UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND APPLIED SCIENCES

School of Chemistry

Investigations into Open Notebook Science: directed evolution as a model for exploring online laboratory notebooks.

by

Jennifer Rose Hale

Thesis for the degree of Master of Philosophy

February 2013

ABSTRACT

Directed evolution is used by protein engineers to alter the properties of an enzyme and can be considered as an uphill walk through a fitness landscape. The process is used to mimic natural evolution, though natural adaptive evolution is poorly understood. There are currently several linked theories as to how adaptive evolution occurs: through adaptation of the promiscuous functions of enzymes, conformational dynamism and neutral drift. Neutral drift is the incorporation of mutations that have little or no effect on the protein in its current environment but which may have adaptive potential if the selection pressure changes. Neutral drift may also provide alternative routes through the fitness landscape of a protein that allows it to reach new peaks of fitness not accessible by direct selection.

In this thesis the question of whether neutral drift is a useful technique to employ in directed evolution is explored. The E. coli β-glucuronidase is evolved into a β-galactosidase in four rounds of neutral drift, i.e. selecting for the native activity. Across four experiments (not all carried out as part of work for this thesis), five key mutations are returned: T509A, S557P, N566S, K568Q and W529L. These mutations are identical to mutations found in standard directed evolution experiments on this enzyme system. The libraries of mutant β-glucuronidases were screened against a variety of alternative substrates and three mutants with β-xylosidase activity were found. These are different mutations to those found in other experiments.

This thesis also explores electronic methods of recording research data including ELNs, blogs and wikis. The chapter also features the development and testing of a blog system named LaBlog. The experiments using neutral drift are used to develop and optimise the blog for molecular biology type work; though the system is flexible and can be optimised for other types of scientific work.

Additionally blogs are compared to wikis, another online recording method, in the context of the Open Notebook Science Challenge: a challenge to measure the solubilities of compounds in organic solvents.

Contents

DECLARATION OF AUTHORSHIP

I, Jennifer Rose Hale

declare that the thesis entitled

Investigations into Open Notebook Science: directed evolution as a model for exploring online laboratory notebooks

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:

  1. Smith, W. S., Hale, J. R., and Neylon, C. Applying neutral drift to the directed evolution of a beta-glucuronidase into a beta-galactosidase: two different evolutionary pathways lead to the same variant. BMC Research Notes. 2011. 4:138 (6th May).

  2. Bradley, J-C., Neylon, C., Guha, R., Williams, A., Hooker, B., Lang, A., Friesen, B., Bohinski, T., Bulger, D., Federici, M., Hale, J., Mancinelli, J., Mirza, K., Moritz, M., Rein, D., Tchakounte, C., and Truong, H. Open Notebook Science Challenge: Solubilities of Organic Compounds in Organic Solvents. 3rd Edition 2010. ISBN: 978-0-557-31801-8.

  3. Bradley, J-C., Neylon, C., Guha, R., Williams, A., Hooker, B., Lang, A., Friesen, B., Bohinski, T., Bulger, D., Federici, M., Hale, J., Mancinelli, J., Mirza, K., Moritz, M., Rein, D., Tchakounte, C., and Truong, H. Open Notebook Science Challenge: Solubilities of Organic Compounds in Organic Solvents. Available from Nature Precedings http://dx.doi.org/10.1038/npre.2010.4243.3 (2010).

  4. Milsted, A. J., Hale, J. R., Frey, J. G., and Neylon C. LabTrove: A Lightweight, Web Based, Laboratory “Blog” as a Route towards a Marked Up Record of Work in a Bioscience Research Laboratory. PLoS ONE. 2013. 8(7): e67460. doi:10.1371/journal.pone.0067460.

Signed: ……..

Date:…….

Acknowledgements

So many people have had an input somewhere in this period whether it has been in a professional or personal capacity. I am so grateful to everyone and send them my sincerest thanks.

Firstly I must thank my supervisor Dr Cameron Neylon for believing in me by taking me on to do postgraduate research in spite of my health not being at its peak at the time; additionally for providing advice, support and encouragement throughout the project. Special thanks also to Dr Wendy Smith and Dr Rob Wood. To Wendy for giving up her time specially to train me in all the techniques needed for the project, also for providing information on the project and allowing me access to her results; and to Rob for various bits of day to day advice and support in the lab. Also to lab members Michael Maynard-Smith, Lilyan Chan and Hannah Feirns for a few lab based giggles. On the computing side of the project, thanks to the Frey group for managing to answer most of my questions, and solve my computer problems.

I have been so lucky to have a network of wonderfully supportive family and friends, so thank you to my Mum and Dad, my sister Victoria, and all other members of the family for everything you have done. Also thank you to all my friends who really are too numerous to mention individually, but you hopefully all know who you are. However, special mention goes to former housemates Laura Aruparayil, Peter Russell, Naomi Clarke and Esther Fung for always being there, especially when things got tough; former colleagues Simon Gerrard and Vicki Powers for their continuous support (and plenty of laughter); and finally Dr Matthew Harmer who has provided me with accommodation on my trips to Southampton after I moved back home.
Thanks everyone I couldn't have made it this far without you!

Definitions and abbreviations

A	Adenine
A	Single letter abbreviation for Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
BBCode	Bulletin Board Code
Blog	Recognised contraction of web log
BSA	Bovine Serum Albumin
C	Cytosine
COSHH	Control of Substances Hazardous to Health
D	Single letter abbreviation for Aspartic acid
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dNTP	deoxynucleoside triphosphate
dPTP	deoxypyrene triphosphate
dTTP	deoxythymidine triphosphate
E	Single letter abbreviation for Glutamic acid
E. coli	*Escherichia coli*
*Eco*RI	*Escherichia coli* restriction endonuclease I
ELN	Electronic Laboratory Notebook
F	Single letter abbreviation for phenylalanine
FACS	Fluorescence Activated Cell Sorting
Fmoc	Fluorenylmethyloxycarbonyl: a protecting group used in organic synthesis
FPLC	Fast Protein Liquid Chromatography
G	Single letter abbreviation for glycine
G	Guanine
GFP	Green Fluorescent Protein
Glu	Glutamic Acid
GRoEL/GRoES	Name of two essential molecular chaperones used to fold proteins
GSSM	Gene Site Saturation Mutagenesis
GUS	Identifier for the β-glucuronidase gene
H	Single letter abbreviation for Histidine
HTML	Hyper Text Mark-up Language
I	Single letter abbreviation for isoleucine

IP	Intellectual Property
ISM	Iterative Saturation Mutagenesis
ISOR	Incorporating Synthetic Oligonucleotides via Gene Reassembly
ITCHY	Incremental Truncation for the Creation of Hybrid Enzymes
K	Single letter abbreviation for lysine
L	Single letter abbreviation for leucine
LaBlog	Laboratory Blog: notebook system developed by Southampton University
LB	Luria-Bertani broth
Lys	Lysine
N	Single letter abbreviation for asparagines
NEB 4	New England Biolabs buffer number 4
*Nco*I	*Nocardia coralline* restriction endonuclease I
ONP-gal	*ortho*Nitrophenyl-β-D-galactopyranoside (2-nitrophenyl-β-D-galactopyranoside)
ONP-glu	*ortho*Nitrophenyl-β-D-glucuronic acid (2-nitrophenyl-β-D-glucuronic acid)
ONS	Open Notebook Science
OPW-PCR	Overlap Primer Walk – Polymerase Chain Reaction
OSCARR	One-pot Simple Methodology for Cassette Randomisation and Recombination
P	Single letter abbreviation for proline
PC	Personal Computer
PCR	Polymerase Chain Reaction
12-pNCA	*para*Nitrophenoxydodecanoic acid
PNP-gal	*para*Nitrophenyl-β-D-galactopyranoside (4-nitrophenyl-β-D-galactopyranoside)
PNP-glu	*para*Nitrophenyl-β-D-glucuronic acid (4-nitrophenyl-β-D-glucuronic acid)
PON	Paraoxonase
Q	Single letter abbreviation for glutamine
R	Single letter abbreviation for Arginine
RACHITT	Random Chimeragenesis in Transient Templates
RID	Random insertion Deletion
Rpm	Revolutions per minute
S	Single letter abbreviation for serine
SCLE	Systematic Catalytic Loop Exchange
SDDSM	Rapid Site Directed Domain Scanning Mutagenesis
SeSaM	Sequence Saturation Mutagenesis
SeSam-Tv+	Transversion Enriched Sequence Saturation Mutagenesis
SOC	Super Optimal Broth containing added glucose. A bacterial growth medium
StEP	Staggered Extension Process
T	Single letter abbreviation for threonine

T	Thymine
TAE buffer	Tris, acetic acid, ethanol buffer
Taq	*Thermus Aquaticus*
TBBL	γ-thiobutylbutyrolactone
TBL	γ-thiobutyrolactone
TEM-1	Gene code for β-lactamase
TEMED	Tetramethylethylenediamine
THF	Tetrahydrofuran
TIM	Triose phosphate Isomerase
TriNEx	Tri-nucleotide Exchange
Tris	Tris(hydroxymethyl)aminomethane
Tyr	Tyrosine
UV	Ultraviolet
V	Single letter abbreviation for valine
W	Single letter abbreviation for tryptophan
WYSIWYG	What You See Is What You Get
X-fuco	5-chloro-4-bromo-3-indolyl-β-D-fucopyranoside
X-gal	5-chloro-4-bromo-3-indolyl-β-D-galactopyranoside
X-galam	5-chloro-4-bromo-3-indolyl-N-acetyl-β-D-galactosaminide
X-glu	5-chloro-4-bromo-3-indolyl-β-D-glucuronic acid
X-gluam	5-chloro-4-bromo-3-indolyl-N-acetyl-β-D-glucosaminide
X-gluco	5-chloro-4-bromo-3-indolyl-β-D-glucopyranoside
X-xyl	5-chloro-4-bromo-3-indolyl-β-D-xylanopyranoside
Y	Single letter abbreviation for Tyrosine

Chapter 1 - An introduction to protein engineering and library construction methods for directed evolution experiments

1.1 Protein engineering and directed evolution

1.1.1 Protein engineering and directed evolution

Nature employs proteins for a vast and diverse range of functions and uses[1]. These include structural roles e.g. proteins in hair, skin, and muscles or cellular scaffolding; receptors, cell signalling and communications e.g. G-protein coupled receptors, ligand gated ion channels, and neurotransmitters; antibodies and immunological response, and catalysis amongst many others. Enzymes are the proteins that are involved in catalysis. Enzymes are able to catalyse a wide range of chemical transformations with enhanced rates and stereo-selectivity[2]. Enzymes are thought to facilitate these transformations by lowering the activation energy of the reaction. The activation energy is lowered through stabilisation of the transition state by reversibly bonding with residues in the active site[3]. It is thought that enzyme dynamics and quantum tunnelling may also have an important role, but these effects are less well understood[4].

Enzymes are being targeted for use in industry as biocatalysts because of their ability to facilitate complex chemical transformations, which may, for example, be useful in synthesising drugs or other chemicals. Additionally enzymes are advantageous because they function under mild conditions and are energy efficient i.e. they act at room temperature and pressure and are active in water. Enzymes are also much less hazardous to the environment and produce few waste products[5]. Through natural adaptation enzymes also possess the ability to catalyse a range of structurally similar substrates; whilst others have catalytic 'promiscuity', which is the ability to catalyse mechanistically different reactions to that of the native activity[1,6]**.** The ability to catalyse a range of substrates is useful in the context of synthesis.

The usability of enzymes in industry is hindered by a number of disadvantages. This is because enzymes have been optimised through biological evolution with respect to their *in vivo* functions[7] not for the *in vitro* needs of industry and medicine. As a result many enzymes are unsuitable for use in their natural state. For example many have regulatory feedback loops which cause the enzyme to temporarily stop functioning when product accumulates, which is a useful characteristic for regulating cellular metabolites, but not ideal for a process required to run constantly[5]. Directed evolution can be used to alter the properties of enzymes to make them more suitable for industry and medicine.

Directed evolution is a method used in protein engineering for evolving different enzyme properties, for example thermostability or substrate selectivity. Directed evolution is a method of mimicking natural evolution in the laboratory. In directed evolution, genetic diversity is introduced into the DNA encoding the protein via mutation, recombination or by a combination of both methods, to produce a library of mutant genes, which are expressed, screened and selected according to the desired activity. The methods used for introducing genetic diversity are discussed in greater detail later in this chapter, whilst the overall processes are reviewed in the literature by Romero and Arnold[8]; Bershtein and Tawfik[9] and Neylon[10] amongst others.

Directed evolution is a particularly useful technique to use as it does not require knowledge of the full structure activity relationship of the protein[4, 11, 12], and can in turn provide details about the structure activity relationship. Randomly mutating the whole gene allows for interactions outside the active site to be included in generating new properties. These interactions are difficult to predict but may be important and should not be ignored[4, 13,14]. These interactions may include holding catalytic residues in the correct orientation, or ensuring the correct charge distribution throughout the molecule as a whole. Often the importance of these interactions is hard to demonstrate when the contribution of each interaction is very modest[13, 14].

There is a wealth of available literature on individual examples of directed evolution experiments, and many reviews/overviews of the reasons for carrying out these experiments, which is far too extensive to be covered in detail. Presented here are some of the main categories of directed evolution experiments provided in the literature. Dougherty and Arnold[15] review uses of directed evolution in synthetic biology including construction of synthetic genetic circuits based on transcriptional regulators e.g. bacterial toggle switches, pulse generators and oscillators; as well as engineering of non-native metabolic pathways. Wen, Nair and Zhao[16] review advances in protein engineering for biofuel production from lignocelluloses; this includes improving the catalytic efficiency of cell wall hydrolases and biofuel synthesis pathways. Turner[17] reviews advances in biocatalysis; including biosynthesis of pharmaceutical and agrochemical building blocks previously prepared by chemical synthesis. Other examples include enzymes being evolved for improved biomedical and industrial utility[14] such as thermostability in biotechnology industry[11, 18], kinetics, substrate/product specificities[19], chemozymic single enantiomer synthesis[20], enzyme fitness for synthesis, optimisation to work in organic solvents, improvements in fluorescence maturation and half life for cell imaging, catalytic antibodies[2], new restriction endonuclease cutting sites[2, 7], vaccines, and even plant improvement[19]. However enzymes are also evolved for basic research purposes[19]. Many other experiments seek to design and build scaffolds with *de novo* activities[2, 3, 5, 21- 24].

Despite directed evolution being widely used in the laboratory to mimic natural evolution; natural adaptive evolution is poorly understood even though it is a fundamental biological process[6, 25, 26]. From examination of databases of protein sequences, it can be shown that a small number of

much simpler, catalytically inefficient, broad specificity polypeptides (otherwise described as proto-enzymes[25]) diversified into the millions of distinct protein sequences in the current environment[25, 27]. However there are different theories, which are similar and can all be linked together, as to how adaptive evolution can occur.

One theory considers promiscuous activity of enzymes as a major driving force for adaptation. From this theory an enzyme, E, has evolved to recognise substrate A, but also has low level promiscuous activity towards alternative substrate B. In the event of a change in environment, activity towards substrate B confers a selective advantage. Over expression of the promiscuous function confers an improvement in fitness. Continued selection pressure evolves specificity for substrate B with eventual loss of ancestral function towards substrate A[26]. This theory can be seen in pictorial form (figure 1.1).

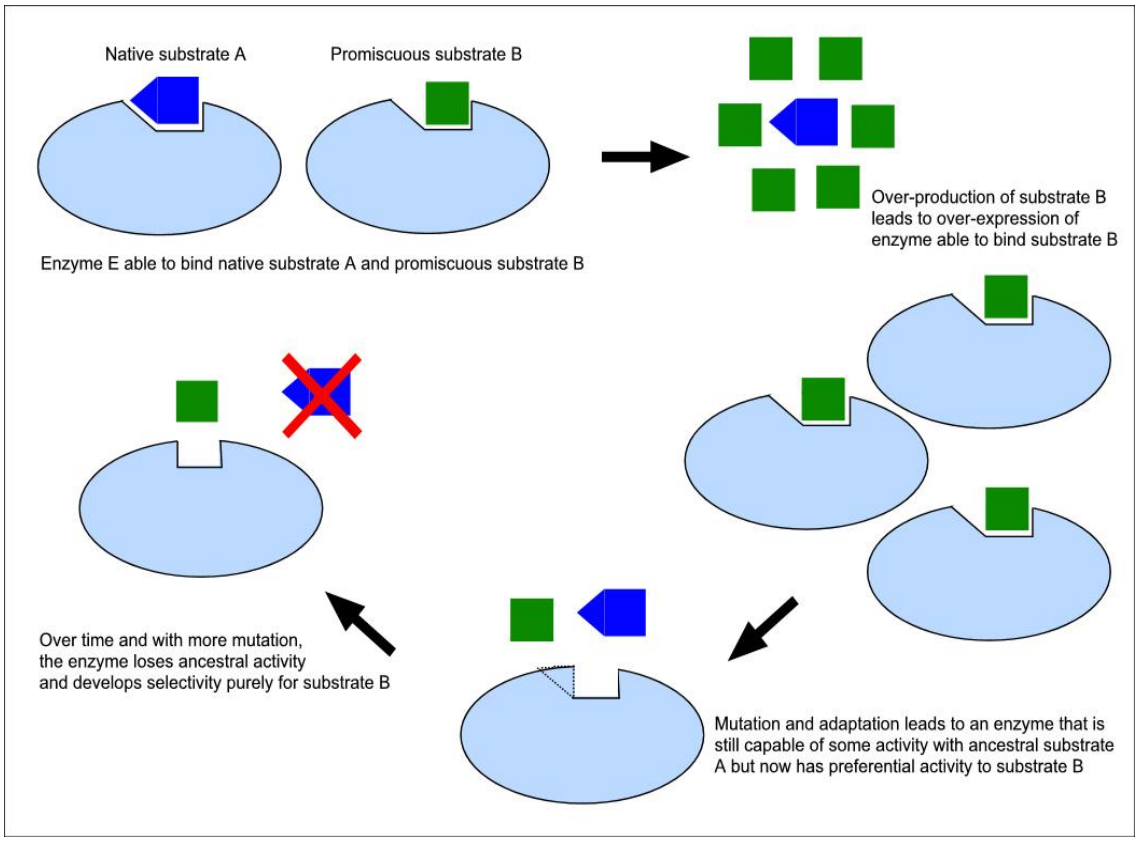


Figure 1.1: pictorial representation of adaptive evolution of an enzyme for changed substrate specificity through the promiscuous activity of the enzyme towards alternative substrates. Over production of an alternative substrate causes adaptation to the enzyme to catalyse the alternative substrate. Over time further mutation causes loss of ancestral activity.

This theory is closely linked to the theory of conformational dynamism[27, 28]. Conformational dynamism states that enzymes are not static in structure but fluctuate between a number of

different conformations in an equilibrium, with the conformation of the native activity having the most stable state. From this theory (figure 2.2) an enzyme, E, in conformation one binds native substrate A, but is able to bind substrate B when in conformation three (where conformations one to three have been arbitrarily numbered for the purposes of the diagram). A mutation to the enzyme increases the stability of conformation three and alters the equilibrium in favour of the promiscuous activity without significant loss of native activity. With time further mutation changes the selectivity to that of the alternative substrate[27, 28].

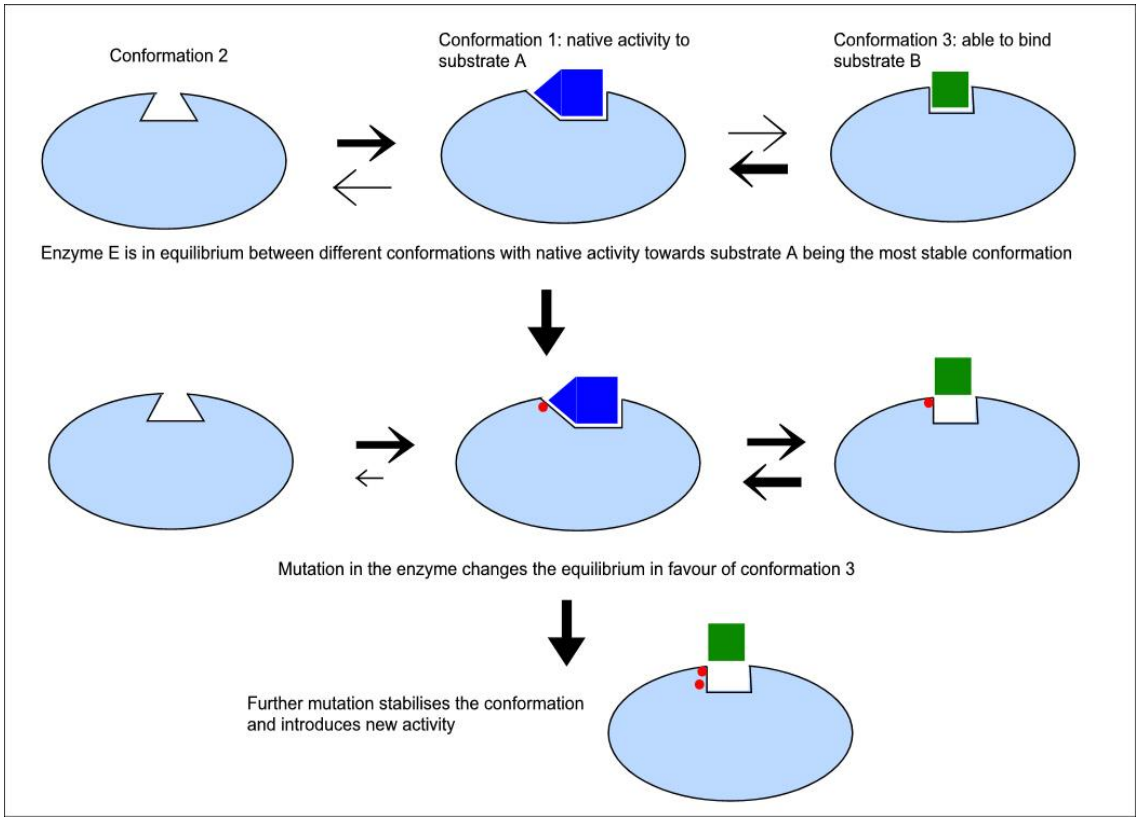


Figure 1.2: the conformational dynamism theory of enzyme evolution. An enzyme has a number of different conformations in equilibrium. A mutation in the enzyme alters the equilibrium in favour of a conformation able to bind an alternative substrate. Further mutation stabilises the conformation introducing the new activity and eventual loss of the original activity.

A further theory is Kimura's neutral theory which noted that in protein families, gene sequences accumulate mutations at an approximately constant rate over time. From this it was inferred that fixation of random neutral mutations accounted for the vast majority of sequence changes[27, 29], though subtle recalibration of protein properties also accounts for some sequence changes[1, 30]. Neutral mutations are mutations which have little or no effect on the protein in its current environment. Neutral mutations however, may have adaptive potential and allow adaptive events to occur more frequently. The theories outlined above infer that acquisition of new

protein functions occurs in steps produced by single beneficial mutations, with the potential for the initial steps to occur before selection for the new function[1, 27, 31, 32].

The adaptive evolution of enzymes is different between bacteria and eukaryotic systems. For eukaryotic systems, gene duplication may play an important role by allowing adaptation and enzyme function to be uncoupled. Gene duplication causes an extra copy of a gene to be produced in a cell and passed on to all cell progeny. This allows both a working copy and a redundant copy to be present. The working copy allows a functional enzyme to be produced and the redundant copy is free to undergo mutation and adaptation[27, 33]. However, many of these duplicates lead to loss of function[27]. Adaptive mutations in isolation are rare[1, 34] as many mutations are potentially deleterious by undermining protein stability or folding. These adaptive mutations which are deleterious in isolation may however be able to be included if other neutral mutations are first introduced into the gene; or through the introduction of stabilising residues which can bring residues into line with residues shared by the common ancestor or the consensus residue across a protein family[29, 32]. The effect that one mutation has on another in the context of protein fitness is termed an epistatic interaction.

Some proponents of the conformational dynamism theory accept the theories of neutral mutations and promiscuous functions, but propose that gene duplication cannot be important to adaptive evolution, due to the loss of function to most duplicates produced by mutation[28]. Conformational dynamism is also thought to address a paradox of adaptive evolution: that of neutrality versus plasticity. Proteins need to be robust in response to mutations to maintain structure and function (neutrality); but also need to react to mutations to adapt (plasticity).

1.1.2 Fitness landscapes and neutral drift

Evolution of proteins can be visualised in terms of a fitness landscape. A fitness landscape is a multidimensional plot of sequences versus fitness, where the fitness is defined as how well a protein performs the target function under the desired conditions[8, 31]. Within this landscape regions of higher elevations represent desirable proteins and directed evolution can allow peaks to be climbed in single uphill steps which equate to the introduction of a single beneficial amino acid mutation. Landscapes vary in appearance from the idealistic single peak Fujiyama landscape to rugged multi-peaked Badlands landscapes (figure 1.3). Fujiyama landscapes are unrealistic for real proteins due to the vast majority of mutations being deleterious to protein function, and the nature of epistatic effects[8].

Fitness landscapes are not necessarily a good visual concept to apply to evolution as they cannot be directly compared to maps of terrestrial landscapes[8]; also fitness landscapes can imply that bad enzymes lie at the base of a peak or that a protein occupies a peak and every amino acid substitution is a downhill evolution[31, 32]. An alternative depiction is the neutral network[32, 34]. A neutral network is a series of genotypes linked by point mutations that share a

common phenotype[34]. Neutral networks highlight the many starting routes and pathways to optimising the required activity which may include both beneficial and neutral mutations[32]. An example of a neutral network for variants of a PON1 enzyme is depicted by Amitai *et al*[32] as part of investigations into neutral mutations.

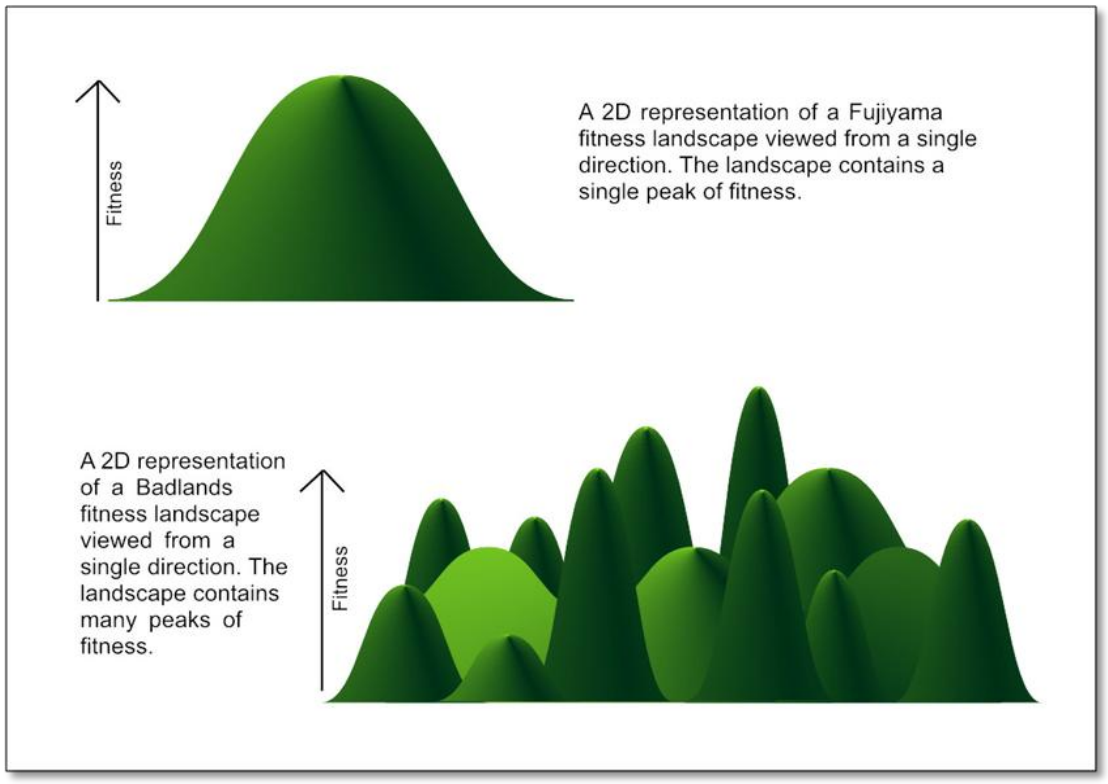


Figure 1.3: The two extremes of fitness landscapes characterised as idealistic Fujiyama landscapes and the more realistic rugged Badlands landscape. The fitness landscape is a multidimensional plot of fitness versus sequence which cannot be accurately matched to the concept of a map or earthly landscape. In this figure the landscape is viewed from a single direction.

As directed evolution experiments are considered to be climbing an uphill peak of a fitness landscape one beneficial mutation at a time, neutral mutations that could potentially allow greater sampling of sequence space and open up new routes through the landscape may be overlooked. Therefore building a directed evolution experiment around neutral drift i.e. deliberately incorporating neutral mutations that maintain native function could produce new mutants with increased fitness beyond what is attainable by standard directed evolution, by escaping local minima in the landscape.

The use of neutral drift in a directed evolution experiment forms the basis of a substantial portion of this thesis. In chapter two, an investigation into whether neutral drift can be used to aid directed evolution is carried out on the enzyme β-glucuronidase. This enzyme is a family 2

glycosyl hydrolase used to cleave a monomer of glucuronic acid from the non-reducing terminus of carbohydrate chains. It is found in both prokaryotes and eukaryotes. For this investigation, the *E. coli* β-glucuronidase is used. The enzyme is only functional as a tetramer, which is composed of four identical monomer units. Each monomer comprises three domains. The active site is located in the third domain at the bottom of a barrel structure called the α/β_8-barrel, which is a common fold found in many enzymes. The crystal structure for E. coli was only recently obtained (2010). Previous to that time, the catalytic, and other important residues were predicted by mapping the sequence onto the crystal structure of human β-glucuronidase. There is approximately 50% sequence similarity between the human and *E. coli* β-glucuronidases. Mapping the *E. coli* sequence onto the human crystal structure predicted the catalytic residues for the *E. coli* β-glucuronidase as being Glu413 and Glu504. In 2010 crystal structures for the E. coli β-glucuronidase were obtained and the catalytic residues confirmed. In the introduction to chapter two, the structure, mechanism and active site of β-glucuronidase is demonstrated in greater detail.

In the investigation, neutral drift will be used to direct the evolution of β-glucuronidase to have increased β-galactosidase activity. The neutral drift libraries are also screened for variants with increased activity towards other glycosyl hydrolases including β-xylosidase and β-fucosidase. There is precedent for increasing the activity towards these substrates, as there have been previous directed evolution experiments to change the substrate selectivity for β-galactopyranosides[12,35] and β-xylopyranosides[25]. In one study which directed the evolution towards β-galactopyranosides[12], the most active mutant was also active towards fucopyranosides. Other studies have increased the thermostability[11,19] and improved the surface chemistry of the enzyme towards aldehyde fixatives[36]. These experiments are also discussed. All the mutations are mapped onto the crystal structure and their location and role analysed. These experiments form valuable controls for the neutral drift experiments by providing information on the sites of previous mutations. The results of the mutations found in the later experiments are discussed in the context of the results of the earlier experiments in deciding whether neutral drift added value to the directed evolution, and also whether neutral drift is or is not an effective method to employ for directed evolution in this system and overall.

1.2 Library Construction methods

For any directed evolution experiment, the method in which the library is constructed is important. There are a variety of methods that can be employed and each has advantages and disadvantages. The methods used to prepare libraries of mutants can be split into two broad categories: non-recombinative and recombinative[10]. Many experiments may use a combination of both methods to construct the library. There is also a third category, computational library preparation, which is often used in tandem with the experimental approaches.

1.2.1 Non-recombinative methods

In non-recombinative methods, diversity is introduced via point mutations, substitution of single amino acids, insertion or deletion of amino acids. This can be achieved by using either site directed or random mutagenesis[10, 37].

Random mutagenesis by error-prone PCR is one of the simplest and most commonly used methods of introducing diversity (figure 1.4). Error prone PCR mimics the natural imperfect replication of DNA[32]. Error prone PCR can allow the mutation rate to be varied through the concentration of the template, the number of cycles and through the agent which causes the polymerase to make errors. Addition of Mn^{2+} instead of Mg^{2+}, the usual co-factor of the polymerase, is one method of introducing errors[10, 30], whilst biasing the dNTPs to have an over representation of dGTP and dCTP is a second method[10]. Use of non-standard dNTP analogues such as 8-oxo-deoxyguanosine triphosphate and dPTP will also introduce errors[10, 30]. Alternatively commercial mutagenesis kits may be used[10]. An example of one such kit is the Genemorph I/II™ system produced by Stratagene.

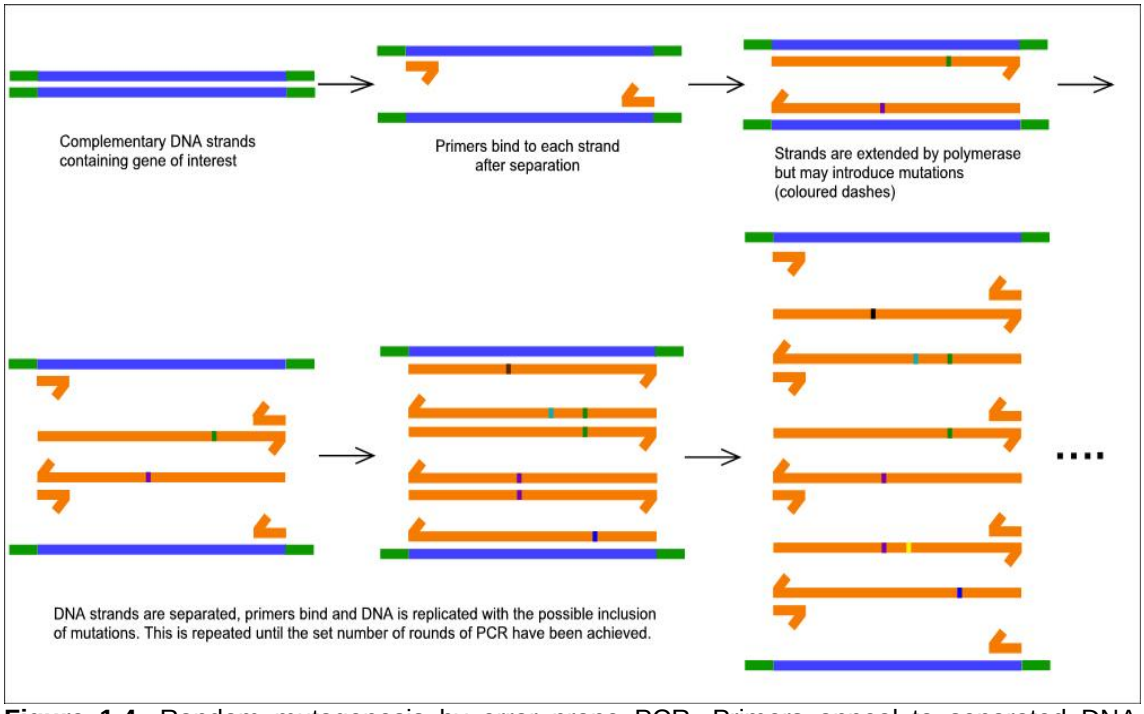


Figure 1.4: Random mutagenesis by error prone PCR. Primers anneal to separated DNA strands. Polymerase enzymes replicate that DNA strand. One or more mutations may be incorporated (coloured dashes). These mutations are carried through to the next cycle of replication building a library of mutations.

The advantages of using error prone PCR is that very little knowledge is required about the structure[11, 38]; and as the whole gene is potentially subject to mutation, it also allows mutations distant from the active site to be sampled. Additionally the low mutation rates produced increase

the probability of finding beneficial mutations, and introducing diversity through random mutagenesis can provide useful structure-function information[12]. Error prone PCR does have disadvantages too. The main disadvantage is that there is a bias to the mutations found in a library produced by error prone PCR. This is caused by several factors.

The first factor is caused by the frequency of the type of mutation that is likely to occur. Single point mutations can be classified as either transitions or transversions. Transition mutations occur when a purine base is swapped for another purine base i.e. A→G or G→A, or a pyrimidine base is swapped for the other pyrimidine base i.e. C→T or T→C. Transversions occur when a purine is swapped for a pyrimidine and vice versa i.e. A→C/T, G→C/T, C→A/G or T→A/G. As demonstrated, there are twice as many possible transversion mutations compared to transition mutations. However, transitions occur more frequently due to the mechanism by which they occur[39]. Transition mutations occur due to tautomerism in DNA bases between the amine and imine form of the base in A and G, and the keto and enol forms of the base in C and T, which allow non-standard base pairings to occur. This produces a transition mutation in the next cycle of replication. Transition mutations are more likely to produce a silent mutation[40]. This is linked to a second factor which is the inherent bias found in the genetic code, leading to certain single point mutations being more likely to lead to a silent mutation. This is because some amino acids (methionine and tryptophan) have only one codon, whilst others (leucine, valine, proline, threonine, alanine and glycine) have four or even six codons (serine and arginine).

A single base substitution leads to the accessibility of an average of only 5.6 amino acid codons[10, 37]. To access other amino acids two or even three mutations in adjacent bases are required. This bias is present and has been optimised under natural replication conditions to ensure that an amino acid change is less likely to cause loss of function[10]. A further factor is caused by an inherent bias found in the polymerases themselves. For example Taq polymerase has a bias towards mutations of A and T. This leads to a number of more common mutations being added to a library.

The final disadvantage is an amplification bias during PCR. As PCR amplifies DNA exponentially, mutations produced earlier in the amplification will be over represented compared to mutations produced in the final cycles. For example a mutation produced in round 1 of the amplification will be represented in 25% of the product[41].

The use of a commercially produced mutagenesis kit such as Genemorph II™ can help eliminate bias from the library as it contains a mixture of two polymerases: Taq, which favours A and T, and Mutazyme which favours C and G mutations. The original Genemorph™ contained only Mutazyme and thus had a bias towards C and G mutations, but was described in the literature[35, 37] as being used in conjunction with Taq mutagenesis to alleviate library bias. It

cannot alleviate the bias caused by degeneracy in the genetic code, and the unlikelihood of two or more adjacent bases being mutated to allow sampling of the full range of amino acids.

Mutations can also be introduced systematically by saturation mutagenesis. There are a number of techniques developed which use saturation mutagenesis. They fall into two groups. The first group targets a smaller number of known residues, whilst the second targets the whole gene. When more information is known about a protein and important residues have been identified, site saturation mutagenesis can be used to introduce all 19 possible amino acid substitutions to these residues. The advantage of this rational method is that a much more manageable number of variants are produced from the library. The disadvantage is that a considerable amount of information must be known about the structure-function relationship of the protein to predict which residues should be mutated[12, 38]. Alternatively site saturation mutagenesis could be used in a later experiment to further optimise sites identified as important during earlier experiments. The use of site saturation mutagenesis is demonstrated in the directed evolution of a β-galactosidase with increased β-fucosidase activity[14] and a β-glucuronidase with increased β-xylosidase activity[25].

Iterative Saturation Mutagenesis (ISM) is another newer form of saturation mutagenesis that has recently been introduced[42]. This method however, requires knowledge of the protein structure in order to determine sites of interest. In ISM, a number of sites for mutation are identified and libraries are produced by carrying out saturation mutagenesis on one site per library. The hits from each library are selected and taken as the templates for a second round of saturation mutagenesis during which a new saturation site is selected. This is continued until each site has been systematically mutated. The technique is also successful when degenerate NNK codons (N = A/C/G/T; K = G/T) or non-degenerate NDT codons (D = A/G/T; T = T) which code only 12 amino acids are used.

It is also possible to introduce all 19 possible amino acid substitutions for every position in a gene using a technique called Gene Site Saturation Mutagenesis™ (GSSM™)[43]. In this technique amino acid substitutions are systematically prepared for every position in a gene using a set of primers containing 32 or 64 degenerate codons of the NNK or NNN system respectively, where N is any base and K is G or T. Unlike site saturation mutagenesis which relies on some knowledge of the structure-function relationship to identify potential mutation sites, GSSM™ does not require this information and is able to sample all sites equally leading to a more balanced library composition. It also demonstrates positions that are more open to mutation and thus provides information on the structure-function relationship.

Sequence Saturation Mutagenesis (SeSaM)[44] is a further technique that has been introduced to aid library construction. The technique uses the non-standard base deoxyinosine to generate random mutations as deoxyinosine is able to form a bond with any of the bases. This allows the

incorporation of random mutations throughout the gene. A new variant of the technique, Transversion-enriched Sequence Saturation Mutagenesis (SeSaM-Tv(+))[45] has been developed to counteract the bias of PCR based library construction by enhancing the number of transversion mutations.

There have been a large number of techniques for site directed mutagenesis have been developed[46]. These techniques ISOR (Incorporating Synthetic Oligonucleotides via Gene Reassembly); OSCARR (One-pot Simple methodology for Cassette Randomisation and Recombination); OPW-PCR (Overlap Primer Walk – Polymerase Chain Reaction); and SDDSM (Rapid Site Directed Domain Scanning Mutagenesis) have very little use due, but have all been developed with the aim of increasing functional diversity in mutant libraries.

Diversity can also be introduced into a library through the insertion and deletion of nucleotides. However, this method is much less common as a method for library construction as insertions and deletions can lead to frame shifts. A frame shift is an effect of the addition or deletion of a base on the reading frame of a DNA sequence, which is used to translate the DNA into protein (figure 1.5). A frame shift changes the amino acid sequence in the translated protein. In many libraries this leads to a large number of truncated, non-functional proteins. However, if whole amino acids can be inserted or deleted the library may produce new and interesting variants. It is possible to produce a library prepared through insertion and deletion using a complex multi-step procedure called Random Insertion Deletion (RID)[10, 41]. This can be used to delete and subsequently insert an arbitrary number of bases and sequences at random along a target gene. This method is much more controlled than the normal random generation of insertions and deletions.

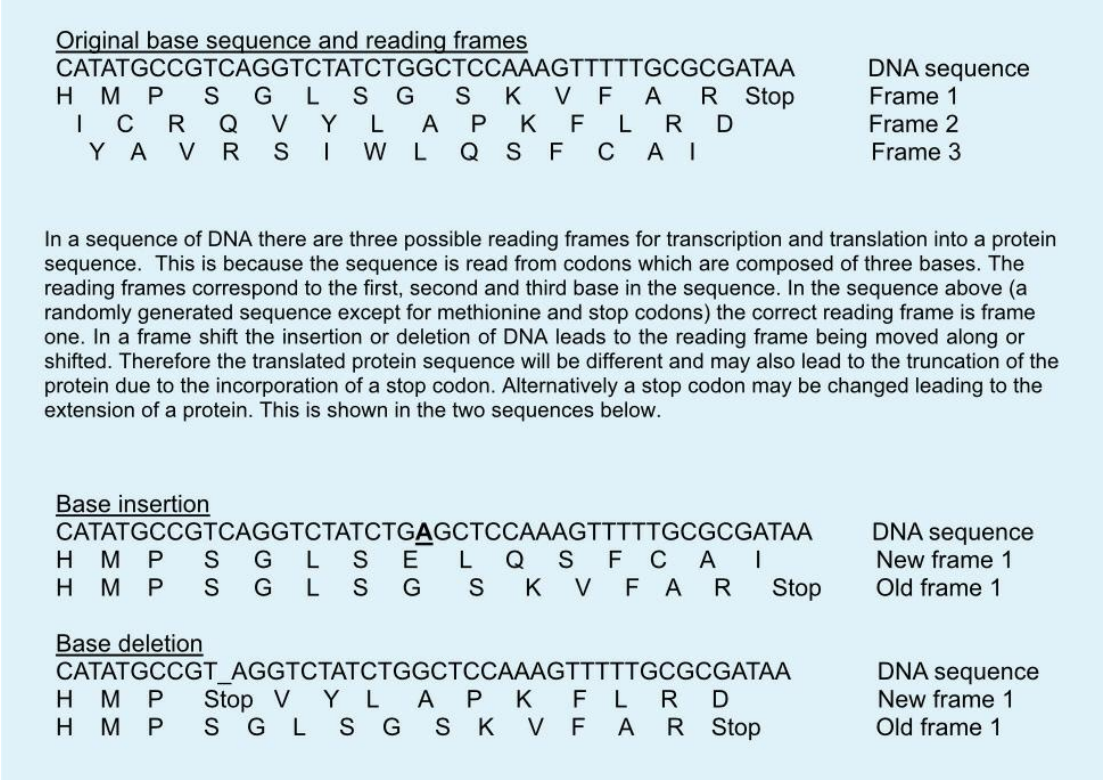


Figure 1.5: explanation and demonstration of the effects of insertion and deletion of DNA bases on the reading frames of a DNA sequence, and on the translated protein of the correct reading frame. The first part of the diagram shows a demonstration DNA sequence and the amino acid sequence translated from each reading frame. The second part of the diagram shows the effect of an additional base (A bold and underlined) into the sequence, and the third part, the effects of the deletion of a base (C represented by an underscore) from the sequence.

One new technique being reported[47] is that of TriNEx, or tri-nucleotide exchange. This is a method that does not use PCR and has similarities to RID and saturation mutagenesis, but is able to sample the whole length of the gene not just selected residue targets. TriNEx allows the deletion and replacement of three contiguous nucleotides with a randomised sequence. In 2/3 of these sequence exchanges, two codons are altered allowing a wider range of molecular diversity to be explored which may not occur through other forms of mutagenesis. As PCR is not used, TriNEx is easier to implement than RID. TriNEx has been shown to effectively sample the whole gene and is less prone to the bias inherent to mutation introduction based on PCR. Modified TriNEx approaches have also been used for whole domain insertion[48, 49] and single amino acid deletions and insertions of non-natural amino acids[50]. In the former example the deletion of three contiguous nucleotides was followed by the insertion of coding sequences for a different protein to produce a library of functional chimeric fusion proteins. In the latter approach contiguous nucleotides were replaced with a TAG stop codon that initiated the addition of a non-natural amino acid in the protein chain.

Of the techniques covered in the non-recombinative library construction category, the most commonly used techniques are error prone PCR and site saturation/site directed mutagenesis.

1.2.2 Recombinative methods

Recombinative construction methods take fragments of genes and recombine them to produce new whole genes in a method that mimics natural homologous recombination[37] (sexual selection[10]). The most popular method is DNA shuffling in which the DNA is fragmented using DNase I, and reassembled under carefully controlled conditions, first without primers to recreate whole gene sequences, then with primers to amplify the genes in a standard PCR reaction (figure 1.6). DNA shuffling can be used alone as a method of constructing the library, or as a method to further optimise the best mutants previously selected from libraries constructed via other methods. The advantages of DNA shuffling are that DNA shuffling may allow beneficial mutations to combine to produce a new improved gene, and also to remove deleterious mutations. Equally DNA shuffling can also combine less beneficial mutations.

One disadvantage of DNA shuffling is that it can generate tight population bottlenecks caused by very fit recombinants, which can limit diversity. However DNA shuffling is still able to enhance the effectiveness of directed evolution by preventing beneficial mutations driving each other to extinction[35]. In the experiment described[35], mutations identified as beneficial for substrate specificity drove each other to extinction in the absence of DNA recombination (sexual selection).

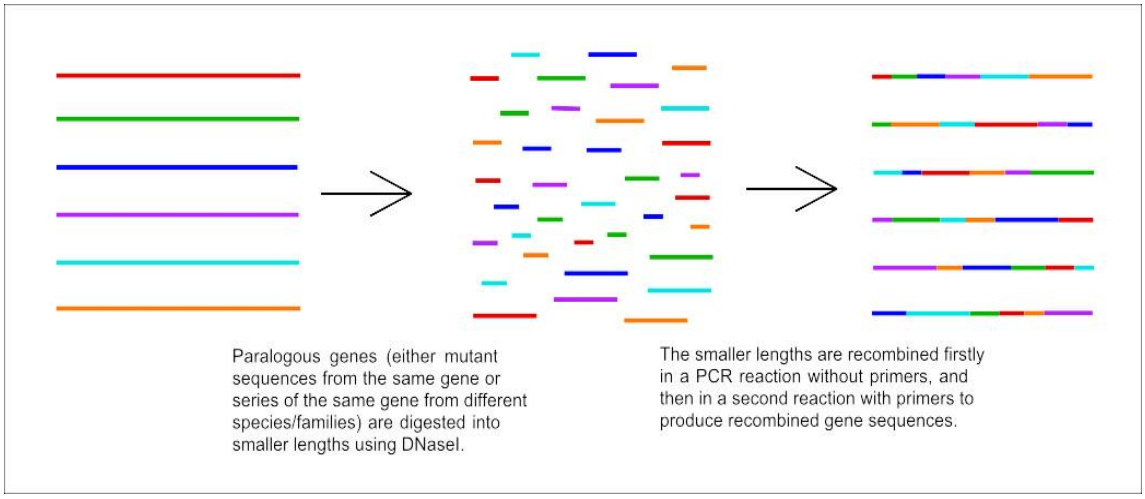


Figure 1.6: DNA shuffling. A series of genes are digested into smaller pieces and re-combined to produce new combined gene sequences. The recombined genes combine beneficial mutations to produce fit recombinants and prevent beneficial mutations driving each other to extinction.

Recently there have been a number of other variant shuffling techniques developed. Kamondi[18] used rational family shuffling to combine two closely related xylanases from the same organism, which live at different ambient temperatures to produce chimeras with increased thermostability. Family shuffling was also used by Kheronsky[51] alongside site directed mutagenesis to introduce consensus or ancestor mutations (the most commonly occurring residue across a family and thought to increase stability) to PON3 enzymes from human, mouse and rabbit. The PON3 variants which were evolved from the introduction of consensus and ancestor mutations had developed new promiscuous activities, without compromising the native activity.

Ochya-Levya[23] developed a technique called Systematic Catalytic Loop Exchange (SCLE) which exchanges the loops found in $(\alpha/\beta)_8$-barrel (TIM barrel) folds. In this technique, one $(\alpha/\beta)_8$-barrel enzyme was selected and a further eight $(\alpha/\beta)_8$-barrel enzymes were selected with wide ranging functions. The loops from these enzymes involved in binding and catalysis were mixed in an attempt to develop brand new protein activities. However, no functional screening was carried out to measure the overall success of the technique. If new activities can be produced then this could prove to be a very important technique for future directed evolution experiments as $(\alpha/\beta)_8$-barrel folds are found in 10% of all enzyme structures and across five of the six classes of enzyme.

Other less commonly used methods for recombinative construction include Staggered Extension Process (STeP) which is a PCR with short elongation times. Partially elongated strands melt and anneal to new templates causing crossovers; Random Chimeragenesis in Transient Templates (RACHITT) and Incremental Truncation for the Creation of Hybrid Enzymes (ITCHY). For laboratories with DNA and organic synthesis experience, trinucleotide phosphoramidite and Fmoc oligonucleotide synthesis can be used. This allows synthesis of sequences with all possible 20 amino acids at a given position. However it also allows the introduction of stop codons[10].

Other methods of introducing mutations include chemical treatment of the DNA with hydroxylamine, and directly damaging the DNA with UV radiation or alkylating agents. In both of these approaches the damaged DNA is incorrectly replicated or repaired leading to the inclusion of mutations. Alternatively mutator strains of bacteria can be used. Mutator strains have defects in one or more DNA repair pathways. Genetic material passing through mutator strains accumulate mutations at higher rates than normal. It has an advantage of being a very simple method of generating mutations, but does have some disadvantages. Firstly mutator strains are extremely slow methods of preparing libraries and secondly, in common with direct damage of DNA, mutagenesis is not confined to the gene of interest. Therefore mutations will also occur in the construct holding the gene and the chromosomal DNA of the host[10, 41].

1.2.3 Other considerations

As well as the method of introducing mutations to the library, consideration as to the rate of mutation and the types of mutation required in the library must also be considered. A new study by Rasila[52] focuses on the number and types of mutations at the nucleotide and amino acid levels by error prone PCR using Taq polymerase; Mutazyme polymerase (in the form of the Genemorph kits); XL1 red mutator strain and hydroxylamine. The findings show that error prone PCR and mutazyme are the most useful for low to medium mutation rate libraries with mutazyme providing the least bias. Hydroxylamine and XL1 red mutator strain can be used when a very low mutation rate is required.

Tokuriki and Tawfik[53, 54] developed a technique to improve the efficiency of the prepared library. The technique increases the number of mutants that survive during a directed evolution experiment using chaperone proteins, chaperonins GroEL and GroES. Chaperonins are a class of protein that aids correct protein folding. It was found that co-expression of the mutants with GroEL and GroES led to the rescue of misfolded mutants and aided their correct folding increasing the number of mutants in the library expressing, and fixing, otherwise deleterious mutations. This technique could prove useful for overcoming some epistatic interactions in which a mutation that is deleterious in isolation is found to be beneficial in the context of other mutations.

1.2.4 Computational libraries

Alternatively libraries can be prepared computationally and the results of these libraries can be used to guide the best mutants to prepare biologically. When detail is already known about the structure and function of a protein, sequence changes and their effects can be predicted using computer algorithms for example Monte Carlo simulations[38]. This can ensure that the practical preparation of the library by site directed mutagenesis produces the most efficient set of mutants and minimises the number of non-functional mutants. Computational protein design methods include structure based modelling of electrostatics, packing, other interactions, as well as data mining in sequence space. Computation can also be used to estimate the number of sequences that will fold to a given protein structure[55]. However computational modelling can omit potentially significant contributions of enzyme dynamics and quantum tunnelling[4].

Computational methods can also be used for the design of novel proteins which have never been seen before and have valuable applications in medicine, research and industry, and this is the ultimate aim of protein engineers. Damborsky[3] and Butterfoss[24] provide in depth reviews of computational methods and advances in computational design such as designing novel protein functions through the introduction of amino acids essential for catalysis into existing protein scaffolds. To do this, the transition states of reactions can be modelled and libraries of protein scaffolds are searched to identify potential binding pockets that tightly bind the transition state, which also retain the required geometry of the functional groups. This job has been aided by the

development of several computer programs. The principles of how each program function are given in the reviews.

In one specific example, a retro-aldolase enzyme was designed that was able to break a carbon-carbon bond in a substrate not found in biological systems[56]. The retro-aldolase was designed using an ensemble of transition states and intermediates predicted to occur from each part of the chemical transformation. In the first step the mechanism was predicted and amino acids able to carry out each step of the mechanism were selected along with water molecules. In the second step, scaffolds able to accommodate the transition states were selected. An extremely large number of scaffolds were scanned and sorted computationally using a program called Rosetta to produce 72 outcomes which fitted energetically and geometrically. Genes for the 72 designs were synthesised and cloned into *E. coli*. From this the protein was expressed, and purified. The proteins had four different active site designs of which functional retro-aldolases were found from two of the designs. These were based on TIM (triose phosphate isomerase) barrels and jelly roll folds. Kinetic parameters were obtained for the synthetic enzymes, but these were many orders of magnitude lower than a natural enzyme. In this example, directed evolution was not used to further optimise the designed enzymes.

In a second example an enzyme was designed to carry out the Kemp elimination[57]. The Kemp elimination is a key synthetic step used for the removal of a proton from a carbon. However there is no enzyme that carries out this process. In a similar manner to the previous example, mechanistic routes and possible amino acids were identified. From this starting point transition states and intermediates were predicted, and possible active sites were built around them by positioning residues to maximise transition state stabilisation and aid the proposed mechanism. The RosettaMatch algorithm was then employed to search for suitable scaffolds able to accommodate the idealised active site. A wide number of protein conformations were able to accommodate the idealised active site, but overall the optimised design focussed on the TIM barrel. TIM barrels are widespread in naturally occurring enzymes. It was therefore not surprising that the potential answer would come from this protein structure. Eight active results were screened and kinetic parameters obtained. The most active design underwent seven rounds of directed evolution to improve the catalytic ability. The observed residue changes often occurred in the residues adjacent to those involved in catalysis and are thought to alter interactions and stability.

Computational design has both advantages and disadvantages. Computational design has an advantage in being able to sample a vast amount of sequence space not immediately accessible from traditional directed evolution experiments. As folding and interactions in the active site can be predicted, non-functional proteins will be eliminated, improving the number of positive hits prepared in the laboratory. For *de novo* design, computational methodology is also able to scan databases of known enzyme scaffolds and folding patterns much more quickly and

effectively than humans alone could. One major disadvantage is that by purely focussing on building an active site, the cumulative effects on catalysis of residues further away from the active site is ignored. The individual effects of these residues may be small and hard to predict but may be important to overall enzyme stability and geometry. The advantage of combining directed evolution with the computational approach is that beneficial mutations to these distant residues are highlighted, which may allow their effects to be predicted and used in future computational design.

1.3 Presentation of experiments and experimental data: the need for accurate record keeping

It is not just the experiments themselves that are important. Any experiment, from any discipline must report the results clearly and accurately in a manner that will allow others to follow and repeat the experiments. This thesis also seeks to cover the different methods in which data can be displayed and the reasons for each method as well as its advantages and disadvantages. Currently data can be displayed and stored on paper, or electronically in Electronic Laboratory Notebooks, wikis or blogs. In chapter four the main features, advantages and disadvantages of each system are discussed. The concept of Open Notebook Science is introduced: explaining how it contributes to transparency in working methods and reporting results, but also how ONS is a concern to many scientists, and the problems, and potential problems that can arise from ONS. This chapter focuses on the development of an electronic lab book in the form of a blog, named LaBLog. LaBLog is different to the way blogs are conventionally used in science, by being the primary notebook. The blog was used to record the data for the directed evolution experiments and the development of the blog to be suitable for use in this, and any scientific context is demonstrated. The advantages that using the blog has brought are also discussed. Likewise the points during the development when the blog did not help research are also noted. Whilst the blog was optimised for use in recording molecular biology experiments, the blog is not a specialised piece of software and can be applied to any experimental context. Though the blog is also a project to demonstrate the benefits of Open Notebook Science, the need for privacy and security for certain work is recognised and the ability of the blog to be a closed system is also highlighted.

In chapter five, a specific example of experimental capture and Open Notebook Science is examined through the ONS Challenge. The Open Notebook Science Challenge does not use the directed evolution experiments but a separate chemistry based project on the measurement of the solubility of organic compounds in non-aqueous solvents. The challenge uses a wiki as the method of recording the results. The project is separate to the lab blog. In this chapter the different methods i.e. blog and wiki are compared and contrasted. By using both systems the advantages and disadvantages of each system can be reasoned. The implications of Open Notebook Science are discussed in this context.

1.4 Aims and objectives

This project aims to test the theory that using neutral drift is more effective at finding variants of β-glucuronidase with increased β-galactosidase activity than by carrying out a standard directed evolution.

To do this a library of variants will be prepared by introducing mutations into the gene through error prone PCR using the GenemorphII Random Mutagenesis Kit. The plasmid DNA will be cloned into *E. coli* and screening against the native substrate mimic 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-glu) which produces a blue precipitate in colonies carrying a functioning β-glucuronidase. Blue colonies, i.e. variants that maintain native activity will be selected and subjected to further mutation. After four rounds of mutation and selection, the resulting library will be screened against the alternative substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). Colonies which are blue will be selected for kinetic analysis and will be sequenced. Sites of mutation will be compared to previous experiments which carried out the directed evolution by the normal means of selection for the new property after each round of mutation. The kinetics of the prepared mutants will also be compared to those of earlier directed evolution experiments to determine whether the variants have increased activity, and in turn whether neutral drift has been more successful than standard directed evolution.

This project also aims to investigate the effect of Open Notebook Science on working practices through the development and testing of a laboratory blog. The results of the neutral drift experiments will be used to test the capabilities of the blog as a tool for enhancing research. All experimental work prepared for the neutral drift project will be added to the blog to test how best to use the blog, and to influence the future direction of blog design. The practice of Open Notebook Science is also explored through the use of a wiki in a separate challenge. The use of the wiki is compared and contrasted with the use of the blog.

A further project seeks to determine whether galactose and/or the synthetic compound 2, 2, 3, 3-tetrafluorogalactose are inhibitors of the enzyme β-galactosidase. To do this various assay methodologies will be tested to find a suitable assay method and to gain meaningful results.

1.5 References

1. Bloom, J. D., Romero, P. A., Lu, Z., and Arnold, F.H.; *Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution*; Biology Direct; 2007; **2**; 17.
2. Tao, H. and Cornish, V. W.; *Milestones in directed enzyme evolution*; Current Opinion in Chemical Biology; 2002; **6**; 858-864.
3. Damborsky, J. and Brezovsky, J.; *Computational tools for designing and engineering biocatalysts*; Current Opinion in Chemical Biology; 2009; **13**; 26-34.
4. Dalby, P. A.; *Optimising enzyme function by directed evolution*; Current Opinion in Structural Biology; 2003; **13**; 1-6.
5. Arnold, F. H.; *Combinatorial and computational challenges for biocatalyst design*; Nature; 2001; **409**; 253-257.
6. O'Loughlin, T. L., Patrick, W. M., and Matsumura, I.; *Natural history as a predictor of protein evolvability*; Protein Engineering, Design & Selection; 2006; **19**; 439-442.
7. Lanio, T., Jeltsch, A., and Pingoud, A.; *Towards the Design of Rare Cutting Restriction Endonucleases: Using Directed Evolution to Generate Variants of Eco*RV *Differing in Their Substrate Specificity by Two Orders of Magnitude*; Journal of Molecular Biology; 1998; **283**; 59-69.
8. Romero, P. A. and Arnold, F. H.; *Exploring protein fitness landscapes by directed evolution*; Nature Reviews: Molecular and Cellular Biology; 2009; **10**; 866-876.
9. Bershtein, S. and Tawfik. D. S.; *Advances in laboratory evolution of enzymes*; Current Opinion in Chemical Biology; 2008; **12**; 151-158.
10. Neylon, C.; *Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution*; Nucleic Acids Research; 2004; **32**; 1448-1459.
11. Flores, H. and Ellington, A. D.; *Increasing the Thermal Stability of an Oligomeric Protein, Beta-glucuronidase*; Journal of Molecular Biology; 2002; **315**; 325-337.
12. Matsumura, I. and Ellington, A. D.; *In vitro Evolution of Beta-glucuronidase into a Beta-galactosidase Proceeds through Non-specific Intermediates*; Journal of Molecular Biology; 2001; **305**; 331-339.
13. Oue, S., Okamoto, A., Yano, T., and Kagamiyama, H.; *Redesigning the Substrate Specificity of an Enzyme by Cumulative Effects of the Mutations of Non-active Site Residues*; Journal of Biological Chemistry; 1999; **274**; 2344-2349.
14. Parikh, M. R. and Matsumura, I.; *Site-saturation Mutagenesis is more Efficient than DNA Shuffling for the Directed Evolution of β-Fucosidase from β-Galactosidase*; Journal of Molecular Biology; 2005; **352**; 621-628.
15. Dougherty, M. J. and Arnold, F. H.; *Directed evolution: new parts and optimized function*; Current Opinion in Biotechnology; 2009; **20**; 486-491.

16. Wen, F., Nair, N. U., and Zhao, H.; *Protein engineering in tailored enzymes and microorganisms for biofuels production*; Current Opinion in Biotechnology; 2009; **20**; 412-419.
17. Turner, N. J.; *Directed evolution drives the next generation of biocatalysts*; Nature Chemical Biology; 2009; **5**; 567-573.
18. Kamondi, S., Szilágyí, A., Barna, L., and Závodszky, P.; *Engineering the thermostability of a TIM-barrel enzyme by rational family shuffling*; Biochemical and Biophysical Research Communications; 2008; **374**; 725-730.
19. Xiong, A-S., Peng, R-H., Cheng, Z-M., Li, Y., Liu, J-G., Zhuang, J., Gao, F., Xu, F., Qiao, Y-S., Zhang, Z., Chen, J-M., and Yao, Q-H.; *Concurrent mutations in six amino acids in β-glucuronidase improve its thermostability*; Protein Engineering, Design and Selection; 2007; **20**; 1-7.
20. El Hawrani, A. S., Sessions, R. B., Moreton, K. M., and Holbrook, J. J.; *Guided Evolution of Enzymes with New Substrate Specificities*; Journal of Molecular Biology; 1996; **264**; 97-110.
21. Arnold, F. H., Wintrode, P. L., Miyazaki, K., and Gershenson, A.; *How enzymes adapt: lessons from directed evolution*; Trends in Biochemical Sciences; 2001; **26**; 100-106.
22. Koide, S.; *Generation of new protein functions by nonhomologous combinations and rearrangements of domains and modules*; Current Opinion in Biotechnology; 2009; **20**; 398-404.
23. Ochoa-Levya, A., Soberón, X., Sánchez, F., Argüello, M., Montero-Morán, G., and Saab-Rincón, G.; Protein Design through Systematic Catalytic Loop Exchange in the $(\alpha/\beta)_8$ Fold; Journal of Molecular Biology; 2009; **387**; 949-964.
24. Butterfoss, G. L. and Kuhlman, B.; *Computer-Based design of Novel Protein Structures*; Annual Review of Biophysical and Biomolecular Structures; 2006; **35**; 49-65.
25. Geddie, M. L. and Matsumura, I.; *Rapid Evolution of β-Glucuronidase Specificity by Saturation Mutagenesis of an Active Site Loop*; Journal of Biological Chemistry; 2004; **279**; 26462-26468.
26. Patrick, W. M., and Matsumura I.; *A Study in Molecular Contingency: Glutamine Phosphoribosylpyrophosphate Amidotransferase is a Promiscuous and Evolvable Phosphoribosylanthranilate Isomerase*; Journal of Molecular Biology; 2008; **377**; 323-336.
27. DePristo, M. A.; *The subtle benefits of being promiscuous: adaptive evolution potentiated by enzyme promiscuity*; HFSP Journal; 2007; **1**; 94-98.
28. Tokuriki, N. and Tawfik, D. S.; *Protein Dynamism and Evolvability*; Science; 2009; **324**; 203-207.
29. Bershtein, S., Goldin, K., and Tawfik, D. S.; *Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins*; Journal of Molecular Biology; 2008; **379**; 1029-1044.

30. Matsumura, I. and Ellington, A. D.; *Mutagenic PCR of Protein-Coding Genes for In Vitro Evolution*; Methods in Molecular Biology 182: In Vitro Mutagenesis, 2nd Edition. Edited by Braman, J.: 2001.
31. Tracewell, C. A. and Arnold, F. H.; *Directed enzyme evolution: climbing fitness peaks one amino acid at a time*; Current Opinion in Chemical Biology; 2009; **13**; 3-9.
32. Bloom, J. D. and Arnold, F. H.; *In the light of directed evolution: Pathways of adaptive protein evolution*; Proceedings of the National Academy of Sciences; 2009; 9995-10000.
33. Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P.; *Essential Cell Biology*; 2nd Edition; 2004; ISBN 0-8153-3481-8; pp 297
34. Amitai, G., Gupta, R. D., and Tawfik, D. S.; *Latent evolutionary potentials under the neutral mutational drift of an enzyme*; HFSP Journal; 2007; **1**; 67-78.
35. Rowe, L. A., Geddie, M. L., Alexander, O. B., and Matsumura, I.; *A Comparison of Directed Evolution Approaches Using the β-Glucuronidase Model System*; Journal of Molecular Biology; 2003; **332**; 851-860.
36. Matsumura, I., Wallingford, J. B., Surana, N. K., Vize, P. D., and Ellington, A. D.; *Directed evolution of the surface chemistry of the reporter enzyme β-glucuronidase*; Nature Biotechnology; 1999; **17**; 696-701.
37. Williams, G. J., Nelson, A. S., and Berry, A.; *Directed evolution of enzymes for biocatalysis and the life sciences*; Cellular and Molecular Life Sciences; 2004; **61**; 3034-3046.
38. Polizzi, K. M., Parikh, m., Spencer, C. U., Matsumura, I., Lee, J. H., Realff, M. J., and Bommarious, M.; *Pooling for Improved Screening of Combinatorial Libraries for Directed Evolution*; Biotechnology Progress; 2006; **22**; 961-967.
39. Carr, S. M.; *Transition vs transversion mutations*; 2010; http://www.mun.ca/biology/scarr/Transitions_vs_Transversions.html. Accessed: 2013-02-21. (Archived by WebCite® at http://www.webcitation.org/6EbZy0tWO).
40. Carr, S. M.; *Induction of transition mutations by spontaneous tautomeric shifts*; 2011; http://www.mun.ca/biology/scarr/Transition_mutations.html. Accessed: 2013-02-21. (Archived by WebCite® at http://www.webcitation.org/6EbaQzSbl).
41. Neylon, C.; *Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: Library construction methods for directed evolution.* Updated revised edition of reference 10 at OpenWetWare.org http://openwetware.org/wiki/Reviews/Directed_evolution/Library_construction. Accessed: 20.02.13. (Archived by WebCite® at http://www.webcitation.org/6Ea5BU8Lp)
42. Reetz, M. T., Kahakeaw, D. and Sanchis, J.; *Shedding light on the efficacy of laboratory evolution based on iterative saturation mutagenesis*; Molecular Biosystems; 2009; **5**; 115-122.

43. Kretz, K. A., Richardson, T. H., Gray, K. A., Robertson, D. E., Tan, X., and Short, J. M.; *Gene Site Saturation Mutagenesis: A Comprehensive Mutagenesis Approach*; Methods in Enzymology; 2004; **388**; 3-11.
44. Wong, T. S., Tee, K. L., Hauer, B., and Schwaneberg, U.; *Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution*; Nucleic Acids Research; 2004; **32** (3); e26.
45. Wong, T. S., Roccatano, D., Loakes, D., Tee, K. L., Schenk, A., Hauer, B., and Schwaneberg, U.; *Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): a random mutagenesis method with consecutive nucleotide changes that complements the bias of error-prone PCR*; Biotechnology Journal; 2008; **3** (1); 74-82.
46. Shivange, A. V., Marienhagen, J. Mundhada, H., Schenk, A., and Schwaneberg, U.; *Advances in generating functional diversity for directed protein evolution*; Current Opinion in Chemical Biology; 2009; **13**; 19-25.
47. Baldwin, A. J., Busse, K., Simm, A. M., and Jones, D. D.; *Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx)*; Nucleic Acids Research; 2008; **36**; e77.
48. Edwards, W.R., Busse, K., Allemann, R. K., and Jones, D. D.; *Linking the functions of unrelated proteins using a novel directed evolution domain insertion method*; Nucleic Acids Research; 2008; **36** (13); e78.
49. Edwards, W. R., Williams, A. J., Morris, J. L., Baldwin, A. J., Allemann, R. K., and Jones, D. D.; *Regulation of β-Lactamase by Remote Binding of Heme: Functional Coupling of Unrelated Proteins through Domain Insertion*; Biochemistry; 2010; 49 (31); 6541-6549.
50. Baldwin, A. J., Arpino, J. A. J., Edwards, W. R., Tippmann, E. M., and Jones, D. D.; *Expanded chemical diversity sampling through whole protein evolution*; Molecular BioSystems; 2009; **5**; 764-766.
51. Kheronsky, O. K., Rosenblat, M., Toker, L., Yacobson, S., Hugenmatter, A., Silman, I., Sussman, J. L., Aviram, M., and Tawfik, D. S.; *Directed Evolution of Serum Paraoxonase PON3 by Family Shuffling and Ancestor/Consensus Mutagenesis, and its Biochemical Characterization*; Biochemistry; 2009; **48**; 6644-6654.
52. Rasila, T. S., Pajunen, M. I., and Savilahti, H.; *Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment*; Analytical Biochemistry; 2009; **388**; 71-80.
53. Tokuriki, N. and Tawfik, D. S.; *Chaperonin overexpression promotes genetic variation and enzyme evolution*; Nature; 2009; **459**; 668-673.
54. Matsumura, I. and Ivanov, A. A.; *How evolving enzymes can beat the heat and avoid defeat*; Nature Chemical Biology; 2009; **5**; 538-539.
55. Armstrong, K. A. and Tidor, B.; *Computationally Mapping Sequence Space To understand Evolutionary Protein Engineering*; Biotechnology Progress; 2008; **24**; 62-73.

56. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Röthlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas III, C. F., Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D.; *De Novo Computational Design of Retro-Aldol Enzymes*; Science; 2008; 319; 1387-1391.
57. Röthlisberger, 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., and Baker, D.; *Kemp elimination catalysts by computational design*; Nature; 2008; 453; 190-195.

Chapter 2 – Investigating the effectiveness of employing neutral drift in the directed evolution of a β-glucuronidase into a β-galactosidase

2.1 Introduction

2.1.1 The structure and function of β-glucuronidase

β-glucuronidase (GUS) is an enzyme produced in both bacterial and eukaryotic systems. β-glucuronidases remove a glucuronic acid unit from a wide variety of polysaccharide compounds. In bacterial species glucuronic acid is used as a carbon source to produce energy for the bacteria, but also has a role in carbohydrate metabolism, vitamin production, and processing bile acids, sterols and xenobiotics in the gut[1].

β-glucuronidase belongs to a group of enzymes called the glycoside hydrolases. The glycoside hydrolases are split into 91 families classified according to sequence similarities. β-glucuronidase belongs to family 2 and also to a subclass called the GH-A clan (4/7 super family) a multifamily class[2]. All the enzymes in this clan retain the same $(\alpha/\beta)_8$ barrel fold (also known as the TIM barrel) and catalytic mechanism, which either involves retention or inversion of configuration[3]. β-glucuronidase acts to produce a product that retains configuration.

Approximately 10% of proteins belong to the $(\alpha/\beta)_8$ barrel fold structural family, which have a wide range of catalytic functions and non-enzymatic roles. The $(\alpha/\beta)_8$-barrel fold is formed from a basic motif of ~ 250 residues. The $(\alpha/\beta)_8$-barrel is a funnel shaped pocket created by loops connecting the C terminal end of β-strands with the amino terminal of the eight α-helices that form the barrel. The active site is usually located at the bottom of the pocket. The substrate binding residues are predominantly within the barrel whilst the catalytic residues are predominantly within the connecting loop regions[4]. In GH-A enzymes, loop 8 is responsible for the specificity of the C5 substituent[2].

β-glucuronidase is a homo-tetrameric protein formed from two asymmetric units each containing two monomers, with each monomer arranged into three structural domains. Domain I, encompassing 180 residues from the N-terminal[1] is formed of a distorted barrel with a jelly roll motif and two β-hairpin motifs[5]. Domain II, encompassing residues 181-273[1], has an immunoglobulin-like β-sandwich structure. Domain III, encompassing residues 274-603, has a C-terminal $(\alpha/\beta)_8$-barrel fold containing the active site and the tetramer interface[1,5] (figures 2.1-2.3). The enzyme is only functional as a tetramer.

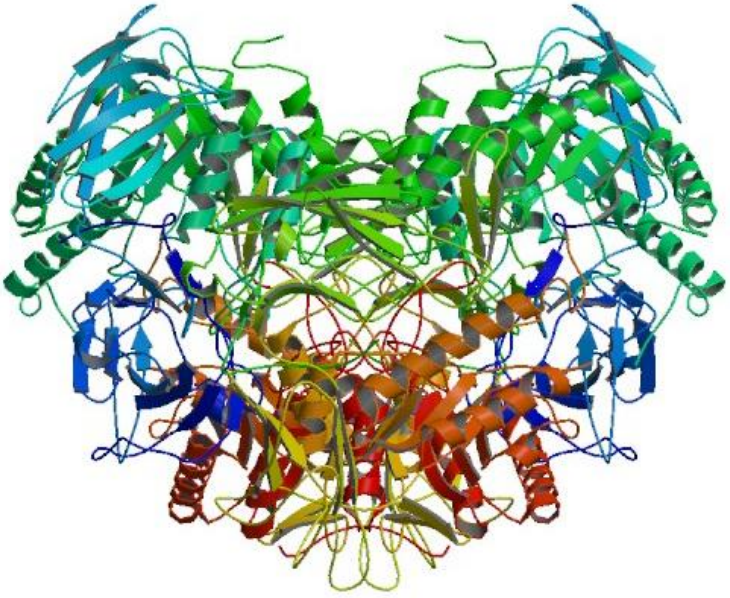


Figure 2.1: The tetrameric structure of E. coli beta-glucuronidase. Image accessed from the protein data bank[1]. The two asymmetric units are shown mirroring each other vertically on the left and right of the image.

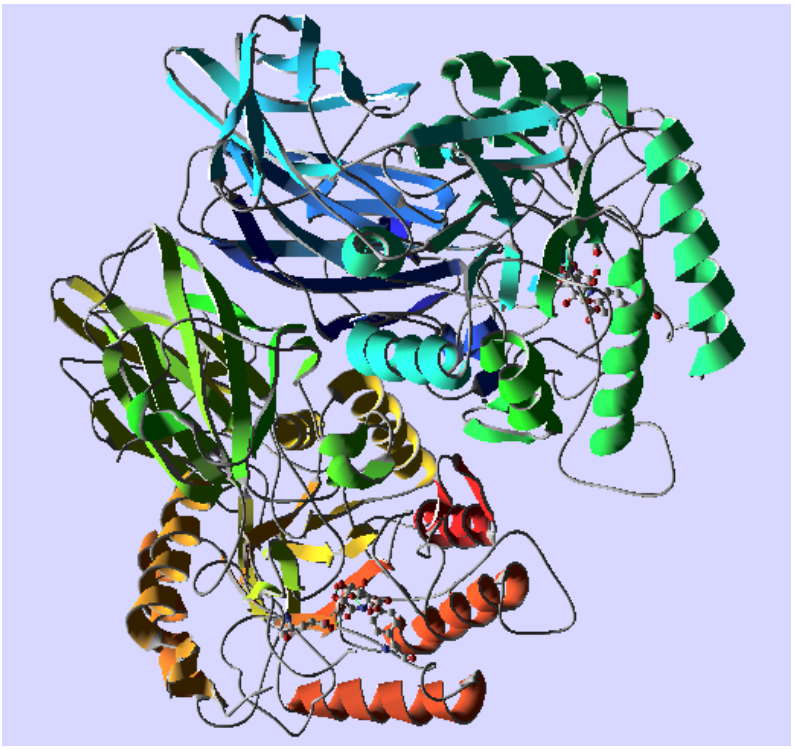


Figure 2.2: The asymmetric dimer unit of β-glucuronidase from *E*. Coli shown with an inhibitor molecule bound in each monomer[1]. The two monomers are diagonally opposite each other with one monomer shown ranging from red to green and the other monomer coloured from green to blue. The picture was acquired from the protein databank, image 3k4d, and edited using the Swiss PDBViewer software Deep View.

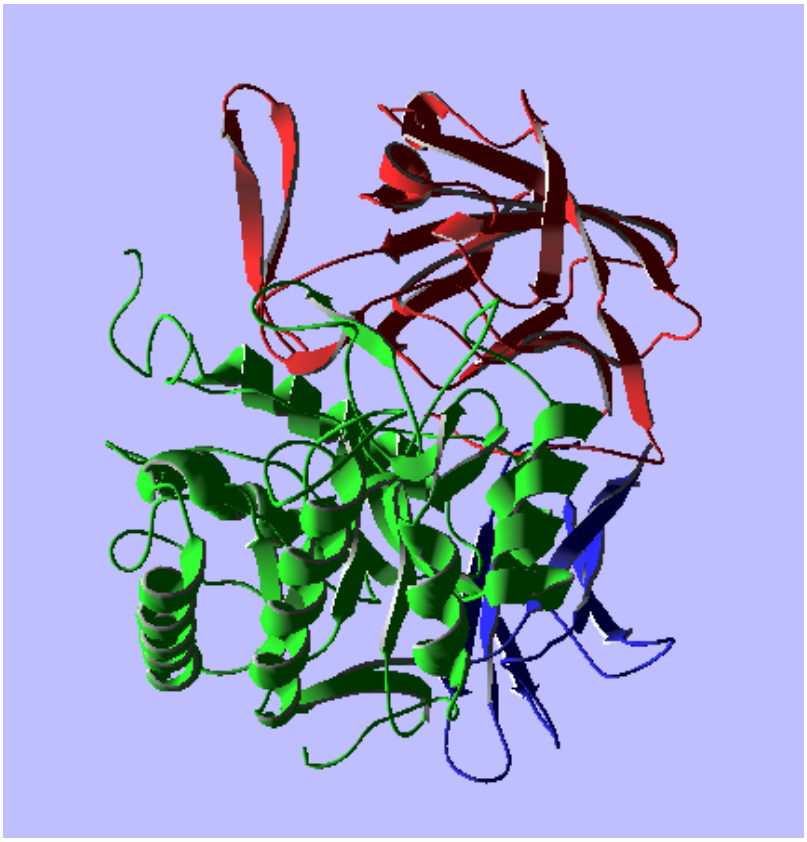


Figure 2.3: single monomer of β-glucuronidase showing the three domains. Domain I is shown in red, domain II in blue and domain III in green. Crystal structure[1] obtained from the protein data bank and edited using Swiss PDBViewer software, Deep View.

There are seven conserved lysine residues across several GUS species (K13, K157, K286, K400, K567, K568 and K578); six of them are interfacial whilst the seventh, K568, is found in the active site[6]. The lysines are thought to be important for the quaternary structure. Modifications to interfacial lysines disrupt the quaternary structure and enzyme function[6]. Mutations to the active site lysine, residue 568, have already been shown in experiments by Matsumura[7] to be deleterious to enzyme function in isolation.

β-glucuronidase catalyses the hydrolysis of the 1,4-glycosidic bond in β-glucuronide carbohydrates, to release a molecule of glucuronic acid. It is also weakly active towards β-galactopyranosides, which differ only in the position and substituents of the C4 and C5. It is also active to β-glucopyranosides, which differ only in the C5 substituent[7] (figure 2.4). β-galactopyranoside is usually cleaved by β-galactosidase, which displays no activity towards β-glucuronides[7]. Though linked by catalytic mechanism and α/β-barrel active site structure, the putative catalytic domains of β-glucuronidase and β-galactosidase are only 25% identical in amino acid sequence and the nucleotide sequence cannot be aligned by GAP algorithm.

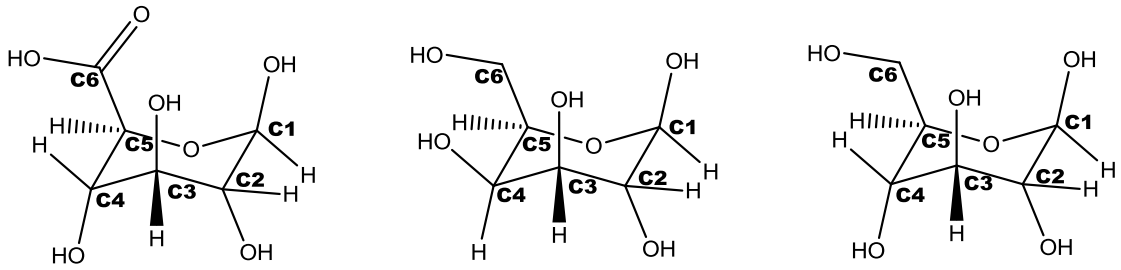


Figure 2.4: chemical structures of β-glucuronic acid and β-galactopyranose, along with the similar β-glucopyranoside. The carbons in the sugars have been annotated to demonstrate the naming and numbering system for carbohydrates.

Both β-glucuronidase and β-galactosidase are thought to have diverged from an ancient common ancestor. This is based on a theory called the 'Patchwork Hypothesis' in which it is thought that the wide variety of specialised enzymes found in the modern environment diverge from a much smaller number of catalytically inefficient broad-specificity enzymes[7].

The putative catalytic residues in β-glucuronidase are Glu413 and Glu504[7]. The proposal bases one glutamate as being the nucleophile and the other glutamate as the acid/base residue in accordance with the proposal for human β-glucuronidase[8] Current proposals for this mechanism suggest that retention of configuration is achieved using a two step mechanism composed of two SN2 reactions. In the first mechanistic step, Glu504 acts as the nucleophile attacking the anomeric centre of the non-reducing terminal sugar monomer to form a covalent bond with the enzyme, and release the rest of the polysaccharide; whilst Glu413 acts as the acid/base residue. This results in inversion of configuration at the anomeric centre (figure 2.5).

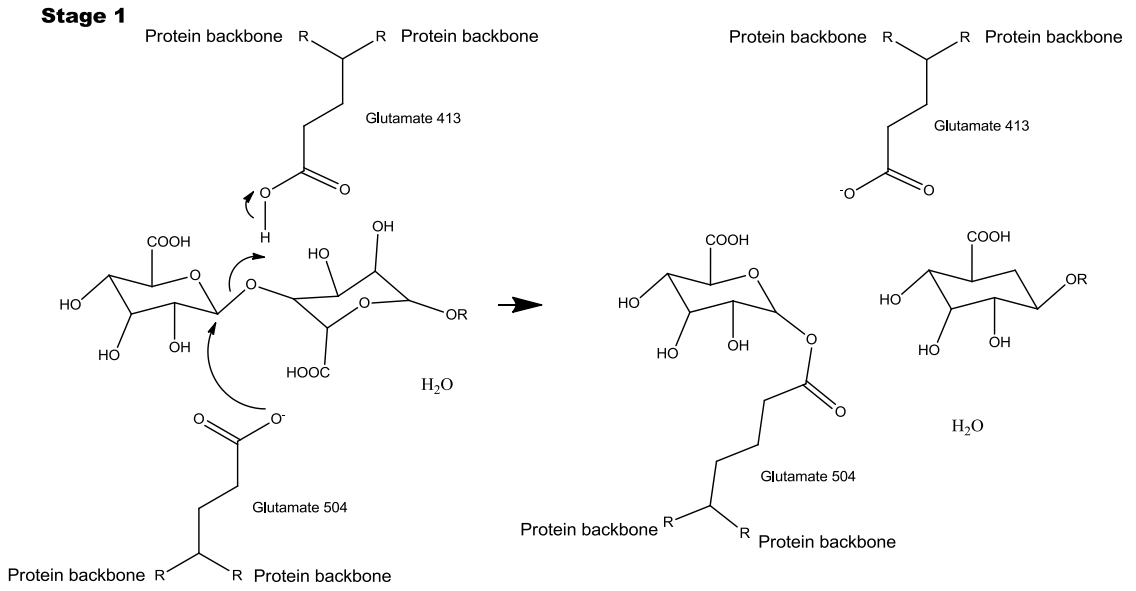


Figure 2.5: the first stage in the mechanism of hydrolysis of the 1, 4-glycosidic bond is an SN_2 reaction resulting in inversion of configuration and a covalent bond being formed between the enzyme and the substrate.

In the second mechanistic step, Glu413 acts as a base to aid water in hydrolysing the glycosyl-enzyme intermediate to release the molecule of glucuronic acid with the original anomeric configuration (figure 2.6).

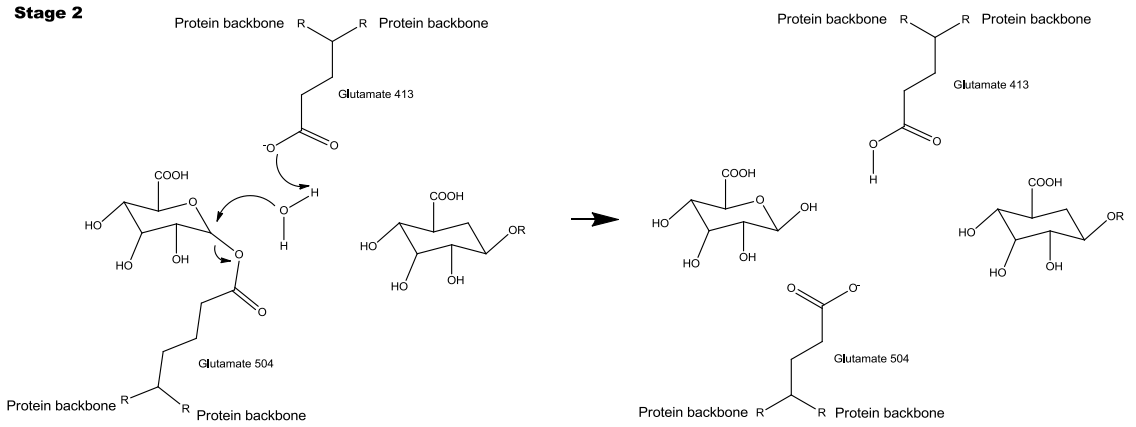


Figure 2.6: the second stage in the mechanism of hydrolysis of the 1,4-glycosidic bond. A second SN_2 reaction converts the product back to its original configuration, and releases the molecule of glucuronic acid from its bond with the enzyme.

Tyr468 is also important for catalysis but is not a catalytic residue[3,8]. The role of the tyrosine is not clear, but it is conserved across several species and known to be important. The proposed

mechanism omits an active role for Tyr468 as it is not considered to be a catalytic residue. Tyr468 is likely to be involved in hydrogen bonding with the substrate and may be involved in stabilising the active site or the enzyme-substrate complex.

A crystal structure (Protein databank file 3k4d[1]) of the active site highlights the position of the glutamate and tyrosine residues in conjunction with the lactam inhibitor molecule glucaro-δ-lactam (figure 2.7). Glucaro-δ-lactam is similar in shape to the natural substrate and is likely to bind in a similar manner. This model shows that the tyrosine lies very close to both catalytic residues and the substrate (figure 2.8). From this it could be inferred that the tyrosine could be forming hydrogen bonds to the substrate holding it in the correct orientation for hydrolysis or forming a hydrogen bond with the deprotonated glutamate.

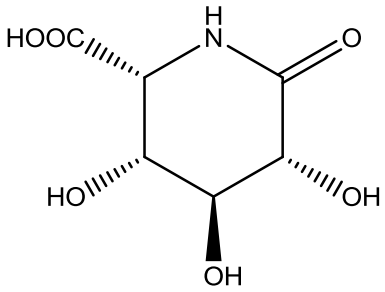


Figure 2.7: the structure of the β-glucuronidase inhibitor glucaro-δ-lactam which has a similar structure to that of the native substrate.

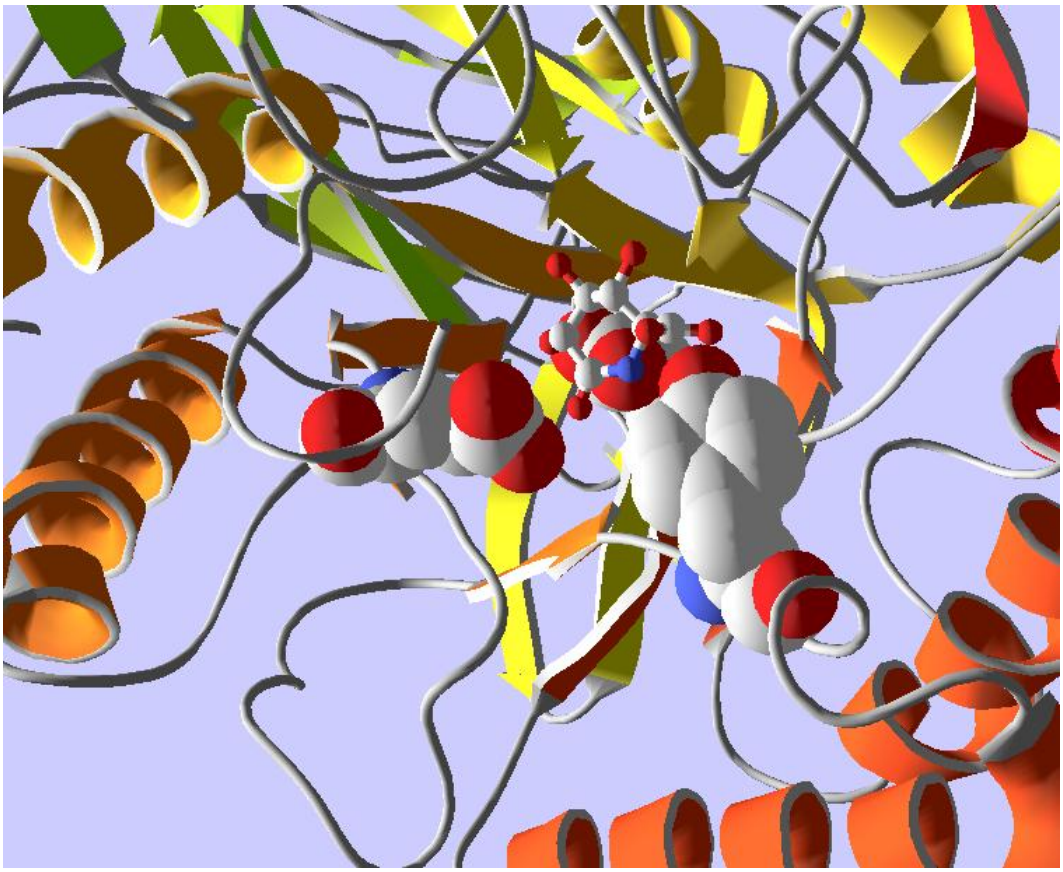


Figure 2.8: active site of β-glucuronidase demonstrating the catalytic glutamate residues and the position of Tyr468 relative to an inhibitor molecule with a very similar structure to that of the native substrate. Picture generated in the Swiss PDBViewer software Deepview from PDB file 3k4d.

Molecular modelling suggests that the specificity of β-glucuronidase for its substrate is due to a strong network of hydrogen bonds between the C4 OH and the C6 carboxylate oxygens with Tyr549, Arg562, Asn566, Lys568 all from loop 8; Tyr468 from loop 6 and Asp163 in domain I. Glu504 is also involved in bonding as well as being the catalytic residue[2]. The location of these residues in relation to the inhibitor molecule glucaro-δ-lactam can be modelled onto a crystal structure (figure 2.9).

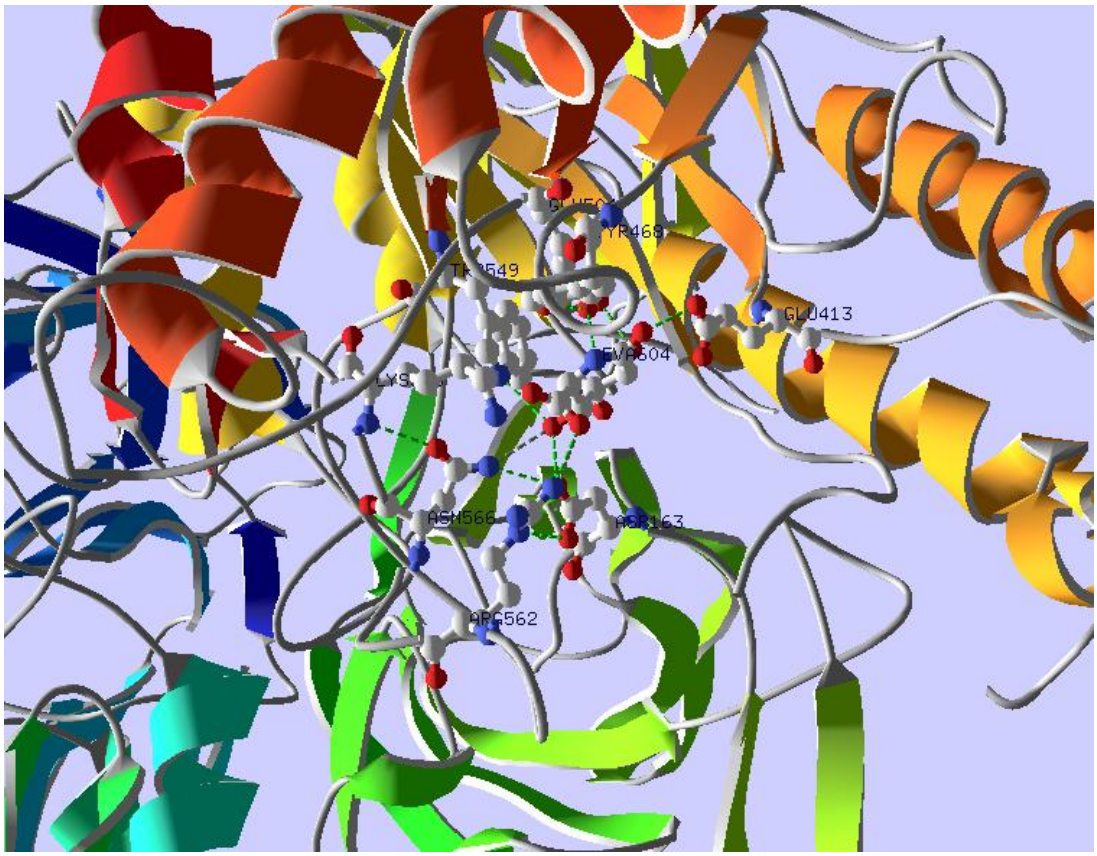


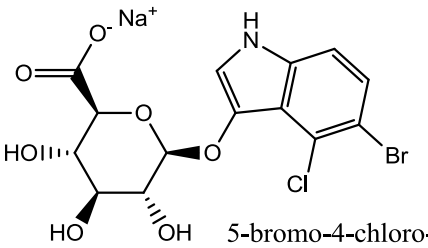
Figure2.9: active site of β-glucuronidase highlighting the location of catalytic residues and other residues proposed to bind with the substrate. In this picture the enzyme is modelled binding to the inhibitor glutaro-δ-lactam.

