON, M. J. 2004. Enhanced fructose oxidase activity in a galactose oxidase variant. *Chembiochem,* 5**,** 972-979.

DEACON, S. E. & MCPHERSON, M. J. 2011. Enhanced Expression and Purification of Fungal Galactose Oxidase in Escherichia coli and Use for Analysis of a Saturation Mutagenesis Library. *Chembiochem,* 12**,** 593-601.

DEHOUCK, Y., KWASIGROCH, J. M., GILIS, D. & ROOMAN, M. 2011. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *Bmc Bioinformatics,* 12**,** 151-162.

DELAGRAVE, S., MURPHY, D. J., PRUSS, J. L. R., MAFFIA, A. M., MARRS, B. L., BYLINA, E. J., COLEMAN, W. J., GREK, C. L., DILWORTH, M. R., YANG, M. M. & YOUVAN, D. C. 2001. Application of a very high-throughput digital imaging screen to evolve the enzyme galactose oxidase. *Protein Engineering,* 14**,** 261-267.

DEMIREL-GULEN, S., LUCAS, M. & CLAUS, P. 2005. Liquid phase oxidation of glycerol over carbon supported gold catalysts. *Catalysis Today,* 102**,** 166-172.

DICKS, J. M., ASTON, W. J., DAVIS, G. & TURNER, A. P. F. 1986. Mediated Amperometric Biosensors For D-Galactose, Glycolate And L-Amino-Acids Based On A Ferrocene-Modified Carbon Paste Electrode. *Analytica Chimica Acta,* 182**,** 103-112.

DISNEY, M. D. & SEEBERGER, P. H. 2004. The use of carbohydrate microarrays to study carbohydrate-cell interactions and to detect pathogens. *Chemistry & Biology,* 11**,** 1701-1707.

DOORES, K. J., GAMBLIN, D. P. & DAVIS, B. G. 2006. Exploring and exploiting the therapeutic potential of glycoconjugates. *Chemistry-a European Journal,* 12**,** 656-665.

DUBOIS, J. L. & KLINMAN, J. P. 2005. Mechanism of post-translational quinone formation in copper amine oxidases and its relationship to the catalytic turnover. *Archives of Biochemistry and Biophysics,* 433**,** 255-265.

ELLIS, K. J. & MORRISON, J. F. 1982. Buffers Of Constant Ionic-Strength For Studying Ph-Dependent Processes. *Methods in Enzymology,* 87**,** 405-426.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography,* 66**,** 486-501.

ESCALETTES, F. & TURNER, N. J. 2008. Directed evolution of galactose oxidase: Generation of enantioselective secondary alcohol oxidases. *Chembiochem,* 9**,** 857-860.

ESTELL, D. A., GRAYCAR, T. P. & WELLS, J. A. 1985. Engineering An Enzyme By Site-Directed Mutagenesis To Be Resistant To Chemical Oxidation. *Journal of Biological Chemistry,* 260**,** 6518-6521.

EVNIN, L. B., VASQUEZ, J. R. & CRAIK, C. S. 1990. Substrate-Specificity Of Trypsin Investigated By Using A Genetic Selection. *Proceedings of the National Academy of Sciences of the United States of America,* 87**,** 6659-6663.

FIRBANK, S. J., ROGERS, M., HURTADO-GUERRERO, R., DOOLEY, D. M., HALCROW, M. A., PHILLIPS, S. E. V., KNOWLES, P. F. & MCPHERSON, M. J. 2003. Cofactor processing in galactose oxidase. *Biochemical Society Transactions,* 31**,** 506-509.

FIRBANK, S. J., ROGERS, M. S., WILMOT, C. M., DOOLEY, D. M., HALCROW, M. A., KNOWLES, P. F., MCPHERSON, M. J. & PHILLIPS, S. E. V. 2001. Crystal structure of the precursor of galactose oxidase: An unusual self-processing enzyme. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 12932-12937.

FORSSKAHL, I., POPOFF, T. & THEANDER, O. 1976. Reactions Of D-Xylose And D-Glucose In Alkaline, Aqueous-Solutions. *Carbohydrate Research,* 48**,** 13-21.

FORSTNER, M., MULLER, A., ROGNAN, D., KRIECHBAUM, M. & WALLIMANN, T. 1998. Mutation of cis-proline 207 in mitochondrial creatine kinase to alanine leads to increased acid stability. *Protein Engineering,* 11**,** 563-568.

FOX, R. J., DAVIS, S. C., MUNDORFF, E. C., NEWMAN, L. M., GAVRILOVIC, V., MA, S. K., CHUNG, L. M., CHING, C., TAM, S., MULEY, S., GRATE, J., GRUBER, J., WHITMAN, J. C., SHELDON, R. A. & HUISMAN, G. W. 2007. Improving catalytic function by ProSAR-driven enzyme evolution. *Nature Biotechnology,* 25**,** 338-344.

FROLOV, A., HOFFMANN, P. & HOFFMANN, R. 2006. Fragmentation behavior of glycated peptides derived from D-glucose, D-fructose and D-ribose in tandem mass spectrometry. *J. Mass Spectrom.,* 41**,** 1459-1469.

FROLOV, A. & HOFFMANN, R. 2010. Identification and relative quantification of specific glycation sites in human serum albumin. *Anal. Bioanal. Chem.,* 397**,** 2349-2356.

FUNKE, S. A., EIPPER, A., REETZ, M. T., OTTE, N., THIEL, W., VAN POUDEROYEN, G., DIJKSTRA, B. W., JAEGER, K. E. & EGGERT, T. 2003. Directed evolution of an enantioselective Bacillus subtilis lipase. *Biocatalysis and Biotransformation,* 21**,** 67-73.

GALLOWAY, J. M. & STANILAND, S. S. 2012. Protein and peptide biotemplated metal and metal oxide nanoparticles and their patterning onto surfaces. *Journal of Materials Chemistry,* 22**,** 12423-12434.

GARCIA, R., BESSON, M. & GALLEZOT, P. 1995. Chemoselective Catalytic-Oxidation Of Glycerol With Air On Platinum Metals. *Applied Catalysis a-General,* 127**,** 165-176.

GATTEGNO, L., PERRET, G., FABIA, F., BLADIER, D. & CORNILLOT, P. 1981. In vivo Aging Of Human-Erythrocytes And Cell-Surface Labeling By D-Galactose Oxidase And Sodium Borotritide. *Carbohydrate Research,* 95**,** 283-289.

GAYTAN, P., YANEZ, J., SANCHEZ, F., MACKIE, H. & SOBERON, X. 1998. Combination of DMT-mononucleotide and Fmoc-trinucleotide phosphoramidites in oligonucleotide synthesis affords an automatable codon-level mutagenesis method. *Chemistry & Biology,* 5**,** 519-527.

GERFEN, G. J., BELLEW, B. F., GRIFFIN, R. G., SINGEL, D. J., EKBERG, C. A. & WHITTAKER, J. W. 1996. High-frequency electron paramagnetic resonance spectroscopy of the apogalactose oxidase radical. *Journal of Physical Chemistry,* 100**,** 16739-16748.

GLASCOCK, C. B. & WEICKERT, M. J. 1998. Using chromosomal lacI(Q1) to control expression of genes on high-copy-number plasmids in Escherichia coli. *Gene,* 223**,** 221-231.

GORL, J., TIMM, M. & SEIBEL, J. 2012. Mechanism-Oriented Redesign of an Isomaltulose Synthase to an Isomelezitose Synthase by Site-Directed Mutagenesis. *Chembiochem,* 13**,** 149-156.

GRIFFITHS, J. S., CHERIYAN, M., CORBELL, J. B., POCIVAVSEK, L., FIERKE, C. A. & TOONE, E. J. 2004. A bacterial selection for the directed evolution of pyruvate aldolases. *Bioorganic & Medicinal Chemistry,* 12**,** 4067-4074.

GROEGEL, D. B. M., LINK, M., DUERKOP, A. & WOLFBEIS, O. S. 2011. A New Fluorescent PET Probe for Hydrogen Peroxide and its Use in Enzymatic Assays for L-Lactate and D-Glucose. *Chembiochem,* 12**,** 2779-2785.

HAMILTON, G. A., ADOLF, P. K., DE JERSEY, J., DUBOIS, G. C., DYRKACZ, G. R. & LIBBY, R. D. 1978. Trivalent Copper, Superoxide, And Galactose Oxidase. *Journal of the American Chemical Society,* 100**,** 1899-1912.

HAN, E., DING, L., QIAN, R. C., BAO, L. & JU, H. X. 2012. Sensitive Chemiluminescent Imaging for Chemoselective Analysis of Glycan Expression on Living Cells Using a Multifunctional Nanoprobe. *Analytical Chemistry,* 84**,** 1452-1458.

HANAI, T., ATSUMI, S. & LIAO, J. C. 2007. Engineered synthetic pathway for isopropanol production in Escherichia coli. *Applied and Environmental Microbiology,* 73**,** 7814-7818.

HANSEN, K. B., YI, H., XU, F., RIVERA, N., CLAUSEN, A., KUBRYK, M., KRSKA, S., ROSNER, T., SIMMONS, B., BALSELLS, J., IKEMOTO, N., SUN, Y., SPINDLER, F., MALAN, C., GRABOWSKI, E. J. J. & ARMSTRONG, J. D., III 2009. Highly Efficient Asymmetric Synthesis of Sitagliptin. *Journal of the American Chemical Society,* 131**,** 8798-8804.

HARMS, H., SCHLOSSER, D. & WICK, L. Y. 2011. Untapped potential: exploiting fungi in bioremediation of hazardous chemicals. *Nature Reviews Microbiology,* 9**,** 177-192.

232

HARRIMAN, A. 1987. Further Comments On The Redox Potentials Of Tryptophan And Tyrosine. *Journal of Physical Chemistry,* 91**,** 6102-6104.

HASHEMIFARD, N., MOHSENIFAR, A., RANJBAR, B., ALLAMEH, A., LOTFI, A. S. & ETEMADIKIA, B. 2010. Fabrication and kinetic studies of a novel silver nanoparticles-glucose oxidase bioconjugate. *Analytica Chimica Acta,* 675**,** 181-184.

HATTON, M. W. C. & REGOECZI, E. 1976. Proteolytic Nature Of Commercial Samples Of Galactose Oxidase - Purification Of Enzyme By A Simple Affinity Method. *Biochimica Et Biophysica Acta,* 438**,** 339-346.

HERSCHLAG, D. 1988. The Role Of Induced Fit And Conformational-Changes Of Enzymes In Specificity And Catalysis. *Bioorganic Chemistry,* 16**,** 62-96.

HOLM, R. H., KENNEPOHL, P. & SOLOMON, E. I. 1996. Structural and functional aspects of metal sites in biology. *Chemical Reviews,* 96**,** 2239-2314.

HOLTON, J. B. 1990. Galactose Disorders - An Overview. *Journal of Inherited Metabolic Disease,* 13**,** 476-486.

HORSMAN, G. P., LIU, A. M. F., HENKE, E., BORNSCHEUER, U. T. & KAZLAUSKAS, R. J. 2003. Mutations in distant residues moderately increase the enantioselectivity of Pseudomonas fluorescens esterase towards methyl 3-bromo-2-methylpropanoate and ethyl 3-phenylbutyrate. *Chemistry-a European Journal,* 9**,** 1933-1939.

HU, W. B., KNIGHT, D., LOWRY, B. & VARMA, A. 2010. Selective Oxidation of Glycerol to Dihydroxyacetone over Pt-Bi/C Catalyst: Optimization of Catalyst and Reaction Conditions. *Industrial & Engineering Chemistry Research,* 49**,** 10876-10882.

HUGHES, J. M. & FINGER, L. W. 1983. The Crystal-Chemistry Of Shcherbinaite, Naturally-Occurring V2O5. *American Mineralogist,* 68**,** 1220-1222.

HUGHES, M. D., NAGEL, D. A., SANTOS, A. F., SUTHERLAND, A. J. & HINE, A. V. 2003. Removing the redundancy from randomised gene libraries. *Journal of Molecular Biology,* 331**,** 973-979.

HUMPHREYS, K. J., MIRICA, L. M., WANG, Y. & KLINMAN, J. P. 2009. Galactose Oxidase as a Model for Reactivity at a Copper Superoxide Center. *Journal of the American Chemical Society,* 131**,** 4657-4663.

IGARASHI, N., MORIYAMA, H., FUJIWARA, T., FUKUMORI, Y. & TANAKA, N. 1997. The 2.8 angstrom structure of hydroxylamine oxidoreductase from a nitrifying chemoautotrophic bacterium, Nitrosomonas europaea. *Nature Structural Biology,* 4**,** 276-284.

INUI, M., SUDA, M., KIMURA, S., YASUDA, K., SUZUKI, H., TODA, H., YAMAMOTO, S., OKINO, S., SUZUKI, N. & YUKAWA, H. 2008. Expression of Clostridium acetobutylicum butanol synthetic genes in Escherichia coli. *Applied Microbiology and Biotechnology,* 77**,** 1305-1316.

ISBELL, H. S. & PIGMAN, W. W. 1968. Mutarotation of Sugars in Solution: Part I: History, Basic Kinetics, and Composition of Sugar Solutions. *Advances in Carbohydrate Chemistry,* 23**,** 11-57.

ISBELL, H. S. & PIGMAN, W. W. 1969. Mutarotation of Sugars in Solution: Part II: Catalytic Processes, Isotope Effects, Reaction Mechanisms, and Biochemical Aspects. *Advances in Carbohydrate Chemistry and Biochemistry,* 24**,** 13-65.

ITO, N., PHILLIPS, S. E. V., STEVENS, C., OGEL, Z. B., MCPHERSON, M. J., KEEN, J. N., YADAV, K. D. S. & KNOWLES, P. F. 1991. Novel Thioether Bond Revealed by a 1.7-A Crystal-Structure of Galactose-Oxidase. *Nature,* 350**,** 87-90.

ITO, N., PHILLIPS, S. E. V., YADAV, K. D. S. & KNOWLES, P. F. 1994. Crystal-structure of a free-radical enzyme, galactose-oxidase. *Journal of Molecular Biology,* 238**,** 794-814.

JANCZYK, M., APPEL, B., SPRINGSTUBBE, D., FRITZ, H. J. & MULLER, S. 2012. A new and convenient approach for the preparation of beta-cyanoethyl protected trinucleotide phosphoramidites. *Organic & Biomolecular Chemistry,* 10**,** 1510-1513.

JIANG, L., ALTHOFF, E. A., CLEMENTE, F. R., DOYLE, L., ROTHLISBERGER, D., ZANGHELLINI, A., GALLAHER, J. L., BETKER, J. L., TANAKA, F., BARBAS, C. F., III, HILVERT, D., HOUK, K. N., STODDARD, B. L. & BAKER, D. 2008. De novo computational design of retro-aldol enzymes. *Science,* 319**,** 1387-1391.

JOCHENS, H., AERTS, D. & BORNSCHEUER, U. T. 2010. Thermostabilization of an esterase by alignment-guided focussed directed evolution. *Protein Engineering Design & Selection,* 23**,** 903-909.