β-Glucuronidase is a useful enzyme system to use in directed evolution experiments. The enzyme has been studied in detail and there is a wealth of literature for reference. It is a convenient system which is demonstrated by β-glucuronidase and β-galactosidase being one of the most popular marker systems used in transgenic plant research[6, 9]. Both have been found to be highly stable and non-toxic to plants and the chromogenic substrates X-glu and X-gal are convenient histochemical markers. As β-galactosidase is found as a background activity in some plants, new β-galactosidase variants with β-glucuronidase activity have been developed and the retention of β-galactosidase activity has been considered to be an advantage[9].

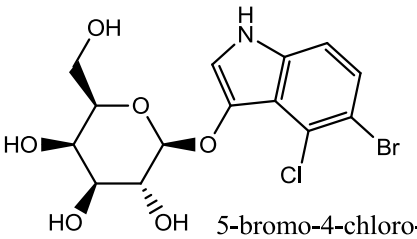
Study of glycoside hydrolases such as β-glucuronidase and β-galactosidase as a means of exploring directed evolution is particularly common as there are many readily available chromogenic substrates, which can be used in simple qualitative and quantitative assays. Examples include the native substrate mimics 2-nitrophenol-β-D-glucuronic acid (ONP-glu) and 4-nitrophenol-β-D-glucuronic acid (PNP-glu), which can be used in quantitative assays. Alternatively the native substrate mimic 5-chloro-4-bromo-3-indolyl-β-D-glucuronic acid (X-glu) can be used for a qualitative blue/white screening assay. These types of substrates are also available for many related pyranoside sugars including galactopyranoside, glucopyranoside and

xylanopyranoside allowing study of the activity to non-native substrates (figure 2.10a and 2.10b).

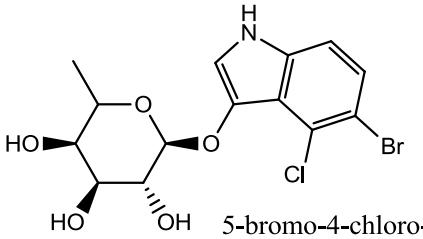
Both types of substrate rely on the ability of the enzyme to hydrolyse the O-R bond on C1 to release a coloured compound. When the bond is hydrolysed the compound is split into the sugar and chromogenic moieties. In the case of the 5-bromo-4-chloro-3-indole compound, this produces a blue precipitate which colours the bacterial colony allowing easy selection from the plate. In the case of the 2-nitrophenol-pyranoside substrate and 4-nitrophenol-pyranoside substrate, the resulting 2-nitrophenol/4-nitrophenol is a yellow coloured compound that has a UV/visible light absorbance that can be measured quantitatively by spectrophotometer and converted to a concentration, allowing convenient measurements to be made of enzyme kinetics.



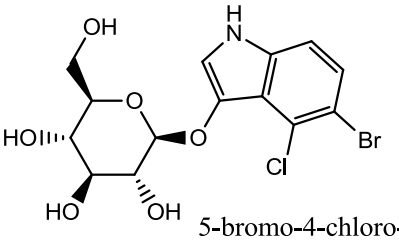
5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid (X-glu)



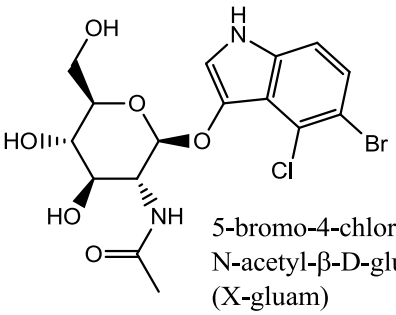
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal)



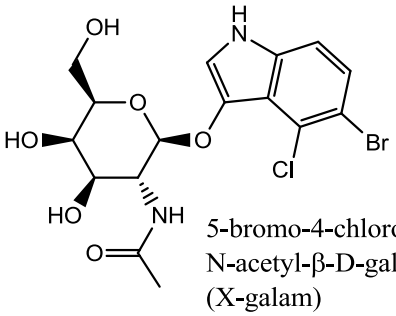
5-bromo-4-chloro-3-indolyl-β-D-fucopyranoside (X-fuco)



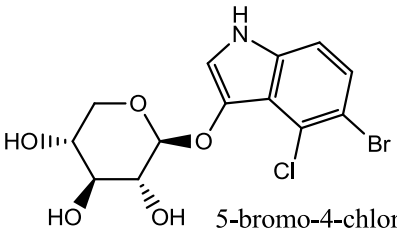
5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (X-gluco)



5-bromo-4-chloro-3-indolyl-N-acetyl-β-D-glucopyranoside (X-gluam)



5-bromo-4-chloro-3-indolyl-N-acetyl-β-D-galactopyranoside (X-galam)



5-bromo-4-chloro-3-indolyl-β-D-xylopyranoside (X-xyl)

Figure 2.10a: a selection of the histochemical 5-bromo-4-chloro-3-indolyl-pyranoside substrates available for studying glycoside hydrolases.

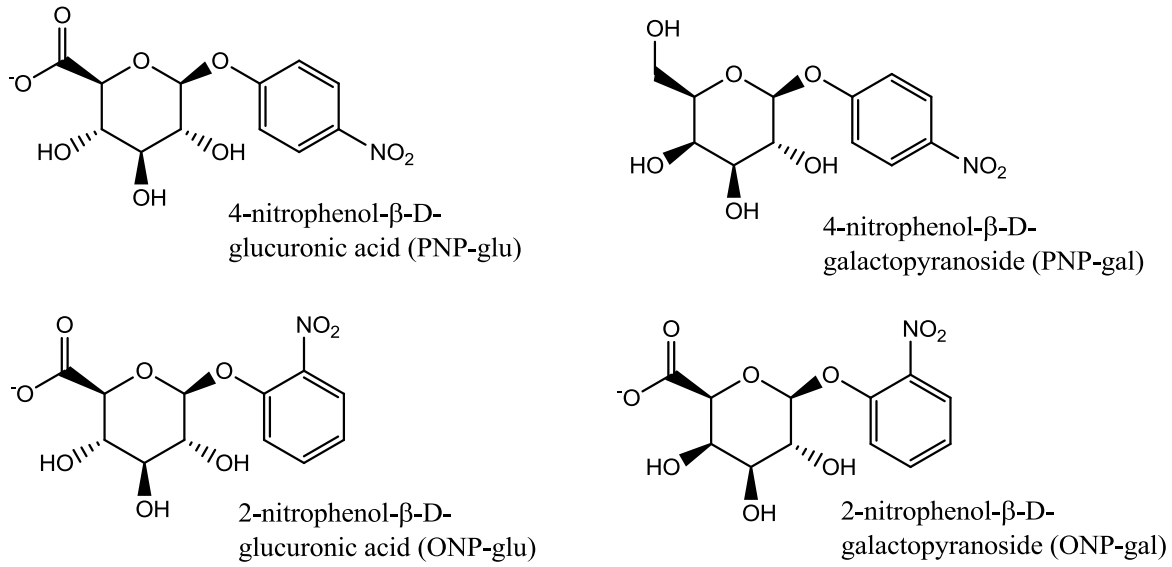


Figure 2.10b: a selection of the chromogenic 2-nitrophenyl and 4-nitrophenyl pyranoside derivatives available for studying the glycoside hydrolases.

2.1.2 Directed evolution experiments on β-glucuronidase

There have been several important classic directed evolution experiments on β-glucuronidase to direct the substrate activity towards β-galactosidase and β-xylosidase. Additionally there have also been directed evolution studies to improve the thermostability and surface chemistry of the enzyme. These experiments will be used to provide key controls and comparisons for the directed evolution experiments encompassing neutral drift.

Matsumura and Ellington[7] directed the evolution of β-glucuronidase mutants with increased β-galactosidase activity in three rounds of random mutagenesis and DNA shuffling. Overall T509A/N566S/S557P/K568Q was the fittest mutant found in the experiment. Other mutations were noted during the experiments including four commonly occurring mutations D508G, T509A/S, S557P and N566S. During the course of the experiment both S557P and N556S became fixed whilst D508G became extinct. The mutation K568Q was also identified during later rounds of the experiment and also became fixed. When these mutations are mapped onto a crystal structure of β-glucuronidase, all of the mutations can be found around the active site and binding pocket (figure 2.11).

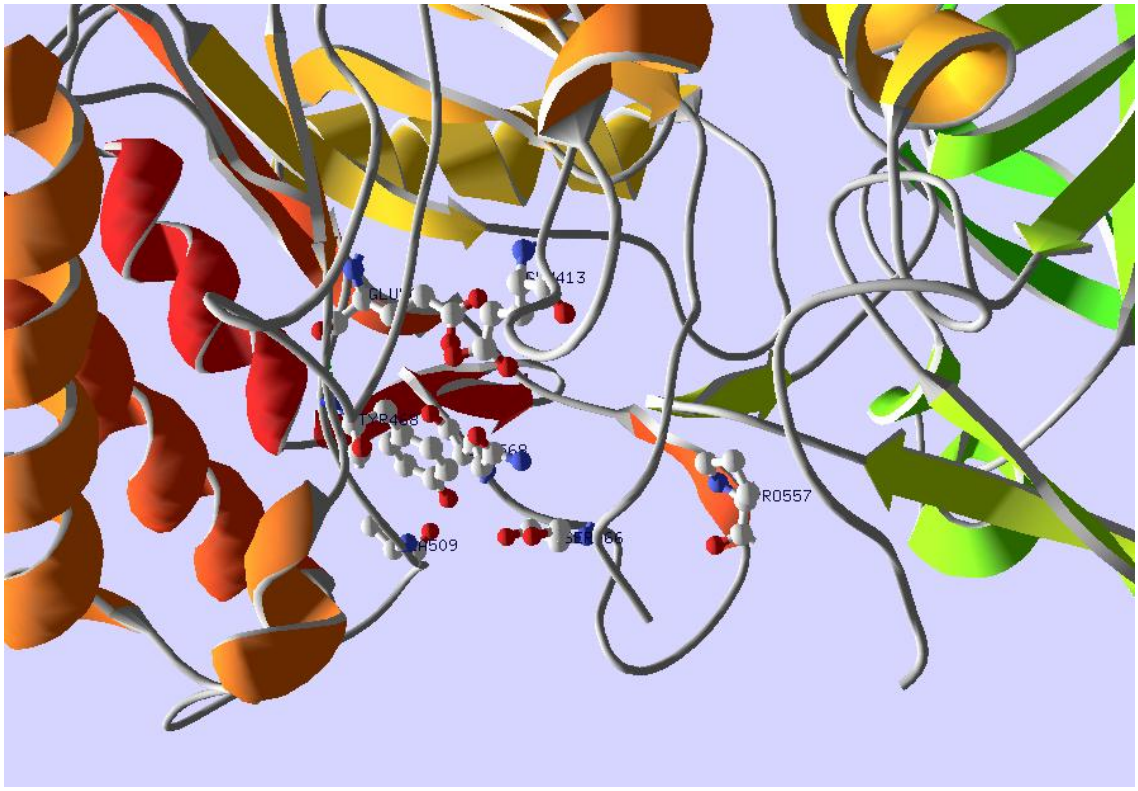


Figure 2.11: 3D rendered image of the crystal structure of β-glucuronidase binding pocket and active site. The catalytic and bonding residues E413, Y468 and E504 are shown along with the mutated residues T509A, S557P, N566S and K568Q.

When these mutations are considered the changes are quite significant. The mutation T509A changes from the nucleophilic larger sized threonine to the smaller non-nucleophilic alanine. This will create more space in the binding pocket and affect the ability to form hydrogen bonds through the loss of an OH group. Both these changes may improve the binding of the new substrate. N566S changes the amide asparagine into the nucleophilic serine. The serine is smaller than the asparagine and the ability to form hydrogen bonds is again changed as $CONH_2$ is changed to OH.

S557P changes serine to proline. Proline is unique amongst the amino acids for being cyclic. As a result, this amino acid is able to change the direction of the backbone. Residue 557 formed the terminal residue of a short β-sheet. The introduction of the cyclic structure of proline forces a directional change to the adjoining loop from the end of the sheet. This change affects the shape of that part of the binding pocket and changes the hydrogen bonding ability. This change is most likely to affect the binding of the new substrate.

The change at residue K678Q loses the positive charge from the lysine and changes an amine to an amide. The lysine at position 568 was already known to be involved in stabilising the quaternary structure of β-glucuronidase so a change to this residue has wide reaching effects

throughout the protein. Mutations of K568 alone cause the enzyme to become non-functional. Glutamine has a shorter, non-charged, side chain length than lysine so is smaller overall. This alters interactions with the neighbouring monomer. This may introduce flexibility into the protein to allow the binding of an alternative substrate, but comes at a cost of loss of overall stability to the protein. As this mutation is only found in the context of other mutations, there is an epistatic effect that offsets the destabilising effect of this mutation.

In a related experiment Rowe and Matsumura[3] directed the evolution of β-glucuronidase mutants with increased β-galactosidase activity in 10 rounds of random mutagenesis, in an "asexual" recombination experiment. The experiment also compared a PNP-gal microplate assay with the X-gal screening assay and found that a wider number of mutations were discovered using the microplate assay. However, many of the mutations drove each other to extinction without the fixing of beneficial mutations by DNA shuffling ("sexual" recombination). The same mutations T509A, S557P, N566S and K568Q were again seen in multiple mutants. In addition the double mutation F365S/W529L was frequently seen. The other mutations noted (S22N,G81S, K257E,Q598R, stop604W, E377K, H162L, S231T, F288L,T384N, V405A, N445I, K567R, A581V, I12V, I560V, and S475C) along with these key sites were plotted onto the crystal structure of β-glucuronidase (figure 2.12).

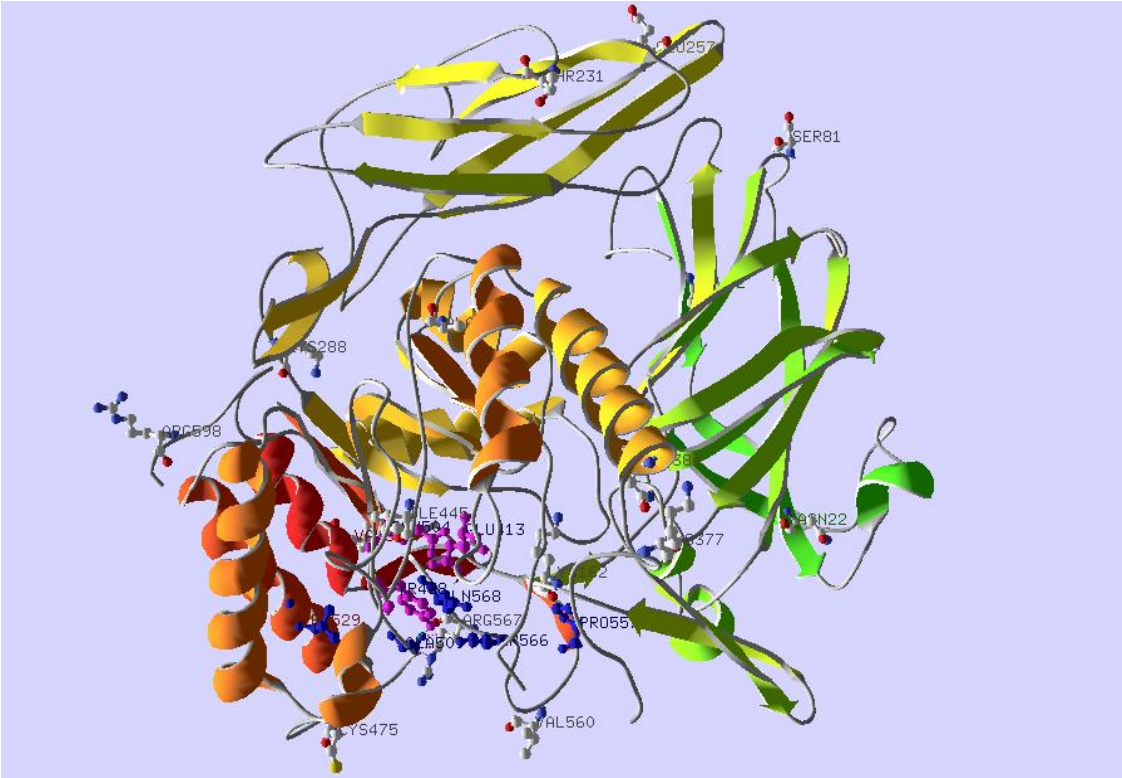


Figure 2.12: crystal structure of β-glucuronidase with mutation sites found by Rowe *et al* highlighted. The key mutations T509A, W529L, S557P, N566S and K568Q have been coloured blue. The catalytic and binding residues E413, Y468 and E504 have been coloured pink. The new key mutation F365S is not shown on this diagram as it appears in a portion of the protein

that was missing from this crystal structure. Stop604W is also not shown as residues beyond 600 are missing from the crystal structure.

As with the previous experiment, most of the mutations are centred round the active site and binding pocket. A few mutations are found distant from this area, but do not appear to interact with adjoining monomers. The differences caused by T509A, S557P, N566S and K568Q have already been discussed. The changes resulting from W529L and F365S can be rationalised. Both F365S and W529L are thought, by Rowe *et al*, to interact with the substrate. W529L changes the planar tryptophan to the smaller hydrophobic leucine. The loss of the NH of tryptophan affects the ability of that residue to form hydrogen bonds, but instead may allow increased flexibility to the structure allowing the new substrate to fit and bind more strongly. F365S loses the hydrophobic phenylalanine, replacing it with a serine which is a much smaller sized residue and containing an OH group able to form hydrogen bonds. This may bind to the new β-galactopyranose substrate, which has some OH groups in the opposite orientation to that of the β-glucuronide. F365 was noted as important for binding of an anti-cancer drug by Wallace et al[1]. The study did not seek to mutate residues but to develop compounds that inhibit the enzyme. The compound is significantly different in structure to the native substrate. Therefore it seems that this residue is important for binding a wide range of substrates. This residue is quite exposed so may guide the substrate into the active site and is also close to the monomer interfaces.

The authors also noted that the F365S/W529L double mutation remained white in an X-gal screen but was active in the microplate assay which utilised PNP-gal. This was attributed to the evolution of two distinct lineages displaying different phenotypes, of which one phenotype demonstrated a preference for PNP-gal over X-gal screening.

Geddie and Matsumura[2] directed the evolution of a β-glucuronidase with increased β-xylosidase activity using saturation mutagenesis of residues S557, N566 and K568, identified as the important residues in the Matsumura[7] experiment detailed above. The technique led to rapid evolution of clones with increased β-xylosidase activity. These were selected, pooled and amplified by whole plasmid PCR. Two fit mutants were identified through the experiment as S557P/N566A/K568F and S557P/N566S/K568Q. It was also found that the fittest mutants were all products of recombination events during whole plasmid PCR. Further rounds of evolution using DNA shuffling and StEP led to only modest improvements. It is interesting that the mutants generated here are so similar to those generated for increase in β-galactosidase activity. β-Xylopyranosides are quite noticeably different in structure to the glucuronides and galactopyranosides as they are missing C6. The six-membered ring is identical among the three sugars but C5 is a CH2 group in the xylopyranosides. Much of the bonding and stabilising of the substrate in the enzyme comes from hydrogen bond interactions between the C6 COOH/OH so

expected changes to protein structure would arise to fill the space and allow new hydrogen bonds to be formed.

Flores and Ellington[5] increased the thermostability of β-glucuronidase in seven rounds of random mutagenesis and DNA shuffling. The improved variants were able to function at up to 80°C whereas the wild type enzyme was inactivated above 65°C. The mutations which appear to be important for thermostability include F51Y, A64V, D185N, Y517F, Y525F, N550S, G559S, K567R, Q585H and G601D. None of these positions correlate with the positions of mutations for altering substrate selectivity. This is attributed to their position within the enzyme, as many were on the surface and thought to stabilise quaternary structure, which is important to thermostability. Residues for changing substrate selectivity are buried deeper within the enzyme. However, N550S is located within the active site, whilst residues K567R and G559S are located on loops important for binding different substrates. K567R is located directly between residues 566 and 568, which are important sites of mutations for substrate selectivity. These residues were plotted onto the crystal structure of β-glucuronidase (figure 2.13) which showed that most of these mutations were on the surface and distant from the active site. Of the mutations, two exchanged tyrosine for phenylalanine and a third phenylalanine for tyrosine. Phenylalanine and tyrosine share some similarity in structure, both having a planar benzene ring at the heart of their side chain. They are thus very similar in size and shape. The tyrosine has the ability to form hydrogen bonds due to the presence of the OH group. However the ability to form hydrogen bonds may not be as important for these residues and their orientations. K567R changes lysine for arginine. This change retains the positive charge of the side chain that may be important for interactions with the other monomers as part of quaternary structure stabilisation and will be important for thermostability. Arginine is also a similarly sized residue and changes the NH_3^+ for a guanidinium ion.

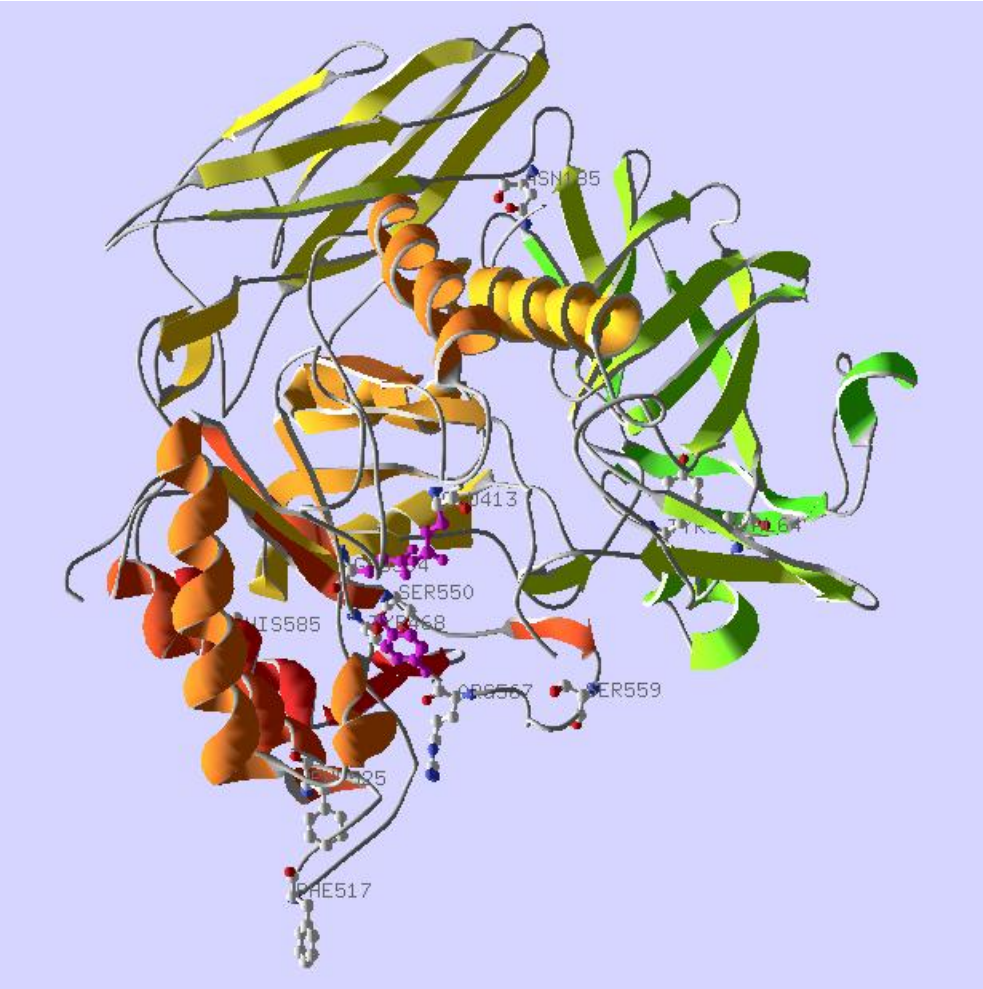


Figure 2.13: crystal structure highlighting mutation sites for residues important for thermostability of β-glucuronidase. The catalytic and binding residues have been coloured purple to highlight their position relative to other mutations. Picture generated using a crystal structure of β-glucuronidase obtained from the Swiss Protein Database and manipulated using the Swiss-PdbViewer.

Matsumura *et al*[6] directed the evolution of an aldehyde resistant β-glucuronidase in three rounds of mutagenesis and DNA shuffling. Resistance to aldehydes prevents the loss of activity of the enzyme during tissue fixing with aldehydes. This allows the protein to be used as a reporter in gene expression studies in animal cells. The improved variant had eight amino acid substitutions: N66D, D151N, A219V, I396T, T480A, Q498R, D508E and K567R. These were plotted onto the crystal structure of β-glucuronidase to determine their position relative to the active site and other mutation sites (figure 2.14). These mutations do not match with those found to be important for substrate selectivity, and only K567R was found to be important for thermostability. This difference in mutation sites is not surprising as this directed evolution selects for a very different property.

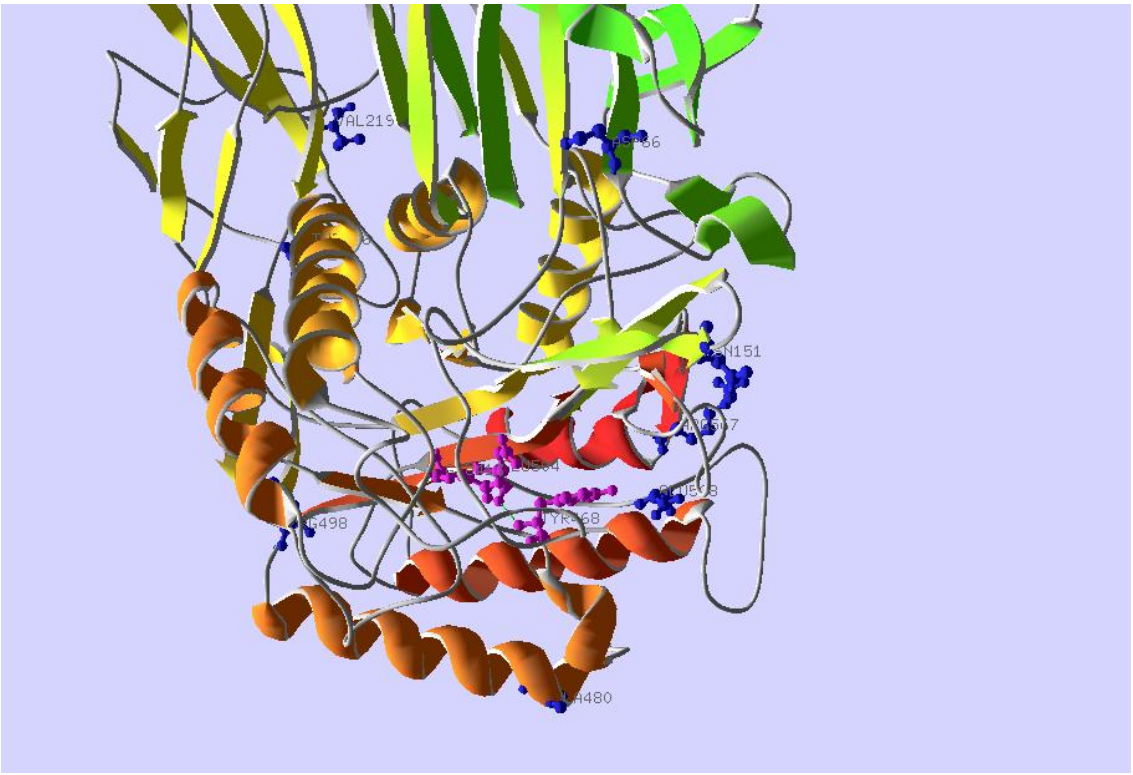


Figure 2.14: crystal structure of β-glucuronidase with the location of mutations important for aldehyde resistance mapped onto the structure. Catalytic and binding residues E413, Y468 and E504 are coloured purple. Mutated residues are coloured blue.

These experiments show that the key mutations for substrate selectivity in β-glucuronidase are F365S, T509A, S557P, N566S, K568Q and W529L. The mutation sites, except F365S and W529L are common to all substrate experiments, which may suggest that there is a local minimum in the fitness landscape. It is possible that conducting the directed evolution using neutral drift will introduce new, different mutations that are able to escape that minimum, allowing further parts of the fitness landscape to be explored that are not accessible by standard direct selection experiments.

2.1.3 Incorporating neutral drift into directed evolution experiments

Neutral drift is thought to play an important part in natural evolution and may prove a useful technique to use in directed evolution experiments. Neutral drift introduces a number of mutations that maintain native functionality of the protein in the existing environment, but which may prove to be beneficial if the selection pressure should change. The question is whether it is possible to use neutral drift to produce variants with new substrate activities, which have a higher activity than those found through direct selection. This thesis seeks to answer this question by using neutral drift in the directed evolution of a β-glucuronidase into a β-galactosidase.

The number of experiments which incorporate neutral drift is small in comparison to the vast numbers of direct selection experiments carried out. There are however, five key experiments which have used neutral drift in the directed evolution of antibiotic resistance in β-lactamases and also for new activities in PON enzymes, including organophosphate breakdown.

Bershtein *et al*[11] carried out 18 rounds of neutral drift (referred to as purifying selection in the paper) on a large library of TEM-1 β-lactamase and found that most of the mutations were predictable; returning the mutant to either an ancestral or consensus sequence. These sequences contained a number of global suppressors that allowed more destabilising mutations to be accepted. Additionally new activities were developed: TEM-1 β-lactamase degrades penicillins thus providing antibiotic resistance, whilst the library mutants were also able to degrade cefotaxime, a "third generation" cephalosporin antibiotic for which TEM-1 β-lactamase usually has negligible activity. From these results, the authors concluded that neutral drift can be used to produce highly diverse and evolvable libraries, which enrich ancestral and consensus mutations. These mutations act as suppressor mutations balancing the effect of potentially deleterious adaptive mutations thus allowing the evolution of new enzyme activities.

An earlier study by Bershtein *et al*[12] on TEM-1 β-lactamase carrying out ten rounds of neutral drift and purifying selection found that there is a threshold of robustness in the number of mutations, and the stability of the mutations that a protein can accommodate which is regulated by epistasis: the effect one mutation exerts on another. Recombination (by sexual or other mechanisms) is of crucial importance in maintaining the threshold. The importance of recombination correlates well with the results of the Rowe experiment in which beneficial mutations drove each other to extinction in the absence of recombination; even though the latter experiment did not use neutral drift.

Gupta and Tawfik[13] prepared two libraries of serum paraoxonase (PON1) mutants. The first library utilised a GFP reporter attached to PON1 which allowed sorting of active neutral mutants by FACS. This library contained $>10^6$ mutants. The second library was prepared on a much smaller scale and was selected according to enzymatic activity alone using a lactone derivative TBBL. The neutral libraries were then used in conventional directed evolution experiments for improving the activity with non-native organophosphate substrates. Again the authors concluded that neutral drift was a suitable method for producing libraries of variable sizes. Additionally they assert that libraries prepared by neutral drift can be used to identify mutations that alter the active site of an enzyme and improve the target function.

An earlier related experiment by Amitai *et al*[14] also produced a library of GFP labelled PON1 variants; with each variant classified as neutral, nearly neutral or possibly neutral with respect to the expression levels of protein and the activity with TBBL. This produces a neutral network from the library. A neutral network is a series of genotypes linked by point mutations that share

a common phenotype. The experiment was designed to sample variants to the perimeter of the neutral network; and investigate the evolutionary potential of different genotypes which share a common phenotype. Screening of the library against promiscuous substrates found increased activity to new substrates including synthetic compounds, and concluded that neutral mutations could facilitate adaptation under changing circumstances by expanding the activity range of an enzyme and reducing the number of mutation required to reach a new function.

Bloom et al[15, 16] prepared a library of neutrally evolved P450 enzyme variants. Mutants had to retain >75% activity in hydroxylating 12-pNCA to fit the neutral criteria. The library was then subjected to screening against a range of other substrates with decreasing similarity of structure to 12-pNCA, which all had a degree of hydroxylation activity by the parent enzyme, and the promiscuous activities were analysed. The results showed that the mutants tended to either show an overall increase or decrease in activity towards all substrates, and to have developed a preference to either fused ring or phenolic ether substrates. From these results the authors propose that neutral drift allows for the gathering of latent mutations which may become useful if the selection pressure changes. Another proposal is that neutral drift does not cause unpredictable shifts in activities, and also samples underlying “equilibrium” distributions of activities.

2.1.4 Aims of this experiment

This experiment aims to direct the evolution of a β-glucuronidase with increased β-galactosidase activity, which has increased activity compared to mutants found in other direct selection experiments. In order to do this an approach using neutral drift will be employed. The whole gene will be subjected to mutagenic PCR to sample variation throughout the protein. Mutants will initially be selected by retention of β-glucuronidase activity through a qualitative screen with the histochemical substrate X-glu. This will ensure that neutral mutations, not those deleterious to β-glucuronidase activity are carried through. After four rounds of mutagenic PCR and selection for original activity, the mutants will be screened for the new β-galactosidase activity.

The mutants obtained will be sequenced and compared with those found through other direct selection experiments. The aim is to find new beneficial mutations not found by direct selection.

2.2 Results and discussion

The experiments in this thesis seek to repeat and expand on work that was carried out by Dr W. Smith. However, the results of those experiments are important to discuss in context with the newly generated results. The libraries of β-glucuronidase mutants created by W. Smith are used in the experiments screening against a range of alternative substrates which was work carried out towards this thesis. For the purposes of identification the libraries prepared by W. Smith are described as libraries WS1 and WS2, and the libraries prepared by J. Hale as libraries JH1 and JH2. A library is defined in this context as one directed evolution experiment encompassing three rounds of mutation and selection for wild-type (β-glucuronidase) activity followed by a fourth round of mutation and selection for alternative activity (β-galactosidase). They are distinct libraries as there are subtle differences in the method of preparation and numbers of mutants. Initially the intention was to use the same protocols for the preparation of libraries JH1 and JH2 as was used to prepare libraries WS1 and WS2, but changes to proprietary products required changes to the protocols as they were no longer effective.

2.2.1 Changes to experimental procedures between the preparation of libraries WS1 and WS2, and JH1 and JH2

Mutagenesis by error prone PCR is the method by which genetic diversity is introduced into the libraries for these experiments. As outlined in the introduction, error prone PCR relies on the infidelity of certain DNA polymerases, and unnatural conditions to introduce errors when copying the DNA. This produces a large number of copies of altered DNA sequences. For WS1 and WS2[17], random mutagenesis was carried out across the rounds using a mixture of Taq mutagenesis and the commercial Genemorph™ random mutagenesis kit. This helped balance the mutational biases of the two enzymes. Taq DNA polymerase has a mutational bias towards A and T, whilst Mutazyme, the enzyme found in the Genemorph™ kit has a bias towards C and G. The Genemorph™ kit has been updated to GenemorphII™, which has lost the mutational bias by combining the properties of Taq and Mutazyme into one new proprietary enzyme, MutazymeII. This eliminates the need for using two methods of mutagenesis therefore GenemorphII alone is used for the preparation of JH1 and JH2.

Restriction digests are a method of cutting DNA which can be utilised in cloning to allow deletion and insertion of lengths of DNA into a gene or plasmid. Restriction endonucleases are a class of enzymes that recognise certain DNA sequences and cut the double stranded DNA inside that sequence to leave either a blunt ended or sticky ended product. A sticky ended product has the DNA cut asymmetrically inside the sequence to leave an overhang on one DNA strand: for example the cutting sequence for *Eco*RI is 5'-GAATTC-3' which cuts to leave 5'-G_AATTC-3'. A blunt end has the cut straight across the two strands: for example the enzyme *Zra*I which cuts the sequence 5'-GACGTC-3' to leave 5'-GAC GTC-3'. In this system the restriction endonucleases *Eco*RI (5'-GAATTC-3') and *Nco*I (5'-CCATGG-3') are used. Both enzymes cut

with sticky ends. This allows the DNA to be cut in two places to leave different overhangs, and is used to cut both the plasmid and gene fragment. In this system using two different cutting sequences prevents linearised plasmid DNA from reforming a circularised plasmid and ensures that only the correct gene fragments can be inserted. Restriction digests, when carried out according to the protocol devised by W. Smith, had produced poor results including only partially cut products, particularly when digesting the p042 plasmid. This was initially not noticed as the partially cut fragments appeared at almost identical points on an electrophoresis gel, and therefore were taken on to subsequent stages of the cloning procedure. After repeatedly testing restriction digests under a range of altered volumes of reactants without much effect, it was considered that the amount of glycerol in the restriction enzyme buffers was having a detrimental effect on the digestion causing non-specific cutting activity (also known as * activity). Diluting the enzyme 1:1 in water before addition resulted in cleaner digestions. The buffer that the digestions were carried out in was also changed. It had previously been carried out using Roche branded *Eco*RI in Roche *Eco*RI buffer by W. Smith, as double digestions of *Eco*RI and *Nco*I were advised, by that supplier, to be carried out in *Eco*RI buffer. However a change of supplier to New England Biolabs had occurred in the interim and it was found through experiment that in our laboratories the activity of *Nco*I was sub-optimal in the new *Eco*RI buffer. From test digestions it was found that the activity of *Eco*RI was not compromised by using NEB buffer 4, the standard buffer for *Nco*I reactions, supplied by New England Biolabs. Therefore the buffer was changed to NEB buffer 4.

Ligation is a method for joining DNA together. In cloning it allows a length of DNA such as a gene to be inserted into a plasmid. Ligations use a class of enzymes called DNA ligases which catalyse the formation of the phosphodiester bond between the 5' and 3' ends of the DNA being annealed. The ligations were tested out under a range of conditions changing volumes of linearised plasmid and gene product until the best yields of circularised plasmid were given in transformations. Additionally ligations were tried at different temperatures and lengths of time. From this it was found that doing ligations at room temperature for two hours gave the same number of transformants in the next step as ligating overnight at 4°C. Arguably, ligation at room temperature for two hours is a more efficient method by cutting out one overnight step.

Transformation is a technique for inserting DNA into a bacterium. The piece of DNA is usually in a supercoiled circular form called a plasmid which contains the gene of interest along with other genes and important DNA including the code that signals the bacterial DNA replication machinery to bind. The DNA is forced through the bacterial cell wall and cell membranes using either heat or electricity. Usually the strain of bacteria used will be lacking the gene of interest, and will be a strain optimised to accept foreign DNA and the conditions of heat shock or electricity without dying in the process.

The number of colonies being recovered after transformation into the commercial supercompetent *E. coli* strain XL1 Blue was extremely poor on a number of occasions in spite of optimisation of ligation conditions. After extensive testing with different cell strains and cell competency levels, the cause was eventually discovered to be the tubes used for transformation. Changing from a 1.5 mL microcentrifuge tube to a 15 mL centrifuge tube increased the total colony count by 10^2-10^3. This is being attributed to the change in surface area and thickness of the plastic tubes. This change was made during the preparation of library JH1, which was prepared using 1.5 mL tubes for transformations in rounds 1-3 and changed to 15 mL tubes for round 4. Transformation is carried out in two stages in our laboratory due to the poor results of electroporation and transformation of the plasmid directly into BW25141. BW25141 is a strain of E. coli that does not contain the β-galactosidase and β-glucuronidase genes. The screening cannot be carried out directly in XL1 Blue as this strain contains the β-galactosidase gene and would therefore produce false positive results.

With hindsight, the logical explanation for the overall small number of colonies from transformation is a fault of the protocol itself. By directly transforming the ligation mixture without first purifying the plasmid, the plasmid is found in a high salt based buffer solution along with any remaining linear sections of DNA. The salts interfere with electroporation by causing the electricity to arc across the solution, killing the cells in the process, which was why electroporation directly into BW25141 was not possible. Additionally transformation relies on the ability of the heat and cold to disrupt the cell wall and membrane to allow the passage of the plasmid into the bacterium. Salts in the solution will affect the cell wall and be transported through cell membranes. Transformation of the ligation mixture would also have allowed a small amount of linear DNA to pass into the bacteria. Transformation could have been highly optimised by first cleaning the ligation mixture to remove salts and linear DNA. There is evidence to suggest that this was a factor in transformations as the positive control plate prepared at each transformation, which used stock p042 plasmid, always contained vastly more colonies (>1000 colonies).

2.2.2 Results from libraries WS1 and WS2 (discussion of previous experiments[17])

Libraries WS1 and WS2 were prepared by carrying out mutagenic PCR on the β-glucuronidase gene contained in plasmid p042 via standard Taq mutagenesis or use of the commercially available Genemorph Random Mutagenesis Kit. The mutated gene was digested using EcoRI and NcoI and ligated into a digested pBAD vector containing the same digestion sites. The resulting plasmids were transformed into E. coli XL1 Blue and all surviving colonies harvested and combined. The plasmid was extracted and transformed, by electroporation, into the E. coli strain BW25141. This strain, unlike XL1 Blue, does not contain the β-galactosidase or β-glucuronidase genes. The BW25141 transformants were grown on a plate containing the chromogenic substrate X-glu. All colonies displaying a blue colour, indicating the presence of a

functional β-glucuronidase were selected and the plasmids harvested to be used in the next round of the cycle. This cycle was carried out a further three times. After the fourth cycle of mutagenesis and cloning, the BW25141 transformants were screened against the chromogenic substrate X-gal. Colonies exhibiting a blue colour indicate the presence of β-galactosidase activity. As the bacterial strain contains no inherent β-galactosidase activity, the blue coloration must be derived from β-glucuronidase mutants with increased β-galactosidase activity. Blue colonies were selected individually. From these, samples of DNA were prepared for sequencing and a larger culture of bacteria prepared from which the proteins were expressed and purified ready for further study. The overall process and key stages for the generation of the library is summarised in a flow diagram in figure 2.15.

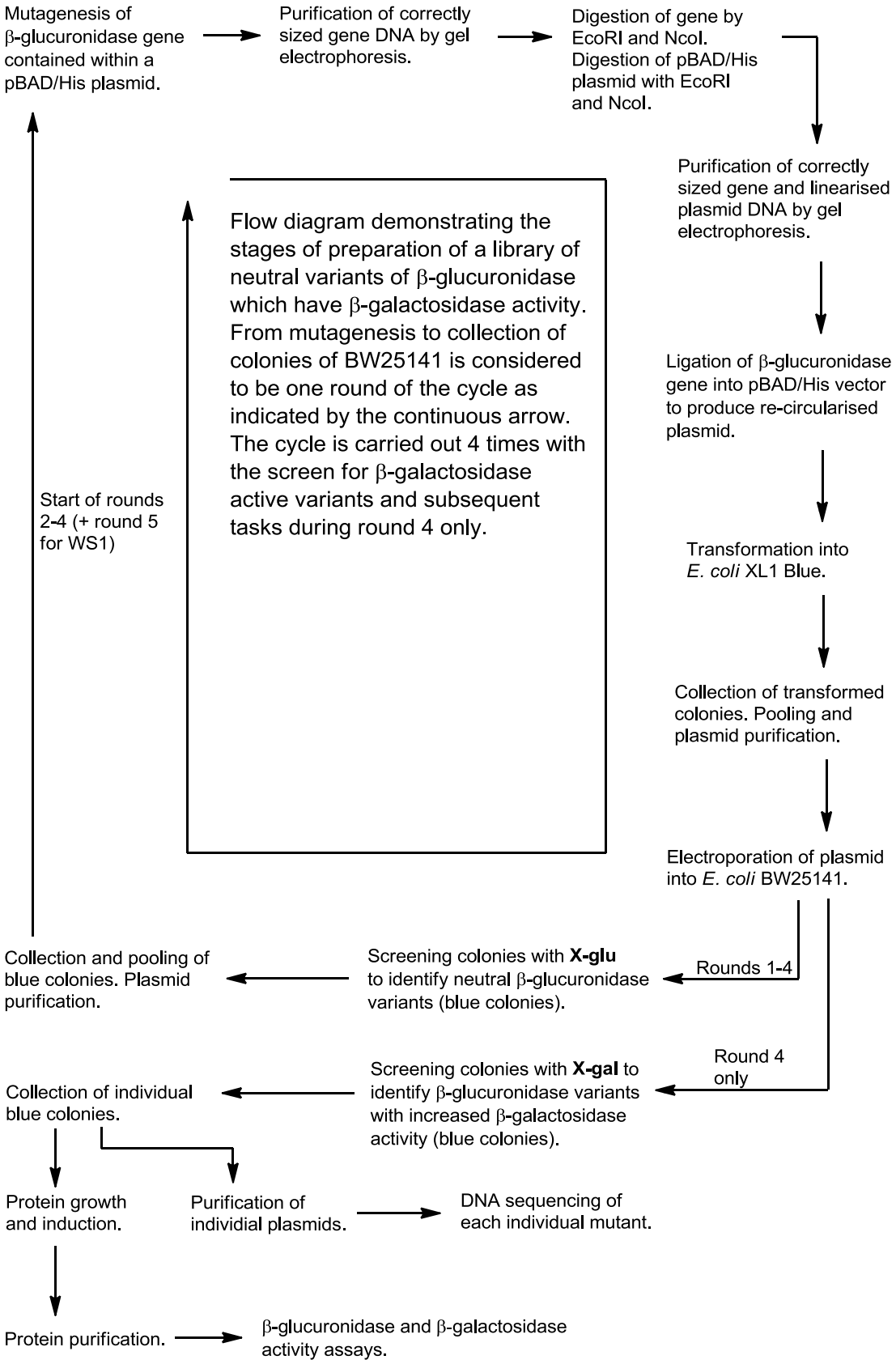


Figure 2.15: flow diagram representing the stages of neutral variant library generation used in this project.

In library WS1, β-galactosidase active mutants were found that contained mutations at positions D508, T509 and N566: all sites previously identified by Matsumura and Ellington[7]. These mutants were all identified from selecting colonies coloured blue by the product of the chromogenic substrate X-gal. Variants containing mutants other than at these key sites were also selected from the blue coloration of colonies. These colonies were selected qualitatively for being blue and there was reportedly little difference in the intensity of blue colour produced by colonies containing key mutations compared to those not containing key mutations. After a fifth round of random mutagenesis two new variants were identified: D56N/D508G/Y533C/S557A and G80V/A167S/W529L. These new variants add S557, the fourth important mutation site identified by Matsumura[7], and the W529L mutation identified by Rowe[3] to the results. Notably, the W529L mutation produced a blue colony from these experimental conditions, whereas Rowe[3] reported it to be a white colony as a result of their experimental conditions. Additionally D508G is seen. This mutation was frequently seen in the early rounds of Matsumura and Ellington's experiment, but was driven to extinction. As a result it was considered not to be an important mutation. The first appearance of this mutation at a later point in the experiment could suggest it does have some importance, but it is also an easy change of amino acid to make through point mutation. Aspartic acid to glycine is done by the point mutation of an A to a G, a transition mutation which occurs more frequently than a transversion mutation.

The K568Q mutation from Matsumura's experiment was not found. This is expected since K568Q is deleterious to β-glucuronidase activity in isolation and therefore would have been eliminated from a neutral variant screen. After shuffling, most of the mutations already considered to be unimportant were lost, K568Q appeared and S557A reverted to S557P. The kinetic parameters of the variants generated in this screen correlated with those calculated by Matsumura and Ellington[7].

The results found from screening libraries WS1 and WS2 were disappointing from the perspective that the use of neutral drift did not find new variants with increased β-galactosidase activity compared to those already discovered. However, as reported above, some interesting results were discovered. It is possible that by repeating the experiment again with a significantly larger library size further results may be discovered. If no new mutants are found it is to be expected that the results would again corroborate with those previously seen. The work for this thesis seeks to explore this by repeating the experiment using different library sizes. Library JH1 is a smaller library compared to WS1 and WS2, whilst library JH2 is a larger library.

2.2.3 The preparation and results of library JH1

The β-glucuronidase gene inside plasmid p042 was successfully mutated using the primers OSOT165 and OSOT166 and the GenemorphII™ Random Mutagenesis Kit. The product was digested using *EcoR*I and *Nco*I before being ligated into pBAD/His using T4 DNA Ligase. The resulting plasmid was transformed into supercompetent XL1 Blue with approximately 0.1%

transformation efficiency per plate compared to the positive control. The resultant colonies were harvested and the plasmid extracted using the Wizard*Plus*® SV DNA Minipreps Kit. The plasmid was transformed into BW25141 by electroporation and the transformants screened on X-glu coated LB ampicillin arabinose plates. Blue colonies, which indicate the presence of β-glucuronidase activity, were selected, pooled and the plasmid extracted using the Wizard*Plus*® SV DNA Minipreps Kit. The cycle of mutagenesis, digestion, ligation, transformation and selection was repeated a further three times. During the screening stage of the fourth round the BW25141 transformants were also screened for β-galactosidase activity using X-gal coated LB ampicillin arabinose plates. This is the same procedure as shown in figure 2.15. Table 2.1 gives the details of the exact plasmid names and numbers of XL1 Blue colonies counted at each round of the cycle.

Round Number	Plasmid Code	Number of XL1 Blue colonies	Number of BW25141 colonies	Activity being selected for (substrate)
1	pWSS042 (gusA)	200-250	679 (575 blue)	β-glucuronidase (X-glu)
2	pJRH4880/35 (mixed sequences)	478	654 (397 blue)	β-glucuronidase (X-glu)
3	pJRH4880/41 (mixed sequences)	652	4428 (3450 blue)	β-glucuronidase (X-glu)
4	pJRH4880/53 (mixed sequences)	609	193 (0 blue)	β-galactosidase (X-gal)

Table 2.1: table giving the code name for the plasmid used in mutagenesis reactions at the beginning of each round of the cycle. The table also shows the numbers of XL1 Blue colonies counted and selected after the first stage of transformation, the number of BW25141 colonies and the number of those which were blue selected colonies after the second round of transformation. The selection criterion is also listed.

From round one, 679 BW25141 colonies were screened against X-glu, and 575 blue colonies displaying native β-glucuronidase activity were selected. This is equivalent to 85% neutral mutants. During the production of libraries WS1 and WS2 the optimum number of neutral mutants had been found to be 30-50%. This indicated that the rate of mutagenesis was too low, and that less plasmid needed to be used for round two as the DNA was too concentrated. Overall the number of BW25141 colonies obtained was considered to be suitable for doing a screen of a small library.

From round two, 654 colonies were screened against X-glu and 397 blue colonies displaying native β-glucuronidase activity were selected. This is equivalent to 61% neutral mutants. The number of neutral mutants after round two suggested the rate of mutagenesis was more appropriate but could be further optimised.

From round three, 4428 colonies were screened against X-glu and 3450 blue colonies displaying native β-glucuronidase activity were selected. The results were equivalent to 87% neutral mutants. The reason for the sudden increase in the number of BW25141 colonies is unclear. The results suggest that electroporation was more efficient than what was normally achieved. It could be due to new electroporation cuvettes, plates or cells, but these details were not noted in the experimental. It was unlikely to be a contamination event as negative controls were free from contaminant colonies.

After screening the colonies in round four for β-glucuronidases displaying β-galactosidase activity, no β-galactosidase active mutants were found. 193 colonies were screened with X-gal and none were blue. The low number of colonies returned is surprising given the results of previous electroporations. However, reviewing the laboratory notebook this seems to be linked to a problem that was occurring with electroporations at that time, as a number of unsuccessful attempts at electroporation and screening had been carried out. Though it was not noted in the laboratory notebook if all the failed electroporations came from the same batch of cuvettes or cells, it is possible that the poor results were caused by a faulty batch of electroporator cuvettes or cells. The overall result of no β-galactosidase active mutants is not surprising as the number of colonies taken through at each XL1 Blue transformation was small, and the mutation rate was shown to be too low by the blue colony ratios of 85 and 87% during the screening and selection. As there is degeneracy in the genetic code for amino acids, single point mutations are likely to result in silent mutations in amino acids. In a small library with a low mutation rate, this leads to an inherent bias to wild type β-glucuronidases being selected and carried through to each subsequent round.

Gupta and Tawfik[12] make the claim that it should be possible to have a minimum library size of 200 mutants using neutral drift and still get results, as long as the mutation rate is sufficiently low enough. This library at 200-600 mutants (generated from the transformation into XL1 Blue) was larger than their minimum library size of 200, and had a low mutation rate as shown by the percentage of blue colonies (β-glucuronidases displaying native activity) during rounds 1-3 of selection; however in this instance no active mutants were found. These claims do not seem to take into account that low mutation rates will favour wild-type mutants being carried through the selection, particularly as the mutants are being selected for retaining original functionality.

2.2.4 The preparation and results of screening library JH2

Library JH2 was prepared according to the same protocol as library JH1. At the transformation step, the transformation efficiency per plate was approximately 5% compared to the control plate. The details of the plasmid name and numbers of XL1 Blue colonies from each round are given in table 2.2.

Round Number	Plasmid Name	Number of XL1 Blue colonies	Number of BW25141 colonies	Activity being selected for (substrate)
1	pWSS042 (gusA) (mixed sequences)	1533	3724 (2825 blue)	β-glucuronidase (X-glu)
2	pJRH5025/27 (mixed sequences)	5743	3488 (3064 blue)	β-glucuronidase (X-glu)
3	pJRH5025/34 (mixed sequences)	1788	5634 (3262 blue)	β-glucuronidase (X-glu)
4	5025/41 (mixed sequences)	1382	~6000 (~22 blue) 4637 (1902 blue)	β-galactosidase (X-gal) β-glucuronidase (X-glu)

Table 2.2: table giving the code name for the plasmid used in mutagenesis reactions at the beginning of each round of the cycle. The table also shows the numbers of XL1 Blue colonies counted and selected after the first stage of transformation, the number of BW25141 colonies and the number of those which were blue selected colonies after the second round of transformation. The selection criterion is also listed.

From round one, 3724 colonies were screened against X-glu and 2825 blue colonies displaying native β-glucuronidase activity were selected. This is equivalent to 75% neutral mutants. The proportion of neutral mutants was high, suggesting the mutation rate was low. The experiments to prepare libraries WS1 and WS2 had identified producing 30-50% blue colonies (neutral mutants) as being an adequate mutation rate. This suggested that for this round, the mutagenesis rate was too low. If the mutagenesis rate is too low then the library of variants will be biased towards wild type sequences. However, increasing the mutagenesis rate too far will produce a library of non-functional glucuronidases, so the balance must be carefully maintained. It was decided to maintain that mutagenesis rate for the next round.

From round two, 3488 colonies were screened against X-glu and 3064 blue colonies displaying native β-glucuronidase activity were selected. This is equivalent to 87% neutral mutants. This was an even higher proportion of neutral mutants than screened in the first round, which would have been biasing the library. It was decided that the mutation rate was too low and that the amount of the plasmid used in the mutagenesis step would be halved.

From round three, 5634 colonies were screened against X-glu and 3262 blue colonies displaying native β-glucuronidase activity were selected. This is equivalent to 58% neutral mutants. This mutation rate was deemed to be better and so an equivalent amount of plasmid was used in round four. 58% is slightly higher than ideal, but not to an extent as to make a vast difference.

From round four, 4637 colonies were screened against X-glu and 1902 blue colonies displaying native β-glucuronidase activity were selected. This was equivalent to 41% neutral mutants. The plasmid was extracted as before to leave plasmid pJRH5025/46. Transformants were also screened for new β-galactosidase activity against X-gal. Approximately 6000 colonies were screened, of which two colonies had a noticeably blue colour and approximately 20 colonies had a very weak blue colour. The two blue colonies and six of the very pale blue colonies were selected. From this cell stocks and plasmids were prepared and named 5025/46.b1; 5025/46.b2; 5025/46.wb1; 5025/46.wb2; 5025/46.wb3; 5025/46.wb4; 5025/46.wb5, and 5025/46.wb6 prefixed by either pJRH for the plasmid or BW25141 for the cell stock. The suffixes ‘b’ and ‘wb’ referred to the intensity of the blue colour, with ‘b’ representing blue and ‘wb’ representing weak blue.

The eight selected X-gal positive colonies were sequenced, and the mutants were found to be:

5025/46.b1	G245A/(N324N)/K370R/W529L
5025/46.b2	V67D/N166T/(N324N)/K370R/N412D/W529L
5025/46.wb1	A44E/P76L/G356V
5025/46.wb2	(P113P)/S127R/C133R/K155R/I187F/(H192H)/P239Q/Q243H/S258R
5025/46.wb3	F365S/L583Q (gene did not sequence in full)
5025/46.wb4	Wild type
5025/46.wb5	Wild type (Q598Q)
5025/46.wb6	(Q70Q/V85V) (gene did not sequence in full)

It was surprising that two of the selected mutants were sequenced as being wild type β-glucuronidases, with a third possibly being wild type. The wild type β-glucuronidase had previously shown as a white colony in β-galactosidase screens. It is possible that a wild type β-glucuronidase can exhibit a blue colouring as β-glucuronidase does have a small amount of β-galactosidase activity, and other experiments[3, 7] have reported that wild-type β-glucuronidases produce a pale blue colouring in X-gal screens. As it was approximately only 20/6000 colonies

which showed a pale blue coloration, and three were found to contain non-synonymous mutations, this is equivalent to a rate of ≤ 0.3% false positive results. It would have been better to have selected all of the very pale blue colonies to see if any of these returned wild type sequences, and also to have kept a record of all the mutations obtained. However, the focus was on colonies that were likely to have greatly increased β-galactosidase activity, which was considered to be colonies displaying a stronger, darker blue colour.

BW25141 containing mutant G245A/K370R/W529L (mutant 5025/46.b1) was grown in LB medium and the protein expression induced with the addition of arabinose. The protein was extracted and purified from the cells via FPLC and a nickel column into a high imidazole buffer. The pure fractions were exchanged into imidazole free buffer.

To measure the kinetic parameters of the selected mutant, an assay was used which measured the new β-galactosidase activity using a 96 well plate in an automated spectrophotometer. Increasing concentrations of protein from 100 nM to 10 μM were used to hydrolyse the substrate 4-nitrophenol-β-D-galactopyranoside (PNP-gal) of concentrations from 100 μM to 15 mM. The assay was carried out twice. The first assay produced poor results. From these results, the assay was adapted to use only the highest protein concentrations of 5 and 10 μM which had provided sensible looking results. The absorbance of the resulting product, 4-nitrophenol, was measured at 405 nm over a 5000 second time course. Each absorbance measurement was converted into a concentration using the Beer-Lambert law, $A=\varepsilon cl$, where A is the absorbance, ε is the extinction coefficient given in literature of 17400 mol $dm^{-3}.cm^{-1}$, and the path length, l, calculated from the diameter of a well and the volume of reaction, using the formula for the volume of a cylinder (volume=$\pi.radius^2$.height), where height is equivalent to path length. These concentrations were used to plot graphs of time versus 4-nitrophenol concentration (figures 2.16 and 2.17).

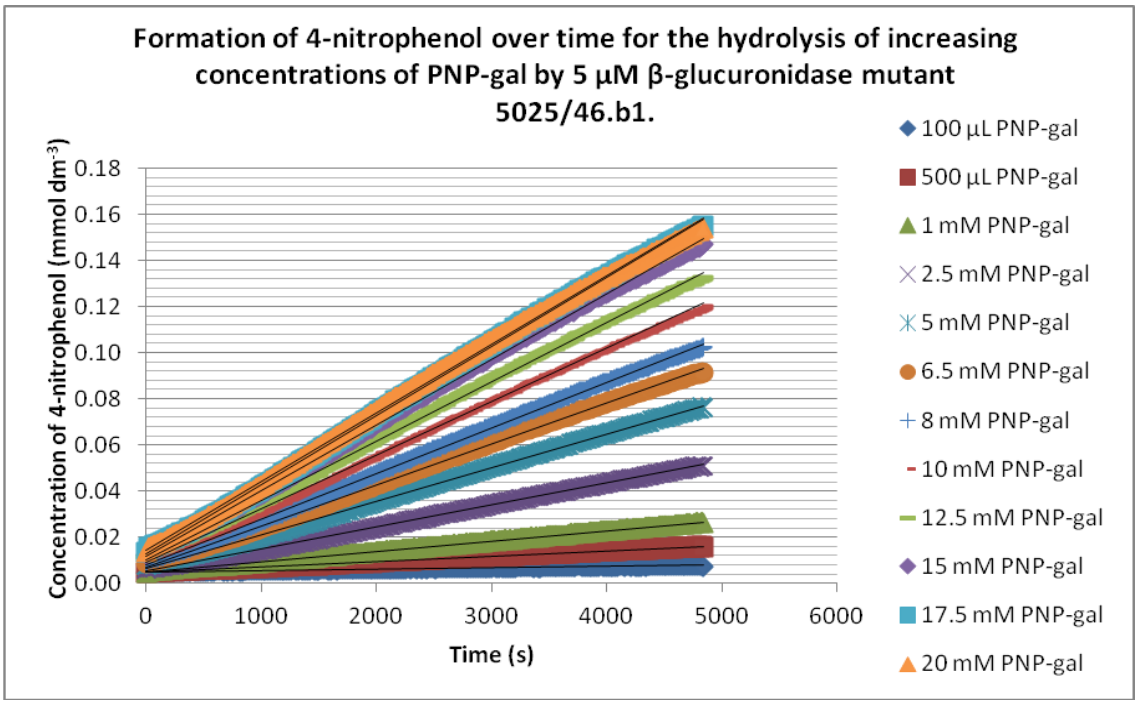


Figure 2.16: formation of 4-nitrophenol over time for the hydrolysis of increasing concentrations of PNP-gal by 5µM β-glucuronidase mutant 5025/46.b1. Absorbance was measured at 405 nm using a spectrophotometer and converted to concentration using Beer-Lambert Law.

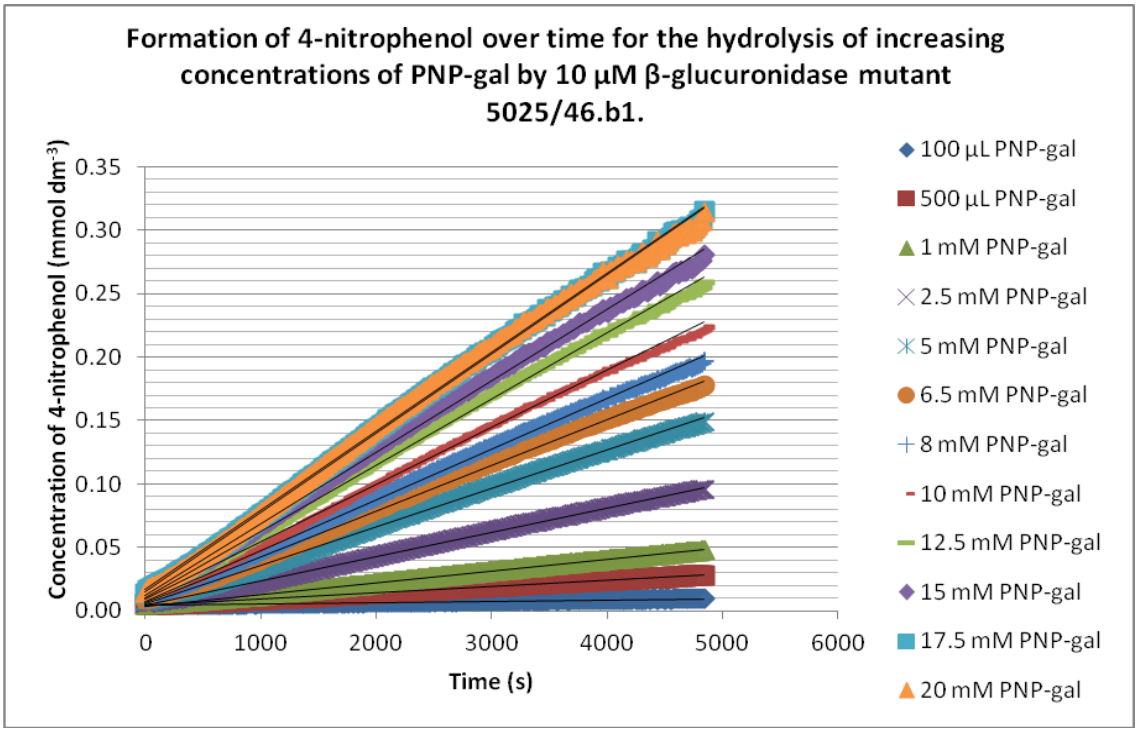


Figure 2.17: formation of 4-nitrophenol over time for the hydrolysis of increasing concentrations of PNP-gal by 10 µM β-glucuronidase mutant 5025/46.b1. Absorbance was measured at 405 nm using a spectrophotometer and converted to concentration using Beer-Lambert Law.

As expected there was a positive correlation between 4-nitrophenol concentration and time. The results were very clean and there were no obviously anomalous or outlying results generated. The gradient of each data series was calculated and used to represent the rate of product formation in mmol $dm^3.s^{-1}$ (table 2.3).

PNP-gal concentration (mM)	**Rate of 4-nitrophenol production by 5 µM β-glucuronidase mutant 5025/46.b1 (mmol $dm^{-3}.s^{-1}$)**	**Rate of 4-nitrophenol production by 10 µM β-glucuronidase mutant 5025/46.b1 (mmol $dm^{-3}.s^{-1}$)**
0.1	5.92×10^{-7}	1.07×10^{-6}
0.5	2.32×10^{-6}	4.98×10^{-6}
1	4.45×10^{-6}	9.11×10^{-6}
2.5	9.53×10^{-6}	1.93×10^{-5}
5	1.46×10^{-5}	3.03×10^{-5}
6.5	1.78×10^{-5}	3.61×10^{-5}
8	1.99×10^{-5}	3.99×10^{-5}
10	2.32×10^{-5}	4.53×10^{-5}
12.5	2.58×10^{-5}	5.22×10^{-5}
15	2.86×10^{-5}	5.62×10^{-5}
17.5	3.00×10^{-5}	6.26×10^{-5}
20	2.97×10^{-5}	6.22×10^{-5}

Table 2.3: Rate of 4-nitrophenol production from the hydrolysis of substrate mimic PNP-gal by 5 µM and 10 µM β-glucuronidase which was calculated from the gradient of each data series plotting 4-nitrophenol concentration against time.

The rates were used to produce a Hanes-Woolf plot (figure 2.18). The Hanes-Woolf plot is a linear derivation of the Michaelis-Menten equation that fits the equation for a straight line y=mx+c in the form:

$$\frac{[s]}{v} = \frac{1}{V_{max}}[s] + \frac{K_M}{V_{max}}$$

From this graph, the V_{max}, K_M, k_{cat} and k_{cat}/K_M can be calculated. In error the assay was only carried out once at each protein concentration so the error of each data point cannot be calculated.

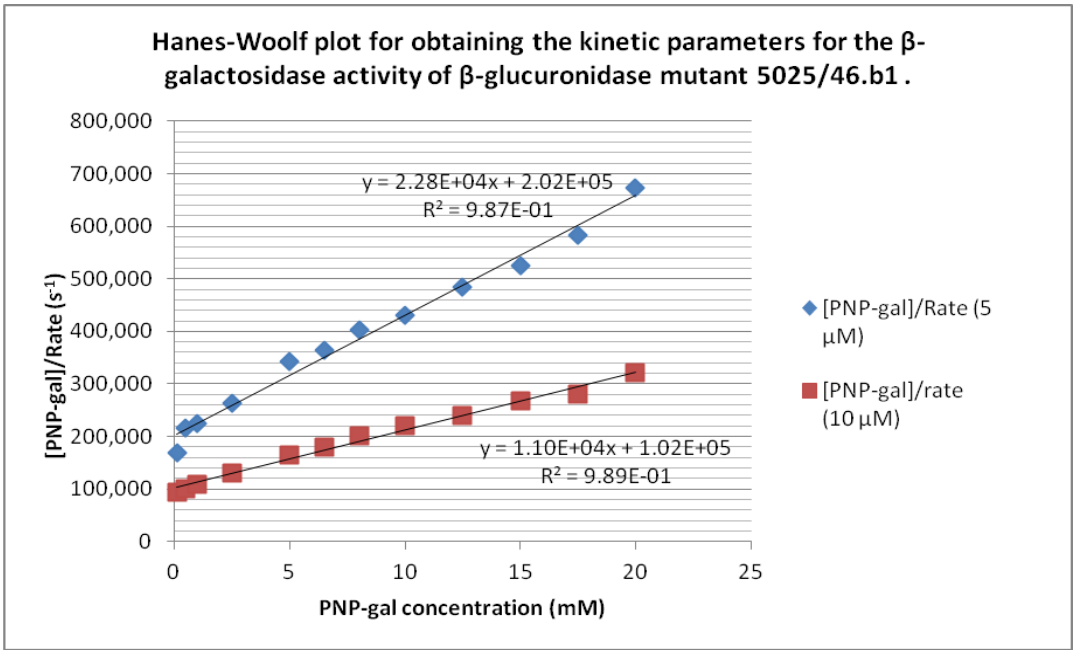


Figure 2.18: Hanes-Woolf plot of the rates obtained from the hydrolysis of PNP-gal by β-glucuronidase mutant 5025/46.b1.

The Vmax is the predicted maximum rate that an enzyme can be expected to achieve.

V_{max} was calculated for the 5 µM concentration mutant β-glucuronidase as:

$$Graph\ gradient = \frac{1}{V_{max}} \therefore V_{max} = \frac{1}{graph\ gradient}$$

$$V_{max} = \frac{1}{2.28 \times 10^{4}} = 4.39 \times 10^{-5}\ mmol\ dm^{-3}.s^{-1}$$

V_{max} was calculated for the 10 µM concentration mutant β-glucuronidase as:

$$V_{max} = \frac{1}{1.10 \times 10^{4}} = 9.09 \times 10^{-5}\ mmol\ dm^{-3}.s^{-1}$$

From this the K_M for the 5 µM concentration mutant β-glucuronidase was calculated as:

$$Y\ intercept = \frac{K_M}{V_{max}} \therefore K_M = Y\ intercept \times V_{max}$$

$$K_M = 2.02 \times 10^{5}\ s^{-1} \times 4.39 \times 10^{-5}\ mmol\ dm^{-3}.s^{-1} = 8.86\ mM$$

The K_M for the 10 µM concentration mutant β-glucuronidase was calculated as:

$$K_M = 1.02 \times 10^{5}\ s^{-1} \times 9.09 \times 10^{-5}\ mmol\ dm^{-3}.s^{-1} = 9.27\ mM$$

The K_M, the Michaelis-Menten constant is an equilibrium constant relating to when 50% of active sites are filled. It is also a measure of the affinity of an enzyme for its substrate; therefore the smaller the value for K_M, the greater the affinity of the enzyme for its substrate.

The k_{cat} is a rate constant relating to how many reactions can theoretically be catalysed by the enzyme per second. It is also known as the turnover number.
The k_{cat} for the 5 µM concentration mutant β-glucuronidase was calculated as:

$$k_{cat} = \frac{V_{max}}{[E]} = \frac{4.39 \times 10^{-5}\ mmol\ dm^{-3}.s^{-1}}{0.005\ mmol\ dm^{-3}} = 0.00878\ s^{-1}$$

The k_{cat} for the 10 µM concentration mutant β-glucuronidase was calculated as:

$$k_{cat} = \frac{9.09 \times 10^{-5}\ mmol\ dm^{-3}s^{-1}}{0.01\ mmol\ dm^{-3}} = 0.00909\ s^{-1}$$

The k_{cat}/K_M is a measure of how efficient the enzyme is, with a larger value denoting a more efficient enzyme.
The k_{cat}/K_M for the 5 µM concentration mutant β-glucuronidase was calculated as:

$$\frac{k_{cat}}{K_M} = \frac{0.00878\ s^{-1}}{8.86\ mM} = 9.91 \times 10^{-4}\ s^{-1}mM^{-1}$$

The k_{cat}/K_M for the 10 µM concentration mutant β-glucuronidase was calculated as:

$$\frac{k_{cat}}{K_M} = \frac{0.00909\ s^{-1}}{9.27\ mM} = 9.81 \times 10^{-4}s^{-1}mM^{-1}$$

Though there is some difference between the parameters obtained from the two different enzyme concentrations, the parameters are relatively close. Had the assay been carried out in triplicate a mean could have been generated for each parameter and the errors properly calculated.

The residual β-glucuronidase activity of the mutant was also measured by a similar assay which measured the residual β-glucuronidase activity using a 96 well plate in an automated spectrophotometer. Increasing concentrations of protein from 2 nM to 20 nM were used to hydrolyse the substrate 4-nitrophenol-β-D-glucuronic acid (PNP-glu) of concentrations from 0.01 mM to 5 mM. The absorbance of the resulting product, 4-nitrophenol, was measured at 405 nm over a 5000 second time course. This provided results which were above the range of the spectrophotometer (listed as “over” in the associated machine output file). Therefore the absorbance measurements for the first 1500 seconds were used for the analysis. Each absorbance measurement was converted into a concentration using the Beer-Lambert law, $A=\varepsilon cl$, where A is the absorbance, ε is the extinction coefficient given in literature of 17400 mol $dm^{-3}.cm^{-1}$, and the path length, l, calculated from the diameter of a well and the volume of

reaction, using the formula for the volume of a cylinder (volume=π.radius^2.height), where height is equivalent to path length. These concentrations were used to plot graphs of time versus 4-nitrophenol concentration (figures 2.19 - 2.22).

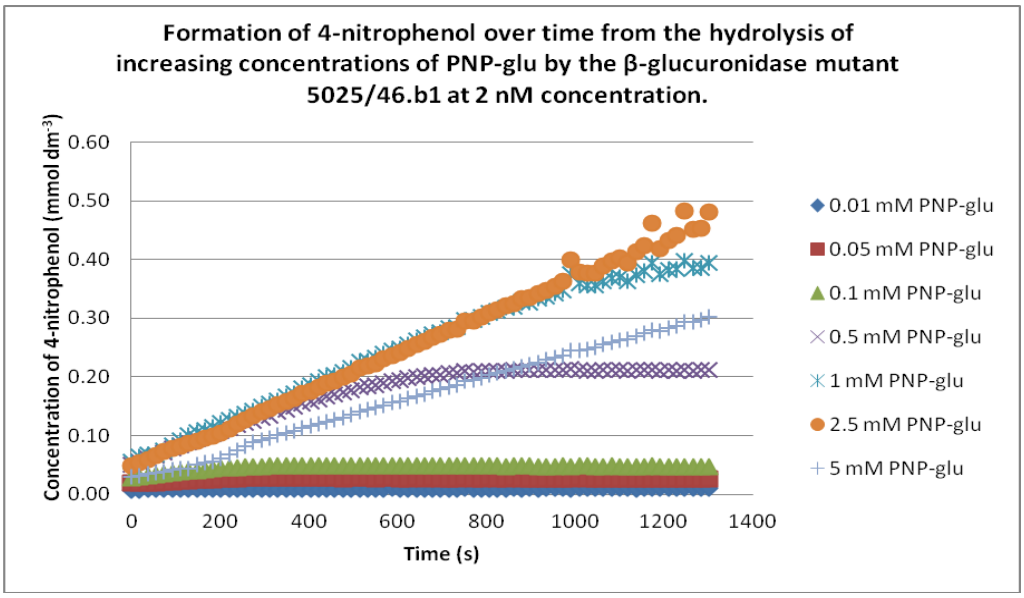


Figure 2.19: formation of 4-nitrophenol over time for the hydrolysis of increasing concentrations of PNP-glu by the β-glucuronidase mutant 5025/46.b1 at a concentration of 2 nM. Absorbance measured by spectrophotometer at 405 nm and converted to concentration using Beer-Lambert Law.

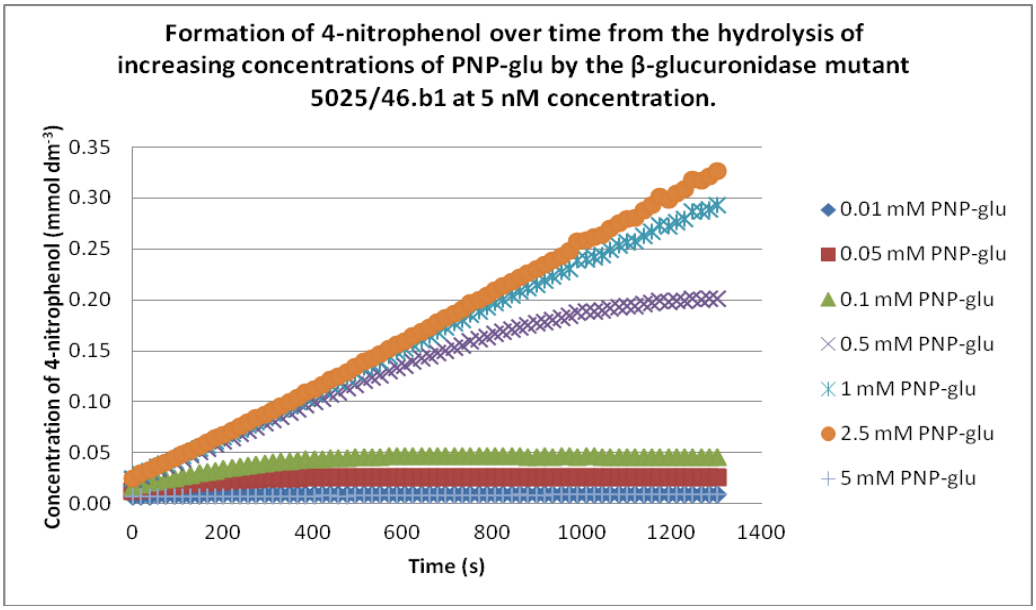


Figure 2.20: formation of 4-nitrophenol over time for the hydrolysis of increasing concentrations of PNP-glu by the β-glucuronidase mutant 5025/46.b1 at a concentration of 5 nM. Absorbance measured by spectrophotometer at 405 nm and converted to concentration using Beer-Lambert Law.

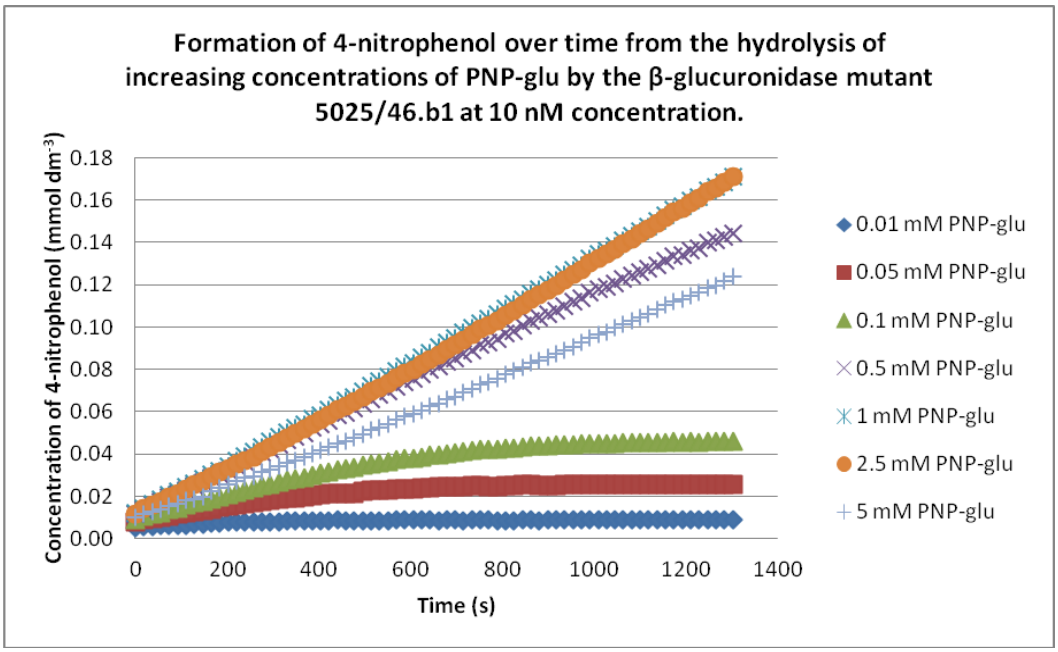


Figure 2.21: formation of 4-nitrophenol over time for the hydrolysis of increasing concentrations of PNP-glu by the β-glucuronidase mutant 5025/46.b1 at a concentration of 10 nM. Absorbance measured by spectrophotometer at 405 nm and converted to concentration using Beer-Lambert Law.

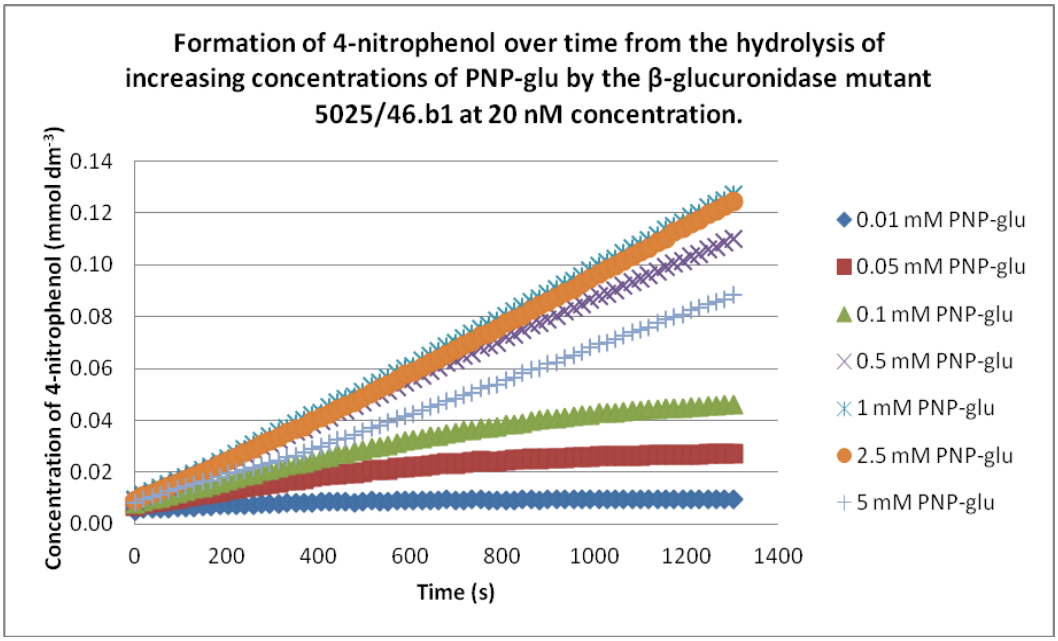


Figure 2.22: formation of 4-nitrophenol over time for the hydrolysis of increasing concentrations of PNP-glu by the β-glucuronidase mutant 5025/46.b1 at a concentration of 20 nM. Absorbance measured by spectrophotometer at 405 nm and converted to concentration using Beer-Lambert Law.

These graphs showed that in a number of cases the substrate had been fully depleted as the 4-nitrophenol concentration reached a steady value. The data was used for the first ~250 seconds, which included the linear portion of the reactions and used to plot graphs of time versus 4-nitrophenol concentration (figures 2.23 - 2.26). The results showed that across the assay, the hydrolysis involving 5 mM PNP-gal had produced unusual results as less 4-nitrophenol was produced compared to the amount produced from lower concentrations of substrate. Additionally the first measured concentration of 4-nitrophenol produced from the 1 mM substrate was higher than that of the 2.5 and 5 mM substrate at most protein concentrations. There is noticeably an unexpected result that relatively less 4-nitrophenol is produced overall as the β-glucuronidase concentration increases. It would be expected that a higher enzyme concentration would produce more product overall.

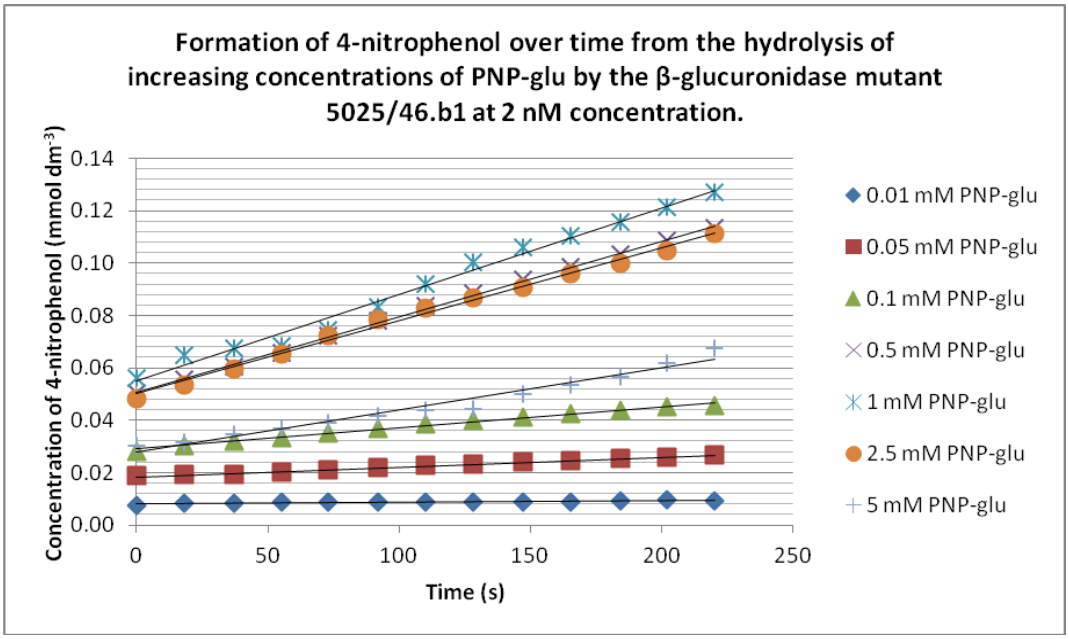


Figure 2.23: formation of 4-nitrophenol over time for the hydrolysis of increasing concentrations of PNP-glu by the β-glucuronidase mutant 5025/46.b1 at a concentration of 2 nM. Absorbance measured by spectrophotometer at 405 nm and converted to concentration using Beer-Lambert Law.

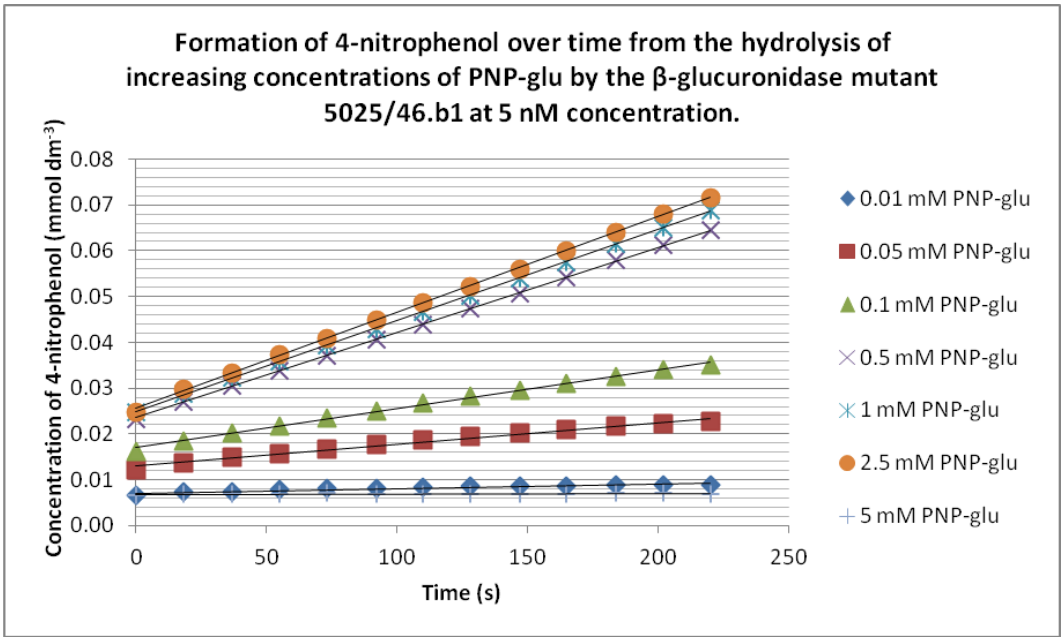


Figure 2.24: formation of 4-nitrophenol over time for the hydrolysis of increasing concentrations of PNP-glu by the β-glucuronidase mutant 5025/46.b1 at a concentration of 5 nM. Absorbance measured by spectrophotometer at 405 nm and converted to concentration using Beer-Lambert Law.

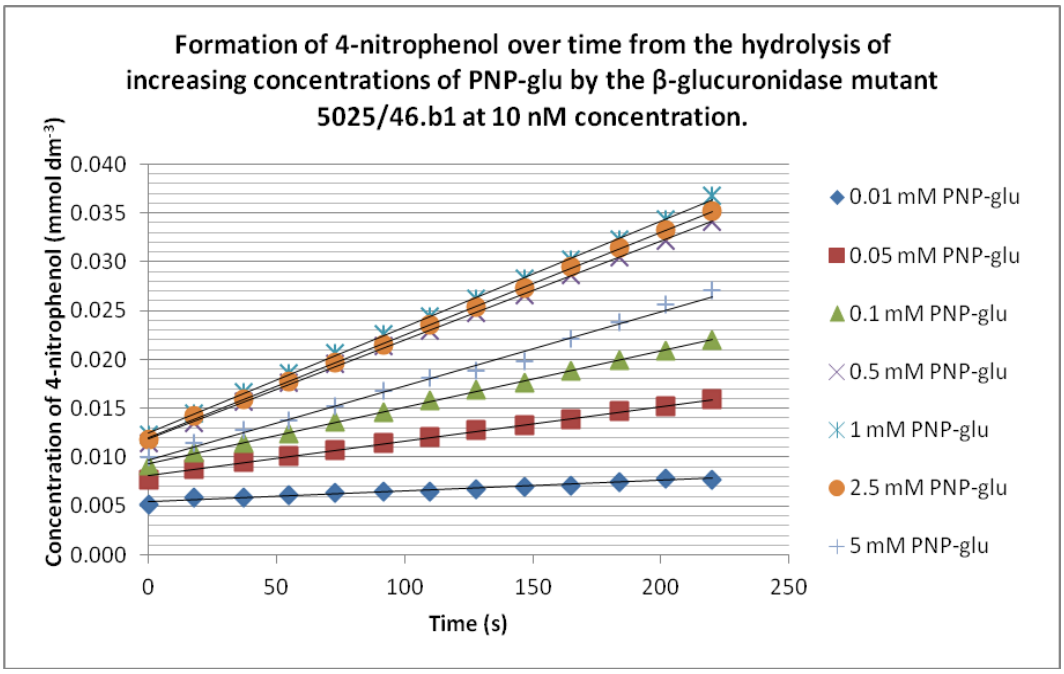


Figure 2.25: formation of 4-nitrophenol over time for the hydrolysis of increasing concentrations of PNP-glu by the β-glucuronidase mutant 5025/46.b1 at a concentration of 10 nM. Absorbance measured by spectrophotometer at 405 nm and converted to concentration using Beer-Lambert Law.

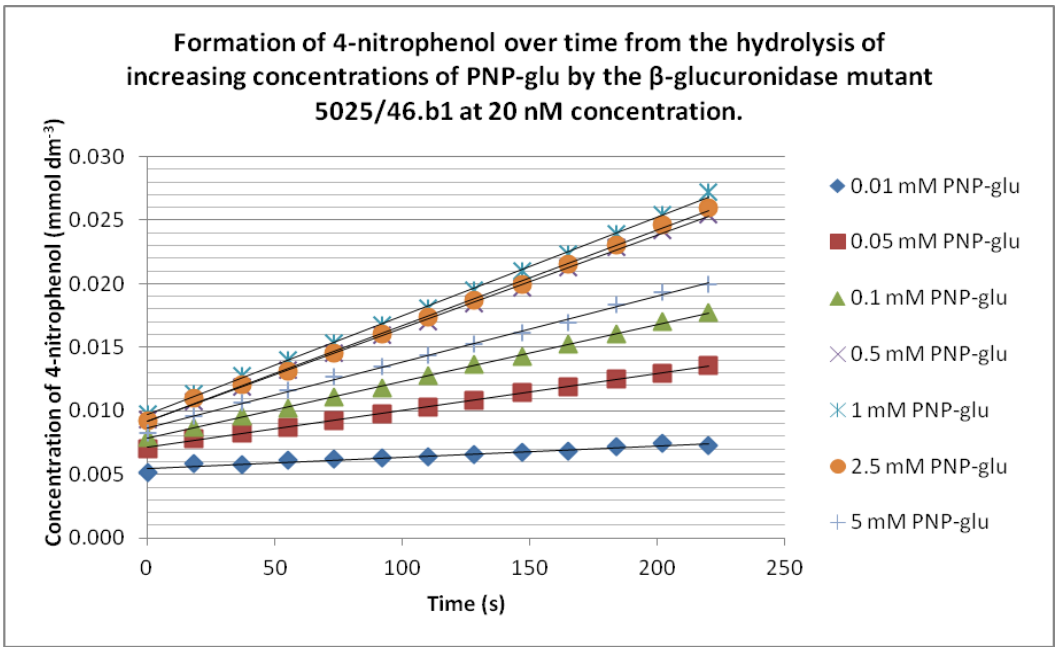


Figure 2.26: formation of 4-nitrophenol over time for the hydrolysis of increasing concentrations of PNP-glu by the β-glucuronidase mutant 5025/46.b1 at a concentration of 20 nM. Absorbance measured by spectrophotometer at 405 nm and converted to concentration using Beer-Lambert Law.

The gradient of each data series was calculated which is equivalent to a rate of the reaction (table 2.4). These were used to produce a Hanes-Woolf plot (figure 2.27) and to calculate the kinetic parameters of the enzyme for residual β-glucuronidase activity.