JOCHENS, H. & BORNSCHEUER, U. T. 2010. Natural Diversity to Guide Focused Directed Evolution. *Chembiochem,* 11**,** 1861-1866.

JOCHENS, H., STIBA, K., SAVILE, C., FUJII, R., YU, J. G., GERASSENKOV, T., KAZLAUSKAS, R. J. & BORNSCHEUER, U. T. 2009. Converting an Esterase into an Epoxide Hydrolase. *Angewandte Chemie-International Edition,* 48**,** 3532-3535.

JOHNSON, D. T. & TACONI, K. A. 2007. The glycerin glut: Options for the value-added conversion of crude glycerol resulting from biodiesel production. *Environmental Progress,* 26**,** 338-348.

JOHNSON, K. A. 2008. Role of induced fit in enzyme specificity: A molecular forward/reverse switch. *Journal of Biological Chemistry,* 283**,** 26297-26301.

JOJIMA, T., INUI, M. & YUKAWA, H. 2008. Production of isopropanol by metabolically engineered Escherichia coli. *Applied Microbiology and Biotechnology,* 77**,** 1219-1224.

KAIM, W. & RALL, J. 1996. Copper - A ''modern'' bioelement. *Angewandte Chemie-International Edition in English,* 35**,** 43-60.

KANYONG, P., PEMBERTON, R. M., JACKSON, S. K. & HART, J. P. 2013. Development of an amperometric screen-printed galactose biosensor for serum analysis. *Analytical Biochemistry,* 435**,** 114-119.

KARLSSON, A., PARALES, J. V., PARALES, R. E., GIBSON, D. T., EKLUND, H. & RAMASWAMY, S. 2003. Crystal structure of naphthalene dioxygenase: Side-on binding of dioxygen to iron. *Science,* 299**,** 1039-1042.

KAYUSHIN, A., KOROSTELEVA, M. & MIROSHNIKOV, A. 2000. Large-scale solid-phase preparation of 3 '-unprotected trinucleotide phosphotriesters - Precursors for synthesis of trinucleotide phosphoramidites. *Nucleosides Nucleotides & Nucleic Acids,* 19**,** 1967-1976.

KELLEHER, F. M. & BHAVANANDAN, V. P. 1986. Re-examination Of The Products Of The Action Of Galactose-Oxidase - Evidence For The Conversion Of Raffinose To 6''-Carboxyraffinose. *Journal of Biological Chemistry,* 261**,** 1045-1048.

KELLY-FALCOZ, F., GREENBERG, H. & HORECKER, B. L. 1965. Galactose Oxidase - Studies On Structure And Role Of Disulfide Linkages. *Journal of Biological Chemistry,* 240**,** 2966-2970.

KERSTEN, P. J. & KIRK, T. K. 1987. Involvement Of A New Enzyme, Glyoxal Oxidase, In Extracellular H2O2 Production By Phanerochaete chrysosporium. *Journal of Bacteriology,* 169**,** 2195-2201.

KILLE, S., ACEVEDO-ROCHA, C. G., PARRA, L. P., ZHANG, Z.-G., OPPERMAN, D. J., REETZ, M. T. & PABLO ACEVEDO, J. 2013. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *Acs Synthetic Biology,* 2**,** 83-92.

KIM, M. I., SHIM, J., LI, T., WOO, M.-A., CHO, D., LEE, J. & PARK, H. G. 2012. Colorimetric quantification of galactose using a nanostructured multi-catalyst system entrapping

galactose oxidase and magnetic nanoparticles as peroxidase mimetics. *Analyst,* 137**,** 1137-1143.

KOSHLAND, D. E. 1994. The Key-Lock Theory And The Induced Fit Theory. *Angewandte Chemie-International Edition,* 33**,** 2375-2378.

KOSMAN, D. J. 1984. Galactose Oxidase. *In:* LONTIE, R. (ed.) *Copper Proteins and Copper Enzymes.* CRC Press.

KOSMAN, D. J., ETTINGER, M. J., WEINER, R. E. & MASSARO, E. J. 1974. Molecular-Properties Of Copper Enzyme Galactose Oxidase. *Archives of Biochemistry and Biophysics,* 165**,** 456-467.

KREBBER, A., BURMESTER, J. & PLUCKTHUN, A. 1996. Inclusion of an upstream transcriptional terminator in phage display vectors abolishes background expression of toxic fusions with coat protein g3p. *Gene,* 178**,** 71-74.

KRIES, H., BLOMBERG, R. & HILVERT, D. 2013. De novo enzymes by computational design. *Current Opinion in Chemical Biology,* 17**,** 221-228.

KUIPERS, R. K., JOOSTEN, H.-J., VAN BERKEL, W. J. H., LEFERINK, N. G. H., ROOIJEN, E., ITTMANN, E., VAN ZIMMEREN, F., JOCHENS, H., BORNSCHEUER, U., VRIEND, G., DOS SANTOS, V. A. P. M. & SCHAAP, P. J. 2010. 3DM: Systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins-Structure Function and Bioinformatics,* 78**,** 2101-2113.

KWIATKOWSKI, L. D. & KOSMAN, D. J. 1973. Role Of Superoxide Radical In Mechanism Of Action Of Galactose Oxidase. *Biochemical and Biophysical Research Communications,* 53**,** 715-721.

LAEMMLI, U. K. 1970. Cleavage Of Structural Proteins During Assembly Of Head Of Bacteriophage-T4. *Nature,* 227**,** 680-685.

LEEMHUIS, H., KELLY, R. M. & DIJKHUIZEN, L. 2009. Directed Evolution of Enzymes: Library Screening Strategies. *Iubmb Life,* 61**,** 222-228.

LEEMHUIS, H., STEIN, V., GRIFFITHS, A. D. & HOLLFELDER, F. 2005. New genotype-phenotype linkages for directed evolution of functional proteins. *Current Opinion in Structural Biology,* 15**,** 472-478.

LITTLECHILD, J., GARCIA-RODRIGUEZ, E., DALBY, A. & ISUPOV, M. 2002. Structural and functional comparisons between vanadium haloperoxidase and acid phosphatase enzymes. *Journal of Molecular Recognition,* 15**,** 291-296.

LIU, T. N. E., MTIMKULU, T., GEIGERT, J., WOLF, B., NEIDLEMAN, S. L., SILVA, D. & HUNTERCEVERA, J. C. 1987. Isolation And Characterization Of A Novel Nonheme Chloroperoxidase. *Biochemical and Biophysical Research Communications,* 142**,** 329-333.

LIU, W. F. & WANG, P. 2007. Cofactor regeneration for sustainable enzymatic biosynthesis. *Biotechnology Advances,* 25**,** 369-384.

LIU, X. C. & DORDICK, J. S. 1999. Sugar-containing polyamines prepared using galactose oxidase coupled with chemical reduction. *Journal of the American Chemical Society,* 121**,** 466-467.

LU, A. H., SALABAS, E. L. & SCHUTH, F. 2007. Magnetic nanoparticles: Synthesis, protection, functionalization, and application. *Angewandte Chemie-International Edition,* 46**,** 1222-1244.

LUTZ, S. 2010. Beyond directed evolution-semi-rational protein engineering and design. *Current Opinion in Biotechnology,* 21**,** 734-743.

MACPHERSON, I. S. & MURPHY, M. E. P. 2007. Type-2 copper-containing enzymes. *Cellular and Molecular Life Sciences,* 64**,** 2887-2899.

MAHESHRI, N. & SCHAFFER, D. V. 2003. Computational and experimental analysis of DNA shuffling. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 3071-3076.

MAHMOUD, K. M. A. G., SPOONER, R. K., DEACON, S. & MCPHERSON, M. J. 2000. Domain I of galactose oxidase is not required for catalytic activity. *Biochemical Society Transactions,* 28**,** A72-A72.

MANNINO, S., COSIO, M. S. & BURATTI, S. 1999. Simultaneous determination of glucose and galactose in dairy products by two parallel amperometric biosensors. *Italian Journal of Food Science,* 11**,** 57-65.

MATSUMOTO, M. & OHASHI, K. 2003. Effect of immobilization on thermostability of lipase from Candida rugosa. *Biochemical Engineering Journal,* 14**,** 75-77.

MAZUR, A. W. & HILER, G. D. 1997. Chemoenzymic approaches to the preparation of 5-C-(hydroxymethyl)hexoses. *Journal of Organic Chemistry,* 62**,** 4471-4475.

MCCOY, J. G., BAILEY, L. J., BITTO, E., BINGMAN, C. A., ACETI, D. J., FOX, B. G. & PHILLIPS, G. N. 2006. Structure and mechanism of mouse cysteine dioxygenase. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 3084-3089.

MCLOUGHLIN, S. Y., JACKSON, C., LIU, J. W. & OLLIS, D. 2005. Increased expression of a bacterial phosphotriesterase in Escherichia coli through directed evolution. *Protein Expression and Purification,* 41**,** 433-440.

MCPHERSON, M. & MOLLER, S. 2006. *PCR*, Taylor & Francis Group.

MCPHERSON, M. J., OGEL, Z. B., STEVENS, C., YADAV, K. D. S., KEEN, J. N. & KNOWLES, P. F. 1992. Galactose-Oxidase Of Dactylium-Dendroides - Gene Cloning And Sequence-Analysis. *Journal of Biological Chemistry,* 267**,** 8146-8152.

MCPHERSON, M. J., STEVENS, C., BARON, A. J., OGEL, Z. B., SENEVIRATNE, K., WILMOT, C., ITO, N., BROCKLEBANK, I., PHILLIPS, S. E. V. & KNOWLES, P. F. 1993. Galactose-Oxidase - Molecular Analysis And Mutagenesis Studies. *Biochemical Society Transactions,* 21**,** 752-756.

MEAGHER, R. B. 2000. Phytoremediation of toxic elemental and organic pollutants. *Current Opinion in Plant Biology,* 3**,** 153-162.

MERINO-MONTIEL, P., LOPEZ, O., ALVAREZ, E. & FERNANDEZ-BOLANOS, J. G. 2012. Synthesis of conformationally-constrained thio(seleno)hydantoins and alpha-triazolyl lactones from D-arabinose as potential glycosidase inhibitors. *Tetrahedron,* 68**,** 4888-4898.

MESSERSCHMIDT, A., HUBER, R., POULOS, T. & WIEGHARDT, K. 2001. *Handbook of Metalloproteins,* England, John Wiley & Sons, Ltd.

MINASIAN, S. G., WHITTAKER, M. M. & WHITTAKER, J. W. 2004. Stereoselective Hydrogen Abstraction by Galactose Oxidase. *Biochemistry,* 43**,** 13683-13693.

MIYAZAKI, J., NAKAYA, S., SUZUKI, T., TAMAKOSHI, M., OSHIMA, T. & YAMAGISHI, A. 2001. Ancestral residues stabilizing 3-isopropylmalate dehydrogenase of an extreme thermophile: Experimental evidence supporting the thermophilic common ancestor hypothesis. *Journal of Biochemistry,* 129**,** 777-782.

MOHANTY, J. G., JAFFE, J. S., SCHULMAN, E. S. & RAIBLE, D. G. 1997. A highly sensitive fluorescent micro-assay of H2O2 release from activated human leukocytes using a dihydroxyphenoxazine derivative. *Journal of Immunological Methods,* 202**,** 133-141.

MOORE, G. L. & MARANAS, C. D. 2002. Predicting out-of-sequence reassembly in DNA shuffling. *Journal of Theoretical Biology,* 219**,** 9-17.

MOORE, G. L., MARANAS, C. D., LUTZ, S. & BENKOVIC, S. J. 2001. Predicting crossover generation in DNA shuffling. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 3226-3231.

MORELL, A. G., VAN DEN HAMER, C. J. A., SCHEINBERG, I. H. & ASHWELL, G. 1966. Physical and Chemical Studies on Ceruloplasmin. *Journal of Biological Chemistry,* 241**,** 3745-3749.

MORLEY, K. L. & KAZLAUSKAS, R. J. 2005. Improving enzyme properties: when are closer mutations better? *Trends in Biotechnology,* 23**,** 231-237.

MURAKAMI, H., HOHSAKA, T. & SISIDO, M. 2002. Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. *Nature Biotechnology,* 20**,** 76-81.

MURZIN, A. G. 1992. Structural Principles For The Propeller Assembly Of Beta-Sheets - The Preference For 7-Fold Symmetry. *Proteins-Structure Function and Genetics,* 14**,** 191-201.

NAGASAWA, T. & YAMADA, H. 1989. Microbial Transformations Of Nitriles. *Trends in Biotechnology,* 7**,** 153-158.

NAKAMURA, C. E. & WHITED, G. M. 2003. Metabolic engineering for the microbial production of 1,3-propanediol. *Current Opinion in Biotechnology,* 14**,** 454-459.

NATHAN, D. M. 1993. Long term complications of Diabetes mellitus. *New England Journal of Medicine,* 328**,** 1676-1685.

NESS, J. E., KIM, S., GOTTMAN, A., PAK, R., KREBBER, A., BORCHERT, T. V., GOVINDARAJAN, S., MUNDORFF, E. C. & MINSHULL, J. 2002. Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nature Biotechnology,* 20**,** 1251-1255.

NEYLON, C. 2004. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Research,* 32**,** 1448-1459.

NGAMSOM, B., HICKEY, A. M., GREENWAY, G. M., LITTLECHILD, J. A., MCCREEDY, T., WATTS, P. & WILES, C. 2010. The development and evaluation of a conducting matrix for the electrochemical regeneration of the immobilised co-factor NAD(H) under continuous flow. *Organic & Biomolecular Chemistry,* 8**,** 2419-2424.

NIELSEN, H., BRUNAK, S. & VON HEIJNE, G. 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Engineering,* 12**,** 3-9.

NISHIHARA, T., IWABUCHI, T. & NOHNO, T. 1994. A T7 Promoter Vector With A Transcriptional Terminator For Stringent Expression Of Foreign Genes. *Gene,* 145**,** 145-146.

NOBLES, M. K. & MADHOSINGH, C. 1963. Dactylium dendroides (bull.) Fr. Misnamed as Polyporus circinatus Fr. *Biochemical and Biophysical Research Communications,* 12**,** 146-147.

NOVOGRODSKY, A. & KATCHALSKI, E. 1973. Induction Of Lymphocyte Transformation By Sequential Treatment With Neuraminidase And Galactose Oxidase. *Proceedings of the National Academy of Sciences of the United States of America,* 70**,** 1824-1827.

OGEL, Z. B., BRAYFORD, D. & MCPHERSON, M. J. 1994. Cellulose-triggered sporulation in the galactose oxidase-producing fungus Cladobotryum (Dactylium) dendroides NRRL-2903 and its reidentification as a species of Fusarium. *Mycological Research,* 98**,** 474-480.

OKELEY, N. M. & VAN DER DONK, W. A. 2000. Novel cofactors via post-translational modifications of enzyme active sites. *Chemistry & Biology,* 7**,** R159-R171.

OSTERMEIER, C., HARRENGA, A., ERMLER, U. & MICHEL, H. 1997. Structure at 2.7 angstrom resolution of the Paracoccus denitrificans two-subunit cytochrome c oxidase complexed with an antibody F-V fragment. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 10547-10553.