PNP-glu (mM)	Rate of 4-nitrophenol production by 2 nM β-glucuronidase mutant 5025/46.b1 (mmol $dm^{-3}.s^{-1}$)	Rate of 4-nitrophenol production by 5 nM β-glucuronidase mutant 5025/46.b1 (mmol $dm^{-3}.s^{-1}$)	Rate of 4-nitrophenol production by 10 nM β-glucuronidase mutant 5025/46.b1 (mmol $dm^{-3}.s^{-1}$)	Rate of 4-nitrophenol production by 20 nM β-glucuronidase mutant 5025/46.b1 (mmol $dm^{-3}.s^{-1}$)
0.01	6.47×10^{-6}	9.52×10^{-6}	1.09×10^{-5}	8.74×10^{-6}
0.05	3.74×10^{-5}	4.66×10^{-5}	3.58×10^{-5}	2.91×10^{-5}
0.1	8.05×10^{-5}	8.46×10^{-5}	5.75×10^{-5}	4.47×10^{-5}
0.5	2.89×10^{-4}	1.86×10^{-4}	1.01×10^{-4}	7.34×10^{-5}
1	3.31×10^{-4}	1.99×10^{-4}	1.08×10^{-4}	7.74×10^{-5}
2.5	2.79×10^{-4}	2.10×10^{-4}	1.05×10^{-4}	7.51×10^{-5}
5	1.60×10^{-4}	1.61×10^{-4}	7.60×10^{-5}	5.22×10^{-5}

Table 2.4: Rate of 4-nitrophenol production from the hydrolysis of substrate mimic PNP-glu by 2, 5, 10 and 20 nM β-glucuronidase calculated from the gradient of each data series plotting 4-nitrophenol concentration against time.

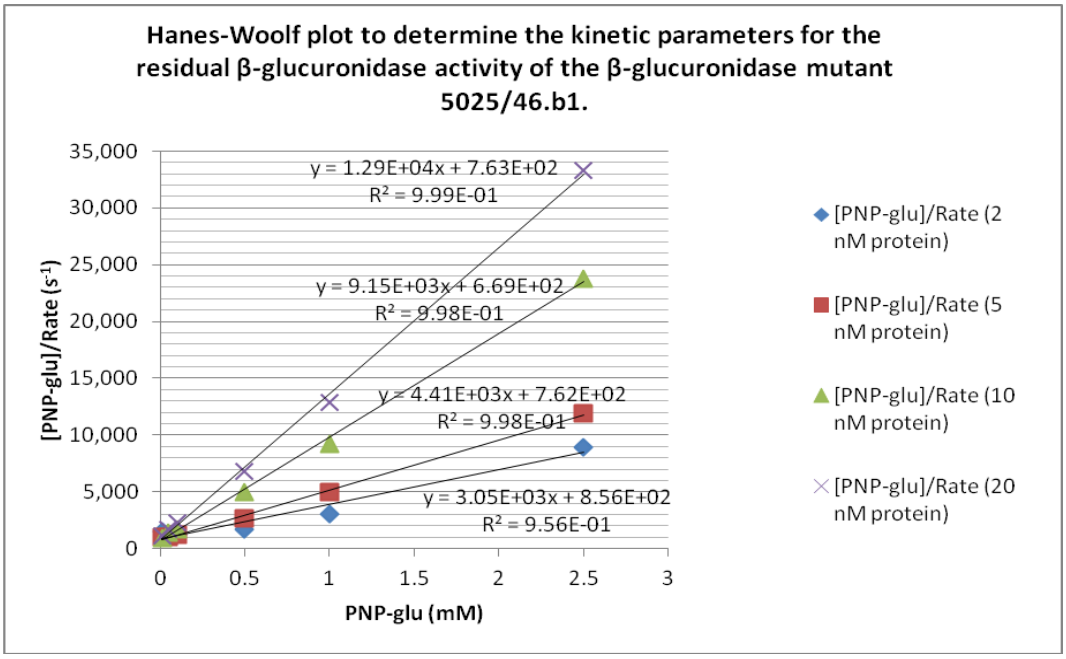


Figure 2.27: Hanes-Woolf plot of the rates obtained from the hydrolysis of PNP-glu by the mutant β-glucuronidase 5025/46.b1.

Vmax was calculated for each protein concentration:

2 nM:

$$V_{max} = \frac{1}{gradient} = \frac{1}{3.05 \times 10^{3}} = 3.28 \times 10^{-4}\ mmol\ dm^{-3} s^{-1}$$

5 nM:

$$V_{max} = \frac{1}{gradient} = \frac{1}{4.41 \times 10^{3}} = 2.27 \times 10^{-4}\ mmol\ dm^{-3} s^{-1}$$

10 nM:

$$V_{max} = \frac{1}{gradient} = \frac{1}{9.15 \times 10^{3}} = 1.09 \times 10^{-4}\ mmol\ dm^{-3} s^{-1}$$

20 nM:

$$V_{max} = \frac{1}{gradient} = \frac{1}{1.29 \times 10^{4}} = 7.75 \times 10^{-5}\ mmol\ dm^{-3} s^{-1}$$

K_M was calculated for each protein concentration:

2 nM:

$$K_M = Y\ intercept \times V_{max} = 8.56 \times 10^{2} \times 3.28 \times 10^{-4} = 0.28\ mM$$

5 nM:

$$K_M = Y\ intercept \times V_{max} = 7.62 \times 10^{2} \times 2.27 \times 10^{-4} = 0.17\ mM$$

10 nM:

$$K_M = Y\ intercept \times V_{max} = 6.69 \times 10^{2} \times 1.09 \times 10^{-4} = 0.073\ mM$$

20 nM:

$$K_M = Y\ intercept \times V_{max} = 7.63 \times 10^{2} \times 7.75 \times 10^{-4} = 0.59\ mM$$

The k_{cat} was calculated for each protein concentration:

2 nM

$$k_{cat} = \frac{V_{max}}{[E]} = \frac{3.28 \times 10^{-4}}{0.002} = 0.164\ s^{-1}$$

5 nM

$$k_{cat} = \frac{V_{max}}{[E]} = \frac{2.27 \times 10^{-4}}{0.005} = 0.0454\ s^{-1}$$

10 nM

$$k_{cat} = \frac{V_{max}}{[E]} = \frac{1.09 \times 10^{-4}}{0.010} = 0.0109\ s^{-1}$$

20 nM

$$k_{cat} = \frac{V_{max}}{[E]} = \frac{7.75 \times 10^{-4}}{0.02} = 0.0388\ s^{-1}$$

The k_{cat}/K_M was calculated for each protein concentration:

2nM

$$\frac{k_{cat}}{K_M} = \frac{0.164}{0.28} = 0.586\ s^{-1}.mmol\ dm^{-3}$$

5 nM

$$\frac{k_{cat}}{K_M} = \frac{0.0454}{0.17} = 0.267\ s^{-1}.mmol\ dm^{-3}$$

10 nM

$$\frac{k_{cat}}{K_M} = \frac{0.0109}{0.073} = 0.149\ s^{-1}.mmol\ dm^{-3}$$

20 nM

$$\frac{k_{cat}}{K_M} = \frac{0.0388}{0.059} = 0.657\ s^{-1}.mmol\ dm^{-3}$$

From these results it can be seen that there is no consensus amongst the results at different protein concentrations. This is a problem that was not noticed at the time, but is consistent with the odd results of the assay overall. The assay should have been run in triplicate and repeated again in order to rule out problems with the assay and allow the errors to be calculated.

The W529L mutant from library WS1, had previously undergone kinetic analysis and was found to have a K_M of 7.3 mM, k_{cat} of 0.063 s^{-1} and k_{cat}/K_M of 8.6 $mM^{-1}.s^{-1}$ in relation to β-galactosidase activity, and K_M of 280 μM, k_{cat} of 83 s^{-1} and K_{cat}/K_M of 300 000 $s^{-1}.M^{-1}$ in relation to native β-glucuronidase activity[16]. In comparison the W529L/K370R/G245A mutant has a K_M of 9.21 mM, k_{cat} of 0.00909 s^{-1} and a k_{cat}/K_M of 0.98 $mM^{-1}.s^{-1}$ in relation to β-galactosidase activity, showing that it is less efficient at catalysing β-galactosides than the mutant containing W529L alone.

Wild type β-glucuronidase has a K_M of 2.7 mM, k_{cat} of 0.006 s^{-1} and k_{cat}/K_M of 2.2 $s^{-1}.M^{-1}$ with regards to β-galactoside activity, and K_M of 260 μM, k_{cat} of 109 s^{-1} and k_{cat}/K_M of 410 000 $s^{-1}.M^{-1}$ with regards to native β-glucuronidase activity (table 2.5). Comparing the wild type to the two mutants for β-galactosidase activity highlights that the K_M increases, thus the affinity for the enzyme to the new substrate decreases compared to the wild type. This demonstrates that the enzyme is not well adapted for β-galactosidase activity, particularly compared to the T509A/S557P/N566S/K568Q found by Matsumura which has decreased K_M for β-galactosidase activity compared to wild type β-glucuronidase. The K_M for the G245A/K370R/W529L mutant is further increased compared to the K_M of the single W529L mutant. The k_{cat} for W529L was improved towards β-galactosidase activity by a power of 10, but the G245A/K370R/W529L mutant has very little improvement. The great difference is seen in the k_{cat}/K_M, a measure of the efficiency of the enzyme. The efficiency of the enzyme is greatly diminished by the mutations, especially in the G245A/K370R/W529L mutant which decreases k_{cat}/K_M from 2.2 $s^{-1}.M^{-1}$ to

0.00098 $s^{-1}.M^{-1}$. Overall, from these comparisons, the K370R and G245A mutations conferred no benefits; in fact they had a deleterious effect to galactosidase activity.

Mutant	**β-glucuronidase activity**			**β-galactosidase activity**		
	K_M (mM)	**k_{cat} (s^{-1})**	**k_{cat}/K_M ($s^{-1}.M^{-1}$)**	**K_M (mM)**	**k_{cat} (s^{-1})**	**k_{cat}/K_M ($s^{-1}.M^{-1}$)**
Wild type	0.260	109	410000	2.7	0.006	2.2
W529L	0.28	83	300000	7.3	0.063	0.0086
5025/46.b1 (G245A/K370R /W529L)	Not enough number agreement.			9.21	0.00909	0.00098

Table 2.5: comparative constants for the wild type, W529L and G245A/K370R/W529L mutants obtained from an assay measuring the UV/Vis absorbance at 405 nm using a plate reader. The absorbances were converted to concentrations and graphs plotted of concentration over time. The rates were used to generate a Hanes-Woolf plot from which the constants were obtained.

Kinetic analysis was not carried out on any other mutants from this library. Kinetic analysis should have been carried out on the other mutants. There is no explanation in the laboratory notebook as to why this was not done. Mutants 5025/46.b1 and 5025/46.b2 both share the K370R and W529L mutations in common; however, there are two further mutation sites in 5025/46.b2: V67D and N412D. Though V67D is distant from the active site, this mutation is significantly different changing the hydrophobic valine residue for the negatively charged aspartic acid, whilst N412D is the adjacent residue to E413, which is one of the catalytic residues. A change from the amide asparagine to the negatively charged aspartic acid is likely to have had interesting effects. Potentially it may have allowed new hydrogen bonds to be formed with the β-galactopyranoside, which improve the catalytic activity.

As the aim was to find new and improved β-galactosidase active mutants, the remaining mutants were thought to be unimportant due to the pale blue colour generated in the X-gal screen.

2.2.5 Comparison of mutation sites between experiments

As discussed earlier, the mutation W529L has previously been found by both our experiments and those of Rowe[3]. In Rowe's experiment, which was built around screening and selection with PNP-gal, the mutants containing W529L (I12V/F365S/W529L/S557P/I560V and F365S/S475C/W529L/S557P) were found to be white in a secondary screen with X-gal. Unlike Rowe, under the conditions used in our X-gal screen this mutation appeared blue. Rowe suggested that there is a phenotypic preference for screening method/substrate with certain genotypes producing a phenotype selective to PNP-gal over X-gal, particularly in the context of

F365S/W529L. It is possible that a mutation in the enzyme could affect the binding ability of either X-gal or PNP-gal since neither substrate shares the exact same shape as the native substrate. Both X-gal and PNP-gal have the galactopyranoside sub-unit attached to a planar ring system rather than an identical galactopyranoside sub-unit as found in the native substrate. Therefore mutations which alter the shape of the active site or binding pocket could cause a phenotypic preference for a certain substrate. However in this instance it seems unlikely to have been the W529L mutation site that dictates a preference for PNP-gal. F365S was also identified in this library in a variant which manifested as a pale blue colony when screened with X-gal. From this it seems unlikely that F365S causes a phenotypic preference for PNP-gal over X-gal. As F365S was not identified in conjunction with W529L, it is possible that both mutations have to be present for the effect to take place. If the F365S/W529L double mutation does have a phenotypic preference for PNP-gal over X-gal, and was generated during the preparation of the library, it would not have been identified from this experiment due to the primary screen being with X-gal.

Unlike W529L and F365S, none of the other mutations identified in this experiment (A44E, V67D, P76L, S127R, C133R, K155R, N166T, I187F, P239Q, Q243H, G245A, S258R, G356V, K370R, N412D and L583Q) had previously been found in any of the direct selection experiments for changing substrate selectivity. A44E and K370 have been identified as sites of mutation by Flores and Ellington[5] in their experiments to improve the thermostability of β-glucuronidase. In that experiment K370E was found along with A44E, but neither were found to be an important mutation for conferring thermostability.

The mutation sites identified across all the variants from library JH2 were plotted onto the crystal structure of β-glucuronidase (figure 2.28). The active site was highlighted with the catalytic residues E413, E504 and the important residue Y468 (shown in pink). All other residues were added and coloured purple to highlight. F365 is not shown as the crystal structure is missing eight amino acids including this residue.

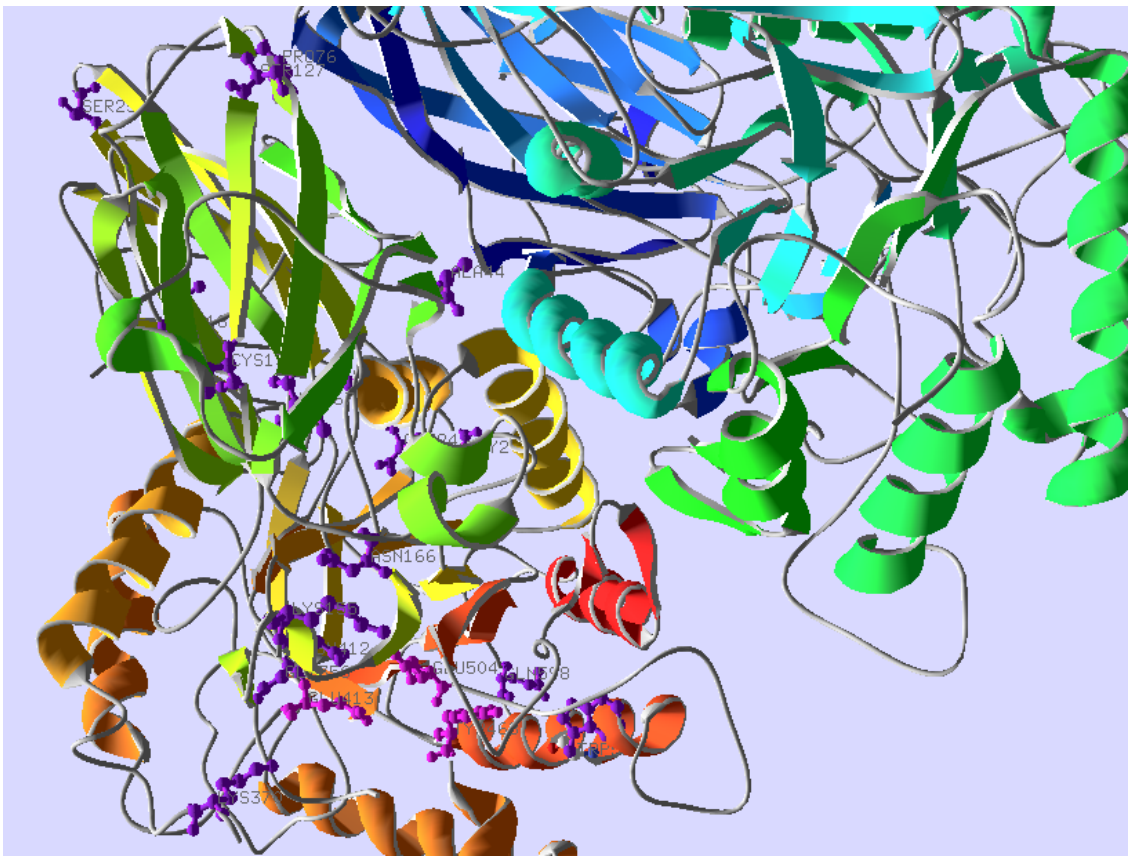


Figure 2.28: crystal structure of β-glucuronidase demonstrating the location of all residues found in galactosidase active variants. Catalytic and important residues E413, Y468 and E504 are shown in pink. All other residues are shown in purple. Picture generated using the Swiss pdbViewer. Green and blue rubbon structure is formed from another monomer and demonstrates the proximity of some residues to the monomer interface.

Residues A44, S127, and P76 are located distant from the active site at the mirrored interface with the other monomer. S258, I187, C133, V67, Q243, G245, and P239 are located distant from the active site. The remaining mutations were located close to the active site in the loops of the α/β-barrel. Of the changes to residues in and around the active site, the change from lysine to arginine in K155R and K370R does not change the overall structure significantly. Lysine and arginine are similar in chain length and arginine retains the positive charge found in lysine. K370R appears to have the side chain reaching out from the monomer; therefore it could be involved in stabilising the quaternary structure of the enzyme. The change from asparagine to threonine in N166T changes the ability of the residue to form hydrogen bonds and creates more space as threonine has a smaller chain length. This could be important in accommodating the change of position of the 4'-OH from below to above the plane of the ring in the substrate. G356 lies very close to the catalytic residues. The change from glycine to valine in G356V introduces

a new hydrophobic residue to the active site. The active site is located at the bottom of the barrel where more hydrophobic residues would be expected. Adding further hydrophobic residues may increase the stability of the active site. The change from asparagine to aspartic acid in N412D changes the neutral amide to the negatively charged aspartic acid. N412D is directly next to the catalytic residue E413, and also appears from the crystal structure picture to reach into the active site. Therefore it could contribute to binding the substrate through hydrogen bonds forming between the carboxylic acid and alcohol group on C6 of the substrate. It could also contribute to catalysis. C133R loses a cysteine from the protein and potentially a disulphide bridge, which may allow more conformational flexibility in the enzyme, which improves binding to the alternative substrate. Likewise the change from proline to leucine in P76L may introduce new flexibility. The cyclic structure of proline is used to change direction of the backbone in proteins, so the loss of the proline increases the movement of the backbone. The change from tryptophan to leucine in W529L loses a potential hydrogen bonding site and introduces a smaller, hydrophobic residue to the α/β-barrel. It is difficult to rationalise the effects of other mutations which are distant from the active site and do not appear to have any interactions with the substrate or catalytic residues, but it is well known that residues distant from the active site may exert an effect throughout the protein.

It is not possible to assess from the results whether any of the other mutations besides W529L are important for activity. The kinetic data from the G245A/K370R/W529L mutant suggests that the G245A and K370R mutations were slightly deleterious to activity towards galactopyranosides. From colour the other mutations were not as important to activity as W529L.

It is interesting to note that this experiment did not produce any of the common important mutations found previously (T509A, S557P, N566S and K568Q). This is perhaps due to the use of the GenemorphII kit for mutagenesis, which eliminates most of the bias produced by using Taq or GenemorphI alone; even if both methods of mutagenesis have previously been used.

Both darker coloured blue mutants (5025/46.b1 and 5025/46.b2) contain both the W529L and K370R mutations, along with the silent mutation N324N. Particularly given that both mutants also contain the same silent mutation, it is likely that both mutants came from a common ancestor produced during the drift. In order to see if it was possible to find out in which round the W529L/K370R mutation had been produced, each round from library JH2 was screened against X-gal.

Plasmid stocks, prepared from the colonies harvested from XL1 Blue transformations, were transformed into BW25141 cells by electroporation and cultured on LB Ampicillin arabinose plates spread with X-gal. A surprising result was produced: many of the colonies screened were blue. This was repeated and many of the colonies were pale blue. Each round of libraries WS1

and WS2 were also screened on two occasions. On the first of these, the screened colonies remained white and no blue colonies were found. On the second occasion, a large number of the colonies had developed a very pale blue coloration indicating some β-galactosidase activity. The number of blue colonies meant that it would be impractical to locate the source round of the double K370R/W529L mutant from library JH2, since locating its origin would have involved sequencing every blue colony. This wasn't possible due to the numbers produced (>6000 colonies across both screening experiments on JH2). It was also impractical to sequence blue colonies from libraries WS1 and WS2 as similar numbers of pale blue colonies were produced.

Wild type β-glucuronidase had previously been shown, during these experiments and in literature, to produce a pale blue coloration when screened with X-gal, therefore it is possible that many of the blue colonies observed were neutral β-glucuronidases exhibiting β-galactosidase activity at levels similar to that of the wild type β-glucuronidase. This seems to correlate with the apparent inconsistency of the colonies containing plasmid from a single stock screening as white and blue on different occasions. However it is possible that some of the mutants could have been new variants with activity for β-galactosidase, but due to the inability to sequence every colony or distinguish activity by colour alone, it was not possible to find this out.

This assay was the equivalent to doing a standard directed evolution experiment by screening each round for the new activity. The electroporations for the assay used plasmid from each round of library generation, which contained the β-glucuronidase gene that had been mutated, but not yet selected for β-glucuronidase activity. Therefore it is possible that some of the colonies that were β-galactosidase active may not have been β-glucuronidase active if screened for that activity. Such variants that contain mutations which confer β-galactosidase activity, but are deleterious to β-glucuronidase activity in isolation, would have been eliminated from the library, in the neutral drift experiment, at the subsequent step of the cycle: the X-glu screening.

2.2.6 Neutral libraries can be screened against a wide variety of substrates

The advantage of this experimental method is that it produces a neutral library of variants which can be screened for activity towards other related compounds. In order to investigate whether the neutral library also contained variants with activity towards other pyranose sugars, libraries WS1, WS2 and JH2 were screened against five selected pyranose sugar substrates that had a range of structures with similarities and differences compared to β-glucuronic acid and β-galactopyranose. Like the previous experiments, the qualitative blue/white screening approach with 5-bromo-4-chloro-3-indolyl-pyranosides. The five pyranose sugars were 5-bromo-4-chloro-3-indolyl-β-D-fucopyranoside (X-fuco), 5-bromo-4-chloro-3-indolyl-N-acetyl-β-D-glucosaminide (X-gluam), 5-bromo-4-chloro-3-indolyl-N-acetyl-β-D-galactosaminide (X-galam), 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (X-gluco) and 5-bromo-4-chloro-3-indolyl-β-D-xylanopyranoside (X-xyl) (figure 2.10a, page 33).

Library JH1 was omitted from the screen as it was thought unlikely to contain any significant results, as the numbers of variants in the library was so small that most of the library was likely to be biased towards wild type, as previously discussed.

Fucopyranosides were selected as an alternative substrate for which activity may have evolved as it shares similarities in structure to β-galactopyranosides as well as differences. β-fucopyranosides share the same orientation of the 4' OH as β-galactopyranosides, but opposite to that of the β-glucuronides. Additionally the 5' functional group is changed from a carboxylic acid to a methyl group (figure 2.29).

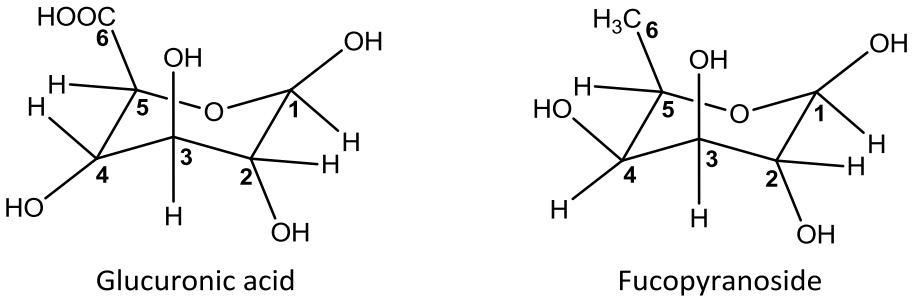


Figure 2.29: the structures of the pyranoside substrate compared to glucuronic acid demonstrating the differences at C4 and C6.

Wild-type β-glucuronidase does not exhibit β-fucosidase activity; neither did any of the round 4 mutants from libraries WS1, WS2 or JH2. Matsumura and Ellington[7] screened their direct selection experiment against fucosides and retrieved activity with the N566S, T509A/D531E/S557P/N566S, and T509A/S557P/N566S/K568Q variants. Searching the literature did not reveal any other experiments in which β-glucuronidase has been evolved to have fucosidase activity and therefore comparisons with other experiments cannot be drawn. There has been one experiment[18] in which a β-galactosidase from *E. coli* was evolved to have increased β-fucosidase activity. The structure of fucopyranoside is much closer to glactopyranoside as the orientations of the OH group on each carbon is identical, which may be why that particular transformation was chosen.

Glucosaminides and galactosaminides have a significantly different structure to glucuronic acid as both substrates have been substituted at the 2' OH by N-acetyl. Additionally C6 has been converted from COOH to CH_2OH. The orientation of the C4' OH in glucosaminide matches that of glucuronic acid, whilst the orientation of the C4' OH galactosaminide matches that of galactopyranoside (figure 2.30). Wild type β-glucuronidase demonstrates activity to the substrate without the N-acetyl modification at C2, so using these substrates provides a comparison between modified and unmodified C2 pyranosides on activity.

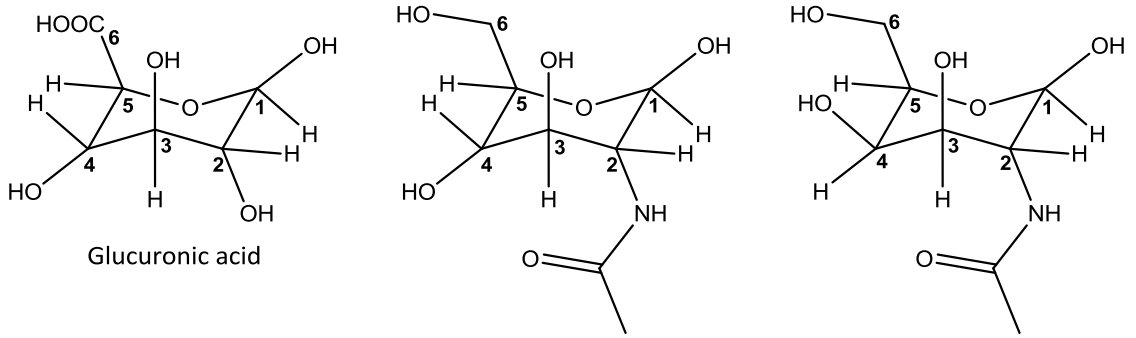


Figure 2.30: a comparison of the structures of glucuronic acid with glucosaminide and galactosaminide showing the orientation of each OH group and the N-acetyl modification at C2.

Wild type β-glucuronidase does not exhibit β-glucosaminide or β-galactosaminidase activity. This is anticipated. It would be reasonable to presume that the C2 modification significantly alters the shape and size of the substrate. Though the OH on C2 is not thought to be directly involved in binding and stability, the changes affect the ability of the substrate to form hydrogen bonds with the enzyme, and the modification may also cause steric hindrance with other residues located around the active site and binding pocket. Round 4 mutants from libraries WS1, WS2 or JH2 did not demonstrate activity towards glucosaminides or galactosaminides.

The structures of glucuronic acid and glucopyranoside are very similar differing only at the 5'-position: in glucuronic acid it is COOH and in glucopyranoside CH_2OH (figure 2.31).

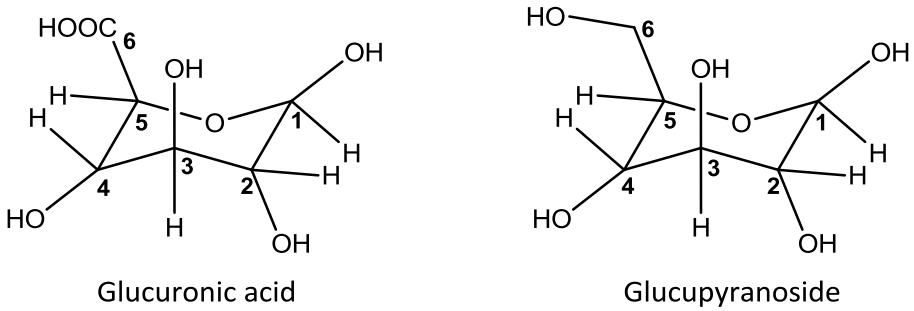


Figure 2.31: a comparison of the structures of glucuronic acid and glucopyranoside showing the similarities in the structures at C1-5 and the difference at C6.

Evidence shows that there is a strong network of hydrogen bonds between the carboxylate of glucuronic acid and the enzyme active site[2]. The change to a CH_2OH group potentially affects the number of hydrogen bonds that are able to be formed, but the substrate is still able to bind and be hydrolysed. The wild type β-glucuronidase and some of the mutants from round 4 of libraries WS1, WS2 and JH2 exhibited activity towards glucopyranoside. The numbers of

mutants active towards glucopyranoside diminished between rounds one and four of the libraries (Approximately1/3 – 1/2 of colonies lost activity completely as shown by the introduction of white colonies to the screen). Neutral variants that had lost activity towards glucopyranoside must have incurred mutations that affected the ability of glucopyranoside to bind as the catalytic residues of the enzyme remain identical, and mechanism of catalysis is the same between substrates. It would have been interesting to have selected and sequenced these colonies to have found out what the mutations were. This idea was unfortunately overlooked.

The structure of xylopyranoside differs from glucuronic acid at the 5' position, being just H_2 instead of the COOH (figure 2.32). It is reasonable to expect that the pyran ring will bind to the active site as per glucuronic acid but the lack of a group on the 5' carbon may mean that there are fewer stabilising hydrogen bonds formed, hence the weak activity of the wild type.

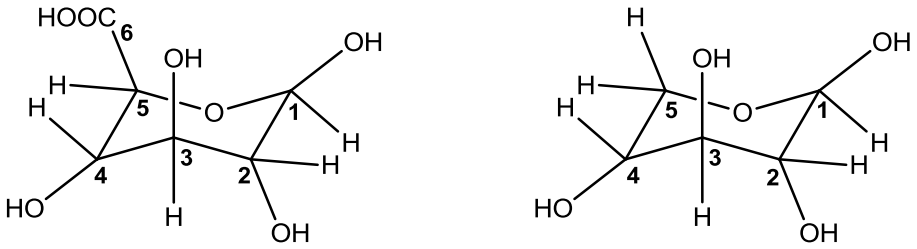


Figure 2.32: a comparison of the structures of glucuronic acid and xylopyranoside showing the similaries in structure at positions C1-C4 and the difference at C5 and lack of C6 in xylopyranoside.

Wild type β-glucuronidase exhibits weak β-xylosidase activity as shown by a very pale blue coloration of the colonies. A number of the mutants from round 4 of libraries WS1, WS2 and JH2 exhibited much stronger activity according to the shade of the blue colonies. The new active mutants may well have modifications to the active site that allow the xylopyranoside to bind better.

For comparison, a sample of unselected mutants (i.e. had been mutated but not screened for β-glucuronidase activity) from round 1 of libraries WS1, WS2 and JH2 were also screened against the alternative substrates: the colonies did not show any activity towards fucosides, glucosaminides or galactosaminides as demonstrated by the plates of white colonies. This is a logical result since the wild type β-glucuronidase showed no activity with these substrates and neither did the round 4 mutants.

The screened colonies continued to demonstrate activity towards glucopyranosides, though the introduction of white colonies to the screen demonstrated that some activity had been lost in

one round of mutagenesis. Other colonies exhibited a paler blue colour and it could be assumed that mutation has occurred to weaken the activity towards glucopyranoside by these variants.

The screened colonies continued to demonstrate weak activity towards xylopyranosides though a number of white colonies, indicating a lack of activity, were also present. The wild type β-glucuronidase exhibits weak xylosidase activity and most of the blue colonies had activity comparable to the wild type by colour shade comparison. However ~10-20 colonies exhibited increased xylosidase activity. It would have been interesting to have selected these colonies and sequenced them to find out what the mutations were, but this was again overlooked in favour of colonies selected from screening round four libraries.

From the screen of round four of each of the three libraries, eight β-xylosidase active colonies were selected for further investigation by sequencing. The results of sequencing were unexpected. Of the eight colonies, three selected mutants came from libraries WS1 and WS2 and five from JH2. These mutants were sequenced as:

1. F88S/I143V/I170V/A323T/A397E
2. A44V/Q385L/F448H/T509A
3. Q40R/A44G/Q70P/V91G/W98G/I130T/A552P/N566S (G594G)
4. (S475S)
5. (A299A)

Colonies 6-8. Wild type – no silent mutations present

The five wild type sequences returned from library JH2 were unexpected as there was a definite difference in shade of colour of the blue colony on the plate and a wild-type blue colony but no sequence changes. The most likely causes would appear to be a difference in protein expression levels or plasmid copy number between experiments. These are plausible reasons: an increase in expression levels in library JH2 would increase the amount of protein available to hydrolyse the substrate. Thus the colony would appear to be relatively darker blue in colour than the control plasmid, even though the sequences were identical. As the experimental procedures were modified between production of libraries WS1and WS2 and JH2, this could account for the difference in copy number and protein expression level.

Other causes for the result such as experimental error were considered. The first of these was if there was a difference in the concentration of X-xyl applied to the plates. As all the plates were prepared from the same batch of X-xyl and at the same time, this was dismissed as being unlikely. Another possible error would have been to have accidentally transformed with wild type plasmid but again this seems unlikely due to the difference in shade of colour between the mutants from round four of JH2 and the wild type β-glucuronidase on the positive control plate.

These results should have been cross checked by re-transforming the purified plasmid from the selected colonies and screening against X-xyl to ensure that the results of screening were consistent. The clones should then have been selected individually and grown to produce protein samples. These should have then been measured in a quantitative assay involving PNP-xyl against wild type β-glucuronidase. From this kinetic parameters could have been calculated and a more accurate representation of the activity of each variant demonstrated.

Analysing the mutation sites that were revealed shows that residue A44 appeared in two independent variants as different mutations: glycine and valine. While the glycine causes the loss of the CH_3 side chain of alanine, the valine increases the length of the side chain, but maintains the hydrophobic nature of that side chain. Plotting this residue onto the crystal structure of β-glucuronidase (figure 2.33), shows that it does not interact directly with the active site so is unlikely to affect bonding with the substrate or active site substrate bonding residues. A44E was identified as a mutation in a variant with activity towards galactospyranosides in library JH2, which suggests that this residue has some importance for substrate selectivity. A44E was also identified as a residue in an experiment to increase the thermostability of β-glucuronidase though was not found to be important. Though this residue is distant from the active site it does appear to be important for a number of functions in β-glucuronidase.

The mutations T509A and N566S were also found, though in independent variants. These are the residues previously identified as being important for β-galactosidase activity. Therefore inclusion of mutations at these sites corroborates the evidence from earlier experiments that these residues are important for enzyme activity with alternative substrates.

Looking at the overall location of all mutations found from the neutral libraries that contribute to xylosidase activity, a picture emerges of the mutations being situated in two distinct locations. The first location is roughly circling the active site with the mutations being located on the loops that contribute to the α/β- barrel structure. The second location is not in contact with the active site, and is distant from it. It is found at the mirror-image interface of two monomers of the tetrameric protein. The importance of this second cluster is hard to determine, but would seem to have some relevance as eight of the seventeen mutations found were clustered in this area. As it is located at the interface of the two monomers, it may have some importance in stabilising quaternary structure, though it is interfacial lysines which have previously been determined as important for that role.

Analysing the mutations made in this distant cluster, the mutation ofQ40R could allow the formation of an ionic bond to nearby acidic residues on either monomer, which would stabilise structure; the change of phenylalanine to serine in F88S could allow the formation of new hydrogen bonds. The mutation of glutamine to proline in Q70P is a major change, but may allow for increased flexibility in that part of the structure. It also creates space in the protein as proline

is much smaller than the amide side chain of glutamine. Likewise the change from tryptophan to glycine in W98G creates space. Other residue mutations produce much less significant changes to residue structure.

Around the active site, residue changes such as I143V make little difference to structure whereas A397E creates new opportunities for hydrogen bonds to be formed from the aspartic acid. The location of this mutation in the structure makes it unlikely to interact directly with the substrate or allow stabilising bonds to be formed with it. A552P may allow increased flexibility to the active site and binding loops, but again from location is unlikely to interact directly with the substrate.

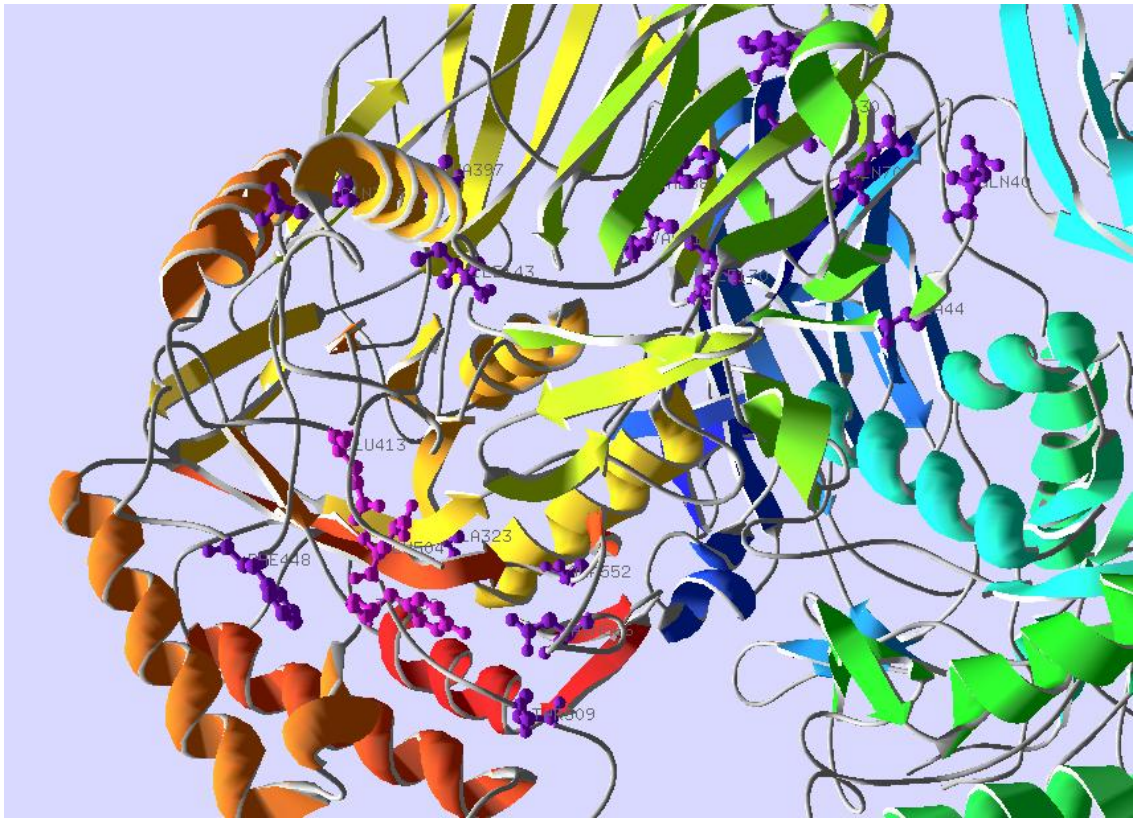


Figure 2.33: close up of the crystal structure of β-glucuronidase demonstrating the locations of the active site catalytic residues (pink) and all mutation sites increasing β-xylosidase activity generated from the neutral libraries (purple). Seven of the eight helices contributing to the α/β-barrel structure are visible surrounding the active site. The two locations of clusters of mutations can be seen. Ribbon structure coloured in bright green to red is from one monomer. Ribbon structure in dark blue to dark green is from the adjoining monomer. Image was generated using the Swiss-PdbViewer.

Geddie and Matsumura[2] previously carried out an experiment to direct the evolution of β-glucuronidases with β-xylosidase activity. That experiment chose to focus on saturation

mutagenesis of residues 557, 566 and 568 to produce xylosidase active mutants. From the results that were generated two fit variants were produced: S557P/N566A/K568F (PAF) and S557P/N566S/K568Q (PSQ). Mutations outside of the sites selected for GSSM, which occurred during long PCR in round 1 were driven to extinction, leading to the conclusion that the sites were neutral or deleterious to xylosidase activity. These mutation sites were listed as Q598Stop, D531E, and A580V. A further PSQ clone was found to contain S22N, G81S, K257E, T509A, Q598R and stop604W. This is identical to a fit clone for β-galactosidase activity from their experiments and is thought to be a result of cross contamination. Apart from this clone, the literature states that sequencing was only carried out on each variant from position 1320 of the gene onwards. This corresponds to residue 412 in the protein; therefore any mutations in the protein occurring before residue 412 would not have been recorded. This does not allow direct comparison with all the mutation sites generated by the neutral drift library screening. The only mutation site in common between the two experiments is N566S (excluding T509A, which was found in a variant not likely to be a genuine xylosidase active mutant in the Geddie experiment). N566S is likely to be important for substrate specificity and its presence in the neutral drift library is likely to confirm its importance for activity towards both substrates.

Due to the Geddie experiment not sequencing the whole gene, it is not possible to directly compare the two studies; however, it is possible that using neutral drift has highlighted other mutations that are important for xylosidase activity. There have been no other studies found in which a β-glucuronidase has been directed into a β-xylosidase so further comparisons cannot be made.

2.2.7 Is neutral drift a useful technique to employ in directed evolution experiments?

The results of the neutral drift experiments have been disappointing in that no new beneficial mutations activity for β-galactosidase were found, which had increased activity over those already discovered by direct selection. In this system, the use of neutral drift did not allow new parts of the fitness landscape to be explored or the escape from local minima in the landscape. This is, perhaps, due to the enzyme system itself which has evolved in this way. Although no new beneficial mutations were discovered, a number of other mutation sites found across other direct selection experiments were found. In this system these particular residues seem to be crucially important to the enzyme for binding and changing activity and thermostability. Therefore the use of neutral drift did also identify the same mutations. It also identified many other sites distant from the active site, which may have small effects singularly but cumulatively have an effect on substrate selectivity perhaps by stabilising the entire protein structure to accommodate the changed residues.

On closer inspection of other experiments employing neutral drift, surprisingly, a similar picture is seen: the mutations given in neutral drift studies are almost identical to those given in

standard directed evolution studies. The best variant found in Gupta and Tawfik's[13] neutral drift study on PON1 had two mutations S193T and T332S which were also found in an earlier traditional directed evolution experiment by Aharoni *et al*[19, 20]. Likewise a neutral drift study on TEM1 β-lactamase by Bershtein[11] produced the same mutants as two earlier traditional directed evolution experiments[12, 21]. For directed evolutionary studies based on random mutagenesis and direct selection it seems that using neutral drift confers no benefit, despite the positive claims made in those experiments.

Neutral drift is not entirely without benefit though. The libraries of β-glucuronidase variants prepared in these experiments were found to contain mutants with activity towards other related substrates. The mutations that were found in this study did not correspond with those found in the previous studies except for two mutation sites. Therefore neutral drift may be useful for preparing libraries when directing the evolution of a range of new enzyme activities.

2.3 Conclusions and further work

Four experiments were carried out to direct the evolution of a β-glucuronidase into a β-galactosidase using neutral drift. Two libraries were previously prepared by Dr W. Smith and two as part of the work for this thesis. No β-galactosidase active mutants were isolated from library JH1, which was concluded to be due to the very small library size and low mutation rate, which would have increased the likelihood of carrying through large numbers of wild type sequences. From library JH2 five β-galactosidase active mutants were isolated, which had altered protein sequences:

1. G245A/K370R/W529L
2. V67D/N166T/K370R/N412D/W529L
3. A44E/P76L/G356V
4. S127R/C133R/K155R/I187F/P239Q/Q243H/S528R
5. F365S/L583Q

Of these mutation sites A44, K370, F365, W529 have been identified in other experiments: A44 and K370 in a thermostability experiment (as A44E and K370E), and F365 and W529L in a direct selection experiment for β-galactosidase activity. W529L was also found during the preparation of neutral drift libraries WS1 and WS2. The four key mutations that had been found in other direct selection experiments (T509A, S557P, N566S and K568Q) were not found in libraries JH1 and JH2, but were in libraries WS1 and WS2. The variant G245A/K370R/W529L was fully characterised, but was found to be a less successful β-galactosidase than the single W529L mutant found in library WS1/WS2; neither were as successful as the T509A/S557P/566S/K568Q variant found by direct selection.

Three of the four neutral drift libraries (WS1, WS2 and JH2) were screened against five alternative activities: β-fucosidase, β-glucosaminidase, β-galactosaminidase, β-glucosidase and β-xylosidase. This identified three variants with β-xylosidase activity:

1. F88S/I143V/I170V/A323T/A397E
2. A44V/Q385L/F448H/T509A
3. Q40R/A44G/Q70P/V91G/W98G/I130T/A552P/N566S

Of these mutation sites A44 was again identified in two separate variants and with a different amino acid change, highlighting that the position is able to accommodate residues with different sizes and charges. Additionally the other known key mutation sites T509A and N566S were also identified in different variants. These mutants were only sequenced and were not fully characterised to analyse the kinetics.

Neutral drift has been indicated as being important in natural evolution and was thought to be useful for directed evolution experiments. However in this system, the most active mutants found for β-galactosidase activity contained the same mutations to those found in conventional directed evolution experiments. The libraries were also screened against a range of other related substrates and three unique variants were identified with activity against β-xylosidases, which contained mutations previously unseen in directed evolution experiments, but also two key mutations previously identified in direct selection experiments, and a third mutation that had previously been identified but considered to be unimportant.

Comparisons with other studies incorporating neutral drift revealed a similar picture where the mutants found correlate with those previously found by direct selection. Neutral drift may confer no extra benefits over standard directed evolution experiments except when a range of new activities are being investigated.

Further work

If this work was being continued, there are a lot of improvements that could be made.

1. Four β-galactosidase active, and three β-xylosidase active variants were not fully characterised. These variants should be screened in a quantitative assay to measure their kinetics against the native β-glucuronidase and new activity.
2. The fully characterised β-galactosidase active variant G245A/K370R/W529L was only screened quantitatively on one occasion. This needs to be screened on a further two or three occasions to allow errors to be calculated.
3. Mutants with increased activity towards β-xylopyranosides were found on screening the neutral library, but there have been no previous direct selection studies for a β-glucuronidase to a β-xylosidase using random mutagenesis and selection. The previous experiment by Geddie[2] used site saturation mutagenesis of selected residues identified in the Matsumura and Ellington experiment[7]. It would be interesting to do the direct

selection for this to see whether the mutants generated are the same as those found in the neutral drift and saturation mutagenesis studies. This could provide further evidence to support or disprove that neutral drift finds the same mutations as traditional directed evolution.

4. The library sizes used in these experiments have also been fairly small. It would be interesting to try the experiment on a much larger library, perhaps by redesigning the experiment to incorporate screening for β-glucuronidase activity by FACS. The current method has been limited in the library sizes produced by the results of transformations into XL1 Blue. Though comparison with Gupta and Tawfik's[13] experiment on PON1 using neutral drift and FACS suggests that its use would make very little difference to the mutants produced, in this system a much larger library size may allow investigation of new parts of the fitness landscape or escape local minima to find new beneficial mutations.

2.4 Experimental

2.4.1 Materials

$MgCl_2$, ligase buffer, ligase, dATP, dCTP, dGTP and dTTP were purchased from Promega. dNTPs were also purchased from Bioline. *Nco*I, *EcoR*I, DNaseI, DNaseI buffer, Vent, Thermopol buffer, NEB4 buffer and BSA were purchased from New England Biolabs. Primers were purchased from Sigma-Genosys and Invitrogen. Agarose was purchased from Melford Laboratories and low melting point agarose from Fisher Scientific. GenemorphII® Random Mutagenesis Kit and XL1 Blue supercompetent *E. coli* were purchased from Stratagene. BW25141 strain was from cell stocks prepared according to Datsenko and Wanner[22]. X-glu was purchased from Acros Organics. X-gal was purchased from Qbiogene. PNP-glu and PNP-gal were purchased from Sigma. Wizard*Plus*® SV DNA Minipreps Kit and Wizard® PCR Preps DNA Purification System were purchased from Promega. GoTaq multimix buffer without magnesium and GoTaq was firstly a free sample and then further purchased from Promega. Plasmid Megapreps were purchased from Qiagen. 5-bromo-4-chloro-3-indolyl-β-D-fucopyranoside (X-fuco), 5-bromo-4-chloro-3-indolyl-β-D-xylanopyranoside (X-xyl), 5-bromo-4-chloro-3-indolyl-N-acetyl-β-D-glucosaminide (X-gluam), 5-bromo-4-chloro-3-indolyl-N-acetyl-β-D-galactosaminide (X-galam) and 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (X-gluco) were purchased from Fluka.

2.4.2 Methods

1. Mutagenesis

Mutagenic PCR of the β-glucuronidase gene was performed using the GenemorphII™ Random Mutagenesis Kit on 50 µL reactions prepared from plasmid p042 (1 µL), 10 x Mutazyme buffer, forward and reverse primers OSOT165 and OSOT166 (final concentration each 4 µM), dNTPs (final concentration 0.8 mM each dNTP 0.2 mM) and Mutazyme (1 µL, final concentration 0.05 U/µL). Thermocycling was carried out using a program containing one cycle of of 94°C for 5 minutes; 1 cycle of 60°C for 5 minutes; 30 cycles of 94°C for 1 minute, 50°C for 1 minute (Ramping = 3°C/s, Gradient = 10°C), 72°C for 2 minutes; and a final cycle of 72°C for 10 minutes. Products were separated and purified on a 1.5% low melting point agarose gel run in 1x TAE buffer (40 mM Tris, 20 mM Acetic acid and 1mM EDTA). The DNA was visualised by UV using a solution of ethidium bromide stain. The correct gene product was excised from the gel using a scalpel, and purified using the Wizard® PCR Preps DNA Purification System.

2. Digestion

Sticky ended restriction endonuclease digestions were carried out simultaneously at the *EcoR*I and *NCo*I sites of the purified gene PCR product. PCR product (45 µL), BSA (6 µL, 1 mg/mL), NEB buffer 4 (6 µL, 50 mM potassium acetate, 20 mM Tris-Acetate, 10 mM magnesium acetate, 1 mM dithiothreitol), *EcoR*I (1 µL of 1:1 dilution, 10000 U/ mL) and *Nco*I (1 µL of 1:1 dilution,

10000 U/mL) were placed in a micro-centrifuge tube and heated in a water bath at 37°C for 3 hours. Products were separated and purified on a 1.5% low melting point agarose gel run in TAE buffer (Tris 40 mM, Acetic acid 20 mM, EDTA 1mM).The DNA was visualised by UV using a solution of ethidium bromide stain. The correct digestion product was excised from the gel using a scalpel and purified using the Wizard® PCR Preps DNA Purification System.

Digestion of p042, a PBAD/His plasmid containing the β-glucuronidase gene, was also carried out to produce a linear plasmid fragment for cloning. P042 (0.5 µL), H_2O (14.5 µL), BSA (2 µL), NEB buffer 4 (2 µL, 50 mM potassium acetate, 20 mM Tris-Acetate, 10 mM magnesium acetate, 1 mM dithiothreitol), *EcoR*I (1 µL of 1:1 dilution, 10000 U/ mL) and *Nco*I (1 µL of 1:1 dilution, 10000 U/mL) were placed in a micro-centrifuge tube and heated in a water bath at 37°C for 3 hours. Products were separated and purified on a 1.5% low melting point agarose gel run in TAE buffer (Tris 40 mM, Acetic acid 20 mM, EDTA 1mM).The DNA was visualised by UV using a solution of ethidium bromide stain. The correct digestion product was excised from the gel using a scalpel and purified using the Wizard® PCR Preps DNA Purification System.

3. Ligation

Digested β-glucuronidase (4 µL), linearised pBAD/His (1 µL), T4 DNA Ligase buffer (1 µL, 300 mM Tris-HCl, 100 mM magnesium chloride, 100 mM DTT, 10 mM ATP), H_2O (3 µL) and T4 DNA Ligase (1 µL, 1-3 U/µL) were placed in 200 µL thin walled PCR tubes. The tubes were incubated for two hours at room temperature to allow the β-glucuronidase gene to be ligated into pBAD/His. The resulting plasmid mixture was used directly without purification.

4. Transformation into supercompetent XL1 Blue *E. coli*

Ligation mixture from the previous procedure (5 µL) and XL1 Blue supercompetent *E. coli* (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI*^{q} *ZΔM15* Tn*10* (Tet^r)], (40 µL)) were chilled on ice for 30 minutes in 15 mL thin walled polypropylene tubes. Transformation tubes were placed in a water bath at 42°C for 45 seconds then immediately being placed back onto ice. Pre-warmed (37°C) SOC medium (250 µL) was added to each transformation and the tubes incubated at 37°C with shaking for one hour. Transformants (250 µL) were spread on LB ampicillin (0.1 mg/mL) plates and incubated at 37°C overnight. All resulting colonies were extracted from the plates into LB medium and incubated with shaking at 37°C for one hour. The culture was spun in a centrifuge to produce a cell pellet. Plasmid DNA was extracted and purified from the cell pellet using the Wizard®*Plus* SV Minipreps DNA Purification System according to the manufacturers' protocol.

5. Transformation via electroporation into electrocompetent BW25141 strain *E. coli*

LB ampicillin (0.1 mg/mL) arabinose (2 mg/mL) plates were pre-prepared with a solution of X-glu (80 µL, 0.048M (20 mg/mL) in DMSO) or X-gal (400 µL, 0.049M (20 mg/mL) in DMSO) and warmed to 37°C.

BW25141 cells (*(ΔaraD-araB)567, ΔlacZ4787(:: rrn B-3), Δ(phoB-phoR)580, λ-, galU95, ΔuidA3 :: pir+, recA1, endA9 (del-ins) :: FRT, rph-1, Δ(rhaD-rhaB)568, hsdR514)* (40 µL)), and plasmid (4 µL) were cooled inside pre-chilled electroporator cuvettes. Each cuvette was submitted to electroporation (200 Ω, 25 µFD, 1.75 kV) followed quickly by the addition of pre-warmed (37°C) SOC medium (950 µL). The transformants were transferred to microcentrifuge tubes and incubated at 37°C with shaking for 1 hour. Transformant (100 µL) was added to the plates and the plates incubated overnight at 37°C.

6. Colony selection

From plates incubated in the presence of X-glu, blue colonies, indicating the presence of an active β-glucuronidase, were counted and picked from the plates. For rounds 1-3 these colonies were pooled and suspended in LB medium (10 mL), spun in a centrifuge to produce a cell pellet and plasmid DNA isolated using the Wizard®*Plus* SV Minipreps DNA purification system according to the manufacturers protocol.

From plates incubated in the presence of X-gal, blue colonies, indicating the presence of β-galactosidase activity, were individually selected and grown to prepare cell and plasmid stocks. Cell stocks were prepared by taking the individual colony and growing up overnight in LB medium (10 mL). Cell culture (1mL) had sterile DMSO (55 µL) added and the stock frozen at -80°C. Plasmid was extracted from the remaining culture using the Wizard*Plus* SV Minipreps system and used to prepare plasmid stocks.

7. Screening with alternative substrates

LB ampicillin (0.1 mg/mL) arabinose (2 mg/mL) plates were pre-prepared with solutions of X-gluco (400 µL, 0.01M (20 mg/mL) in DMSO), X-fuco (400 µL, 0.051M (20 mg/mL) in DMSO), X-xyl (400 µL, 0.053M (20 mg/mL) in DMSO), X-gluam (400 µL, 0.045M (20 mg/mL) in DMSO) and X-galam (400 µL, 0.045M (20 mg/mL) in DMSO) and warmed to 37°C.

BW25141 cells (40 µL), and plasmid (4 µL) were cooled inside pre-chilled electroporator cuvettes. Each cuvette was submitted to electroporation (200 Ω, 25 µFD, 1.75 kV) followed quickly by the addition of pre-warmed (37°C) SOC medium (950 µL). The transformants were transferred to microcentrifuge tubes and incubated at 37°C with shaking for 1 hour. Transformant (100 µL) was added to the plates and the plates incubated for up to 24 hours at 37°C.

Plates containing blue colonies, indicating the presence of a mutant β-glucuronidase with activity against the particular substrate, were counted and were individually selected and grown to prepare cell and plasmid stocks. Cell stocks were prepared by taking the individual colony and growing up overnight in LB medium (10 mL). Cell culture (1mL) had sterile DMSO (55 µL)

added and the stock frozen at -80°C. Plasmid was extracted from the remaining culture using the Wizard*Plus* SV Minipreps system and used to prepare plasmid stocks.

.

8. Sequencing

Plasmid solutions were prepared to contain ~10 µg DNA in water (90 µL) in micro-centrifuge tubes. Ice cold absolute ethanol (200 µL) and sodium acetate (10 µL, 3M, pH 5.2) were added to each sample, and the samples stood on ice for 30-60 minutes.

The samples were centrifuged at 13000 rpm/16000 g for 20 minutes, then the ethanol carefully removed by pipette. 70% ethanol (1µL) was added to each tube and the samples centrifuged at 13000 rpm/16000 g for 5 minutes. The ethanol was removed by pipette and the plasmid samples allowed to air dry for 3-5 hours.

Primer solutions were prepared to a concentration of 10 µM from stock solutions of 100 µM, allowing 10 µL for each reaction required.

The labelled samples were sequenced by MWG Biotech.

9. Preparation of protein samples

A scraping of cell stock was grown overnight in LB medium (10 mL) with ampicillin (10 µL, 100 mg/mL). This was added to LB medium (1 L) containing ampicillin (1 mL, 100 mg/mL) and grown until the OD_{600} was ~0.6. Arabinose (2g) was added to induce protein expression and the culture grown for a further three hours.

The cells were spun down by centrifuge (5500 rpm, 25 minutes at 4°C). Cell pellets were suspended in buffer 1 (50 mM Tris HCl pH 8, 10% glycerol, 4mL/g wet weight of cells) and the cells lysed by sonication (~30 cycles of 15 seconds on, 15 seconds off). The lysate was centrifuged at 11 000 rpm for 30 minutes and the supernatant retained.

The supernatant was purified by FPLC, washing the column with buffer 2 (50 mM Tris-HCl pH 7.4, 0.5M NaCl, 60 mM imidazole) and eluting the protein in buffer 3 (50 mM Tris-HCl pH 7.4, 0.5M NaCl and 0.5 M imidazole). The protein was exchanged into buffer 4 (50 mM Tris-HCl pH 7.4) using a PD10 column and dispensed into 500 µL aliquots for storage at -80°C. Protein concentration was quantified by Bradford reagent and purity by SDS-PAGE using a 10% gel (H2O, 30% acrylamide mix, Tris (1.5 M, pH 8.8), 10% sodium dodecyl sulphate, 10% ammonium persulphate and TEMED). Gel was run in a standard Tris, glycine and SDS running buffer. The gel was stained using Coomassie Blue, heated for one minute then de-stained using a methanol; glycine and acetic acid de-stain solution. The gel was visualised under white light.

10. Protein activity assays

Solutions of PNP-gal were made at concentrations of 200 µM, 1 mM, 2 mM, 10 mM, 20 mM and 30 mM in 50 mM Tris-HCl pH 7.4, which is double the final concentration used in the assay. Solutions of each β-glucuronidase mutant were made in Tris-HCl pH 7.4 to 200 nM, 400 nM, 1 µM, 2 µM, 4 µM, 10 µM and 20 µM, which is double the final concentration used in the assay. PNP-gal of each concentration (50 µL) was added to a 96 well plate, and the reaction initiated with the addition of protein (50 µL). The plate is depicted with final concentrations in figure 2.34. Each well in a row has 50 µL of the listed concentration protein added. Each well in a column has 50 µL of the listed concentration PNP-gal added.

The absorbance of the product in each well was measured at 405 nm over 240 measurement cycles, lasting approximately 5000 seconds.

PNP-gal Protein↓	100 µM	500 µM	1 mM	5 mM	10 mM	15 mM						
0												
100 nM												
200 nM												
500 nM												
1 µM												
2 µM												
5 µM												
10 µM												

Figure 2.34: depiction of a 96 well plate showing how the protein and PNP-gal solutions have been dispensed into the wells. The cells containing the concentration values are just titles and not part of the well plate. The concentration values represent the final concentration in the assay, not the concentrations added to the wells.

For the measurement of residual β-glucuronidase activity, solutions of PNP-glu were prepared in Tris-HCl pH 7.4 to 0.02 mM, 0.1 mM, 0.2 mM, 1 mM, 5 mM and 10 mM. This is double the working concentration used in the assay. Protein solutions were prepared in Tris-HCl pH 7.4 to 4 nM, 10 nM, 20 nM, and 40 nM. This is double the working concentration in the assay. PNP-glu (50 µL) of each concentration was added to a 96 well plate, and the reaction initiated with the addition of protein (50 µL) as depicted in figure 2.35. The reaction was measured over a time course of approximately 5000 seconds.

PNP-glu Protein↓	10 µM	50 µM	100 µM	500 µM	1 mM	2.5 mM	5 mM					
0												
2 nM												
5 nM												
10 nM												
10 nM												
20 nM												

Figure 2.35: depiction of a 96 well plate showing how the protein and PNP-gal solutions have been dispensed into the wells. The cells containing the concentration values are just titles and not part of the well plate. The concentration values represent the final concentration in the assay, not the concentrations added to the wells.

For both assays, data from certain wells was omitted from the analysis if the absorbance exceeded the maximum readable by the machine, or if the absorbance was too low to register.

11. Preparation of electrocompetent BW25141 *E. coli*

A scraping of cell stock was spread over an LB plate (without antibiotics) and incubated at 37°C overnight. From this, a single colony was selected, added to LB media (10 mL), without antibiotics and incubated at 37°C overnight.

1 mL of overnight culture was added to LB medium (100 mL) and incubated at 37°C with shaking until the optical density at 600 nm (OD_{600}) was between 0.5 and 0.6. The culture was split between ice cold 15 mL centrifuge tubes and stored in the cold room at 4°C on ice for 15 minutes.

The culture samples were then centrifuged (7000 rpm, 4°C, 15 minutes) before being transferred to the cold room. The supernatant was removed and each pellet washed and re-suspended in sterile H_2O (10 mL). The suspension was separated by centrifugation (7000 rpm, 4°C, 15 minutes). The washing process was repeated once more to leave clean cell pellets. Each cell pellet was re-suspended in 10% glycerol (100 µL) before being transferred for storage at -80°C.

2.4.3 Oligonucleotide Sequences

OSOT165:
5’-ACGTCGCCCATGGGACACCATCACCATCACCATATGTTACGTCCTGTAGAAAC-3’

OSOT166: 5’-CGCGAATTCTGCAGTCATTGTTTGCCTCCCTGCT-3’

OCN112 (betagluseq-3): 5’-CGATTAAAGAGCTGATAGCG-3’

OCN113 (betagluseq-2): 5’-GCTTTGGTCGTCATGAAGATGC-3’

OCN114 (betagluseq-1): 5**’-**CCGTTTGTGTGAACAACG-3’

OMK164: 5’-ATGCCATAGCATTTTTATCC-3’

2.5 References

1. Wallace, B. D., Wang, H., Lane, K. T., Scott, J. E., Orans, J., Koo, J. S., Venkatesh, M., Jobin, C., Yeh, L-A., Mani, S., and Redinbo, M. R.; *Alleviating Cancer Drug Toxicity by Inhibiting a Bacterial Enzyme*; Science; 2010; **330**; 831-835.
2. Geddie, M. L. and Matsumura, I.; *Rapid Evolution of β-Glucuronidase Specificity by Saturation Mutagenesis of an Active Site Loop*; Journal of Biological Chemistry; 2004; **279**; 26462-26468.
3. Rowe, L. A., Geddie, M. L., Alexander, O. B., and Matsumura, I.; *A Comparison of Directed Evolution Approaches Using the β-Glucuronidase Model System*; Journal of Molecular Biology; 2003; **332**; 851-860.
4. Altamirano, M. M., Blackburn, J. M., Aguayo, C., and Fersht, *A. R.; Directed evolution of new catalytic activity using the α/β-barrel scaffold*; Nature; 2000; **403**; 617-622.
5. Flores, H. and Ellington, A. D.; *Increasing the Thermal Stability of an Oligomeric Protein, Beta-glucuronidase*; Journal of Molecular Biology; 2002; **315**; 325-337.
6. Matsumura, I., Wallingford, J. B., Surana, N. K., Vize, P. D., and Ellington, A. D.; *Directed evolution of the surface chemistry of the reporter enzyme β-glucuronidase*; Nature Biotechnology; 1999; **17**; 696-701.
7. Matsumura, I. and Ellington, A. D.; *In vitro Evolution of Beta-glucuronidase into a Beta-galactosidase Proceeds through Non-specific Intermediates*; Journal of Molecular Biology; 2001; **305**; 331-339.
8. RCSB Protein data bank http://www.rcsb.org/pdb/explore/explore.do?structureId=1BHG. Accessed: 13.11.2010. (Archived by WebCite® at http://www.webcitation.org/5uD4rLaui)
9. Islam, M. R., Tomatsu, S., Shah, G. N., Grubb, J. H., Jain, S., and Sly, W. S.; *Active Site Residues of Human β-glucuronidase*; Journal of Biological Chemistry; 2001; **305**; 331-339.
10. Xiong, A-S., Peng, R-H., Cheng, Z-M., Li, Y., Liu, J-G., Zhuang, J., Gao, F., Xu, F., Qiao, Y-S., Zhang, Z., Chen, J-M., and Yao, Q-H.; *Concurrent mutations in six amino acids in β-glucuronidase improve its thermostability*; Protein Engineering, Design and Selection; 2007; **20**; 1-7.
11. Bershtein, S., Goldin, K., and Tawfik, D. S.; *Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins*; Journal of Molecular Biology; 2008; **379**; 1029-1044.
12. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D. S.; *Robustness-epistatsis link shapes the fitness landscape of a randomly drifting protein*; Nature; 2006; **44**; 929-932.
13. Gupta, R. D. and Tawfik, D. S.; *Directed enzyme evolution via small and effective neutral drift libraries*. Nature Methods; 2008; **5**; 939-942.
14. Amitai, G., Gupta, R. D., and Tawfik, D. S.; *Latent evolutionary potentials under the neutral mutational drift of an enzyme*; HFSP Journal; 2007; **1**; 67-78.

15. Bloom, J. D., Romero, P. A., Lu, Z., and Arnold, F.H.; *Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution*; Biology Direct; 2007; **2**; 17.
16. Bloom, J. D., Lu, Z., Chen, D., Raval, A., Venturelli, O. S., and Arnold, F. H.; *Evolution favours protein mutational robustness in sufficiently large populations*; BMC Biology; 2007; **5**; 29.
17. Smith, W. S., Hale, J. R., and Neylon, C.; *Applying neutral drift to the directed molecular evolution of a β-glucuronidase into a β-galactosidase: Two different evolutionary pathways lead to the same variant;* BMC Research Notes; 2011; **4**; 138-148.
18. Parikh, M. R. and Matsumura, I.; *Site-saturation Mutagenesis is more Efficient than DNA Shuffling for the Directed Evolution of β-Fucosidase from β-Galactosidase*; Journal of Molecular Biology; 2005; **352**; 621-628.
19. Aharoni, A., Gaidukov, L., Yagur, S., Toker, L., Silman, I., and Tawfik, D. S.; *Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization*; Proceedings of the National Academy of Sciences; 2004; **101**; 482-487
20. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S., and Tawfik, D. S.; *High-Throughput Screening of Enzyme Libraries: Thionolactonases Evolved by Fluorescence Activated Sorting of Single Cells in Emulsion Compartments*; Chemistry and Biology; 2005; **12**; 1281-1289.
21. Hecky, J. and Muller, K. M.; *Structural Perturbation and compensation at physiological temperature leads to thermo-stabilization of beta-lactamase*; Biochemistry; 2005; **44**; 12640-12654.
22. Datsenko, K. A. and Wanner, B. L.; *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products**;*** Proceedings of the National Academy of Sciences; 2000, **97**: 6640-5

Chapter 3 - Preliminary investigations into the inhibitor properties of galactose and 2,2,3,3-tetrafluorogalactose on β-galactosidase

3.1 Introduction

The enzyme β-galactosidase is a glycosyl hydrolase like β-glucuronidase featured in the previous chapter. The two enzymes have 25% sequence identity. Like β-glucuronidase, β-galactosidase is a homotetrameric protein which is only active as a tetramer and has an active site found within an α/β- barrel. Unlike β-glucuronidase, the barrel is missing the fifth helix and additionally the sixth beta sheet is distorted. Each monomer is composed of five domains with the active site being predominantly found in domain 3, though the active site is also composed of domains from other subunits. There are four independent binding sites within the enzyme[1,2].
The substrate for β-galactosidase is the disaccharide lactose, which is hydrolysed by the enzyme to leave one molecule of glucose and one molecule of galactose. β-galactosidase is also able to convert lactose to allolactose, which is then ultimately converted to glucose and galactose. The enzyme has both deep and shallow binding modes, with the galactosyl moiety of the sugar binding strongly and deeply within the enzyme. The stabilising bonds are formed mostly with the galactosyl moiety and the glucosyl moiety forms weaker bonds[1,2]. Like β-glucuronidase, β-galactosidase has catalytic glutamic acid residues and shares the same mechanism as discussed in chapter 2.

In this project the aim was to test the hypothesis that galactose and the novel compound 2,2,3,3-tetrafluorogalactose would be inhibitors of β-galactosidase. There is precedent for using fluorinated carbohydrates to test the interactions in the binding sites of enzymes. Systematic replacement of hydroxyl groups with either fluorine or hydrogen can be used to investigate the importance of individual hydroxyl groups on binding[3]. The rationale for why the enzyme would need to be inhibited was not noted in the laboratory research record, although the act of enzyme inhibition by fluorinated compounds would demonstrate important binding information alone, without another reason needed for inhibition. The assays were being carried out on behalf of the research group of Dr B. Linclau at the University of Southampton, who had synthesised the fluorinated compound. In this chapter, the results are not focussed on why inhibition was being sought, but how the assays were developed to reach a point where it could be determined whether inhibition was taking place or not.

3.2 Results and discussion

3.2.1 The initial assays

To test the hypothesis that galactose and the novel compound 2,2,3,3-tetrafluorogalactose ("fluorogalactose") are inhibitors of β-galactosidase, an assay was designed to demonstrate the kinetics of the enzyme with, or without, the presence of inhibitor. The synthetic substrate mimic 4-nitrophenyl-galactopyranoside (PNP-gal) produces a yellow product, 4-nitrophenol, upon hydrolysis by β-galactosidase. The absorbance (at 405 nm) of this compound is measured over the course of the assay. The absorbance readings can be converted to a series of product concentrations, which are used to calculate the kinetic parameters of the enzyme.

An exploratory assay was prepared which was used to gauge the correct substrate and enzyme concentrations to use for later assays. In this assay, 50 μM PNP-gal was hydrolysed by increasing concentrations of β-galactosidase, measured in Weiss Units per mL. The absorbance of the 4-nitrophenol product at 405 nm was measured over a time course of approximately 1800 seconds. The absorbance readings were converted into concentrations (C) using the Beer-Lambert law, $A=\varepsilon Cl$. The path length (l) was calculated by using the formula for the volume of a cylinder: $volume=\pi r^2 l$, using the overall volume in each well and the well diameter, where r is the radius and l is the cylinder height. From this l was calculated as 0.652 cm. The molar extinction coefficient (ε) was the listed literature value for 4-nitrophenol of 17400 mol $dm^{-3}.cm^{-1}$. The calculated concentrations were converted to millimolar concentrations for more effective display, and plotted against time to produce a graph demonstrating the change in concentration over time (figure 3.1).

The results of this assay showed that, as expected, there was a positive correlation between enzyme concentration and the amount of product generated. The assay was carried out in duplicate, but there were some inconsistencies between the concentrations of 4-nitrophenol produced in the two duplicates which can be attributed to the lag time of adding the solutions. The two most concentrated enzyme solutions (0.25 and 0.5 U/mL) produced the clearest data for showing a change in concentration, but the formation of product at the highest enzyme concentration (0.5 U/mL) was not linear. This was probably due to depletion of the PNP-gal during the assay. At low substrate concentrations, the concentration of the substrate can become the rate limiting factor.

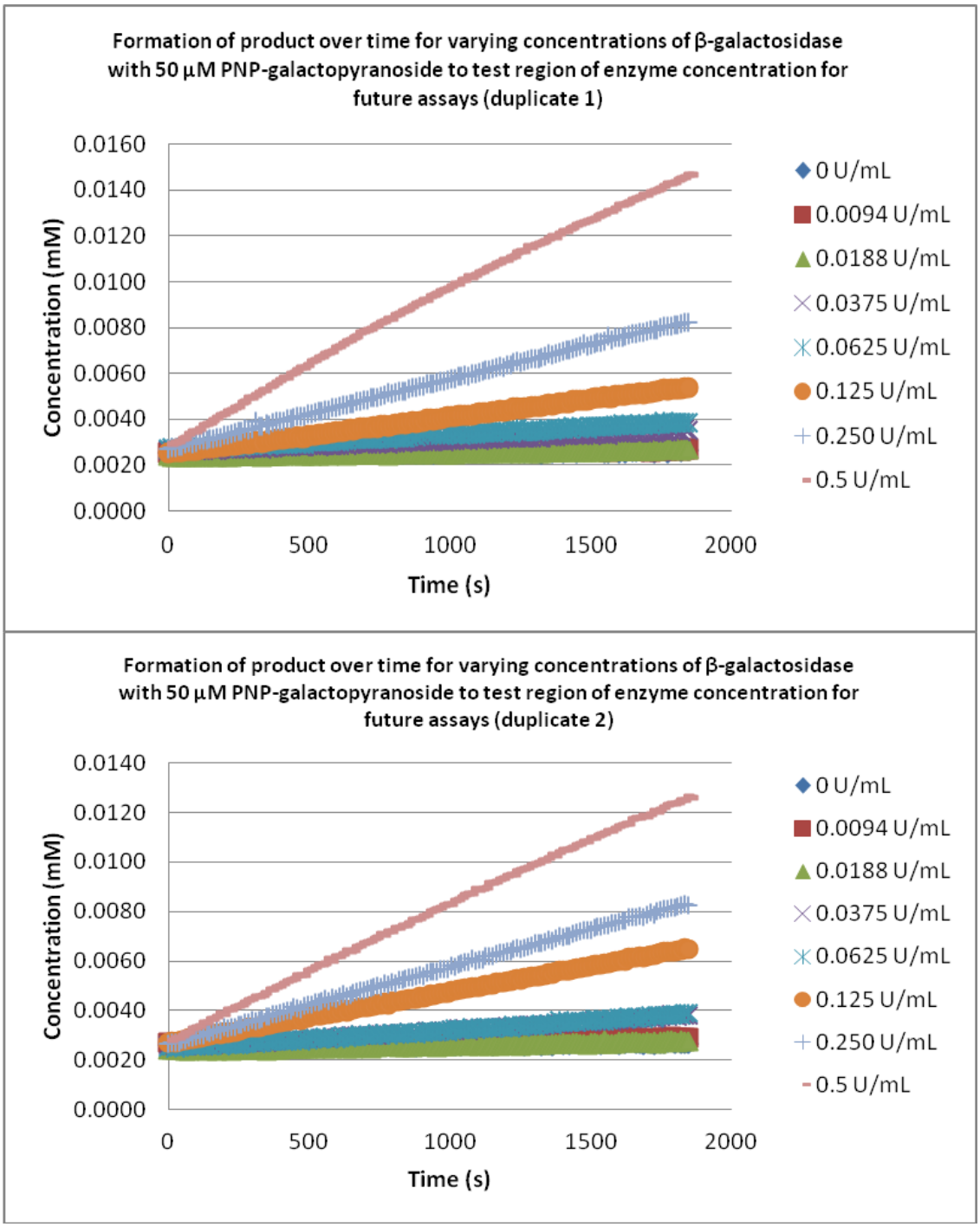


Figure 3.1: formation of 4-nitrophenol from the hydrolysis of 50 µM PNP-gal, catalysed by increasing concentrations of β-galactosidase. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

The preliminary results showed that the assay should be trialled with higher substrate concentrations to see if it eliminated the non-linear nature of the data at the highest enzyme concentrations. The assay was adapted to use both 50 and 100 µM PNP-gal substrate, and was carried out for an extended time course of approximately 3300 seconds. The time course of the experiment was extended in order to help to highlight the point at which the data becomes non-

linear. The rate of reaction will decrease over time as the substrate is depleted, so it is important to ensure that only the initial rates are being used for Michaelis-Menten calculations/kinetics.

The results of this modified assay were consistent with the previous assay as the amount of 4-nitrophenol produced increased as the enzyme concentration increased (figures 3.2 and 3.3). The highest concentrations of enzyme produced the data with the clearest demonstration of change in concentration for the hydrolysis of both 50 and 100 μM PNP-gal. However, the results were not as clean as those in the previous assay. There is also evidence of the hydrolysis not occurring in some wells. This is possibly due to inadequate mixing of the samples in the wells and enzyme not being accurately dispensed due to the small volume (2 μL) compared to the overall assay volume.

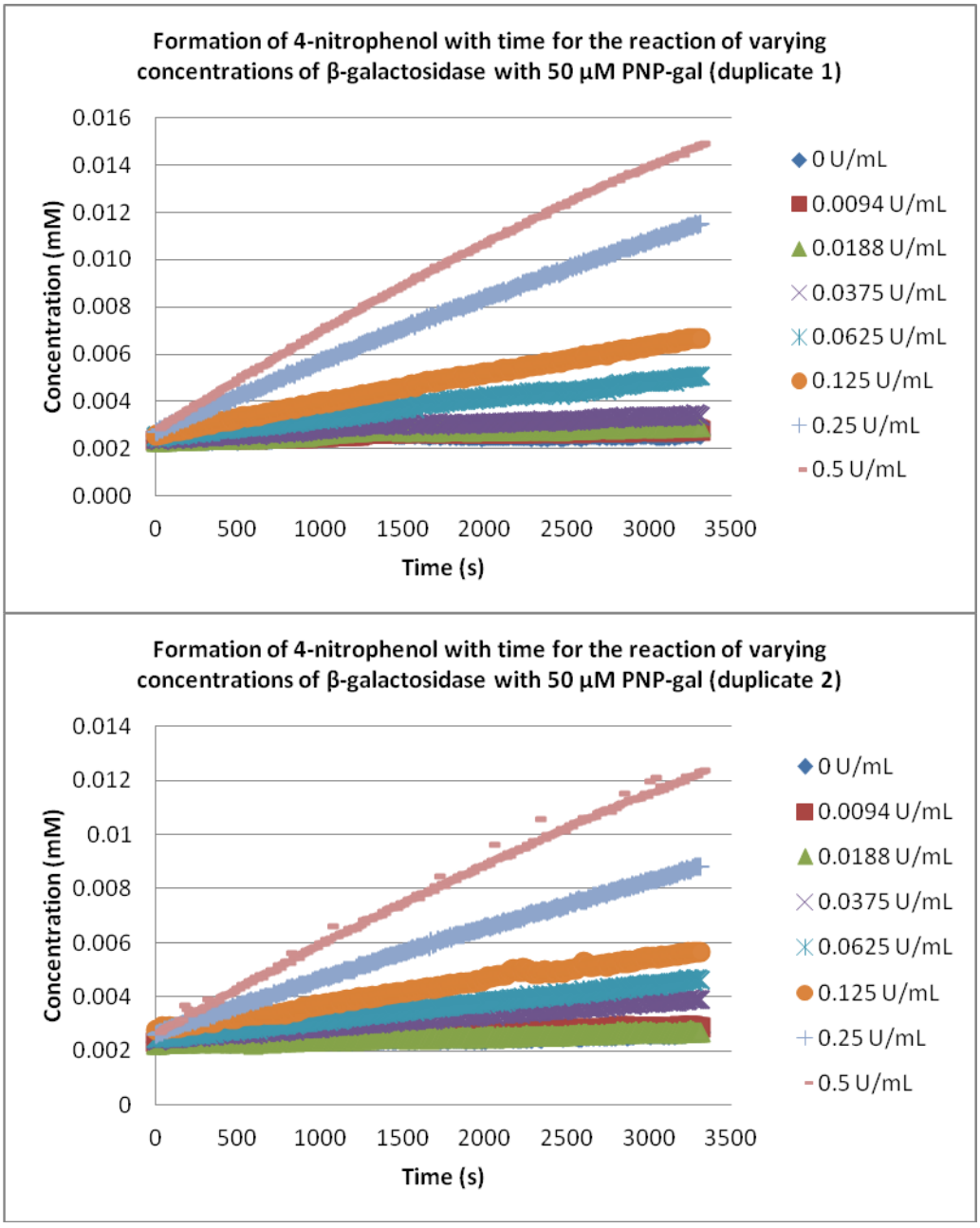


Figure 3.2: formation of 4-nitrophenol from the hydrolysis of 50 µM PNP-gal, catalysed by increasing concentrations of β-galactosidase. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

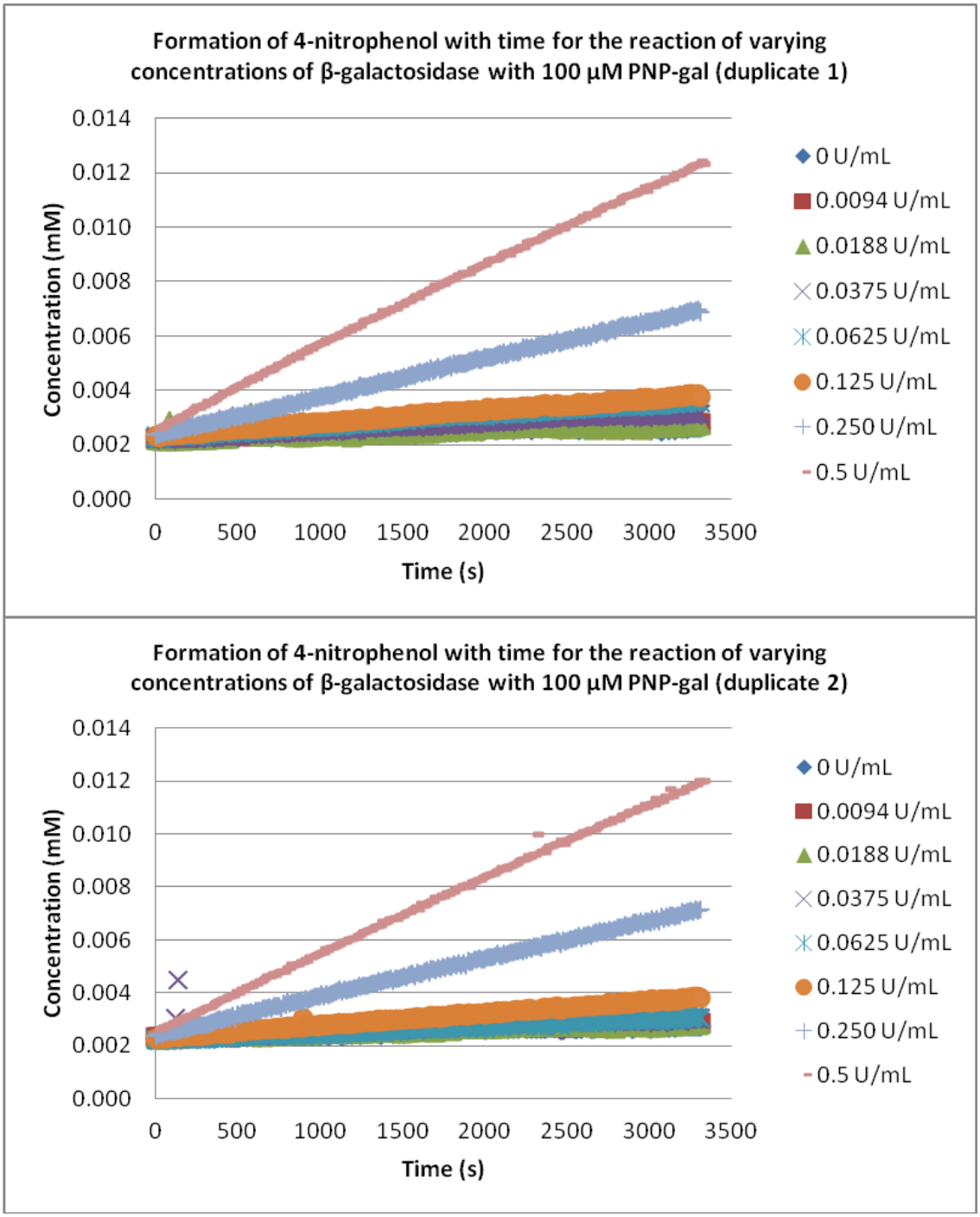


Figure 3.3: formation of 4-nitrophenol from the hydrolysis of 100 µM PNP-gal, catalysed by increasing concentrations of β-galactosidase. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

Unfortunately the data from the exploratory assays was misinterpreted and as a result, the lower substrate concentration 50 µM PNP-gal was selected when 100 µM PNP-gal would have been more suitable. With hindsight, it would have been better to have tried even higher concentrations of substrate, which would have ensured that the substrate would not have been at risk of being rapidly depleted and prevented the substrate from becoming the rate limiting factor.

Using the information obtained from the exploratory assays, the assay was adapted to test whether β-galactosidase is inhibited by galactose. 50 μM PNP-gal was hydrolysed by either 0.25 or 0.5 U/mL β-galactosidase in the presence of increasing concentrations of galactose. The absorption of the 4-nitrophenol product was measured over a 3300 second timecourse. The absorbance was converted to a concentration using the Beer-Lambert equation $A=\varepsilon Cl$ and plotted on graphs to demonstrate the change in concentration with time.

The assay was carried out twice. The first attempt was not successful due to a number of failed reactions and inconsistent data sets (supplementary data). The second attempt produced a better set of data, though not without inconsistencies (figures 3.4 and 3.5). All hydrolyses in the presence of 20 mM galactose, and hydrolyses using the 0.5 U/mL β-galactosidase in the presence of 10 mM galactose have a noticeable decrease in the overall concentration of 4-nitrophenol produced, suggesting that the galactose may be having an inhibitory effect at that concentration.

These results are intriguing in that the 10 mM galactose only appears to have an inhibitory effect against the higher enzyme concentration. It would be normal for the inhibitor to exhibit an equivalent inhibitory effect regardless of the concentration of enzyme. It is the ratio of the concentration of the substrate to the concentration of the inhibitor that is important in competitive inhibition. The most likely explanation for this is that the design of the assay produces more reliable results at higher enzyme concentrations. An alternative explanation could be that if inhibition is taking place, that inhibition is not competitive. It may not even be reversible.

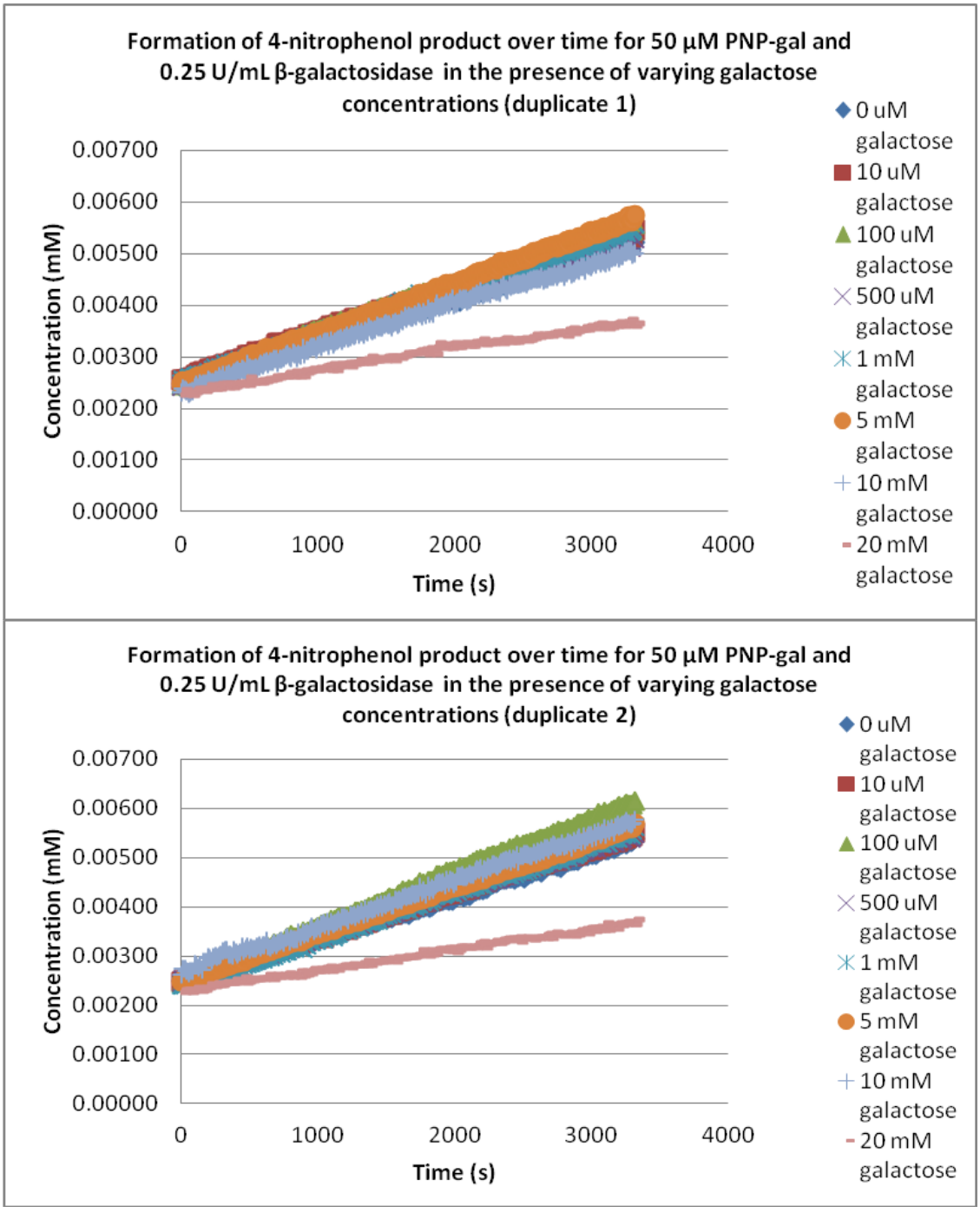


Figure 3.4: formation of 4-nitrophenol from the hydrolysis of PNP-gal catalysed by β-galactosidase in the presence of increasing concentrations of galactose. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration. By looking at the pattern of colours on the graph, the inconsistencies in the data become apparent.

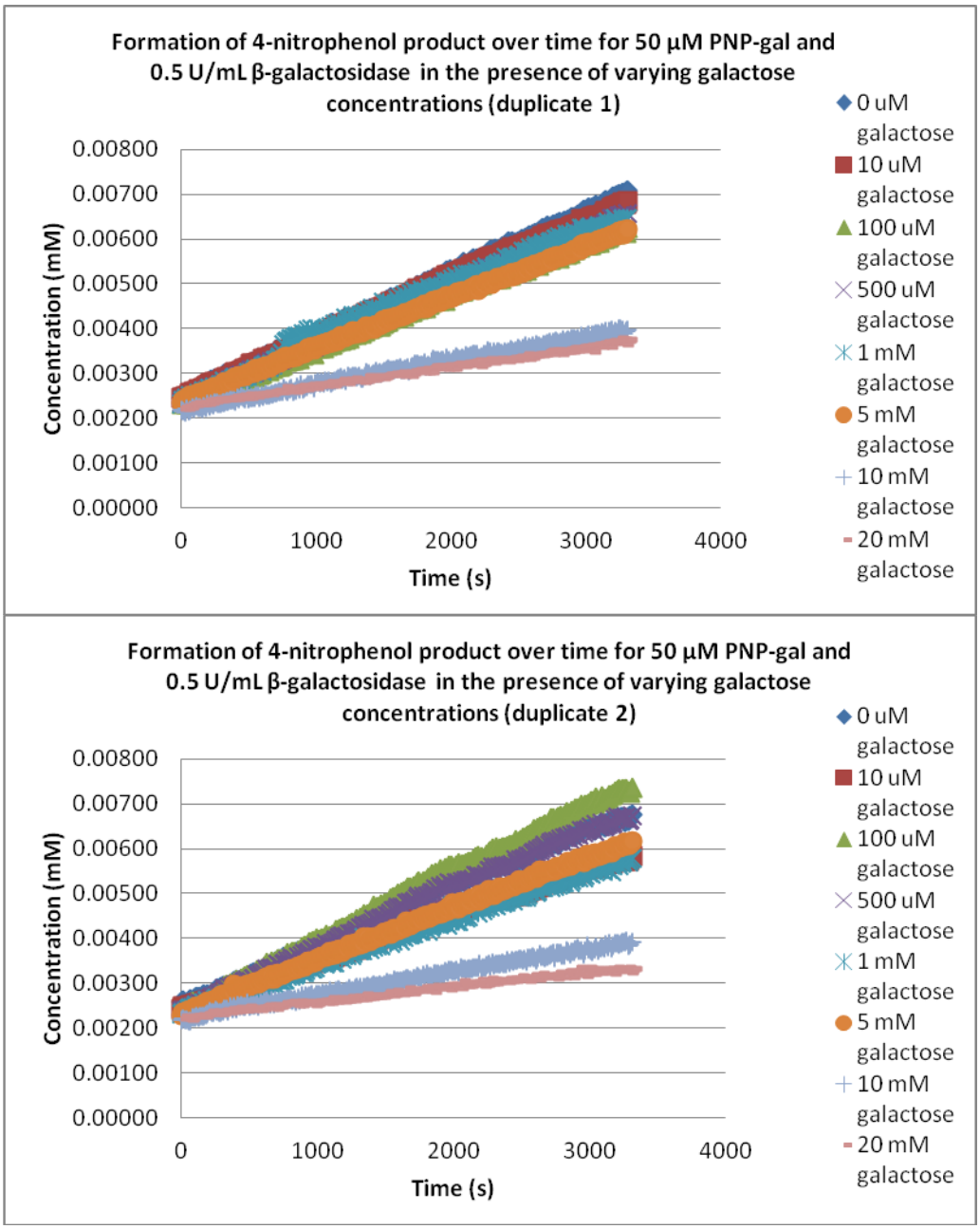


Figure 3.5: formation of 4-nitrophenol from the hydrolysis of PNP-gal catalysed by β-galactosidase in the presence of increasing concentrations of galactose. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration. The order of colours on the graph highlights the inconsistencies in the data. The messier nature of the data can also be observed as there are jerks in some data series.

The gradient for each data series was calculated to produce a rate for the reaction (measured in mmol $dm^{-3}.s^{-1}$) (table 3.1). The reaction rates were used to assess whether inhibition had occurred. There is no obvious evidence of inhibition by galactose. If inhibition had taken place the slope of the graphs i.e. the rate of the reaction would be expected to decrease relative to the

control reaction without galactose. There was not clear evidence of this, though there was definite evidence of an effect at the highest galactose concentrations.