OSTERMEIER, M., SHIM, J. H. & BENKOVIC, S. J. 1999. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nature Biotechnology,* 17**,** 1205-1209.

PANDEY, P., SINGH, S. P., ARYA, S. K., GUPTA, V., DATTA, M., SINGH, S. & MALHOTRA, B. D. 2007. Application of thiolated gold nanoparticles for the enhancement of glucose oxidase activity. *Langmuir,* 23**,** 3333-3337.

PARIKH, M. R. & MATSUMURA, I. 2005. Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of beta-fucosidase from beta-galactosidase. *Journal of Molecular Biology,* 352**,** 621-628.

PARIKKA, K., LEPPANEN, A. S., XU, C. L., PITKANEN, L., ERONEN, P., OSTERBERG, M., BRUMER, H., WILLFOR, S. & TENKANEN, M. 2012. Functional and Anionic Cellulose-Interacting Polymers by Selective Chemo-Enzymatic Carboxylation of Galactose-Containing Polysaccharides. *Biomacromolecules,* 13**,** 2418-2428.

PARIKKA, K. & TENKANEN, M. 2009. Oxidation of methyl alpha-D-galactopyranoside by galactose oxidase: products formed and optimization of reaction conditions for production of aldehyde. *Carbohydrate Research,* 344**,** 14-20.

PASTER, M., PELLEGRINO, J. L. & CAROLE, T. M. 2003. Industrial Bioproducts: Today and Tomorrow. Report prepared by Energetics, Inc., for the United States Department of Energy, Office of Energy Efficiency and Renewable Energy, Office of the Biomass Program, Washington, DC.

PATRICK, W. M. & FIRTH, A. E. 2005. Strategies and computational tools for improving randomized protein libraries. *Biomolecular Engineering,* 22**,** 105-112.

PATRICK, W. M., FIRTH, A. E. & BLACKBURN, J. M. 2003. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Engineering,* 16**,** 451-457.

PAVELKA, A., CHOVANCOVA, E. & DAMBORSKY, J. 2009. HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Research,* 37**,** W376-W383.

PEARSON, A. R. & OWEN, R. L. 2009. Combining X-ray crystallography and single-crystal spectroscopy to probe enzyme mechanisms. *Biochemical Society Transactions,* 37**,** 378-381.

POLYAKOV, K. M., BOYKO, K. M., TIKHONOVA, T. V., SLUTSKY, A., ANTIPOV, A. N., ZVYAGILSKAYA, R. A., POPOV, A. N., BOURENKOV, G. P., LAMZIN, V. S. & POPOV, V. O. 2009. High-Resolution Structural Analysis of a Novel Octaheme Cytochrome c Nitrite Reductase from the Haloalkaliphilic Bacterium Thioalkalivibrio nitratireducens. *Journal of Molecular Biology,* 389**,** 846-862.

QUAN, J. Y. & TIAN, J. D. 2009. Circular Polymerase Extension Cloning of Complex Gene Libraries and Pathways. *Plos One,* 4**,** 1-6.

RANNES, J. B., IOANNOU, A., WILLIES, S. C., GROGAN, G., BEHRENS, C., FLITSCH, S. L. & TURNER, N. J. 2011. Glycoprotein Labeling Using Engineered Variants of Galactose Oxidase Obtained by Directed Evolution. *Journal of the American Chemical Society,* 133**,** 8436-8439.

RANOUX, A., KARMEE, S. K., JIN, J. F., BHADURI, A., CAIAZZO, A., ARENDS, I. & HANEFELD, U. 2012. Enhancement of the Substrate Scope of Transketolase. *Chembiochem,* 13**,** 1921-1931.

REETZ, M. T., BOCOLA, M., CARBALLEIRA, J. D., ZHA, D. X. & VOGEL, A. 2005. Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angewandte Chemie-International Edition,* 44**,** 4192-4196.

REETZ, M. T. & CARBALLEIRA, J. D. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature Protocols,* 2**,** 891-903.

REETZ, M. T., D CARBALLEIRA, J. & VOGEL, A. 2006a. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angewandte Chemie-International Edition,* 45**,** 7745-7751.

REETZ, M. T., KAHAKEAW, D. & LOHMER, R. 2008. Addressing the numbers problem in directed evolution. *Chembiochem,* 9**,** 1797-1804.

REETZ, M. T., SONI, P., FERNANDEZ, L., GUMULYA, Y. & CARBALLEIRA, J. D. 2010. Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chemical Communications,* 46**,** 8657-8658.

REETZ, M. T., WANG, L. W. & BOCOLA, M. 2006b. Directed evolution of enantioselective enzymes: Iterative cycles of CASTing for probing protein-sequence space. *Angewandte Chemie-International Edition,* 45**,** 1236-1241.

REETZ, M. T. & WU, S. 2008. Greatly reduced amino acid alphabets in directed evolution: making the right choice for saturation mutagenesis at homologous enzyme positions. *Chemical Communications***,** 5499-5501.

REYNOLDS, M. P., BARON, A. J., WILMOT, C. M., PHILLIPS, S. E. V., KNOWLES, P. F. & MCPHERSON, M. J. 1995. Tyrosine 495 is a key residue in the active site of Galactose oxidase. *Biochemical Society Transactions,* 23**,** S510-S510.

RO, D. K., PARADISE, E. M., OUELLET, M., FISHER, K. J., NEWMAN, K. L., NDUNGU, J. M., HO, K. A., EACHUS, R. A., HAM, T. S., KIRBY, J., CHANG, M. C. Y., WITHERS, S. T., SHIBA, Y., SARPONG, R. & KEASLING, J. D. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature,* 440**,** 940-943.

ROGERS, M. S., BARON, A. J., MCPHERSON, M. J., KNOWLES, P. F. & DOOLEY, D. M. 2000. Galactose oxidase pro-sequence cleavage and cofactor assembly are self-processing reactions. *Journal of the American Chemical Society,* 122**,** 990-991.

ROGERS, M. S., HURTADO-GUERRERO, R., FIRBANK, S. J., HALCROW, M. A., DOOLEY, D. M., PHILLIPS, S. E. V., KNOWLES, P. F. & MCPHERSON, M. J. 2008. Cross-link formation of the cysteine 228-tyrosine 272 catalytic cofactor of galactose oxidase does not require dioxygen. *Biochemistry,* 47**,** 10428-10439.

ROGERS, M. S., TYLER, E. M., AKYUMANI, N., KURTIS, C. R., SPOONER, R. K., DEACON, S. E., TAMBER, S., FIRBANK, S. J., MAHMOUD, K., KNOWLES, P. F., PHILLIPS, S. E. V., MCPHERSON, M. J. & DOOLEY, D. M. 2007. The stacking tryptophan of galactose oxidase: A second-coordination sphere residue that has profound effects on tyrosyl radical behavior and enzyme catalysis. *Biochemistry,* 46**,** 4606-4618.

ROOT, R. L., DURRWACHTER, J. R. & WONG, C. H. 1985. Enzymatic-Synthesis Of Unusual Sugars - Galactose-Oxidase Catalyzed Stereospecific Oxidation Of Polyols. *Journal of the American Chemical Society,* 107**,** 2997-2999.

ROPER, J. M., RAUX, E., BRINDLEY, A. A., SCHUBERT, H. L., GHARBIA, S. E., SHAH, H. N. & WARREN, M. J. 2000. The enigma of cobalamin (vitamin B-12) biosynthesis in Porphyromonas gingivalis - Identification and characterization of a functional corrin pathway. *Journal of Biological Chemistry,* 275**,** 40316-40323.

ROREM, E. S. & LEWIS, J. C. 1962. A Test Paper For Detection Of Galactose And Certain Galactose-Containing Sugars. *Analytical Biochemistry,* 3**,** 230-235.

ROTHLISBERGER, D., KHERSONSKY, O., WOLLACOTT, A. M., JIANG, L., DECHANCIE, J., BETKER, J., GALLAHER, J. L., ALTHOFF, E. A., ZANGHELLINI, A., DYM, O., ALBECK, S., HOUK, K. N., TAWFIK, D. S. & BAKER, D. 2008. Kemp elimination catalysts by computational enzyme design. *Nature,* 453**,** 190-195.

RUBINO, J. T. & FRANZ, K. J. 2012. Coordination chemistry of copper proteins: How nature handles a toxic cargo for essential function. *Journal of Inorganic Biochemistry,* 107**,** 129-143.

RUIZ, A. I., MALAVE, A. J., FELBY, C. & GRIEBENOW, K. 2000. Improved activity and stability of an immobilized recombinant laccase in organic solvents. *Biotechnology Letters,* 22**,** 229-233.

SAMBROOK, J., FRITSCH, E. & MANIATUS, T. 1989. *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press.

SANNER, M. F. 1999. Python: A programming language for software integration and development. *Journal of Molecular Graphics & Modelling,* 17**,** 57-61.

SAVILE, C. K., JANEY, J. M., MUNDORFF, E. C., MOORE, J. C., TAM, S., JARVIS, W. R., COLBECK, J. C., KREBBER, A., FLEITZ, F. J., BRANDS, J., DEVINE, P. N., HUISMAN, G. W. & HUGHES, G. J. 2010. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science,* 329**,** 305-309.

SAXENA, R. K., ANAND, P., SARAN, S. & ISAR, J. 2009. Microbial production of 1,3-propanediol: Recent developments and emerging opportunities. *Biotechnology Advances,* 27**,** 895-913.

SAYER, C., BOMMER, M., ISUPOV, M., WARD, J. & LITTLECHILD, J. 2012. Crystal structure and substrate specificity of the thermophilic serine: pyruvate aminotransferase from Sulfolobus solfataricus. *Acta Crystallographica Section D-Biological Crystallography,* 68**,** 763-772.

SCHIERLE, C. F., BERKMEN, M., HUBER, D., KUMAMOTO, C., BOYD, D. & BECKWITH, J. 2003. The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the Escherichia coli periplasm via the signal recognition particle pathway. *Journal of Bacteriology,* 185**,** 5706-5713.

SCHLEGEL, R. A., GERBECK, C. M. & MONTGOME.R 1968. Substrate specificity of D-galactose oxidase. *Carbohydrate Research,* 7**,** 193-199.

SCHMIDT, B., SELMER, T., INGENDOH, A. & VONFIGURA, K. 1995. A Novel Amino-Acid Modification In Sulfatases That Is Defective In Multiple Sulfatase Deficiency. *Cell,* 82**,** 271-278.

SCHNELL, R., SANDALOVA, T., HELLMAN, U., LINDQVIST, Y. & SCHNEIDER, G. 2005. Siroheme- and Fe-4-S-4 -dependent NirA from Mycobacterium tuberculosis is a sulfite reductase with a covalent Cys-Tyr bond in the active site. *Journal of Biological Chemistry,* 280**,** 27319-27328.

SCHOEVAART, R. & KIEBOOM, T. 2004. Application of galactose oxidase in chemoenzymatic one-pot cascade reactions without intermediate recovery steps. *Topics in Catalysis,* 27**,** 3-9.

SCHUCKEL, J., RYLOTT, E. L., GROGAN, G. & BRUCE, N. C. 2012. A Gene-Fusion Approach to Enabling Plant Cytochromes P450 for Biocatalysis. *Chembiochem,* 13**,** 2758-2763.

SEGEL, I. H. 1975. *Enzyme Kinetics*, Wiley-Interscience, New York.

SEIFERT, A., ANTONOVICI, M., HAUER, B. & PLEISS, J. 2011. An Efficient Route to Selective Bio-oxidation Catalysts: an Iterative Approach Comprising Modeling, Diversification, and Screening, Based on CYP102A1. *Chembiochem,* 12**,** 1346-1351.

SETHI, M. K., BHANDYA, S. R., MADDUR, N., SHUKLA, R., KUMAR, A. & MITTAPALLI, V. 2013. Asymmetric synthesis of an enantiomerically pure rivastigmine intermediate using ketoreductase. *Tetrahedron-Asymmetry,* 24**,** 374-379.

SHAFFER, P. A. & FRIEDEMANN, T. E. 1930. Sugar activation by alkali. I. Formation of lactic and saccharinic acids. *Journal of Biological Chemistry,* 86**,** 345-374.

SHAMSUDDIN, A. M. & ELSAYED, A. M. 1988. A Test For Detection Of Colorectal-Cancer. *Human Pathology,* 19**,** 7-10.

SHARMA, S. K., SUMAN, PUNDIR, C. S., SEHGAL, N. & KUMAR, A. 2006. Galactose sensor based on galactose oxidase immobilized in polyvinyl formal. *Sensors and Actuators B-Chemical,* 119**,** 15-19.

SIEBER, V., MARTINEZ, C. A. & ARNOLD, F. H. 2001. Libraries of hybrid proteins from distantly related sequences. *Nature Biotechnology,* 19**,** 456-460.

SIEGEL, J. B., ZANGHELLINI, A., LOVICK, H. M., KISS, G., LAMBERT, A. R., CLAIR, J. L. S., GALLAHER, J. L., HILVERT, D., GELB, M. H., STODDARD, B. L., HOUK, K. N., MICHAEL, F.

E. & BAKER, D. 2010. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science,* 329**,** 309-313.

SIMMONS, C. R., LIU, Q., HUANG, Q. Q., HAO, Q., BEGLEY, T. P., KARPLUS, P. A. & STIPANUK, M. H. 2006. Crystal structure of mammalian cysteine dioxygenase - A novel mononuclear iron center for cysteine thiol oxidation. *Journal of Biological Chemistry,* 281**,** 18723-18733.

SINGH, H. & KANFER, J. N. 1980. Quantitation Of Galactocerebrosides And Sulfatides By Galactose-Oxidase And Sodium Borotritide. *Analytical Biochemistry,* 109**,** 27-31.

SINGH, R. K., TIWARI, M. K., SINGH, R. & LEE, J. K. 2013. From Protein Engineering to Immobilization: Promising Strategies for the Upgrade of Industrial Enzymes. *International Journal of Molecular Sciences,* 14**,** 1232-1277.

SPADIUT, O., OLSSON, L. & BRUMER, H., III 2010. A comparative summary of expression systems for the recombinant production of galactose oxidase. *Microbial Cell Factories,* 9**,** 68-80.

SPURLINO, J. C., LU, G. Y. & QUIOCHO, F. A. 1991. The 2.3-A Resolution Structure Of The Maltose-Binding Or Maltodextrin-Binding Protein, A Primary Receptor Of Bacterial Active-Transport And Chemotaxis. *Journal of Biological Chemistry,* 266**,** 5202-5219.

STEMMER, W. P. C. 1994a. DNA Shuffling By Random Fragmentation And Reassembly - In-Vitro Recombination For Molecular Evolution. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 10747-10751.

STEMMER, W. P. C. 1994b. Rapid Evolution Of A Protein In-Vitro By DNA Shuffling. *Nature,* 370**,** 389-391.

STUBBE, J. & VAN DER DONK, W. A. 1998. Protein radicals in enzyme catalysis. *Chemical Reviews,* 98**,** 705-762.