Galactose concentration (μM)	**Rate (duplicate 1) for 0.25 U/mL β-galactosidase (mmol $dm^{-3}.s^{-1}$)**	**Rate (duplicate 2) for 0.25 U/mL β-galactosidase (mmol $dm^{-3}.s^{-1}$)**	**Rate (duplicate 1) for 0.5 U/mL β-galactosidase (mmol $dm^{-3}.s^{-1}$)**	**Rate (duplicate 2) for 0.5 U/mL β-galactosidase (mmol $dm^{-3}.s^{-1}$)**
0	8.16×10^{-7}	8.48×10^{-7}	1.38×10^{-6}	1.31×10^{-6}
10	8.52×10^{-7}	9.03×10^{-7}	1.32×10^{-6}	1.05×10^{-6}
100	9.22×10^{-7}	1.11×10^{-6}	1.18×10^{-6}	1.53×10^{-6}
500	8.33×10^{-7}	9.18×10^{-7}	1.27×10^{-6}	1.34×10^{-6}
1000	9.12×10^{-7}	9.55×10^{-7}	1.25×10^{-6}	1.05×10^{-6}
5000	9.69×10^{-7}	9.48×10^{-7}	1.14×10^{-6}	1.15×10^{-6}
10000	8.16×10^{-7}	9.28×10^{-7}	5.43×10^{-7}	5.16×10^{-7}
20000	4.18×10^{-7}	4.17×10^{-7}	4.37×10^{-7}	3.34×10^{-7}

Table 3.1: a comparison of the rates for the reactions in the presence of increasing concentrations of galactose at each enzyme concentration, calculated from the gradients of each data series for the assay.

To confirm these results, graphs were plotted of log [galactose] versus rate (slope). In the case of competitive inhibition this graph will produce a sigmoid curve directed from top left to bottom right, which can be used to obtain an IC_{50} value. The IC_{50} is the half maximal inhibitory concentration of an inhibitor, and is a measure of the effectiveness of a compound as an inhibitor. The graphs did not return clear results though there was evidence of a sigmoid curve for one duplicate of the assay using 0.5 U/mL β-galactosidase and 50 μM PNP-gal when the anomalous result for 100 μM galactose is ignored (figure 6). As there was only limited evidence of inhibition from one set of data of the four, the results were considered unreliable and further assays were carried out to test for inhibition.

The IC_{50} is not the best measure of inhibition to use, but is the measure of inhibition that can be obtained from the design of this assay. The IC_{50} is relative and may change depending on the experimental conditions. The most suitable measure of inhibition is the K_i, a constant value for the concentration of inhibitor at which 50% of binding sites are filled. In order to obtain a value for K_i, the assay should have been designed to have used multiple substrate concentrations with a fixed inhibitor concentration, whereas this assay used a fixed substrate concentration with multiple inhibitor concentrations.

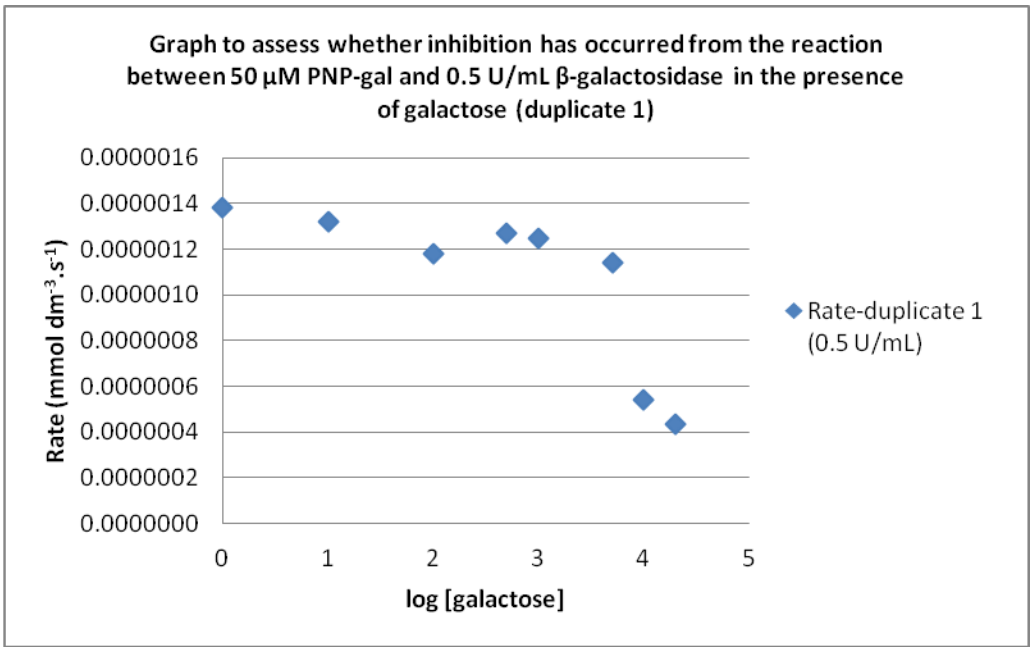


Figure 3.6: IC_{50} dose-response plot to determine whether galactose is an inhibitor of β-galactosidase for the hydrolysis of 50 µM PNP-gal by 0.5 U/mL β-galactosidase. The log_{10} of each galactose concentration was calculated and plotted against the rate of formation of product for that inhibitor concentration

Although the assay did not demonstrate that galactose is an inhibitor of β-galactosidase, the assay itself appeared to be suitable for the purpose. As a result the concentrations of inhibitor were adjusted in order to test whether fluorogalactose is an inhibitor of β-galactosidase. 50 µM PNP-gal was hydrolysed by either 0.25 or 0.5 U/mL β-galactosidase in the presence of increasing concentrations of fluorogalactose over a 3300 second timecourse. The absorbance of the 4-nitrophenol product was measured, converted to a concentration and plotted onto a graph to demonstrate the change in concentration with time.

The assay was unsuccessful (supplementary data). Hydrolysis had not proceeded in some wells, and the results were not linear in all wells where hydrolysis had taken place. It was thought that this may be an effect of the fluorogalactose rather than the PNP-gal substrate being depleted as the 4-nitrophenol concentration did not reach a plateau.

The assay was repeated under the same conditions. Generally the repeated assay was more successful as hydrolysis had taken place under most conditions. However, there were still a number of unusual results (figures 3.7 and 3.8). The data for the unusual results share some similarities in pattern to those seen in the previous failed assay. This could suggest that the fluorogalactose has some effect that is not expected under normal competitive inhibition. Hydrolyses in the wells containing the highest concentrations of fluorogalactose did not have a linear increase in product concentration with time: the gradient of the slope appears to decrease at around 2000 seconds. This could be an inhibitory effect of the fluorogalactose, but it isn't the

result that was anticipated. The anticipated result would show the overall gradient being decreased compared to the gradient of the hydrolysis without inhibitor present. It is possible that as the substrate is depleted, the fluorogalactose is able to demonstrate an inhibitory effect causing the change in gradient. In order to have tested this theory it would have been better to try the assay again with higher concentrations of substrate present. This would have eliminated any effect of substrate depletion by allowing true competitive inhibition from the outset.

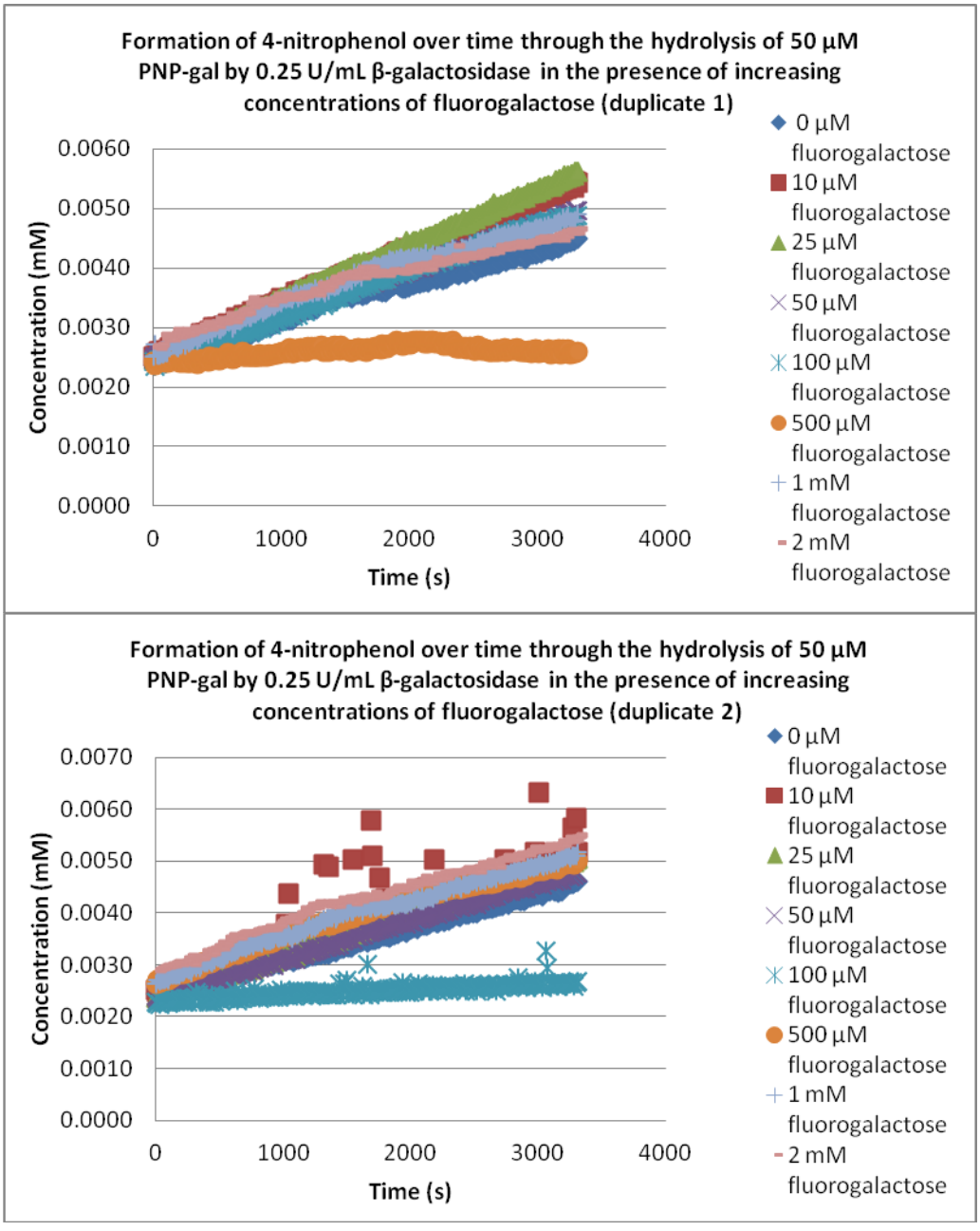


Figure 3.7: formation of 4-nitrophenol from the hydrolysis of PNP-gal catalysed by β-galactosidase in the presence of increasing concentrations of fluorogalactose. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

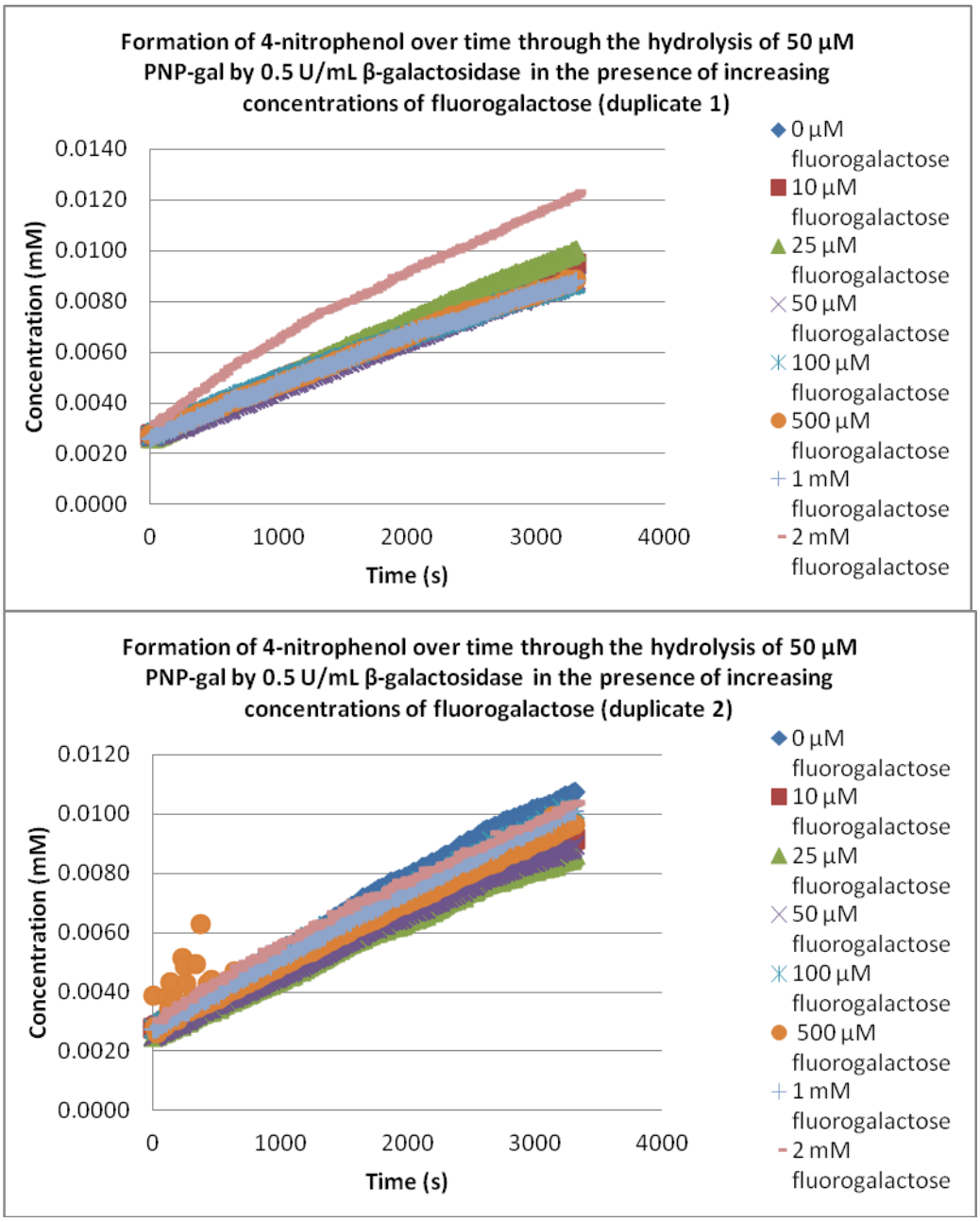


Figure 3.8: formation of 4-nitrophenol from the hydrolysis of PNP-gal catalysed by β-galactosidase in the presence of increasing concentrations of fluorogalactose. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

In isolation, the graphs demonstrating formation of 4-nitrophenol over time do not provide evidence for or against inhibition and demonstrate no clear patterns. The results were further analysed by calculating the gradient for each data series to produce a rate. Overall, there was no clear pattern to the data and the presence of fluorogalactose both increased and decreased the rate of product formation relative to the control reaction without fluorogalactose (table 3.2). These results do not imply that inhibition has occurred. To confirm this, a plot of log

[fluorogalactose] versus rate would normally be plotted. However this would have been pointless for this assay due to the unusual results generated.

Fluorogalactose concentration (μM)	**Rate (duplicate 1) for 0.25 U/mL β-galactosidase (mmol $dm^{-3}.s^{-1}$)**	**Rate (duplicate 2) for 0.25 U/mL β-galactosidase (mmol $dm^{-3}.s^{-1}$)**	**Rate (duplicate 1) for 0.5 U/mL β-galactosidase (mmol $dm^{-3}.s^{-1}$)**	**Rate (duplicate 2) for 0.5 U/mL β-galactosidase (mmol $dm^{-3}.s^{-1}$)**
0	6.14×10^{-7}	6.37×10^{-7}	1.99×10^{-6}	2.44×10^{-6}
10	8.65×10^{-7}	8.41×10^{-7}	2.07×10^{-6}	1.97×10^{-6}
25	9.34×10^{-7}	8.24×10^{-7}	2.26×10^{-6}	1.88×10^{-6}
50	7.31×10^{-7}	7.39×10^{-7}	1.89×10^{-6}	1.96×10^{-6}
100	7.74×10^{-7}	1.11×10^{-7}	1.79×10^{-6}	2.22×10^{-6}
500	6.22×10^{-8}	6.99×10^{-7}	1.85×10^{-6}	2.01×10^{-6}
1000	6.84×10^{-7}	7.19×10^{-7}	1.86×10^{-6}	2.23×10^{-6}
2000	5.45×10^{-7}	7.46×10^{-7}	2.65×10^{-6}	2.17×10^{-6}

Table 3.2: a comparison of the rates for the reactions in the presence of increasing concentrations of fluorogalactose at each enzyme concentration, calculated from the gradients of each data series for the assay.

In order to compare the inhibitory ability of galactose and fluorogalactose in the same assay, the assay conditions were adapted. The assay for each inhibitor was still carried out in duplicate, but was now only carried out using 0.25 U/mL β-galactosidase. In retrospect this seems to be a misguided decision to have taken as the results were clearer and more reliable at the higher enzyme concentration. This decision is likely to be based on earlier misinterpretation of the data.

Additionally the assay was adapted to include extra concentrations of both galactose and fluorogalactose. This increased the highest concentration of fluorogalactose used in the assay, but only added intermediate concentrations of galactose. The maximum usable concentration of fluorogalactose was limited by the overall molarity and total amount of sample available. However, this should not have been a problem for the galactose which widely available. Again it seems to have been an ill thought out decision not to increase the concentration of galactose used in the assay to test whether full competitive inhibition could be achieved at higher concentrations.

For this assay it is also evident a minor change had accidentally been made to the spectrophotometer measurement parameters. This was noticed after the event as the number of measurement cycles were kept the same, but the overall number of seconds decreased to approximately 3000 seconds instead of 3300 seconds. Comparing the machine output files for

this assay with the previous assay highlights that the number of reads had been decreased from 100 to 10 and the kinetic interval from 13 to 12. Therefore the amount of time spent shaking during each cycle was decreased slightly and correlates well with the decrease in time. Apart from these changes, the method used was identical to that for previous assays. The assay was not successful (data not shown). From the general spread of the data and minimal change in absorbance, and therefore concentration, it seemed most likely that that the enzyme had degraded. Degradation is possible from repeated freeze-thawing of the enzyme.

As the assays were not providing evidence of inhibition and were giving unexpected results, it was thought that the results might be affected by the order of addition of the reactants. The assay conditions were amended to incubate the β-galactosidase in the presence of either galactose or fluorogalactose, before adding the PNP-gal. All other assay conditions were kept the same. Thus as for previous assays, 50 μM PNP-gal was hydrolysed by 0.25 U/mL β-galactosidase in the presence of either galactose or fluorogalactose. The absorbance of the 4-nitrophenol product was measured at 405 nm over a 3000 second time course. Each absorbance was converted to a concentration and the concentrations plotted onto a graph against time.

The assay was successful but the results were not very clean. For the assay in the presence of galactose, the two duplicates have very different patterns of results (figure 3.9). One duplicate has a number of odd results. The hydrolysis in the presence of 1 mM galactose produced a negative slope, which should be impossible as the measured amounts of product cannot decrease. Hydrolyses in the presence of 17 mM and 20 mM galactose using 0.25 U/mL β-galactosidase are not linear. In this example it does not appear to be an effect of inhibition by galactose, nor an effect of substrate depletion as it is only found in one duplicate. It also does not follow the previous patterns.

Similar results are also seen in the assays containing fluorogalactose (figure 3.10). In these assays, hydrolyses in the presence of 500 μM fluorogalactose, 700 μM fluorogalactose and 1 mM fluorogalactose using 0.25 U/mL β-galactosidase are not linear. This could be an effect of inhibition, or a combination of inhibition and substrate depletion. It is more likely as an explanation than for the galactose as it is observed across both duplicates.

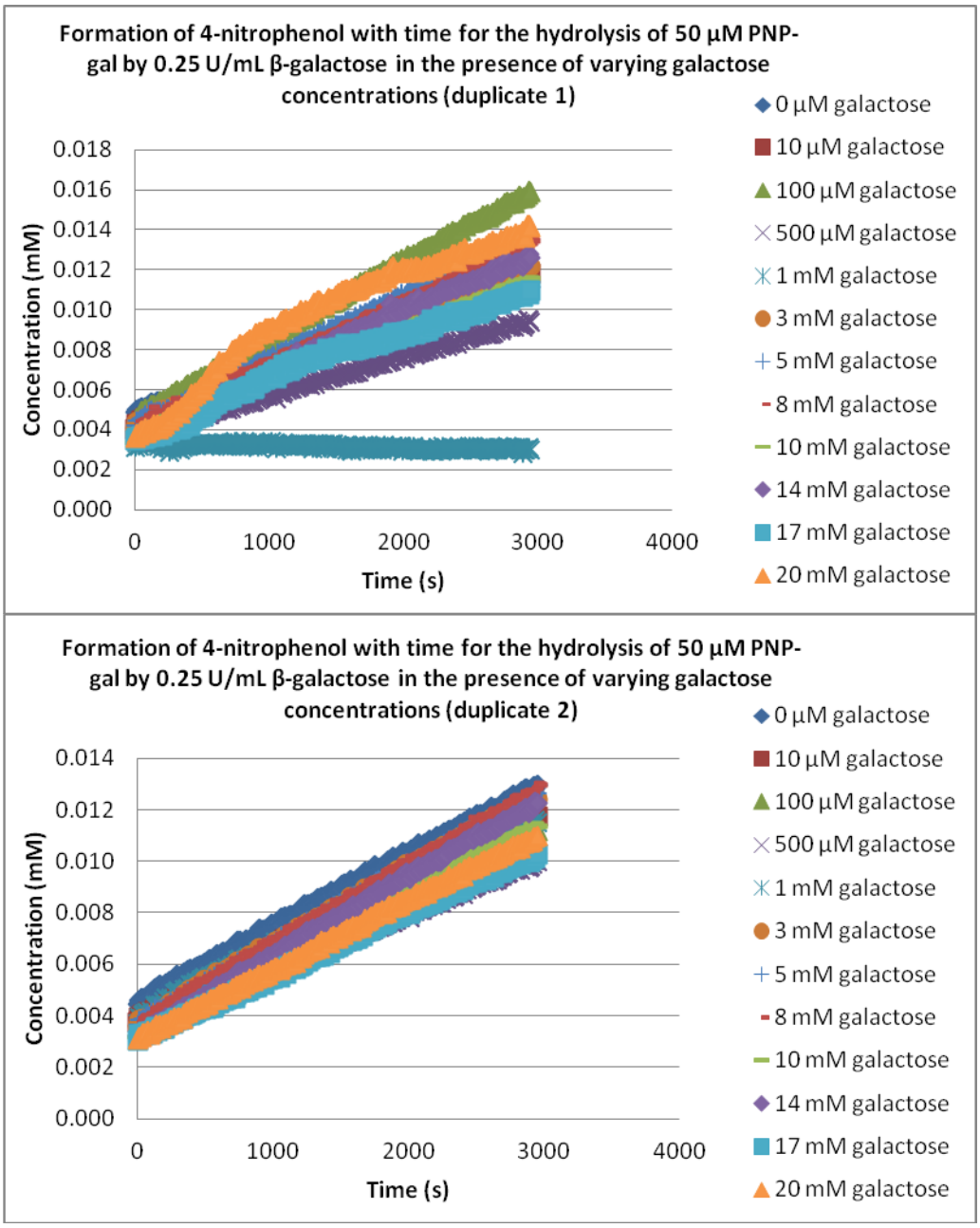


Figure 3.9: formation of 4-nitrophenol from the hydrolysis of PNP-gal catalysed by β-galactosidase in the presence of increasing concentrations of galactose. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

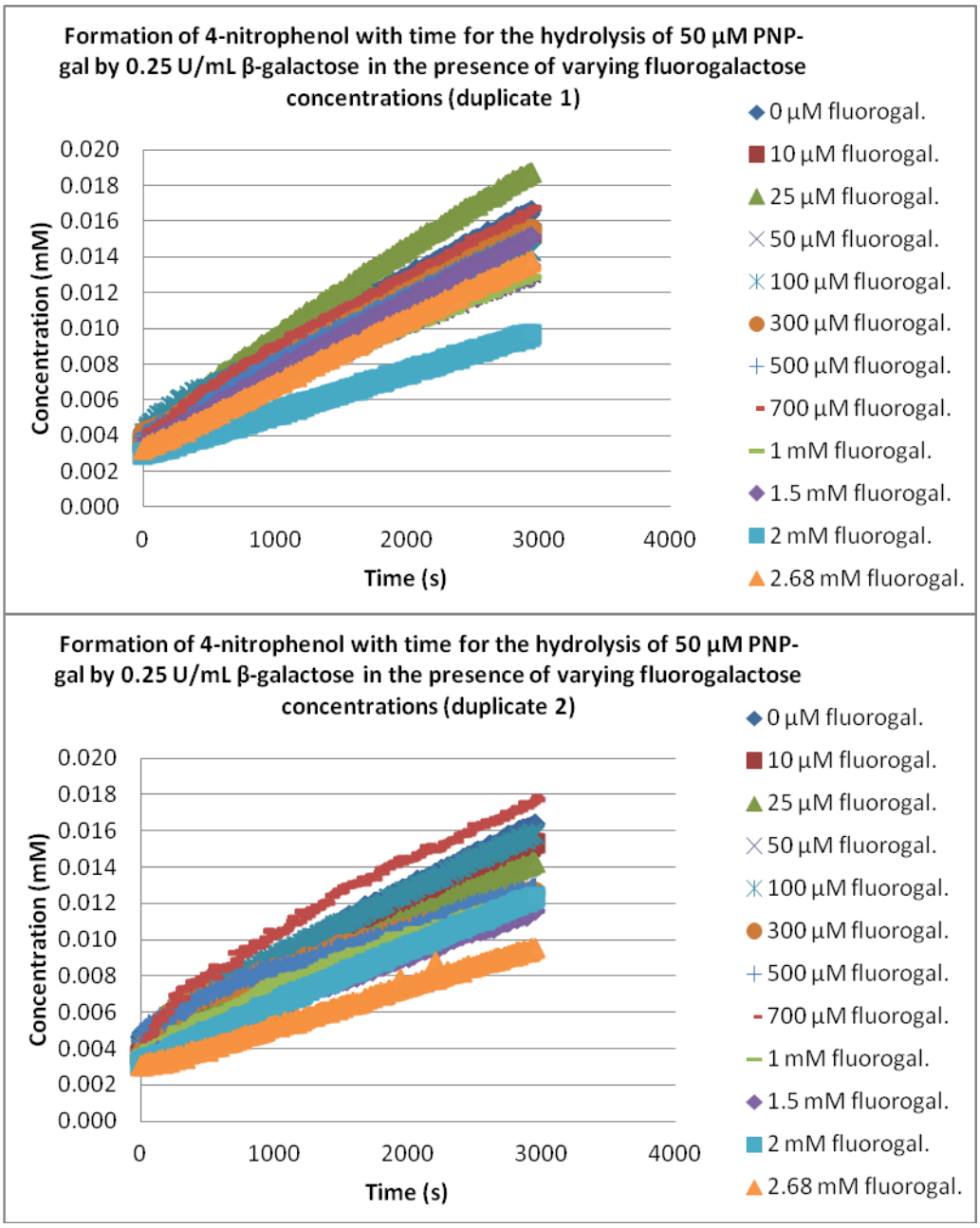


Figure 3.10: formation of 4-nitrophenol from the hydrolysis of PNP-gal catalysed by β-galactosidase in the presence of increasing concentrations of fluorogalactose. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

Although there are some noticeable exceptions, the assay results did suggest that the presence of the inhibitor had a small effect on the slope. The gradient of each slope was calculated to produce a rate (table 3.3).

Galactose concentration (mM)	**Rate 1 (mmol $dm^{-3}.s^{-1}$)**	**Rate 2 (mmol $dm^{-3}.s^{-1}$)**
0	2.09×10^{-6}	2.84×10^{-6}
0.01	2.56×10^{-6}	2.68×10^{-6}
0.1	3.85×10^{-6}	2.56×10^{-6}
0.5	1.92×10^{-6}	2.24×10^{-6}
1	1.04×10^{-7}	2.98×10^{-6}
3	2.72×10^{-6}	2.55×10^{-6}
5	3.17×10^{-6}	3.02×10^{-6}
8	3.06×10^{-6}	3.07×10^{-6}
10	2.63×10^{-6}	2.78×10^{-6}
14	3.17×10^{-6}	3.08×10^{-6}
17	2.50×10^{-6}	2.45×10^{-6}
20	3.42×10^{-6}	2.67×10^{-6}
Fluorogalactose concentration (µM)	**Rate 1 (mmol $dm^{-3}.s^{-1}$)**	**Rate 2 (mmol $dm^{-3}.s^{-1}$)**
0	4.16×10^{-6}	3.92×10^{-6}
10	3.93×10^{-6}	3.92×10^{-6}
25	4.98×10^{-6}	3.57×10^{-6}
50	3.19×10^{-6}	2.84×10^{-6}
100	3.25×10^{-6}	3.75×10^{-6}
300	3.94×10^{-6}	2.50×10^{-6}
500	3.75×10^{-6}	2.54×10^{-6}
700	4.15×10^{-6}	4.19×10^{-6}
1000	3.30×10^{-6}	2.95×10^{-6}
1500	3.99×10^{-6}	2.84×10^{-6}
2000	2.29×10^{-6}	3.11×10^{-6}
2680	3.60×10^{-6}	2.26×10^{-6}

Table 3.3: calculated rates for the hydrolysis of 50 µM PNP-gal by β-galactosidase in the presence of galactose or fluorogalactose.

Each rate was plotted against log [galactose] or log [fluorogalactose] as relevant. Though some results had suggested that there may have been a decrease in rate, there was no reversed sigmoid curve or any other identifiable trend (supplementary data). The increased number of

galactose concentrations did not aid detection of a trend. Although the concentrations were spread on a normal scale, on a logarithmic scale the measurement points were closely spaced.
The updated assay method has an unexpected disadvantage. The opening measured concentration is spread over a slightly wider range than in previous assays. This was an unexpected change given that the substrate could be easily, and more quickly, dispensed using a multichannel pipette. This should have decreased the time between the addition of all the solutions to the plate and taking the first measurements, which in turn should have produced a cleaner dataset. This was not possible when initiating the hydrolysis by the addition of enzyme, as there was not a suitably sized multichannel pipette available for dispensing 2 µL.

3.2.2 Updating the design of the assay

Reviewing the results of the previous assays it was considered that the assay design was faulty: the previous assays had used exceedingly small volumes of β-galactosidase and inhibitor (2 µL) in 200 µL of PNP-gal. It was thought that this may have been contributing to the inaccurate dispensation into wells and inconsistent results. The assay was redesigned to mix larger volumes of β-galactosidase and inhibitor (25 µL) with the PNP-gal (150 µL). Additionally the β-galactosidase and inhibitor were incubated first before the reaction was initiated with the addition of PNP-gal. The shake parameters were again adjusted slightly which increased the assay measurement time to approximately 4200 seconds.

The changes caused the reactions to have much closer concentrations at the first measurement point. However, the assay still had many problems with reactions not occurring in a number of wells and other reactions appearing to cease measurement early (figures 11 and 12). As with previous assays, overall the data was very inconsistent and did not show the same pattern of data between duplicates or evidence of inhibition. For all hydrolyses in the presence of galactose none of the data appears to be linear, there is evidence of a slight sigmoid curve. This is also observed in some hydrolyses in the presence of fluorogalactose.

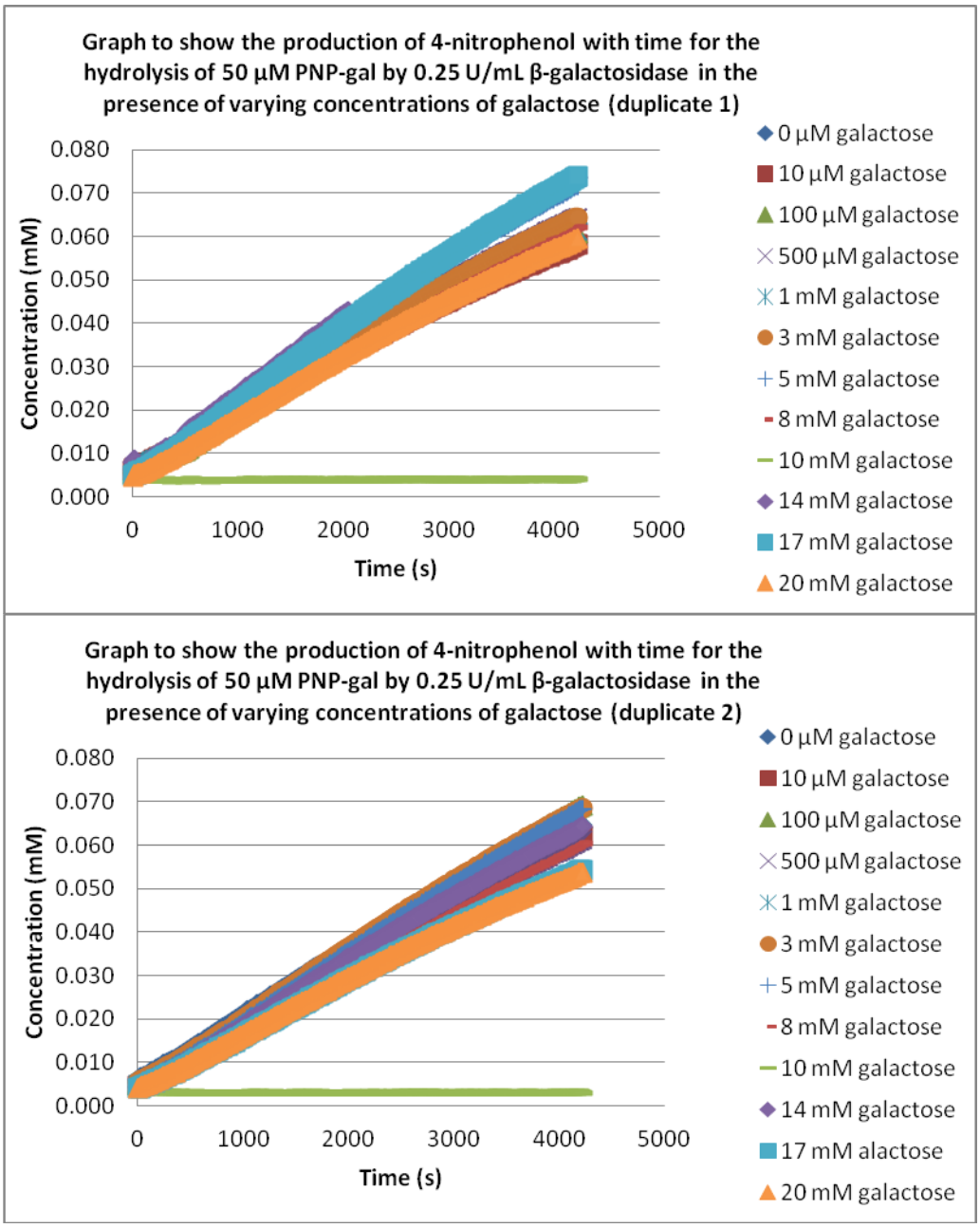


Figure 3.11: formation of 4-nitrophenol from the hydrolysis of PNP-gal catalysed by β-galactosidase in the presence of increasing concentrations of galactose. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

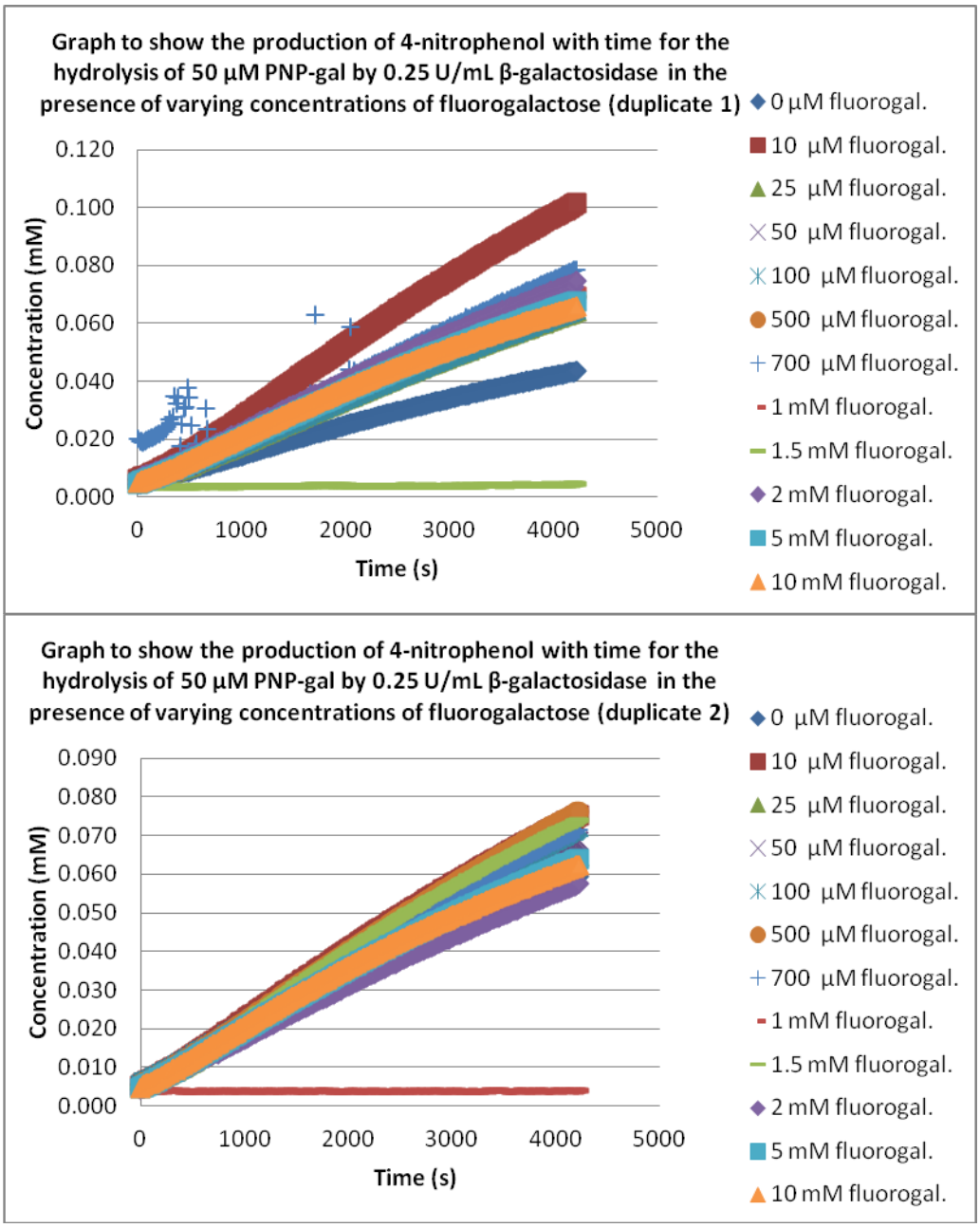


Figure 3.12: formation of 4-nitrophenol from the hydrolysis of PNP-gal catalysed by β-galactosidase in the presence of increasing concentrations of fluorogalactose. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

To confirm the results did not show a pattern of inhibition, the gradient of each data series was calculated. These were plotted against log [inhibitor] in order to check whether inhibition was present. As predicted from the initial graphs there was no pattern of inhibition: the graphs did not demonstrate a sigmoid curve or any obvious trend (supplementary data). The assay was unable to determine whether galactose or fluorogalactose were acting as inhibitors.

3.2.3 Ensuring beta-galactosidase conforms to Michaelis-Menten kinetics.

As the results of the previous assays had not been showing any expected trends, it was considered that there might be a problem with the β-galactosidase. In order to check the β-galactosidase, a standard kinetic assay was designed to test the hydrolysis of increasing concentrations of PNP-gal using 0.5 U/mL β-galactosidase. The absorbance of 4-nitrophenol was measured at 405 nm and the resulting absorbances converted to a series of concentrations using the Beer-Lambert law, $A=\varepsilon Cl$.

As expected there was a positive correlation between increased substrate concentration and the amount of 4-nitrophenol produced, and hence the rate (figure 13). The gradients of each data series were calculated to produce the rate (table 4). Plotting substrate concentration against rate produces a characteristic curve that fits Michaelis-Menten kinetics and shows extremely good correlation between the rates in the two duplicates (figure 14). The graph did also highlight that further measurement points were required in the higher substrate concentration values, and that possibly too many data points were used in the lowest substrate concentration values.

Concentration of PNP-gal (mM)	**Rate – duplicate 1 (mmol $dm^{-3}.s^{-1}$)**	**Rate – duplicate 2 (mmol $dm^{-3}.s^{-1}$)**
0.0001	4.45×10^{-8}	2.80×10^{-8}
0.0005	7.54×10^{-8}	3.59×10^{-8}
0.001	7.80×10^{-8}	4.96×10^{-8}
0.01	3.18×10^{-7}	3.22×10^{-7}
0.025	7.63×10^{-7}	7.74×10^{-7}
0.05	1.54×10^{-6}	1.55×10^{-6}
0.1	2.96×10^{-6}	3.03×10^{-6}
0.25	6.41×10^{-6}	6.49×10^{-6}
0.5	1.32×10^{-5}	1.37×10^{-5}
0.75	1.90×10^{-5}	1.93×10^{-5}
1	2.36×10^{-5}	2.43×10^{-5}
10	6.93×10^{-5}	6.95×10^{-5}

Table 3.4: Rates for the hydrolysis of increasing concentrations of PNP-gal by 0.5 U/mL β-galactosidase

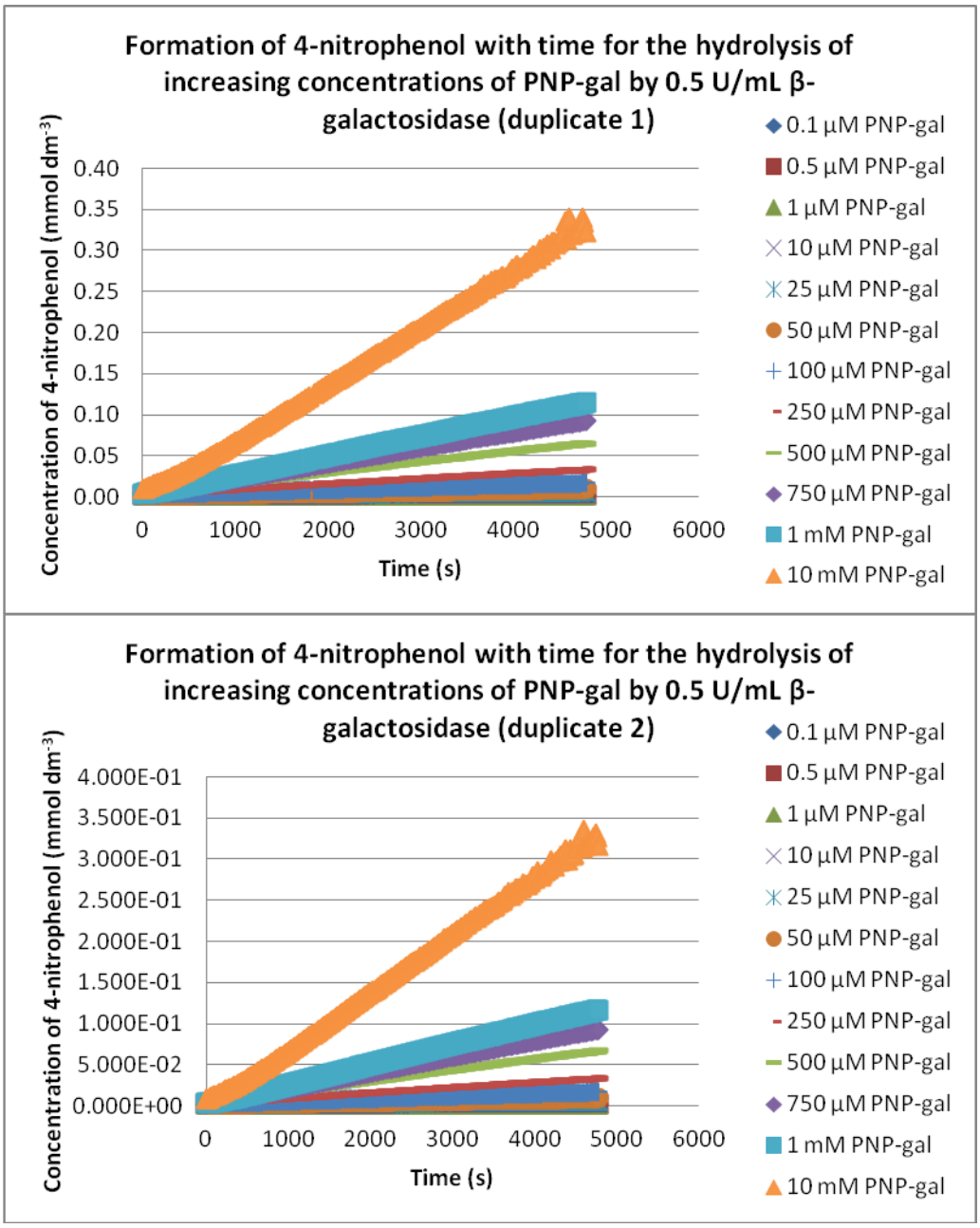


Figure 3.13: formation of 4-nitrophenol with time for the hydrolysis of increasing concentrations of PNP-gal by 0.5 U/mL β-galactosidase. Absorbance was measured at 405 nm using a spectrophotometer and the converted to concentration using Beer-Lambert law.

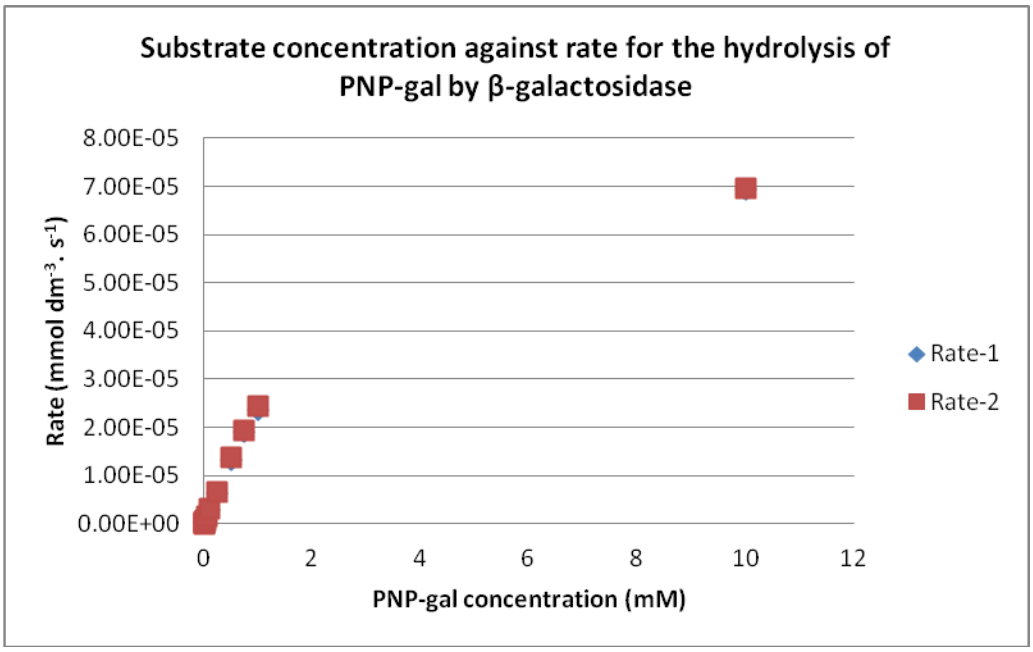


Figure 3.14: plot of substrate concentration versus rate for the hydrolysis of PNP-gal by β-galactosidase. The assay was carried out in duplicate (listed as rate 1 and rate 2) and there is extremely good correlation between the two sets of rates.

The rates generated were used to plot a Hanes-Woolf plot which derives from a linear form of the Michaelis-Menten equation. This graph can be used to determine the kinetic parameters of an enzyme. In the Hanes-Woolf plot the substrate concentration ([s]) is plotted against substrate concentration/rate ([s]/v) and should produce a straight line. Plotting Hanes-Woolf using the data obtained from this assay produced a graph that was not linear (figure 3.15). The graph contained a number of data points crowded about the y axis and an odd point corresponding to the 10 mM substrate concentration. This was unexpected given that the initial graphs of time versus concentration had produced data that appeared to be correct, and the plot of substrate concentration versus rate had also given a characteristic curve. As previously noted the assay does not use a good spread of substrate concentrations, therefore it was decided to repeat the assay using more substrate concentrations between 1 and 10 mM.

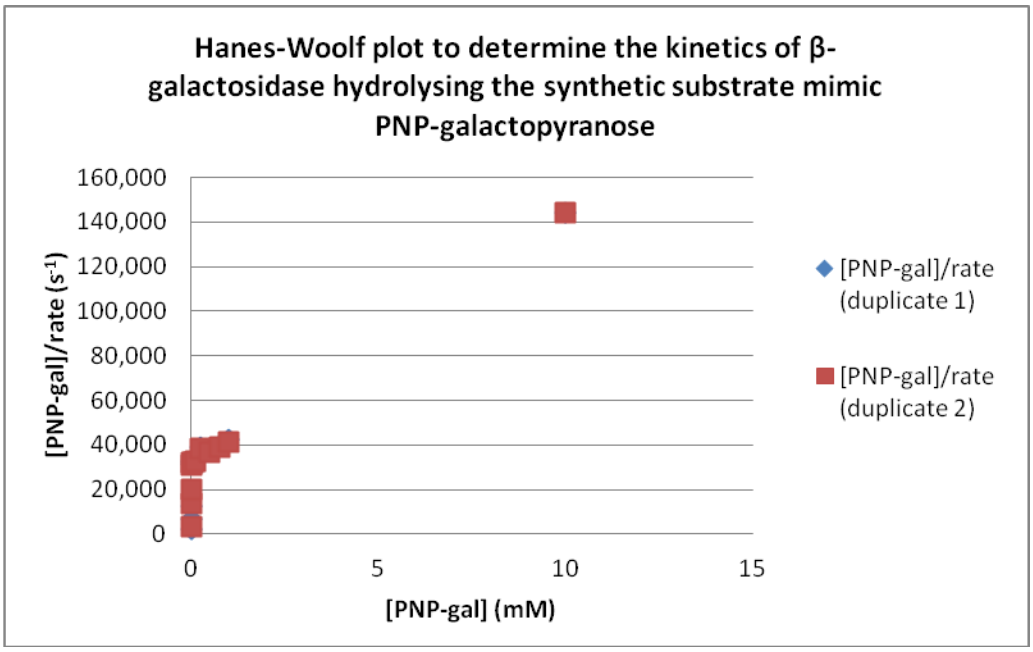


Figure 3.15: Hanes-Woolf plot of [s] vs [s]/v for the determination of the kinetic parameters of β-galactosidase. The two data sets were generated from the two times the absorbance readings were carried out and show extremely good correlation.

The assay was repeated using the same conditions as previously but with the addition of extra measurements between 1 and 10 mM PNP-gal. As in the previous assay, there was a positive correlation between increasing substrate concentration and the amount of 4-nitrophenol generated over time (supplementary data). The gradient of each data series was calculated to generate a rate (table 3.5). The rates were used to plot both a graph of substrate concentration versus rate (figure 3.16) and a Hanes-Woolf plot (figure 3.17).

PNP-gal concentration (mM)	Rate 1 (mmol $dm^{-3}.s^{-1}$)	Rate 2 (mmol $dm^{-3}.s^{-1}$)
0.0001	5.34×10^{-8}	2.46×10^{-8}
0.0005	3.99×10^{-8}	3.44×10^{-8}
0.001	4.55×10^{-8}	2.90×10^{-8}
0.01	7.18×10^{-8}	5.41×10^{-8}
0.025	9.57×10^{-8}	8.76×10^{-8}
0.05	1.46×10^{-7}	1.46×10^{-7}
0.1	2.67×10^{-7}	4.52×10^{-7}
0.25	6.05×10^{-7}	6.11×10^{-7}
0.5	1.13×10^{-6}	1.10×10^{-6}
0.75	1.59×10^{-6}	1.56×10^{-6}
1	1.97×10^{-6}	1.96×10^{-6}
2	3.08×10^{-6}	3.14×10^{-6}
4	4.25×10^{-6}	4.40×10^{-6}
6	4.90×10^{-6}	4.97×10^{-6}
8	5.44×10^{-6}	5.53×10^{-6}
10	5.72×10^{-6}	5.81×10^{-6}

Table 3.5: Rates for the hydrolysis of increasing concentrations of PNP-gal by 0.5 U/mL β-galactosidase

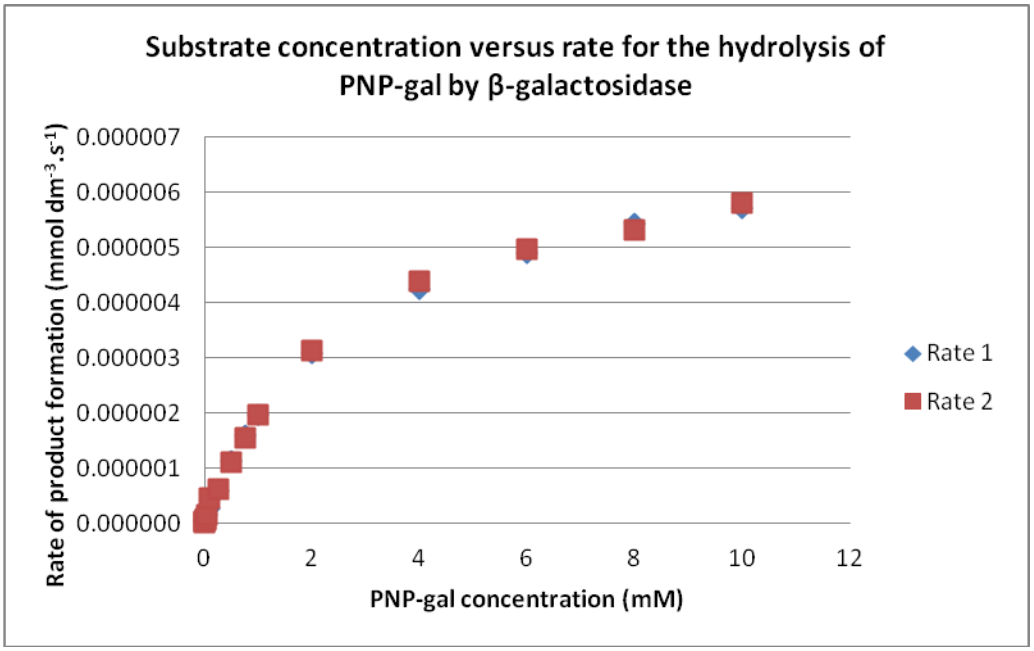


Figure 3.16: plot of substrate concentration versus rate to demonstrate the effect of substrate concentration on rate of 4-nitrophenol formation.

As in the previous assay the plot of PNP-gal concentration versus rate produces an almost perfect curve expected from an enzyme behaving according to Michaelis-Menten kinetics. The curve does not reach Vmax. In order to have reached V max higher concentrations of substrate should have been used. The graph also demonstrates that too many substrate concentrations were used in the lowest regions as a number of data points are crowded and overlapping. From these results and the time versus concentration graphs, a Hanes-Woolf plot should demonstrate linear data. This was, however, not the case as the results were not linear (figure 17). As can be seen there are a number of points on a straight line with a number of points crowded on the y axis.

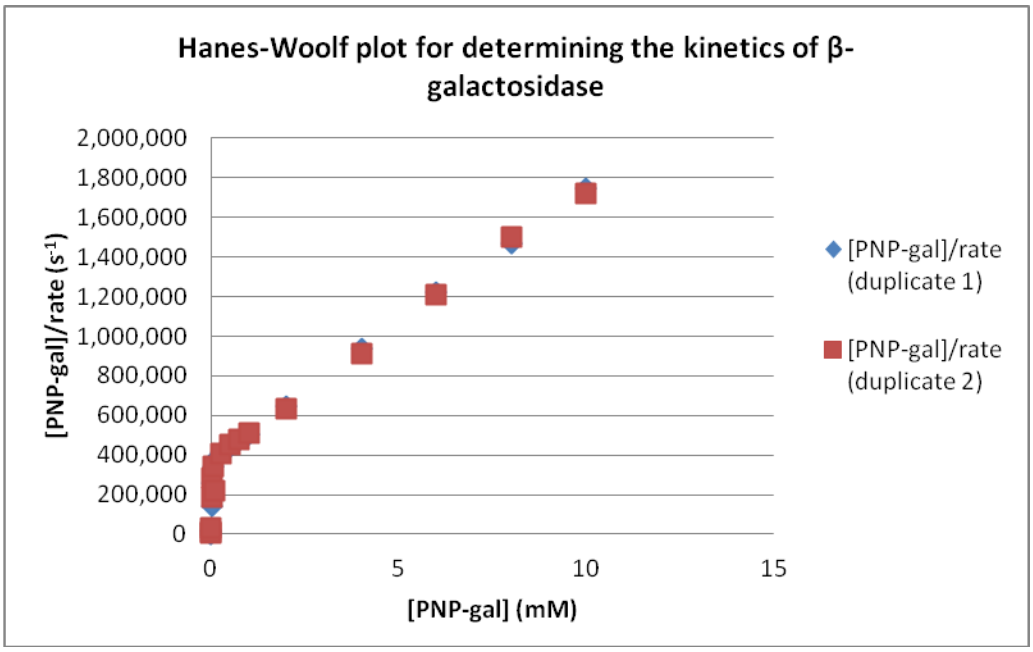


Figure 3.17: Hanes-Woolf plot of [s] vs [s]/v for the determination of the kinetic parameters of β-galactosidase. The two data sets were generated from the two times the absorbance readings were carried out and show extremely good correlation.

At the time when the assay was first carried out the results were a puzzle and unexpected. From the literature, various explanations such as allosteric interactions were excluded as classic Michaelis-Menten behaviour is observed in other studies. Allosteric interactions occur when an enzyme has multiple binding sites and binding of the first substrate molecule increases binding at the other sites. Allostery was a possibility in this enzyme as it has four binding sites. Allostery was easily ruled out by plotting substrate concentration versus rate. In allosteric enzymes this graph exhibits a sigmoid curve, however it was not sigmoid when plotted using the experimental data obtained in these assays.

It was also considered at the time that the enzyme was not pure enough for the purposes of the assay and that a contaminant was also able to bind and hydrolyse the substrate. However, on looking more closely at the results, a much simpler explanation fits the given pattern of results.

At the lowest substrate concentrations small errors in measurement are magnified. In this assay a number of measurements have been made using very dilute solutions of PNP-gal substrate (0.1-10 μM). The measured absorbance of the product from each of these substrate concentrations is therefore very small and hence the rate is also small. These results are likely to be much less accurate and reliable than the results from each hydrolysis at higher substrate concentrations. Any errors are further magnified on plotting the Hanes-Woolf plot leading to a number of data points sat on or around the y axis and not forming part of the straight line. Removing these data points leads to the characteristic Hanes-Woolf plot (figure 3.18).

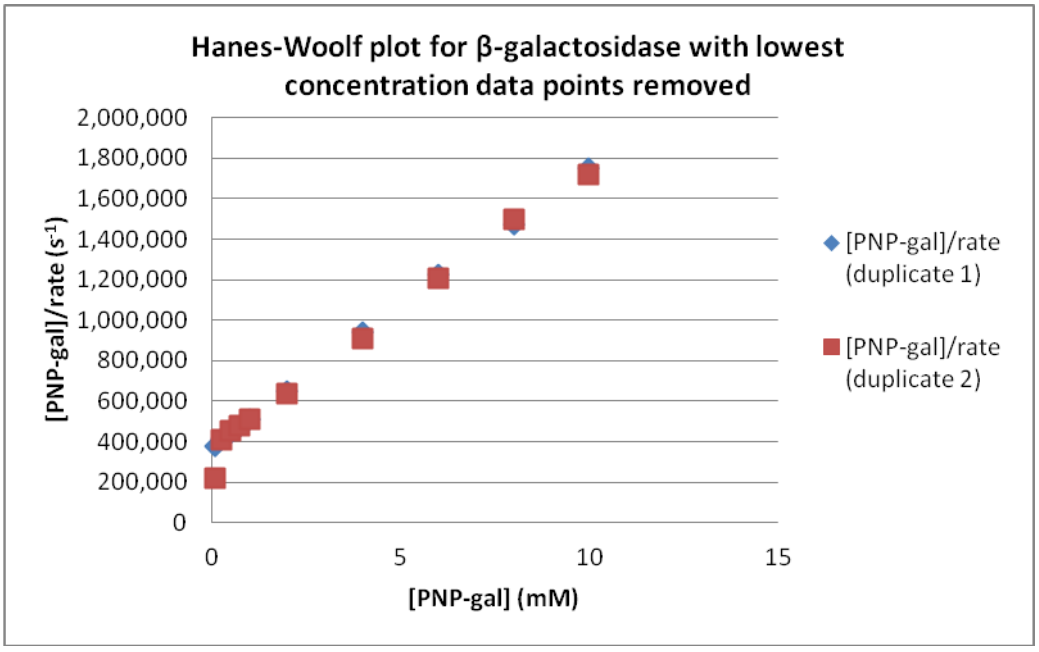


Figure 3.18: Hanes-Woolf plot for calculating kinetic parameters of β-galactosidase. This plot has omitted the very low substrate concentrations which are less reliable.

From this the gradients of the slopes and the equation of each line were calculated and from these the kinetic parameters could be calculated. As the point of the assay was to ensure that the enzyme was behaving according to Michaelis-Menten kinetics, and not exhibiting properties of co-operativity caused by allosteric binding sites, but not to actually calculate the kinetic parameters, these were not calculated. After solving the problem of the apparently anomalous data points and producing a typical Hanes-Woolf plot, it was assumed that the enzyme was behaving correctly and pure enough for the purpose. Therefore it was decided to return to assays exploring the inhibitory properties of galactose and fluorogalactose.

The assay was again redesigned to include higher concentrations of both fluorogalactose and galactose. The highest fluorogalactose concentration was substantially increased from 10 mM to 30 mM while the galactose concentration was increased to 100 mM. 50 μM PNP-gal was hydrolysed by 0.25 U/mL β-galactosidase in the presence of increasing concentrations of either

galactose or fluorogalactose. The β-galactosidase was first incubated in the presence of either galactose or fluorogalactose before the reaction was initiated with the addition of PNP-gal. The absorbance of the 4-nitrophenol product was measured over time and converted to a series of concentrations using the Beer-Lambert law. These were plotted against time on graphs (figures 3.19 and 3.20). In error each of these assays was only carried out only once, not in duplicate as per previous assays.

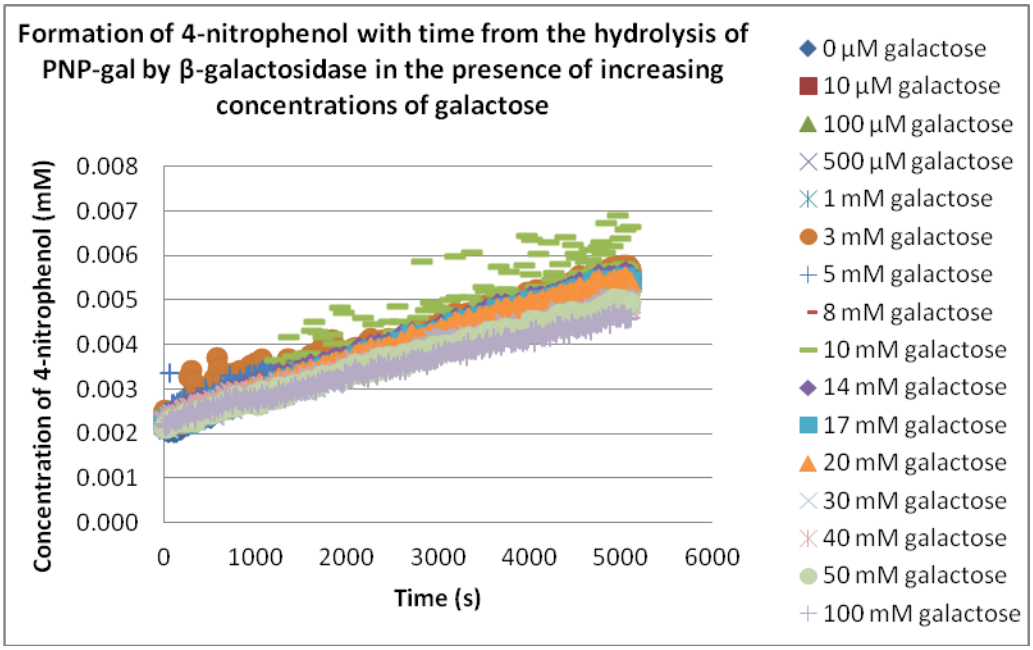


Figure 3.19: formation of 4-nitrophenol from the hydrolysis of PNP-gal catalysed by β-galactosidase in the presence of increasing concentrations of galactose. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

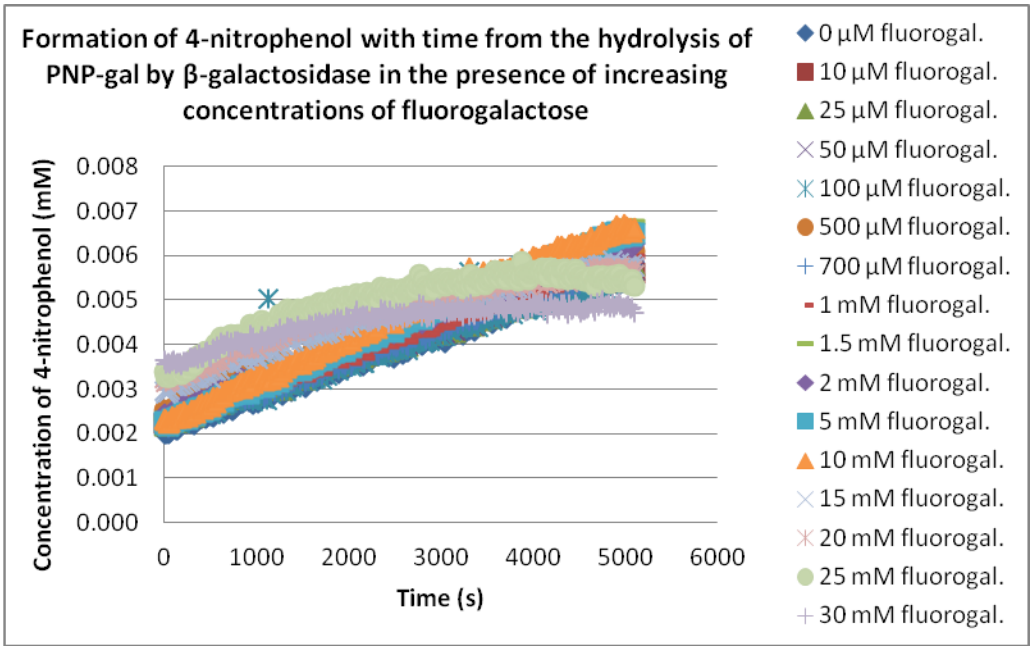


Figure 3.20: formation of 4-nitrophenol from the hydrolysis of PNP-gal catalysed by β-galactosidase in the presence of increasing concentrations of fluorogalactose. The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

The results of the assay in the presence of fluorogalactose show some very interesting results (figure 3.20). The results for wells containing the higher concentrations of fluorogalactose are not linear, and are much more noticeably curved than previous assays with the amount of 4-nitrophenol produced reaching a plateau. In this assay the fluorogalactose is obviously having an effect at higher concentrations, but it is not the expected result of the galactose decreasing the amount of 4-nitrophenol produced overall in the time period, and hence the rate of production.

The gradient of each slope for the assay in the presence of galactose was calculated to generate a rate (table 6). The rate was used to generate a dose response curve by plotting log [galactose] versus rate (figure 3.21). The rate at which 4-nitrophenol was produced was very small in all cases in this assay so it is possible that the assay was not working correctly. The β-galactosidase could have partially expired or a dilution error could have led to either the β-galactosidase or the PNP-gal being wrong, for example much more dilute than anticipated. Alternatively incubating the β-galactosidase with the inhibitor could be having some effect. The dose response curve confirms that there appears to be no significant inhibition by the galactose. If inhibition is occurring the graph should exhibit a sigmoid curve.

Galactose concentration (mM)	Rate (mM.s^{-1})
0	6.91 x 10-7
0.01	6.34 x 10-7
0.1	6.59 x 10-7
0.5	6.60 x 10-7
1	6.44 x 10-7
3	6.17 x 10-7
5	5.59 x 10-7
8	6.37 x 10-7
10	7.86 x 10-7
14	6.35 x 10-7
17	6.44 x 10-7
20	6.32 x 10-7
30	5.47 x 10-7
40	4.96 x 10-7
50	5.31 x 10-7
100	4.63 x 10-7

Table 3.6: Rates calculated for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of increasing concentrations of galactose.

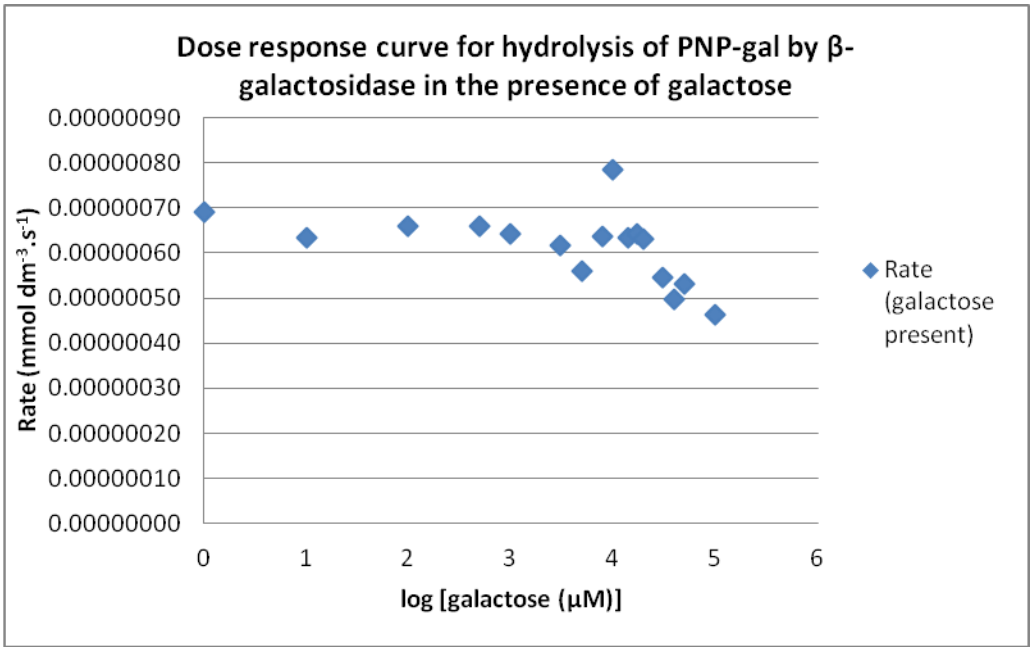


Figure 3.21: IC_{50} dose-response plot to determine whether galactose is an inhibitor of β-galactosidase for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase.

As a number of the assays in the presence of fluorogalactose did not produce linear results, the time versus concentration graph was re-plotted to only include approximately the first 1000 seconds, which was the extent of the linear portion of those results (figure 3.22).

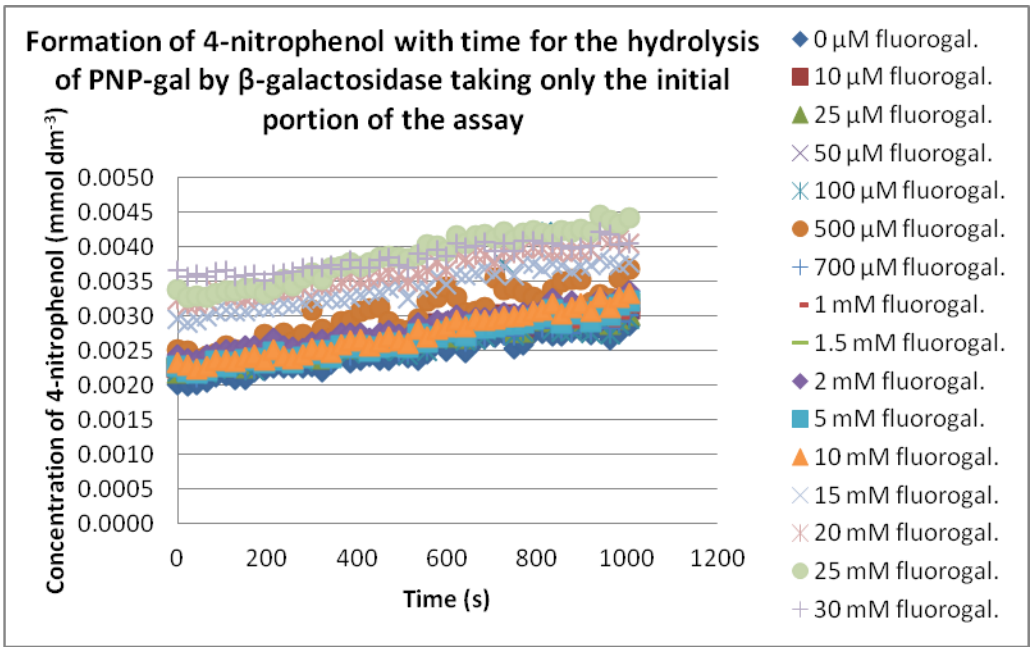


Figure 3.22: formation of 4-nitrophenol from the hydrolysis of PNP-gal catalysed by β-galactosidase in the presence of increasing concentrations of fluorogalactose (using approximately the first 1000 seconds of the assay results). The absorbance of 4-nitrophenol was measured at 405 nm using a spectrophotometer and the absorbance converted to concentration.

This graph highlights that there is a large spread of opening concentrations. This could be due to the time lag of adding the PNP-gal to initiate the reaction, though this seems unlikely as the solution was added with a multichannel pipette. Also if it was due to the time lag then it would expected that the wells containing the highest concentrations of fluorogalactose that would be at lower concentration as these would have been added later due to the assay set-up. However, the order of decreasing concentration does not correspond to this order. The highest opening concentrations correspond to the highest fluorogalactose concentrations. As the highest opening concentration of product corresponds with the wells containing the highest concentration of fluorogalactose, it could be an effect of incubating the β-galactosidase with the fluorogalactose. However this doesn't make sense as if it was acting as an inhibitor, then again the opening concentration might reasonably be expected to be lower as competitive inhibition takes place.

The graph also highlights that the data for the well containing 500 µM fluorogalactose had very uneven increase in product concentration, with apparent gaps in the measurement.

The gradients of each data series were calculated to produce a rate (table 7) and the rates used to plot a dose response curve (figure 3.23). The dose response curve does not show inhibition taking place. There is no sigmoid curve and no clear trend to the data.

Fluorogalactose concentration (mM)	**Rate (mmol $dm^{-3}.s^{-1}$)**
0	8.19×10^{-7}
0.01	8.37×10^{-7}
0.025	8.19×10^{-7}
0.05	7.99×10^{-7}
0.1	8.62×10^{-7}
0.5	1.13×10^{-6}
0.7	7.93×10^{-7}
1	8.27×10^{-7}
1.5	9.55×10^{-7}
2	8.17×10^{-7}
5	9.44×10^{-7}
10	1.04×10^{-6}
15	9.03×10^{-7}
20	9.48×10^{-7}
25	1.21×10^{-6}
30	6.09×10^{-7}

Table 3.7: the rates for the first 1000 seconds of the production of 4-nitrophenol from the hydrolysis of PNP-gal by β-galactosidase.

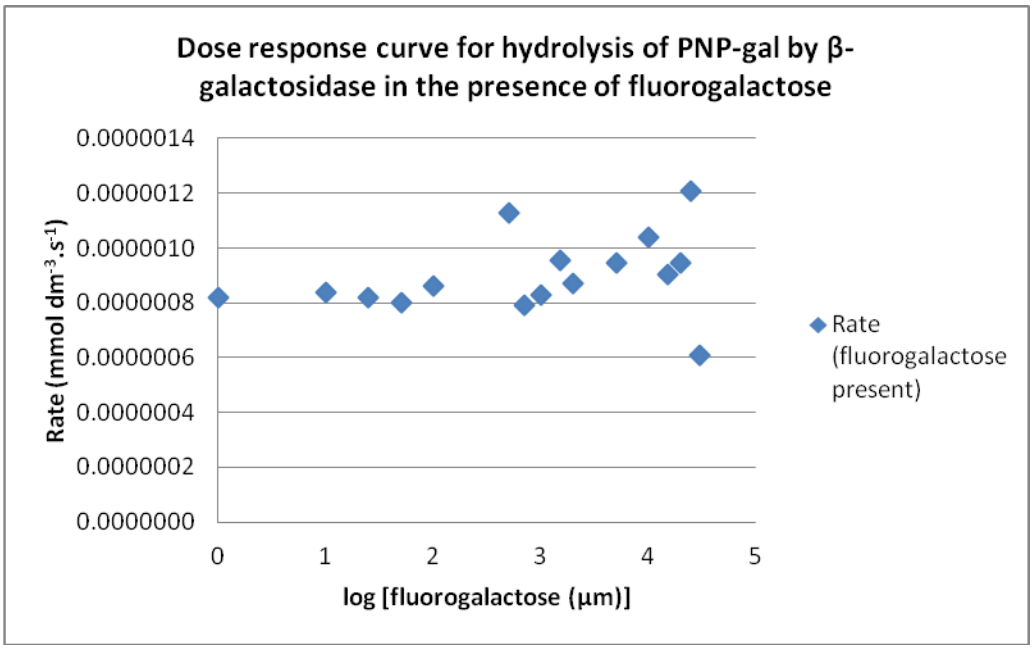


Figure 3.23: IC_{50} dose-response plot to determine whether fluorogalactose is an inhibitor of β-galactosidase for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase.

3.2.4 A major overhaul of the assay

Based on these assay results, the design of the assay was radically overhauled. The assay was increased in complexity by carrying out measurements of the absorbance of the 4-nitrophenol product for the hydrolysis of 10, 25, 50 and 100 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of increasing concentrations of galactose. This allows the data to be analysed in two ways. The data can be continued to be analysed as previously by generating a rate and plotting log [inhibitor] against rate to produce a dose response plot; but also as standard kinetic assays by plotting the rate obtained from different substrate concentrations against a fixed inhibitor concentration. The new assay encompassed PNP-gal concentrations of 10, 25, 50 and 100 µM with a range of galactose concentrations from 0 to 100 mM. The kinetic interval of the spectrophotometer control program was also increased, possibly accidentally, causing the assay to run for approximately 5700 seconds. With hindsight, it would have been better to have used higher substrate concentrations, for example 50, 100, 250 and 500 µM, but as for the previous assays the meaning of the generated data had been misinterpreted, therefore very low substrate concentrations were selected.

The results of the assay, as expected, showed a positive correlation between 4-nitrophenol concentration and time, and also between 4-nitrophenol concentration and substrate concentration (supplementary data). The results for the hydrolysis of 10 µM PNP-gal were very messy as there was a large spread in the data and only a small increase in concentration. As previously discussed this would be expected as the substrate is very dilute as is the enzyme which leads to a magnification in errors and is in the less reliable range of the spectrophotometer. The results for the hydrolysis at the remaining substrate concentrations

were much better. The gradient of each data series was calculated to generate a rate (table 3.8).

Galactose concentration (mM)	**Rate (mmol $dm^{-3}.s^{-1}$) (10 µM PNP-gal)**	**Rate (mmol $dm^{-3}.s^{-1}$) (25 µM PNP-gal)**	**Rate (mmol $dm^{-3}.s^{-1}$) (50 µM PNP-gal)**	**Rate (mmol $dm^{-3}.s^{-1}$) (100 µM PNP-gal)**
0	2.52×10^{-7}	4.20×10^{-7}	7.03×10^{-7}	1.69×10^{-6}
0.01	1.18×10^{-7}	3.18×10^{-7}	7.01×10^{-7}	1.71×10^{-6}
0.1	1.28×10^{-7}	3.24×10^{-7}	6.40×10^{-7}	1.64×10^{-6}
0.5	1.21×10^{-7}	3.51×10^{-7}	6.61×10^{-7}	1.69×10^{-6}
1	1.54×10^{-7}	3.64×10^{-7}	7.74×10^{-7}	1.63×10^{-6}
3	1.41×10^{-7}	3.54×10^{-7}	6.76×10^{-7}	1.72×10^{-6}
5	1.37×10^{-7}	3.43×10^{-7}	7.65×10^{-7}	1.71×10^{-6}
8	1.49×10^{-7}	3.55×10^{-7}	6.74×10^{-7}	1.71×10^{-6}
10	4.22×10^{-8}	3.96×10^{-7}	6.59×10^{-7}	1.73×10^{-6}
14	1.45×10^{-8}	3.69×10^{-7}	6.79×10^{-7}	1.72×10^{-6}
17	1.35×10^{-7}	3.27×10^{-7}	6.61×10^{-7}	1.67×10^{-6}
20	1.28×10^{-7}	3.23×10^{-7}	6.19×10^{-7}	1.66×10^{-6}
30	1.21×10^{-7}	2.84×10^{-7}	5.33×10^{-7}	1.48×10^{-6}
40	9.61×10^{-8}	2.53×10^{-7}	4.89×10^{-7}	1.44×10^{-6}
50	9.95×10^{-8}	2.68×10^{-7}	4.91×10^{-7}	1.60×10^{-6}
100	1.09×10^{-7}	2.60×10^{-7}	4.64×10^{-7}	1.40×10^{-6}

Table 3.8: rates of the hydrolysis of different concentrations of PNP-gal by β-galactosidase in the presence of increasing concentrations of galactose.

The rates were used to plot a series of dose response curves (supplementary data). Unlike previous assays there is very limited evidence of an overall trend of a reverse sigmoid shape in the data for the hydrolysis of 50 and 100 µM PNP-gal substrate by β-galactosidase (figures 3.24-3.25).

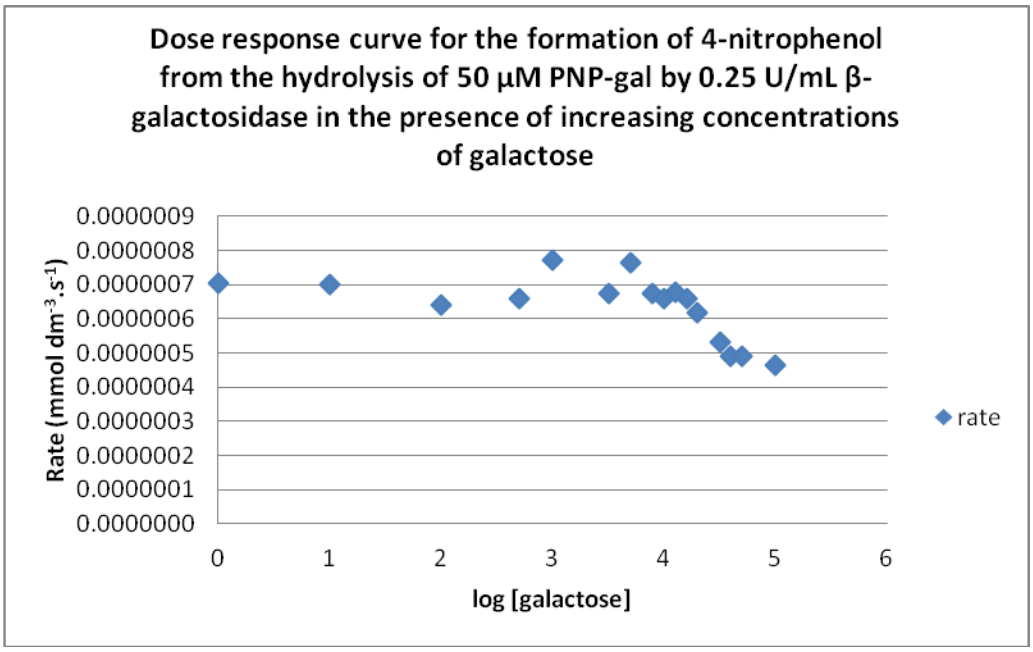


Figure 3.24: IC_{50} dose-response plot to determine whether galactose is an inhibitor of β-galactosidase for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase.

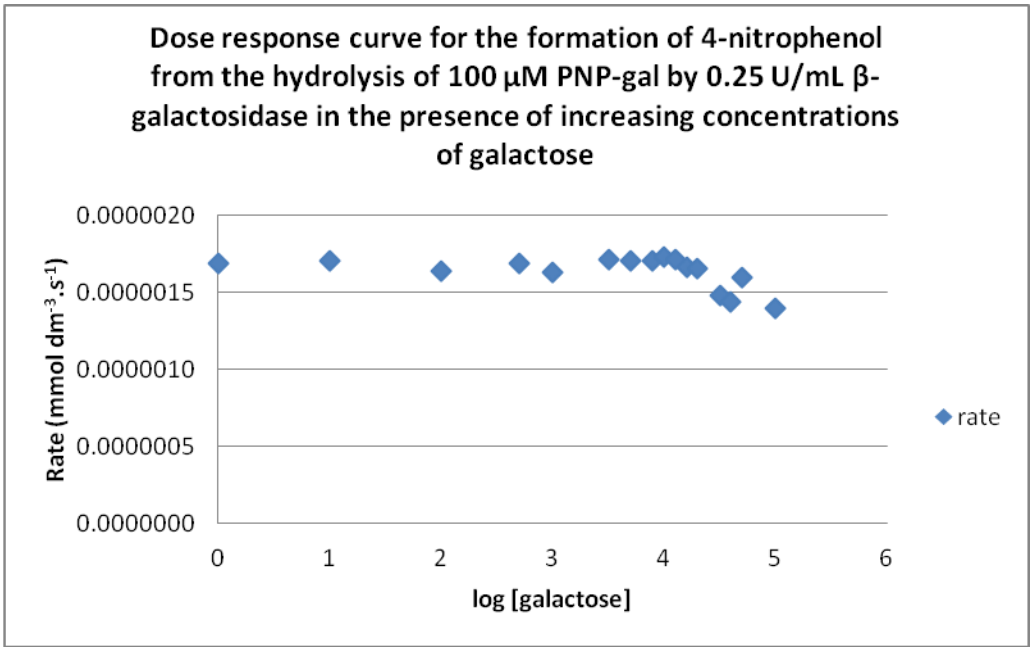


Figure 3.25: IC_{50} dose-response plot to determine whether galactose is an inhibitor of β-galactosidase for the hydrolysis of 100 µM PNP-gal by 0.25 U/mL β-galactosidase.

The assay in the presence of 10 µM PNP-gal was not likely to show any sensible trends as the values are so small and are likely to be unreliable as a result with magnification of errors. Figure 24 has two outliers for the results in the presence of 1mM and 5mM galactose (points 3 and 3.7 on the log scale), which when omitted leave a reversed sigmoid shape. The middle of the vertical portion of the curve can be approximated from the scale to be a log [galactose] of 4.4. Reversing the log value of this produces a value of 25.12 mM. This is also corroborated by

the curve in figure 3.25. Again it has an outlier for the value at 4.7 (50 mM galactose), but when omitted produces the correctly shaped curve. Taking the middle of the vertical portion of the curve leads to the same log value of 4.4 which is equivalent to 25.12 mM. This provides an IC_{50} value of 25.12 mM which can be linked to the K_i by the Cheng-Prusoff equation:

$$K_i = \frac{IC_{50}}{1 + \frac{[s]}{K_M}}$$

In order to convert the IC50 to a K_i, the K_M for the enzyme is required. This style of assay also allows the results to be interpreted as a standard enzyme kinetic assay, which will allow calculation of the K_M and other kinetic parameters. To do this time versus product concentration graphs were plotted as previously but for increasing substrate concentrations at a fixed inhibitor concentration. This produced a series of 16 graphs, which all showed the expected positive correlation between increased substrate concentration and increased product concentration at each point in time (supplementary data).

The gradient of each slope, which was previously obtained (table 3.8), and is equivalent to the rate were used to produce a series of Hanes-Woolf plots. Plotting these graphs showed unexpected results – the graphs were not always linear, but also did not follow the pattern of the previous apparently non-linear Hanes-Woolf graphs. The graphs did not follow a consistent pattern either. A number of the graphs were essentially linear, but with a negative slope, whilst others followed a shape more closely linked to a hyperbola or elliptical curve. A small number of graphs had a U shape (supplementary data). The odd results must be a result of the small substrate concentrations being measured, as seen previously with the magnified errors affecting the graph shape. As a result of the odd shapes for these graphs, the kinetic parameters of the enzyme cannot be calculated and cannot be used in the Cheng-Prusoff equation to convert the IC_{50} into a K_i.

In spite of the inability to gain meaningful data to convert the IC_{50} to a K_i, the value of 25 mM for the IC_{50} of galactose is not unreasonable. A literature value of 24 mM has been recorded for the Ki of galactose against a native E. coli β-galactosidase[4]. The K_i and IC_{50} are sufficiently linked and likely to be close. Though the kinetic parameters weren't able to be determined, it does show that the assay had the potential to produce meaningful results.

This newer, more complicated, assay was adapted for testing fluorogalactose as an inhibitor, by adjusting the inhibitor concentrations to suit those of the fluorogalactose. The absorbance of the yellow product, 4-nitrophenol, from the hydrolysis of PNP-gal by β-galactosidase was measured over time. As for the equivalent galactose assay, PNP-gal concentrations of 10, 25, 50 and 100 μM were used with a range of fluorogalactose concentrations from 0- 30 μM. Each absorbance was converted to a concentration using the Beer-Lambert Law and the concentrations plotted on a graph against time.

As expected there was a positive correlation between concentration of 4-nitrophenol and time. There was no evidence of a decrease in the overall amount of product as the fluorogalactose concentration increased, which would have suggested inhibition. As with previous assays the results were not very clean and the correlation was not linear for all results (supplementary data). For the hydrolysis of 10 μM PNP-gal, the results had a wide spread and minor increase in concentration of the product, which is to be expected given the very low concentration of the substrate. There is a jump in all the data at around 400 seconds. As this seems to be relatively consistent across all data series, it can be reasonably expected that this is an artefact of the assay. The results for the hydrolysis of 25 μM PNP-gal were cleaner with less spread in the data, and a greater change of overall concentration recorded; as would be reasonably expected from the increased substrate concentration. There is no evidence of inhibition taking place. The change in product concentration is not linear at the highest concentrations of fluorogalactose. Though less noticeable, there is very limited evidence of the same jerk in data at around 400 seconds. This corroborates with the assay at lower substrate concentration and suggests that some fault occurred with the spectrophotometer at that point. As the jerk is less noticeable it also adds more evidence to the argument that at very low concentration errors are magnified and the spectrophotometer is below reliable recording range.

The results for the hydrolysis of 50 μM PNP-gal are much less cleaner than the previous assay. There are a number of non-linear results and very jerky data. The results of the hydrolysis in the presence of 20, 25 and 30 μM fluorogalactose are obviously not linear. The results of the hydrolysis of 100 μM PNP-gal are again much cleaner and less jerky. The formation of products for the hydrolysis in the presence of 15, 20, 25 and 30 μM fluorogalactose is not linear. It is obvious from the non-linear result of this hydrolysis that the fluorogalactose is having some effect, but it is not causing inhibition in the expected manner, which would be to cause the production of less product overall for the same point in time compared to no fluorogalactose being present. It also does not appear to be the result of the substrate being depleted as it occurs more noticeably at the higher substrate concentrations. It is possible that it is the structure of the fluorogalactose itself which leads to the change in pattern.

As with previous assays, the gradient of each slope, which is equivalent to the rate, was calculated (table 3.9). These were used to plot graphs of log [fluorogalactose] against rate to assess whether inhibition was occurring (supplementary data). The gradients for the curved lines are inaccurate but have been used as there is not an obvious linear portion to measure. The graphs did not show inhibition. There was no sigmoid curve. The general overall shape of the pattern was the same across all four graphs.

Fluorogalactose (μM)	**Rate – 10 μM PNP-gal (mmol $dm^{-3}.s^{-1}$)**	**Rate – 25 μM PNP-gal (mmol $dm^{-3}.s^{-1}$)**	**Rate – 50 μM PNP-gal (mmol $dm^{-3}.s^{-1}$)**	**Rate – 100 μM PNP-gal (mmol $dm^{-3}.s^{-1}$)**
0	1.79×10^{-7}	5.13×10^{-7}	8.49×10^{-7}	1.76×10^{-6}
0.01	1.43×10^{-7}	4.00×10^{-7}	7.55×10^{-7}	1.51×10^{-6}
0.025	1.54×10^{-7}	4.21×10^{-7}	7.77×10^{-7}	1.71×10^{-6}
0.05	1.69×10^{-7}	4.37×10^{-7}	7.60×10^{-7}	1.77×10^{-6}
0.1	1.76×10^{-7}	4.47×10^{-7}	8.17×10^{-7}	1.85×10^{-6}
0.5	1.58×10^{-7}	4.17×10^{-7}	8.07×10^{-7}	1.95×10^{-6}
0.7	1.72×10^{-7}	4.65×10^{-7}	9.38×10^{-7}	2.07×10^{-6}
1	1.90×10^{-7}	5.05×10^{-7}	9.07×10^{-7}	2.06×10^{-6}
1.5	2.32×10^{-7}	4.95×10^{-7}	8.99×10^{-7}	2.05×10^{-6}
2	1.89×10^{-7}	4.83×10^{-7}	1.21×10^{-6}	2.02×10^{-6}
5	1.90×10^{-7}	5.33×10^{-7}	9.76×10^{-7}	2.16×10^{-6}
10	2.31×10^{-7}	5.70×10^{-7}	1.08×10^{-6}	2.18×10^{-6}
15	2.70×10^{-7}	6.55×10^{-7}	1.14×10^{-6}	2.07×10^{-6}
20	2.26×10^{-7}	6.15×10^{-7}	1.25×10^{-6}	2.01×10^{-6}
25	2.36×10^{-7}	6.60×10^{-7}	1.29×10^{-6}	2.08×10^{-6}
30	2.57×10^{-7}	6.31×10^{-7}	1.09×10^{-6}	1.54×10^{-6}

Table 3.9: rates of the hydrolysis of different concentrations of PNP-gal by β-galactosidase in the presence of increasing concentrations of fluorogalactose.

The results of the assay were also analysed according to a normal kinetic assay by measuring the change in concentration of the product produced from the hydrolysis of four different substrate concentrations in the presence of a fixed inhibitor concentration. This produced 16 graphs all consistently showing a positive correlation between increasing substrate concentration and increasing product concentration over time. The graphs also highlighted more clearly some of the previously mentioned unusual results, such as non-linear results, outlier results and jerky data. It also highlighted other data series which had odd results (supplementary data). The previously calculated gradients of each data series were used to plot Hanes-Woolf graphs of substrate versus [s]/v.