STUDIER, F. W. 2005. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification,* 41**,** 207-234.

STUDIER, F. W. & MOFFATT, B. A. 1986. Use Of Bacteriophage-T7 RNA-Polymerase To Direct Selective High-Level Expression Of Cloned Genes. *Journal of Molecular Biology,* 189**,** 113-130.

SUN, L. H., BULTER, T., ALCALDE, M., PETROUNIA, I. P. & ARNOLD, F. H. 2002. Modification of galactose oxidase to introduce glucose 6-oxidase activity. *Chembiochem,* 3**,** 781-783.

SUN, L. H., PETROUNIA, I. P., YAGASAKI, M., BANDARA, G. & ARNOLD, F. H. 2001. Expression and stabilization of galactose oxidase in Escherichia coli by directed evolution. *Protein Engineering,* 14**,** 699-704.

TALY, V., KELLY, B. T. & GRIFFITHS, A. D. 2007. Droplets as microreactors for high-throughput biology. *Chembiochem,* 8**,** 263-272.

TAWFIK, D. S. & GRIFFITHS, A. D. 1998. Man-made cell-like compartments for molecular evolution. *Nature Biotechnology,* 16**,** 652-656.

TAYLOR, P. J., KMETEC, E. & JOHNSON, J. M. 1977. Design, Construction, And Applications Of A Galactose Selective Electrode. *Analytical Chemistry,* 49**,** 789-794.

TAYLOR, T. C. & ANDERSSON, I. 1997. The structure of the complex between rubisco and its natural substrate ribulose 1,5-bisphosphate. *Journal of Molecular Biology,* 265**,** 432-444.

TERPE, K. 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology,* 60**,** 523-533.

TERPE, K. 2006. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology,* 72**,** 211-222.

THOMPSON, A. P. 1931. Platinum vs vanadium pentoxide as catalysts for sulfuric acid manufacture. *Transactions of the American Institute of Chemical Engineers,* 27**,** 264-309.

TOMASINI-JOHANSSON, B. R., JOHNSON, I. A., HOFFMANN, F. M. & MOSHER, D. F. 2012. Quantitative microtiter fibronectin fibrillogenesis assay: use in high throughput screening for identification of inhibitor compounds. *Matrix Biology,* 31**,** 360-367.

TORRES-SALAS, P., MATE, D. M., GHAZI, I., PLOU, F. J., BALLESTEROS, A. O. & ALCALDE, M. 2013. Widening the pH Activity Profile of a Fungal Laccase by Directed Evolution. *Chembiochem,* 14**,** 934-937.

TRESSEL, P. & KOSMAN, D. J. 1980. O,O-Dityrosine In Native And Horseradish Peroxidase-Activated Galactose-Oxidase. *Biochemical and Biophysical Research Communications,* 92**,** 781-786.

TSUJINO, Y., YOKOO, Y. & SAKATO, K. 1991. Hair Coloring And Waving Using Oxidases. *Journal of the Society of Cosmetic Chemists,* 42**,** 273-282.

VAN DER MEER, R. A., JONGEJAN, J. A. & DUINE, J. A. 1989. Pyrroloquinoline Quinone As Cofactor In Galactose-Oxidase (EC-1.1.3.9). *Journal of Biological Chemistry,* 264**,** 7792-7794.

VAN DER PALEN, C. J. N. M., REIJNDERS, W. N. M., DE VRIES, S., DUINE, J. A. & VAN SPANNING, R. J. M. 1997. MauE and MauD proteins are essential in methylamine metabolism of Paracoccus denitrificans. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology,* 72**,** 219-228.

VAN DER PALEN, C. J. N. M., SLOTBOOM, D. J., JONGEJAN, L., REIJNDERS, W. N. M., HARMS, N., DUINE, J. A. & VAN SPANNING, R. J. M. 1995. Mutational Analysis Of Mau Genes Involved In Methylamine Metabolism In Paracoccus-denitrificans. *European Journal of Biochemistry,* 230**,** 860-871.

VAN HERCKE, T., AMPE, C., TIRRY, L. & DENOLF, P. 2005. Reducing mutational bias in random protein libraries. *Analytical Biochemistry,* 339**,** 9-14.

VAN LEEUWEN, J. G. E., WIJMA, H. J., FLOOR, R. J., VAN DER LAAN, J. M. & JANSSEN, D. B. 2012. Directed Evolution Strategies for Enantiocomplementary Haloalkane Dehalogenases: From Chemical Waste to Enantiopure Building Blocks. *Chembiochem,* 13**,** 137-148.

VAN POELJE, P. D. & SNELL, E. E. 1990. Pyruvoyl-Dependent Enzymes. *Annual Review of Biochemistry,* 59**,** 29-59.

VAN SCHIJNDEL, J. W. P. M., BARNETT, P., ROELSE, J., VOLLENBROEK, E. G. M. & WEVER, R. 1994. The Stability And Steady-State Kinetics Of Vanadium Chloroperoxidase From The Fungus Curvularia-Inaequalis. *European Journal of Biochemistry,* 225**,** 151-157.

VAZQUEZ-FIGUEROA, E., CHAPARRO-RIGGERS, J. & BOMMARIUS, A. S. 2007. Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. *Chembiochem,* 8**,** 2295-2301.

VELLIEUX, F. M. D., HUITEMA, F., GROENDIJK, H., KALK, K. H., JZN, J. F., JONGEJAN, J. A., DUINE, J. A., PETRATOS, K., DRENTH, J. & HOL, W. G. J. 1989. Structure Of Quinoprotein Methylamine Dehydrogenase At 2.25-A Resolution. *Embo Journal,* 8**,** 2171-2178.

VETTING, M. W., D'ARGENIO, D. A., ORNSTON, L. N. & OHLENDORF, D. H. 2000. Structure of Acinetobacter strain ADP1 protocatechuate 3,4-dioxygenase at 2.2 angstrom resolution: Implications for the mechanism of an intradiol dioxygenase. *Biochemistry,* 39**,** 7943-7955.

VILTER, H. 1984. Peroxidases From Phaeophyceae - A Vanadium(V)-Dependent Peroxidase From Ascophyllum-Nodosum .5. *Phytochemistry,* 23**,** 1387-1390.

VOET, D. & VOET, J. G. 1995. Biochemistry. USA: John Wiley & Sons, Inc.

VOLLENBROEK, E. G. M., SIMONS, L. H., VAN SCHIJNDEL, J. W. P. M., BARNETT, P., BALZAR, M., DEKKER, H., VAN DER LINDEN, C. & WEVER, R. 1995. Vanadium Chloroperoxidases Occur Widely In Nature. *Biochemical Society Transactions,* 23**,** 267-271.

VOSS, S. & SKERRA, A. 1997. Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein Engineering,* 10**,** 975-982.

VRBOVA, E., PECKOVA, J. & MAREK, M. 1992. Preparation And Utilization Of A Biosensor Based On Galactose-Oxidase. *Collection of Czechoslovak Chemical Communications,* 57**,** 2287-2294.

WACHTER, R. M. & BRANCHAUD, B. P. 1996. Molecular Modeling Studies on Oxidation of Hexopyranoses by Galactose Oxidase. An Active Site Topology Apparently Designed To Catalyze Radical Reactions, Either Concerted or Stepwise. *J. Am. Chem. Soc.,* 118**,** 2782-2789.

WANG, D., BUSHNELL, D. A., WESTOVER, K. D., KAPLAN, C. D. & KORNBERG, R. D. 2006. Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis. *Cell,* 127**,** 941-954.

WANG, S. X., MURE, M., MEDZIHRADSZKY, K. F., BURLINGAME, A. L., BROWN, D. E., DOOLEY, D. M., SMITH, A. J., KAGAN, H. M. & KLINMAN, J. P. 1996. A crosslinked cofactor in lysyl oxidase: Redox function for amino acid side chains. *Science,* 273**,** 1078-1084.

WATANABE, K., OHKURI, T., YOKOBORI, S. & YAMAGISHI, A. 2006. Designing thermostable proteins: Ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree. *Journal of Molecular Biology,* 355**,** 664-674.

WHITTAKER, J. W. 2005. The radical chemistry of galactose oxidase. *Archives of Biochemistry and Biophysics,* 433**,** 227-239.

WHITTAKER, J. W. & WHITTAKER, M. M. 1998. Radical copper oxidases, one electron at a time. *Pure and Applied Chemistry,* 70**,** 903-910.

WHITTAKER, M. M., BALLOU, D. P. & WHITTAKER, J. W. 1998. Kinetic isotope effects as probes of the mechanism of galactose oxidase. *Biochemistry,* 37**,** 8426-8436.

WHITTAKER, M. M., DEVITO, V. L., ASHER, S. A. & WHITTAKER, J. W. 1989. Resonance Raman Evidence For Tyrosine Involvement In The Radical Site Of Galactose-Oxidase. *Journal of Biological Chemistry,* 264**,** 7104-7106.

WHITTAKER, M. M., KERSTEN, P. J., CULLEN, D. & WHITTAKER, J. W. 1999. Identification of catalytic residues in glyoxal oxidase by targeted mutagenesis. *Journal of Biological Chemistry,* 274**,** 36226-36232.

WHITTAKER, M. M., KERSTEN, P. J., NAKAMURA, N., SANDERSLOEHR, J., SCHWEIZER, E. S. & WHITTAKER, J. W. 1996. Glyoxal oxidase from Phanerochaete chrysosporium is a new radical-copper oxidase. *Journal of Biological Chemistry,* 271**,** 681-687.

WHITTAKER, M. M. & WHITTAKER, J. W. 1988. The Active-Site Of Galactose-Oxidase. *Journal of Biological Chemistry,* 263**,** 6074-6080.

WHITTAKER, M. M. & WHITTAKER, J. W. 1993. Ligand Interactions With Galactose-Oxidase - Mechanistic Insights. *Biophysical Journal,* 64**,** 762-772.

WHITTAKER, M. M. & WHITTAKER, J. W. 2000. Expression of recombinant galactose oxidase by Pichia pastoris. *Protein Expression and Purification,* 20**,** 105-111.

WILKINSON, D., AKUMANYI, N., HURTADO-GUERRERO, R., DAWKES, H., KNOWLES, P. F., PHILLIPS, S. E. V. & MCPHERSON, M. J. 2004. Structural and kinetic studies of a series of mutants of galactose oxidase identified by directed evolution. *Protein Engineering Design & Selection,* 17**,** 141-148.

WILLIAMS, G. J., ZHANG, C. & THORSON, J. S. 2007. Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. *Nature Chemical Biology,* 3**,** 657-662.

WILLIES, S., ISUPOV, M. & LITTLECHILD, J. 2010. Thermophilic enzymes and their applications in biocatalysis: a robust aldo-keto reductase. *Environmental Technology,* 31**,** 1159-1167.

WILMOT, C. M. & DAVIDSON, V. L. 2009. Uncovering novel biochemistry in the mechanism of tryptophan tryptophylquinone cofactor biosynthesis. *Current Opinion in Chemical Biology,* 13**,** 469-474.

WONG, T. S., TEE, K. L., HAUER, B. & SCHWANEBERG, U. 2004. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Research,* 32**,** e26.

WOODLEY, J. M. 2013. Protein engineering of enzymes for process applications. *Current Opinion in Chemical Biology,* 17**,** 310-316.

XIAO, G. H., YING, S. H., ZHENG, P., WANG, Z. L., ZHANG, S. W., XIE, X. Q., SHANG, Y. F., ST LEGER, R. J., ZHAO, G. P., WANG, C. S. & FENG, M. G. 2012. Genomic perspectives on the evolution of fungal entomopathogenicity in Beauveria bassiana. *Scientific Reports,* 2**,** 2-10.

YAGODKIN, A., AZHAYEV, A., ROIVAINEN, J., ANTOPOLSKY, M., KAYUSHIN, A., KOROSTELEVA, M., MIROSHNIKOV, A., RANDOLPH, J. & MACKIE, H. 2007. Improved synthesis of trinucleotide phosphoramidites and generation of randomized oligonucleotide libraries. *Nucleosides Nucleotides & Nucleic Acids,* 26**,** 473-497.

YALPANI, M. & HALL, L. D. 1982. Some chemical and analytical aspects of polysaccharide modifications .2. A high-yielding, specific method for the chemical derivatization of galactose-containing polysaccharides - oxidation with galactose-oxidase followed by reductive amination. *Journal of Polymer Science Part a-Polymer Chemistry,* 20, 3399-3420.

YANG, G. Y. & SHAMSUDDIN, A. M. 1996. Gal-GalNAc: A biomarker of colon carcinogenesis. *Histology and Histopathology,* 11**,** 801-806.

YI, D., DEVAMANI, T., ABDOUL-ZABAR, J., CHARMANTRAY, F., HELAINE, V., HECQUET, L. & FESSNER, W. D. 2012. A pH-Based High-Throughput Assay for Transketolase: Fingerprinting of Substrate Tolerance and Quantitative Kinetics. *Chembiochem,* 13**,** 2290-2300.

YOU, L. & ARNOLD, F. H. 1996. Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. *Protein Engineering,* 9**,** 77-83.

YOUNG, P. R. 1989. An Improved Method For The Detection Of Peroxidase-Conjugated Antibodies On Immunoblots. *Journal of Virological Methods,* 24**,** 227-235.

ZAFAR, M. N., PITTMAN, S., CHERCHI, M. & CATOVSKY, D. 1981. Stimulation Of Chronic Lymphatic-Leukemia Cells By Pokeweed Mitogen After Treatment With Neuraminidase-Galactose Oxidase. *Clinical and Experimental Immunology,* 44**,** 124-128.

ZHA, D. X., EIPPER, A. & REETZ, M. T. 2003. Assembly of designed oligonucleotides as an efficient method for gene recombination: A new tool in directed evolution. *Chembiochem,* 4**,** 34-39.

ZHAO, H. M., GIVER, L., SHAO, Z. X., AFFHOLTER, J. A. & ARNOLD, F. H. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology,* 16**,** 258-261.

ZHAO, H. M. & VAN DER DONK, W. A. 2003. Regeneration of cofactors for use in biocatallysis. *Current Opinion in Biotechnology,* 14**,** 583-589.

ZHENG, P., XIA, Y. L., XIAO, G. H., XIONG, C. H., HU, X., ZHANG, S. W., ZHENG, H. J., HUANG, Y., ZHOU, Y., WANG, S. Y., ZHAO, G. P., LIU, X. Z., ST LEGER, R. J. & WANG, C. S. 2011. Genome sequence of the insect pathogenic fungus Cordyceps militaris, a valued traditional chinese medicine. *Genome Biology,* 12**,** R116-136.

ZHOU, M. J., DIWU, Z. J., PANCHUKVOLOSHINA, N. & HAUGLAND, R. P. 1997. A stable nonfluorescent derivative of resorufin for the fluorometric determination of trace hydrogen peroxide: Applications in detecting the activity of phagocyte NADPH oxidase and other oxidases. *Analytical Biochemistry,* 253**,** 162-168.

ZONDEK, B. & ASCHHEIM, S. 1927. Hypophysenvorderlappen und Ovarium. *Archiv für Gynäkologie,* 130**,** 1-45.