Similarly to the previous assay, the results did not follow the normal expected pattern of Hanes-Woolf plots, again displaying a variety of curves (supplementary data). This is likely to be due to the very low substrate concentrations used for the assay. Also the assay was carried out using an insufficient number of substrate concentrations. This is highlighted by plotting a standard

enzyme kinetic graph of substrate concentration versus rate, which produces a straight line rather than the characteristic curve (figure 3.26).

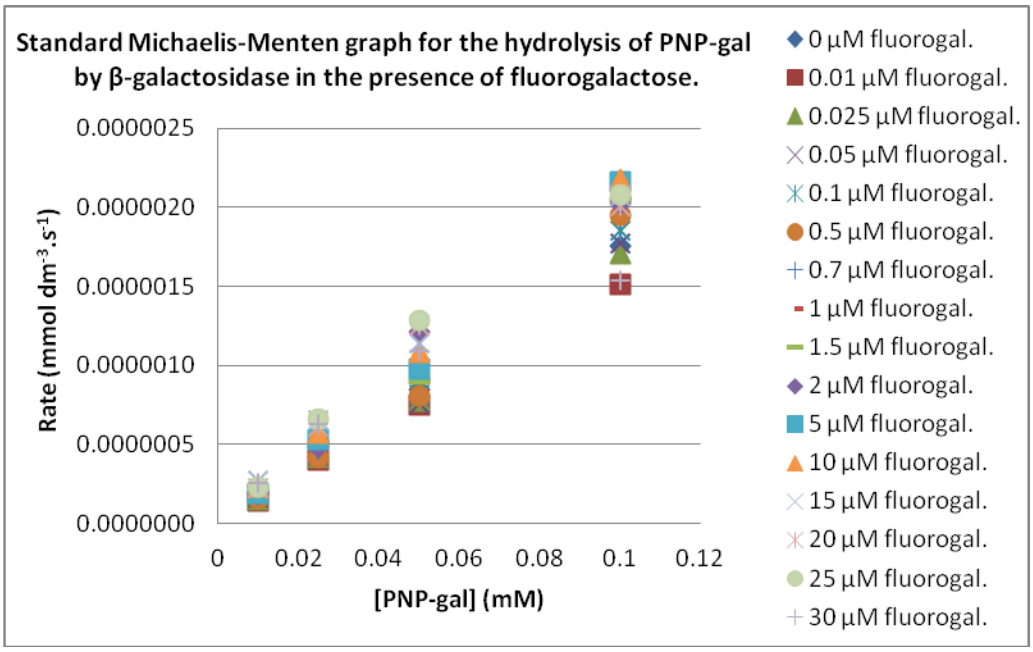


Figure 3.26: Michaelis-Menten graph demonstrating that there was an insufficient number of substrate concentrations used, which means that a curve is not produced.

It was not possible to proceed further with the assays as the sample of fluorogalactose had been used, and it was not possible to obtain further sample from the Linclau group. Therefore the assays were left in an unfinished state with no clear picture of whether inhibition had occurred or not.

3.2.5 Issues with the assays and future directions

There were a number of faults with the design of all of these assays:

- The concentrations of substrates and enzyme used in the assay were too small. This led to very small changes in absorbance and therefore overall concentrations. Had the assay used larger concentrations of enzyme and substrate, the results would have been more reliable. There would have been much more measurable changes in absorbance and concentration and the errors of calculation would not have been magnified as greatly.
- The enzyme was not measured by concentration. The enzyme came supplied as U/mg which was converted to U/mL when the solution was prepared. The enzyme should have been converted to a proper concentration. This would have allowed all kinetic parameters to have been correctly calculated as currently kcat and kcat/Km cannot be calculated without a definitive value for [E]. It would also have highlighted to use more sensible concentration values for the enzyme.

- The earlier assays were not designed well and they could only measure the IC50 rather than the more accurate measure of inhibition, the ki. The final two assays were at least able, in theory, to measure both sets of parameters.
- No assay was carried out in triplicate, which would have allowed for the standard deviation and standard error to be calculated. Whilst some assays were at least carried out in duplicate others were only carried out once and therefore no errors could be calculated for any results.
- The range of inhibitor concentrations used were either too small or did not have sufficient spread. This is highlighted particularly when using the log scale as many of the numbers were very closely spaced. The concentrations also did not go up to a high enough values in many cases, which was why numbers and patterns could not be determined for many of the assays.

Based on these criticisms, if the work was to be carried out in the future, all assays would need to be carried out in triplicate and the design of the assays would need to be changed. The final two assays had the potential to measure meaningful kinetic data but couldn't in practice. By using that same style of assay but using much greater enzyme, substrate and inhibitor concentrations, it is likely that a more definite pattern would have been observed.

3.3 Experimental

3.3.1 General experimental

Tris buffer, galactose and β-galactosidase were purchased from Fisher Scientific. PNP-gal was purchased from Sigma-Aldrich. 2,2,3,3,-Tetrafluorogalactose was synthesised by R. Timofte in the School of Chemistry at the University of Southampton and was given freely for this work.

3.3.2 Assay Methods

Method 1

Solutions of β-galactosidase were prepared in Tris buffer at 0.94 U/mL, 1.88 U/mL, 3.75 U/mL, 6.25 U/mL, 12.5 U/mL, 25 U/mL and 50 U/mL which is 100 times the concentration of the solutions in the assay. A Solution of PNP-gal in Tris buffer at 50 μM was also prepared. PNP-gal solution (200 μL) was aliquoted into a 96 well plate as described. The plate was incubated until it reached 37°C and the β-galactosidase solutions (2 μL) added as described. The absorbance of the product was measured at 405 nm for 150 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2										
B1	B2										
C1	C2										
D1	D2										
E1	E2										
F1	F2										
G1	G2										
H1	H2										

Final concentrations in each well were 50 μM PNP-gal plus:

A1, A2: 0 U/mL β-galactosidase **B1, B2**: 0.0094 U/mL β-galactosidase

C1, C2: 0.0188 U/mL β-galactosidase **D1, D2**:0.0375 U/mL β-galactosidase

E1, E2: 0.0625 U/mL β-galactosidase **F1, F2**: 0.125 U/mL β-galactosidase

G1, G2: 0.25 U/mL β-galactosidase **H1, H2**: 0.50 U/mL β-galactosidase

Method 2

Solutions of β-galactosidase in Tris buffer were prepared at 0.94 U/mL, 1.88 U/mL, 3.75 U/mL, 6.25 U/mL, 12.5 U/mL, 25 U/mL and 50 U/mL which is 100 times the concentration of the solutions in the assay. Solutions of PNP-gal in Tris buffer at 50 μM and 100 μM were also prepared. PNP-gal solutions (200 μL) were aliquoted into a 96 well plate as decribed. The plate was incubated until it reached 37°C and the β-galactosidase solutions (2 μL) added as

described. The absorbance of the product was measured at 405 nm for 240 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2	A3	A4								
B1	B2	B3	B4								
C1	C2	C3	C4								
D1	D2	D3	D4								
E1	E2	E3	E4								
F1	F2	F3	F4								
G1	G2	G3	G4								
H1	H2	H3	H4								

Blue coloured cells contain 50 µM PNP-gal, red coloured cells contain 100 µM PNP-gal in addition to the following all at final concentration:

A1, A2, A3, A4: 0 U/mL β-galactosidase

B1, B2, B3, B4: 0.0094 U/mL β-galactosidase

C1, C2, C3, C4: 0.0188 U/mL β-galactosidase

D1, D2, D3, D4: 0.0375 U/mL β-galactosidase

E1, E2, E3, E4: 0.0625 U/mL β-galactosidase

F1, F2, F3, F4: 0.125 U/mL β-galactosidase

G1, G2, G3, G4: 0.25 U/mL β-galactosidase

H1, H2, H3, H4: 0.50 U/mL β-galactosidase

Method 3

A solution of PNP-gal in Tris buffer at 50 µM was prepared along with solutions of β-galactosidase in Tris buffer at 0.25 and 0.5 U/mL. Solutions of galactose were prepared in distilled water at concentrations 1 mM, 10 mM, 50 mM, 100 mM, 500 mM, 1 M and 2 M. The β-galactosidase and galactose concentrations are 100 times the final assay concentration. PNP-gal solution (200 µL) was aliquoted into wells as shown along with solutions of galactose (2 µL). The plate was incubated until it reached 37°C and the β-galactosidase solutions (2 µL) added as described. The absorbance of the product was measured at 405 nm for 240 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2	A3	A4								
B1	B2	B3	B4								
C1	C2	C3	C4								
D1	D2	D3	D4								
E1	E2	E3	E4								
F1	F2	F3	F4								
G1	G2	G3	G4								
H1	H2	H3	H4								

All wells contain a final concentration of 50 μM PNP-gal along with 0.25 U/mL β-galactosidase in blue coloured wells or 0.5 U/mL β-galactosidase in red wells in addition to:

A1, A2: 0.25 U/mL β-galactosidase, 0 mM galactose

A3, A4: 0.50 U/mL β-galactosidase, 0 mM galactose

B1, B2: 0.25 U/mL β-galactosidase, 0.01 mM galactose

B3, B4: 0.50 U/mL β-galactosidase, 0.01 mM galactose

C1, C2: 0.25 U/mL β-galactosidase, 0.1 mM galactose

C3, C4: 0.50 U/mL β-galactosidase, 0.1 mM galactose

D1, D2: 0.25 U/mL β-galactosidase, 0.5 mM galactose

D3, D4: 0.50 U/mL β-galactosidase, 0.5 mM galactose

E1, E2: 0.25 U/mL β-galactosidase, 1 mM galactose

E3, E4: 0.50 U/mL β-galactosidase, 1 mM galactose

F1, F2: 0.25 U/mL β-galactosidase, 5 mM galactose

F3, F4: 0.50 U/mL β-galactosidase, 5 mM galactose

G1, G2: 0.25 U/mL β-galactosidase, 10 mM galactose

G3, G4: 0.50 U/mL β-galactosidase, 10 mM galactose

H1, H2: 0.25 U/mL β-galactosidase, 20 mM galactose

H3, H4: 0.50 U/mL β-galactosidase, 20 mM galactose

Method 4

A solution of PNP-gal in Tris buffer at 50 μM was prepared along with solutions of β-galactosidase in Tris buffer at 0.25 and 0.5 U/mL. Solutions of 2,2,3,3-tetrafluorogalactose were prepared in Tris buffer at concentrations 1 mM, 2.5 mM, 5 mM, 10 mM, 50 mM, 100 mM and 200 mM. The β-galactosidase and 2,2,3,3-tetrafluorogalactose concentrations are 100 times the final assay concentration. PNP-gal solution (200 μL) was aliquoted into wells as shown along with solutions of 2,2,3,3-tetrafluorogalactose (2 μL). The plate was incubated until it reached 37°C and the β-galactosidase solutions (2 μL) added as described. The absorbance of the product was measured at 405 nm for 240 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2	A3	A4								
B1	B2	B3	B4								
C1	C2	C3	C4								
D1	D2	D3	D4								
E1	E2	E3	E4								
F1	F2	F3	F4								
G1	G2	G3	G4								
H1	H2	H3	H4								

A1, A2: 50 μM PNP-gal, 25 U/mL β-galactosidase, 0 mM 2,2,3,3-tetrafluorogalactose
A3, A4: 50 μM PNP-gal, 50 U/mL β-galactosidase, 0 mM 2,2,3,3-tetrafluorogalactose
B1, B2: 50 μM PNP-gal, 25 U/mL β-galactosidase, 1 mM 2,2,3,3-tetrafluorogalactose
B3, B4: 50 μM PNP-gal, 50 U/mL β-galactosidase, 1 mM 2,2,3,3-tetrafluorogalactose
C1, C2: 50 μM PNP-gal, 25 U/mL β-galactosidase, 2.5 mM 2,2,3,3-tetrafluorogalactose
C3, C4: 50 μM PNP-gal, 50 U/mL β-galactosidase, 2.5 mM 2,2,3,3-tetrafluorogalactose
D1, D2: 50 μM PNP-gal, 25 U/mL β-galactosidase, 5 mM 2,2,3,3-tetrafluorogalactose
D3, D4: 50 μM PNP-gal, 50 U/mL β-galactosidase, 5 mM 2,2,3,3-tetrafluorogalactose
E1, E2: 50 μM PNP-gal, 25 U/mL β-galactosidase, 10 mM 2,2,3,3-tetrafluorogalactose
E3, E4: 50 μM PNP-gal, 50 U/mL β-galactosidase, 10 mM 2,2,3,3-tetrafluorogalactose
F1, F2: 50 μM PNP-gal, 25 U/mL β-galactosidase, 50 mM 2,2,3,3-tetrafluorogalactose
F3, F4: 50 μM PNP-gal, 50 U/mL β-galactosidase, 50 mM 2,2,3,3-tetrafluorogalactose
G1, G2: 50 μM PNP-gal, 25 U/mL β-galactosidase, 100 mM 2,2,3,3-tetrafluorogalactose
G3, G4: 50 μM PNP-gal, 50 U/mL β-galactosidase, 100 mM 2,2,3,3-tetrafluorogalactose
H1, H2: 50 μM PNP-gal, 25 U/mL β-galactosidase, 200 mM 2,2,3,3-tetrafluorogalactose
H3, H4: 50 μM PNP-gal, 50 U/mL β-galactosidase, 200 mM 2,2,3,3-tetrafluorogalactose

Method 5

A solution of PNP-gal in Tris buffer at 50 μM was prepared along with a solution of β-galactosidase in Tris buffer at 0.25 U/mL. Solutions of 2,2,3,3-tetrafluorogalactose were prepared in Tris buffer at concentrations 1 mM, 2.5 mM, 5 mM, 10 mM, 30 mM, 50 mM, 70 mM, 100 mM, 150 mM, 200 mM and 268 mM. Solutions of galactose in distilled water were also prepared to concentrations of 1 mM, 10 mM, 50 mM, 100 mM, 300 mM, 500 mM, 800 mM, 1 M, 1.4 M, 1.7 M and 2 M. The galactose, β-galactosidase and 2,2,3,3-tetrafluorogalactose concentrations are 100 times the final assay concentration. PNP-gal solution (200 μL) was aliquoted into wells as shown along with solutions of 2,2,3,3-tetrafluorogalactose (2 μL) or galactose. The plate was incubated until it reached 37°C and the β-galactosidase solutions (2 μL) added as described. The absorbance of the product was measured at 405 nm for 240 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12

All alpha-numeric numbered wells contain 50 μM PNP-gal and 0.25 U/mL β-galactosidase. Additionally

A1, B1: 0 mM galactose **A2, B2**: 1 mM galactose **A3, B3**: 10 mM galactose **A4, B4**: 50 mM galactose **A5, B5**: 100 mM galactose **A6, B6**: 300 mM galactose **A7, B7**: 500 mM galactose **A8, B8**: 800 mM galactose **A9, B9**: 1 M galactose **A10, B10**: 1.4 M galactose **A11, B11**: 1.7 M galactose **A12, B12**: 2 M galactose

C1, D1: 0 mM fluorogalactose **C2, D2**: 1 mM fluorogalactose

C3, D3: 2.5 mM fluorogalactose **C4, D4**: 5 mM fluorogalactose

C5, D5: 10 mM fluorogalactose **C6, D6**: 30 mM fluorogalactose

C7, D7: 50 mM fluorogalactose **C8, D8**: 70 mM fluorogalactose

C9, D9: 100 mM fluorogalactose **C10, D10**: 150 mM fluorogalactose

C11, D11: 200 mM fluorogalactose **C12, D12**: 268 mM fluorogalactose

Method 6

A solution of PNP-gal in Tris buffer at 50 μM was prepared along with a solution of β-galactosidase in Tris buffer at 0.25 U/mL. Solutions of 2,2,3,3-tetrafluorogalactose were prepared in Tris buffer at concentrations 1 mM, 2.5 mM, 5 mM, 10 mM, 30 mM, 50 mM, 70 mM, 100 mM, 150 mM, 200 mM and 268 mM. Solutions of galactose in distilled water were also prepared to concentrations of 1 mM, 10 mM, 50 mM, 100 mM, 300 mM, 500 mM, 800 mM, 1 M, 1.4 M, 1.7 M and 2 M. The galactose, β-galactosidase and 2,2,3,3-tetrafluorogalactose concentrations are 100 times the final assay concentration. β-galactosidase solutions (2 μL) was aliquoted into wells as shown along with solutions of 2,2,3,3-tetrafluorogalactose (2 μL) or galactose. The plate was incubated until it reached 37°C and the PNP-gal solution (200 μL) added to each well. The absorbance of the product was measured at 405 nm for 240 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12

All alpha-numeric numbered wells contain 50 μM PNP-gal and 0.25 U/mL β-galactosidase. Additionally

A1, B1: 0 mM galactose **A2, B2**: 1 mM galactose **A3, B3**: 10 mM galactose **A4, B4**: 50 mM galactose **A5, B5**: 100 mM galactose **A6, B6**: 300 mM galactose **A7, B7**: 500 mM galactose **A8, B8**: 800 mM galactose **A9, B9**: 1 M galactose **A10, B10**: 1.4 M galactose **A11, B11**: 1.7 M galactose **A12, B12**: 2 M galactose

C1, D1: 0 mM fluorogalactose **C2, D2**: 1 mM fluorogalactose

C3, D3: 2.5 mM fluorogalactose **C4, D4**: 5 mM fluorogalactose

C5, D5: 10 mM fluorogalactose **C6, D6**: 30 mM fluorogalactose

C7, D7: 50 mM fluorogalactose **C8, D8**: 70 mM fluorogalactose

C9, D9: 100 mM fluorogalactose **C10, D10**: 150 mM fluorogalactose

C11, D11: 200 mM fluorogalactose **C12, D12**: 268 mM fluorogalactose

Method 7

A solution of PNP-gal in Tris buffer at 65 μM was prepared along with a solution of β-galactosidase in Tris buffer at 4 U/mL. Solutions of 2,2,3,3-tetrafluorogalactose were prepared in Tris buffer at concentrations 80 μM, 200 μM, 400 μM, 800 μM, 4 mM, 5.6 mM, 8 mM, 12 mM, 16 mM, 40 mM and 80 mM. Solutions of galactose in distilled water were also prepared to concentrations of 80 μM, 800 μM, 4 mM, 8 mM, 24 mM, 40 mM, 64 mM, 80 mM, 112 mM, 136 mM, and 160 mM. The galactose, β-galactosidase and 2,2,3,3-tetrafluorogalactose concentrations are 8 times the final assay concentration. β-galactosidase solutions (25 μL) was aliquoted into wells as shown along with solutions of 2,2,3,3-tetrafluorogalactose (25 μL) or galactose (25 μL). The plate was incubated until it reached 37°C and the PNP-gal solution (150 μL) added to each well. The absorbance of the product was measured at 405 nm for 240 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12

All alpha-numeric numbered wells contain final concentration 50 μM PNP-gal and 0.25 U/mL β-galactosidase. Additionally

A1, B1: 0 mM galactose **A2, B2**: 10 μM galactose **A3, B3**: 100 μM galactose**A4, B4**: 500 μM galactose **A5, B5**: 1 mM galactose **A6, B6**: 3 mM galactose

A7, B7: 5 mM galactose **A8, B8**: 8 mM galactose **A9, B9**: 10 mM galactose

A10, B10: 14 mM galactose **A11, B11**: 17 mM galactose **A12, B12**: 20 mM galactose

C1, D1: 0 mM fluorogalactose **C2, D2**: 10 μM fluorogalactose

C3, D3: 25 μM fluorogalactose **C4, D4**: 50 μM fluorogalactose

C5, D5: 100 μM fluorogalactose **C6, D6**: 500 μM fluorogalactose

C7, D7: 700 μM fluorogalactose **C8, D8**: 1 mM fluorogalactose

C9, D9: 1.5 mM fluorogalactose **C10, D10**: 2 mM fluorogalactose

C11, D11: 5 mM fluorogalactose **C12, D12**: 10 mM fluorogalactose

Method 8

Solutions of PNP-gal were prepared in Tris-HCl buffer at concentrations of 0.3 μM, 0.65 μM, 1.3 μM, 13 μM, 32.5 μM, 65 μM, 130 μM, 325 μM, 650 μM, 975 μM, 1.3 mM and 13 mM. A solution of β-galactosidase of 1 U/mL was also prepared. The solutions of PNP-gal were added to a 96 well plate as shown in the table and the reactions initiated with the addition of β-galactosidase. The absorbance of the product was measured at 405 nm for 240 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12

All wells contain final concentration as listed:

A1, B1: 0.1 µM PNP-gal, 0.25 µM β-galactosidase

A2, B2: 0.5 µM PNP-gal, 0.25 µM β-galactosidase

A3, B3: 1 µM PNP-gal, 0.25 µM β-galactosidase

A4, B4: 10 µM PNP-gal, 0.25 µM β-galactosidase

A5, B5: 25 µM PNP-gal, 0.25 µM β-galactosidase

A6, B6: 50 µM PNP-gal, 0.25 µM β-galactosidase

A7, B7: 100 µM PNP-gal, 0.25 µM β-galactosidase

A8, B8: 250 µM PNP-gal, 0.25 µM β-galactosidase

A9, B9: 500 µM PNP-gal, 0.25 µM β-galactosidase

A10, B10: 750 µM PNP-gal, 0.25 µM β-galactosidase

A11, B11: 1 mM PNP-gal, 0.25 µM β-galactosidase

A12, B12: 10 mM PNP-gal, 0.25 µM β-galactosidase

Method 9

Solutions of PNP-gal were prepared in Tris-HCl buffer at concentrations of 0.3 µM, 0.65 µM, 1.3 µM, 13 µM, 32.5 µM, 65 µM, 130 µM, 325 µM, 650 µM, 975 µM, 1.3 mM, 2.6 mM, 5.2 mM, 7.8 mM, 10.4 mM and 13 mM. A solution of β-galactosidase of 1 U/mL was also prepared. The solutions of PNP-gal were added to a 96 well plate as shown in the table and the reactions initiated with the addition of β-galactosidase. The absorbance of the product was measured at 405 nm for 240 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
A13	A14	A15	A16	-	-	-	-	-	-	-	-
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
B13	B14	B15	B16	-	-	-	-	-	-	-	-

All wells contain final concentration as listed:

A1, B1: 0.1 μM PNP-gal, 0.25 μM β-galactosidase

A2, B2: 0.5 μM PNP-gal, 0.25 μM β-galactosidase

A3, B3: 1 μM PNP-gal, 0.25 μM β-galactosidase

A4, B4: 10 μM PNP-gal, 0.25 μM β-galactosidase

A5, B5: 25 μM PNP-gal, 0.25 μM β-galactosidase

A6, B6: 50 μM PNP-gal, 0.25 μM β-galactosidase

A7, B7: 100 μM PNP-gal, 0.25 μM β-galactosidase

A8, B8: 250 μM PNP-gal, 0.25 μM β-galactosidase

A9, B9: 500 μM PNP-gal, 0.25 μM β-galactosidase

A10, B10: 750 μM PNP-gal, 0.25 μM β-galactosidase

A11, B11: 1 mM PNP-gal, 0.25 μM β-galactosidase

A12, B12: 2 mM PNP-gal, 0.25 μM β-galactosidase

A13, B13: 4 mM PNP-gal, 0.25 μM β-galactosidase

A14, B14: 6 mM PNP-gal, 0.25 μM β-galactosidase

A15, B15: 8 mM PNP-gal, 0.25 μM β-galactosidase

A16, B16: 10 mM PNP-gal, 0.25 μM β-galactosidase

-: water blank.

Method 10

A solution of PNP-gal in buffer (50 mM Tris-HCl, 10% glycerol) of 65 μM was prepared along with a solution of β-galactosidase in Tris/glycerol buffer at 2 U/mL. Solutions of 2,2,3,3-tetrafluorogalactose were prepared in Tris buffer at concentrations 80 μM, 200 μM, 400 μM, 800 μM, 4 mM, 5.6 mM, 8 mM, 12 mM, 16 mM, 40 mM, 80 mM, 120 mM, 160 mM, 200 mM and 240 mM. Solutions of galactose in distilled water were also prepared to concentrations of 80 μM, 800 μM, 4 mM, 8 mM, 24 mM, 40 mM, 64 mM, 80 mM, 112 mM, 136 mM, 160 mM, 240 mM, 320 mM, 400 mM and 800 mM. The galactose, β-galactosidase and 2,2,3,3-tetrafluorogalactose concentrations are 8 times the final assay concentration. β-galactosidase solutions (25 μL) was aliquoted into wells as shown along with solutions of 2,2,3,3-tetrafluorogalactose (25 μL) or galactose (25 μL). The plate was incubated until it reached 37°C

and the PNP-gal solution (150 µL) added to each well. The absorbance of the product was measured at 405 nm for 240 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2	A3	A4								
B1	B2	B3	B4								
C1	C2	C3	C4								
D1	D2	D3	D4								
E1	E2	E3	E4								
F1	F2	F3	F4								
G1	G2	G3	G4								
H1	H2	H3	H4								

All alpha-numeric numbered wells contain final concentration 50 µM PNP-gal and 0.25 U/mL β-galactosidase. Additionally

A1: 0 mM galactose
B1: 10 µM galactose
C1: 100 µM galactose
D1: 500 µM galactose
E1: 1 mM galactose
F1: 3 mM galactose
G1: 5 mM galactose
H1: 8 mM galactose
A2: 10 mM galactose
B2: 14 mM galactose
C2: 17 mM galactose
D2: 20 mM galactose
E2: 30 mM galactose
F2: 40 mM galactose
G2: 50 mM galactose
H2: 100 mM galactose
A3: 0 mM fluorogalactose
B3: 10 µM fluorogalactose
C3: 25 µM fluorogalactose
D3: 50 µM fluorogalactose
E3: 100 µM fluorogalactose
F3: 500 µM fluorogalactose
G3: 700 µM fluorogalactose
H3: 1 mM fluorogalactose
A4: 1.5 mM fluorogalactose
B4: 2 mM fluorogalactose
C4: 5 mM fluorogalactose
D4: 10 mM fluorogalactose
E4: 15 mM fluorogalactose
F4: 20 mM fluorogalactose
G4: 25 mM fluorogalactose
H4: 30 mM fluorogalactose

Method 11

Solutions of PNP-gal in buffer (50 mM Tris-HCl, 10% glycerol) of 13 µM, 32.5 µM, 65 µM and 130 µM were prepared along with a solution of β-galactosidase in Tris/glycerol buffer at 2 U/mL. Solutions of galactose in distilled water were also prepared to concentrations of 80 µM, 800 µM, 4 mM, 8 mM, 24 mM, 40 mM, 64 mM, 80 mM, 112 mM, 136 mM, 160 mM, 240 mM, 320 mM, 400 mM and 800 mM. The galactose and β-galactosidase concentrations are 8 times the final assay concentration. PNP-gal solution (150 µL) was aliquoted into wells as shown along with solutions of galactose (25 µL). The plate was incubated until it reached 37°C and the β-

galactosidase solution (25 µL) added to each well. The absorbance of the product was measured at 405 nm for 240 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2	A3	A4	A5	A6	A7	A8				
B1	B2	B3	B4	B5	B6	B7	B8				
C1	C2	C3	C4	C5	C6	C7	C8				
D1	D2	D3	D4	D5	D6	D7	D8				
E1	E2	E3	E4	E5	E6	E7	E8				
F1	F2	F3	F4	F5	F6	F7	F8				
G1	G2	G3	G4	G5	G6	G7	G8				
H1	H2	H3	H4	H5	H6	H7	H8				

All alpha-numeric numbered wells contain final concentration 0.25 U/mL β-galactosidase. Cells in red contain final concentration 10 µM PNP-gal, in blue 25 µM PNP-gal, in green 50 µM PNP-gal, and pink 100 µM PNP-gal. Additionally

A1, A3, A5, A7: 0 mM galactose
B1, B3, B5, B7: 10 µM galactose
C1, C3, C5, C7: 100 µM galactose
D1, D3, D5, D7: 500 µM galactose
E1, E3, E5, E7: 1 mM galactose
F1, F3, F5, F7: 3 mM galactose
G1, G3, G5, G7: 5 mM galactose
H1, H3, H5, H7: 8 mM galactose
A2, A4, A6, A8: 10 mM galactose
B2, B4, B6, B8: 14 mM galactose
C2, C4, C6, C8: 17 mM galactose
D2, D4, D6, D8: 20 mM galactose
E2, E4, E6, E8: 30 mM galactose
F2, F4, F6, F8: 40 mM galactose
G2, G4, G6, G8: 50 mM galactose
H2, H4, H6, H8: 100 mM galactose

Method 12

Solutions of PNP-gal in buffer (50 mM Tris-HCl, 10% glycerol) of 13 µM, 32.5 µM, 65 µM and 130 µM were prepared along with a solution of β-galactosidase in Tris/glycerol buffer at 2 U/mL. Solutions of 2,2,3,3-tetrafluorogalactose were prepared in Tris buffer at concentrations 80 µM, 200 µM, 400 µM, 800 µM, 4 mM, 5.6 mM, 8 mM, 12 mM, 16 mM, 40 mM, 80 mM, 120 mM, 160 mM, 200 mM and 240 mM. The fluorogalactose and β-galactosidase concentrations are 8 times the final assay concentration. PNP-gal solution (150 µL) was aliquoted into wells as shown along with solutions of fluorogalactose (25 µL). The plate was incubated until it reached 37°C and the β-galactosidase solution (25 µL) added to each well. The absorbance of the product was measured at 405 nm for 240 cycles at a temperature of 37°C.

Plate was organised as shown:

A1	A2	A3	A4	A5	A6	A7	A8				
B1	B2	B3	B4	B5	B6	B7	B8				
C1	C2	C3	C4	C5	C6	C7	C8				
D1	D2	D3	D4	D5	D6	D7	D8				
E1	E2	E3	E4	E5	E6	E7	E8				
F1	F2	F3	F4	F5	F6	F7	F8				
G1	G2	G3	G4	G5	G6	G7	G8				
H1	H2	H3	H4	H5	H6	H7	H8				

All alpha-numeric numbered wells contain final concentration 0.25 U/mL β-galactosidase. Cells in red contain final concentration 10 μM PNP-gal, in blue 25 μM PNP-gal, in green 50 μM PNP-gal, and pink 100 μM PNP-gal. Additionally

A1, A3, A5, A7: 0 mM fluorgalactose

B1, B3, B5, B7: 0.01 μM fluorogalactose

C1, C3, C5, C7: 0.025 μM fluorogalactose

D1, D3, D5, D7: 0.05 μM fluorogalactose

E1, E3, E5, E7: 0.1 μM fluorogalactose

F1, F3, F5, F7: 0.5 μM fluorogalactose

G1, G3, G5, G7: 0.7 μM fluorogalactose

H1, H3, H5, H7: 1 μM fluorogalactose

A2, A4, A6, A8: 1.5 μM fluorogalactose

B2, B4, B6, B8: 2 μM fluorogalactose

C2, C4, C6, C8: 5 μM fluorogalactose

D2, D4, D6, D8: 10 μM fluorogalactose

E2, E4, E6, E8: 15 μM fluorogalactose

F2, F4, F6, F8: 20 μM fluorogalactose

G2, G4, G6, G8: 25 μM fluorogalactose

H2, H4, H6, H8: 30 μM fluorogalactose

3.4 References

1. Matthews, B. W.; *The structure of E. coli β-galactosidase*; C. R. Biologies; 2005; 328; 549-556.
2. Jacobson, R. H., Zhang, X-J., DuBose, R. F., and Matthews, B. W.; *Three dimensional structure of β-galactosidase from E. coli*; Nature; 1994; **369**; 761-766.
3. Ioannou, A., Cini, E., Timofte, R. S., Flitsch, S. L., Turner, N. J., and Linclau, B.; *Heavily fluorinated carbons as enzyme substrates:oxidation of tetrafluorinated galactose by galactose oxidase*; Chemical Communications; 2011; **47**; 11228-11230.
4. Lo, S., Dugdale, M. L., Jeerh, N., Ku, T., Roth, N. J., and Huber, R. E.; Protein J.; 2010; **1**; 26-31.

Chapter 4 - LaBLog, the laboratory blog: a new approach to open notebook science through the development and testing of a blog, which records the workings of a molecular biology laboratory.

4.1 Introduction

In this chapter the focus is moved away from the results being produced, but to how those results are recorded and stored. Every experiment carried out by a researcher should be recorded in some manner which is meaningful, allowing another researcher to replicate the method and validate the results. Every researcher has a unique style of recording data reflecting individual personalities, but there is a recognised vocabulary for scientific research which facilitates the work of another to be followed and understood. The standard method for recording experiments in many laboratories worldwide is paper. In this chapter electronic systems including ELNs, blogs and wikis are introduced and the advantages and disadvantages of the systems noted. Additionally the concept of making scientific data available through open notebook science is explored. From this, the development and use of an electronic lab blog as the open, primary research record is discussed.

4.1.1 An introduction to electronic laboratory notebooks (ELNs)

For hundreds of years, paper has been a vitally important commodity. Used for communication and documentation, paper has been the traditional medium on which scientific experiments have been recorded. Paper is portable, robust, flexible, and able to capture writing and pictures[1]. However, paper is not an infallible medium: certain solvents can dissolve ink, paper discolours and inks fade with age. Large amounts of paper are also cumbersome to store, using valuable space.

In the modern laboratory environment, many machines already output data in electronic format. These digital results are often manually copied out, or printed off and pasted into the paper lab book[2]. A system where all the data from different machines can be stored in one place would provide many advantages by lessening the amounts of paper being stored, whilst protecting against damage to records through spillage or discoloration of ink and paper caused by age. Electronic notebooks (ELNs) and digitised systems are a potential answer to this problem.

Electronic notebooks and digitised systems have been in existence since the 1990s. However, there has only been a large increase in the use of Electronic Laboratory Notebooks (ELNs) since 2004[2]. Much of this increase is in commercial laboratories rather than academia, which

has been predicted to have only 4% uptake[3]. Prior to that time, there was limited uptake of such systems[2,4]. This is because the systems that were in existence were found to be difficult to use or conferred no benefits over paper. There were also serious concerns with maintaining Intellectual Property (IP) rights and filing for patents. Many of these concerns were due to misconceptions of the law since electronic records have been admissible in court under US law for 30 years[5]. ELN technology has progressed and now offers digital signatures and other verification devices which have eased these concerns. In many cases digital signatures have increased the credibility of IP claims as the electronic system actually forces signatures to be added more regularly than many paper records were signed.

There are currently a large number of ELN systems available. Researching the features of these ELNs, it is possible to categorise them into three main classes: discipline specific, which have various versions of the program for different types of science; generic, which are designed to be suitable for all types of science; and custom, designed for a specific requirement. Additionally some of these systems rely on security stationery to enhance the security of electronic documentation. Details of the name and main features of each program are summarised in appendix 1.

There are a number of common features to these programs and a number of differences. Most of the available ELNs use relatively prescriptive form and window based entries, whilst a few are based on web forms. The differences are between the capabilities of the systems to completely replace the paper note book; to store and archive data, and to be secure without the need for separate security stationery. The different combinations of capabilities lead to the varying strengths and weaknesses of these programs.

One of the main issues with all of these systems is that they are biased towards one scientific discipline. Where the program is designed to be generic it often has a series of add-ons and plug-ins designed to enhance the specificity for a particular discipline, or convert it to an alternative discipline. This is not ideal in an industry where collaboration between different disciplines and departments is paramount.

4.1.2 ELNs in academia

Whilst ELN systems are being used widely in industry, and many of the major pharmaceutical companies now subscribe to their use, uptake and development has been limited within academia, though there are specific instances of their use[6]. There are two main reasons for this: suitability and cost. Most of the available ELNs are designed for use by the large pharmaceutical companies and are unsuitable for the multi-discipline science carried out in academic labs. Whilst suitability is of importance, the main factor against implementation in academic settings is cost. The costs of deploying an ELN system do not match the amount of

money available to academic researchers. For example at the beginning of 2013, Labtrack[7] has a wide variety of pricing schemes encompassing single user academic licences from $590 to ten user academic full licences at $9350. It also has 'pay as you go' services from $200 per month. In addition there are other costs incurred for software training and installation.

In much of academia, experimental record keeping is carried out on paper, usually in the form of exercise books, notebooks or duplicate books. Results generated in digital format are printed out and glued into the notebook. Where electronic records are kept, these are most likely to be in the form of text documents such as papers and reports, or spreadsheets containing lists of chemicals. Like paper records these records are also likely to be kept by an individual, though some research groups may have communal data storage.

There are few examples of an ELN being developed in an academic environment. The Smart Tea project developed by Southampton University is one of only a few examples to be found of academic ELNs. Smart Tea[1,4,8] is an ELN designed for use in a synthetic and analytical chemistry lab. At the time of its conception, Smart Tea took a different route to the ELNs that had preceded it by being internet based, not stand alone. This allowed the system to be accessed using tablet PCs which, being smaller and lighter than laptops, are more portable and can be moved around the lab. The added capability of a tablet PC to accept input through a stylus increased the flexibility of the system.

A more recent example is that of the University of Houston, who have developed an internet based ELN specifically for use by the students on an undergraduate practical biotechnology course[3].

A different approach was taken by collaboration between several structural biology laboratories across the UK to develop an extended ELN including a Protein Information Management System[9].

4.1.3 Blogs and wikis: an alternative approach to ELNs

An alternative approach to that of ELNs is a blog or wiki. These internet based forms of electronic communication have changed the way information can be stored and shared. Both wikis and blogs have the ability to store and archive data like ELNs, and being internet based are widely accessible.

Wikis: collaborative websites allowing editing by multiple users[10] are becoming a popular means of sharing resources. Wikis have been in existence since 1994, and were developed originally as a tool for quick information sharing between programmers with the concept initially being called the 'wiki wiki web'. Wiki was chosen as it is the Hawaiian word for quick. Out of this original site, there are now hundreds of wiki sites on the internet[11]. Arguably the most famous wiki site to be developed is Wikipedia which is now used as a primary information source online.

Sites such as myExperiment[12] allow protocols to be written up and produce workflows with the aid of software called Taverna. Similarly sites such as OpenWetWare[13] contain multiple lab communities who share protocols, and host blogs and courses. These wikis act as group repositories allowing information to be shared between a small number of researchers in one laboratory or a much larger community.

Wikis can also be used as the primary research notebook. Wikis are easy to set up and use as there are specific wiki hosting services available, such as wikispaces.com[14]. Examples of wikispaces users are the Bradley group at Drexel University, Philadelphia with a wiki called Usefulchem[15], and the research group of Gus Rosania[16], from the University of Michigan.

An alternative to the wiki is a weblog, more frequently referred to as a blog. Blogs are websites or parts of websites which often contain the possibility for readers to comment. They developed from the earlier online communities of bulletin boards and newsgroups with the earliest blogs being recorded in 1994[17]**.** Most blogs in a scientific context are not used as the primary research record but as a place to discuss major news about that scientist's research, or thoughts about general research, news and many other articles of interest. Some blogs are actually used as a notebook, however these blogs still have an element of 'presentation' about them i.e. they still show a summary of the results of work or are used as a place to think or describe the activity of the lab but not the full record of the work.

4.1.4 The advantages and disadvantages of electronic systems

Electronic notebooks have compelling advantages over paper based books. Data can be shared in real time; work can be distributed between teams, who may not be in the same location; research can be commented upon; searching and indexing can be automated, and electronic storage costs little money[16]. Most laboratory instrumentation is now producing electronic output, and many laboratory activities in research and design rely on computer systems. Electronic notebooks have also been found to increase productivity in the lab by decreasing the amount of time spent on administration of paperwork and improve the overall quality of experiment recording.

There are also some disadvantages to electronic systems. As computers and software develop, file types also change. A concern is that electronic records may be unreadable in the future if the accepted and recognised file types change. Another concern is with system availability. Networks are known to have outages and it is a concern that the system would not be useable during the period of the outage. A system that is able to be used even if there is no network available is preferred by many companies[19]. Of particular concern, especially to internet based systems, is the issue of security.

Despite the many differences between the styles and operational set-ups of ELNs they share a common feature with traditional paper records: they have security as a central aim, and control who has access to the data. This is because traditionally, scientific research is kept private until publication in peer reviewed journals, or patents and IP have been secured. There is much unpublished data, and even the data shown in peer reviewed literature is just a selection of the data produced. There is unwillingness in academia, particularly among biological and medical scientists, to publish data as it is produced over concerns of "being scooped". "Being scooped" is a term to describe the theft of data, which allows another group to find an answer and publish ahead of the original scientists working on a particular problem. The fear of failing to be the first publisher is a particular problem in fields such as the biological sciences where publication in peer reviewed journals acts as an indicator for the level of success of a scientist[20].

4.1.5 The growth of Open Notebook Science

Recently there has been a move towards open science by some sections of the scientific community. Open science is a blanket term covering the advancement of open-access publication; the act of publishing and allowing that data to be read and used without charge; open-source development and sharing of software, along with open data. Also under this term is Open Notebook Science (ONS). ONS is the publication of the primary research record, i.e. the research notebook in whatever form it may take, online as the material is generated. The term was first used by Jean-Claude Bradley in September 2006 to highlight the context within open science[20,21]. The motivation for this is to allow "dark data", the data from failed, less significant and unpublished work, to be made public and is the logical extension of transparent science.

ONS can enable scientists to refer to, and draw on the experience of others. Detailed protocols, results and failures can be obtained. This improves communication, increases the speed of research and saves time wasted on repeating failed experiments. ONS does have some instances where its use is limited. This is in research that ultimately will be patented. Pre-publication on the internet is disclosure and voids the possibility of patenting that research. ONS involves disclosure and thus has caused concerns that research reported by ONS will not be suitable for publication in traditional peer-reviewed literature. In most cases these concerns are needless, as journals which accept conference proceedings and pre-prints will usually allow ONS. There are a few journals which will not accept ONS data, but the majority of ONS practitioners will choose to publish in journals open to the idea of ONS[20].

The increase in use of ONS has led to increased use of blogs and wikis as a research record, and in some cases the primary research record. Additionally the importance of social networking sites cannot be underestimated in open notebook science. Social networking sites such as Twitter[22] and Friendfeed[23] are emerging as sites which people are using to compare and discuss scientific data. The previously mentioned site myExperiment[12] also has a social networking capacity. The discussion of scientific data on social networking sites has led to new

concerns for the research being presented at conferences. This data, which may not yet be in the public domain, is finding its way onto the internet in minutes as conference attendees blog, “tweet” (a term used to describe short messages added to the Twitter site), record and photograph the sessions. Some conference organisers are now introducing embargoes on blogging, “tweeting” and photographing conferences in an attempt to keep data private. Others worry that conference presenters will choose only to speak on published results making conferences boring and heralding the end of conferences[24-26].

Geneticist Daniel MacArthur was one such blogger at the CSHL conference who inadvertently broke the rules by blogging. His blog covers many of the conference experiences as well as debate about the situation after the event. As a consequence of the debate caused by the use of blogging, contributors to the 2010 conference were asked to stipulate whether their presentation may be discussed on Twitter, with an overwhelming majority agreeing[27].

There are a number of practitioners of ONS. Of these, the notebooks of Raf Aerts, Anthony Salvagno and Steve Koch are all found within the OpenWetWare[13] wiki, whilst Caleb Morse also has a notebook on the site, but it is only used for some data. Other ONS practitioners such as Garrett Lisi have developed their own sites[28] that are not on specialist wiki hosting services, whilst Rosie Redfield[29] and Michael Barton[30] are good examples of scientific blogs showing a mixture of thoughts and results. Carl Boettiger[31] has developed an entire website containing both his primary notebook and other aspects of his scientific life.

The mathematician Michael Nielson has developed a different approach to practising ONS through the way he uses his blog[32]. In this example the blog is used both as a place to report news and thoughts, and also as a place to pose mathematical questions and research, allowing other mathematicians to collaborate. This is common to a number of mathematicians who all have similar blogs. Michael Nielson poses the question of how much science could be done online and routine problems solved, just by sharing information and problems on his blog[33]. These thoughts and ideas are at the core of open science.

4.1.6 The concept of the blog as the primary research record

Here at Southampton, building on the successes, and learning from the less successful areas of the Smart Tea project, a different approach to electronic recording methods has been developed. The Smart Tea framework was very prescriptive and would have taken complete restructuring to make it useable for other disciplines. Instead, continuing the successes of using an internet based platform and tablet PCs, a new primary research notebook in the form of a blog has been developed. The blog also subscribes to the concept of Open Notebook Science by publishing the data at the time of creation. The blog has been developed in the context of recording the biosciences of the preceding chapters: that of the directed evolution of a β-

glucuronidase into a β-galactosidase using neutral drift, and assessing whether galactose and 2,2,3,3-tetrafluorogalactose are inhibitors of β-galactosidase.

The blog that has been developed, named LaBlog, functions identically to other blogs with the addition of posts by the blog author(s) and the ability for other users to comment on posts. However, unlike other blogs, LaBlog is intended to function as a place in which to put the full experimental detail, including all the raw and analysed data. An experiment can be followed throughout its entire course with the aid of two-directional tagging and linking. The results are not in a presented form like other blog examples previously cited, but much more like a paper notebook, showing all the experimental detail from methods to results, be they successes or failures. LaBlog also emulates both a paper notebook and a conventional blog by allowing the author and others to add thoughts and notes on the work. Additionally LaBlog allows for complete reports or papers to be added, as well as data from any machine through the ability to upload data of any file format. By allowing data of any file format to be accepted and allowing whole experiments to be added, it also emulates some of the ELN systems currently on the market. It is able to have this added functionality by not being hosted on conventional blogging sites allowing for a much more structured database system tailored to the use as primary notebook.

4.2 Results and discussion

The research and development of the blog system has been a two way process of end-user requirements directing the implementation of features, and the implemented features being tested to the break point by the user to discover how things worked, what didn't work and why. This chapter talks about the end-user directing the course of blog development, and testing how the blog worked in practice in the context of a bioscience laboratory. This is therefore research into the models and formats in which every-day work in a bioscience laboratory can be recorded most effectively and efficiently. The blog features were developed and built by a separate software engineering team, therefore no details have been provided as to the software and hardware that the blog is built on, nor any examples of code.

The data referred to in this chapter is data used in the preparation of the neutral drift libraries discussed in chapter 2 and also covers a large amount of the optimisation work before the libraries were successfully prepared. It also refers to work relevant to the assays testing whether galactose and fluorogalactose are inhibitors of β-galactosidase. The data is in the form of text, jpegs, Microsoft Excel files, PDF and FASTA file formats. Data was added and uploaded directly to the blog by being typed into the add post window. Data files that were not text were uploaded from the local drive of the computer using the upload data feature.

4.2.1 The development of LaBLog. Phase 1 – an introduction to blogging using the "beta-glu" blog.

The "beta-glu" blog (http://blogs.chem.soton.ac.uk/beta_glu)[34] is an example of the first version of the LaBLog system. It was started approximately nine months after practical laboratory research had been taking place. Therefore a large amount of the research had taken place before results had started to be recorded. This corresponds to most of the work featured in chapter three on the betagalactosidase inhibition assays, although some parts of that work is referenced on the beta-glu blog.

4.2.1.1 Connecting what is written on paper to what can be written on screen.

The recording of an experiment in a laboratory notebook is a skill. Some lab books may have handwriting that is illegible, whilst others can contain so little detail as to be unintelligible. A lab book can be a very personal item and what makes sense to one person, and be sufficient to prompt a memory can be meaningless to another. The amount of detail and information recorded for an experiment varies between different people.

What is considered to be the important information to record, and the method by which experiments are recorded has been at the forefront of developing the laboratory blog. The paper lab book was at the heart of the first foray into using a blog as a research notebook.

A blog is very much like a piece of paper in that the writer starts out with a blank page. For example, when writing up experiments in a paper lab book, the natural instinct was to treat each page as an experiment, starting with the title and safety information. Below that the reagents and quantities were written in a table type structure with the method placed as a separate piece of text underneath. This can be seen in figure 4.1, a representative page from the paper lab book for a restriction digestion that formed part of the results of chapter 2.

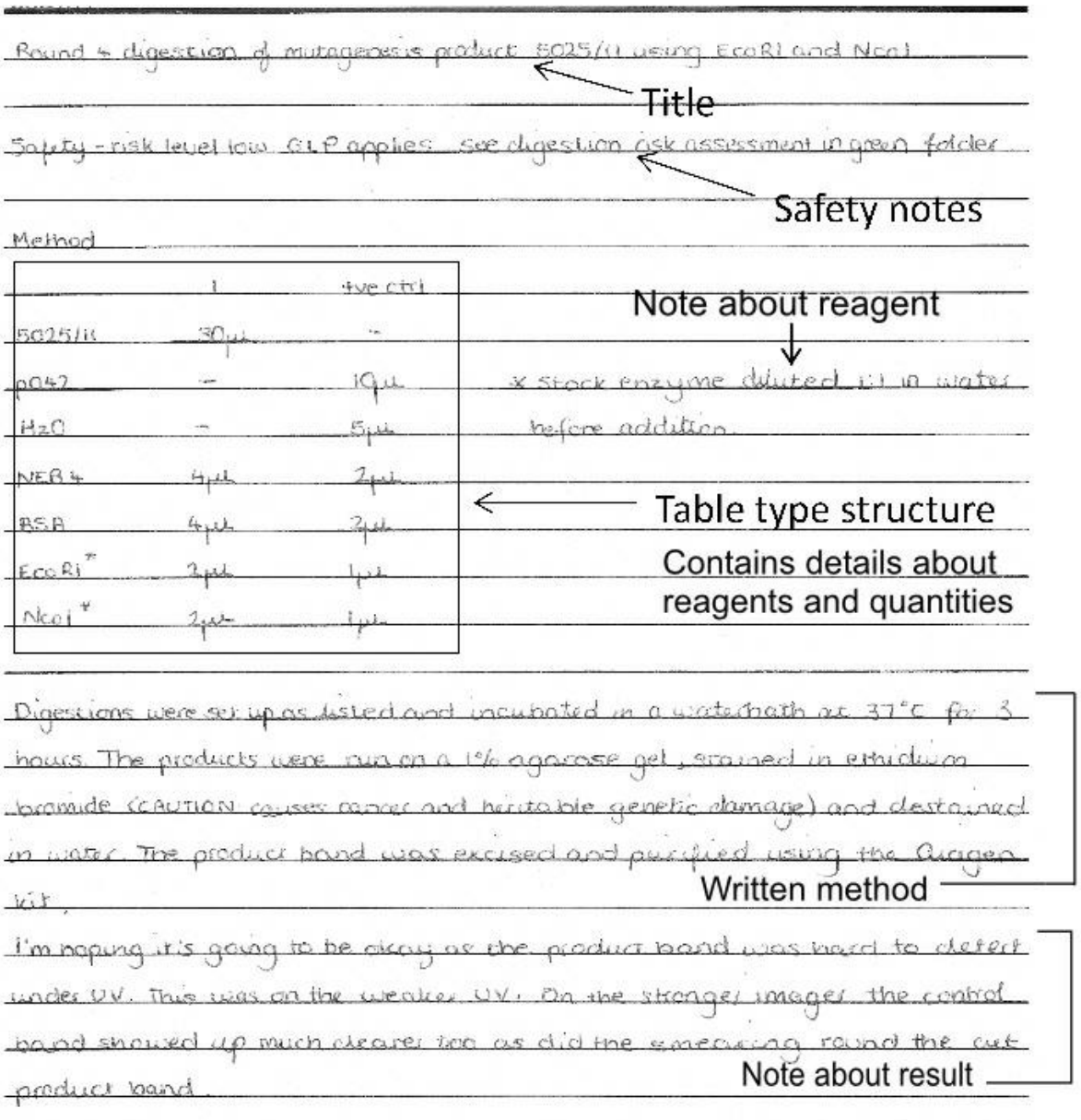


Figure 4.1: Representative page from written lab book for a standard restriction digest procedure. The figure has been annotated with descriptions of various areas of text.

At the outset of the use of the blog, the natural instinct was to consider a post as if it were identical to a page in a paper lab book i.e. treating the post as if it was one discrete experiment starting with reagents and finishing with results. However the first problem encountered is that a

blank text editor window does not have the same degree of flexibility as a blank sheet of paper with respect to formatting. Paper gives almost total freedom of text placement, format and space utilisation. In contrast, computerised page content must follow rigid rules and protocols for preparing visible formats. Whilst ordinary text can usually just be typed into the post editor, other structures such as tables can only be created by writing instructions using a code. The code contains the instructions for displaying a table, whilst the data within the table must be written inside the code.

Initially the blog was unable to display tables and the software engineers had to enable the creation of tables based on user request. For the blog, the code chosen was BBCode which is similar to HTML. BBCode was a personal choice due to previous experience in using that language. HTML could equally have been chosen, particularly as HTML is also required to display characters and formats not found as standard on an alpha-numeric keyboard, for example the subscript 2 in H_2O and the Greek character mu as µ. These commands are found inside < > brackets or involve prefixing characters with an ampersand (&) and suffixing with a semi colon (;), whilst BBCode commands are found inside square brackets.

For example, to display the table style structure shown in figure 4.1 as a table on the blog, the code that would need to be written to the post is:

```
[table][mrow]Reagent[mcol]1[mcol]+ve ctrl[/mrow]
[row]5025/11[col]30 &mu;L[col]-[/row]
[row]p042[col]-[col]10 &mu;L[/row]
[row]H<sub>2</sub>O[col]-[col]5 &mu;L[/row]
[row]NEB4[col]4 &mu;L[col]2 &mu;L[/row]
[row]BSA[col]4 &mu;L[col]2 &mu;L[/row]
[row]EcoRI[col]2 &mu;L[col]1 &mu;L[/row]
[row]NcoI[col]2 &mu;L[col]1 &mu;L[/row][/table]
```

This method of writing is time consuming, and would have been difficult for a user without prior knowledge and experience in the use of HTML and BBCode to have written. However the added work load and time penalty is offset by the advantage of having structured data.

Tables are useful for displaying molecular biology experiments because they convey a lot of information in a simple and clear manner. In the example shown above, the reaction, which is only being carried out in duplicate, requires the addition of seven different reagents in varying volumes to two different reaction vessels. If the experiment was written in a purely textual form, it is likely to be much harder to follow and may lead to replication errors. As the number of experiments increases this would become even more confusing. For example when doing the enzyme activity assays, such as those carried out on mutant 5025/46.b1 in chapter 2, there are

up to 96 reactions being carried out in parallel. This type of experiment is much better suited to a table depicting the 96-well plate than a written list of what went into each well.

4.2.1.2 Categorising posts into sections and using metadata.

In a paper laboratory notebook, experiments are naturally always arranged chronologically by date. The pages of the lab book follow on with no gaps being allowed. A blog is different as the news feed is usually organised to show the most recent post first i.e. chronologically reversed. A blog is also more flexible as the posts can also be organised into sections, which may sort the posts by type rather than date. At the outset of using the blog, both the directed evolution project on β-glucuronidase (chapter 2), and the inhibitor investigations on β-galactosidase (chapter 3) were in progress concurrently. Therefore the posts were initially categorised into two sections corresponding to the two projects, with the sections being called “beta-glucuronidase” and “beta-galactosidase preparation and assays”. All posts were assigned into one of those particular sections, which aided following the work flow by categorising which experiment a particular procedure belonged to, as by selecting the relevant section, the newsfeed could be ordered chronologically by experiment.

Shortly after the blog was initiated, the beta-galactosidase project was stopped and newer posts on the beta-glucuronidase project began to be categorised into new, more specific, sections which were added to the beta-glu blog as the research expanded, rather than just added to the “beta-glucuronidase” section. The number of sections increased to accommodate these new posts, leading to the blog becoming difficult to navigate and posts being more difficult to assign (figure 4.2). This is a natural process as the blog data model mentally evolves and is a process which makes reorganisation difficult. For an ordinary blog covering a wide variety of subjects this many sections may not have been a problem, but for a primary research notebook it was a problem.

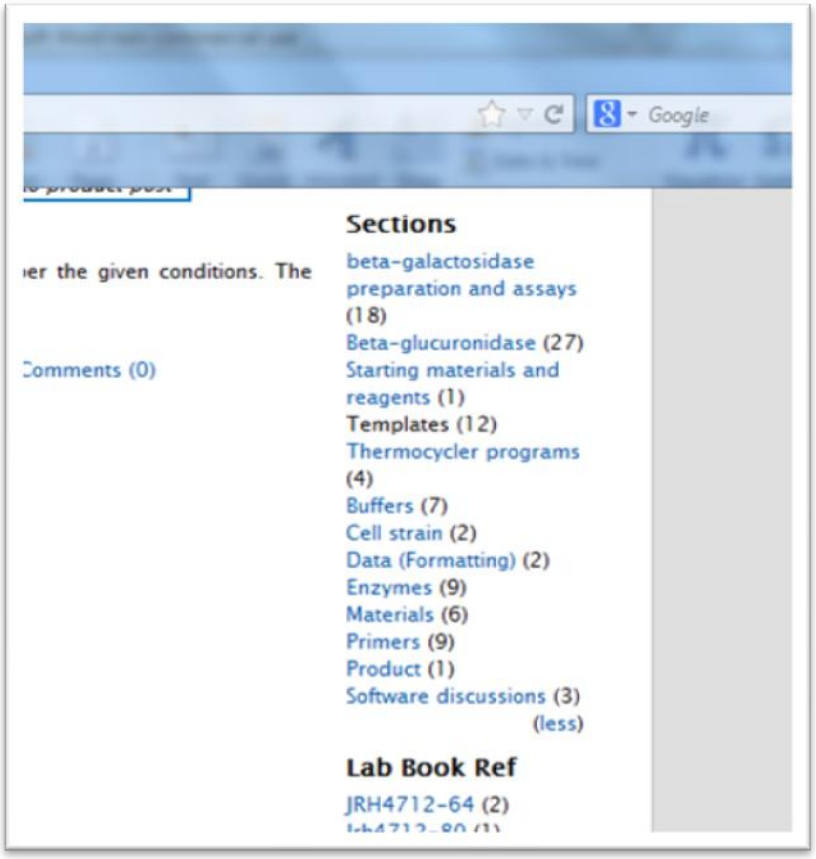


Figure 4.2: close up section of a screenshot of the beta-glu blog showing the number of sections on the blog. These are listed under the heading sections and amount to 13 categories in total (beta-galactosidase..., Beta-glucuronidase, starting materials..., templates, Thermocycler programs, buffers, cell strain, Data, enzymes, materials, primers, product, and software discussions)

The main method of organising and categorising blog posts is through the use of user-defined categories, and tags comprised of key-value pairs. As users are able to define categories, it can lead to a tendency to overload specific categories. While a post can only have one category, it can have multiple key-value pairs associated with it.

Early use of the blog tended to overload the categories (in this case the category defined as “sections”), and not use the tags effectively. As a consequence of the categorising and tagging being ineffective it made it difficult to link sequences of work together in a meaningful manner.

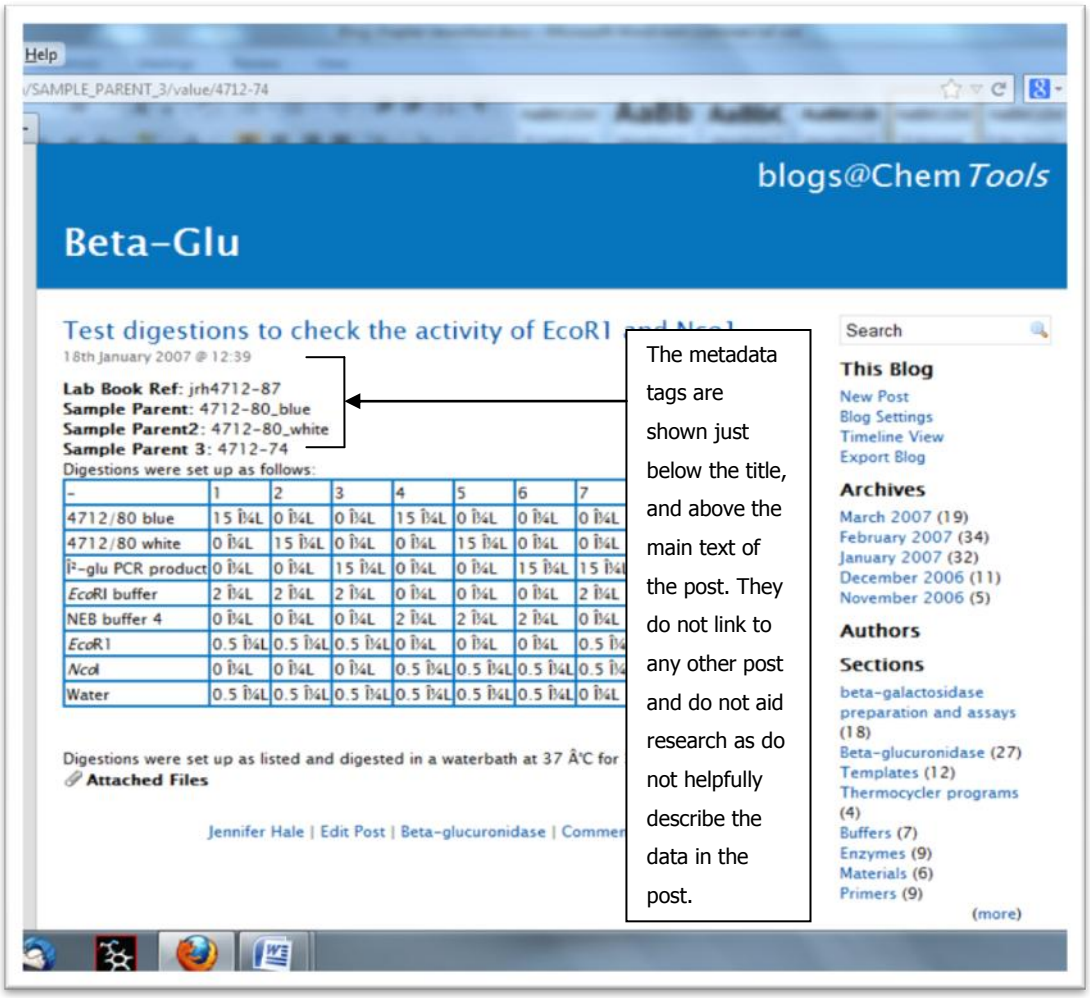
For instance, at this stage experiments were still being written up in the paper lab book as well as on the blog. Due to this the duplicated paper laboratory book page was selected as the key-value pair for tagging experiments. This was not a sensible tag as laboratory book references do not help to categorise data effectively. A laboratory book reference may be a useful piece of information if you have the book, but it doesn’t provide any useful information about the experiment to an outside user such as what type of experiment it is and other information regarding what happened during the experiment. Additionally a book reference is an infinite list,

the number of references increase as the number of experiments increase. Therefore the list of key-value pairs was infinite.

Within a complex set of experiments, such as those of the directed evolution, it was clear that a goal for the blog was to enable a researcher to be able to track through the process of one research “thread” among many “threads” in progress. For example on the beta-glu blog this would be to track between experiments regarding the directed evolution and experiments regarding the inhibition assays. Early attempts to implement this tracking led to tagging the data as “sample parent” and “product”. Categorising data as sample parent and product attempted to link experiments together and show how they were related through the procedure by which a sample is changed from a reagent (“sample parent”) to a product. In a series of experiments, the product of one procedure may become the reagent for another. Therefore tagging with these categories should have made it easier to get an overall picture of how the samples fitted together and were used for each procedure and throughout the cycle. In practice this did not work as the wrong value was assigned to the tag key. The key was “sample parent” or product, but the given value assigned to each key was its associated paper lab book reference.

This meant that like using lab book reference as a key value pair, the list was infinite and meaningless. Though a sample was now a sample parent or product or both, it still didn’t tell the reader or user anything useful about the data such as what it was exactly, only that it linked to the paper reference. This did not help research in general, only the transition to understanding recording experiments on the blog alone (though this was not happening at that stage).

As the new tagging system was used, it also became clear that the tag system was causing a lot of confusion in the linking of sample parents and products. This confusion highlighted a fundamental flaw in the organisation. One example is shown by reactions such as ligations which combine two sample parents (in this example the ‘insert’ and the ‘backbone’) to produce one product. It was impossible to tag both the insert and the backbone as the sample parent, so further tags of sample parent 2, and in some examples sample parent 3 had to be added as a solution to this problem. Figure 4.3 demonstrates a digestion reaction where above the method are the tags sample parent, sample parent 2 and sample parent three with three lab book references. This highlights the confusion in tagging the data and putting it in the context of the overall experiment.



-	1	2	3	4	5	6	7
4712/80 blue	15 Î¼L	0 Î¼L	0 Î¼L	15 Î¼L	0 Î¼L	0 Î¼L	0 Î¼L
4712/80 white	0 Î¼L	15 Î¼L	0 Î¼L	0 Î¼L	15 Î¼L	0 Î¼L	0 Î¼L
Î²-glu PCR product	0 Î¼L	0 Î¼L	15 Î¼L	0 Î¼L	0 Î¼L	15 Î¼L	15 Î¼
*Eco*RI buffer	2 Î¼L	2 Î¼L	2 Î¼L	0 Î¼L	0 Î¼L	0 Î¼L	2 Î¼L
NEB buffer 4	0 Î¼L	0 Î¼L	0 Î¼L	2 Î¼L	2 Î¼L	2 Î¼L	0 Î¼L
*Eco*R1	0.5 Î¼L	0.5 Î¼L	0.5 Î¼L	0 Î¼L	0 Î¼L	0 Î¼L	0.5 Î¼
*Nco*I	0 Î¼L	0 Î¼L	0 Î¼L	0.5 Î¼L	0.5 Î¼L	0.5 Î¼L	0.5 Î¼
Water	0.5 Î¼L	0.5 Î¼L	0.5 Î¼L	0.5 Î¼L	0.5 Î¼L	0.5 Î¼L	0 Î¼L

Figure 4.3: screenshot of a post on the beta-glu blog where the metadata tags and descriptors are useless for categorising the data due to being assigned what are, in essence, arbitrary numbers. These are lab book ref, sample parent, sample parent 2 and sample parent 3. There are a number of random symbols on view within the table. This is due to an upgrade of html, which has led to html rendering differently to when it was first entered.

The solution to this organisational problem came from the realisation that the different parts of the experiment could be linked together using *links*, while the post categories and key-value pairs should focus on what the post represents. Therefore the connection between posts was shown by links and finally the data in a post could be categorised by what it represented.

4.2.1.3 Expanding the capabilities of the blog.

One of the ideas of the blog and its capabilities as more than an ordinary lab book was to add far more information than is found in a typical paper record. Thus the blog can be used as a database, with materials such as enzymes, buffers, cell strains, and plasmids having information stored about them. For products such as enzymes, a new blog entry is made each time a new batch is opened. By linking a post to that particular batch, the sorting and searching facilities of the blog become a powerful tool. For example if certain experiments are not working

and they can all be linked to a common batch of material, it may indicate that the particular batch has a problem. Storing information on materials allows the blog to act as a repository or catalogue for multiple researchers.

In section 4.1.2.1 the use of tables was discussed for clear presentation of certain types of data that are frequently repeated, and the time penalty incurred from preparing those tables. There are quite a few standard types of posts that include tables which are frequently repeated. To make the blog more useful, it was realised that there was a need to have a quicker, more efficient way of preparing tables than by manually coding a table each time one was required. The obvious answer to this is to use templates: sets of preset forms designed to be relevant for each type of material or procedure. A template would automatically generate the required post type, allowing the user to just fill in the relevant data.

From personal use of the blog and thinking about how templates could be effectively used, an idea emerged that a template could become very powerful if it was also able to be used to implement the links between posts. The vision became realised through a design for procedure templates that would contain drop down menus with links to other posts allowing selection of relevant materials, reagents and products. The first template ideas based on this vision are shown in figure 4.4. This idea for how templates could work was passed to the software engineers for implementation. Unfortunately, the implementation of that style of template was not possible using that particular blog technology. At that stage, in order to use templates, the table structure was ready prepared and the data was added by editing the underlying code. This made them difficult to use and did not enhance practical research at this point.

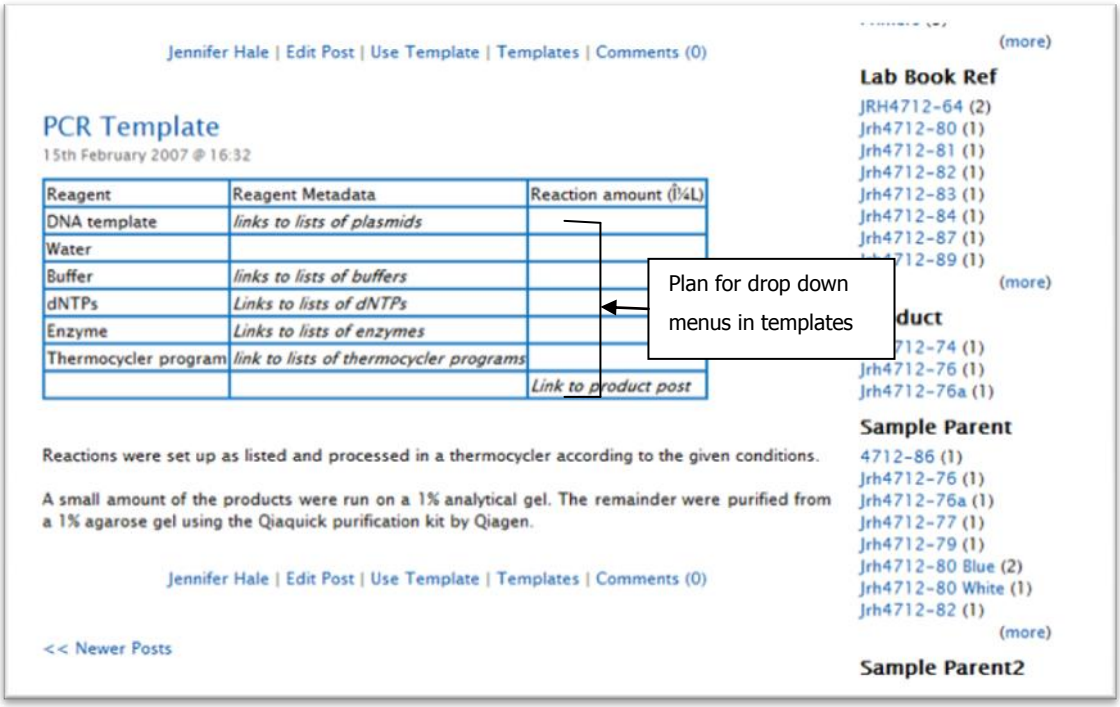


Figure 3.4: an example of an early template as it was envisioned to appear and work. Under the reagent metadata column, the idea was to have clickable boxes which would produce drop down lists of other related posts.

4.2.1.4 Moving forward – lessons from the beta-glu blog

Researching the best way to use and organise the blog, and what features needed to be implemented had highlighted a number of issues.

1. Tables are useful for displaying molecular biology research. Writing the table into each post is, however, time consuming and as a result does not enhance research. Templates are a solution for saving time. If the templates can be linked to other posts, this increases the power and capabilities of the blog.
2. Assigning posts with sections and metadata has to be done in the right way or the information is lost and linkages cannot be made. From these considerations it was envisioned that the logical extension of the linking of posts to materials was to link each individual step of the experiment together to produce an entire chain, or web, of experiments. This would show explicitly where and how each material, product and procedure was being used throughout the course of the experiment. In order for this to be possible each piece of data used in the research including physical entities such as buffers, enzymes or a plate of colonies, in addition to procedures and products, would be assigned a post and given a single metadata tag that would allow it to be linked to any other post. The new approach that would link all the posts together was designated as ‘one pot-one post’.

In order to move forward, a new blog was started using improved blog technology and with some post structure pre-determined.

4.2.2 The development of LaBlog. Phase two – a new blog with a more clearly defined structure.

A new blog was started which was given the name “Investigations into neutral drift” (http://blogs.chem.soton.ac.uk/neutral_drift)[35]. This blog will be called neutral drift blog from here on in. Though this blog looks and appears to function identically to the previous blog, the newer blog was built on improved technology. From the outset, blog posts were given much more structure with clearly defined section categories labelled materials, procedure, product, notes, safety and templates. The metadata was defined as a single category called post type, and post types were tagged into discrete articles by the material type or the procedure name e.g. PCR, plasmid, buffer, enzyme or digestion.

4.2.2.1 What is a post and what makes a procedure an experiment?

The ‘one pot, one post’ approach initially had some problems. The approach logically implied that every reactant, procedure and product should be linked together and thus each entity should have its own post. However when procedures were carried out in parallel this led to an extremely large number of posts being generated. Not only did this burden the user with lots of extra work to prepare each procedure post, it did not aid research. In fact it made it much harder to follow the course of the research due to the number of posts created. This is demonstrated by the ligation reaction given the experiment number 4880/18. The ligation was carried out on four separate pots, which would produce four pots of product. Each of these was given its own post and the procedure was written up four times with each procedure separated by a delineating number e.g. 4880/18.1 (figure 4.5).

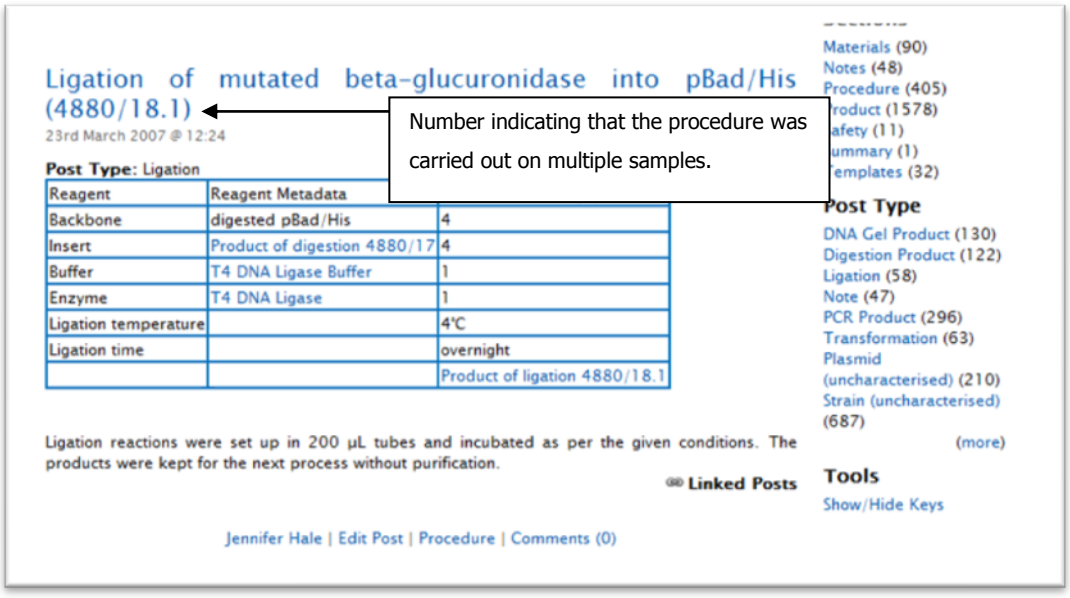
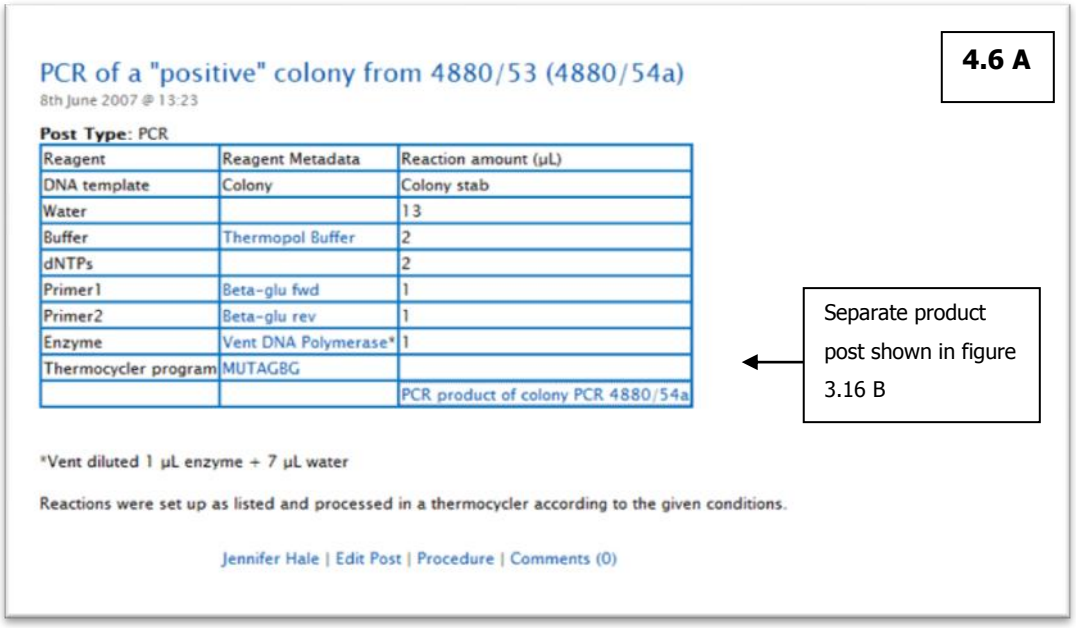
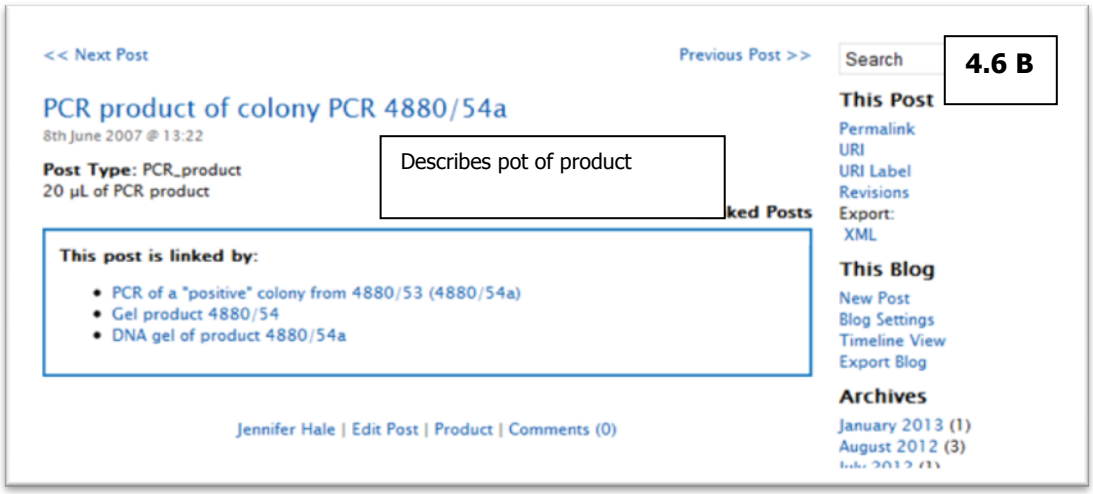


Figure 4.5: a screenshot showing one copy of a procedure post for a ligation that was carried out in parallel four times. As a result the procedure was written up four times with each copy being annotated as 4880/18.1, 4880/18.2, 4880/18.3, and the control as 4880/18.4 pBad only.

It became immediately obvious that it made far more sense to have one procedure post linking to all the products produced and their associated posts, even if that meant not rigidly following the one pot-one post rule.

What also had to be determined was what constituted an experiment, and therefore an individual procedure post. For example, procedures such as PCRs and digestions are carried out in two parts. In the first part the DNA is amplified to produce an impure mixture containing the correct DNA product and other materials. This is separated and purified by gel electrophoresis into the pure DNA product and other side products. Initially sticking rigidly to the one pot-one post format meant that the gel electrophoresis was given its own separate procedure post(s). Like the issues with creating multiple procedure posts, having split procedures caused problems and did not aid the research due to the added burden of writing the extra posts. Additionally it increased the difficulty in following what procedure was linked to what product, and what subsequently happened to that product. This is demonstrated in a series of three posts which follow the path of a PCR sample from initial procedure to gel product (figure 4.6). Trying to write the procedures in this way highlighted that it wasn't the correct way to manage the posts and links between them. For a procedure such as a PCR or a digestion, one experiment, and therefore one post involves linking together the two halves of the procedure for it to make sense as a whole and produce a more logical linkage.





Parts A and B of figure 4.6.

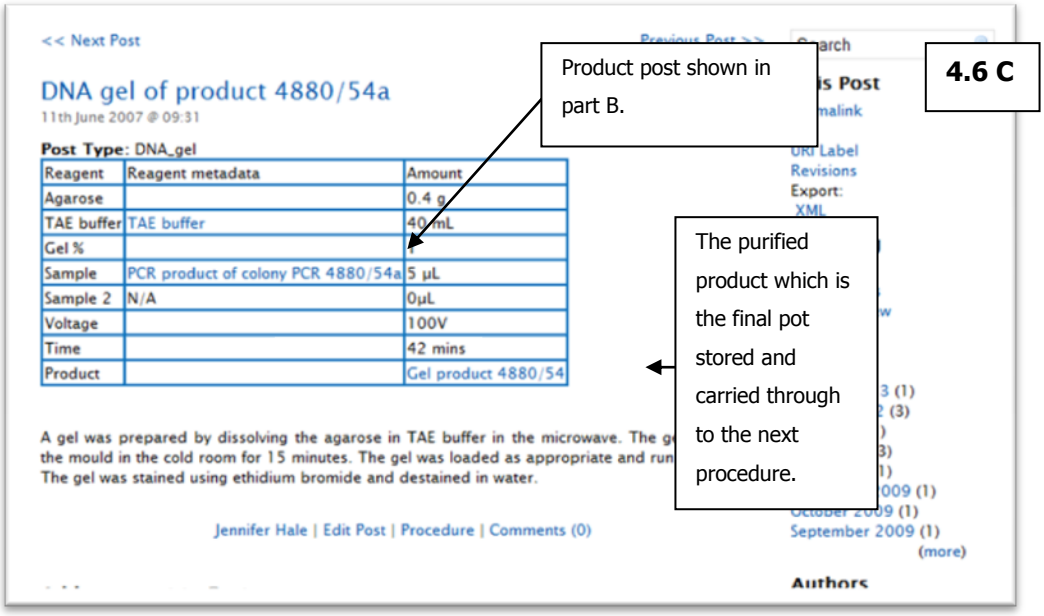


Figure 4.6: Part A. The initial PCR procedure post which has a product post called “PCR product of colony PCR 4880/54a”. Clicking through the product post link leads to the post shown in part B, which is the product before it has been purified. Following through the box of linked posts, the DNA gel procedure can be found which produces a gel product, which would be the purified product (part C).

The question of what constitutes an experiment and how to organise the data also highlighted another issue with regards to the number and types of posts featuring on the main news feed. The news feed showed every post including products and materials. This made the newsfeed unmanageable to read as it was interspersed with posts that were not immediately relevant. It also made it harder to follow the course of the research. When interacting with the system it is mainly the procedure posts and notes that will need to be read by the user and any other visitors to the notebook. The software engineers took this criticism and adapted the news feed so that only procedures and notes were displayed. Materials, products and other types of posts can still be accessed directly using the navigation menu.

4.2.2.2 The importance of working templates

The use of templates was a key feature to implement in the blog to aid and benefit use of the blog and thus the practical research. However templates would only benefit the research when they worked correctly. Finding and implementing new technology that allowed increased automation of the templates would be the key to making the blog a powerful tool for research. While the software engineers were developing the code for automation, it was important to elucidate the best data structure to improve the usability of the templates once automation could be implemented. This structure would also aid the research by making it easier to link together related posts and follow the course of a sample. Useable templates for the one pot-one post approach took several rounds of design optimisation, particularly for procedure posts.

Templates for reagents were much easier to design correctly as they are much simpler. The reagent templates follow a standard structure laid out during use of the beta-glu blog in which the reagent is named with the properties and associated data listed e.g. ingredients of an enzyme storage buffer, expiry date etc.

The first procedure templates envisaged on the beta-glu blog have already been shown to be unsuitable as they were not functional as templates, requiring editing of the code of a post to add data. They were also unsuitable because they didn't fit the one pot-one post approach. These templates treated a procedure as a single event whether it was carried out singularly or in parallel.

A second generation of procedure templates were designed to combine the original template idea shown on the beta-glu blog and the new one pot-one post approach to record keeping. These templates did not get the data structure correct so were impractical to use. The problem with the data structure of these templates was the addition of the term 'metadata' to the template to introduce the link between reagents and products. This is an artefact of thought processes carried over from the data structure of the original beta-glu blog in which the concept of metadata (in this case tagging with key-value pairs) and its use was misunderstood. An example of the use of the word metadata is found in experiment 4880/40, a transformation reaction (figure 4.7). The template has extra columns to accommodate the metadata linkage column, making the displayed post too wide to be visible on screen, and as a result difficult to follow.

The solution to this confusion was eventually found during template optimisation and is based on linkage by post titles. Other mistakes included trying to condense all of the information into a table rather than having a clear written experimental method. This made the work difficult to follow in retrospect as there wasn't enough information given to repeat the work carried out. This does not benefit research for the primary blog user or anyone else who might have been trying to follow the work.

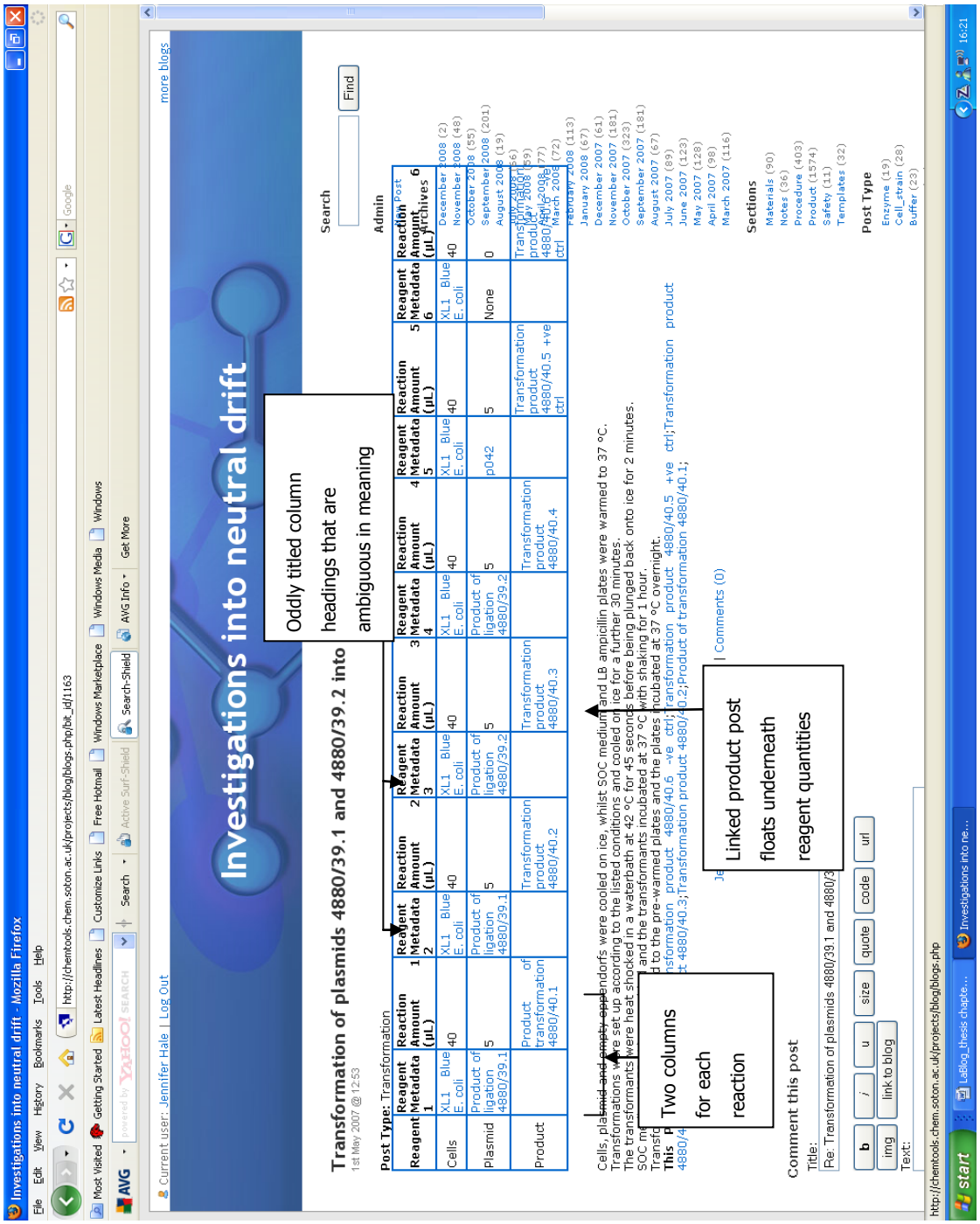


Figure 4.7: An example of a second generation template for a transformation reaction that did not display the data in the best way. The template has extra columns for reagent metadata 1-6 which increase the number of columns, extending the template over the navigation menu. The product post is oddly situated below the values for reagent amounts and the whole post does not flow well demonstrating the incorrect structure of the data.

The answer to structuring the data in the procedures was solved by switching the data for rows and columns, and using effective metadata. Although this change seems trivial, many of the big changes to blog use have been based on small interface changes. It highlights how difficult it is to work out correctly in advance what is considered to be right with regards to data structure and visual display. As has previously been shown (figure 4.7), in the second generation templates, the reagents for each procedure were listed in a single column which was followed by a number of columns relating to the number of times the procedure was to be carried out. Each individual multiple of the procedure used two columns: the first column for the metadata and the second for the quantity. If rows and columns were switched, then a single multiple of the procedure could be read across one row, and the row reads much more clearly and logically. By using post titles as the descriptor for the item, and reordering the data within the templates, the linking of samples and procedures was increased. This also removed the need for the ambiguously defined 'reagent metadata' column. This fulfils one of the aims of the blog of linking procedures to batches of reagents, which allows monitoring of reagents to highlight when that batch expires and experiments stop working.

The new procedure posts were much clearer to follow and to use, and would form the basis of the automated templates (figure 4.8). Figure 4.8 is a transformation procedure almost identical to that in figure 4.7. The main difference is that the table is now much narrower and the number of times the procedure is carried out is clearly numbered.

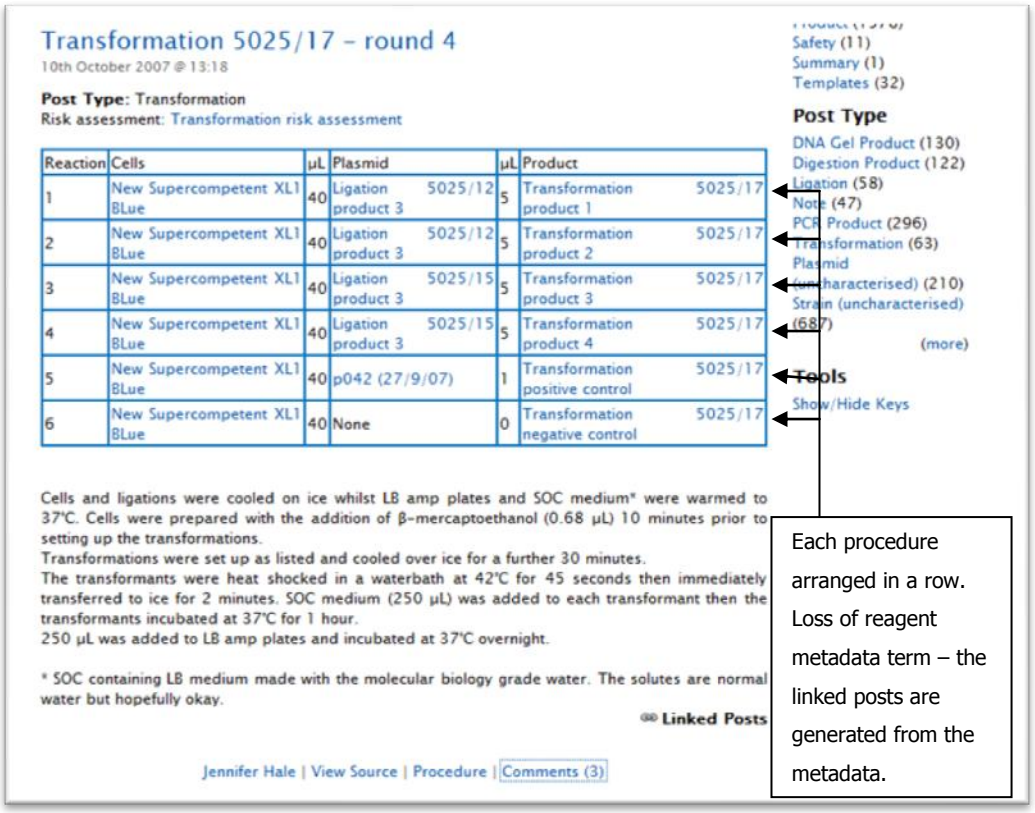


Figure 4.8: an example of the new improved template structure which has arranged the data by rows rather than columns, and contains much more direct data linkages without highlighting that it is metadata which has allowed the generation of those linkages.

4.2.3 Phase three – the introduction of automated templates

The common factor with all of the designed templates was the need to edit the template at BBCode level via the "edit post" box to input data. The change in the usability of the blog came with the introduction of template automation and the implementation of the drop down menu as had been the vision in the original template design. This allowed the addition of posts to the blog containing data by directly using the template, rather than the addition of an empty template which had the data added afterwards. This coupled with the optimised data structure allowed use of the blog to increase in effectiveness.

Once created, the new templates could be used through a new form generated from the 'use template' button on the blog, and did not need to be edited at the BBCode level. The templates included drop down menus for selecting other blog posts, thereby allowing easy selection of the correct post to be added in each cell requiring a link.

4.2.3.1 Creating the templates

In section 4.2.1.1 the basic commands used to code a table in BBCode was introduced. The new templates are still based on these basic commands, but extra commands are used to generate the added functionality. Double brackets [[ ]] are used as markers which produce editable boxes and lists when the template is viewed as a form. The double brackets also store links when the template is saved as a post. The double brackets have instructions inside them which describe the desired function. For example the command [[box]] would be used to prepare a text box that might be used for specifying the quantity of a reagent. This command would be written inside a particular cell of a table. Leaving the cell blank will produce a box that cannot be edited without using the 'edit post' function, whilst writing text in the cell would produce a cell containing text that cannot be edited without using the 'edit post' function.

In the same way that [[box]] is used to produce a box, a dropdown list is prepared using double brackets and the metadata tags. For example to produce a list of buffers the command [[Post_type:Buffer]] must be used. Post type is the metadata key and Buffer is the value of that key. For lists where one or more post types could apply for example 'Plasmid' and 'plasmid (uncharacterised)' then the command [[Post_type:plasmid%]] would be used. The % sign selects from all metadata key values containing the word plasmid. These things can all be demonstrated in the following code for a table that would produce part of a transformation template:

```
[table][mrow]Reaction[mcol]Cells[mcol]Quantity[mcol]Plasmid[mcol]Quant
ity[mcol]Product[/mrow]
[row]1[col][[Post_type:cell_strain]][col][[box]][col][[Post_type:plasm
id%]][col][[box]][col][[Post_type:strain_(uncharacterised)]][/row]...
```

The words reaction, cells, quantity, plasmid and product would show up in the form and would not be editable from the form. Cells given the command [[box]] would be empty and editable, whilst those given the command [[Post_type:]] would produce drop down lists.

An example of part of a template post displayed on the blog (figure 4.9) shows the commands [[box]] and [[Post_type:]], plus boxes with pre-set data/titles added. In this example compared to the code above, the term "µL" has been used rather than "quantity". The template also demonstrates how commands in double brackets do not have to be used within a table. At the bottom of the post, [[Post_type:strain_(uncharacterised)]] is used to form the first words of a sentence of free text as the post it is linked to can also form the first words of the sentence. The command has also been formatted to allow the text to show up bold in the final displayed post. This example shows the post as it would be displayed on the blog in the news feed or through selecting templates. To access the template as a useable form, the "use template" function would be selected from the post (not shown).

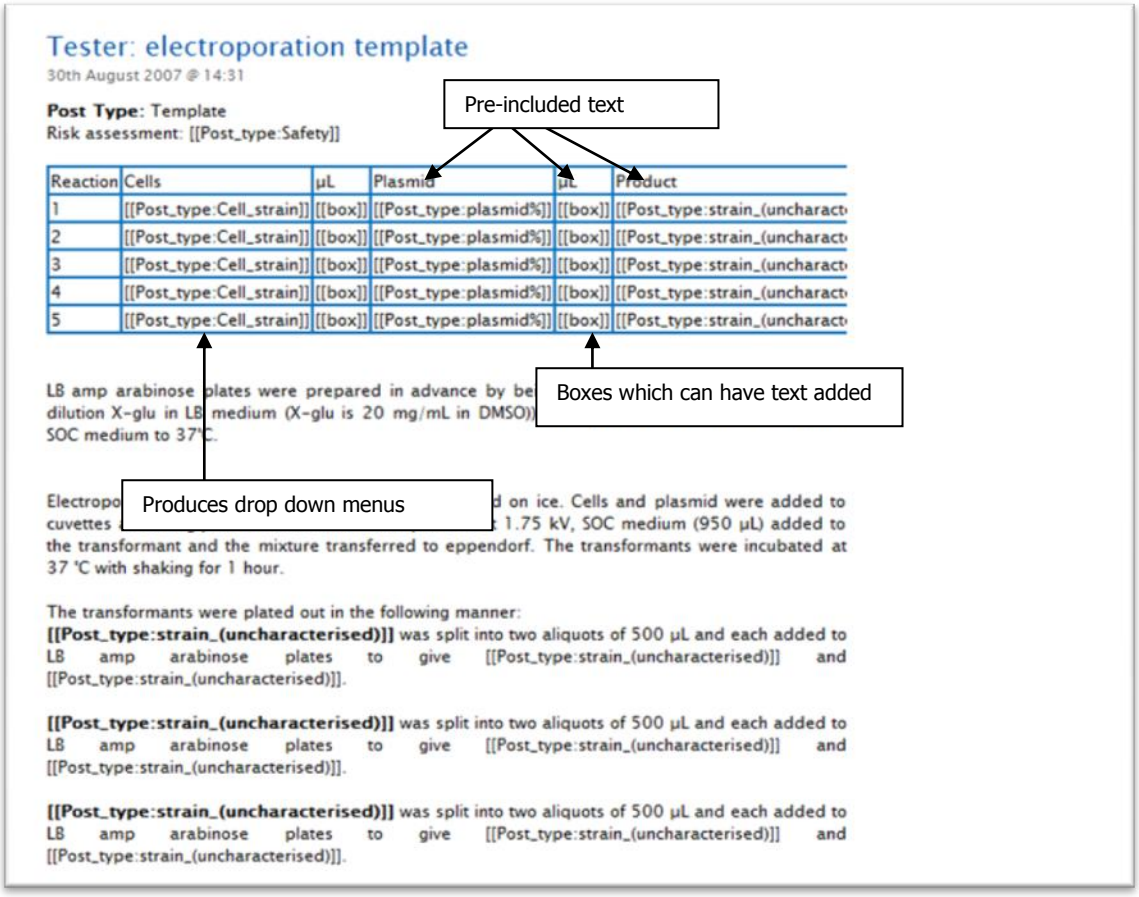


Figure 4.9: part of a template for an electroporation procedure showing how the drop down box and text box functionality is prepared using [[]] and non-editable text is prepared by pre-writing in the post during template creation.

It took time to prepare and create the new templates ready for use, but the time was not considered to be a burden due to the advantages of using working templates, and the time that would be gained back for future practical research.

4.2.3.2 Posts must be named systematically to aid the use of automated templates

When a drop down menu is selected in the form, each post in the list appears by name. If the name is ambiguous, it is much more difficult to select the correct post. This slows down research as time is consumed checking on the blog for the exact name of the required post. This problem highlighted the importance of how post titles were named and what data they contained, as the titles of the posts must provide adequate information for the correct post to be selected when using the menu. Therefore a new system for naming posts systematically to introduce a greater degree of clarity was implemented.

Previously post names had been randomly assigned usually with the paper lab book reference being a major part of the title. This was changed to a system assigning each post an experiment number, experiment name and round number to indicate which round of the cycle the experiment belonged to. This is because there are four repeated rounds in each experiment. If the post was a repeat of a previous experimental post, an attempt number was also added. This was a matter of user choice and each individual blog user can define post titles as they choose. In this instance experiments continued to be assigned an experiment number based on the system used in the paper laboratory book, though the use of the paper book had been abandoned.

The naming convention was started just prior to starting the generation of library JH2. Therefore the course of experiments from that library is much easier to follow as will be shown later.

4.2.4 Other changes to the neutral drift blog

One of the advantages of using the blog system is that it can store other related data. Therefore it was considered that the blog could also store risk assessments and COSHH forms. Safety is a big part of any experiment, and when using the paper lab book, a brief risk assessment was written at the beginning of each experiment, whilst the main COSHH forms and risk assessments were kept in a file in the laboratory. Safety data had not been added to the beta-glu blog or to the neutral drift blog in the opening months.

As COSHH forms and risk assessments naturally fit into tables and can have a standard template prepared, safety templates were designed that fulfilled the role of a standard COSHH form and experiment risk form. The paper copies were transferred to the blog and a drop down menu added to each experiment to allow selection of the relevant safety data. Though tables and templates were designed for this process, it would be equally acceptable to scan the paper COSHH assessments and upload the scanned images to the blog. Although by using a template, new assessments can be directly added to the blog and the system of safety documentation maintains consistency.

As the blog is readable by the public, a disclaimer was added to all assessments reminding the reader that the assessments are specific to the location where the experiment is carried out and may not apply in other countries and situations.

One further important feature to be added to the blog during this phase was the ‘link to’ button, available in the edit post screen. The ‘link to’ button brings up a window listing all other blog posts with the most recently posts listed first. This feature allows a blog post to be linked to any other blog post, not just posts that link through drop down menus. This made a huge difference to the usability of the blog as before automation and the ‘link to’ button, linking between any posts required copying and the URL from a second copy of the blog and pasting into the

required post. This is because it is not possible to navigate between posts without losing unsaved data from the ‘edit post’ box when it is in use. Therefore a second open copy of the blog was required to solve this issue. The easiest solution to having two copies open simultaneously was to have dual monitors and have one copy open on each screen. This was not a practical solution when viewed in the context of the overall aims of the blog, but made the blog easier to use by avoiding navigation issues caused by swapping between browser windows on the same screen. Once automation and the ‘link to’ button had been added, the blog could be used effectively with just one screen. This minor alteration to the user interface improved the general usability of the system.

As has been seen throughout, it has often been minor changes to the user interface that have improved the usability of the blog for example templates, clear post titles, the ‘link to’ button and drop down menus. All these changes have made it easier to link data together, even though it may not have initially been obvious that there was a need for a link between that data. These changes to the interface have usually come about through user comment and idea. It highlights how important a two way development process is. The developers wouldn’t know the features were needed without the users requesting or suggesting them.

4.2.5 Idiosyncrasies in the data structure

One of the most unique features of the blog is its freeform nature. A blog can be individually organised to be in the form considered most suitable by that user. The non-prescriptive nature of the blog makes it versatile, but it can also lead to idiosyncrasies in individual blogs if the data structure has not been correctly organised. An example of where the data structure is still wrong can be found in the neutral drift blog where the digestion procedure template is not suitable for digestion of plasmids as the drop down menu has been designed to sort suitable posts based on the post type ‘DNA_gel_product’. This is because for many digestions the DNA will be the gel purified product of a mutagenesis or PCR which is classified as ‘DNA_gel_product’. However, when a plasmid is digested, it will be registered on the blog as ‘plasmid’ or ‘plasmid_ uncharacterised’. This is not available to select from the menu and the correct plasmid must be added manually by editing the post.

This idiosyncrasy does not affect another user reading the blog and in most instances it would be unlikely that more than one person would be adding posts to a blog, since each blog is a personal record of laboratory work. This was, however, time consuming during research to write up as quite frequently a small sample of plasmid was digested as a control reaction to check that digestions were working. This problem could be solved if the blog was being restarted from the beginning. The post types could be labelled slightly differently by prefixing each post type with labels such as DNA, protein, or bacteria. Plasmids are also DNA and therefore if both were prefixed with DNA, then the search term for digestions could be sorted by [[Post-Type:DNA_%]]

and that would select all types of DNA based posts thereby solving the problem of not all possible DNA types being selected when the drop down list was prepared in the template.

This idea of prefixing with DNA or protein is already seen on the Sortase Cloning blog (http://blogs.chem.soton.ac.uk/sortase_cloning)[33], written by Dr C. Neylon. On that blog entities are categorised with increased specificity using DNA_, protein_ and Ecoli_ before each sub-category to indicate the nature of the entity. This came about as a result of a discussion between us that using prefixes would solve the particular example of idiosyncratic data. The implementation on another blog proved that the issue could be overcome.

The difference in metadata tags demonstrates how differently people think about a problem and approach the solution. It also indicates the strong interactions between the metadata structure and the templates, and how this affects blog use. If the metadata structure is too tightly or wrongly described, the lack of flexibility can lead to the templates not being an effective tool for adding data to the blog.

4.2.6 Using the blogs in practice

The two blogs were in use for just over two years of the time that practical research was taking place. The beta-glu blog was used for around five months during which only 101 posts were made, and the research being carried out in the laboratory was also being recorded in parallel in a paper lab book. Of the posts that were added, a number of these were exploratory posts deciding how the blog should and could be used rather than recording actual physical lab based research. The low number of posts overall is perhaps testimony to the difficulties of using the first blog effectively.

The neutral drift blog was used for approximately 21 months of research, and has been used on several occasions during thesis preparation to record other details and notes. In the 21 months of use whilst practical lab research was taking place over 1900 posts were added to the blog. Of these posts 405 posts were procedures and over 1500 were products. The blog was additionally used for approximately two weeks during this time by Dr Wendy Smith as a visiting researcher. In the time she was present she added 237 posts, which is testament to the ease of use of the blog, and its intuitive nature, in spite of having to adapt to data structures that she did not design. The number of posts in each section and of each type, and the number of posts by a particular person is shown on the right hand navigation menu of each blog (figure 4.10).

The use of the paper lab book was also totally abandoned during the use of the neutral drift blog, and corresponded approximately to the time that template automation was implemented.

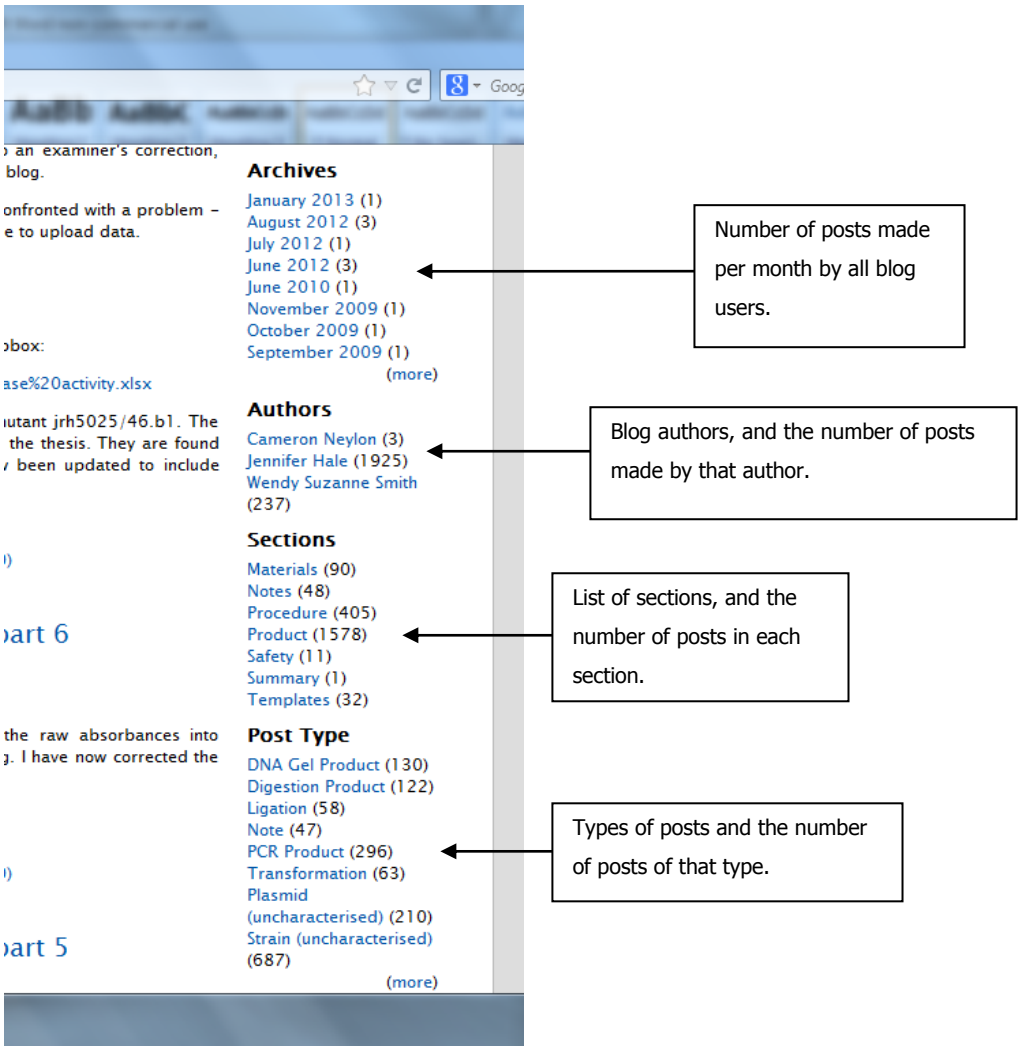


Figure 4.10: screenshot of a section of the right hand navigation bar on the blog with the date archive, authors, sections and post types listed and the number of posts of that type listed in brackets.

The blog was used as the sole laboratory record for the production of library JH2 described in chapter 2, and also was used to record the production of library JH1 in tandem with a paper lab book. The ability of posts to be linked together and for experimental courses be followed can be demonstrated in a post on the neutral drift blog. In the post a summary of all the experiments relating to practical work to be submitted for a paper was summarised (figure 4.11a and 4.11b).

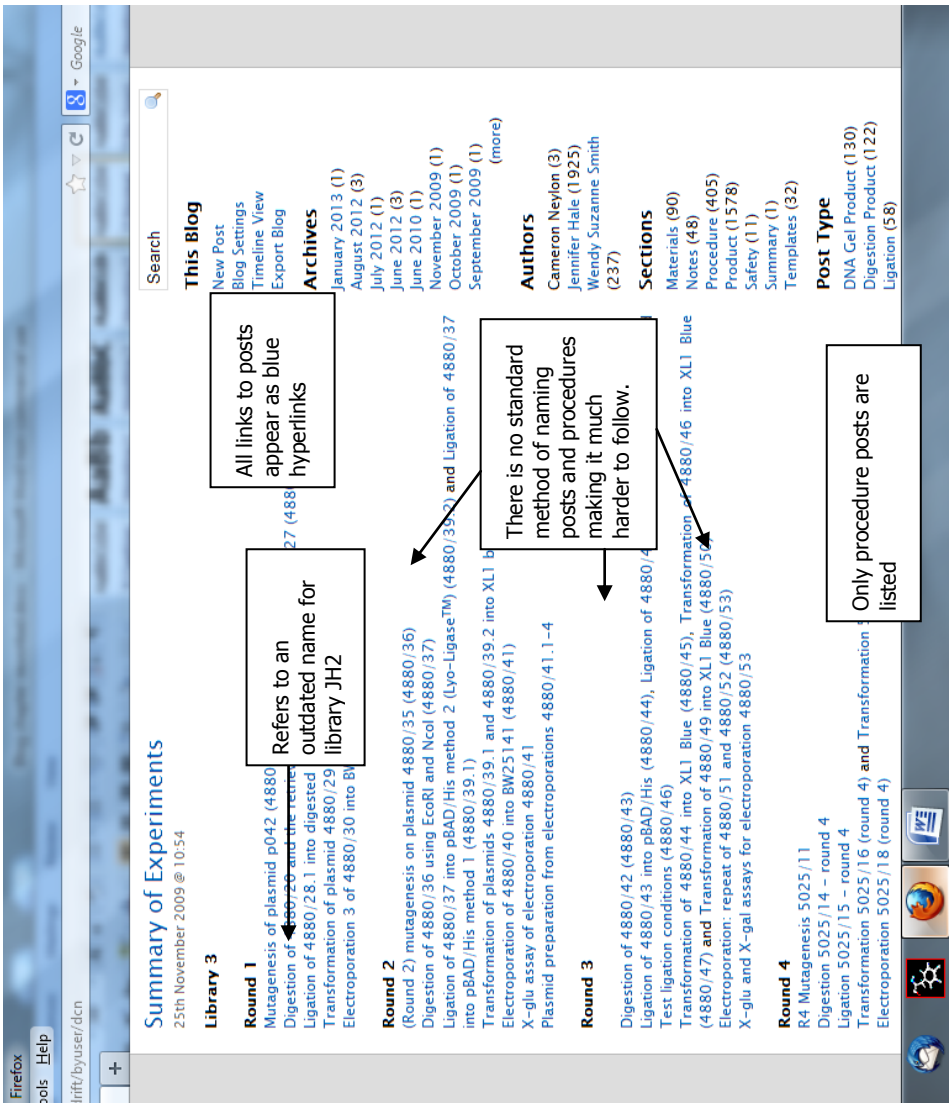


Figure 4.11a: top half of a post detailing every experiment relating to practical work submitted for a paper. The post was too long to display in full on the screen and is continued in the figure 4.11b.

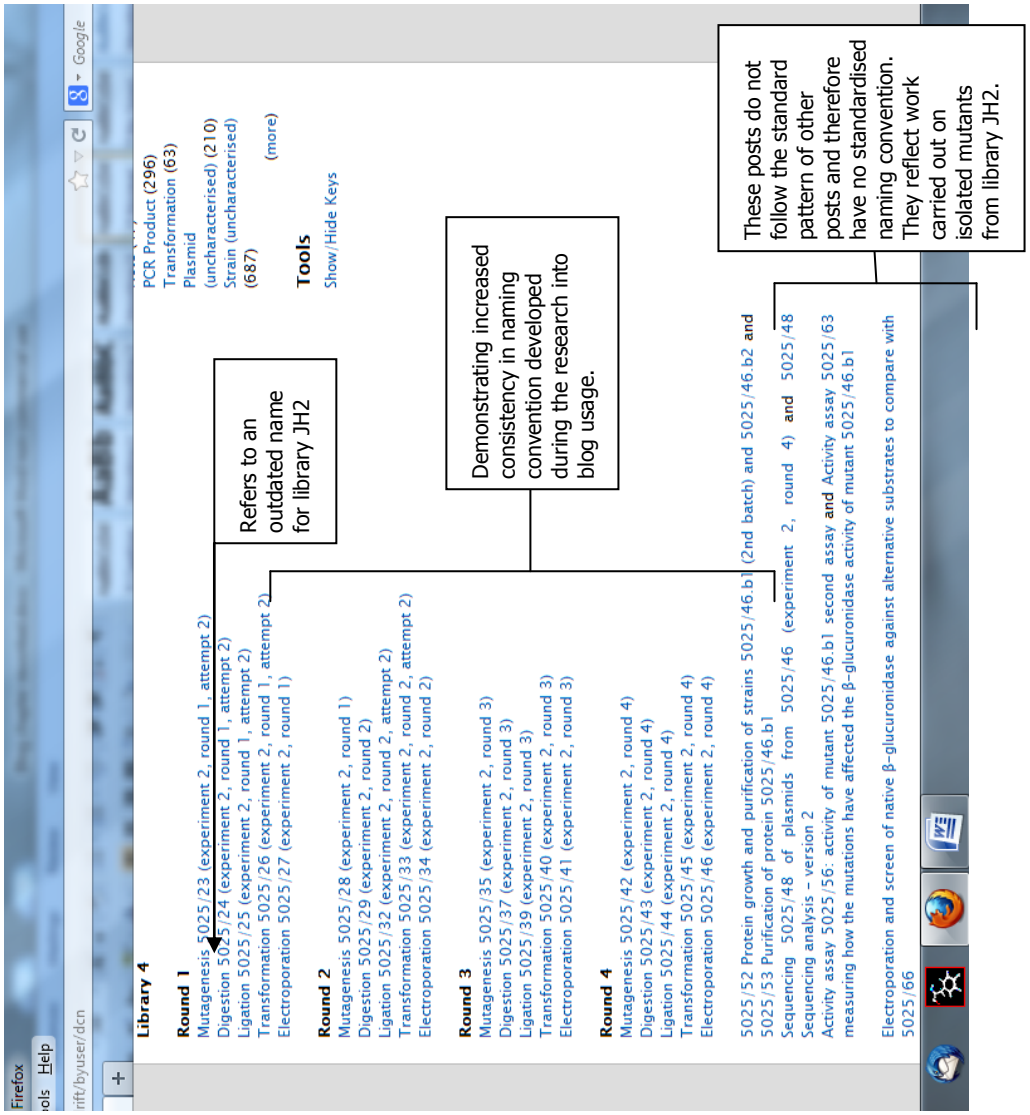


Figure 4.11b: the lower part of the blog post shown in 4.11a (which was not visible on screen) detailing all the practical work referenced in a paper submission. The work is easier to follow in this section due to the introduction of a more systematic naming convention.

One of the most important advantages personally found using the blog has been the ability to attach almost limitless data to a post. This has been particularly useful for associating a photo of a particular Petri dish of colonies with the pot of transformant that it originated from. This is an advantage of the one pot-one post system and the traceability of samples back to the raw starting materials. It has the ability to highlight whether overall a procedure didn't work or whether an individual sample failed. It also provides a visible guide to what the written result described e.g. what a plate containing 400 colonies actually looks like.

In contrast in the paper notebook there was limited space for adding results and so maybe only one photo summarising a whole set of plates would be glued in with no reference in the text as to how many transformations were carried out and how each plate was related to a certain sample. Also as Southampton use carbon copy duplicate books as the laboratory notebook two

copies of every photo would be required: one to stick in the copy book and one to associate with the detachable sheet.

Another advantage is that a photo can be viewed in colour on the blog whereas the photo would have been printed in black and white (by necessity due to available printers) to be glued into the notebook. As much of the work has relied on distinguishing between blue and white bacterial colonies, the ability to see in colour is very important to any observer looking at the results.

This ability to attach photos has also been useful for posts which have produced a gel image. The gel image can be directly associated with the experiment that it related to and could be annotated directly on the blog as LaBLog provides functionality for annotating images. Again it means the gel image doesn't require printing and sticking into the notebook, and notes can be made overlaying the gel image perhaps highlighting an odd result or noting features of the gel. This was harder to do when appearing as a black and white print out as pen could be difficult to read over black areas of the image.

The ability to accept data of any file type has allowed data generated from many different sources to be added. For example the raw Microsoft Excel files from the plate reader, and FASTA files and PDF from sequencing. These would have had to be listed as a computer file reference in a paper lab book as it wouldn't have been possible to print out and add the results. On the blog files of any type can be opened if the computer has the appropriate software. The files can also be downloaded too if a copy is required locally on a computer.

A further advantage of having files stored remotely on the blog is that when properly backed up they are safe from destruction. At one point, the gel imager computer in the lab crashed irreparably and required reformatting before Windows could be reloaded. This led to the loss of all the gel images stored on it. By backing all the files up on the blog, these gel images were not lost.

Another advantage of using the blog has come from the increased rigidity of categorising data. Before the blog was started, and in the early stages of blog use, there was no consistency to how an experiment was named. By devising a more rigid framework for naming posts and procedures systematically it encouraged better, more stringent, record keeping. For example by introducing an attempt number, it was much easier to keep track of where work was progressing and where it was failing. It was also easier to see where the experiment needed to be restarted from to reach the same point again. This is because certain procedures use almost the entire sample prepared in the previous procedure whilst others use only a small amount. This can be demonstrated using the summary of experiments for the preparation of library JH2 (shown previously in figure 4.11b). In round one the mutagenesis, digestion, ligation and transformation were all listed as attempt 2, but the electroporation is listed as the first attempt. This suggests

that something had gone wrong in the first transformation, but in this case it required starting from the beginning with another mutagenesis reaction and working through to reach the same point again. In round two only the ligation and transformation was listed as attempt 2. In this instance more ligations could be set up from the digestion product so the entire round did not need to be restarted.

4.2.7 Blog security and comparisons with other systems

In the introduction, the issue of security of work was discussed. It is a concern of many scientists that work remains private and secure to prevent potential issues with filing for Intellectual Property and patents, as well as to prevent "scooping". As a result many of the ELNs are closed systems and have digital signatures built in for verification to aid IP claims. The LaBLog system is able to address these concerns by providing security permissions for a blog. Though the aim of using a blog was to show the benefits of open notebook science, the developers have understood the need for work to remain protected and private where necessary. As a result individual blogs of the LaBLog site can be given different levels of security. Some blogs on the site are accessible to be read by anyone, and allow all registered posters to post, whilst others are set to be available to certain people only. The variation in security is identified by padlock symbols on the blog menu page (figure 4.12).

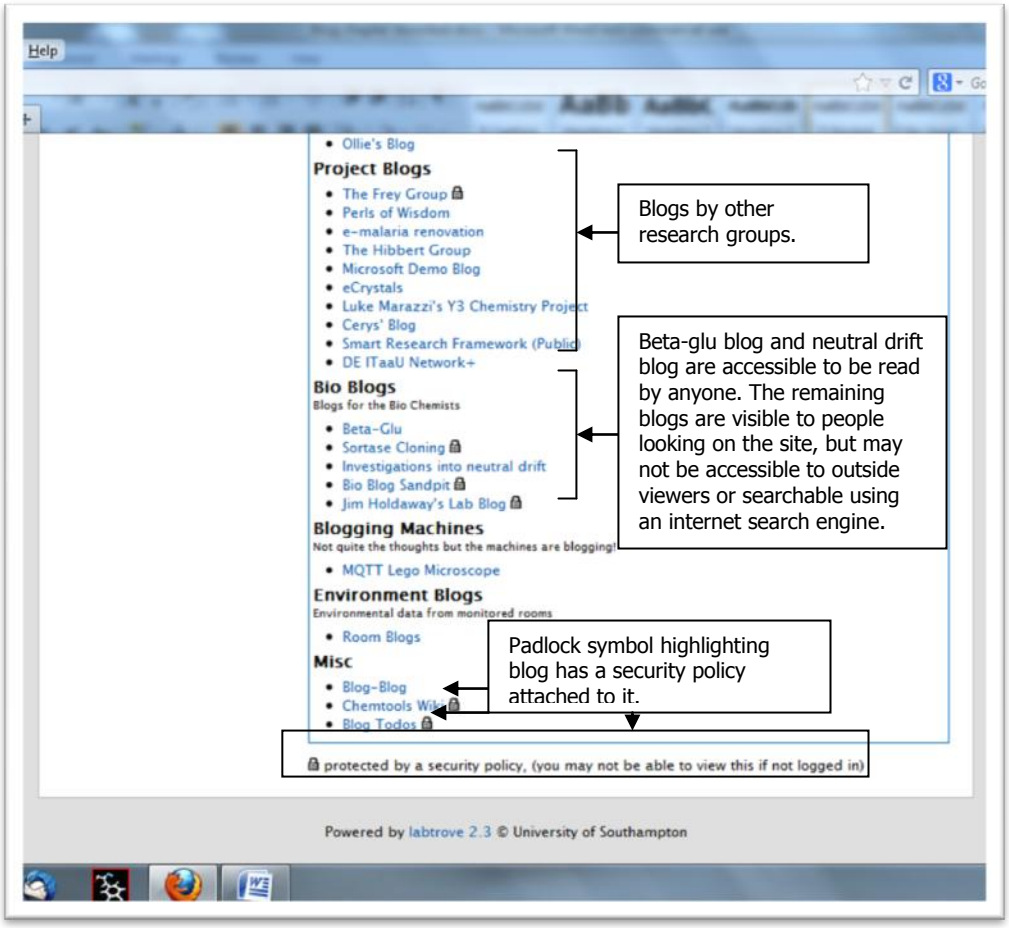


Figure 4.12: screenshot of part of the contents page of the LaBLog system highlighting blogs that are open and those which have higher levels of security. The screenshot was prepared

whilst logged out of the LaBLog system to prevent the titles of other blogs with higher levels of protection from being seen.

Furthermore, all posts have records of any edits and changes. Each time an edit is made, the reason for that edit must be provided before altered posts can be submitted. The blog also provides control as to who can make the changes (usually only the post writer) and a record of all versions of the post is kept by the database with revisions of a post being accessible under certain circumstances. To allow a revision to a post is the digital equivalent to striking through a mistake on a page in a paper notebook. Mistakes occur in blog posts for many reasons – typographical errors and genuine mistakes in reporting. In an ordinary paper notebook such mistakes would be struck through and possibly initialled to acknowledge the change. Allowing revisions to a post has the same effect.

Security on all blogs, even those readable by the public, is maintained by only allowing registered users to comment and post on blogs. This prevents malicious alterations being made to posts and keeps comments to those users with a legitimate interest in the work. Having these systems of security introduces a degree of flexibility whilst addressing genuine concerns regarding intellectual property and patents.

Overall the blog has many advantages, though a number of disadvantages remain. These weaknesses stem from the data structure being not fully correct and some design flaws of the LaBLog system regarding the way in which data is added. Overall the use of the blog has brought benefits over the paper system through increased storage space for associated data, and clear linking together of procedures and products to follow through the entire course of an experiment.

It is not possible to directly compare the LaBLog system with the proprietary ELN systems available. The systems are closed and most are pay to use so access is limited. The features of these programs that are known about have been researched from the published information on the websites of the companies providing these systems.
The extremely freeform nature of the blog does have an advantage over ELN systems in that it can be used in any context. As was demonstrated in the introduction, many of the ELN systems available tend to be biased towards one type of science, even if they are marketed as being generic. An example of how generic the LaBLog system is could be to consider how the system could be used in the context of an organic synthesis laboratory. The system could be used to write up a chemical synthesis procedure and link to the products and reagents used in that synthesis. Again this would be a highly useful system as the blog could act as a repository of chemicals held by a research group; hold safety data related to the chemicals and the procedures, and also allow a whole series of syntheses to be followed through. The synthesis of many compounds is a multi-step procedure that would benefit from being linked together. The

blog would also be able to hold other data including chemical structure images, NMR spectra, Mass Spectra, IR spectra, and X-ray crystallography data all linked in one place and attached to the post for the relevant compound.

4.2.8 The future of the blog

The future of the blog could take several directions. Improvements could be made to the current blog system taking account of improved technology and following up on user suggestions; a brand new blog system could be built on new or different technology, and finally further investigations could be made to allow the system to be used on other forms of technology, particularly hand held devices. In this section these directions are considered.

4.2.8.1 Improvements to the current blog

As has previously been mentioned, the use of templates and certain features of the blog require a certain amount of knowledge of html and BB Code. A user with limited computer knowledge and/or experience being presented with a new blog may struggle to begin using it, particularly to prepare templates. One solution may be to have some templates ready installed on the blog. However, this immediately makes the blog become geared towards a certain discipline, which defeats the purpose of the blog to be completely suitable for any discipline.

An alternative solution that could be considered is to add a selection of blank tables to the blog which the user could edit, adding a smaller number of markers to each cell of the table in order to turn it into a functioning template for the future. A list of the types of codes likely to be required such as [[box]] or [[metadata key:metadata value]] could be provided with and on the blog, and would not be too difficult to learn. As a user gains experience and knowledge of the code, preparing custom made templates would become easier.

A further alternative, and advantage to any blog user, even an experienced blog user could be to develop a template maker that is similar to the function in Microsoft Word for making a table, where the number of columns and rows for the table is selected by the user and the table then appears ready to be edited. In the same manner as the previous alternative it is then only the cell markers that would need to be added. A template maker which provides this function would also be an advantage as it would save time by removing the step of writing the code for the table itself.

Some minor alterations to the blog would improve the experience of using the blog. The first of these would be to add a notification system within the blog to alert a user when a new post or comment has been added. Notification is currently possible by using an external RSS feed, but the use of an external feed may not be practical for all users as it requires the blog to be viewed through a separate interface, which shows all the posts that have been added or updated, even those made by the poster. Certain RSS feed viewers, such as Yahoo Pipes, can be configured

to ignore the originators own posts by creating a filter. However, this may be beyond the computing experience of some blog users.

A notification system within the blog has many advantages by removing the need to view the blog through multiple interfaces and set up extra software tools to filter out unnecessary feeds. This keeps it simpler overall for users with less knowledge of available technologies. The technology to have an internal notification system is possible as most forums which are also run on database software have such a notification system, which can take the form of a coloured spot by the title of the post, or a change in colour of the post title.

Another small change that would make a big difference to the usability of the blog would be the addition of navigation arrows at the top of the page. Currently there are navigation buttons at the bottom of the page. The navigation buttons need to remain at the bottom of the page so that the page does not need to be scrolled up after searching all the posts on that page, but if navigation buttons were installed at the top of the page then it would no longer be necessary to scroll down each page when searching through a long list of posts when the user knows the post they are looking for is not to be found on that page.

One other problem with the blog is that despite tags linking posts together it can be difficult to follow the overall picture of an experiment or how certain posts link together. A useful tool to add which would enhance the blog would be a ‘web diagram’: a pictorial representation of all the posts and the links between them with the ability to click on the diagram and bring up selected posts. This picture would also allow identification of broken links and missed tags as these posts would show up as being isolated. A potential web diagram design showing the sorts of posts and the links between posts can be seen (figure 4.23).

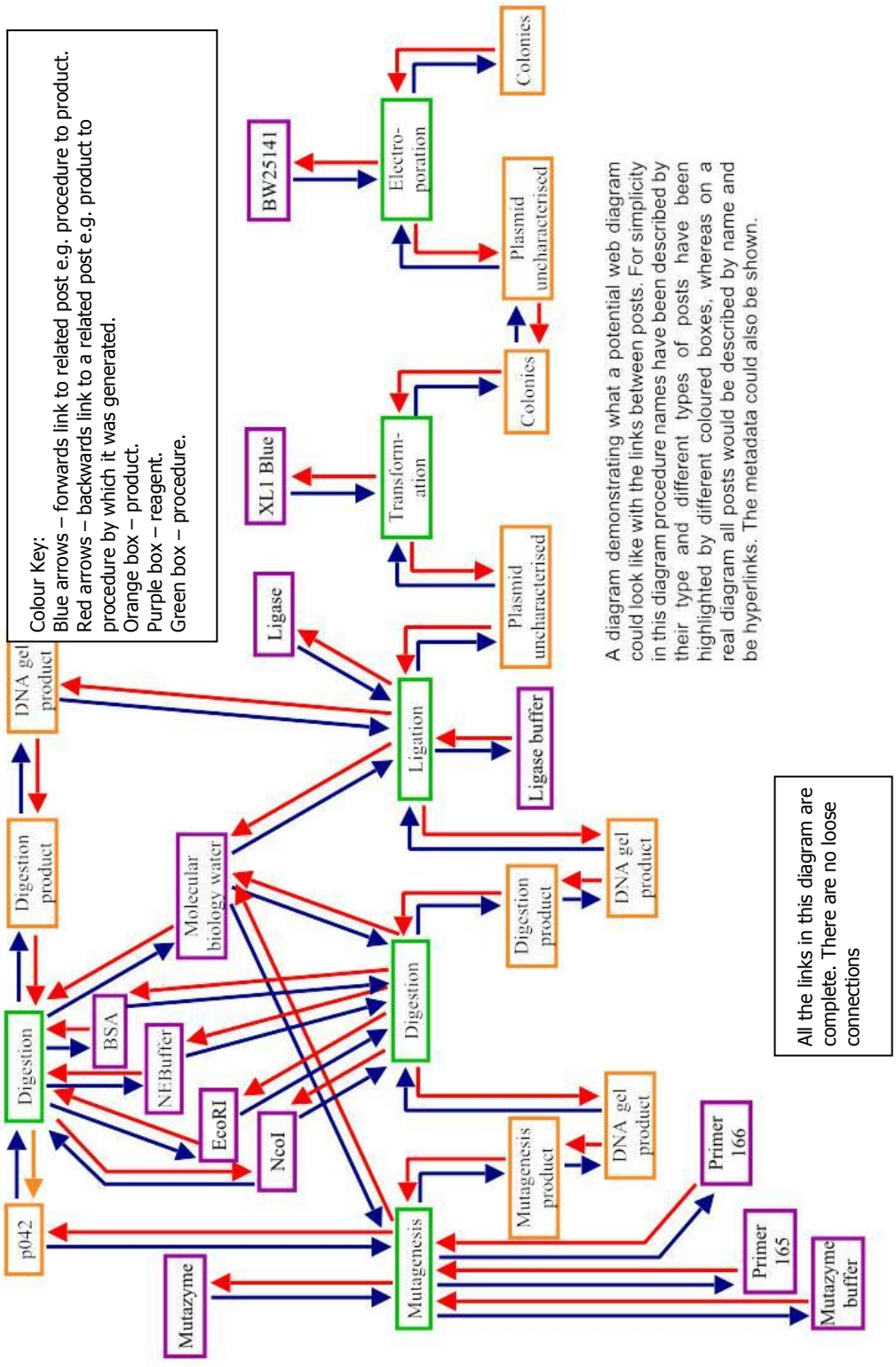


Figure 4.23: a suggestion for a web diagram in which blog posts are shown with the posts that they connect to. It could be generated from a single blog post or requested to get a general overview. Isolated or partially isolated posts would highlight missing links. The diagram would not necessarily be colour coded as some items can act both as reagents and products.

4.2.8.2 Adapting LaBLog for use on new technologies

One potential study for the future could be to study the feasibility of regular use of LaBLog on smart phones and hand held computers. Whilst tablet PCs are smaller and more portable than laptops and desktops, a smart phone or handheld computer is even smaller for easy portability, and are becoming more widely used. Many of the smart phones also include technologies such as WiFi, data connections and reasonable quality digital cameras which could allow work to be videoed or photographed and directly uploaded to the blog without first being transferred to a computer. An example of how this could be useful is shown by the neutral drift project which generated plates of blue and white bacteria. These were photographed on a conventional camera and the photos transferred to computer before being uploaded to the blog, whereas a smart phone with a camera of reasonable standard could have directly taken and uploaded the photographs to the blog. Also the ability to photograph or video any aspect of the work – even the laboratory itself, and upload it straight to the blog without an intermediate step would be very useful. It is far more practical in this circumstance to grab a mobile phone than always have a camera ready for use in the lab.

The introduction of the Apple iPad[34] and other tablets based on the Android operating system provides another means of portable computing and a device which is conveniently sized for using in a laboratory. Both iPads and tablets are smaller and lighter than a tablet PC, but larger than a handheld computer or phone. Like smart phones they also usually incorporate a camera. This provides the functionality of a smart phone with the advantages of a larger screen area to read and use the blog, as well as wireless portability allowing it to be taken to anywhere within the lab. Tablets and the iPad are growing more popular, particularly as they decrease in price. The cost of an iPad can match that of a high specification laptop[35,36] and the expenditure may not be considered cost effective compared to the extra features available on a laptop.

Despite these potential problems and the improvements that it would be desirable to make, the blog is currently a complete product that could be used in any scientific laboratory context with just a small amount of time required for blog setup including template/table design and organisation.

4.3 Conclusions

4.3.1 Key features of LaBLog

The LaBLog system is a unique electronic laboratory notebook system that has been developed to fit with the ideals of open notebook science, whilst retaining the ability to create closed notebooks which require login to view. LaBLog is designed to be completely freeform and useable by any scientific discipline. Features of the blog include the ability to add posts either through the freeform new post box or a pre-designed template. Once added, posts can be edited if mistakes are made. This is a move away from other systems and is because all revised

versions of a post are saved and stored in the database. Additionally revisions cannot be made without providing a reason, and all revisions of a post can be accessed. LaBLog can store pictures and data files, which can also be linked to the main body of the text. Additionally, picture files can be annotated using the specially designed annotation tools. LaBLog also has a search function which provides search and retrieval of old experiments, and details of previous batches of chemicals.

4.3.2 Key stages of development from a front-end perspective

LaBLog has been developed in three main phases and across two blogs. During phase one the blog (beta-glu) was first being used in a laboratory context and three main points of importance were discovered:

- A table is a good way to display molecular biology experiments as it provides a lot of information in a clear and concise manner.
- Tables involve a time penalty in preparation thus templates for the main types of reactions would be a good feature to have for preparing large numbers of similar posts.
- For the blog to function efficiently, the data structure must be accurate. Data, which takes the form of individual posts, must be correctly divided into sections and accurately described with correct, sensible metadata.

Through these points of importance it was realised that the data structure of the beta-glu blog was wrong; which was why the blog was becoming less practical to use. Realising that the data structure was wrong led to in depth consideration of how the data should be organised. As a result the one pot-one post method of describing the data structure was developed, which gave each type of material, product and procedure its own identity through individual posts. Additionally the first templates were also envisaged in which reagents, and products could all be linked together through the procedure posts; whilst the overall research could be linked together by showing how the products of one procedure could become the reagent for the next procedure.

Phase two brought in the new blog ‘investigations into neutral drift’ which introduced the new data structure of one pot-one post. The way posts were categorised and assigned metadata was different from the outset in order to satisfy the one pot-one post structure. During this phase the key points of importance were that:

- The metadata structure is vital to preparing templates that are functional. The new metadata structure for posts fitted the research much more accurately.
- Using one pot-one post can create problems with deciding what constitutes an experiment and when something needs its own post. It is not immediately obvious what one experiment actually is.

- Separation of procedures into smaller component parts, or repetition of identical procedures to fit one pot-one post increased the burden on the user and made the research more difficult to follow.

Through these points of importance it was highlighted that one logical experiment is not necessarily just one procedure. For example whilst a mutagenesis reaction might in reality consist of the mutagenesis followed by gel electrophoresis to purify the product, the two parts go hand in hand and when split up make less sense. It also does not make sense to write up an identical procedure multiple times to satisfy the one pot-one post approach when the procedure is carried out in parallel. It is more logical to have one procedure post taking multiple starting materials and linking to multiple products. Once the logistics of what constitutes an experiment and a procedure were solved, and how the metadata related to the data, it also became much easier to define how a template should be organised.

Phase three saw the introduction of functioning templates with increased automation. The templates were finally optimised, though not quite accurately reflecting the structure of all work occurring in the laboratory due to one small error in structuring the data. This phase also saw a complete transfer from using both paper and blog, to blog alone for recording the work. Hindsight has shown that there is a better method for describing the data which would eliminate the error. Systematic post naming was also devised during this phase, which provides a more consistent method of assigning post names and allows for easier selection of relevant posts from the drop down boxes within templates.

Through the process of user centred design it has become apparent that at the beginning as users we didn't know how to structure the data. Through a process of trial and error the data structure began to fall into place. All of the changes to the blog – tables, templates, drop down menus, links between posts, naming conventions and minor changes to the user interface made it easier to link data together or to tag data with information. These changes helped identify the relationship between pieces of data, which was demonstrated with links and the relationship between data and what that data represented, which was demonstrated with tagging. By solving the data structure, the laboratory research became strongly linked together and a much clearer work record kept. Combined together, these features all lead to a powerful and unique product, different to any other form of ELN, blog or wiki used for recording scientific experiments.
There is still potential for development of the blog to increase the ease of use, and to remove some bugs and problems. However the blog is completely usable in its current form.

4.3.3 Coda

Since the original preparation of this chapter, it has been noted that a number of the suggestions listed in the future work section have now been implemented. Therefore a user or visitor to the blog will now find navigation arrows at both the top and bottom of the screen.

Additionally there has been a change to the way templates are used and prepared which reflects the wish for a Microsoft Word style table maker for templates. A number of other improvements and changes have also been made. Therefore a number of points discussed in this chapter are now largely irrelevant to the blog in its current form. However they are important to discuss in the context of what occurred at the time and the impact that the blog had on research.

4.4 References

1. Hughes, G., Mills, H., De Roure, D., Frey. J. G., Moreau, L., Schraefel, M. C., Smith, G., and Zaluska, E.; *The semantic smart laboratory; a system for supporting the chemical eScientist*; Organic and Biomolecular Chemistry; 2004; **2**; 3284-3293.
2. Taylor, K.; *The status of electronic laboratory notebooks for chemistry and biology.;* Current Opinion in Drug Discovery and Development; 2006; **9**(3); 348-353.
3. Iyer, R., and Kudrle, W.; *Implementation of an Electronic Lab Notebook to integrate research and education in an undergraduate biotechnology program*; Technology Interface International Journal; 2012; **12**; 5-12
4. Schraefel, M.C., Hughes, G. V., Mills, H. R., Smith, G., Payne, T. R., and Frey, J.; *Breaking the Book: Translating the Chemistry Lab Book into a Pervasive Computing Lab Environment*; CHI; 2004; **6**; 25-32.
5. Sandercock, C. G.; *In Focus: Understanding the legal issues surrounding electronic laboratory notebook (ELN) records*; 2008; http://www.contur.com/customers/infocus.asp#article2; Accessed 08.07.10 (Archived by WebCite® at http://www.webcitation.org/5r4i6l8dh).
6. ConturELN; http://www.contur.com/shared/pdf/ConturELN_Case_study_Academic.pdf; Accessed 08.07.10 (Archived by WebCite® at www.webcitation.org/5r4sLjvCQ).
7. LABTrack; http://www.labtrack.com/Tab1b.html; Accessed on 08.07.10 (Archived by WebCite® at http://www.webcitation.org/5r4mXBRPz).
8. Schraefel, M. C., Hughes, G. V., Mills, H. R., Smith, G., and Frey, J.; *Making Tea: Iterative Design Through Analogy*; DIS; 2004: 49-58.
9. Morris, C., Pajon, A., Griffiths, S. L., Daniel, E., Savitsky, M., Lin, B., Diprose, J. M., Wilter da Silva, A., Pilicheva, K., Troshin, P., Van Niekerk, J., Isaacs, N., Naismith, J., Nave, C., Blake, R., Wilson, K. S., Stuart, D. I., Henrik, K., and Esnouf, R. M.; *The Protein Information Management System (PiMS): a generic tool for any structural biology research laboratory*; Acta Crystallographica; 2011; **D67**; 249-260.
10. OED; http://www.oxforddictionaries.com/view/entry/m_en_gb0975445#m_en_gb0975445; Accessed 08.07.10 (Archived by WebCite® at www.webcitation.org/5r4vUX0dh).
11. Wikipedia; http://en.wikipedia.org/wiki/History_of_wikis; Accessed 08.07.10 (Archived by WebCite® at www.webcitation.org/5r4vrWyZ8).
12. myExperiment; http://www.myexperiment.org/; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5jr8d5J).
13. OpenWetWare; http://openwetware.org/wiki/Main_Page; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5jao0iM).
14. Wikispaces; http://www.wikispaces.com; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5m0Xb4P).
15. Usefulchem; http://usefulchem.wikispaces.com/All__Reactions; Accessed 09.07.10 (Archived by WebCite® at http://www.webcitation.org/5r5mNHMgl).

16. Gus Rosania; http://1cellpk.wikispaces.com/; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5mgnlQX).
17. Wikipedia; http://en.wikipedia.org/wiki/History_of_blogs; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5nWVXuR).
18. Talbott, T., Peterson, M., Schwidder, J., and Myers, J. D.; *Adapting the Electronic Notebook for the Semantic Era*; IEEE; 2005; 136-143.
19. Drake, D.; *ELN implementation challenges*; Drug Discovery Today; 2007; **12**; 647-649.
20. Bradley, J-C.; http://en.wikipedia.org/wiki/Open_notebook_science; Accessed 08.07.10 (Archived by WebCite® at www.webcitation.org/5r4uULx08).
21. Bradley, J-C.; http://drexel-coas-elearning.blogspot.com/2006/09/open-notebook-science.html; Accessed 09.06.10 (Archived by WebCite® at www.webcitation.org/5qM7gfNKs).
22. Twitter; http://twitter.com/; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5kFNRFh).
23. Friendfeed; http://friendfeed.com; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5kQBG8q).
24. Brumfiel, G.; *Breaking the convention?*; Nature; 2009; **459**; 1050-1051.
25. *How to stop blogging*; Nature; 2009; **460**; 152.
26. Nature Chemical Biology; 2009; **5**; 601.
27. MacArthur, D.; http://scienceblogs.com/geneticfuture/2010/05/what_a_difference_a_year_makes.php; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5lcyt8Z).
28. Lisi, G.; http://www.deferentialgeometry.org/; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5nDzM56).
29. Redfield, R.; http://rrresearch.blogspot.com/; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5o6DJpe).
30. Barton, M.; http://www.michaelbarton.me.uk/research/; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5oGUimz).
31. Boettiger, C.; http://www.carlboettiger.info/index.html. Accessed: 25.02.13. (Archived by WebCite® at http://www.webcitation.org/6EhapVLCW)
32. Nielson, M.; http://michaelnielsen.org/blog/; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5ok1bjR).
33. Nielson, M.; http://michaelnielsen.org/blog/connecting-scientists-to-scientists/; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5otYb96).
34. Hale, J.; Beta-glu; http://blogs.chem.soton.ac.uk/beta_glu; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5pL3R6S).
35. Hale, J.; Investigations into neutral drift; http://blogs.chem.soton.ac.uk/neutral_drift; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5pb1mjr).
36. Neylon, C.; Sortase Cloning; http://blogs.chem.soton.ac.uk/sortase_cloning; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r5yy97GC).

37. Apple; http://www.apple.com/uk/ipad/; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r61GgXZcw).
38. Apple store; http://store.apple.com/uk/browse/home/shop_ipad/family/ipad?mco=OTY2ODA0NQ; Accessed 09.07.10 (Archived by WebCite® at www.webcitation.org/5r61qOsEs).
39. Ebuyer.com; http://www.ebuyer.com/store/Computer/cat/Laptops; Accessed 09.07.10.

Chapter 5 - The Open Notebook Science Challenge: a comparison between blog and wiki based open notebook science systems

5.1 Introduction

In the previous chapter regarding the development and testing of an electronic laboratory blog (LaBlog), the concept of Open Notebook Science was introduced along with examples of the types of ONS that are being practiced including laboratory blogs and wikis. Blogs are a particular type of website which are maintained usually by a single person and include commentaries, pictures, and news, whilst others may function as an online journal or diary. Wikis are collaborative websites managed by communities of people. This chapter looks in more detail at a specific example of ONS in practice and how the social and scientific challenges of managing the data differ to those found on the LaBlog system.

The Open Notebook Science Challenge[1] is a specific example in the use of Open Notebook Science. The challenge was launched in September 2008[2] by Jean-Claude Bradley and collaborators. The challenge is to develop methods for the measurement of the solubility of organic reagents in non-aqueous solvents; and to record the measurements through Open Notebook Science capturing as much detail as is considered important. The nature of this detail may be extra data, or a wider variety of data than that scientist would usually capture.

The reasoning for choosing to measure solubility is that there are very few open databases and sources of solubility information for solubility in other solvents besides water, despite most organic reactions being carried out in non-aqueous solvents. A search of the literature also yields few instances of the explicit measurement of solubility, though some literature articles and databases make reference to aqueous solubilities. Other literature will provide non-quantitative references to solubility in non-aqueous solvents including freely soluble or sparingly soluble. This challenge provides not only a valuable service in reporting solubilities in a freely available form, but shows full experimental workings allowing other labs to use the method and corroborate results. The results are also being added to a free chemistry search engine and database called ChemSpider[3].

The importance of this work was recognised by the scientific community with the addition of sponsorship from Submeta for ten prizes of $500 to be awarded to contributors to the challenge who were considered to have made a contribution worthy of recognition. Additionally Sigma-Aldrich sponsored chemicals, and Nature sponsored three prize winners a yearlong subscription to the journal. Similarly a repeat of the challenge was held in 2010 with five prizes of $500 from the Royal Society of Chemistry[1].

The compounds that measurements are being focussed around are the components of the Ugi reaction: carboxylic acids, amines, isocyanides and aldehydes or ketones. The Ugi reaction is being used by the Bradley laboratory as part of the synthesis of potential anti-malarial compounds. However other types of compounds have been measured as well during the challenge.

The results of the Open Notebook Science Challenge are recorded on a wiki[1] hosted by wikispaces.com, which is the same hosting service as used by the Usefulchem wiki referenced in the previous chapter. The data organisation of the wiki will be discussed in greater detail during this chapter.

5.2 Results and Discussion

5.2.1 Solubility measurements – results generated during the challenge

There have been a vast number of solubility measurements generated during the ONS Challenge. These have been generated by various different people. The following section details the results of measurements personally contributed to the challenge. These included measurements of solids and liquids at room temperature, as well as a study on the effects of temperature on solubility. In general, solubility for solid solutes was calculated by preparing a saturated solution of the solute in the given solvent, drying a set proportion of the supernatant and calculating the solubility from the amount of residue remaining (table 5.1). Small changes to individual experimental methods were made to adapt to circumstances: for example, changing size of measurement vessel due to a solute being so soluble that the solution volume expanded beyond the maximum volume of the vessel. Also some measurements were not possible due to effects of the solvent reacting with the solute, and in one example, the solvent reacting with the plastic measurement vessel. Additionally the method was optimised to improve the reliability and ensure that the solutions were saturated. The solubility of liquid solutes was calculated from the density to provide the maximum solution concentration (table 5.2). The measurements listed in tables 5.1 and 5.2 only refer to measurements for that compound obtained personally. Accessing the database of measurements on the challenge wiki demonstrates that some compounds have many more measurements made for them, and a large number of compounds which are not listed here.

Solute Name	**Maximum solubility of the given solute in ethanol**	**Maximum solubility of the given solute in methanol**	**Maximum solubility of the given solute in THF**
Glucose	0.17 M 0.013 M 0.0011 M	0.056 M 0.052 M 0.02 M	0.008 M 0.017 M 0.0022 M
Mannitol	0.016 M 0.011 M 0.01 M	0.016 M 0.017 M 0.005 M	0.011 M 0.027 M 0.01 M
NaCl	0.257 M 0.017 M 0	0.017 M 0.188 M 0.006 M	2.31 M 0.060 M 0.003 M
Vanillin	2.23 M*	3.197 M*	3.29 M*
Sodium n-octanoate (caprylic acid sodium salt)	0.07 M	0.55 M	0.006 M

2-hydroxypropane-1,2,3-tricarboxylic acid monohydrate (citric acid monohydrate)	1.78 M	2.27 M	1.52 M
2-oxopentanedioic acid (2-oxoglutaric acid)	0.89 M* 2.99 M	2.48 M* 5.02 M	1.54 M* 3.12 M
3-(4-hydroxy-3,5-dimethoxyphenyl)prop-2-enoic acid (sinapic acid)	0.08 M* 0.79 M	0.11 M* 1.04 M	0.44 M* 0.45 M
4-methylbenzoic acid (p-toluic acid)	0.76 M* 0.79 M	0.72 M* 0.84 M	1.46 M* 1.75 M
Benzotriazole-5-carboxylic acid	0.03 M* 0.042 M	0.03 M* 0.039 M	0.02 M
8-bromo-octanoic acid	3.75 M	3.28 M	3.56 M
4-dimethylaminocinnamaldehyde	0.05 M 0.059 M	0.074 M 0.073 M	0.42 M
2-(4-hydroxyphenyl) azobenzoic acid	0.04 M 0.126 M	0.067 M 0.045 M	0.55 M
4-hydroxyphenyl pyruvic acid	0.711 M	0.907 M	-
4-nitrophenylacetic acid	0.705 M	1.286 M	-
1-adamantylamine	0.82 M	0.56 M	-
6-aminocaproic acid	0.003 M	0.006 M	-
10-chloro-9-anthraldehyde	0.004 M	0.003 M	0.076 M
4-hydroxybenzaldehyde	2.396 M	3.508 M	0.909 M* 3.58 M
4-hydroxy-3-methylbenzaldehyde	2.146 M	2.782 M	3.052 M
Benzoic acid	2.5 M	2.86 M	2.81 M
Undecanoic acid	3.94 M	4.0 M	-
Citric acid (anhydrous)	1.86 M	3.08 M	1.80 M

Table 5.1: solubility measurements of solids obtained personally during the challenge. Measurements marked * are not likely to be accurate due to the technique used to attain solution saturation and in many cases have been marked as “DO NOT USE” on the wiki solubility database.

Solute Name	Maximum solubility calculated from the density of the solute
3,4-dimethylbenzaldehyde	7.54 M
3,5-dimethylbenzaldehyde	7.44 M
4-ethylbenzaldehyde	7.30 M

Table 5.2: maximum solubility measurements of liquids obtained personally during the challenge, calculated from the density of the solute.

From the tables it can be seen that a number of the measurements are marked as being inaccurate (highlighted with * in the table). This is due to the methods used to prepare saturated solutions. In these methods a truly saturated solution was not obtained, which was caused by inadequate mixing: either from not mixing for long enough, or not using an adequate method of mixing e.g. shaking samples by hand or inverting them end over end. The methods used to prepare the saturated solutions were optimised through experiment, in order to ensure that full saturation was achieved. This was done by changing from mixing samples by shaking to using a vortex to mix, and vortexing for an increased length of time.

A study of the effect of temperature on the solubility of vanillin was also carried. The study was carried out because it was noted that there were differences between the values obtained for the solubility of vanillin across different labs, and it was hypothesised that the reason for the difference was due to the ambient temperature in the different locations. The ambient temperatures were thought to be different as the locations are geographically different being situated in the USA and UK, but are also likely to be different due to room size, equipment in the labs generating heat, and the effects of heating and air conditioning systems. In order to test the hypothesis a study was carried out to measure the solubility of vanillin at room temperature (usually measured as 23°C), 30°C and 40°C. It was hypothesised that the solubility of vanillin would increase as the solvent temperature increased but would stop increasing as the solvent neared boiling point and vapour was lost. The measurements did show a trend of increasing solubility with temperature (table 5.3) and that it was not a linear relationship (figure 5.1), suggesting that the solubility would stop increasing as the solvent boiling temperature was reached.

Temperature at which solubility was measured.	Maximum solubility of vanillin in ethanol	Maximum solubility of vanillin in methanol
23° C	2.38 M* 2.58 M* 2.59 M* 2.48 M	3.98 M* 3.76 M* 4.00 M* 4.15 M
30° C	2.86 M* 2.80 M* 2.88 M* 3.06 M	4.14 M* 3.48 M* 3.69 M* 4.63 M
40° C	2.97 M* 3.40 M* 3.45 M* 3.37 M	3.91 M* 2.74 M* 3.71 M* 4.90 M

Table 5.3: solubility measurements obtained for vanillin at different temperatures. Measurements marked * are not likely to be accurate due to the technique used to attain solution saturation.

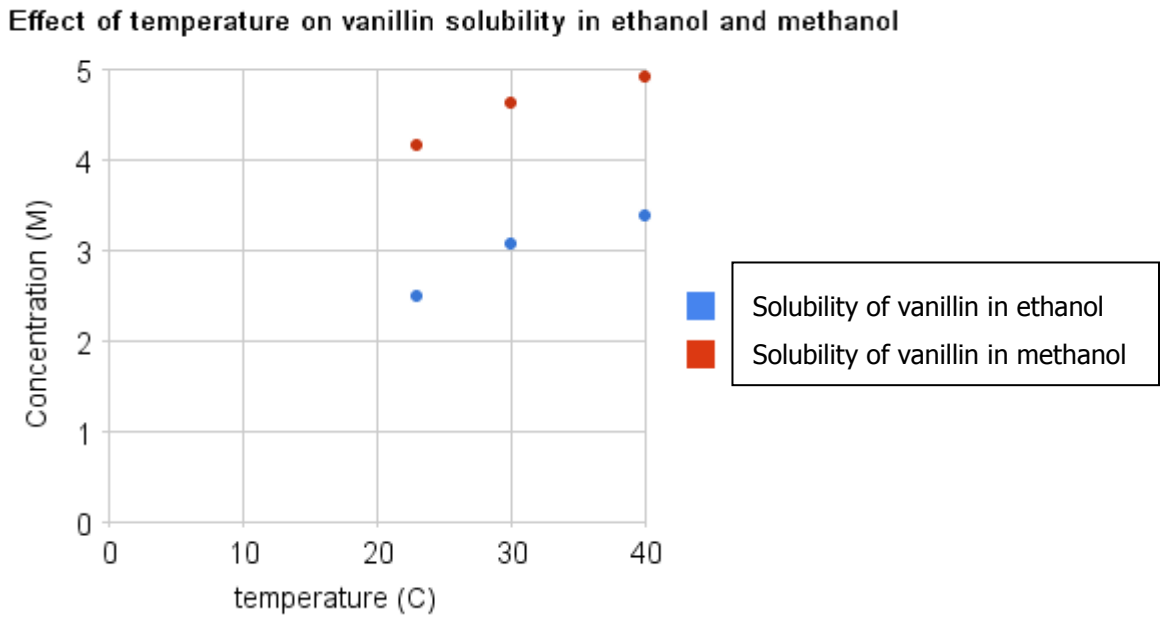


Figure 5.1: showing the trend in solubility with temperature of vanillin in ethanol and methanol (blue = ethanol, red = methanol)

A key part of the challenge has been in the demonstration of the advantages of using Open Notebook Science. There have been two examples involving measurements that were

personally obtained, which have demonstrated the power of collaboration and open notebook science in achieving results and highlighting errors. The first of these involves the measurement of the solubility of sodium n-octanoate, which was initially listed as caprylic acid/n-octanoic acid in the measurements. Another participant in the challenge had already measured the solubility of n-octanoic acid and knew it to be a liquid, whilst the measured compound was a solid. Collaborating on the wiki allowed the mistake to be highlighted and investigated (figure 5.2). The container was traced and revealed that it was the sodium salt: the container had sodium salt written in much smaller text underneath the acid name. Such a mistake may have gone unnoticed without the collaboration and the notebook being available for everyone to read.

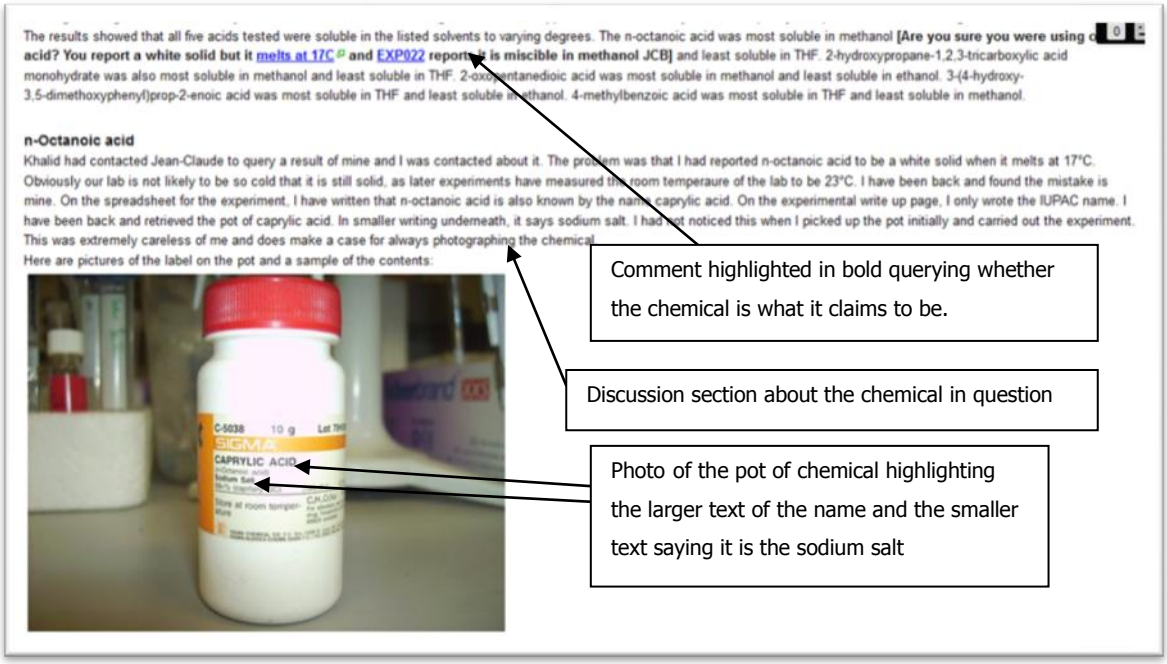
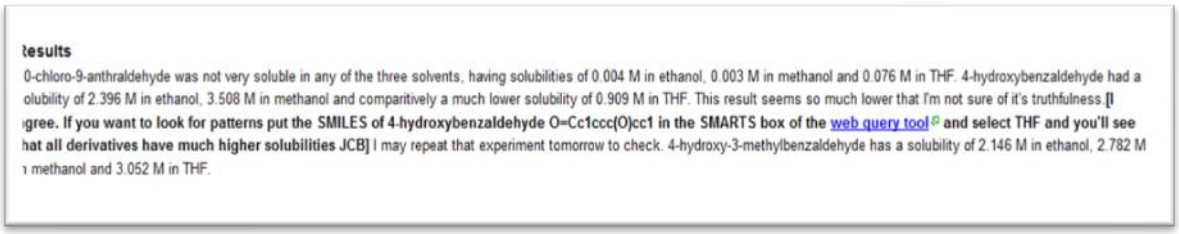


Figure 5.2: section of a screenshot for the experimental write up for an experiment on the wiki. In it, the state of matter of a chemical had been queried in relation to another measurement made elsewhere. The full page can be found at http://onschallenge.wikispaces.com/JennyHale-2.

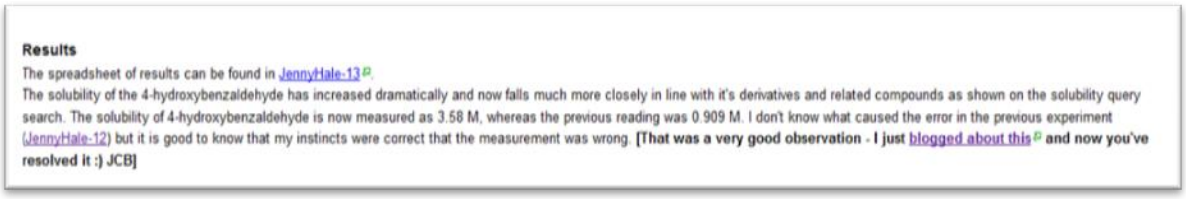
The second example is shown by the measurement of the solubility of 4-hydroxybenzaldehyde in THF. The measurement calculated the solubility to be 0.909 M. This result did not fit when compared to the results of the solubility in ethanol and methanol, nor with other measurements on compounds of similar structure (figure 5.3). This was highlighted by using a solubility search facility developed by a challenge collaborator which showed all the measurements for that compound calculated during the challenge. Repeating the experiment provided a result that was much more closely aligned to other results (figure 5.4). Having a collaborator able to develop tools for searching and querying measurements allows participants or any interested person to search the data quickly and effectively.



tesults

0-chloro-9-anthraldehyde was not very soluble in any of the three solvents, having solubilities of 0.004 M in ethanol, 0.003 M in methanol and 0.076 M in THF. 4-hydroxybenzaldehyde had a olubility of 2.396 M in ethanol, 3.508 M in methanol and comparitively a much lower solubility of 0.909 M in THF. This result seems so much lower that I'm not sure of it's truthfulness.**[I igree. If you want to look for patterns put the SMILES of 4-hydroxybenzaldehyde O=Cc1ccc(O)cc1 in the SMARTS box of the web query tool and select THF and you'll see hat all derivatives have much higher solubilities JCB]** I may repeat that experiment tomorrow to check. 4-hydroxy-3-methylbenzaldehyde has a solubility of 2.146 M in ethanol, 2.782 M n methanol and 3.052 M in THF.

Figure 5.3: a portion of a screenshot from the experimental write up measuring the solubility of 4-hydroxybenzaldehyde in THF. The whole experiment can be found at http://onschallenge.wikispaces.com/JennyHale-12.



Results

The spreadsheet of results can be found in JennyHale-13.

The solubility of the 4-hydroxybenzaldehyde has increased dramatically and now falls much more closely in line with it's derivatives and related compounds as shown on the solubility query search. The solubility of 4-hydroxybenzaldehyde is now measured as 3.58 M, whereas the previous reading was 0.909 M. I don't know what caused the error in the previous experiment (JennyHale-12) but it is good to know that my instincts were correct that the measurement was wrong. **[That was a very good observation - I just blogged about this and now you've resolved it :) JCB]**

Figure 5.4: a portion of a screenshot from the experimental write up re-measuring the solubility of 4-hydroxybenzaldehyde in THF. This time the measurement is in line with other measurements. The whole experiment can be found at http://onschallenge.wikispaces.com/JennyHale-13.

5.2.2 Comparing blogs and wikis: the similarities, differences and suitability for the science

As the ONS Challenge has had the results recorded on a wiki, it provides an opportunity to compare directly with recording results on the blog; as well as a wider comparison of blogs and wikis. Before the ONS Challenge was officially started, some initial measurements were carried out to check that the equipment available was suitable for the purpose and that the methodology would allow solubility to be calculated. As the wiki page was not yet operational, these initial measurements were recorded on the blog. These pre-challenge solubility measurements were carried out twice on glucose, mannitol and sodium chloride. The ability to write data to both the blog and the wiki allows direct comparison of the methods used to display the data and write up the experiments.

The organisation of blogs and wikis are quite different. A blog site may contain many blogs, but an individual blog usually has only one author. However, other authors may also contribute posts if the posts are directly linked to the project. In contrast a wiki site may contain many wikis with each individual wiki being used or managed by one or more people.

On accessing the wiki for the challenge, a guest or user reaches a home page which gives background information on the challenge. However, there is not an equivalent page on the LaBlog system. This has a 'dashboard' page which lists all the individual blogs, but no blog has a dedicated page describing the work it contains. This is not a criticism of either system and the homepage has a very important role on the ONSC wiki in describing what is going on for a visitor and potential collaborator. There is less need for this function on the LaBLog site, though there is potential for such a page to exist. From the homepage of the wiki there is a navigation bar on the left hand side which takes the user to a page called list of experiments. This is a form of contents page providing a list of the experiment titles and links to each individual experiment page. By contrast accessing a blog takes the user directly to a news feed page showing the titles and first few lines of each experimental post.

5.2.2.1 Pre-challenge solubility measurements posted on the blog

As has been stated, a small number of measurements were carried out before the challenge officially started to test the methodology and equipment available. The results were written on the blog as the wiki page was not operational.

The data structure on the relevant blog of the LaBlog system has been optimised for the molecular biology work that was the main research focus, and is based around a one pot-one post data structure. When presented with solubility measurements there was an issue of how to display the data as a post, or multiple posts, on the blog. The approach used to post solubility data on the blog was wrong. The solubility experiments were forced into the data structure for the directed evolution project even though they are very different. The one pot-one post structure is not necessarily wrong in this context. It makes sense to be able to track a sample through all the different stages of measurement. However, the implementation of the one pot-one post data structure should have been different. There were a series of mistakes – first the procedure was split up into three smaller procedures each given its own post. These were preparation of a saturated solution as one post, drying the supernatant as a second, and the third post calculating the solubility (figure 5.5). These should have all been one single procedure post. A second mistake was in trying to force the experiments into a template structure containing tables. By doing this, a lot of experimental detail was lost and it just didn't make sense. A third mistake meant the posts were not sensibly or properly linked together after they were created producing protracted and disjointed experiments.

The science for the molecular biology involved a series of linked procedures that were part of an overall chain of procedures; a wider project, with the product of each procedure being an independent product such as a length of DNA, a protein, or a bacterial colony that could be stored and used elsewhere. By contrast the products of each step in the solubility measurements were not the final 'correct' product. Only the piece of data giving the solubility was the final product. It was through not making the correct judgements about the data that the mistake in data structure on the blog occurred. It was, however, not overly important that the

data was structured wrongly on the blog as the major part of the challenge was to be carried out on the wiki and the only two experiments posted on the blog did not contain data fully relevant to the challenge. The data generated was recorded in the results during preparation of this chapter.

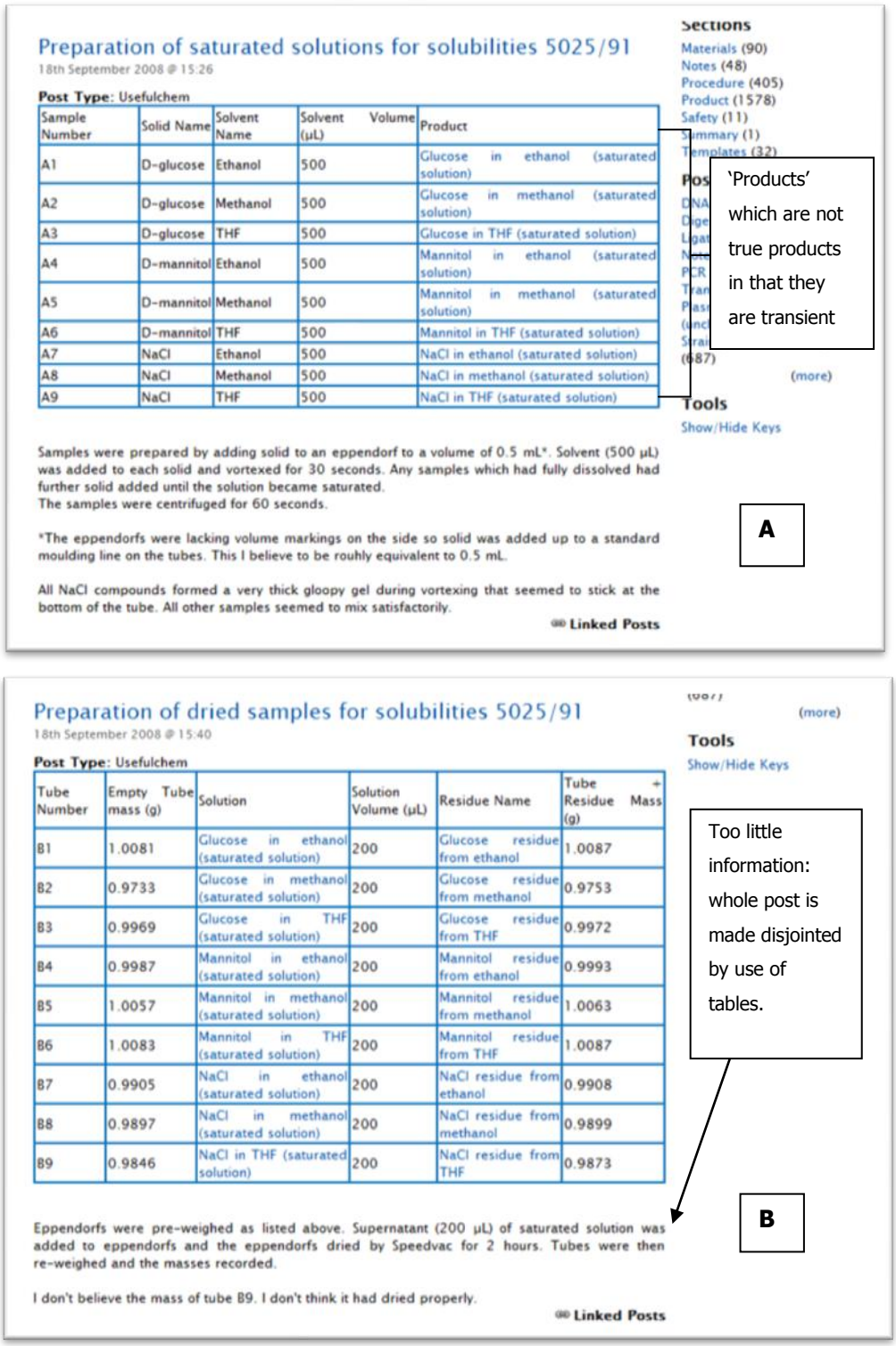


Figure 5.5: Parts A and B showing close ups of screenshots of the first two posts of the three posts that contribute to each solubility experiment on the blog. Full post pages available at http://blogs.chem.soton.ac.uk/neutral_drift/14831/Preparation_of_saturated_solutions_for_solubilities_502591.html and http://blogs.chem.soton.ac.uk/neutral_drift/14844/Preparation_of_dried_samples_for_solubilities_502591.html.

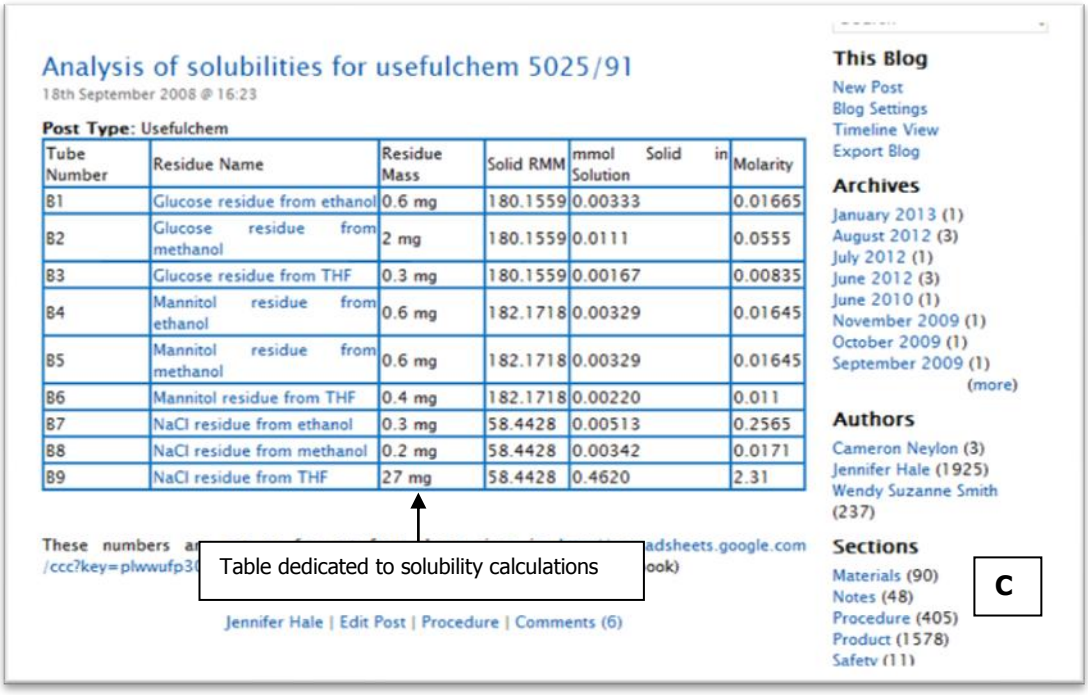


Figure 5.5: Part C showing a close up of a screenshot of the third post which makes up the complete set of posts for one solubility experiment on the blog. Full post page is available at http://blogs.chem.soton.ac.uk/neutral_drift/14858/Analysis_of_solubilities_for_usefulchem_502591.html.

5.2.2.2 Solubility measurements recorded on the wiki

In contrast to the experiments recorded on the blog, the measurements recorded on the wiki are displayed in a very different style. Each experiment has one page on the wiki along with the analysis of the solubility on a linked online spreadsheet provided by Googledocs[4]. The experiments take the form of a short report including a brief introduction with the objectives of the measurements, method, results and conclusions (figure 5.6). The report also contains a time log detailing the progress of the experiment (figure 5.7) and is linked to an associated spreadsheet (figure 5.8). Compared to experiments posted on LaBlog, the reports on the wiki contain more details about the experiment, even when the experimental methods are identical to those of the experiments posted on the blog.

The experiments were written in this way to emulate the style of the first test solubility experiment carried out by ONSC judges posted on the Usefulchem wiki[5], and to obey the rules of the challenge which asked for experiments to be written in this format. Using Googledocs[4], or any, online spreadsheet has advantages over using a conventional spreadsheet program: being online, the spreadsheet is viewable through any internet browser and is not reliant on spreadsheet software compatibility. Additionally as data files do not have to be uploaded, an experiment takes up less space on the wiki as the spreadsheet is stored on the server of the spreadsheet provider.

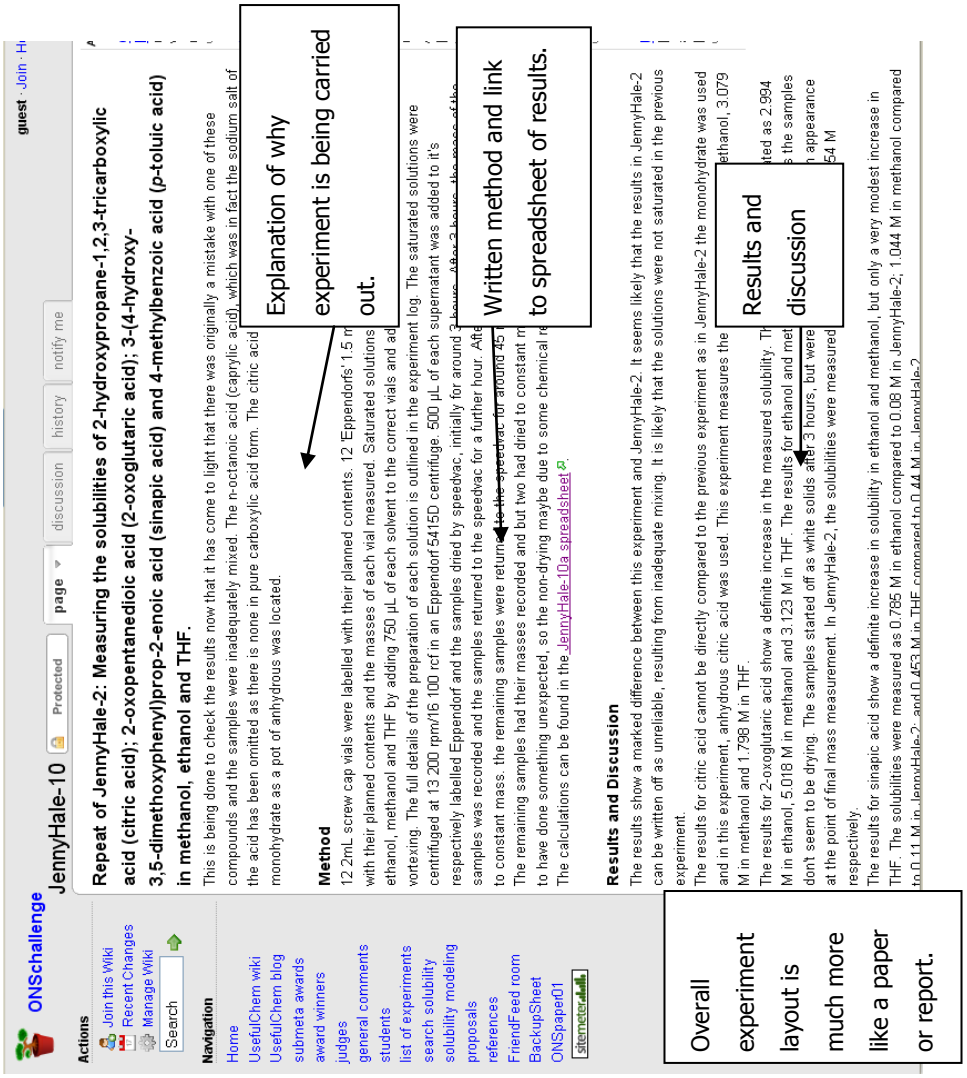


Figure 5.6: A screenshot showing an experimental report on the ONSC wiki. The page is structured much more like a paper or report with an introduction, method, results and discussion section. There is a lot more text and much more information about the experiment has been recorded than an equivalent experiment on the blog.

Figure 5.7: Screenshot of the remainder of the report shown in figure 4.6 showing the experiment log, which details the times when things occurred.

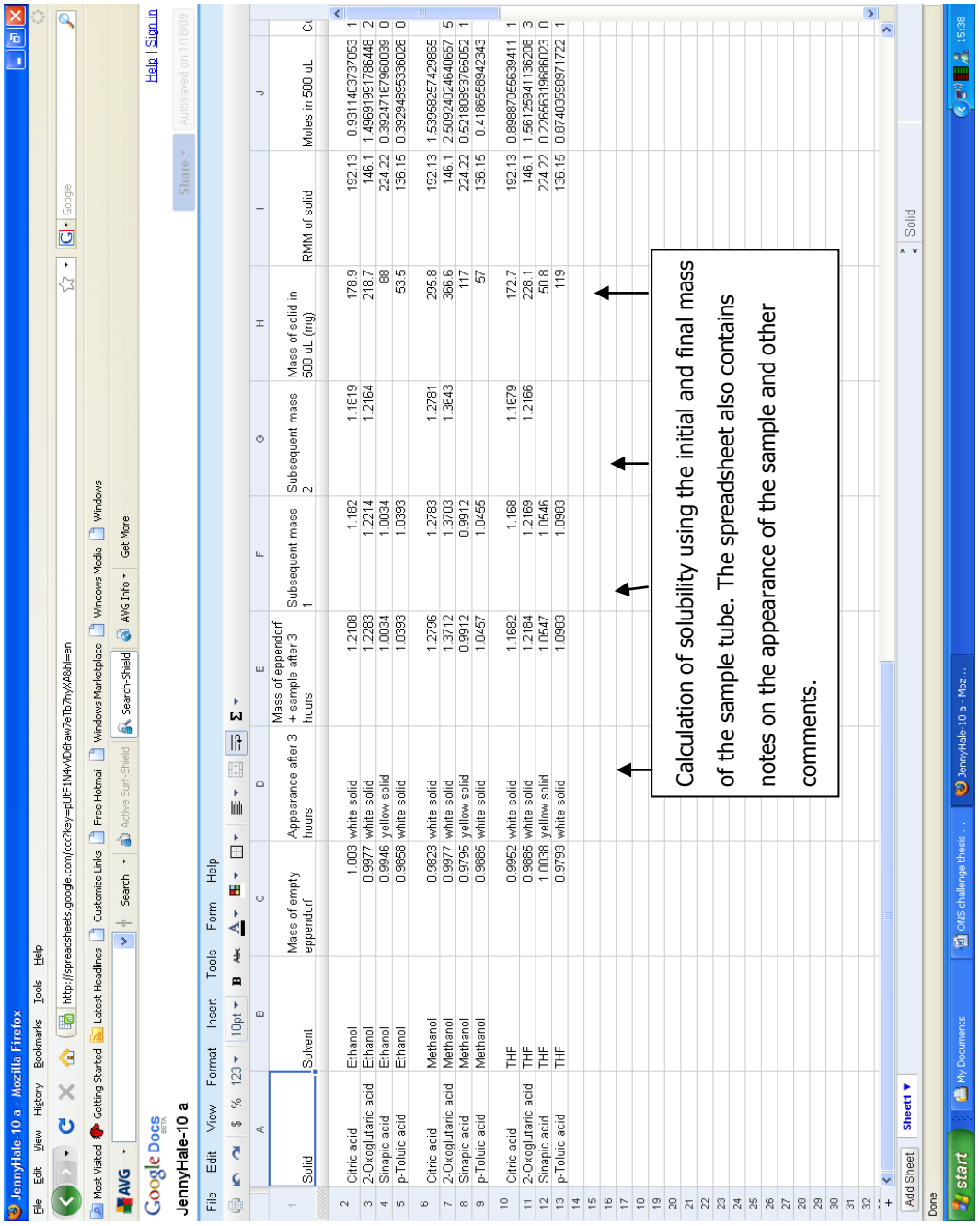


Figure 5.8: screenshot of the relevant Googledocs spreadsheet associated with the experiment shown in the previous two figures.

The posts on the blog compared to the reports on the wiki highlight a big difference between the styles and methods of the write-ups between the blog and wiki systems. Both the blog and the wiki require the work to be published as soon as possible after it occurs, but whereas the wiki contains a lot of editing and revision to produce the final version of the report, the blog is a lot more 'raw' in appearance and often just contains the experimental detail and results. The presentational styles are very different but equally valid as both are adhering to the principles of Open Notebook Science by openly publishing the work at the time it was carried out.

The wiki is better suited to preparing a more refined style of presentation/report as it has more formatting features. The wiki allows for choice of font and colour, numbers and bullet points; as well as styling for text size, bold, italic and underlined text. Additionally it has a table function in the style of that used in Microsoft Word. By contrast the blog is a lot simpler allowing only one font and colour, nor bullet points or numbered lists. Text size is editable using a button called size. As discussed in the previous chapter, tables can only be added by coding a table in BBCode.

There are other differences too between the blog and the wiki for this project and in a general science context. Both the wiki and the blog use comments and discussions about the experiments, but the way in which they are used and positioned are implemented differently between the two systems. Unlike the blog which has a specific section for comments and discussions between users underneath each post, on the ONSC wiki comments are added directly to the experimental page wherever they are deemed to be relevant. Comments are usually written in bold, or occasionally italics, inside square brackets along with the initials of the commenter. Although this is the standardised form for posts in the challenge, there are no set rules for commenting on wikis. The writing of comments in bold text inside square brackets is also the standard method for commenting on experiments on the Usefulchem wiki, and could have been chosen as both wikis are administered by the same person. Although adding comments directly to the experiment has been chosen as the standard method, there is also a separate discussion tab associated with each page of the wiki, which could be used.

Whatever system is used it is important to set up a standardised convention for interaction in order to maintain the usability of the system. Blogs are much more obviously set up for using the comment feature as a means of interaction, whereas the wiki forces the users to create an individual set of rules. However, the lack of a specified comment area in a wiki was not a hindrance to Jean-Claude Bradley who moved from using a blog to using the wiki.

Both systems are similar in that they rely on RSS feeders or emails to notify users of the addition of comments and do not have an internal notification system. Efficient notification of changes to the blog requires the blog to be viewed through an RSS feed, whilst the wiki

requires users to request notification to individual pages or add up to four separate RSS feeds which notify on four different types of changes. It is possible to add internal notification to the blog but it may not be possible to add such a feature to the wiki. Both systems are usable without relying on notification, but it can lead to comments being overlooked.

In general there is a lot less commenting on the blog both in terms of observations about the experiment and the wider implications of that experiment. On the wiki a judge or collaborator may write a comment on the experimental page suggesting a result is odd or that a piece of information is missing. These types of comments can act as prompts for the write up to be edited and improved. The time log found in experimental reports on the wiki also provides a place of commentary about what actually happened during the experiment; for example the loss of a sample that could not be replaced.

The time log is not a necessary part of each experiment and it would be possible to make observations on the work without it being present. For example explaining the reason for a lost sample could still be added to the method. In many respects the time log caused work to take longer as it was necessary to keep stopping to look at a watch/clock in order to keep logging the time.

A second major difference between the LaBlog system and the ONSC wiki is the way in which page security is managed. Maintaining security, or integrity, of work is a very important part of any science, not just Open Notebook Science, but the problem of access to posts is of particular importance to online notebooks on blogs and wikis. In the LaBlog system, only the person who wrote the post has access to that post to modify and edit it. All post edits made also have to have a reason for that edit listed before the post changes will be accepted; in addition to the blog keeping a machine readable copy of each version of a post in the database. This provides a reasonable level of security to posts, whilst allowing genuine editing to occur.

In contrast wiki users are able to access pages written by other users in order to add comments. Additionally the reason for edit box is optional to use and page changes will be saved without filling in a reason. Though there would be a level of trust that a member of a group would not alter another member's work, there is always the potential for unintentional edits such as accidentally hitting a key on the keyboard whilst adding a comment.

Whilst it may be possible for the wiki administrator to set page permissions so that only a page creator can access that page, the level of security of pages may be considered to be lower. Some scientists may be concerned that using a wiki could make it possible for one user to edit another user's posts. Restricting commenting and discussion to the discussions tab found on each wiki page may increase the security of pages without losing the ability to comment, which is a valuable tool for collaboration both within and between lab groups.

The LaBlog system has an advantage over the wiki system through the use of metadata tagging, as this allows the data to be assigned extra description. For example there have been various methods used to obtain the solubility by different participants of the challenge including saturated solution methods, UV-visible spectroscopy and NMR. If the work was being displayed on the blog, the data structure could be organised so that posts could also be tagged with the method. This would easily allow a person looking at the blog to select measurements by method. Another tag could indicate whether the solute is a solid or a liquid. Chemicals including solutes and solvents could also be listed with individual posts including batch numbers or re-ordering details and linked to measurement experiments, in a manner similar to that employed by the molecular biology blogs on the LaBlog system. The individual materials posts could also be linked to online resources such as Chemspider to allow retrieval of key data such as melting and boiling points.

In contrast the ONSC wiki has an advantage over the LaBlog system in allowing for collaboration. The wiki can be used from anywhere in the world as it is hosted on an external server, whereas the LaBlog system is hosted on a local server inside an institution, and therefore requires a local institution login address. It should be noted that it is possible for a blog system to be hosted on an external server, whilst a wiki can be hosted on a local institution server; therefore it is only in the context of these two systems that the wiki has an advantage for collaboration.

When blogs and wikis are compared generally for different types of research group and science, there are examples where a wiki is advantageous compared to a blog; and where a blog is advantageous compared to a wiki. There are also conditions where the type of system is not as important. An example of blogs being more suitable than wikis for recording the science is when each member of a research group has a different research project. For example the neutral drift and sortase blogs from the molecular biology blogs on the LaBlog system; these come from the same lab group but detail some similar and some different experiments on two proteins serving different functions. It would be confusing to a viewer to have the work mixed in together on one blog or wiki, even if it was made explicit who had carried out the work.

An example of a wiki being more suitable than a blog for the science would be when multiple people contribute to the same experiment. For example this can be seen on the Usefulchem wiki[6] experiments 235, 238 and 243 where two people are involved in a single synthesis. If each member had separate blogs then the experiment could potentially have been written up twice.

An example of when both systems could be used would be a project like the ONS Challenge where all the participants are working on the same project, but each contributing individual experiments to that project. Though it would be unusual for a blog to have a large number of

authors, there is no reason why the results could not be recorded on a single blog. It wouldn't be logical to have the results added to multiple blogs kept by each member of the challenge so for this example a wiki might be considered to be the more appropriate system.
The science being recorded on the Usefulchem wiki could also be displayed on blogs, despite the previous example, as each member of the group is usually contributing individual experiments that all are concerned with synthesising a library of similar products.

5.3 Conclusions

The ONS Challenge has highlighted the importance of keeping an accurate laboratory notebook as through that data errors can be highlighted and others may accurately reproduce work. Data produced during the challenge has contributed to the scientific knowledge available for chemists when preparing reactions. The data has also been used in a book[8] and paper[9] in addition to Chemspider, which caters for a wide audience of scientists.

The ONS Challenge has also allowed an in-depth comparison between open notebook science on blogs and wikis to be made. There are some key differences between blogs and wikis in the structure of the websites, the display of the data, the security of the pages and the facilities for comment and discussion. A summary of the key features, similarities and differences can be seen in table 5.4.

These differences make blogs and wikis suitable for different types of research groups and work. Wikis may be more appropriate if the users are in different physical locations; are all contributing to a common project or multiple people are working on a single experiment. The ONSC wiki worked because the researchers were willing to add pages and make comments. This is quite unusual as many internet communities suffer from lack of user participation. It is known that the number of people who read the content of the web is large compared to the number of contributing users. This is often known as the 90-9-1 rule, in which 90% of the population read the content, 9% contribute to the content and 1% create the content. Perhaps this was a result of the people who belonged to the project being the people who would ordinarily contribute and create web content, but it may have been to do with the structured format of the challenge and how to create and comment on wiki content.

Blogs may be more appropriate if each member of a research group is carrying out a different project. However some research could be displayed on either system. When all the factors are considered, the collaborative nature of the ONS challenge, with participants in different countries suggests that an externally hosted wiki is a better recording system to use than a system such as LaBlog which is hosted on a local institution server. Some of the features available on the LaBlog system but not on the wiki, such as tagging data, could have allowed the data to be classified by method or other descriptors; and provided extra information to

viewers and participants of the challenge. However the blog is less practical for collaboration and structure for having multiple users adding data to a single blog.

	Blog	**Wiki**
Site and Data Structure	Usually one blog per person. Main posts seen are either experiments or parts of experiments. Posts are tagged with metadata and linked together through tagging.	One wiki site with many people. Each experiment given a new page on the wiki. No tagging or linking of posts.
Display	Basic formatting features to allow bold, italic and underlined text. More elaborate formatting requires HTML. Display much more 'raw' in appearance.	Many more advanced formatting features without needing to use HTML allows reports to be written in a more formal style.
Commenting	Separate section underneath each post.	Comments made within wiki pages. Also discussion tab available but not used on ONSC wiki.
Security	Only post writer can access post to edit.	All wiki users can access pages to edit (such as adding comments).

Table 5.4: a comparison of the similarities and differences between key features of the LaBlog system and the ONSC wiki. There is also a comparison of some general differences between blogs and wikis.

5.4 Experimental

5.4.1 General experimental

Reagents were obtained from Sigma-Aldrich, Acros Organics or Fisher Scientific. We thank Sigma-Aldrich for the kind gift of chemicals donated as part of the challenge.

5.4.2 Experimental methods

Method 1

This method was used to measure the solubility of glucose, mannitol and salt in methanol, ethanol and THF. Samples were prepared for each combination of solute and solvent.

Solid was added to an Eppendorf microcentrifuge tube to a volume of approximately 500 µL. Solvent (500 µL) was added to each solid and the samples vortexed for 30 seconds. Samples which were not saturated had further solid added until the solution became saturated. The saturated solutions were spun in a microcentrifuge (16, 100 g, 1 minute) to separate the solid from the supernatant. Supernatant (200 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, for two hours. From the final mass, the number of moles in the sample was calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

Method 2

This method was used to measure the solubility of glucose, mannitol and salt in methanol, ethanol and THF. Samples were prepared for each combination of solute and solvent.

Solid was added to an Eppendorf microcentrifuge tube to a volume of approximately 500 µL. Solvent (500 µL) was added to each solid and the samples vortexed for 30 seconds. Samples which were not saturated had further solid added until the solution became saturated. The saturated solutions were spun in a microcentrifuge (16, 100 g, 1 minute) to separate the solid from the supernatant. Supernatant (200 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, for two hours. The mass of the samples was recorded and the samples dried by Speedvac for a further two hours. From the final mass, the number of moles in the sample was calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

Method 3

This method was used to measure the solubility of glucose, mannitol, salt and vanillin in methanol, ethanol and THF. Samples were prepared for each combination of solute and solvent.

Solvent (1 mL) was added to an Eppendorf microcentrifuge tube. Solid was added until the solution became saturated. The saturated solutions were spun in a microcentrifuge (16, 100 g, 1 minute) to separate the solid from the supernatant. Supernatant (500 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, until constant mass was achieved (3 hours to overnight). From the final mass, the number of moles in the sample was calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

Method 4

This method was used to measure the solubility of n-octanoic acid (sodium salt) (sodium caprylate), 2-hydroxypropane-1,2,3-tricarboxylic acid monohydrate (citric acid monohydrate), 2-oxopentanedioic acid (2-oxoglutaric acid), 3-(4-hydroxy-3,5-dimethoxyphenyl)prop-2-enoic acid (sinapic acid), and 4-methyl benzoic acid in methanol, ethanol and THF. Samples were prepared for each combination of solute and solvent.

Solvent (750 µL) was added to an Eppendorf microcentrifuge tube. Solid was added until the solution became saturated. The saturated solutions were spun in a microcentrifuge (16, 100 g, 1 minute) to separate the solid from the supernatant. Supernatant (500 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, until constant mass was achieved (3 hours to overnight). From the final mass, the number of moles in the sample was calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

Method 5

This method was used to measure the solubility of benzotriazole-5-carboxylic acid, 8-bromo-octanoic acid, 4-dimethylaminocinnamaldehyde and 2-(4-hydroxyphenyl)-azobenzoic acid in ethanol, methanol and THF.

Solvent (750 µL) was added to an Eppendorf microcentrifuge tube. Solid was added until the solution became saturated. Due to the increase in volume caused by the addition of solid, the solution of 8-bromo-octanoic acid was transferred to a glass vessel of larger volume. The saturated solutions, except 8-bromo-octanoic acid, were spun in a microcentrifuge (16, 100 g, 1 minute) to separate the solid from the supernatant. Supernatant (500 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, until constant mass was achieved (3 hours to overnight). Some samples remained in a gel formation but became crystalline solids on being left open to air for two days. From the final mass, the number of moles in the sample was calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

Method 6

This method was used to measure the solubility of vanillin in ethanol and methanol at different temperatures. Samples were prepared for each combination of solvent and temperature.

Heat blocks were prepared at 30°C and 40°C. Solvent (750 µL) was added to an Eppendorf microcentrifuge tube. Vanillin was added to each tube. One tube was kept at room temperature (23°C) whilst the others were put at 30°C and 40°C until the solutions became saturated. Saturation was ensured by inverting the tube end over end by hand in a relatively slow speed action. Supernatant (500 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, until constant mass was achieved (3 hours to overnight). From the final mass, the number of moles in the sample was calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

Method 7

This method used to measure the solubility of vanillin in ethanol and methanol at different temperatures and also using two different mixing techniques. Samples were prepared for each combination of solvent and temperature.

Heat blocks were prepared at 30°C and 40°C. Solvent (750 µL) was added to a screw cap microcentrifuge vial. Vanillin was added to each vial. One vial was kept at room temperature whilst the others were put at 30°C and 40°C. Vanillin was added in portions until the solutions became saturated. Saturation was ensured by inverting the tube end over end by hand in a relatively slow speed action or mixed using a vortex. The saturated solutions were spun in a microcentrifuge (16, 100 g, 1 minute) to separate the solid from the supernatant. Supernatant (500 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, until constant mass was achieved (3 to 5 hours). From the final mass, the number of moles in the sample was calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

Method 8

This method was used to measure the effect of mixing times on the solubility of vanillin in ethanol and methanol at room temperature. Samples were prepared for both vanillin in ethanol and vanillin in methanol.

Solvent (750 µL) was added to a screw cap microcentrifuge vial (2 mL volume). Vanillin was added to each tube and the tubes vortexed for one minute in 10 second intervals. This process was repeated until solid vanillin remained in both tubes and both tubes had been mixed for at least 3 minutes. The saturated solutions were spun in a microcentrifuge (16, 100 g, 1 minute) to separate the solid from the supernatant. Supernatant (500 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, until constant mass was achieved (3 to 5 hours). From the final mass, the number of moles in the sample was

calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

Method 9

This method was used to measure the effect of temperature on the solubility of vanillin. Samples were prepared for vanillin in ethanol at room temperature, 30°C and 40°C, and for vanillin in methanol at room temperature, 30°C and 40°C.

Heat blocks were set at 30°C and 40°C. . Solvent (750 µL) was added to a screw cap microcentrifuge vial (2 mL volume). Vanillin was added to each tube and the tubes vortexed for one minute in 10 second intervals. This process was repeated until solid vanillin remained in each tube and each tube had been mixed for at least 3 minutes. Temperature was maintained by heating each tube between rounds of mixing. The saturated solutions were spun in a microcentrifuge (16, 100 g, 1 minute) to separate the solid from the supernatant. Supernatant (500 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, until constant mass was achieved (3 to 5 hours). From the final mass, the number of moles in the sample was calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

Method 10

This method was used to measure the solubility of 1-adamantylamine, 6-aminocaproic acid, 4-hydroxyphenylpyruvic acid, 4-nitrophenylacetic acid, citric acid, sinapic acid, 2-oxoglutaric acid, 4-methylbenzoic acid, benzotriazole-5-carboxylic acid, 4-dimethylaminocinnamaldehyde, 2-(4-hydroxyphenyl)azobenzoic acid, 10-chloro-9-anthraldehyde, 4-hydroxybenzaldehyde and 4-hydroxy-3-methylbenzaldehyde in ethanol and methanol. The solubility in THF was also measured for some samples.

Samples were prepared for each combination of solute and solvent. Solvent (750 µL) was added to a screw cap microcentrifuge vial (2 mL volume). Solid was added to each tube and each tube vortexed to mix for 1 minute in 10 second intervals. More solid was added and the process repeated until all samples were considered to be saturated. All samples had been mixed by vortex for at least 3 minutes. The saturated solutions were spun in a microcentrifuge (16, 100 g, 1 minute) to separate the solid from the supernatant. Supernatant (500 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, until constant mass was achieved (3 to 4 hours). From the final mass, the number of moles in the sample was calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

Method 11

This method was used to measure the solubility of benzoic acid, and undecanoic acid in ethanol, methanol and THF, and the solubility of 4-hydroxybenzaldehyde in THF. Samples were prepared for each combination of solute and solvent.

Solvent (750 µL) was added to a screw cap microcentrifuge vial (2 mL volume). Solid was added to each tube and each tube vortexed to mix for 1 minute in 10 second intervals. More solid was added and the process repeated until all samples were considered to be saturated. In the case of undecanoic acid, the samples had to be transferred to 15 mL tubes due to volume increase caused by level of solubility. Undecanoic acid in THF could not be prepared due to an inability to find a suitable vessel for preparing the solution (THF melted 15 mL centrifuge tubes and exploded a glass vial). All samples had been mixed by vortex for at least 3 minutes to ensure saturation. The saturated solutions, except those of undecanoic acid, were spun in a microcentrifuge (16, 100 g, 1 minute) to separate the solid from the supernatant. Supernatant (500 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, until constant mass was achieved (3 to 4 hours). From the final mass, the number of moles in the sample was calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

Method 12

This method was used to measure the solubility of the liquid aldehyde 3,4-dimethylbenzaldehyde in ethanol, methanol and THF. Samples were prepared for each combination of solute and solvent.

Solvent (200 µL) was added to a 1.5 mL glass sample vial. The aldehyde was added drop-wise to the solvent. The two layers were gently swirled and were deemed to comprise two immiscible layers. The lower layer had 200 µL aliquotted and the mass of the solution measured. The mass of the solvent was subtracted from this figure to leave the lass of the aldehyde. Solubility was calculated in the normal way. This assay method was incorrect and the results were nonsensical.

The density of the aldehyde was used to calculate the maximum concentration of the aldehyde and hence the maximum concentration solution that could be made from that aldehyde.

Method 13

This method was used to measure the solubility of the liquid aldehydes 3,5-dimethylbenzaldehyde and 4-ethylbenzaldehyde in ethanol, methanol and THF. Samples were prepared for all combinations of aldehyde and solvent.

Solvent (400 µL) was added to a 7.5 mL glass sample vial. Each aldehyde (600 µL) was added to the solvent. The vial was vortexed for 10 seconds. Complete miscibility was obtained. The vial was vortexed for a futher 50 seconds in 10 second intervals to ensure the solution was thoroughly mixed. As complete miscibility was obtained, the densities of the aldehydes were

used to calculate the maximum concentration of each aldehyde and hence the maximum concentration solution that could be prepared from each aldehyde.

Method 14

This method was used to measure the solubility of 8-bromo-octanoic acid in ethanol. Ethanol (300 µL) was added to a 2 mL screw capped microcentrifuge tube. 8-Bromo-octanoic acid was added in portions and the solution vortexed for 1 minute in 10 second intervals until the solution became saturated. The saturated solution was spun in a microcentrifuge (16, 100 g, 1 minute) to separate the solid from the supernatant. Supernatant (200 µL) was transferred to a pre-massed Eppendorf and dried using a Speedvac, a centrifuge attached to a vacuum pump, for 3 hours. The sample was still liquid and was returned to speedvac for one hour. The sample was still liquid so was placed in the freezer for 20 minutes. Sample was returned to speedvac for a further 10 minutes. From the final mass, the number of moles in the sample was calculated. This was used to calculate the molarity of the sample and hence the maximum solubility of the solute in the given solvent was calculated.

5.4.3 List of experiments on the LaBLog system and ONSC wiki

1. http://blogs.chem.soton.ac.uk/neutral_drift/14831/Preparation_of_saturated_solutions_for_solubilities_502591.html
2. http://blogs.chem.soton.ac.uk/neutral_drift/14844/Preparation_of_dried_samples_for_solubilities_502591.html
3. http://blogs.chem.soton.ac.uk/neutral_drift/14858/Analysis_of_solubilities_for_usefulchem_502591.html
4. http://blogs.chem.soton.ac.uk/neutral_drift/17383/Preparation_of_saturated_solutions_for_solubilities_502596.html
5. http://blogs.chem.soton.ac.uk/neutral_drift/18377/Rewrite_of_Preparation_of_dried_samples_for_solubilities_502596.html
6. http://blogs.chem.soton.ac.uk/neutral_drift/18398/Analysis_of_solubilities_502596.html
7. http://onschallenge.wikispaces.com/JennyHale-1
8. http://onschallenge.wikispaces.com/JennyHale-2
9. http://onschallenge.wikispaces.com/JennyHale+3
10. http://onschallenge.wikispaces.com/JennyHale4
11. http://onschallenge.wikispaces.com/JennyHale-5
12. http://onschallenge.wikispaces.com/JennyHale-6
13. http://onschallenge.wikispaces.com/JennyHale-7
14. http://onschallenge.wikispaces.com/JennyHale-8
15. http://onschallenge.wikispaces.com/JennyHale-9
16. http://onschallenge.wikispaces.com/JennyHale-10
17. http://onschallenge.wikispaces.com/JennyHale-11
18. http://onschallenge.wikispaces.com/JennyHale-12

19. http://onschallenge.wikispaces.com/JennyHale-13
20. http://onschallenge.wikispaces.com/JennyHale-14
21. http://onschallenge.wikispaces.com/JennyHale-15
22. http://onschallenge.wikispaces.com/JennyHale-16

5.5 References

1. Bradley, J-C.; The Open Notebook Science Challenge; http://onschallenge.wikispaces.com; Accessed 20.07.10 (Archived by http://www.webcitation.org at http://www.webcitation.org/5rMWdPFUx).
2. Bradley, J-C.; The Open Notebook Science Challenge; http://usefulchem.blogspot.com/2008/09/open-notebook-science-challenge.html.; Accessed 20.07.10 (Archived by http://www.webcitation.org at http://www.webcitation.org/5rMXh7u01).
3. Chemspider; http://chemspider.com; Accessed 20.07.10 (Archived by http://www.webcitation.org at http://www.webcitation.org/5rMXzS5U4).
4. Googledocs; https://www.google.com/accounts/ServiceLogin?service=writely&passive=1209600&continue=http://docs.google.com/&followup=http://docs.google.com/&ltmpl=homepage; Accessed 20.07.10 (Archived by http://www.webcitation.org at http://www.webcitation.org/5rMYbS9Z8).
5. Bradley, J-C. and Neylon, C.; http://usefulchem.wikispaces.com/EXP207; Accessed 16.07.10 (Archived by http://www.webcitation.org at http://www.webcitation.org/5rGPwGusP).
6. Usefulchem; http://usefulchem.wikispaces.com/All+Reactions; Accessed 19.07.10 (Archived by http://www.webcitation.org at http://www.webcitation.org/5rLGhwQdE).
7. ONS Challenge; http://onschallenge.wikispaces.com/list+of+experiments; Accessed 20.07.10 (Archived by http://www.webcitation.org at http://www.webcitation.org/5rMmvKgBv).
8. Bradley, J-C., Neylon, C., Guha, R., Williams, A., Hooker, B., Lang, A., Friesen, B., Bohinski, T., Bulger, D., Federici, M., Hale, J., Mancinelli, J., Mirza, K., Moritz, M., Rein, D., Tchakounte, C., and Truong, H. Open Notebook Science Challenge: Solubilities of Organic Compounds in Organic Solvents. 3rd Edition 2010. ISBN: 978-0-557-31801-8.
9. Bradley, J-C., Neylon, C., Guha, R., Williams, A., Hooker, B., Lang, A., Friesen, B., Bohinski, T., Bulger, D., Federici, M., Hale, J., Mancinelli, J., Mirza, K., Moritz, M., Rein, D., Tchakounte, C., and Truong, H. Open Notebook Science Challenge: Solubilities of Organic Compounds in Organic Solvents. Available from Nature Precedings http://dx.doi.org/10.1038/npre.2010.4243.3 (2010).

Chapter 6 - Overall conclusions and future directions

This thesis has investigated three projects which are seemingly independent yet intrinsically linked. In the first project, the role of neutral drift in directed evolution has been explored through the directed evolution of the *E.* coli β-glucuronidase into a β-galactosidase. Theoretical approaches suggested that neutral drift could have some value in directed evolution experiments for increasing the number of routes through a fitness landscape and allow new variants to be found that were able to escape from local minima in the landscape produced by direct selection.

The directed evolution was carried out using four rounds of neutral drift i.e. screening and selecting for native activity, before screening and selecting variants with increased β-galactosidase activity. Two main variants were found which contained mutations G245A/K370R/W529L and V67D/K370R/W529L. The W529L mutation had previously been discovered using conventional direct selection techniques. The results of these experiments were compared with other directed evolution and neutral drift experiments in the literature and a similar picture emerged. The mutations found when using neutral drift were in most cases identical to that found by direct selection, leading to the conclusion that neutral drift does not have as much value for directed evolution experiments as had been predicted.

The libraries, including two previously prepared libraries, were also screened against a range of alternative substrates. Three mutants with increased β-xylosidase activity were found. These mutants had a range of mutations and only one of these had previously been found in a directed evolution experiment (T509A). However as there is no literature to suggest a purpose led directed evolution to prepare a β-xylosidase from a β-glucuronidase, no reasonable comparisons can be made. It is hypothesised that neutral drift may have some value for preparing libraries which are to be screened against a wide variety of substrates.

In the second project the neutral drift experiments from the first project were used to develop and optimise an electronic laboratory record in the form of an online blog (LaBlog). The blog underwent three distinct phases in its development. These phases highlighted the importance of the data (and metadata) structure; how the data structure was not the same as the data display and the link between the data structure and display. It also highlighted how there are many different ways to display the data of an experiment and the best way may not be the most obvious way.

In the third project, the Open Notebook Science Challenge, the measurement of the solubilities of various solutes in organic solvents was used as a tool to compare and contrast keeping laboratory records on a wiki and on a blog (using the LaBlog system designed in the previous

project). This project highlighted how different scientific disciplines require different methods of recording data: trying to fit the solubility data to the molecular biology data model was unsuccessful; but the project also analysed how the successful structure of the wiki could be applied to the blog to make it suitable for that science, and in doing so how it could be applied to any type of science. The challenge also demonstrated the power of collaboration and the use of Open Notebook Science by highlighting when results did not fit solubility patterns or other recorded results just by being able to read the whole laboratory record.

Linking these projects together highlighted that the method and levels of detail used for recording the work carried out in the laboratory is equally as important as the work carried out. Recording extra data allowed experiments to be linked by connections beyond the capability of a paper laboratory notebook, as well as highlighting the importance of the quality of data that is recorded in order to allow other scientists to repeat work or highlight mistakes.

Future directions

There are a number of directions in which the practical work for the neutral drift project could be taken. More neutral drifts could be carried out in other enzyme systems to assess if the results always compare to standard direct selection experiments; or further experiments could be carried out on the β-glucuronidase system. Firstly a much larger library could be used in the directed evolution of a β-glucuronidase to a β-galactosidase by altering the conditions to allow Fluorescence Activated Cell Sorting (FACS) or secondly β-glucuronidase could undergo two directed evolutions using both direct selection and neutral drift to develop either β-xylosidase or β-fucosidase activity. The literature provided examples of β-glucuronidase variants with activity towards these substrates, but not directed evolution experiments designed to evolve these activities. The results of such directed evolution experiments could be used to further deduce whether neutral drift will play an important role in future directed evolution experiments.

The LaBlog system has now largely been optimised towards the molecular biology project work and is a fully functioning blog. Many of the suggestions for future improvements to the blog are being added or methods for adding the improvements are being devised at the time of writing. However the main challenges are to increase automation as well as to provide a system that can be used under any scientific situation.

The ONS challenge will continue to record solubility measurements of a wide range of compounds. Other methods will be devised to record measurements that are more effective for use with compounds of limited solubility, or are liquid at room temperature. Statistical analyses of the results will provide models of solubility and allow greater prediction of suitable solvents for reactions and the behaviour of compounds in solvents.

Overall these projects have demonstrated the use of molecular biology and chemistry in Open Notebook Science, and compared and contrasted the different methods available. In addition they have highlighted some of the benefits of collaborative science and added to the scientific understanding of directed enzyme evolution.

Appendix 1 – ELN program details

Program Name	Program Class	Does it fully replace paper?	Style of user interface	Features
OpenELN	Discipline Specific	No. The program replaces paper for the lab work but is reliant on security stationery for witnessing	Form and window based entry	Storage, search and retrieval of ELN documents. Audit trail and archiving service.
Symyx	Discipline Specific but developing to be generic	Yes	Form and window based entry	Storage, search and retrieval of ELN documents. Audit and archiving included within the program.
ChemBioOffice	Generic	Yes	Form and window based entry	Storage, search and retrieval of ELN documents. Audit and archiving included within the program. Plug ins allow program to be made discipline specific.
LABTrack	Generic	Yes	Internet browser	Storage, search and retrieval of ELN documents. Audit and archiving included within the program.
eNovator	Generic	Yes	Web based interface	Accepts all data formats. Can capture freeform and structured data.
NoteBookMaker	Generic	Yes	Window based Freeform entry.	IP protection. Can accept data copied and pasted from other programs.
Nexxis	Custom	Yes	Web based forms	For analytical laboratories. Integration and automation of routine procedures with lab book experiment writing capabilities.

Labscape Lab assistant	Custom	No	Window based	For cellular biology. Aids and prompts scientist. Results printed and pasted into lab book afterwards
CERF	Custom	Yes	Window and form based entry	Similar to other programmes
iPadELN (This is different to the portable device made by Apple)	Generic	Yes	Window , form and browser based entry	Free. Works on both Windows and Mac operating systems
ConturELN	Generic	Yes	Window and form based entry	Templates for ease of use. Designed to be usable without computing knowledge.
ScripSafe	Secure stationery	No	Paper	Enhance ELN systems. Doubly watermarked paper, Chemically active papers. Latent images that appear on photocopying.
PatentPad	Secure stationery	No	Paper	Enhance ELN systems. Doubly watermarked paper, Chemically active papers. Latent images that appear on photocopying.
PatentSafe	Secure stationery registration	No	N/A	Registration service for security papers.

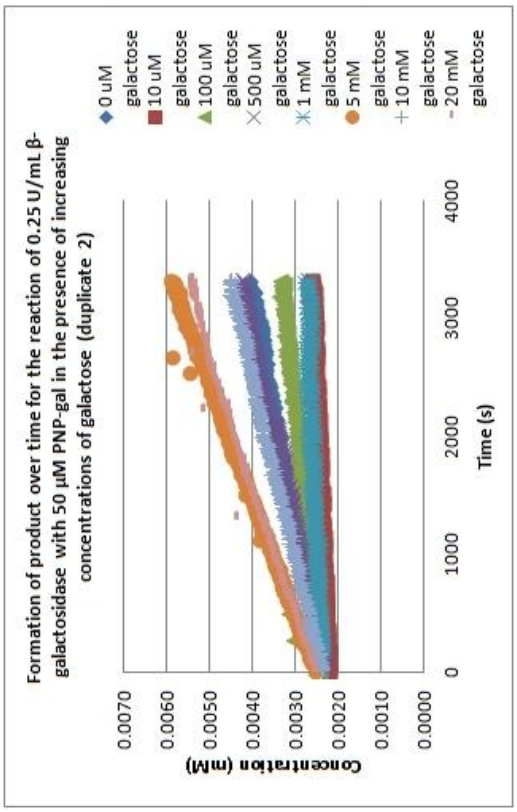
Table 3.1: ELN systems

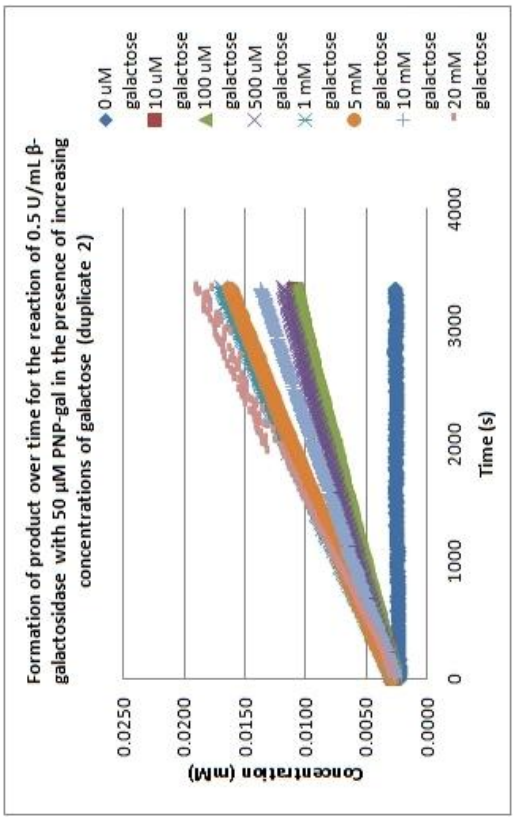
References

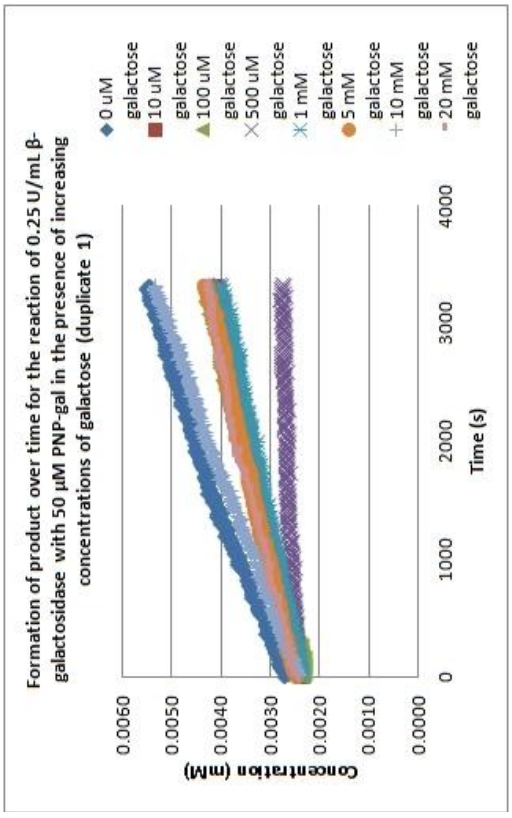
1. SCRIP-SAFE; *Keeping a World of Information Safe, Secure and Convenient*; http://www.scrip-safe.com; Accessed 08.07.10 (archived by WebCite® at http://www.webcitation.org/5r4k8ZhmZ).
2. Symyx; http://www.symyx.com/products/software/lab-notebooks/index.jsp; Accessed 08.07.10 (Archived by WebCite® at http://www.webcitation.org/5r4kY3JO7).
3. OpenELN; http://www.amphora-research.com/products/openeln; Accessed 08.07.10 (Archived by WebCite® at http://www.webcitation.org/5r4l06BbO).

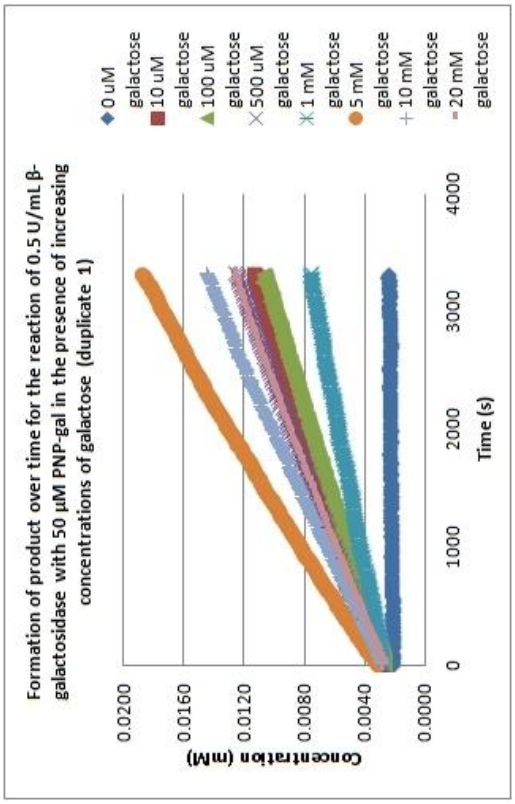
4. ChemBioOffice; http://www.cambridgesoft.com/software/ChemBioOffice/; Accessed 08.07.10 (Archived by WebCite® at http://webcitation.org/5r4lLFfrF).
5. Scoffin, R.; *The New Wave in Electronic Laboratory Notebook Systems*; Chemical Biology and Drug Design; 2006; **67**; 184-185.
6. LABTrack; http://www.labtrack.com/Tab1b.html; Accessed on 08.07.10 (Archived by WebCite® at http://www.webcitation.org/5r4mXBRPz).
7. eNovator: http://www.kinematik.com/index.cfm/page/industrysolutions; Accessed on 08.07.10 (Archived by WebCite® at http://www.webcitation.org/5r4mr8pw5).
8. NoteBookMaker; http://www.notebookmaker.com/NoteBook/Intro.html; Accessed on 08.07.10 (Archived by WebCite® at http://www.webcitation.org/5r4nJyP8R).
9. Nexxis; http://www.labtronics.com/electronic_laboratory_notebook.htm; Accessed 08.07.10 (Archived by WebCite® at http://www.webcitation.org/5r4oPrfQH).
10. Arnstein, L., Hung, C-Y., Franza, R., Zhou, Q. H., Borriello, G., Consolvo, S., and Su, J.; *Labscape: A Smart Environment For The Cell Biology Laboratory*; Pervasive Computing; 2002; **July-September**; 13-21.
11. Arnstein, L. Grimm, R., Hung, C-Y., Kang, J-H., La Marca, A., Look, G., Sigurdsson, S. B., Su. J., and Borriello, G.; *Systems Support for Ubiquitous Computing: A Case Study of Two Implementations of Labscape*; Proceedings of the First international Conference on Pervasive Computing; 2002; 15.
12. CERF; http://rescentris.com/cerf.aspx; Accessed 08.07.10 (Archived by WebCite® at http://www.webcitation.org/5r4pnLv1e).
13. iPad ELN; http://www.ipadeln.com/index.html; Accessed 08.07.10 (Archived by WebCite® at http://www.webcitation.org/5r4qBl3Ji).
14. ConturELN; http://www.contur.com/products/products.asp; Accessed 08.07.10 (Archived by WebCite® at www.webcitation.org/5r4qaRLk9).
15. ConturELN; http://www.contur.com/shared/pdf/ConturELN_Case_study_Biopharmaceuticals.pdf; Accessed 08.07.10 (Archived by WebCite® at www.webcitation.org/5r4qn4uIt).
16. Kihlen, M.; *Electronic lab notebooks – do they work in reality?*; Drug Discovery Today; 2005; **10**; 1205-1207.
17. PatentPad; http://www.amphora-research.com/products/patentpad; Accessed 08.07.10 (Archived by WebCite® at www.webcitation.org/5r4rluel9).
18. PatentSafe; http://www.amphora-research.com/products/patentsafe; Accessed 08.07.10 (Archived by WebCite® at www.webcitation.org/5r4ryJp2U).

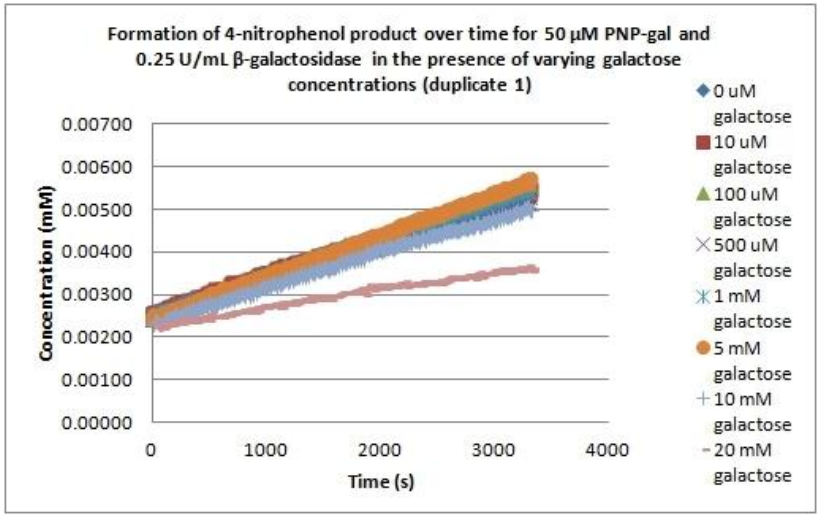
Supplementary data – graphs from the β-galactosidase inhibition assays

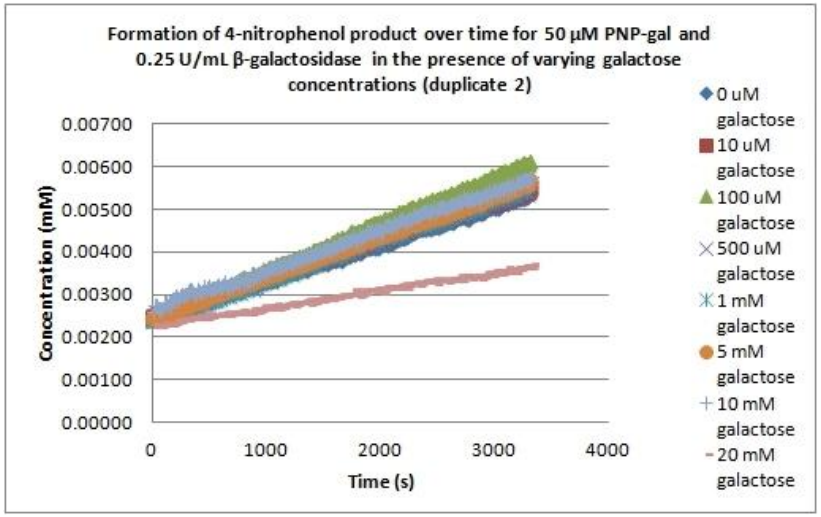


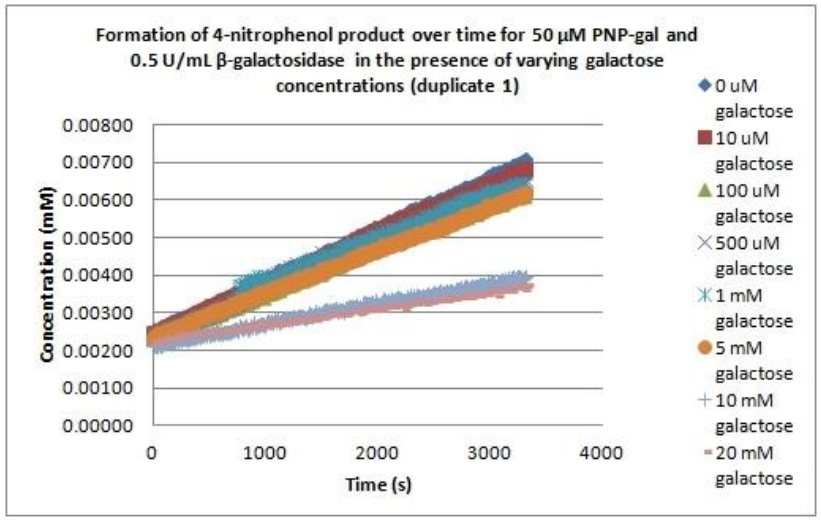


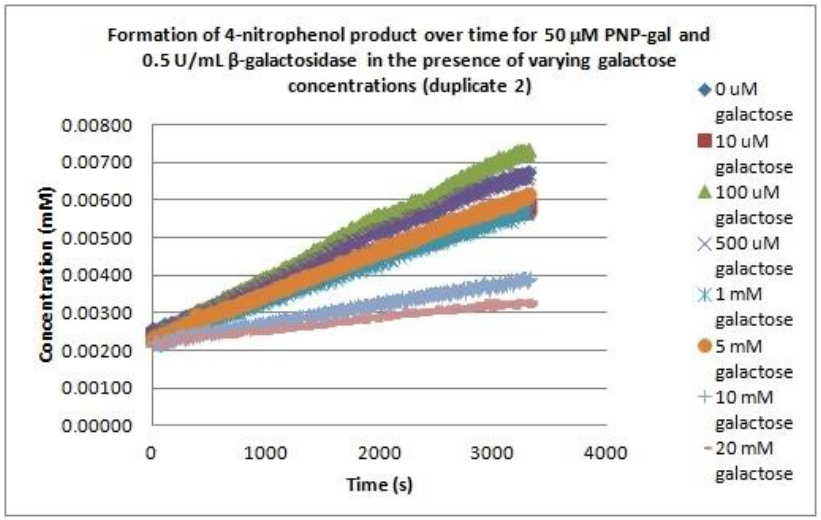


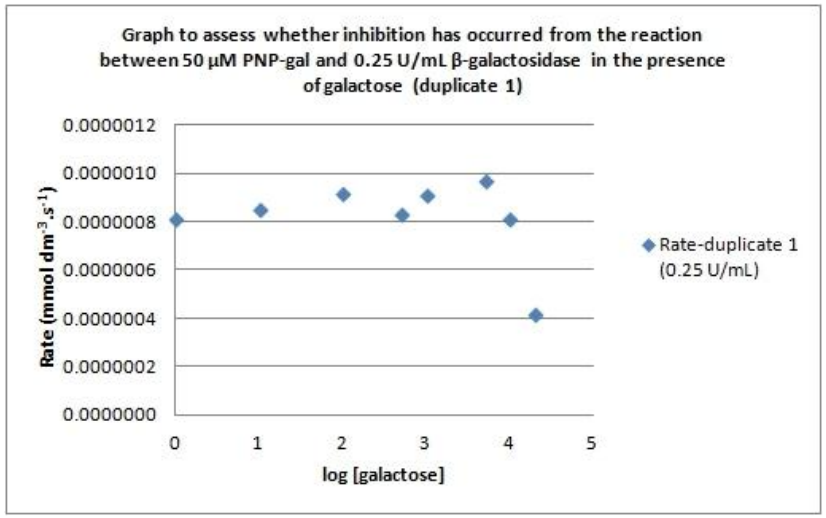


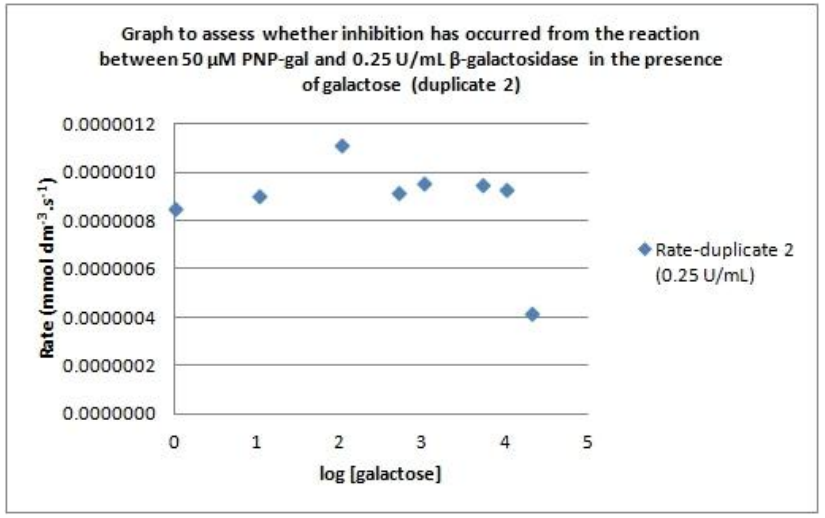

Formation of 4-nitrophenol product over time for 50 µM PNP-gal and 0.25 U/mL β-galactosidase in the presence of varying galactose concentrations (duplicate 1)
Concentration (mM)
0.00700
0.00600
0.00500
0.00400
0.00300
0.00200
0.00100
0.00000
0
1000
2000
3000
4000
Time (s)
0 uM galactose
10 uM galactose
100 uM galactose
500 uM galactose
1 mM galactose
5 mM galactose
10 mM galactose
20 mM galactose

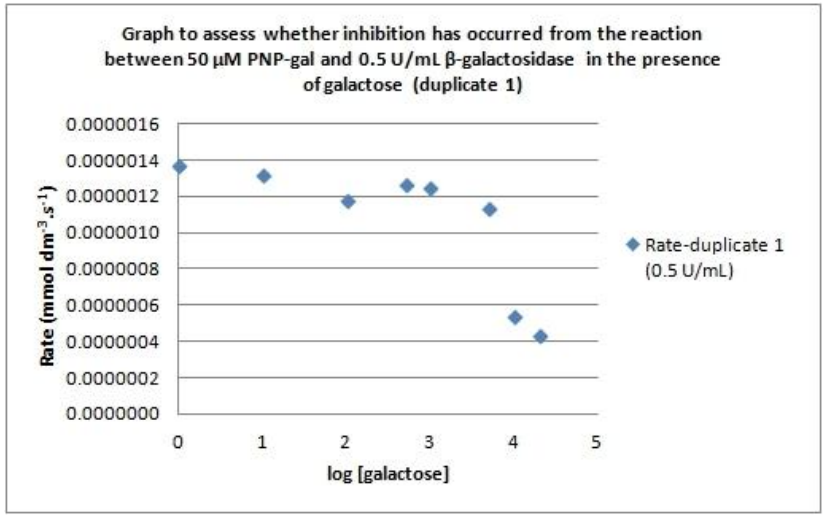

Formation of 4-nitrophenol product over time for 50 µM PNP-gal and 0.25 U/mL β-galactosidase in the presence of varying galactose concentrations (duplicate 2)
Concentration (mM)
0.00700
0.00600
0.00500
0.00400
0.00300
0.00200
0.00100
0.00000
0
1000
2000
3000
4000
Time (s)
0 uM galactose
10 uM galactose
100 uM galactose
500 uM galactose
1 mM galactose
5 mM galactose
10 mM galactose
20 mM galactose

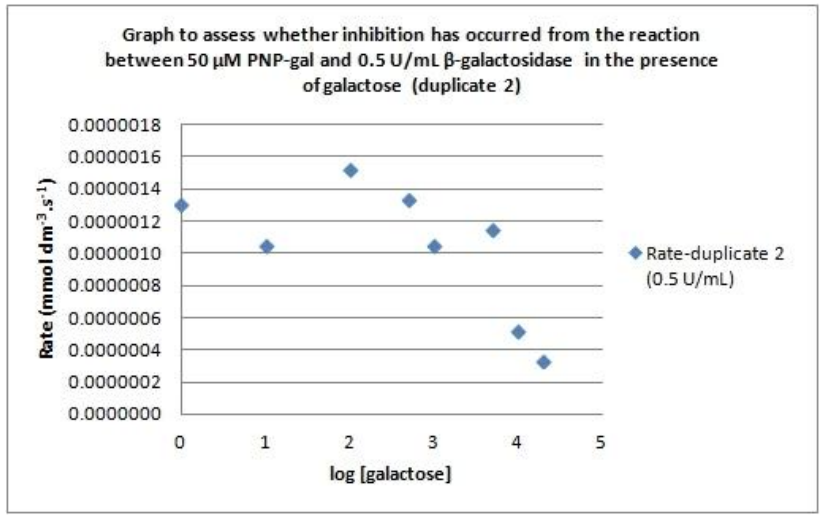

Formation of 4-nitrophenol product over time for 50 µM PNP-gal and 0.5 U/mL β-galactosidase in the presence of varying galactose concentrations (duplicate 1)
Concentration (mM)
0.00800
0.00700
0.00600
0.00500
0.00400
0.00300
0.00200
0.00100
0.00000
0
1000
2000
3000
4000
Time (s)
0 uM galactose
10 uM galactose
100 uM galactose
500 uM galactose
1 mM galactose
5 mM galactose
10 mM galactose
20 mM galactose

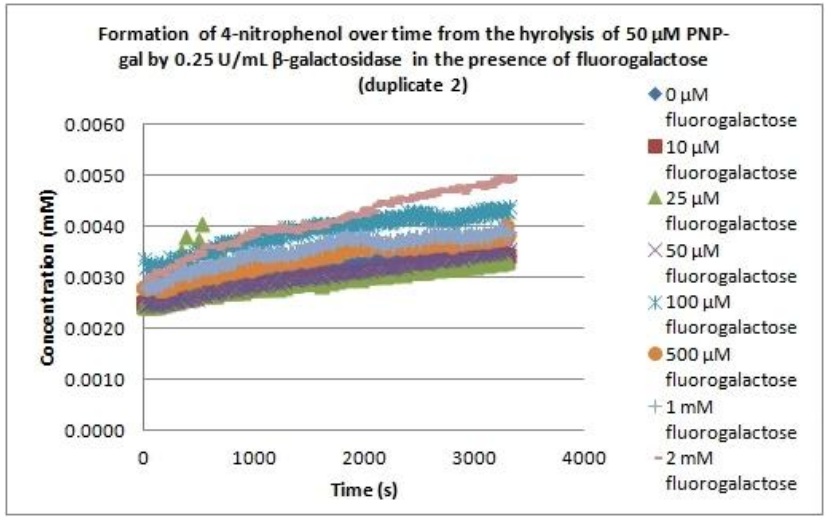

Formation of 4-nitrophenol product over time for 50 µM PNP-gal and 0.5 U/mL β-galactosidase in the presence of varying galactose concentrations (duplicate 2)
Concentration (mM)
0.00800
0.00700
0.00600
0.00500
0.00400
0.00300
0.00200
0.00100
0.00000
0
1000
2000
3000
4000
Time (s)
0 uM galactose
10 uM galactose
100 uM galactose
500 uM galactose
1 mM galactose
5 mM galactose
10 mM galactose
20 mM galactose

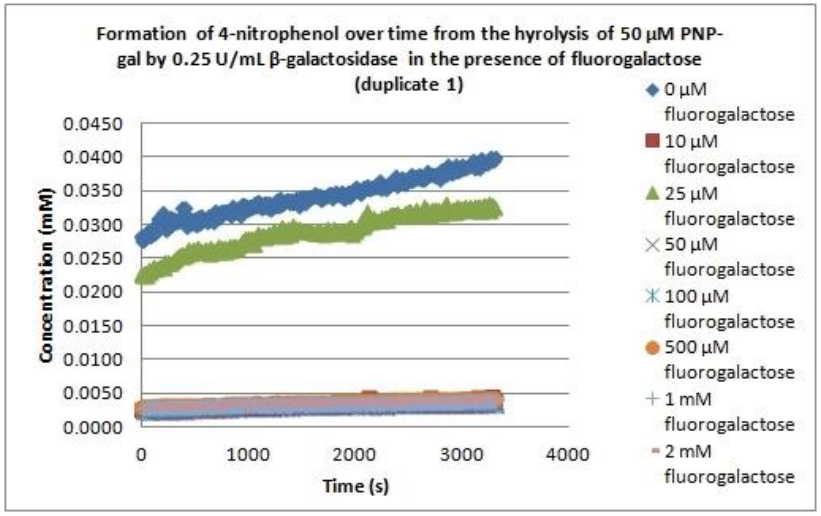

Graph to assess whether inhibition has occurred from the reaction between 50 µM PNP-gal and 0.25 U/mL β-galactosidase in the presence of galactose (duplicate 1)
Rate (mmol dm-3.s-1)
0.0000012
0.0000010
0.0000008
0.0000006
0.0000004
0.0000002
0.0000000
0
1
2
3
4
5
log [galactose]
Rate-duplicate 1 (0.25 U/mL)

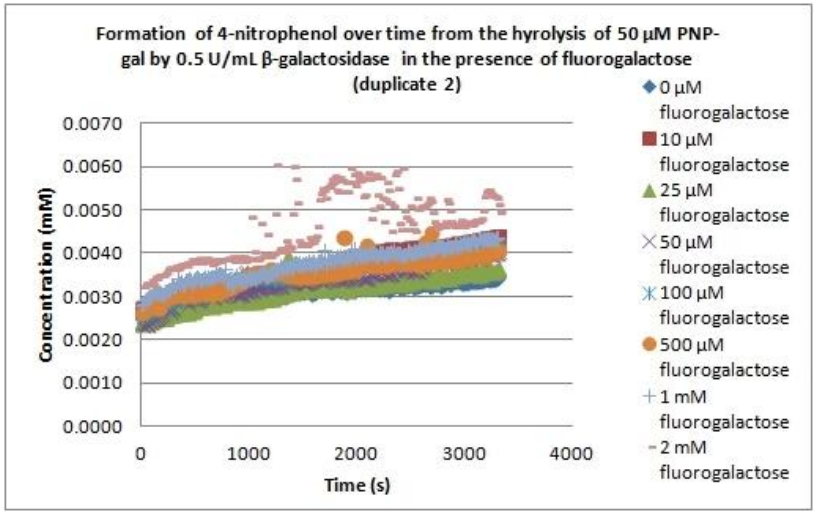

Graph to assess whether inhibition has occurred from the reaction between 50 µM PNP-gal and 0.25 U/mL β-galactosidase in the presence of galactose (duplicate 2)
Rate (mmol dm-3.s-1)
0.0000012
0.0000010
0.0000008
0.0000006
0.0000004
0.0000002
0.0000000
0
1
2
3
4
5
log [galactose]
Rate-duplicate 2 (0.25 U/mL)

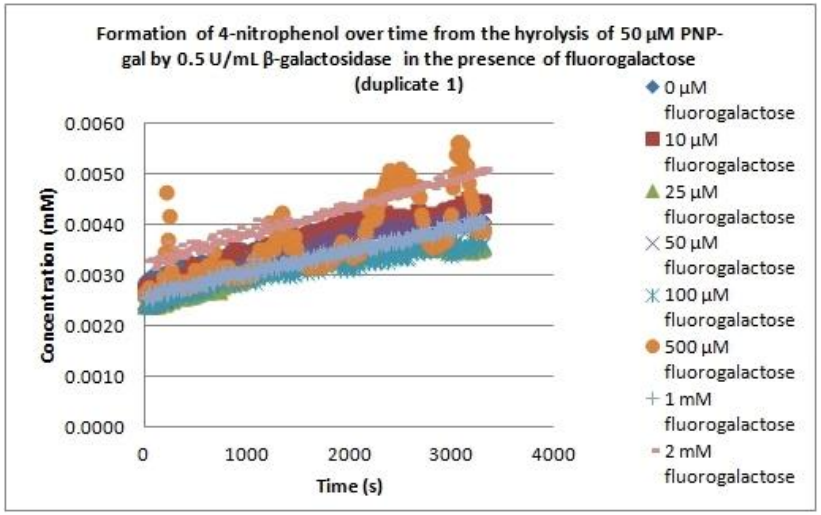

Graph to assess whether inhibition has occurred from the reaction between 50 µM PNP-gal and 0.5 U/mL β-galactosidase in the presence of galactose (duplicate 1)
Rate (mmol dm-3.s-1)
0.0000016
0.0000014
0.0000012
0.0000010
0.0000008
0.0000006
0.0000004
0.0000002
0.0000000
0
1
2
3
4
5
log [galactose]
Rate-duplicate 1 (0.5 U/mL)

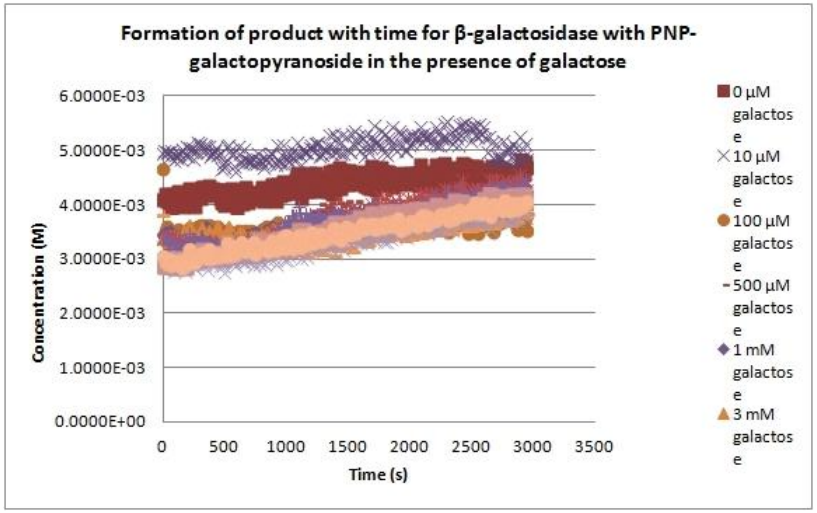

Graph to assess whether inhibition has occurred from the reaction between 50 µM PNP-gal and 0.5 U/mL β-galactosidase in the presence of galactose (duplicate 2)
Rate (mmol dm-3.s-1)
0.0000018
0.0000016
0.0000014
0.0000012
0.0000010
0.0000008
0.0000006
0.0000004
0.0000002
0.0000000
0
1
2
3
4
5
log [galactose]
Rate-duplicate 2 (0.5 U/mL)

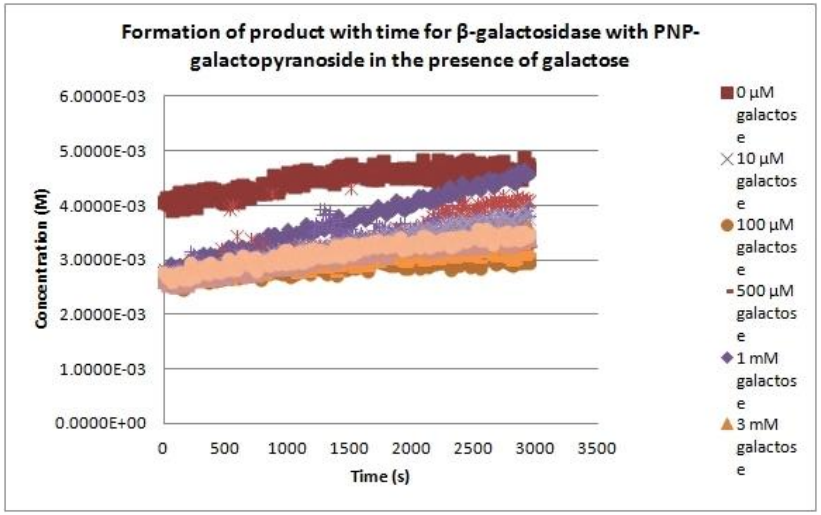

Formation of 4-nitrophenol over time from the hyrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of fluorogalactose (duplicate 2)
Concentration (mM)
0.0060
0.0050
0.0040
0.0030
0.0020
0.0010
0.0000
0
1000
2000
3000
4000
Time (s)
0 µM fluorogalactose
10 µM fluorogalactose
25 µM fluorogalactose
50 µM fluorogalactose
100 µM fluorogalactose
500 µM fluorogalactose
1 mM fluorogalactose
2 mM fluorogalactose

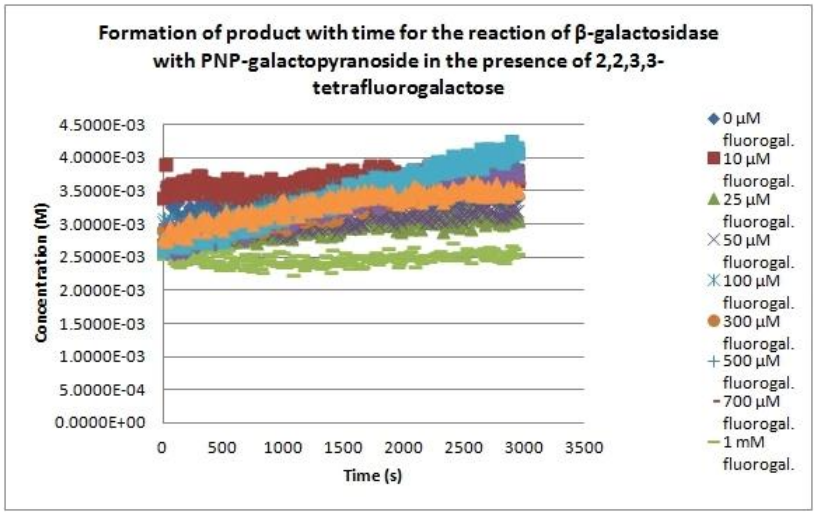

Formation of 4-nitrophenol over time from the hyrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of fluorogalactose (duplicate 1)
Concentration (mM)
0.0450
0.0400
0.0350
0.0300
0.0250
0.0200
0.0150
0.0100
0.0050
0.0000
0
1000
2000
3000
4000
Time (s)
0 µM fluorogalactose
10 µM fluorogalactose
25 µM fluorogalactose
50 µM fluorogalactose
100 µM fluorogalactose
500 µM fluorogalactose
1 mM fluorogalactose
2 mM fluorogalactose

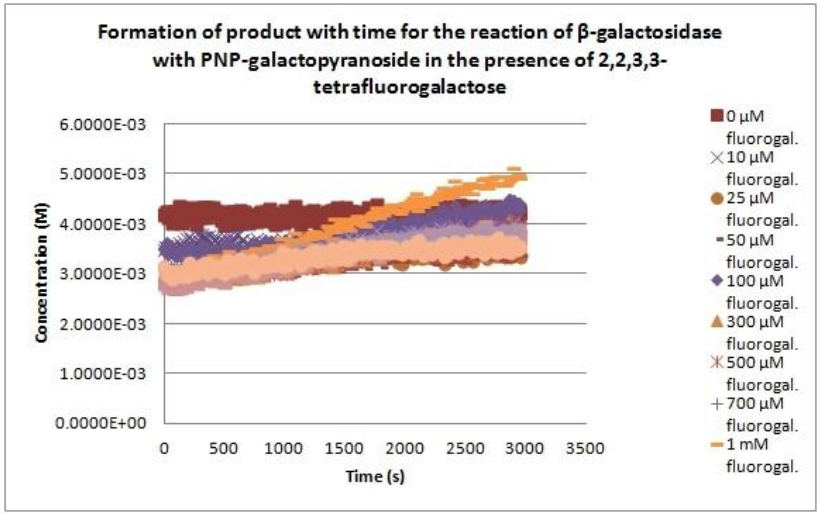

Formation of 4-nitrophenol over time from the hyrolysis of 50 µM PNP-gal by 0.5 U/mL β-galactosidase in the presence of fluorogalactose (duplicate 2)
Concentration (mM)
0.0070
0.0060
0.0050
0.0040
0.0030
0.0020
0.0010
0.0000
0
1000
2000
3000
4000
Time (s)
0 µM fluorogalactose
10 µM fluorogalactose
25 µM fluorogalactose
50 µM fluorogalactose
100 µM fluorogalactose
500 µM fluorogalactose
1 mM fluorogalactose
2 mM fluorogalactose

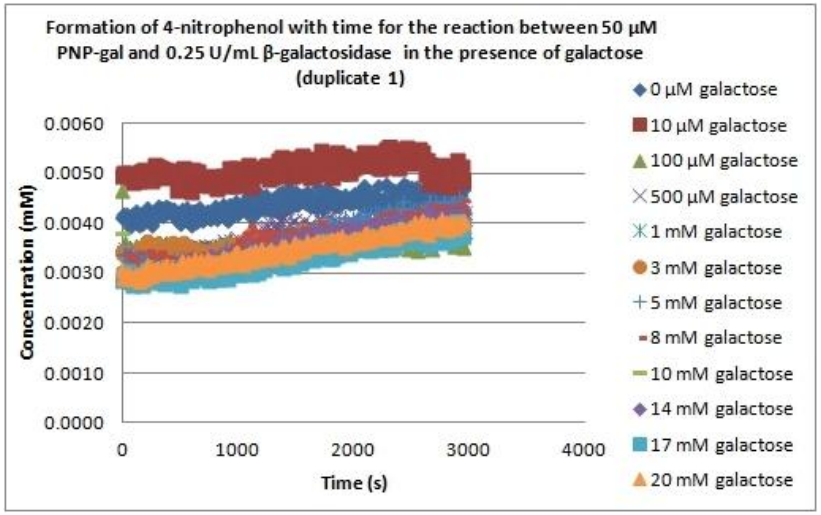

Formation of 4-nitrophenol over time from the hyrolysis of 50 µM PNP-gal by 0.5 U/mL β-galactosidase in the presence of fluorogalactose (duplicate 1)
Concentration (mM)
0.0060
0.0050
0.0040
0.0030
0.0020
0.0010
0.0000
0
1000
2000
3000
4000
Time (s)
0 µM fluorogalactose
10 µM fluorogalactose
25 µM fluorogalactose
50 µM fluorogalactose
100 µM fluorogalactose
500 µM fluorogalactose
1 mM fluorogalactose
2 mM fluorogalactose

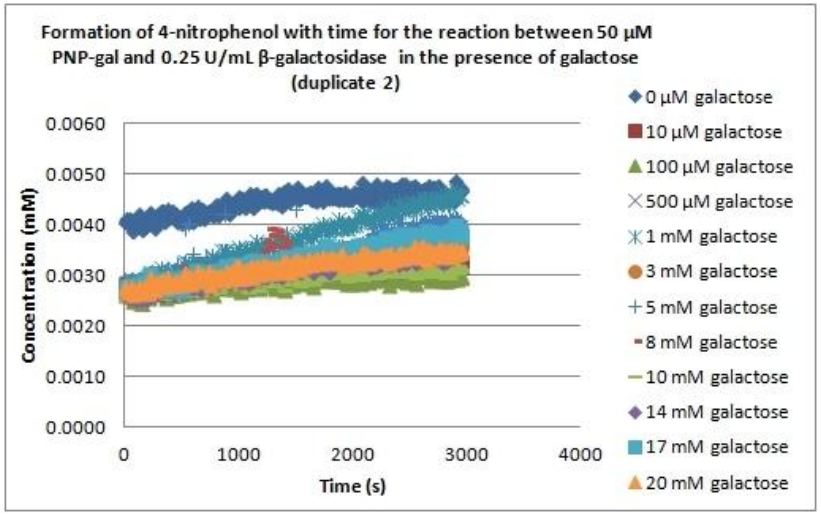

Formation of product with time for β-galactosidase with PNP-galactopyranoside in the presence of galactose
Concentration (M)
6.0000E-03
5.0000E-03
4.0000E-03
3.0000E-03
2.0000E-03
1.0000E-03
0.0000E+00
0
500
1000
1500
2000
2500
3000
3500
Time (s)
0 µM galactose
10 µM galactose
100 µM galactose
500 µM galactose
1 mM galactose
3 mM galactose

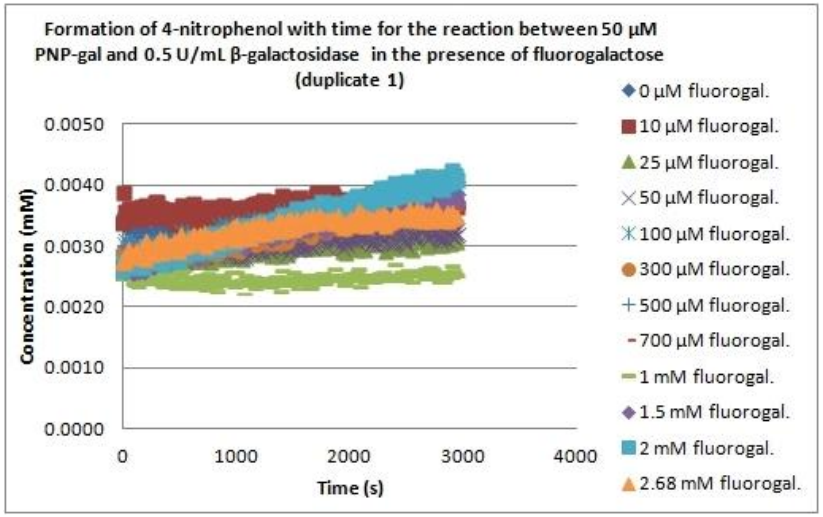

Formation of product with time for β-galactosidase with PNP-galactopyranoside in the presence of galactose
Concentration (M)
6.0000E-03
5.0000E-03
4.0000E-03
3.0000E-03
2.0000E-03
1.0000E-03
0.0000E+00
0
500
1000
1500
2000
2500
3000
3500
Time (s)
0 µM galactose
10 µM galactose
100 µM galactose
500 µM galactose
1 mM galactose
3 mM galactose

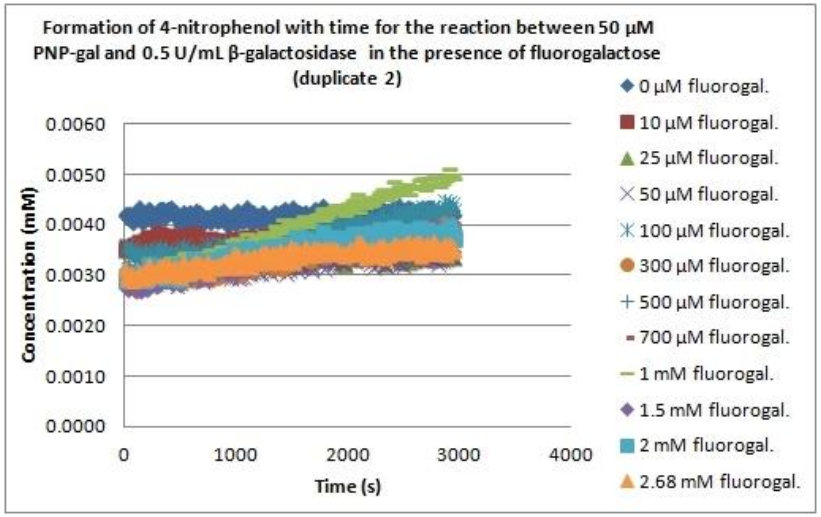

Formation of product with time for the reaction of β-galactosidase with PNP-galactopyranoside in the presence of 2,2,3,3-tetrafluorogalactose
Concentration (M)
4.5000E-03
4.0000E-03
3.5000E-03
3.0000E-03
2.5000E-03
2.0000E-03
1.5000E-03
1.0000E-03
5.0000E-04
0.0000E+00
0
500
1000
1500
2000
2500
3000
3500
Time (s)
0 µM fluorogal.
10 µM fluorogal.
25 µM fluorogal.
50 µM fluorogal.
100 µM fluorogal.
300 µM fluorogal.
500 µM fluorogal.
700 µM fluorogal.
1 mM fluorogal.

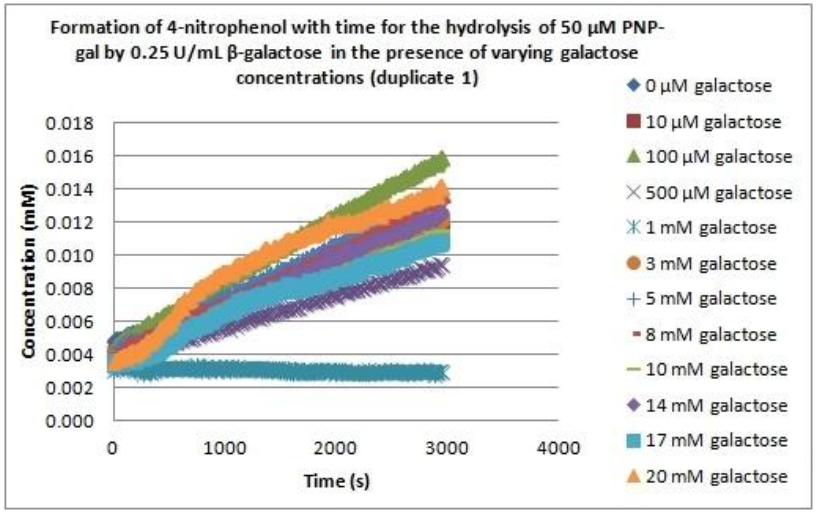

Formation of product with time for the reaction of β-galactosidase with PNP-galactopyranoside in the presence of 2,2,3,3-tetrafluorogalactose
Concentration (M)
6.0000E-03
5.0000E-03
4.0000E-03
3.0000E-03
2.0000E-03
1.0000E-03
0.0000E+00
0
500
1000
1500
2000
2500
3000
3500
Time (s)
0 µM fluorogal.
10 µM fluorogal.
25 µM fluorogal.
50 µM fluorogal.
100 µM fluorogal.
300 µM fluorogal.
500 µM fluorogal.
700 µM fluorogal.
1 mM fluorogal.

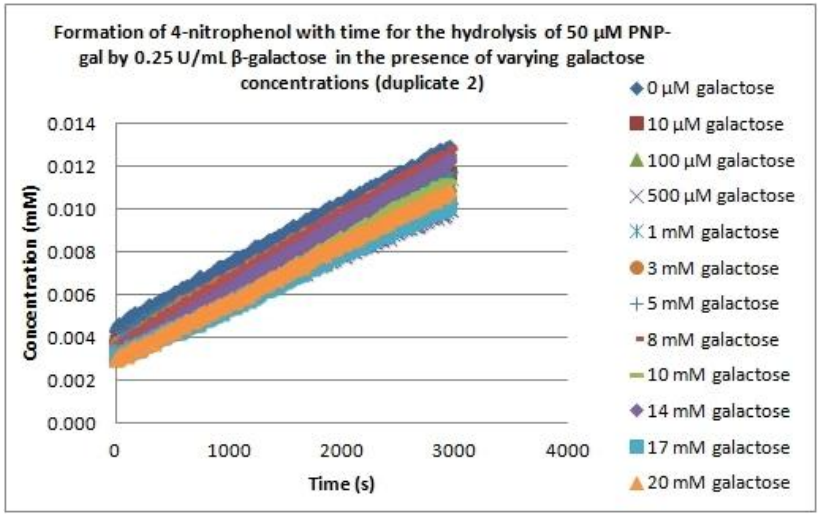

Formation of 4-nitrophenol with time for the reaction between 50 µM PNP-gal and 0.25 U/mL β-galactosidase in the presence of galactose (duplicate 1)
Concentration (mM)
0.0060
0.0050
0.0040
0.0030
0.0020
0.0010
0.0000
0
1000
2000
3000
4000
Time (s)
0 µM galactose
10 µM galactose
100 µM galactose
500 µM galactose
1 mM galactose
3 mM galactose
5 mM galactose
8 mM galactose
10 mM galactose
14 mM galactose
17 mM galactose
20 mM galactose

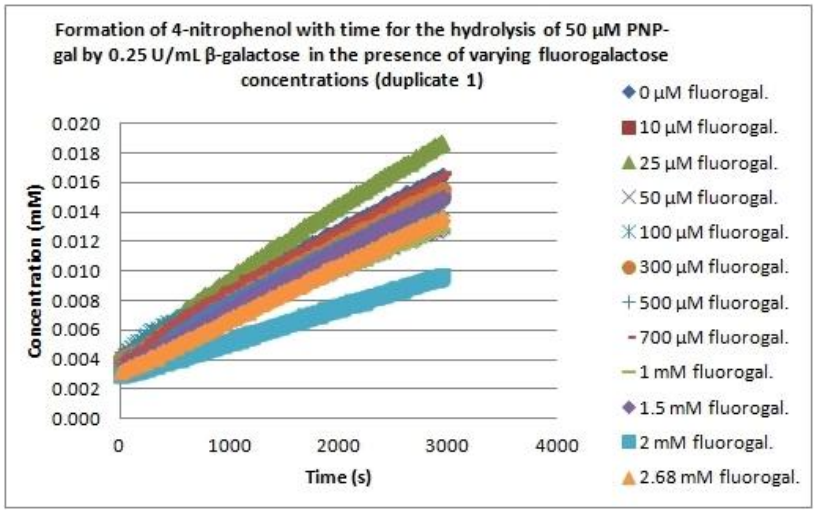

Formation of 4-nitrophenol with time for the reaction between 50 µM PNP-gal and 0.25 U/mL β-galactosidase in the presence of galactose (duplicate 2)
Concentration (mM)
0.0060
0.0050
0.0040
0.0030
0.0020
0.0010
0.0000
0
1000
2000
3000
4000
Time (s)
0 µM galactose
10 µM galactose
100 µM galactose
500 µM galactose
1 mM galactose
3 mM galactose
5 mM galactose
8 mM galactose
10 mM galactose
14 mM galactose
17 mM galactose
20 mM galactose

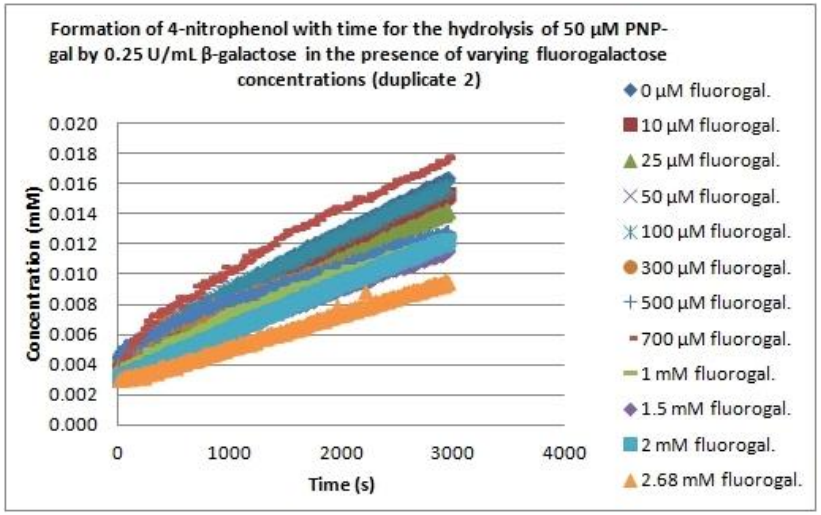

Formation of 4-nitrophenol with time for the reaction between 50 µM PNP-gal and 0.5 U/mL β-galactosidase in the presence of fluorogalactose (duplicate 1)
Concentration (mM)
0.0050
0.0040
0.0030
0.0020
0.0010
0.0000
0
1000
2000
3000
4000
Time (s)
0 µM fluorogal.
10 µM fluorogal.
25 µM fluorogal.
50 µM fluorogal.
100 µM fluorogal.
300 µM fluorogal.
500 µM fluorogal.
700 µM fluorogal.
1 mM fluorogal.
1.5 mM fluorogal.
2 mM fluorogal.
2.68 mM fluorogal.

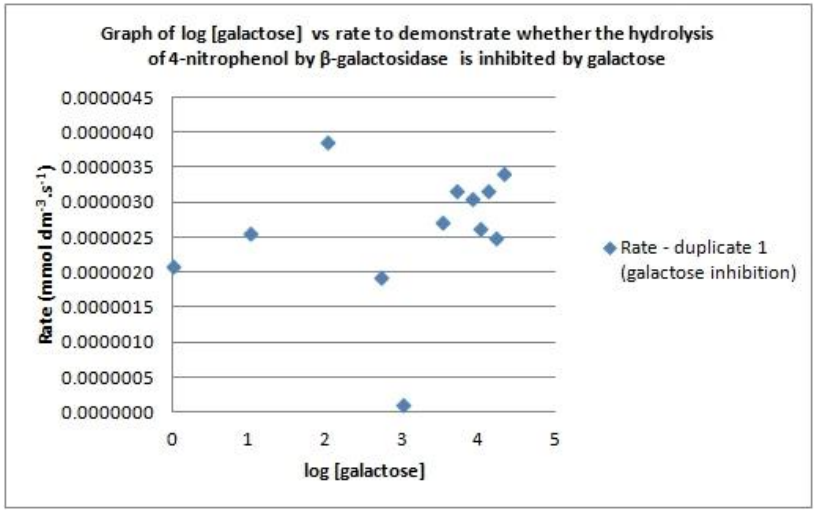

Formation of 4-nitrophenol with time for the reaction between 50 µM PNP-gal and 0.5 U/mL β-galactosidase in the presence of fluorogalactose (duplicate 2)
Concentration (mM)
0.0060
0.0050
0.0040
0.0030
0.0020
0.0010
0.0000
0
1000
2000
3000
4000
Time (s)
0 µM fluorogal.
10 µM fluorogal.
25 µM fluorogal.
50 µM fluorogal.
100 µM fluorogal.
300 µM fluorogal.
500 µM fluorogal.
700 µM fluorogal.
1 mM fluorogal.
1.5 mM fluorogal.
2 mM fluorogal.
2.68 mM fluorogal.

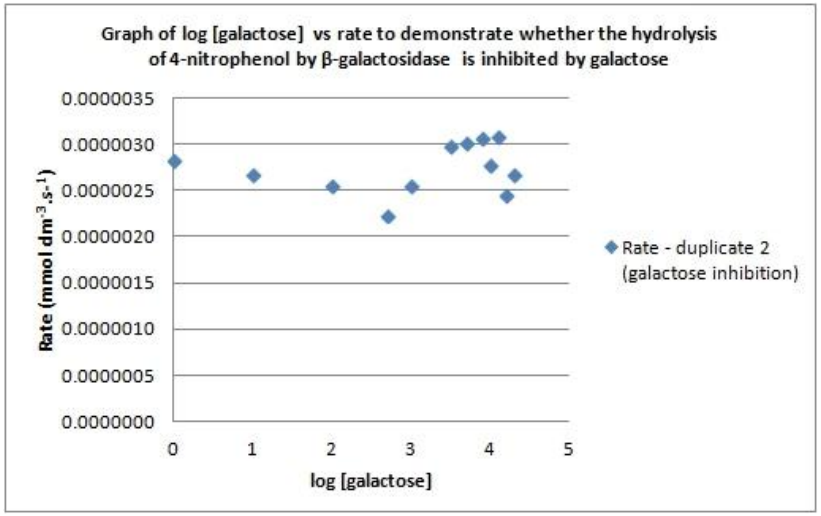

Formation of 4-nitrophenol with time for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactose in the presence of varying galactose concentrations (duplicate 1)
Concentration (mM)
Time (s)
0 µM galactose
10 µM galactose
100 µM galactose
500 µM galactose
1 mM galactose
3 mM galactose
5 mM galactose
8 mM galactose
10 mM galactose
14 mM galactose
17 mM galactose
20 mM galactose

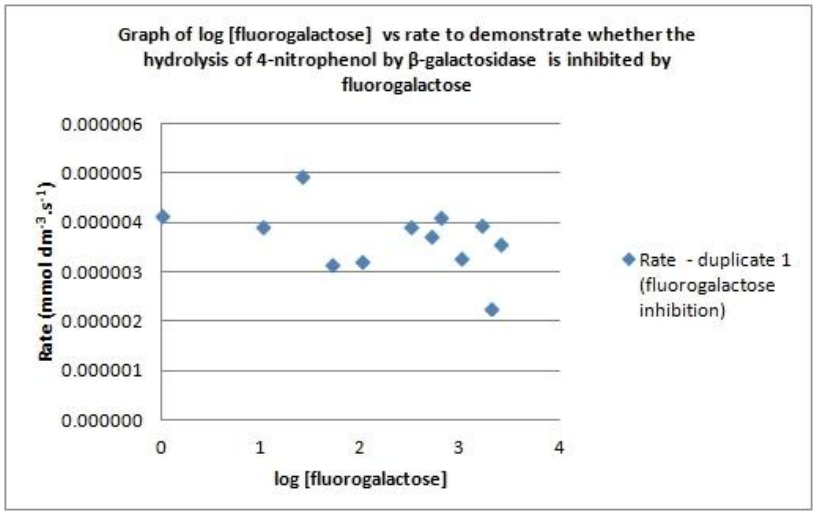

Formation of 4-nitrophenol with time for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactose in the presence of varying galactose concentrations (duplicate 2)
Concentration (mM)
Time (s)
0 µM galactose
10 µM galactose
100 µM galactose
500 µM galactose
1 mM galactose
3 mM galactose
5 mM galactose
8 mM galactose
10 mM galactose
14 mM galactose
17 mM galactose
20 mM galactose

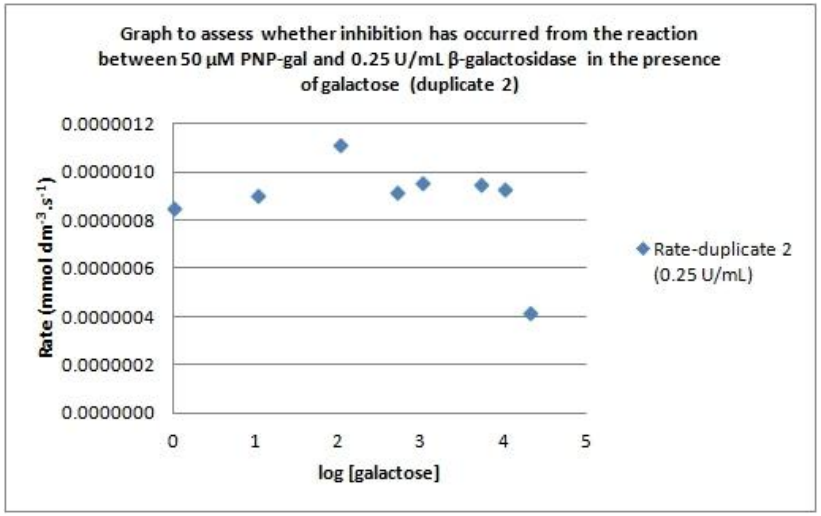

Formation of 4-nitrophenol with time for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactose in the presence of varying fluorogalactose concentrations (duplicate 1)
Concentration (mM)
Time (s)
0 µM fluorogal.
10 µM fluorogal.
25 µM fluorogal.
50 µM fluorogal.
100 µM fluorogal.
300 µM fluorogal.
500 µM fluorogal.
700 µM fluorogal.
1 mM fluorogal.
1.5 mM fluorogal.
2 mM fluorogal.
2.68 mM fluorogal.

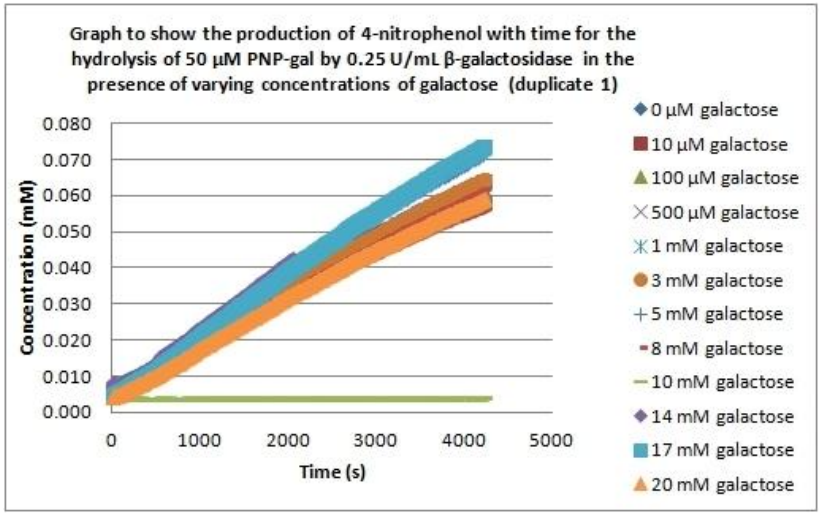

Formation of 4-nitrophenol with time for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactose in the presence of varying fluorogalactose concentrations (duplicate 2)
Concentration (mM)
Time (s)
0 µM fluorogal.
10 µM fluorogal.
25 µM fluorogal.
50 µM fluorogal.
100 µM fluorogal.
300 µM fluorogal.
500 µM fluorogal.
700 µM fluorogal.
1 mM fluorogal.
1.5 mM fluorogal.
2 mM fluorogal.
2.68 mM fluorogal.

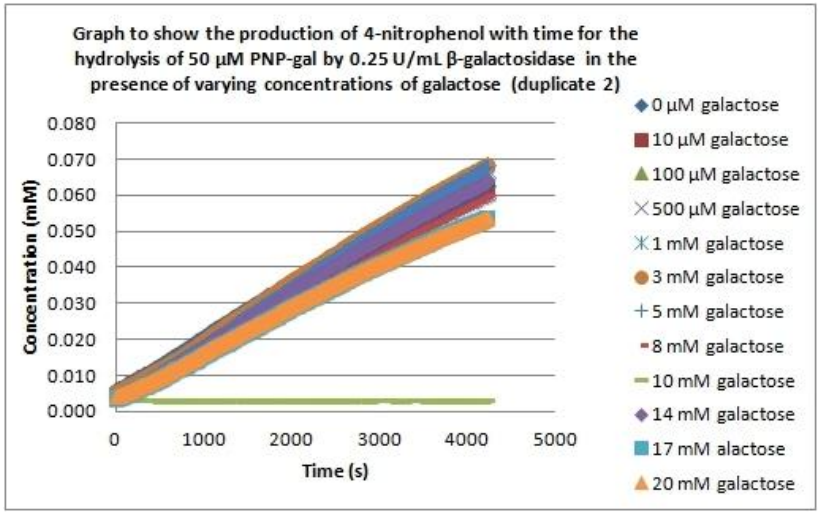

Graph of log [galactose] vs rate to demonstrate whether the hydrolysis of 4-nitrophenol by β-galactosidase is inhibited by galactose
Rate (mmol dm-3.s-1)
0.0000045
0.0000040
0.0000035
0.0000030
0.0000025
0.0000020
0.0000015
0.0000010
0.0000005
0.0000000
0
1
2
3
4
5
log [galactose]
Rate - duplicate 1 (galactose inhibition)

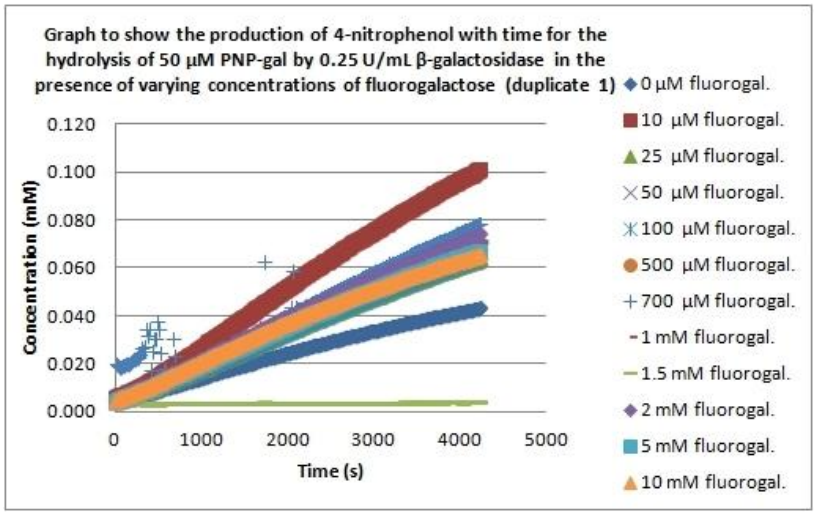

Graph of log [galactose] vs rate to demonstrate whether the hydrolysis of 4-nitrophenol by β-galactosidase is inhibited by galactose
Rate (mmol dm-3.s-1)
0.0000035
0.0000030
0.0000025
0.0000020
0.0000015
0.0000010
0.0000005
0.0000000
0
1
2
3
4
5
log [galactose]
Rate - duplicate 2 (galactose inhibition)

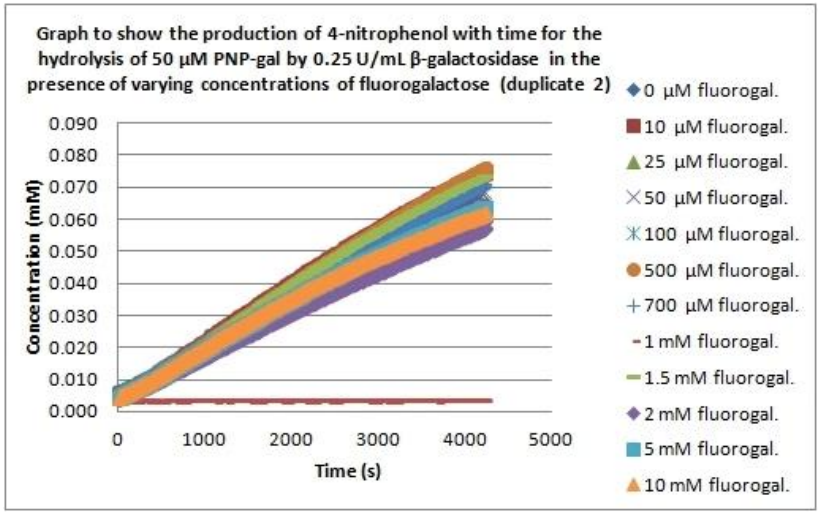

Graph of log [fluorogalactose] vs rate to demonstrate whether the hydrolysis of 4-nitrophenol by β-galactosidase is inhibited by fluorogalactose
Rate (mmol dm-3.s-1)
0.000006
0.000005
0.000004
0.000003
0.000002
0.000001
0.000000
0
1
2
3
4
log [fluorogalactose]
Rate - duplicate 1 (fluorogalactose inhibition)

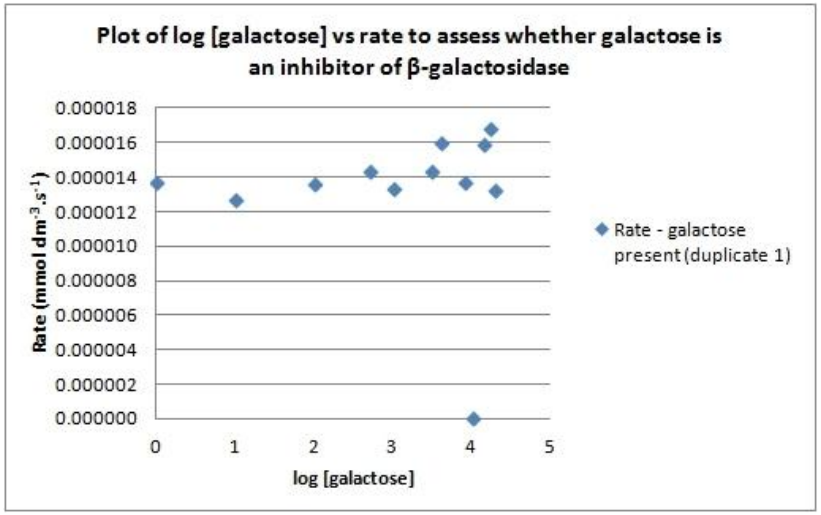

Graph to assess whether inhibition has occurred from the reaction between 50 µM PNP-gal and 0.25 U/mL β-galactosidase in the presence of galactose (duplicate 2)
Rate (mmol dm-3.s-1)
0.0000012
0.0000010
0.0000008
0.0000006
0.0000004
0.0000002
0.0000000
0
1
2
3
4
5
log [galactose]
Rate-duplicate 2 (0.25 U/mL)

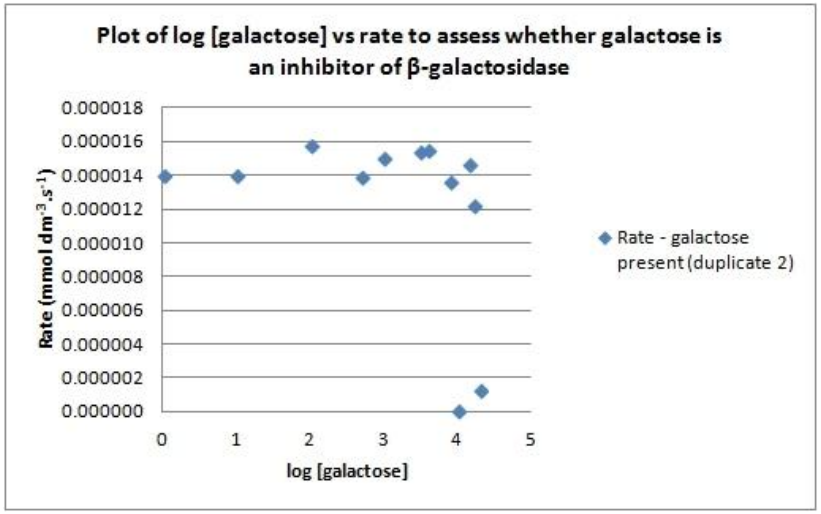

Graph to show the production of 4-nitrophenol with time for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of varying concentrations of galactose (duplicate 1)
Concentration (mM)
0.080
0.070
0.060
0.050
0.040
0.030
0.020
0.010
0.000
0
1000
2000
3000
4000
5000
Time (s)
0 µM galactose
10 µM galactose
100 µM galactose
500 µM galactose
1 mM galactose
3 mM galactose
5 mM galactose
8 mM galactose
10 mM galactose
14 mM galactose
17 mM galactose
20 mM galactose

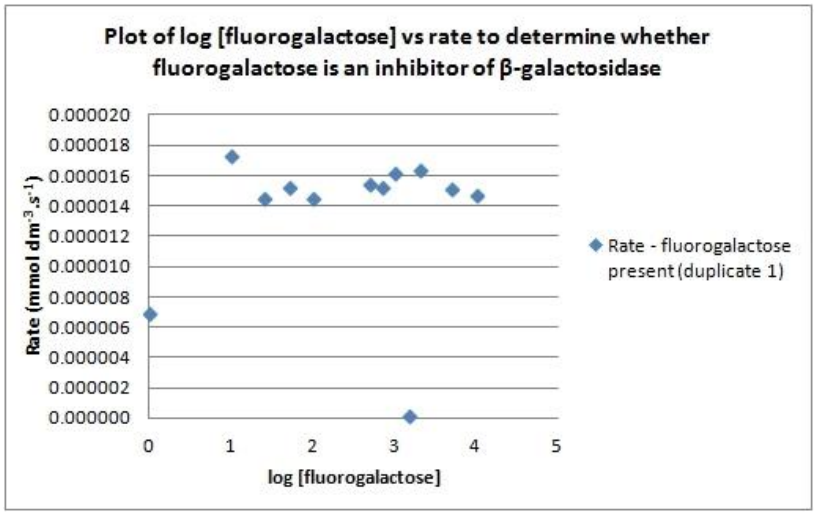

Graph to show the production of 4-nitrophenol with time for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of varying concentrations of galactose (duplicate 2)
Concentration (mM)
0.080
0.070
0.060
0.050
0.040
0.030
0.020
0.010
0.000
0
1000
2000
3000
4000
5000
Time (s)
0 µM galactose
10 µM galactose
100 µM galactose
500 µM galactose
1 mM galactose
3 mM galactose
5 mM galactose
8 mM galactose
10 mM galactose
14 mM galactose
17 mM alactose
20 mM galactose

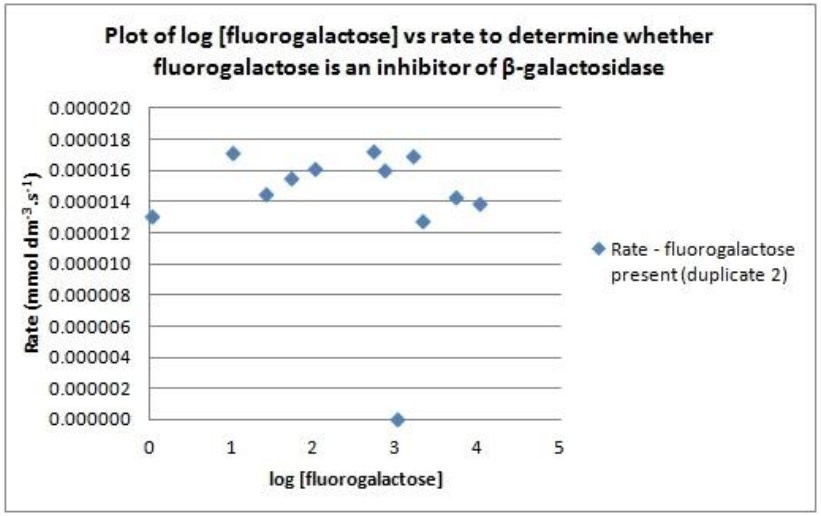

Graph to show the production of 4-nitrophenol with time for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of varying concentrations of fluorogalactose (duplicate 1)
Concentration (mM)
0.120
0.100
0.080
0.060
0.040
0.020
0.000
0
1000
2000
3000
4000
5000
Time (s)
0 µM fluorogal.
10 µM fluorogal.
25 µM fluorogal.
50 µM fluorogal.
100 µM fluorogal.
500 µM fluorogal.
700 µM fluorogal.
1 mM fluorogal.
1.5 mM fluorogal.
2 mM fluorogal.
5 mM fluorogal.
10 mM fluorogal.

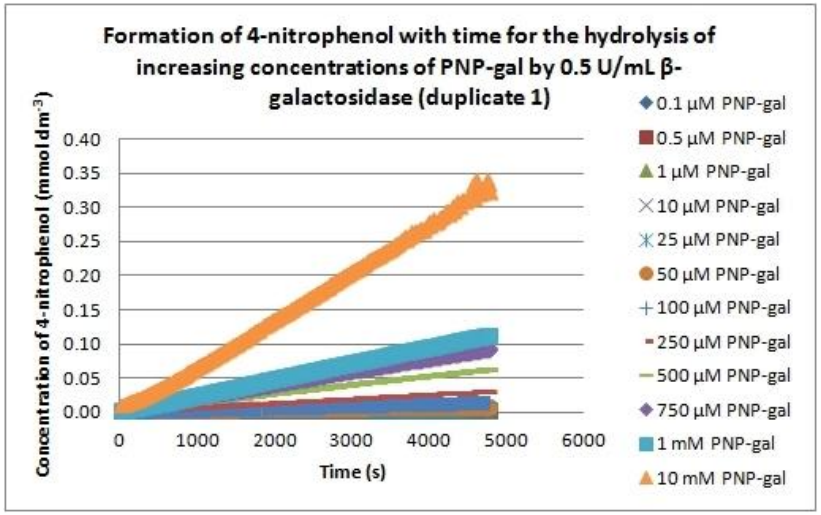

Graph to show the production of 4-nitrophenol with time for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of varying concentrations of fluorogalactose (duplicate 2)
Concentration (mM)
0.090
0.080
0.070
0.060
0.050
0.040
0.030
0.020
0.010
0.000
0
1000
2000
3000
4000
5000
Time (s)
0 µM fluorogal.
10 µM fluorogal.
25 µM fluorogal.
50 µM fluorogal.
100 µM fluorogal.
500 µM fluorogal.
700 µM fluorogal.
1 mM fluorogal.
1.5 mM fluorogal.
2 mM fluorogal.
5 mM fluorogal.
10 mM fluorogal.

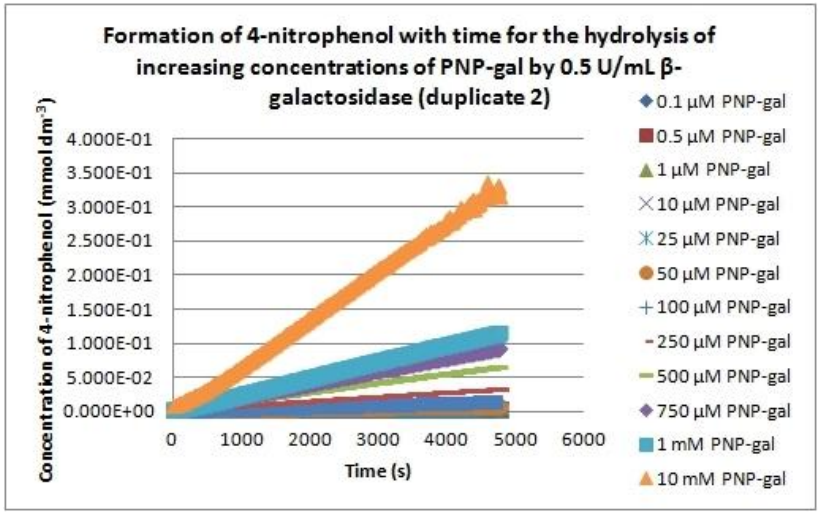

Plot of log [galactose] vs rate to assess whether galactose is an inhibitor of β-galactosidase
Rate (mmol dm-3.s-1)
0.000018
0.000016
0.000014
0.000012
0.000010
0.000008
0.000006
0.000004
0.000002
0.000000
0
1
2
3
4
5
log [galactose]
Rate - galactose present (duplicate 1)

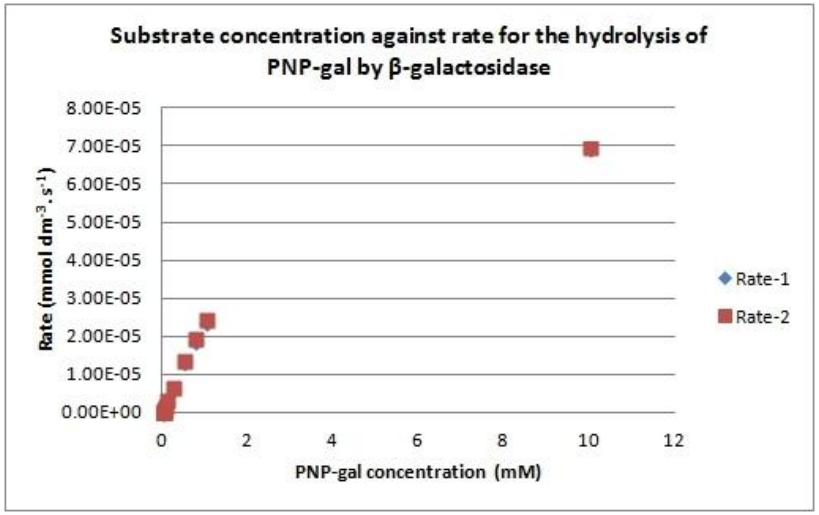

Plot of log [galactose] vs rate to assess whether galactose is an inhibitor of β-galactosidase
Rate (mmol dm-3.s-1)
0.000018
0.000016
0.000014
0.000012
0.000010
0.000008
0.000006
0.000004
0.000002
0.000000
0
1
2
3
4
5
log [galactose]
Rate - galactose present (duplicate 2)

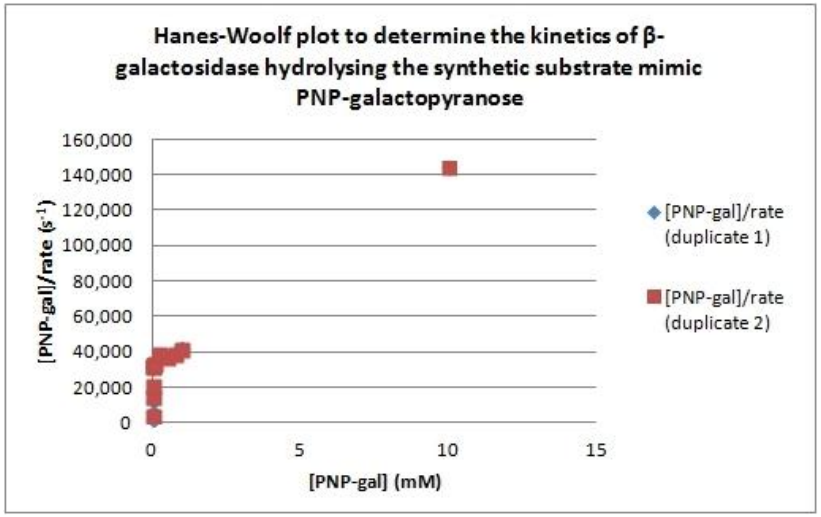

Plot of log [fluorogalactose] vs rate to determine whether fluorogalactose is an inhibitor of β-galactosidase
Rate (mmol dm-3.s-1)
0.000020
0.000018
0.000016
0.000014
0.000012
0.000010
0.000008
0.000006
0.000004
0.000002
0.000000
0
1
2
3
4
5
log [fluorogalactose]
Rate - fluorogalactose present (duplicate 1)

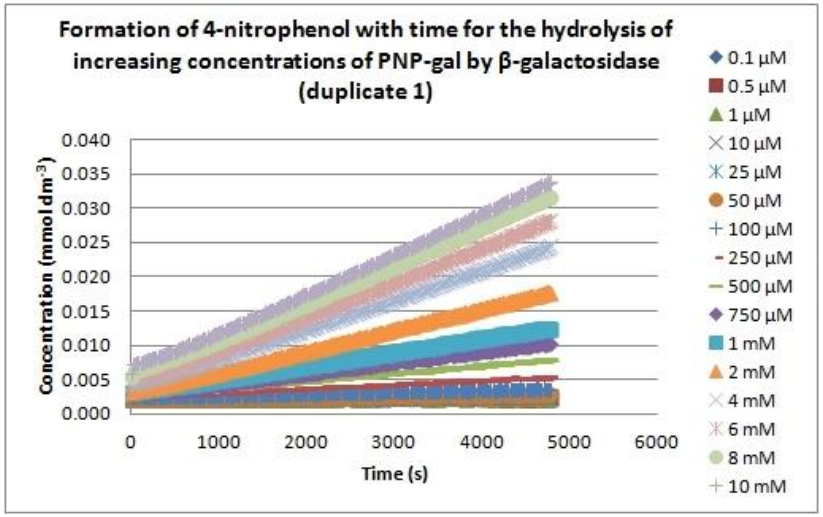

Plot of log [fluorogalactose] vs rate to determine whether fluorogalactose is an inhibitor of β-galactosidase
Rate (mmol dm-3.s-1)
0.000020
0.000018
0.000016
0.000014
0.000012
0.000010
0.000008
0.000006
0.000004
0.000002
0.000000
0
1
2
3
4
5
log [fluorogalactose]
Rate - fluorogalactose present (duplicate 2)

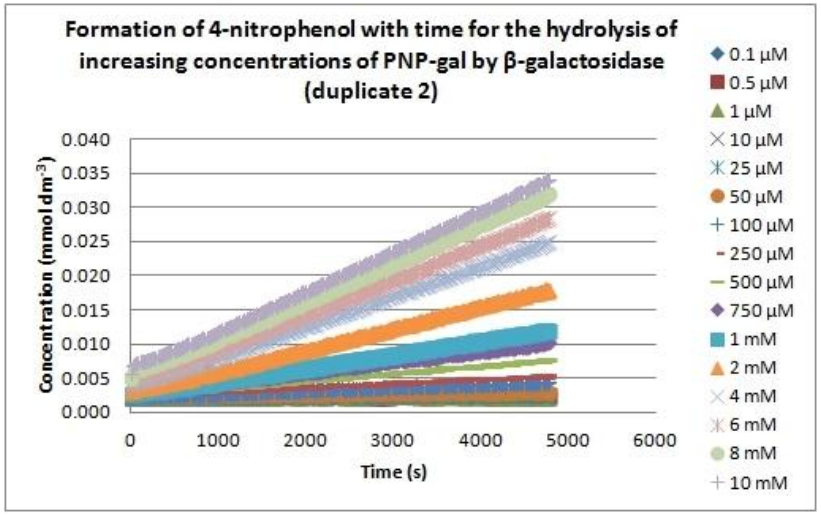

Formation of 4-nitrophenol with time for the hydrolysis of increasing concentrations of PNP-gal by 0.5 U/mL β-galactosidase (duplicate 1)
Concentration of 4-nitrophenol (mmol dm-3)
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
1000
2000
3000
4000
5000
6000
Time (s)
0.1 µM PNP-gal
0.5 µM PNP-gal
1 µM PNP-gal
10 µM PNP-gal
25 µM PNP-gal
50 µM PNP-gal
100 µM PNP-gal
250 µM PNP-gal
500 µM PNP-gal
750 µM PNP-gal
1 mM PNP-gal
10 mM PNP-gal

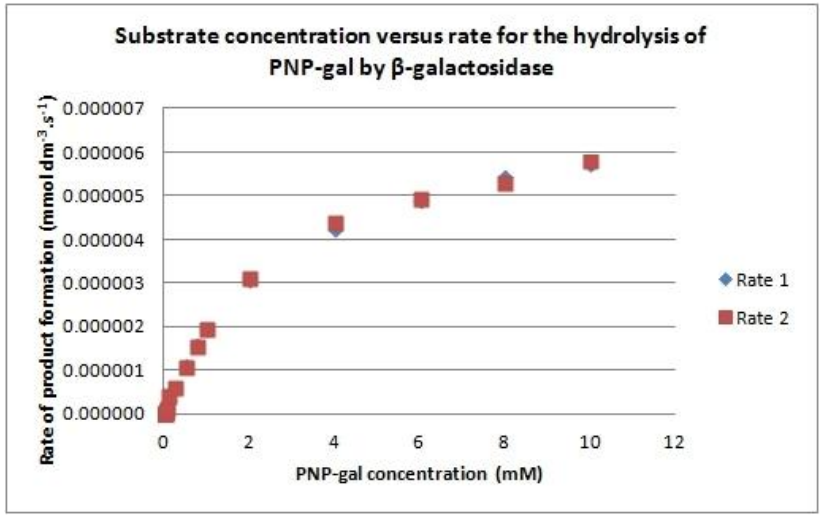

Formation of 4-nitrophenol with time for the hydrolysis of increasing concentrations of PNP-gal by 0.5 U/mL β-galactosidase (duplicate 2)
Concentration of 4-nitrophenol (mmol dm-3)
4.000E-01
3.500E-01
3.000E-01
2.500E-01
2.000E-01
1.500E-01
1.000E-01
5.000E-02
0.000E+00
0
1000
2000
3000
4000
5000
6000
Time (s)
0.1 µM PNP-gal
0.5 µM PNP-gal
1 µM PNP-gal
10 µM PNP-gal
25 µM PNP-gal
50 µM PNP-gal
100 µM PNP-gal
250 µM PNP-gal
500 µM PNP-gal
750 µM PNP-gal
1 mM PNP-gal
10 mM PNP-gal

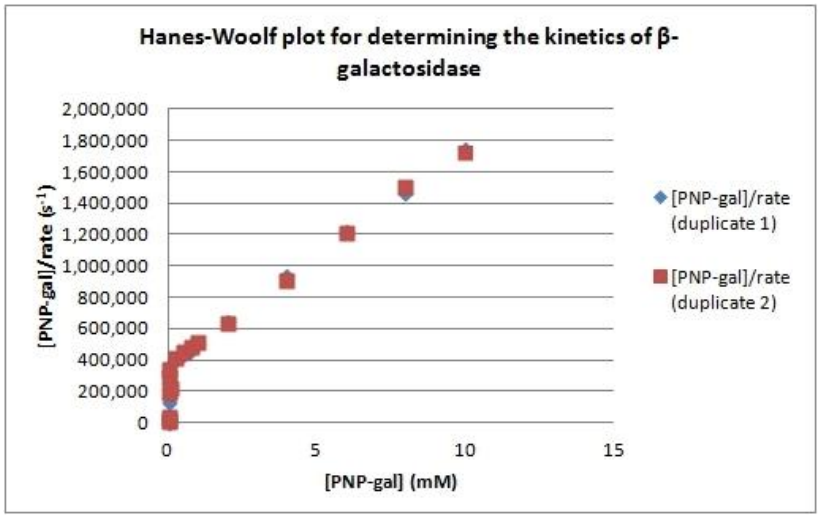

Substrate concentration against rate for the hydrolysis of PNP-gal by β-galactosidase
Rate (mmol dm-3 . s-1)
8.00E-05
7.00E-05
6.00E-05
5.00E-05
4.00E-05
3.00E-05
2.00E-05
1.00E-05
0.00E+00
0
2
4
6
8
10
12
PNP-gal concentration (mM)
Rate-1
Rate-2

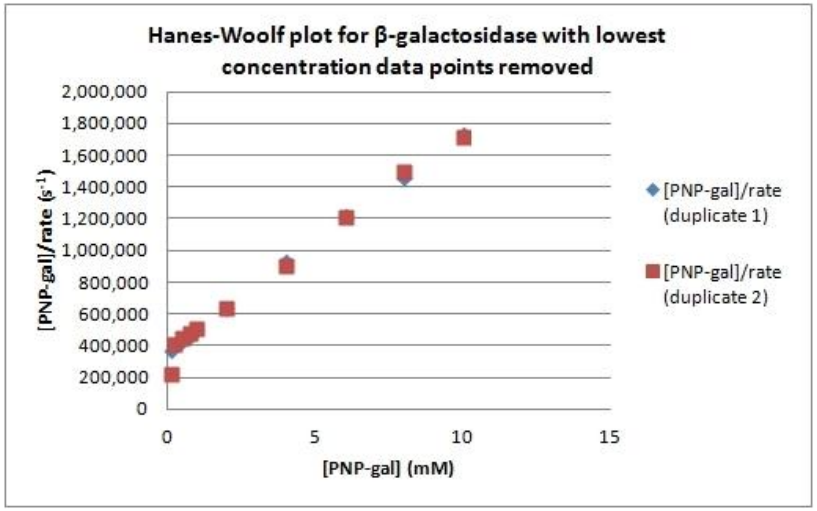

Hanes-Woolf plot to determine the kinetics of β-galactosidase hydrolysing the synthetic substrate mimic PNP-galactopyranose
[PNP-gal]/rate (s-1)
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
0
5
10
15
[PNP-gal] (mM)
[PNP-gal]/rate (duplicate 1)
[PNP-gal]/rate (duplicate 2)

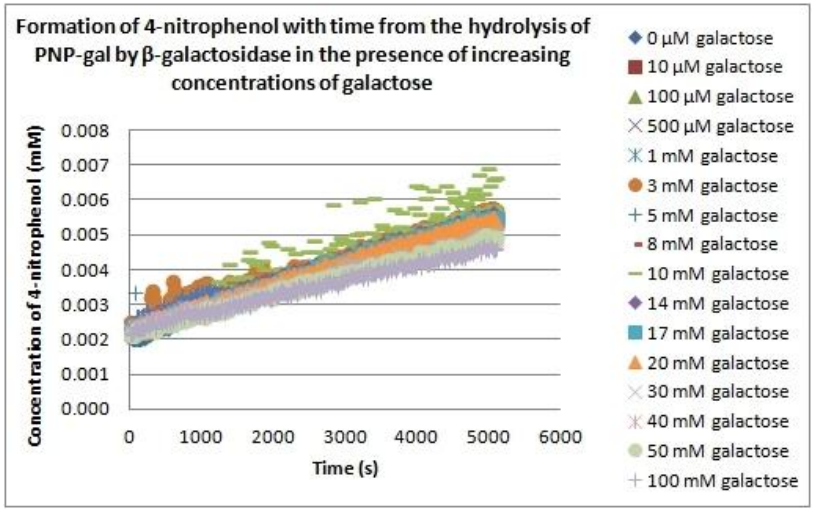

Formation of 4-nitrophenol with time for the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase (duplicate 1)
Concentration (mmol dm-3)
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
1000
2000
3000
4000
5000
6000
Time (s)
0.1 µM
0.5 µM
1 µM
10 µM
25 µM
50 µM
100 µM
250 µM
500 µM
750 µM
1 mM
2 mM
4 mM
6 mM
8 mM
10 mM

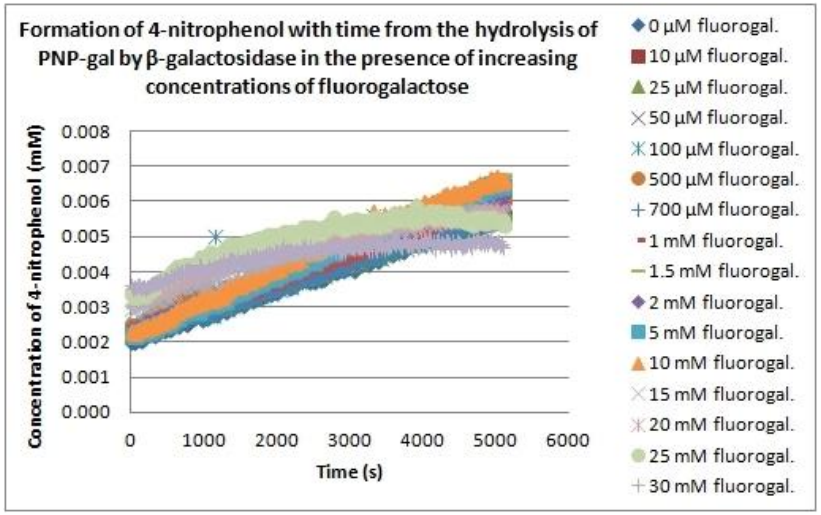

Formation of 4-nitrophenol with time for the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase (duplicate 2)
Concentration (mmol dm-3)
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
1000
2000
3000
4000
5000
6000
Time (s)
0.1 µM
0.5 µM
1 µM
10 µM
25 µM
50 µM
100 µM
250 µM
500 µM
750 µM
1 mM
2 mM
4 mM
6 mM
8 mM
10 mM

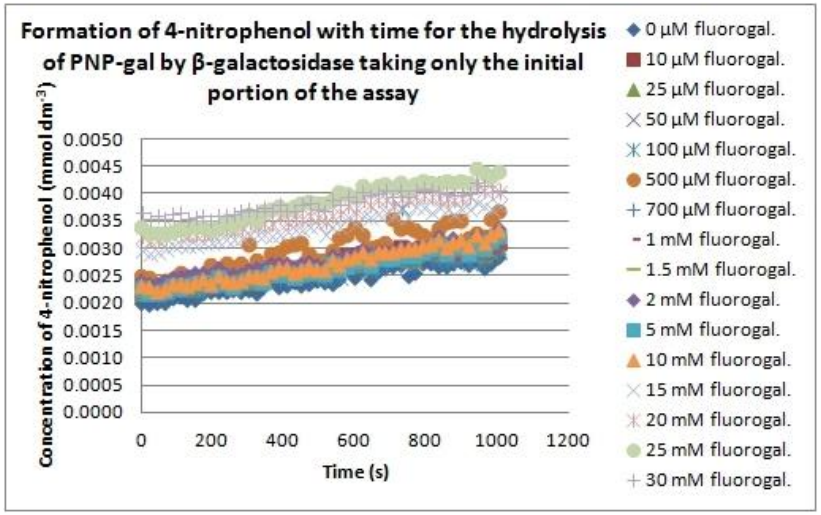

Substrate concentration versus rate for the hydrolysis of PNP-gal by β-galactosidase
Rate of product formation (mmol dm-3.s-1)
0.000007
0.000006
0.000005
0.000004
0.000003
0.000002
0.000001
0.000000
0
2
4
6
8
10
12
PNP-gal concentration (mM)
Rate 1
Rate 2

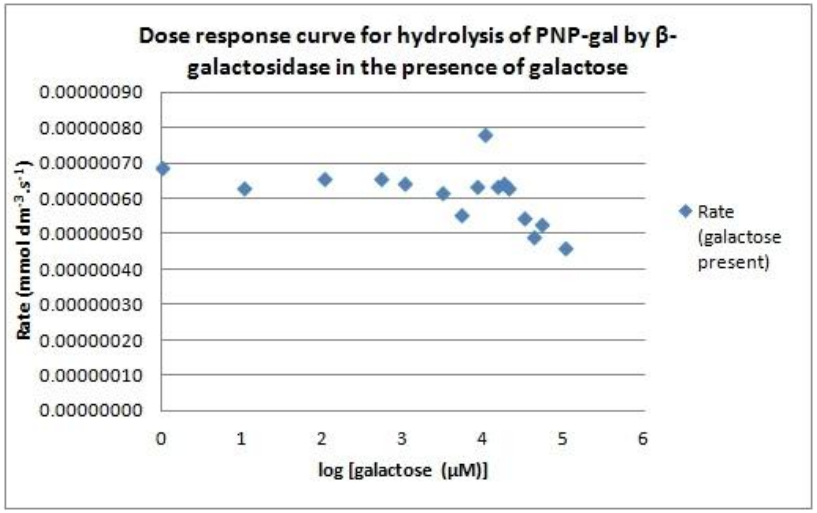

Hanes-Woolf plot for determining the kinetics of β-galactosidase
[PNP-gal]/rate (s-1)
2,000,000
1,800,000
1,600,000
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
0
5
10
15
[PNP-gal] (mM)
[PNP-gal]/rate (duplicate 1)
[PNP-gal]/rate (duplicate 2)

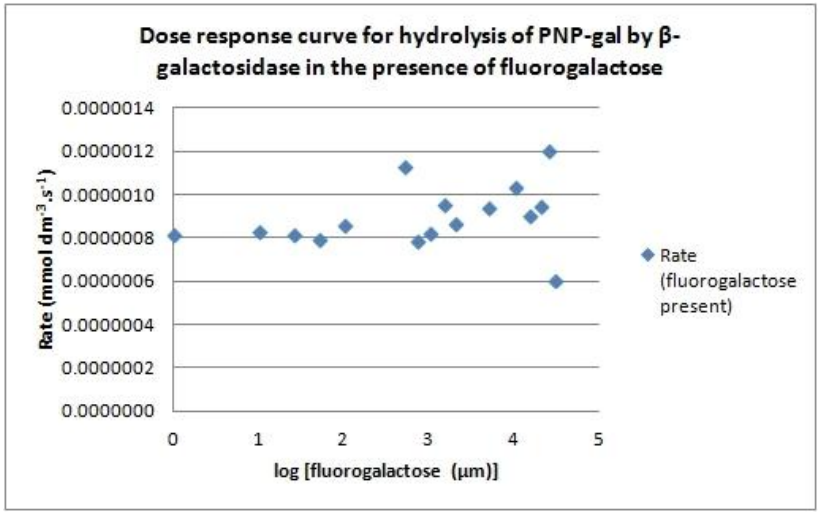

Hanes-Woolf plot for β-galactosidase with lowest concentration data points removed
[PNP-gal]/rate (s-1)
2,000,000
1,800,000
1,600,000
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
0
5
10
15
[PNP-gal] (mM)
[PNP-gal]/rate (duplicate 1)
[PNP-gal]/rate (duplicate 2)

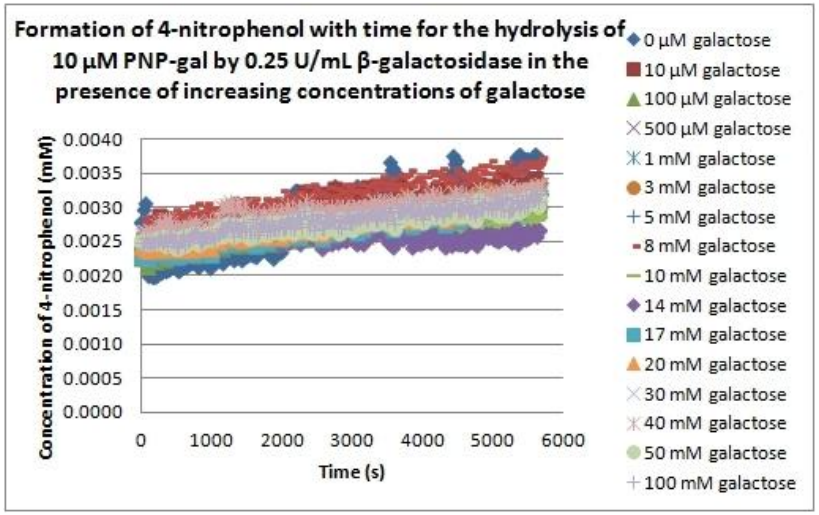

Formation of 4-nitrophenol with time from the hydrolysis of PNP-gal by β-galactosidase in the presence of increasing concentrations of galactose
Concentration of 4-nitrophenol (mM)
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
1000
2000
3000
4000
5000
6000
Time (s)
0 µM galactose
10 µM galactose
100 µM galactose
500 µM galactose
1 mM galactose
3 mM galactose
5 mM galactose
8 mM galactose
10 mM galactose
14 mM galactose
17 mM galactose
20 mM galactose
30 mM galactose
40 mM galactose
50 mM galactose
100 mM galactose

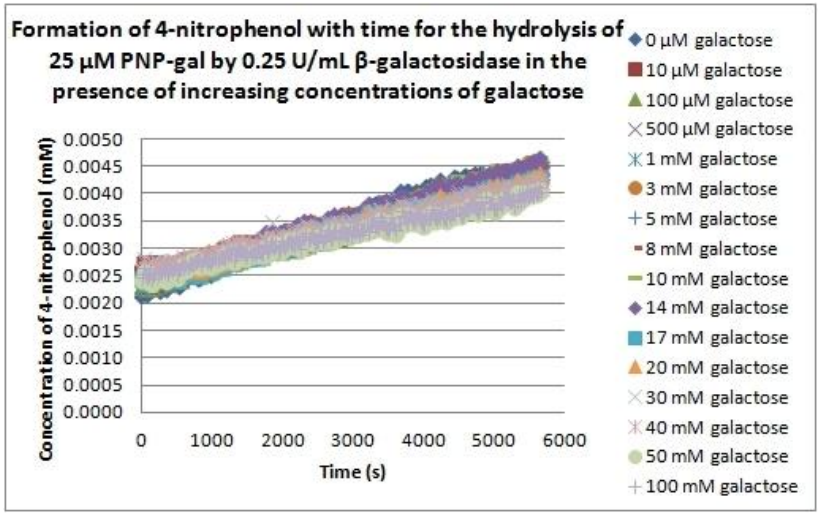

Formation of 4-nitrophenol with time from the hydrolysis of PNP-gal by β-galactosidase in the presence of increasing concentrations of fluorogalactose
Concentration of 4-nitrophenol (mM)
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
1000
2000
3000
4000
5000
6000
Time (s)
0 µM fluorogal.
10 µM fluorogal.
25 µM fluorogal.
50 µM fluorogal.
100 µM fluorogal.
500 µM fluorogal.
700 µM fluorogal.
1 mM fluorogal.
1.5 mM fluorogal.
2 mM fluorogal.
5 mM fluorogal.
10 mM fluorogal.
15 mM fluorogal.
20 mM fluorogal.
25 mM fluorogal.
30 mM fluorogal.

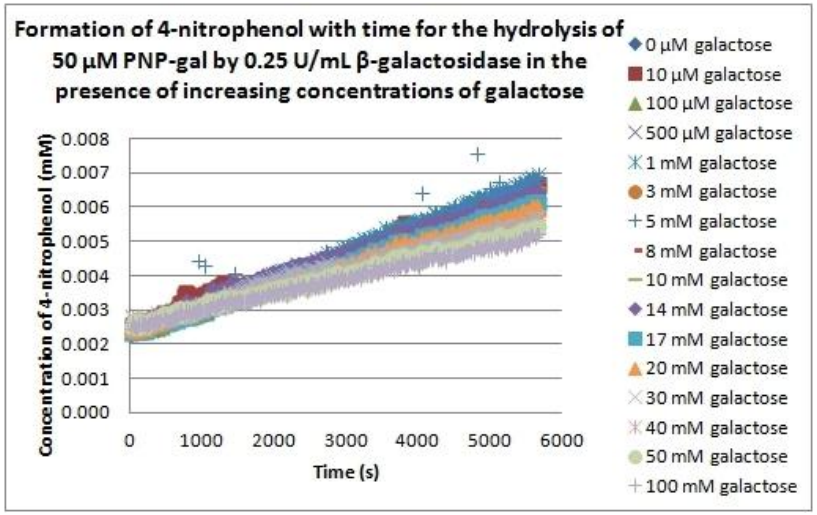

Formation of 4-nitrophenol with time for the hydrolysis of PNP-gal by β-galactosidase taking only the initial portion of the assay
Concentration of 4-nitrophenol (mmol dm-3)
0.0050
0.0045
0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0
200
400
600
800
1000
1200
Time (s)
0 µM fluorogal.
10 µM fluorogal.
25 µM fluorogal.
50 µM fluorogal.
100 µM fluorogal.
500 µM fluorogal.
700 µM fluorogal.
1 mM fluorogal.
1.5 mM fluorogal.
2 mM fluorogal.
5 mM fluorogal.
10 mM fluorogal.
15 mM fluorogal.
20 mM fluorogal.
25 mM fluorogal.
30 mM fluorogal.

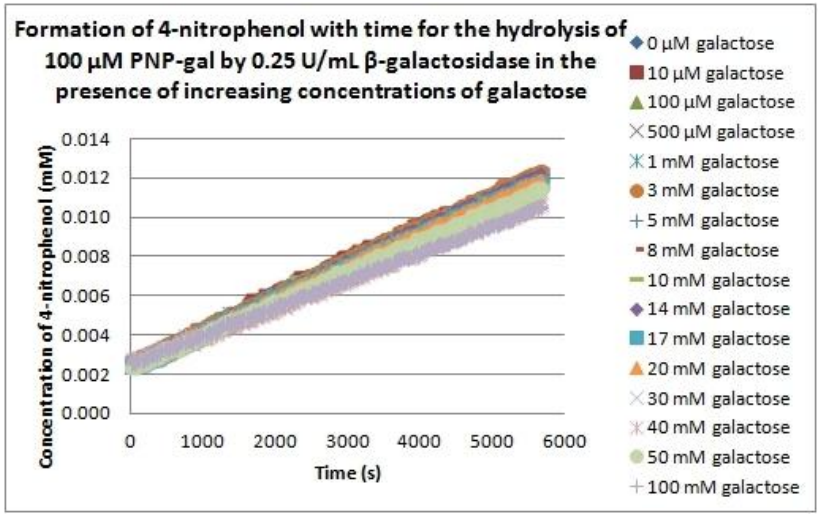

Dose response curve for hydrolysis of PNP-gal by β-galactosidase in the presence of galactose
Rate (mmol dm-3.s-1)
0.00000090
0.00000080
0.00000070
0.00000060
0.00000050
0.00000040
0.00000030
0.00000020
0.00000010
0.00000000
0
1
2
3
4
5
6
log [galactose (μM)]
Rate (galactose present)

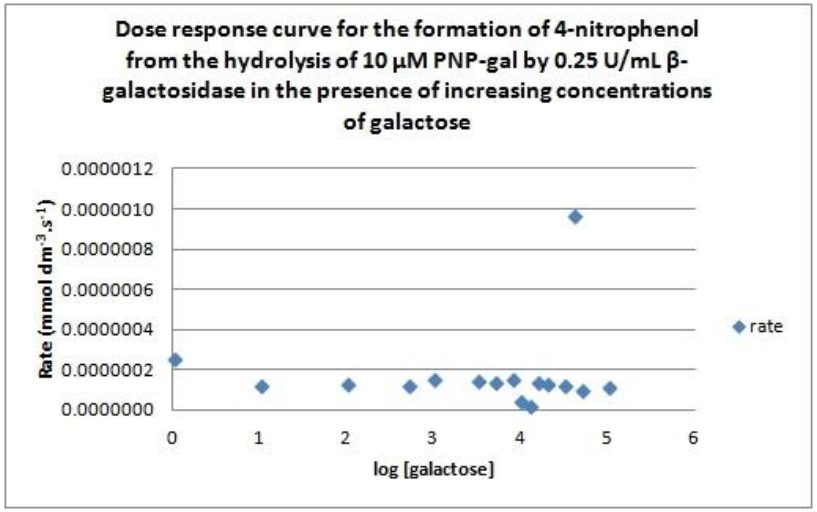

Dose response curve for hydrolysis of PNP-gal by β-galactosidase in the presence of fluorogalactose
Rate (mmol dm-3.s-1)
0.0000014
0.0000012
0.0000010
0.0000008
0.0000006
0.0000004
0.0000002
0.0000000
0
1
2
3
4
5
log [fluorogalactose (μm)]
Rate (fluorogalactose present)

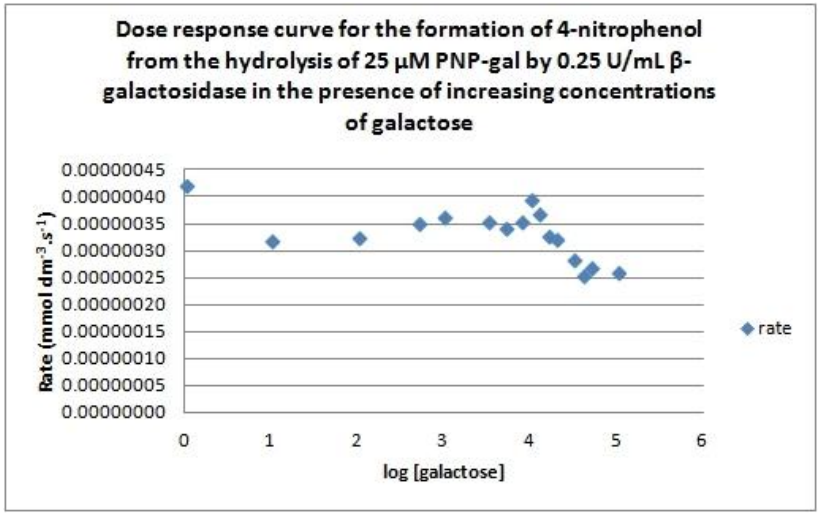

Formation of 4-nitrophenol with time for the hydrolysis of 10 μM PNP-gal by 0.25 U/mL β-galactosidase in the presence of increasing concentrations of galactose
Concentration of 4-nitrophenol (mM)
0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0
1000
2000
3000
4000
5000
6000
Time (s)
0 μM galactose
10 μM galactose
100 μM galactose
500 μM galactose
1 mM galactose
3 mM galactose
5 mM galactose
8 mM galactose
10 mM galactose
14 mM galactose
17 mM galactose
20 mM galactose
30 mM galactose
40 mM galactose
50 mM galactose
100 mM galactose

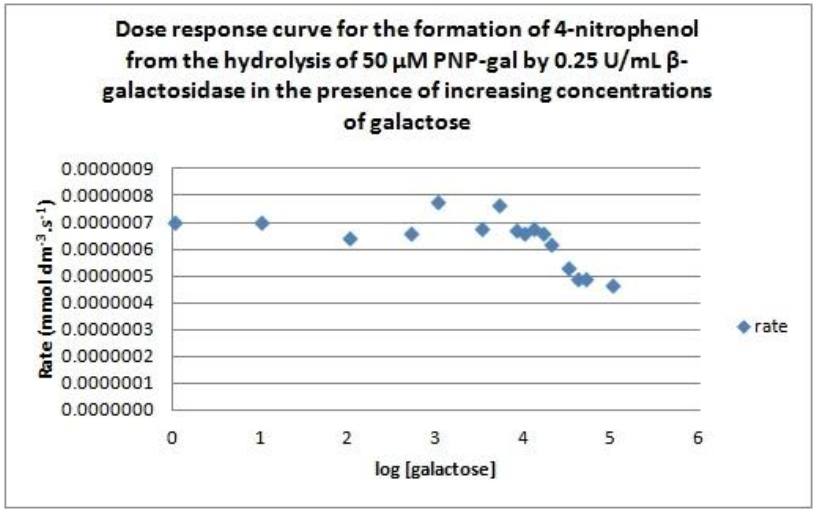

Formation of 4-nitrophenol with time for the hydrolysis of 25 μM PNP-gal by 0.25 U/mL β-galactosidase in the presence of increasing concentrations of galactose
Concentration of 4-nitrophenol (mM)
0.0050
0.0045
0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0
1000
2000
3000
4000
5000
6000
Time (s)
0 μM galactose
10 μM galactose
100 μM galactose
500 μM galactose
1 mM galactose
3 mM galactose
5 mM galactose
8 mM galactose
10 mM galactose
14 mM galactose
17 mM galactose
20 mM galactose
30 mM galactose
40 mM galactose
50 mM galactose
100 mM galactose

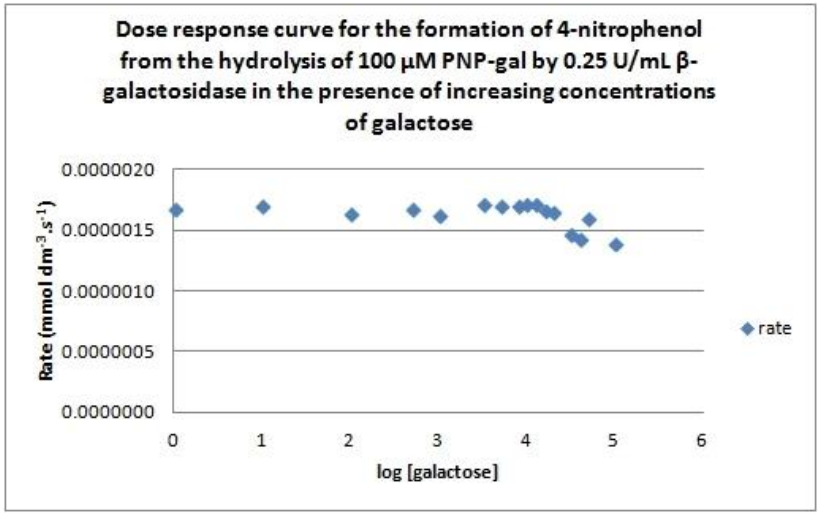

Formation of 4-nitrophenol with time for the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of increasing concentrations of galactose
Concentration of 4-nitrophenol (mM)
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
1000
2000
3000
4000
5000
6000
Time (s)
0 µM galactose
10 µM galactose
100 µM galactose
500 µM galactose
1 mM galactose
3 mM galactose
5 mM galactose
8 mM galactose
10 mM galactose
14 mM galactose
17 mM galactose
20 mM galactose
30 mM galactose
40 mM galactose
50 mM galactose
100 mM galactose

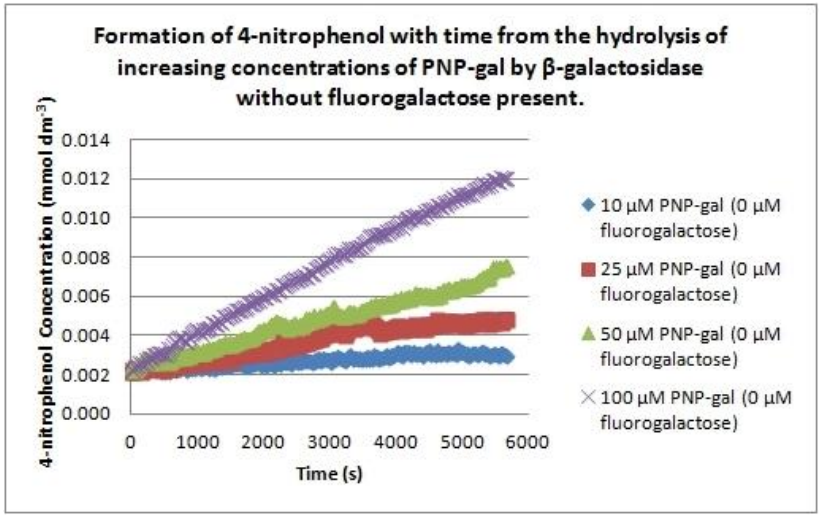

Formation of 4-nitrophenol with time for the hydrolysis of 100 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of increasing concentrations of galactose
Concentration of 4-nitrophenol (mM)
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
1000
2000
3000
4000
5000
6000
Time (s)
0 µM galactose
10 µM galactose
100 µM galactose
500 µM galactose
1 mM galactose
3 mM galactose
5 mM galactose
8 mM galactose
10 mM galactose
14 mM galactose
17 mM galactose
20 mM galactose
30 mM galactose
40 mM galactose
50 mM galactose
100 mM galactose

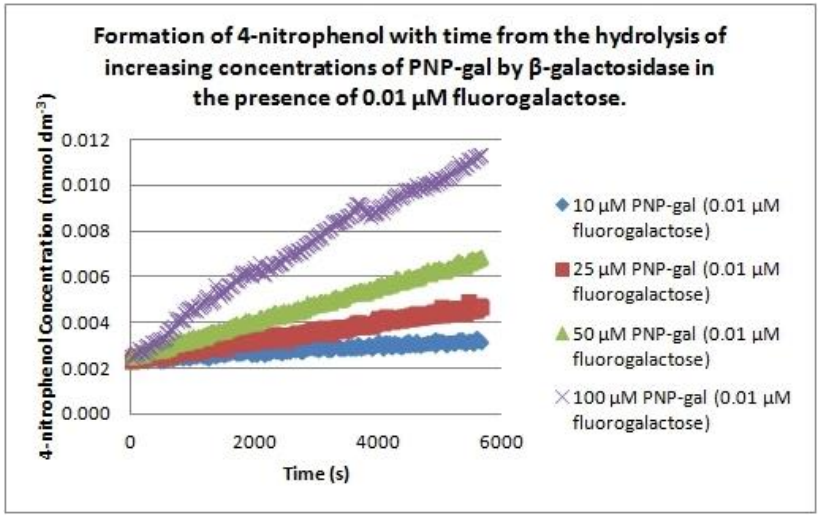

Dose response curve for the formation of 4-nitrophenol from the hydrolysis of 10 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of increasing concentrations of galactose
Rate (mmol dm-3.s-1)
0.0000012
0.0000010
0.0000008
0.0000006
0.0000004
0.0000002
0.0000000
0
1
2
3
4
5
6
log [galactose]
rate

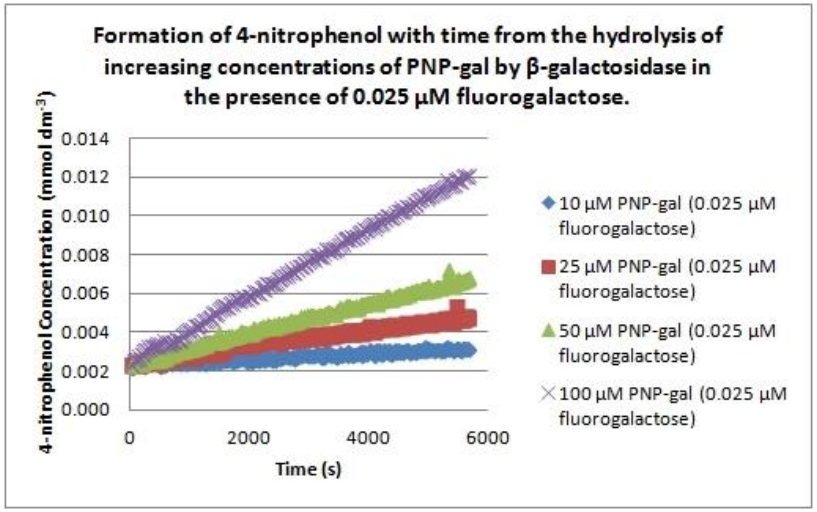

Dose response curve for the formation of 4-nitrophenol from the hydrolysis of 25 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of increasing concentrations of galactose
Rate (mmol dm-3.s-1)
0.00000045
0.00000040
0.00000035
0.00000030
0.00000025
0.00000020
0.00000015
0.00000010
0.00000005
0.00000000
0
1
2
3
4
5
6
log [galactose]
rate

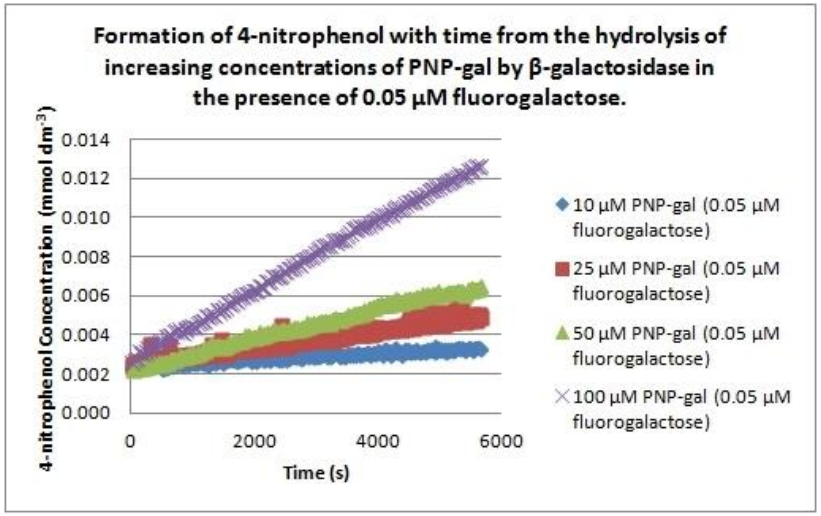

Dose response curve for the formation of 4-nitrophenol from the hydrolysis of 50 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of increasing concentrations of galactose
Rate (mmol dm-3.s-1)
0.0000009
0.0000008
0.0000007
0.0000006
0.0000005
0.0000004
0.0000003
0.0000002
0.0000001
0.0000000
0
1
2
3
4
5
6
log [galactose]
rate

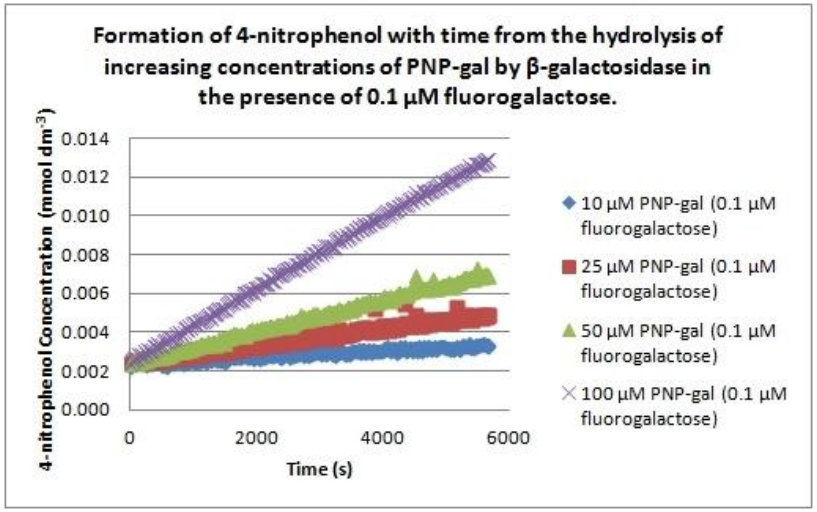

Dose response curve for the formation of 4-nitrophenol from the hydrolysis of 100 µM PNP-gal by 0.25 U/mL β-galactosidase in the presence of increasing concentrations of galactose
Rate (mmol dm-3.s-1)
0.0000020
0.0000015
0.0000010
0.0000005
0.0000000
0
1
2
3
4
5
6
log [galactose]
rate

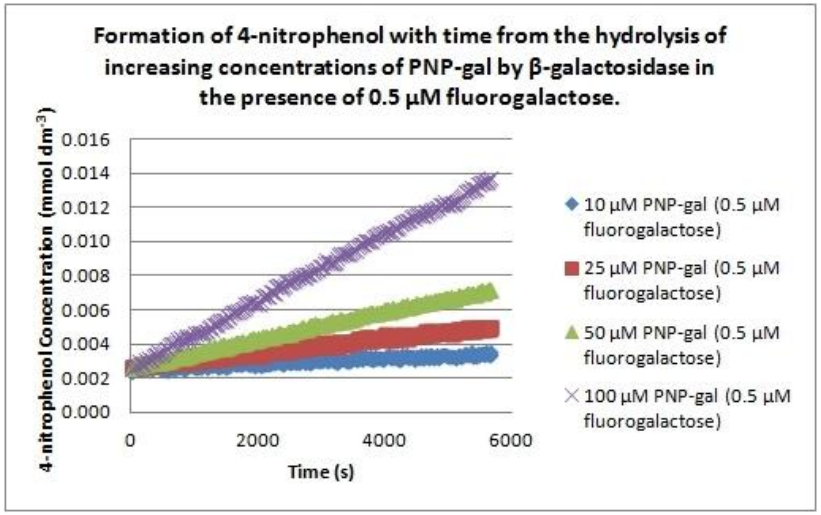

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase without fluorogalactose present.
4-nitrophenol Concentration (mmol dm-3)
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
1000
2000
3000
4000
5000
6000
Time (s)
10 µM PNP-gal (0 µM fluorogalactose)
25 µM PNP-gal (0 µM fluorogalactose)
50 µM PNP-gal (0 µM fluorogalactose)
100 µM PNP-gal (0 µM fluorogalactose)

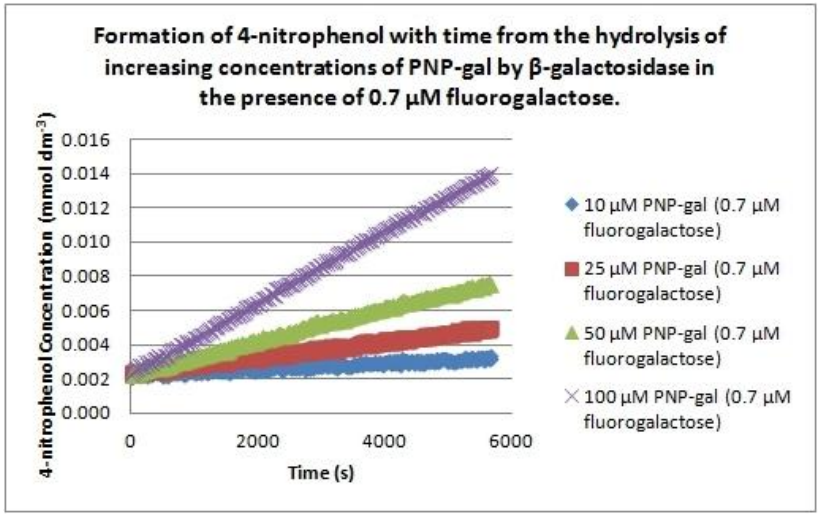

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 0.01 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (0.01 µM fluorogalactose)
25 µM PNP-gal (0.01 µM fluorogalactose)
50 µM PNP-gal (0.01 µM fluorogalactose)
100 µM PNP-gal (0.01 µM fluorogalactose)

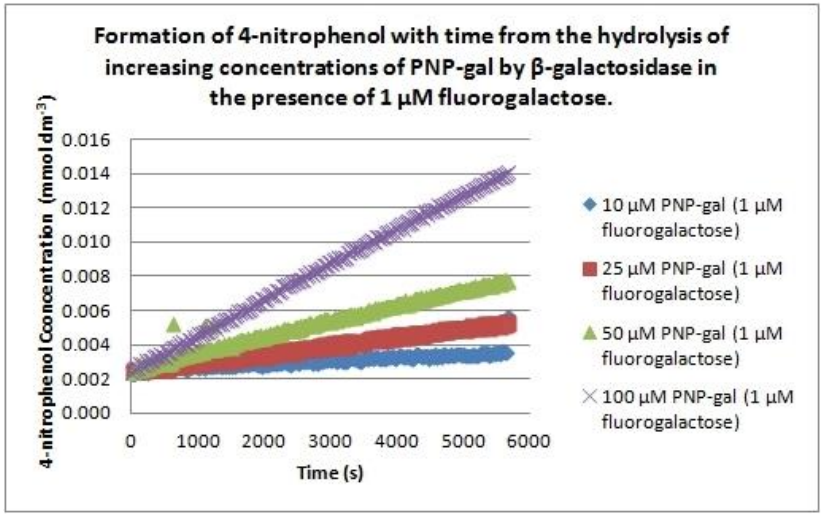

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 0.025 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (0.025 µM fluorogalactose)
25 µM PNP-gal (0.025 µM fluorogalactose)
50 µM PNP-gal (0.025 µM fluorogalactose)
100 µM PNP-gal (0.025 µM fluorogalactose)

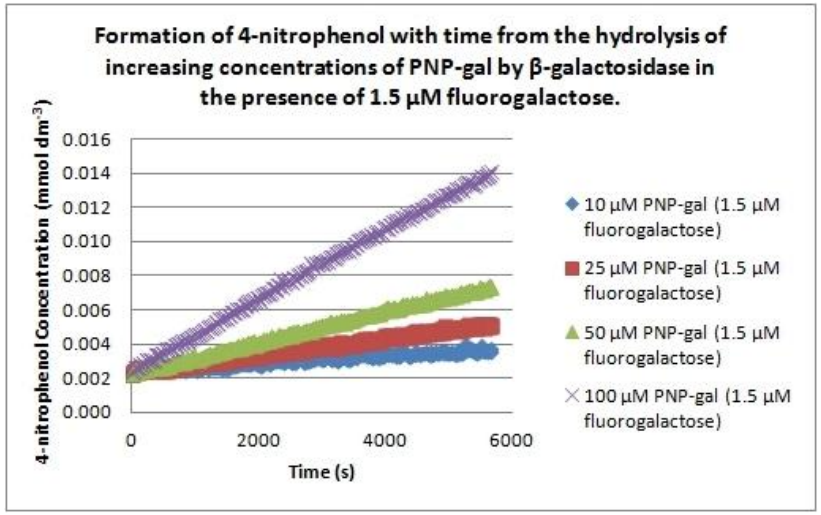

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 0.05 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (0.05 µM fluorogalactose)
25 µM PNP-gal (0.05 µM fluorogalactose)
50 µM PNP-gal (0.05 µM fluorogalactose)
100 µM PNP-gal (0.05 µM fluorogalactose)

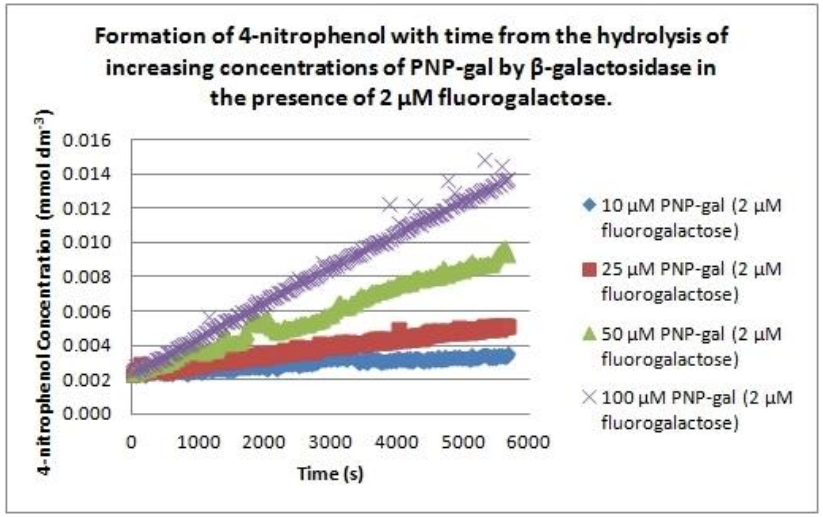

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 0.1 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (0.1 µM fluorogalactose)
25 µM PNP-gal (0.1 µM fluorogalactose)
50 µM PNP-gal (0.1 µM fluorogalactose)
100 µM PNP-gal (0.1 µM fluorogalactose)

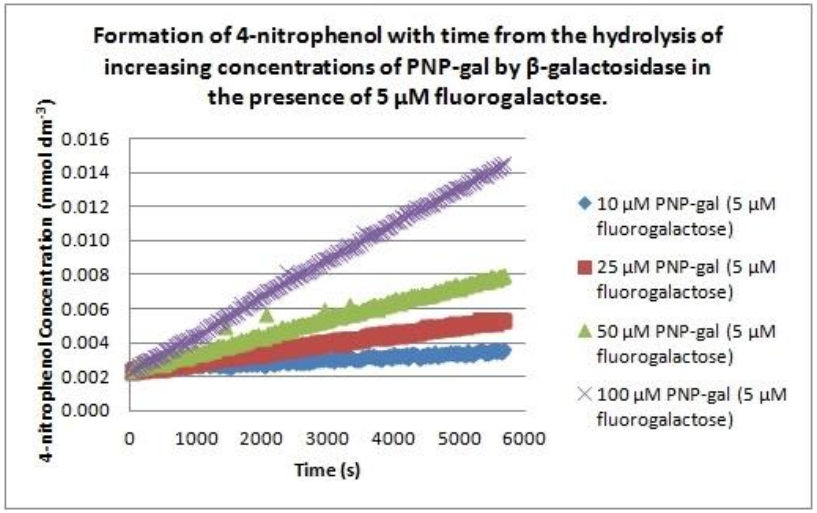

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 0.5 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (0.5 µM fluorogalactose)
25 µM PNP-gal (0.5 µM fluorogalactose)
50 µM PNP-gal (0.5 µM fluorogalactose)
100 µM PNP-gal (0.5 µM fluorogalactose)

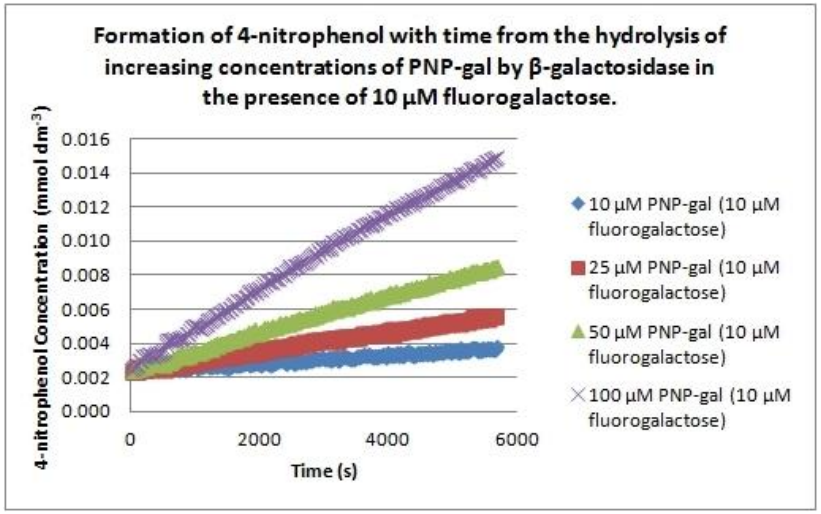

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 0.7 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (0.7 µM fluorogalactose)
25 µM PNP-gal (0.7 µM fluorogalactose)
50 µM PNP-gal (0.7 µM fluorogalactose)
100 µM PNP-gal (0.7 µM fluorogalactose)

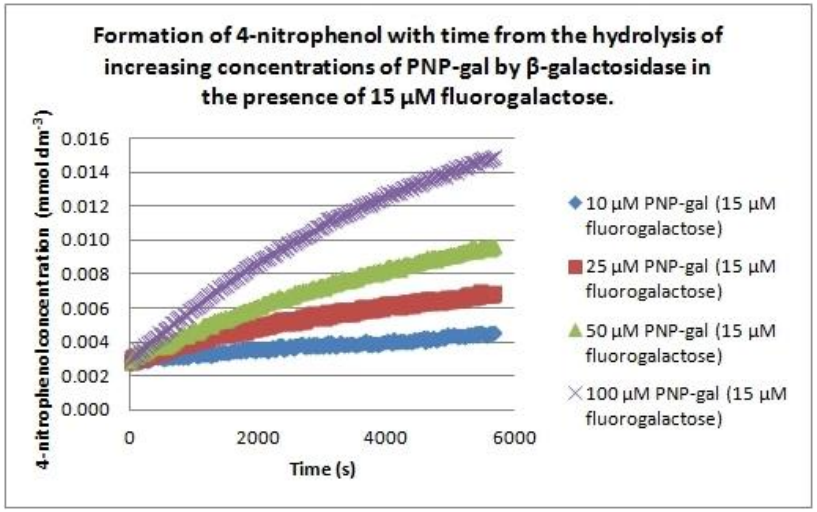

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 1 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
1000
2000
3000
4000
5000
6000
Time (s)
10 µM PNP-gal (1 µM fluorogalactose)
25 µM PNP-gal (1 µM fluorogalactose)
50 µM PNP-gal (1 µM fluorogalactose)
100 µM PNP-gal (1 µM fluorogalactose)

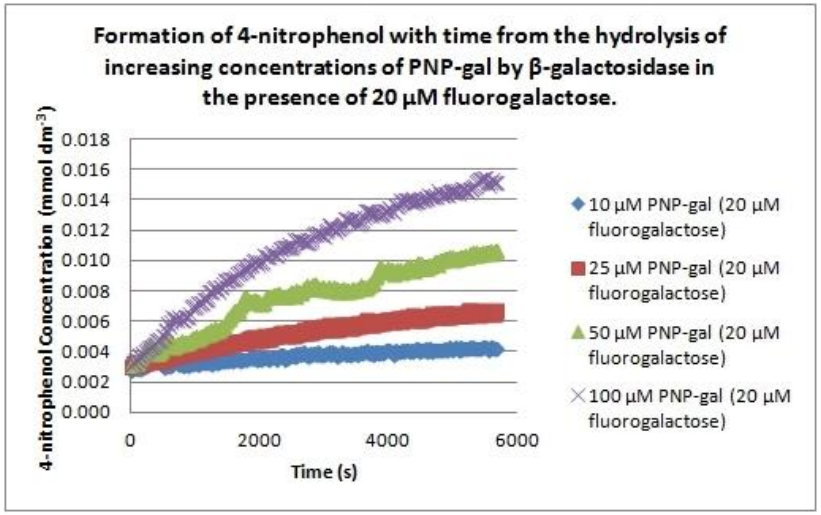

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 1.5 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (1.5 µM fluorogalactose)
25 µM PNP-gal (1.5 µM fluorogalactose)
50 µM PNP-gal (1.5 µM fluorogalactose)
100 µM PNP-gal (1.5 µM fluorogalactose)

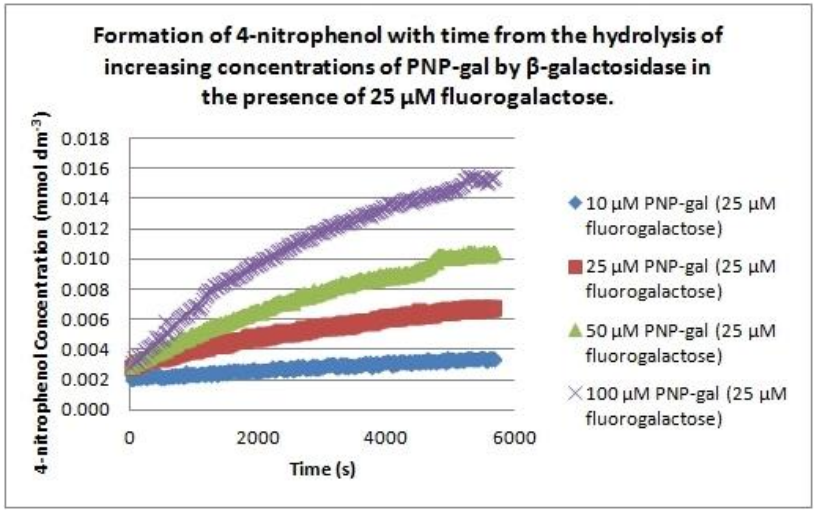

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 2 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
1000
2000
3000
4000
5000
6000
Time (s)
10 µM PNP-gal (2 µM fluorogalactose)
25 µM PNP-gal (2 µM fluorogalactose)
50 µM PNP-gal (2 µM fluorogalactose)
100 µM PNP-gal (2 µM fluorogalactose)

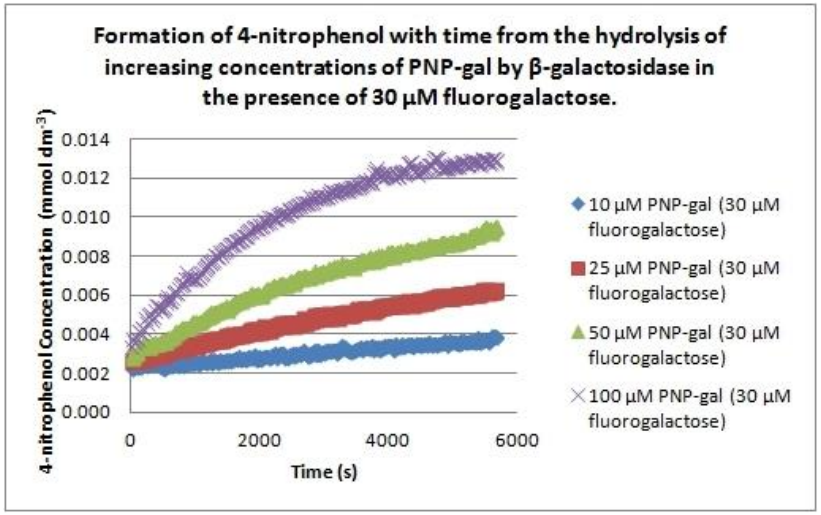

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 5 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
1000
2000
3000
4000
5000
6000
Time (s)
10 µM PNP-gal (5 µM fluorogalactose)
25 µM PNP-gal (5 µM fluorogalactose)
50 µM PNP-gal (5 µM fluorogalactose)
100 µM PNP-gal (5 µM fluorogalactose)

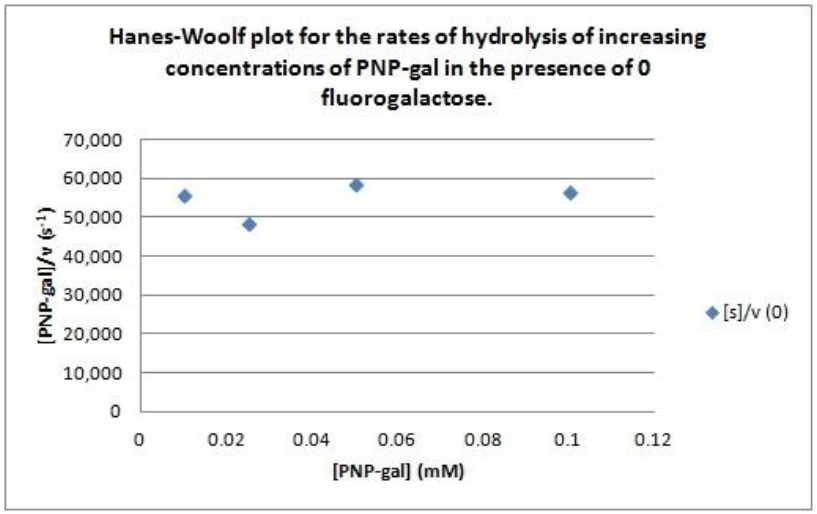

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 10 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (10 µM fluorogalactose)
25 µM PNP-gal (10 µM fluorogalactose)
50 µM PNP-gal (10 µM fluorogalactose)
100 µM PNP-gal (10 µM fluorogalactose)

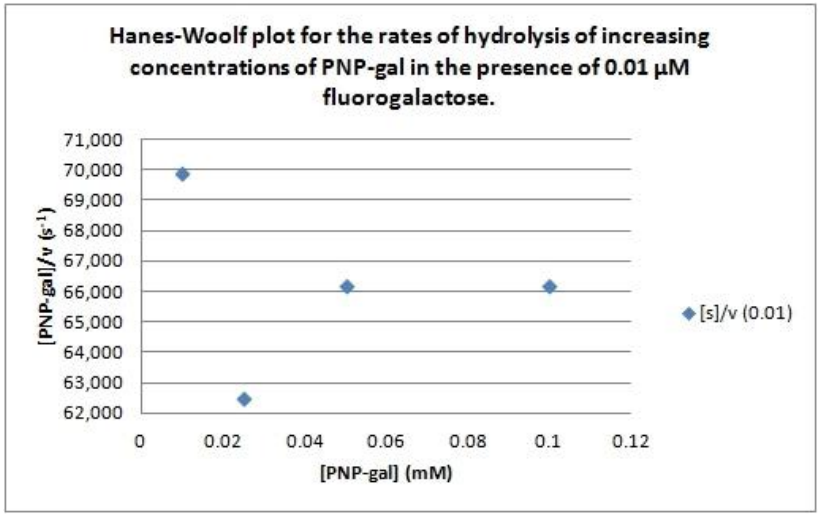

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 15 µM fluorogalactose.
4-nitrophenol concentration (mmol dm-3)
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (15 µM fluorogalactose)
25 µM PNP-gal (15 µM fluorogalactose)
50 µM PNP-gal (15 µM fluorogalactose)
100 µM PNP-gal (15 µM fluorogalactose)

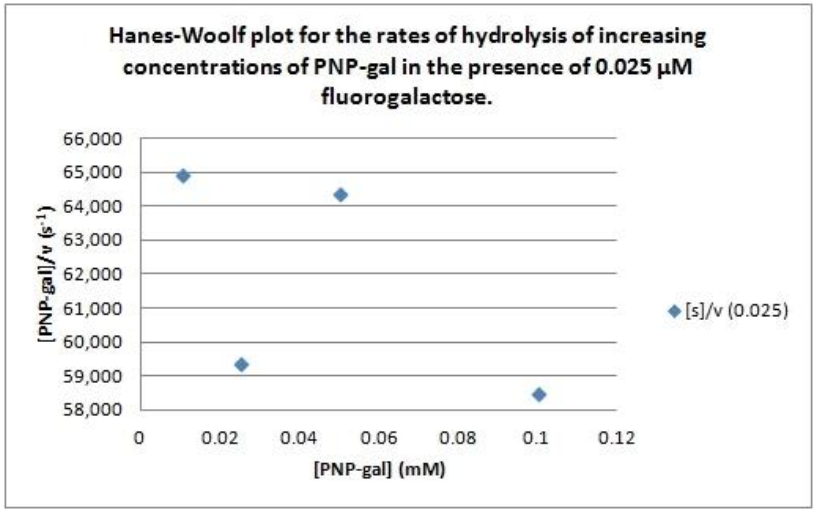

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 20 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (20 µM fluorogalactose)
25 µM PNP-gal (20 µM fluorogalactose)
50 µM PNP-gal (20 µM fluorogalactose)
100 µM PNP-gal (20 µM fluorogalactose)

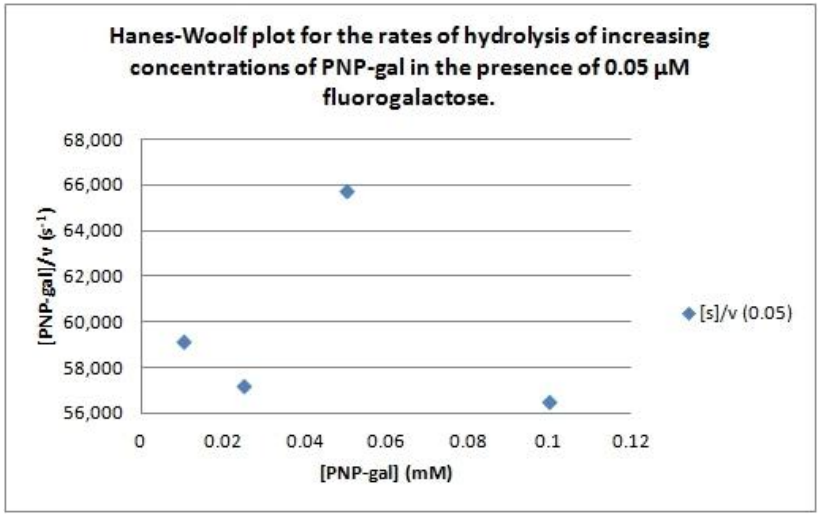

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 25 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (25 µM fluorogalactose)
25 µM PNP-gal (25 µM fluorogalactose)
50 µM PNP-gal (25 µM fluorogalactose)
100 µM PNP-gal (25 µM fluorogalactose)

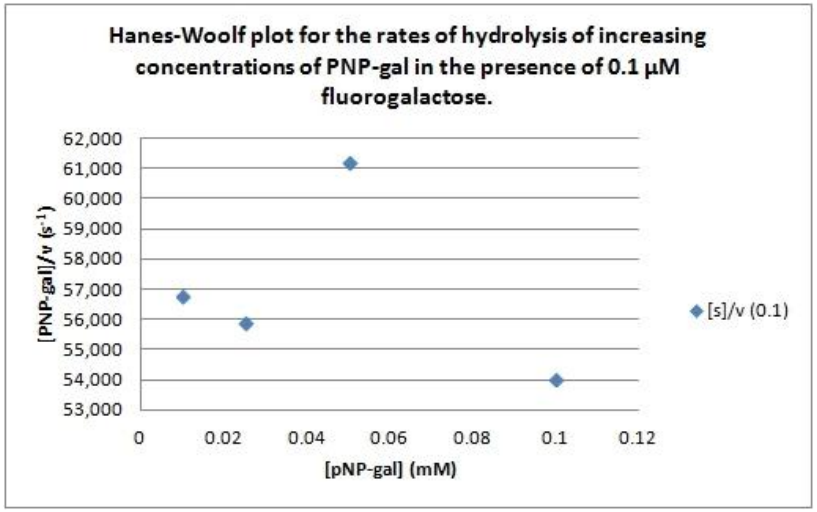

Formation of 4-nitrophenol with time from the hydrolysis of increasing concentrations of PNP-gal by β-galactosidase in the presence of 30 µM fluorogalactose.
4-nitrophenol Concentration (mmol dm-3)
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
2000
4000
6000
Time (s)
10 µM PNP-gal (30 µM fluorogalactose)
25 µM PNP-gal (30 µM fluorogalactose)
50 µM PNP-gal (30 µM fluorogalactose)
100 µM PNP-gal (30 µM fluorogalactose)

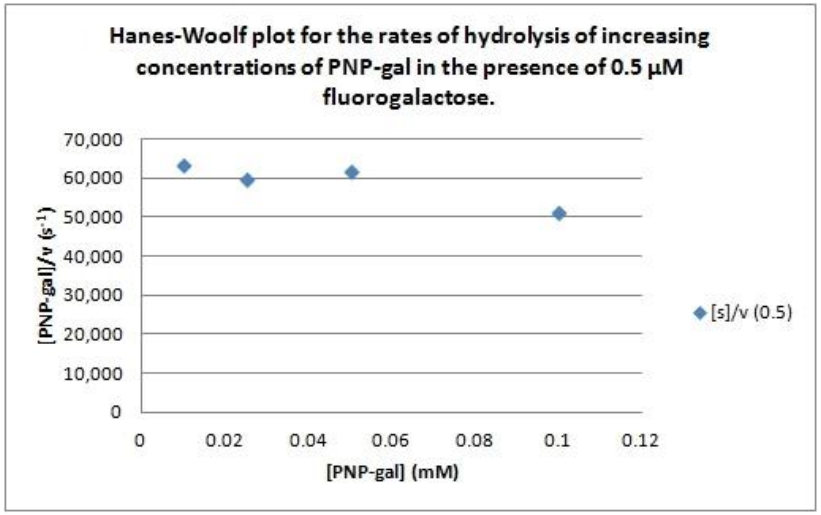

Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 0 fluorogalactose.
[PNP-gal]/v (s-1)
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (0)

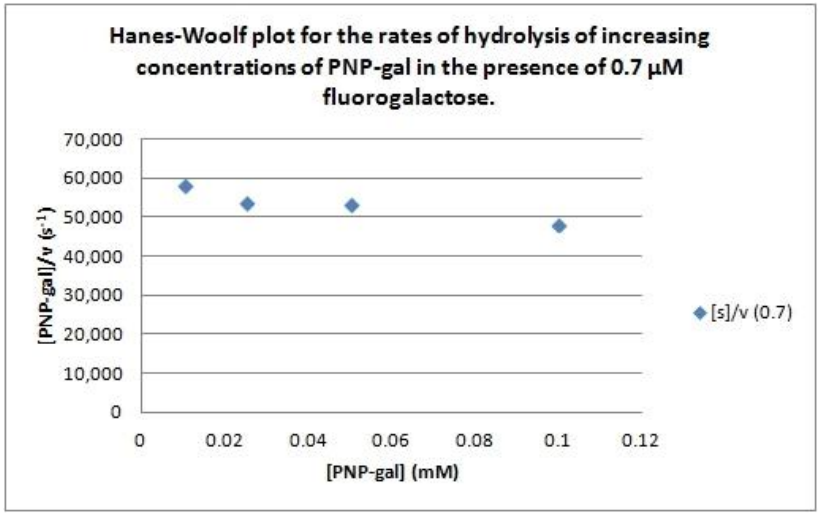

Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 0.01 µM fluorogalactose.
[PNP-gal]/v (s-1)
71,000
70,000
69,000
68,000
67,000
66,000
65,000
64,000
63,000
62,000
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (0.01)

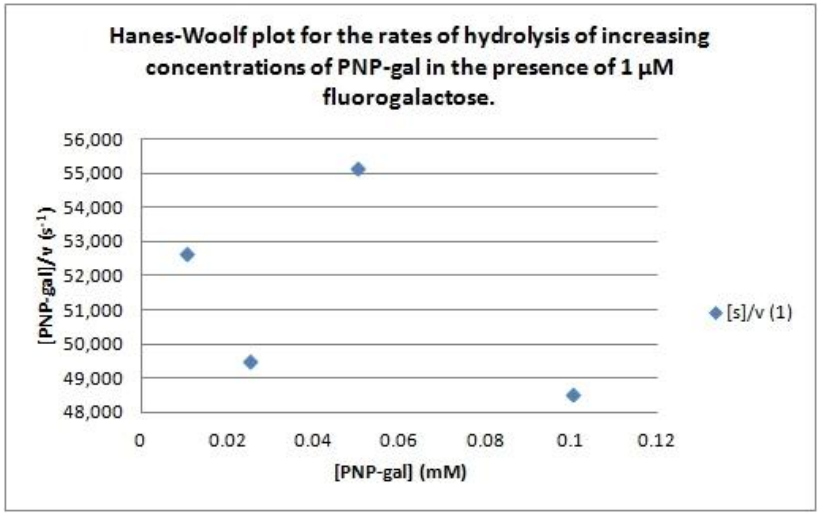

Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 0.025 µM fluorogalactose.
[PNP-gal]/v (s⁻¹)
66,000
65,000
64,000
63,000
62,000
61,000
60,000
59,000
58,000
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (0.025)

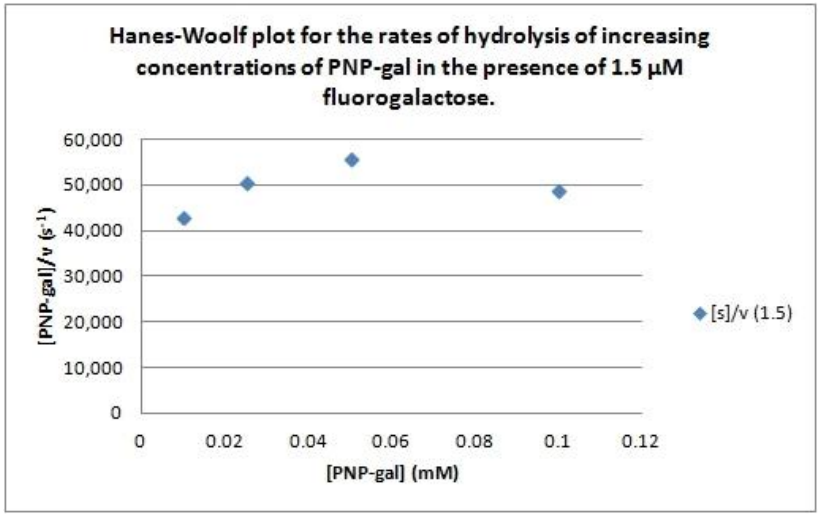

Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 0.05 µM fluorogalactose.
[PNP-gal]/v (s⁻¹)
68,000
66,000
64,000
62,000
60,000
58,000
56,000
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (0.05)

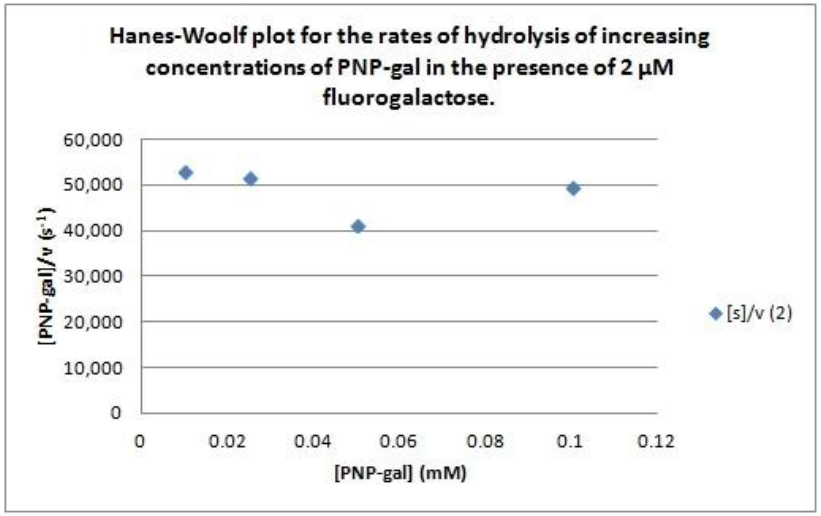

Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 0.1 µM fluorogalactose.
[PNP-gal]/v (s⁻¹)
62,000
61,000
60,000
59,000
58,000
57,000
56,000
55,000
54,000
53,000
0
0.02
0.04
0.06
0.08
0.1
0.12
[pNP-gal] (mM)
[s]/v (0.1)

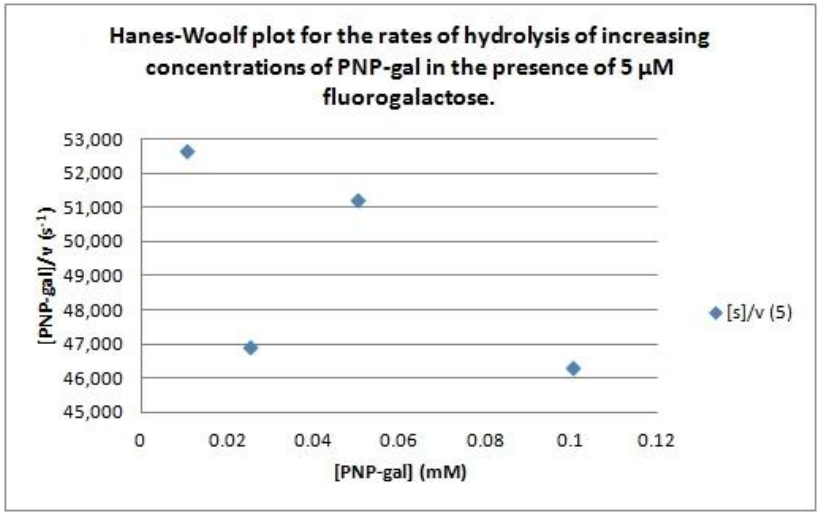

Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 0.5 µM fluorogalactose.
[PNP-gal]/v (s⁻¹)
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (0.5)

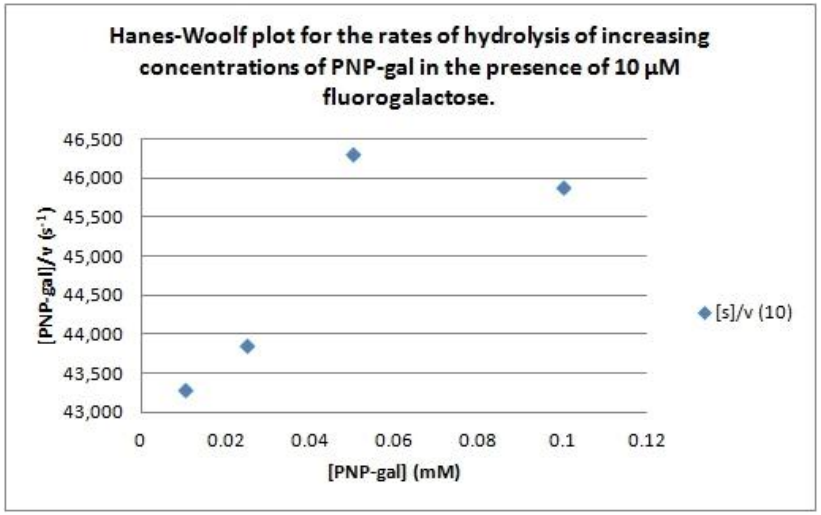

Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 0.7 µM fluorogalactose.
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
[PNP-gal]/v (s-1)
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (0.7)

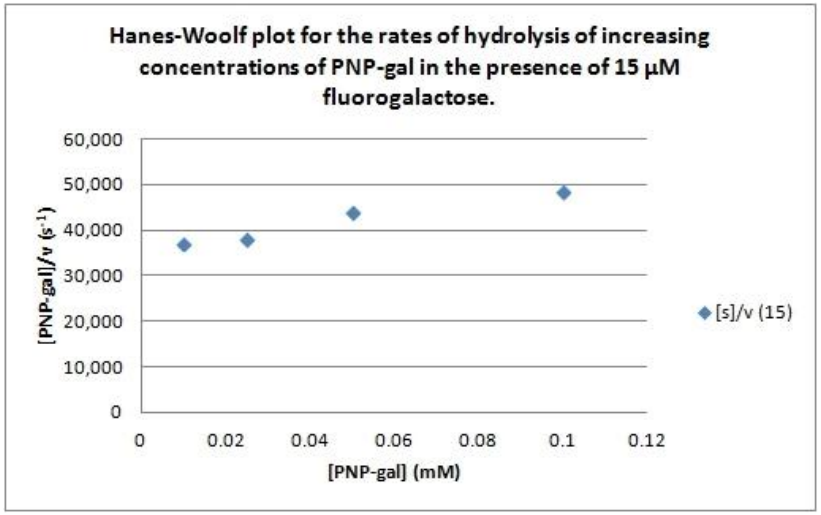

Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 1 µM fluorogalactose.
56,000
55,000
54,000
53,000
52,000
51,000
50,000
49,000
48,000
[PNP-gal]/v (s-1)
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (1)

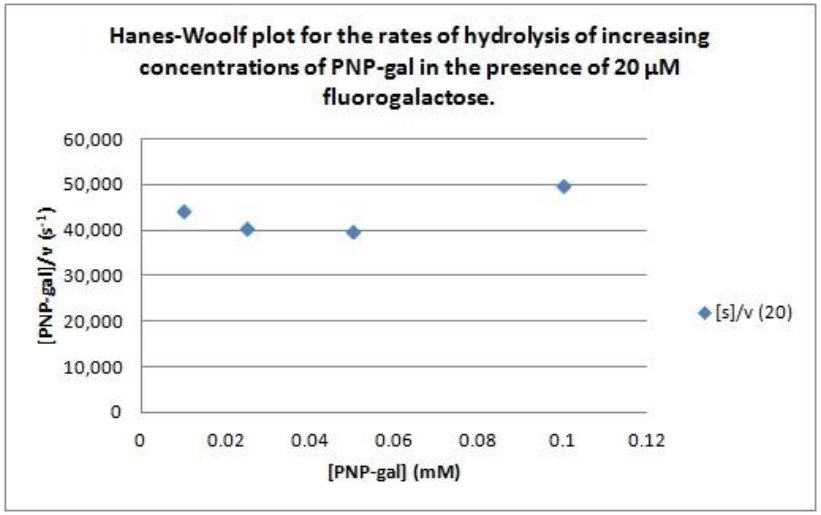

Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 1.5 µM fluorogalactose.
60,000
50,000
40,000
30,000
20,000
10,000
0
[PNP-gal]/v (s-1)
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (1.5)

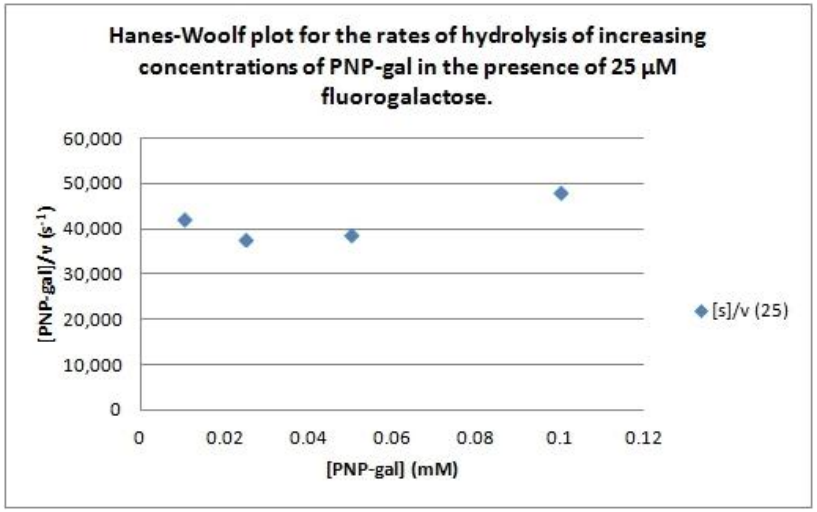

Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 2 µM fluorogalactose.
60,000
50,000
40,000
30,000
20,000
10,000
0
[PNP-gal]/v (s-1)
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (2)

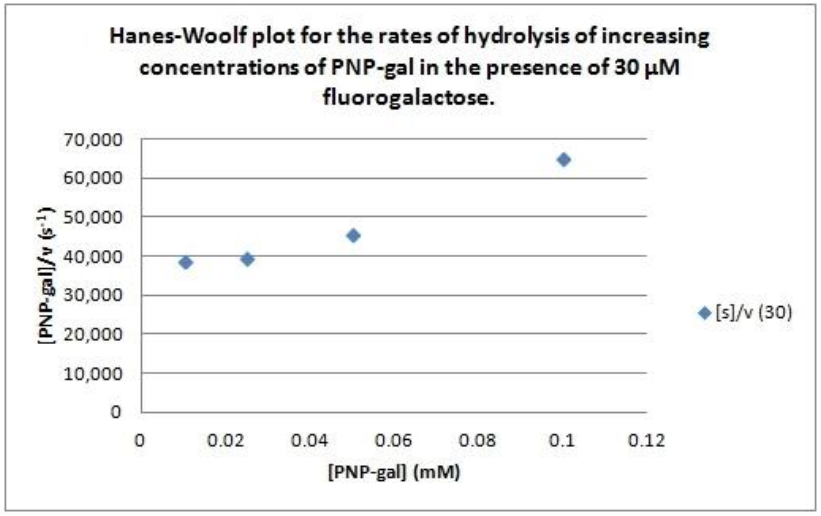

Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 5 µM fluorogalactose.
[PNP-gal]/v (s-1)
53,000
52,000
51,000
50,000
49,000
48,000
47,000
46,000
45,000
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (5)


Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 10 µM fluorogalactose.
[PNP-gal]/v (s-1)
46,500
46,000
45,500
45,000
44,500
44,000
43,500
43,000
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (10)


Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 15 µM fluorogalactose.
[PNP-gal]/v (s-1)
60,000
50,000
40,000
30,000
20,000
10,000
0
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (15)


Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 20 µM fluorogalactose.
[PNP-gal]/v (s-1)
60,000
50,000
40,000
30,000
20,000
10,000
0
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (20)


Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 25 µM fluorogalactose.
60,000
50,000
40,000
30,000
20,000
10,000
0
[PNP-gal]/v (s⁻¹)
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (25)


Hanes-Woolf plot for the rates of hydrolysis of increasing concentrations of PNP-gal in the presence of 30 µM fluorogalactose.
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
[PNP-gal]/v (s⁻¹)
0
0.02
0.04
0.06
0.08
0.1
0.12
[PNP-gal] (mM)
[s]/v (30)